PSL UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à Université Paris-Dauphine

A decision support system adapted to the constraints and the challenges of decision support in customary medical consultations

Soutenue par
Antoine RICHARD
Le 06.04.2021

École doctorale n°543
École Doctorale SDOSE

Spécialité
Informatique

Composition du jury :

Daniela GRIGORI
Full Professor — *President*
Université Paris Dauphine - PSL

Sylvie DESPRES
Full Professor — *Reporter*
Université Sorbonne Paris 13

Pascale ZARATÉ
Full Professor — *Reporter*
Université de Toulouse 1 Capitole

Pavlos DELIAS
Associate Professor — *Examinator*
International Hellenic University

Delphine MAUCORT-BOULCH
Full Professor — *Examinator*
Université Claude-Bernard Lyon 1

Brice MAYAG
Associate Professor — *Supervisor*
Université Paris Dauphine - PSL

Yves MEINARD
CNRS Researcher — *Co-director*
Université Paris Dauphine - PSL

Alexis TSOUKIÀS
CNRS Senior Researcher — *Co-director*
Université Paris Dauphine - PSL

Dauphine UNIVERSITÉ PARIS | PSL

Foreword

The present manuscript is an English version of the official one, written in French, intended for English-speaking readers. This is a digest of the works presented in the official manuscript, gathering the different works published during the thesis. Its structure is the same as the one of the official manuscript. Therefore, publications are distributed into parts and chapters of this English version corresponding to the parts and chapters of the French version which treat the same aspects of this thesis. A translation of the remaining chapters is being provided to present the works that were not included in the publications. Besides, each part of this English version starts with a short introduction explaining the context of each publication presented and the improvements made since their publishing.

I would like to specify that a part of this thesis occured during the pandemic wave of SARS-CoV-2, still ongoing during the writing of this manuscript. This particularly impacted the clinical trials of the decision support system developed during this thesis.

Contents

List of Figures

List of Tables

List of acronyms

Chapter 1

Introduction

Decision support is an activity that consists in helping a decision-maker to improve her/his decisions, through a better understanding of the stakes of the decisions, a more thoughtful examination of the relevant data, and/or a more rigorous utilization of relevant theories and practices. Decision support is usually provided upon demand, but clients requesting decision support are often themselves knowledgeable, at least to some extent, about the topic concerning which they ask decision support.

Moreover, clients requesting decision support are often taken by third parties and/or the general public to be responsible for the decisions they make. In such cases, the task of the decision analyst (or decision support provider) is delicate in the sense that s/he risks infringing upon the expertise and responsibility of the decision-maker.

The case of attempts at providing decision support to physicians in customary consultations is paradigmatic. Physicians are experts in medical matters, and they are responsible for the medical decisions they make. Numerous decision support tools are developed in the literature and in practice in hospitals to provide them with decision support. However, how can one make sure that these tools do not infringe upon physicians' expertise and responsibility?

The works presented in this thesis, funded by Hopsis[1] and developed in collaboration with the employees of the Civil Hospitals of Lyon, aim contribute to answering this question, in the specific case of decision support for customary medical consultations.

1.1 Context

Founded in 1802, the Civil Hospitals of Lyon, or *Hospices Civils de Lyon* (HCL), are a group of 14 hospitals in and around the city of Lyon (cf. Figure 1.1). The HCL includes also 3 computer science departments, or *Directions du Système d'Information et de l'Informatique* (DSII), dedicated to the development and the maintenance of software used by clinicians. They have been developing, since 2015, a software called Easily®, allowing clinicians, among other things, to have quick access to patients' information. Hopsis is an economic interest group, or *Groupement d'Intérêt Économique* (GIE), and it is

[1] www.hopsis.org

dedicated to the funding and the distribution of Easily®. More details about the functioning of Easily®are given in Chapter 2.

Since 2017, Easily®is being used daily by clinicians of the HCL and the DSII strive to develop new software to develop with the aim to support clinicians during their activities. For example, the "O2" family of software was developed to simplify clinical research by constituting a cohort of patients and by analyzing their data. Ducray et al. (2020) give a recent example of the use of such tools.

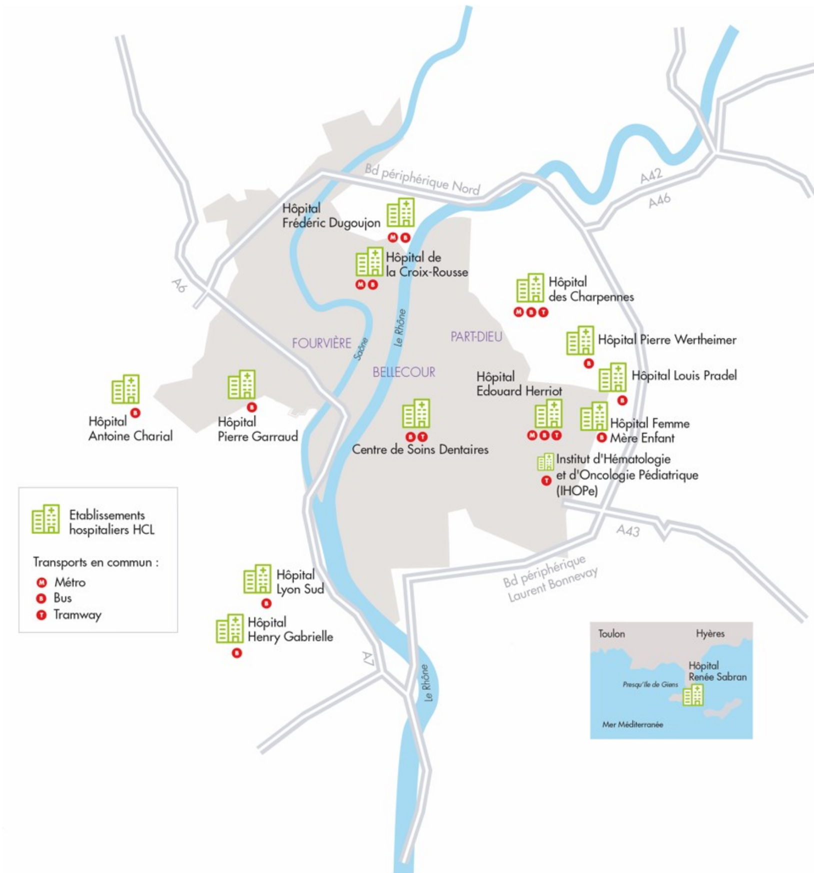


Figure 1.1: The group of 14 hospitals composing the Civils Hospitals of Lyon

Currently, close to 20.000 clinicians work in the HCL, including close to 5.000 physicians. As presented in Figure 1.2, the HCL reported close to 1.250.000 patients coming every year. Medical consultations are composing a large part of physicians' activity with close to one million consultations reported every year.

Medical consultations accordingly constitute a central aspect of patient care processes at the HCL. These consultations are useful for patient follow-up, but also to establish diagnoses used as a basis for patient care processes. During these consultations, physicians make decisions repeatedly and medical errors can occur.

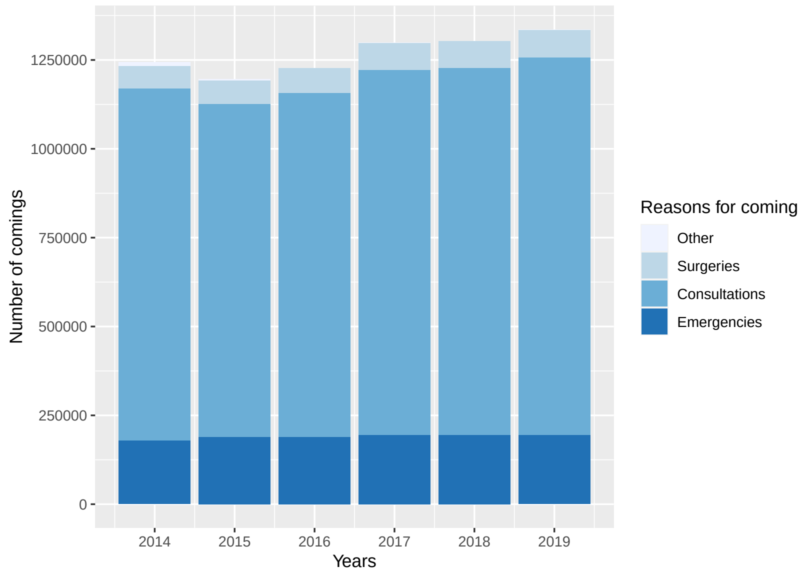


Figure 1.2: Main reasons to come to the HCL from 2014 to 2019

To minimize clinicians' workload, and the associated risk of medical errors, the HCL aim to develop decision support systems. Due to the central aspect of medical consultations in activities of the HCL, we have decided to focus on the support of this type of activity. We aim to propose a decision support system dedicated to providing support to physicians during their customary medical consultations.

1.2 Problematics

As introduced previously, the main objective of this thesis is to propose a decision support system dedicated to medical consultations. This brings us to the central problem treated in this thesis: "How to support physicians during their medical consultation?". Knowing that consultations are customary activities for physicians, for which they can be considered to be competent, trying to answer this question quickly brings other questions, which are just as fundamental.

First, we have to ask ourselves: "Do physicians need any support during their consultations?". Indeed, as mentioned previously, physicians are medical experts and they are competent concerning the decision process of their consultations. Therefore, one might argue that physicians don't need support and that proposing a decision support system is useless in such situations.

This brings about questions on which approaches are currently used to support physicians during their medical activities, customary or not. In other terms, we need to ask ourselves: "What kind of systems are currently provided to physicians to support them during their work?". This question leads to three other questions: "Do decision support systems currently provided to physicians have a beneficial impact on physicians' performance and/or on patient safety?", "Are decision support systems currently provided to physicians accepted by physicians, and are they well integrated into physicians' workflow?" and "Are decision support systems currently provided to physicians adapted to support customary medical consultations?".

At the end, all these questions raise an even more fundamental question: "What does it mean to support a competent and responsible decision-maker such as a physician?".

1.3 Objectives

The thesis that we argue in this manuscript is that an adapted and acceptable decision support system must respect the know-how of physicians and leave them the responsibility of the decisions taken during consultations, by limiting itself to providing them information on their patients which are necessary for their decision-making. The defense of this thesis is based on an analysis of the different aspects of medical consultations and on an analysis of current medical decision-support systems, but also on clinical trials of a new software designed for the occasion. The type of decision support system that we propose to develop, dedicated to providing to physicians targeted pieces of information about their patients, is then based on a better understanding of the constraints and challenges of providing decision support during customary medical consultations. The works presented in this thesis are accordingly organized around three main axes.

In Part I, we analyzed different approaches used to support physicians, but also clinicians in general, during their activities and to understand whether these approaches can be applied in our use case. This part is mainly based on works presented in Richard et al. (2020b). We first proposed a review of the different kinds of clinical decision support systems proposed over the years. Then we analyzed their impact on clinicians' performance and patient safety, before focusing on their acceptability by clinicians in practice. This analysis allowed us to highlight that, despite the potential beneficial impact of such systems, clinicians are still reluctant to use them. We then studied different reasons that could explain this non-acceptance of decision support systems by physicians. We highlighted, with this study, that the current approaches on which clinical decision support systems are based reflect strong choices concerning the aim of decision support tools. These ideological choices are, in numerous aspects, not adapted to the reality of customary medical activities, which may partially explain the non-acceptance of these systems in practice. We proposed then to determine which approach in terms of decision support would be adapted to support customary medical situations. We concluded that an *adjustive* approach, which aims to adapt to the decision-maker's needs and let her/him a maximum of autonomy, seems to be the more adapted given the constraints underlying decision support during customary medical consultations.

In Part II, we focused on physicians' activities at the HCL and we analyzed specificities of the decision process they follow during their medical consultations. We aimed, in this part, to identify the key elements of these decision processes that are potentially costly to physicians and, then, to identify their needs in terms of decision support. This part reproduces and extends works presented in Richard et al. (2018). In Chapter 2, we presented in detail the creation of Easily®and its specificities. This will allow us to have a better understanding of the tools used by physicians during their consultations, but also a better understanding of the context and the objectives that marked the development of these tools. In Chapter 3, we analyzed physicians' activities, using data collected during their consultations, to highlight recurrent work processes. These analyses are based on field observations of medical consultations, but also process mining of physicians' activity logs. In Chapter 4, we proposed models of the physicians' decision process during medical consultations. These models allowed us to highlight the fact that searching for information about their patients is an important part of their decision processes, essential to take decisions about the care of their patients. However, this search for information implies a time-costly workload for physicians.

Based on this observation, we decided to propose a system devoted to learn and anticipate physicians' needs in terms of patients' information, according to the specificities of the patient currently in consultation. In this way, the system will be able to provide, at the beginning of the consultation, targeted information about the current patient and then reduce the workload of physicians.

In Part III, we developed the different aspects of the decision support system we proposed. The first two chapters are mainly based on works presented in Richard et al. (2020a). Given the fact that we use a learning algorithm to learn which set of information to provide to physicians in which situations, this could generate distrust and a non-acceptance of our decision support system. In Chapter 5, we proposed a set of requirements in terms of "transparency" that we imposed on ourselves to minimize the risk of non-acceptance of our system. We proposed then an evaluation of different systems to determine whether they fulfill our requirements. Our choice fell on an adapted version of the well-known Naive Bayes algorithm. Finally, In Chapter 6, we presented concretely the decision support system developed during this thesis and the results of clinical trials we made at the HCL.

Part I

Understanding the implications of supporting physicians during their consultations

To better understand the meaning of supporting physicians during their medical consultations, we have investigated the current approaches used to support physicians, but also the implications of these approaches when used to support physicians in customary situations such as medical consultations. With these investigations, we determined which approach should be used to support physicians during customary consultations. These reflections led to the publication, in the European Journal on Decision Processes, of the following methodological paper (Richard et al., 2020b).

This work also allows us to introduce different terminologies that we will use in this manuscript concerning software used in medical contexts. Let us specify that the model presented as an example in Richard et al. (2020b), is also developed in Chapter 4.

Noname manuscript No.
(will be inserted by the editor)

What does it mean to provide decision support to a responsible and competent expert?

The case of diagnostic decision support systems

Antoine Richard · Brice Mayag · François Talbot · Alexis Tsoukias · Yves Meinard

Received: date / Accepted: date

Abstract Decision support consists in helping a decision-maker to improve his/her decisions. However, clients requesting decision support are often themselves experts and are often taken by third parties and/or the general public to be responsible for the decisions they make. This predicament raises complex challenges for decision analysts, who have to avoid infringing upon the expertise and responsibility of the decision-maker. The case of diagnosis decision support in healthcare contexts is particularly illustrative. To support clinicians in their work and minimize the risk of medical error, various decision support systems have been developed, as part of information systems that are now ubiquitous in healthcare contexts. To develop, in collaboration with the hospitals of Lyon, a diagnostic decision support system for day-to-day customary consultations, we propose in this paper a critical analysis of current approaches to diagnostic decision support, which mainly consist in providing them with guidelines or even full-fledged diagnosis recommendations. We highlight that the use of such decision support systems by physicians raises responsibility issues, but also that it is at odds with the needs and constraints of customary consultations. We argue that the historical choice to favor guidelines or recommendations to physicians implies a very specific vision of what it means to support physicians, and we argue that the flaws of this vision partially ex-

A. Richard (ORCID: 0000-0001-8677-8910)
Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France
GIE Hopsis Lyon, 3 Quai de Célestins, 69229 Lyon, France

B. Mayag
Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France

F. Talbot
DSII Bron, Hospices Civils de Lyon, 61 Boulevard Pinel, 69672 Bron, France

A. Tsoukias (ORCID: 0000-0001-5772-3988)
Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France

Y. Meinard (ORCID: 0000-0001-5928-8959)
Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France
E-mail: yves.meinard@dauphine.lamsade.fr

plain why current diagnostic decision support systems are not accepted by physicians in their application to customary situations. Based on this analysis, we propose that decision support to physicians for customary cases should be deployed in an "adjustive" approach, which consists in providing physicians with the data on patients they need, when they need them, during consultations. The rationale articulated in this article has a more general bearing than clinical decision support and bears lessons for decision support activities in other contexts where decision-makers are competent and responsible experts.

Keywords Decision Analysis · Decision Support Systems · Diagnostic Decision Support Systems

1 Introduction

Decision support is an activity that consists in helping a decision-maker to improve his/her decisions, through a better understanding of the stakes of the decisions, a more thoughtful examination of the relevant data, or/and a more rigorous utilization of relevant theories and practices. Decision support is usually provided upon demand, but clients requesting decision support are often themselves knowledgeable, at least to some extent, about the topic concerning which they ask decision support. Moreover, clients requesting decision support are often taken by third parties and/or the general public to be responsible for the decisions they make. In such cases, the task of the decision analyst (or decision support provider) is delicate in the sense that s/he risks infringing upon the expertise and responsibility of the decision-maker. The case of attempts at providing decision support to physicians in customary consultations is paradigmatic. Physicians are experts in medical matters and they are responsible for the medical decisions they make, but numerous decision support tools are developed in the literature and in practice in hospitals to provide them with decision support. How can one make sure that these tools do not infringe upon physicians' expertise and responsibility? In this article, we set out to answer this question, based on a literature review and a critical methodological analysis of medical decision support approaches.

The decision support systems that we are about to analyze here are part of the larger set of information systems in healthcare environments, more commonly called Health Information Systems (HISs). HISs have been developed in the last decades mainly to support and improve healthcare processes, decisions, and outcomes of patients. Nowadays HISs are ubiquitous in hospitals and it is difficult to find a hospital without an information system. One can distinguish, among HISs, different kinds of systems dedicated to healthcare support. According to Shortliffe and Cimino (2014)'s review of computer applications in healthcare, one of the first systems developed in healthcare environments corresponded to systems allowing the recording of healthcare information. These are Electronic Health Records (EHRs), including databases, indexing systems, and research systems using healthcare information. With a similar objective, Computer Physician Order Entry (CPOE) (Kuperman and Gibson, 2003), are systems developed to digitize physician's orders.

Another subset of HISs is composed of Clinical Decision Support Systems (CDSSs) (Musen et al., 2014; Berner, 2016). CDSSs include all kinds of tools designed to transmit information to clinicians to help them to make decisions or simply to facilitate their daily processes. The main objective of CDSSs is to minimize the risk of medical errors. CDSSs themselves include a variety of systems. Alert Systems provide alert messages to clinicians when an emergency occurs, e.g. when a hospitalized patient undergoes a heart attack. Alert Systems are also integrated into some CPOEs to prevent mistakes in drug prescriptions and/or drug dosages (Van Der Sijs et al., 2006). Reminder Systems (Garg et al., 2005) are likewise developed to avoid omission errors.

Lastly, Diagnostic Decision Support Systems (or DDSSs) are a subset of CDSSs dedicated to providing support to physicians in their clinical diagnosis. These systems will be our main topic in the present article. According to a recent systematic survey of DDSSs (Yanase and Triantaphyllou, 2019), there are currently two main types of DDSSs:

- DDSSs based on "gold standard" rules or guidelines defined by experts of the domain or health authorities (thereafter: "Guideline-based DDSSs").
  Clinical practice guidelines, including diagnostic guidelines, are lists of instructions to follow in a specific situation. They are generally based on current best practices and can be represented by a flowchart. Fig. 1 shows an example of a flowchart from the MIMS website[1] and based on the guidelines for diabetes treatments produced by the UK's National Institute for Health and Care Excellence (NICE)[2]. Other examples of clinical guidelines can be found on the NICE website[3], on the website of the French "Haute Autorité de Santé" (HAS) [4] or in reports of International Classification of Diseases (World Health Organization et al., 1992).
  Guideline-based DDSSs encompass "expert systems", which integrate "gold-standard" flowcharts/rules into their process to produce full-fledged diagnosis recommendations to physicians (Yanase and Triantaphyllou, 2019), but also systems that prescribe to physicians the steps they should follow to abide by the "gold-standard" (this is the case, for example, of the systems found on the NICE website or the Quick Medical Reference (QMR) linked to the INTERNIST expert system (Miller et al., 1986; Miller, 2010)).
- DDSSs based on Machine Learning (ML) algorithms, or ML-based DDSSs, are used to support diagnoses of specific diseases, with the aim to minimize error rates by treating large amounts of data on patients(Dua et al., 2014; Yanase and Triantaphyllou, 2019).
  ML algorithms are methods used to learn how to approximate a classification function based on a learning dataset. Classification functions could be, for example, functions anticipating the value of an exogenous variable y depending of the value of an endogenous variable x, or functions distinguishing pictures of healthy from pictures of diseased organs by analyzing a matrix of pixels. ML problems are generally divided into three subclasses, depending of the degree of knowledge included in the learning dataset: supervised learning (full knowledge), semi-supervised learning (some pieces of information are not available) and unsupervised learning (no predefined class).
  Many ML algorithms have been proposed to handle these classification problems, from Naive Bayes algorithms to Artificial Neural Networks and Support Vector Machine algorithms. In this paper, we used the term "ML-based DDSSs" to refer to all the DDSSs using one of these ML algorithms.

[1] www.mims.co
[2] www.nice.org.uk/guidance/ng28
[3] www.nice.org.uk
[4] www.has-sante.fr

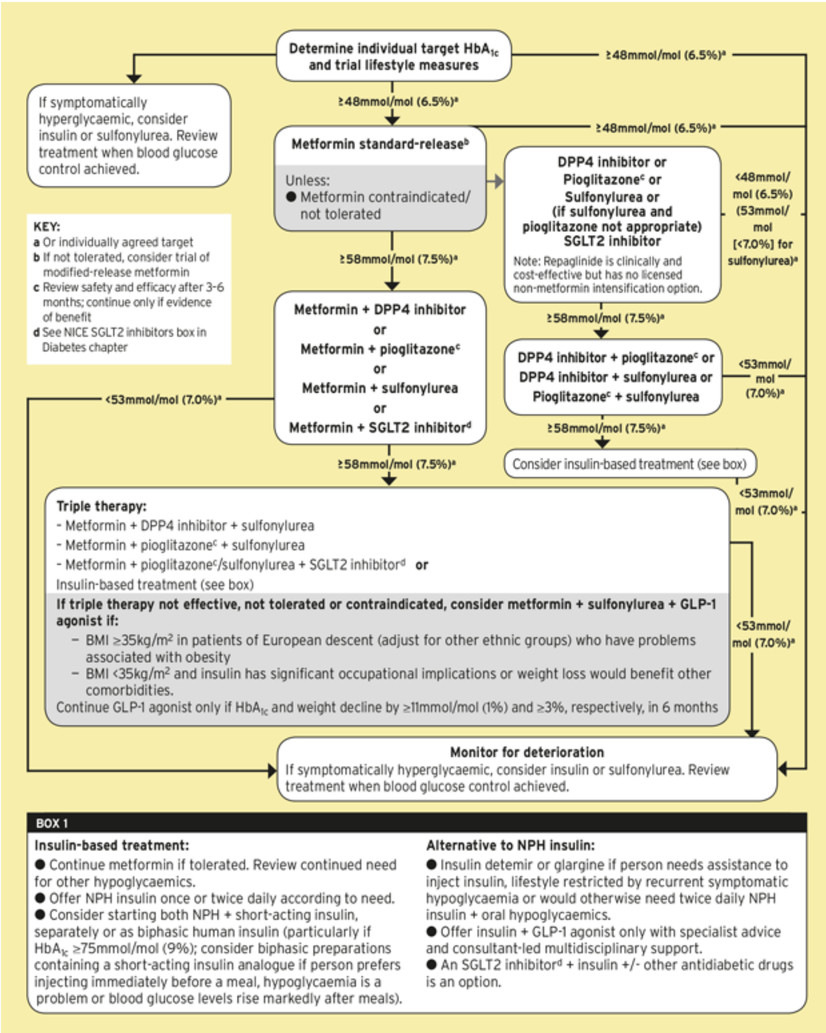


Fig. 1 Summary of NICE's guidance on treatment of type 2 diabetes proposed by the MIMS website[1]

As we will see in this paper, in their application to support customary diagnostic decisions, these DDSSs are currently in a paradoxical situation. On the one hand, their potential usefulness appears unquestionable, but on the other hand, they are generally poorly accepted by physicians. In addition, the use of DDSSs raises responsibility issues and involves patient safety risks. This paradoxical situation reflects, in our view, the more general difficulty to provide decision support to a competent, responsible decision-maker. By analyzing the specific case of DDSSs for customary consultations in detail, we aim to develop a new approach to address this general difficulty. To that end, we analyze here the reasons underlying the current failure of DDSS, and we draw the constructive lessons from this analysis.

By tackling this issue, this article aims to contribute to a broader research program devoted to analyzing the challenges facing decision support approaches and methodologies, as developed mainly in decision sciences and operational research, when they are applied to decisions involved in the design, implementation, and evaluation of public policies (Tsoukiàs et al., 2013; De Marchi et al., 2016). This research program has already produced applications to the evaluation of environmental policies (Jeanmougin et al., 2017), the design of policy options (Ferretti et al., 2019; Pluchinotta et al., 2018, 2019), the development of methodological tools for large scale environmental policies (Choulak et al., 2019), among others. In the wake of these contributions, we endorse the methodological and epistemological approach clarified in Tsoukiàs et al. (2013); Meinard and Tsoukiàs (2019); Meinard and Cailloux (2020)

Our reasoning unfolds in three steps. In section 2, we begin by reviewing historical choices that led to the current development policy of DDSSs and past experiences in the elaboration of DDSSs. Section 3 explores the adverse impact of HISs, CDSSs, and DDSSs, responsibility issues raised by the use of DDSSs, as well as gaps between DDSSs' design and the reality of customary consultations, to highlight potential reasons behind the failure of DDSSs in these situations. Section 4 discusses the conceptual approaches underlying the current DDSSs and sets out to determine which approach should be favored in the case of customary consultations. Section 5 briefly concludes the paper.

2 The paradoxical situation of Diagnostic Decision Support Systems

As introduced in section 1, HISs, such as EHRs and CPOEs, are now ubiquitous in hospitals. Due to this computerization of hospitals, works on CDSSs and DDSSs to support clinicians in their daily practices are on the rise. In this section, we develop a brief historical review of DDSSs and of the impact of the use of CDSSs in practice.

Our analysis is buttressed on a bibliographic review of the systems that have been developed to support physicians during consultations. In order to strengthen the purview of our analysis, we complemented this search by exploring the literature on support systems for clinicians in general. This analysis aims to capture the variety of systems that have been or can be used in practice to support physicians during customary consultations.

We made our research on PubMed with the following request: ("decision support system" or "computer-aided" or "artificial intelligence" or "machine learning" or "expert system") and "consultation". 393 articles were found with this request. This set of articles was to a large extent redundant for our purposes because it contained reviews and meta-analyses of DDSSs, their impact, and their acceptability, which synthesized the relevant information contained in other articles of this initial set. We, therefore, selected these reviews and meta-analyses. Because some fairly recent DDSSs might have been ignored in these reviews and meta-analyses, we also kept papers presenting specific DDSSs dedicated to physicians and published after 2017. We also keept papers including studies of the impact or acceptability of specific DDSS, because of the central role that these notions play in our study.

Applying these criteria to the content of titles and abstracts allowed filtering out 290 articles. Applying these criteria to the full content of the remaining papers then led to selecting 49 articles, including 12 (25%) reviews of systems used in general or in specific healthcare contexts, 27 (55%) papers presenting specific decision support systems (20 including clinical trials or feasibility studies), 10 (20%) studies of the impact of information systems on the performances of physicians, on patient safety or on the acceptability of systems.

2.1 A loss of confidence in physicians' diagnostic skills and know-how

According to Fieschi (1986)'s and Miller (1994)'s overviews of works on DDSSs from 1954 to 1993, early DDSSs were developed to try to reproduce, using computers, the behavior of physicians making a diagnosis. During these early stages of the history of DDSSs, from the 1950s to the late 1970s, most studies were devoted to representing physicians' behavior and possible uses of DDSSs in information systems. According to Miller (1994), the first studies devoted to developing information systems for diagnosis decision support purposes date back to the 1970s. In the early 1980s, the development of DDSSs in differents medical contexts, such as psychiatry (Morelli et al., 1987) or medical consultations (Kulikowski, 1988), was motivated by the development and proliferation of microcomputers, but also by innovations in user interfaces and networks systems. INTERNIST-1, developed by Miller et al. (1986), is an example of DDSSs developed during this period. These systems were generally designed to ask questions to physicians about the symptoms of patients in order to provide diagnostic suggestions to physicians. This "Greek Oracle" model of DDSSs, based on the idea that DDSSs are "magical tools" providing recommendations that physicians must follow, begun to be deprecated in the late 1980s (Miller and Masarie Jr, 1990). In the early 1990s, most studies on DDSSs had switched for explorations of AI methods such as neural networks or fuzzy logic systems, proposing new approaches for diagnosis decision support (Miller, 1994).

In 2000, the Institute of Medicine published the report *To Err Is Human: Building a Safer Health System*. This report, written by Donaldson et al. (2000), was a survey of multiple studies about medical errors, concluding that between 44000 and 98000 people die each year due to preventable medical errors. For comparison, Mokdad et al. (2004), who studied the causes of death in the U.S. in 2000, have reported an estimation of 43000 deaths due to motor vehicle crashes, 75000 deaths due to microbial agents and 29000 deaths due to incidents involving firearms. The *To Err Is Human* report pushed patient safety to the top of the agenda for governments and national healthcare policies.

According to Reider (2016), numerous national policies during the 2000s then set out to improve clinical practice guidelines, to improve education on patient safety, and to develop CDSSs. Other studies on medical and diagnosis errors, such as Leape (2000) and Berner and Graber (2008), bolstered governments in their efforts in this direction. Thereafter, many healthcare information systems were then developed to prevent potential medical errors. This is the case, in particular, of reminders, alert systems, and Guideline-based DDSSs. According to Miller (2016), at that time works on DDSSs also increasingly aimed at supporting physicians' diagnoses by giving them diagnostic recommendations, partly reinstating the deprecated "Greek Oracle" model of DDSSs.

Recent examples of Guideline-based DDSSs that were developed in this dynamic can be found in eIMCI (Bessat et al., 2019) and the ALMANACH project (Bernasconi et al., 2019), both dedicated to improving child health in primary care in developing countries by providing to physicians suggestions of diagnoses or actions according to guidelines of the Integrated Management of Childhood Illness (IMCI). The CHICA system (Anand et al., 2004) of Wishard Memorial Hospital in Indianapolis is another example dedicated to supporting child health in primary care, by generating forms based on patient's data and national guidelines. These forms are used to collect data on patients or to remind physicians of specific actions to do during consultations. Other applications of the CHICA system were developed, for the prevention of maternal depression (Carroll et al., 2013), prevention of suicidal behavior of adolescents (Etter et al., 2018), or prevention of obstructive sleep apnea (Honaker et al., 2018). López et al. (2017) presented a DDSS, called ophtalDSS, dedicated to supporting physicians in primary care to determine ocular diseases. OphtalDSS is based on decision trees and, once an ocular disease is confirmed, it provides adapted national guidelines. Kirby et al. (2018) proposed a DDSS based on guidelines of the American College of Cardiology/American Heart Association (ACC/AHA) to alert physicians when a patient meets criteria for severe aortic diseases and provide them with recommendations of the ACC/AHA. Similarly, Yang et al. (2018) proposed reminder systems, based on patients' allergy background, to prevent hypersensitivity reactions to radiocontrast media, according to the Korean health policy. Gonzalvo et al. (2017) proposed a DDSS based on the CONSORT guidelines, providing treatment recommendations for poly-medicated patients. Another well-known example is DxPlain (Barnett et al., 1987; Hoffer et al., 2005), an early DDSS dedicated to providing recommendations for primary care, which is still available[5].

A recent development in this history relates to the fact that, due to the generalized expansion of HISs, increasing volumes of data about patients are being recorded, flooding physicians under data (Pivovarov and Elhadad, 2015). This large amount of data quickly proved to be too difficult to analyze by human brains (Yanase and Triantaphyllou, 2019). Data mining and machine learning algorithms are better suited to this task, thanks to their distinctive efficiency when it comes to treating large amounts of data, unveiling correlations, approximating risk functions, or solving classification problems. According to Dua et al. (2014); Ozaydin et al. (2016); Miotto et al. (2017); Kulikowski (2019); Yanase and Triantaphyllou (2019), who surveyed ML-based DDSSs, many studies in the 2000s and the 2010s accordingly focussed on diagnoses assisted by machine learning algorithms.

Currently, ML-based DDSSs are being developed for numerous clinical situations. For example, Deig et al. (2019) surveyed ML-based DDSSs used in Radiation Oncology, mainly to assess the risk of bad reactions to treatments based on data on patients, allowing them to adapt treatments to improve outcomes. Peiffer-Smadja et al. (2019) reviewed ML-based DDSSs dedicated to supporting physicians for cases of infectious diseases by providing diagnoses/treatment recommendations, early detection of diseases, or predictions of responses to treatments. Gordon et al. (2018) surveyed the use of ML algorithms to support physicians in genetics, mainly in their analyzes of genetic risk, but also to recommend diagnoses to physicians. De Fauw et al. (2018); Zhang et al. (2018) recently proposed ML-based DDSSs to detect ocular diseases by analyzing retina images. Pearce et al. (2019) proposed a ML-based DDSS to evaluate the risk of emergency for a patient at the time of consultation. Titano et al. (2018) proposed a ML-based DDSS dedicated to anticipating neurological events by analyzing cranial radiographs. Numerous ML-based DDSSs are also dedicated to supporting the detection of tumors, such as breast tumors (Joo et al., 2004), brain tumors (Hollon et al., 2018), or skin tumors (Esteva et al., 2017). Elsner et al. (2018) and Pasquali

[5] `http://www.mghlcs.org/projects/dxplain`

et al. (2020) also reported the use of ML-based DDSSs in teledermatology to support physicians in the detection of skin tumors during teleconsultations. In a review of information systems, Kataria and Ravindran (2018) reported the use of ML-based DDSSs to anticipate responses to treatments or predict the propagation of diseases. In these examples, ML-based DDSSs appear to play the role of extensions of physicians, doing tasks that human physicians cannot perform with the same accuracy.

Based on this brief history of DDSSs, it appears that the *To Err Is Human* report, and the following works, have highlighted the limitations of physicians' diagnostic skills. In response, health authorities have financed the development of "gold-standard" guidelines and Guideline-based DDSSs dedicated to improving physicians' adherence to these guidelines. Early works on DDSSs, from the 1950s to the late 2000s, were mainly focused on these Guideline-based DDSSs. More recently, it appeared that machine-learning algorithms can be more performant than physicians for certain tasks (e.g., identify microscopic melanoma on images). This prompted the development of works on ML-based DDSSs in the last decades. Although works on Guideline-based DDSSs are still being developed, ML-based DDSSs started to dominate the field from the 2010s onwards[6]. In the subsections to come, we investigate whether these tools fulfill their promises by asking the following questions: Is the support provided by Guideline-based or ML-based DDSSs efficient in terms of patient safety? Are Guideline-based and ML-based DDSSs accepted by physicians and patients? Are current approaches of Guideline-based and ML-based DDSSs legitimately applicable to cases in which physicians can be considered to be "competent" and responsible for outcomes of patients?

2.2 Evidence that HISs, CDSSs, and DDSSs are potentially beneficial

In this sub-section, we start by reviewing studies of the impact of HISs on physicians' performances and patient safety, before zooming in on CDSSs and then on DDSSs. Patel et al. (2000) studied the impact of HISs, more specifically of the representation of knowledge in EHRs, not on clinicians' performances but on clinicians' reasoning and behaviors. They showed that a simple computer-based patient record system can have an important impact on physicians' behavior and working processes. In particular, they showed a standardization, through time, of physicians' working processes converging towards the EHR organization. Chaudhry et al. (2006) made a systematic review of the impacts of HISs on quality, efficiency, and costs of medical care, based on 257

[6] Some exceptions exist to the two predominant subsets of DDSSs (Guideline-based and ML-based DDSSs). Gräßer et al. (2017) proposed a DDSS dedicated to providing therapy recommendations, based not on expert guidelines or machine learning algorithms, but on similarity measures between the current case and previous ones, computed for each new cases, without any learning process involved. Whereas this system is akin to ML-based DDSSs, it does not use ML algorithms. Similarly, Giordanengo et al. (2019) proposed a DDSS dedicated to presenting self-collected data on patients and reminders of actions to do to physicians during the consultations of patients with diabetes. In this work, Giordanengo et al. (2019) didn't use the guidelines of any health authority but included physicians in the development process of the DDSS to establish rules to apply in specific situations. In addition, the recommendations established by consensus among the physicians involved are not intended for other physicians, but to developers adding needed features into the DDSS. Lastly, the ML-based DDSS proposed by Simon et al. (2019) does not use ML algorithms to make recommendations but to detect complex concepts in medical documents, facilitating access to information on patients or to reference documents. With this DDSS, Simon et al. (2019) showed that it is possible to use ML algorithms in other ways than by producing recommendations, while still providing support to physicians in practice.

studies. They concluded on the potentially beneficial impact of HISs on clinicians' performances. According to Leape and Berwick (2005) and Wachter (2004), who studied improvements in patient safety five years after *To Err is Human*, but also according to Clancy (2009), who proposed a similar analysis ten years after *To Err is Human*, the first impact of the Institute of Medicine report was the automation of medical error recording. With the introduction of HISs in hospitals, recording medical acts and results became more regulated. In addition, the development of reminders and alert systems helped to reduce potential mistakes. No doubt that such impacts of the ever-increasingly omnipresent HISs on physicians' work and on some aspects of patient outcomes, while not demonstrated before the 2000s, were to some extent perceived by physicians, medical authorities and the general public early on. In this context, the lost confidence epitomized by the *To Err Is Human* report provided a historical opportunity for CDSSs to entrench their usefulness.

Anticipating the call for diagnostic decision support of the *To Err Is Human* report, Johnston et al. (1994) have studied 28 controlled trials of different kinds of CDSSs (computer-assisted dosing, DDSSs, preventive care reminder, and computer-aided quality assurance, etc.) to assess the impact of CDSSs on clinicians' performances. Clinicians' good performances are, in this study, defined as low error rates in drug dosage and diagnosis, but also as the respect of guidelines by clinicians. Based on the few studies they found, Johnston et al. (1994) reported that some CDSSs (especially drug dosage recommendation systems) seem to have a beneficial impact on clinicians' performance. Hunt et al. (1998) similarly studied the effects of CDSSs on physician performances through a systematic review of 68 controlled trials, updated by Garg et al. (2005) with 97 controlled trials. They concluded that many CDSSs can improve clinicians' performances. Kaushal et al. (2003) studied the effects of CDSSs, and more specifically of CPOEs, on medication safety. They showed a potential reduction in the rate of medication errors, due to the use of CDSSs. Slain et al. (2014) analyzed retrospectively one year of use of a CDSS dedicated to supporting nurses in an emergency department. The CDSS was integrated into the workflow of the emergency department and proposed the pre-screening of patients at their arrival. The authors reported a higher triage accuracy and a better transfer of information thanks to the use of the CDSS. Zier et al. (2017) analyzed the use of a CDSS for one year in comparison with three years without CDSS. This CDSS was dedicated to supporting organ donation by early detection of brain death. The authors mentioned an improvement in early detection of brain death and organ donation. There is an exception: Verdoorn et al. (2018), who studied one year of use of a Guideline-based CDSS dedicated to preventing drug-related problems, reported lower performances with the CDSS than without. The authors pointed out the need for improvements of the CDSS.

In the more specific case of DDSS, although there are exceptions, such as Eccles et al. (2002) and Poels et al. (2008), who analyzed controlled trials of Guideline-based DDSSs with scenarios based on customary situations for physicians and reported that DDSSs have no significant impact (either negative or positive, on physicians' performances or workflow), a majority of studies shows a beneficial impact on physicians performances. Heckerling et al. (1991); Chang et al. (1996); Murphy et al. (1996), and Elstein et al. (1996), who made controlled trials on the Iliad expert system (Warner et al., 1988; Warner Jr, 1989), showed that expert systems can improve physicians' diagnosis accuracy in complex cases, in particular in the case of students (Murphy et al., 1996). Taylor et al. (2008) made controlled trials of a Guideline-based DDSS dedicated to supporting physicians in asthma cases. They showed that the DDSS helped physicians to improve their decision process and to decrease the duration of consultations. Watrous et al. (2008) made controlled trials to evaluate the impact of a Guideline-based DDSS dedicated to supporting the detection of heart murmurs during auscultation. They showed an improvement in the sensitivity and specificity of physicians using the DDSS in the classification of murmurs. Carroll et al. (2013) proposed a clinical trial of their Guideline-based DDSS dedicated to supporting the prevention of maternal depression by alerting physicians when a patient meets some criteria. According to the authors, their DDSS showed a potential beneficial impact on patient safety. Kostopoulou et al. (2017) made a controlled trial of a Guideline-based DDSS dedicated to supporting general practitioners by providing a list of potential diagnoses according to data on patients. They showed an improvement in diagnostic accuracy with the DDSS. The authors also mentioned that physicians entered more data on patients when they used the DDSS. Kirby et al. (2018) analyzed the use of a Guideline-based DDSS, dedicated to supporting the prevention of aortic diseases, during one year in 13 hospitals. They showed that their DDSS improved physicians' accuracy but also the clinical outcomes of patients.

Concerning the use of DDSSs in developing countries, Dalaba et al. (2014) studied one year of implementation of a Guideline-based DDSS for child health in healthcare centers in Ghana. The authors reported a decrease in complications and a diminution of deaths after the introduction of the DDSS. Bessat et al. (2019) made clinical trials of a DDSS dedicated to supporting child health in primary care facilities in Burkina Faso. The authors reported improvements in patient safety due to the DDSS. Similarly, Bernasconi et al. (2019) analyzed the impact of the introduction of Guideline-based DDSSs dedicated to child health in hospitals in developing countries. Clinical trials showed that DDSSs improved physicians' accuracy in primary care.

Concerning ML-based DDSSs, even though they are individually able to outperform physicians during sensitivity and specificity tests (Esteva et al., 2017), their impact when used in clinical practices remains understudied (Yanase and Triantaphyllou, 2019). For example, according to Peiffer-Smadja et al. (2019), who reviewed ML-based DDSSs dedicated to infectious diseases, among 60 ML-based DDSSs only three included clinical trials. The feasibility study by Jaroszewski et al. (2019) on a ML-based DDSS dedicated to mental illness prevention showed good results in mental crisis detection. Currently, because ML-based DDSS is still an emerging domain, it remains hazardous to determine if ML-based DDSSs can improve patient safety or beneficially modify physicians' workflow.

To summarize, clinical trials of DDSSs showed a theoretically beneficial impact on physicians' performances and on patient safety. Guideline-based DDSSs are quite performant when used in primary care or in developing countries, situations where "gold-standard" guidelines for specific cases are welcomed. When it comes to ML-based DDSSs, they are currently mainly evaluated on their specificity/sensibility or precision/recall performances (Yanase and Triantaphyllou, 2019). It remains delicate to determine whether current ML-based DDSSs provide physicians with helpful support in practice.

2.3 A questionable acceptability

According to Shortliffe and Cimino (2014), the most ubiquitous tools are Alert Systems, Schedulers, and Electronic Health Records (EHRs). Studying the introduction of an alert system dedicated to HIV prevention, Chadwick et al. (2017) showed that despite "alert fatigue", alert systems are generally accepted by clinicians. However, introducing information systems in clinical contexts remains a difficult task. According to Heeks et al. (1999), who surveyed the potential causes of successes or failures of HISs, even though some HIS succeed, many of them fail. Keen (1994) studied information systems in healthcare contexts and concluded that for every documented success, there are myriads of failures. Pare and Elam (1998), who worked on the introduction of information systems in clinical contexts, argued that many health care institutions have consumed large amounts of money and frustrated countless people in wasted efforts to implement information systems.

Heeks et al. (1999) and Heeks (2006) surveyed different cases of successful or failed HISs' introduction in hospitals. An illustrative example they explored is Beynon-Davies and Lloyd-Williams (1998)'s study of the failure of the introduction of a computer-aided despatch system for the London ambulance service. In this case, failure arose because "*the speed and depth of change were simply too aggressive for the circumstances*". The cancellation of this system caused an estimated waste of £20 million (ca. US$33 million). Another telling example was Guah (1998)'s analysis of the introduction of an expert system for computerized coloscopy in the coloscopy unit of a university hospital, in the UK. This system produced non-significant statistical information for physicians and needed to learn new work processes. The tool was therefore abandoned.

Sittig et al. (2006) studied the factors influencing the acceptability of DDSSs. They reported that a high percentage of CDSS's guidelines and/or recommendations were overridden, or ignored, by physicians. According to Overhage et al. (1997); Tierney et al. (2003) and Weingart et al. (2003), the percentage of DDSSs recommendations overridden by physicians varies between 54% and 91%. Sittig et al. (2006) also reported that physicians were more willing to accept clinical decision support for elderly patients with multiple medications or chronic conditions. Onega et al. (2010), studied the acceptability of DDSSs by radiologists, in comparison with a double reading by another radiologist. The authors surveyed 257 radiologists from different hospitals across the USA. According to their results, the radiologists were more favorable to double reading, even though most of them perceived that DDSSs were better at improving recall rates than double reading. The meta-analysis proposed by Masud et al. (2019), on the use of DDSSs in radiology departments, showed similar results on the low acceptability of DDSSs despite an improvement of performances perceived by radiologists.

Only a handful of studies showed a good acceptability of Guideline-based DDSSs, in very specific situations. This is the case of Porat et al. (2017), who analyzed the acceptability by patients and physicians of a Guideline-based DDSS. 34 general practitioners participated in the study by consulting 12 standardized patients during controlled trials. The authors reported that 74% of GPs found the DDSS useful, even though the use of the DDSS required them to enter more data on patients while interacting with them. Developing countries also constitute a specific case in which guideline-based DDSSs appear to be largely accepted by both physicians and patients, as illustrated by (Dalaba et al., 2014; Bessat et al., 2019; Bernasconi et al., 2019).

Concerning ML-based DDSSs, just like their impact on patient safety or physicians' performances, their acceptability in practice remains understudied (Peiffer-Smadja et al., 2019). Jaroszewski et al. (2019) reported that, during clinical trials of their ML-based DDSS for mental illness prevention, only 28% of participants answered "very likely" to the question presented by the DDSS: "Be honest, how likely are you to try the resources I just shared?". Nadarzynski et al. (2020) studied the acceptability of information systems dedicated to sexual health prevention. The authors reported that, for the first contact, 70% of patients preferred face-to-face consultations. Only 40% of patients found AI-chatbot acceptable.

To sum-up, although there are exceptions in specific situations, it appears that DDSSs are generally poorly accepted in customary situations, where support appears to be redundant with physicians' capabilities (Masud et al., 2019). It hence appears that we are currently in a paradoxical situation. DDSSs appear to be able to improve physicians' performances and patient safety. However, in practice, DDSSs remain poorly accepted in many situations and difficult to integrate into physicians' workflow. It appears also that the intrinsic capacities of a DDSS are not the sole factor determining its usefulness. There is hence a need to better understand why some DDSSs are not well accepted and which features are likely to improve the acceptability of a DDSS in practice.

3 Explaining the paradoxical failure

In section 2, we saw that Clinical Decision Support Systems (CDSSs) are potentially beneficial to minimize medical errors in some cases. However, we also saw that the introduction of a CDSS in a hospital is not without risks or failure and that current Diagnostic Decision Support Systems (DDSSs) are generally not accepted by clinicians, who often ignore DDSS recommendations in their daily practice.

Early explorations of barriers to the use of guidelines contain useful indications on reasons why some decision support tools can be rejected by physicians. Cabana et al. (1999) made an early meta-analysis of 76 studies on the non-acceptability of clinical practice guidelines and reported 7 potential barriers, classified into three categories:

1. External barriers such as the presence of contradictory guidelines, the inability to reconcile patient's preferences with guidelines recommendations, and other environmental factors such as the lack of time or resources.
2. Barriers that affect the attitude of a physician towards guidelines, such as the lack of agreement with specific guidelines or guidelines in general, the inertia of previous practices, the belief that s/he cannot perform guideline recommendations and the belief that performance of guideline recommendations will not lead to desired outcomes.
3. Barriers linked with how knowledgeable physicians are about guidelines, due for example to problems of accessibility of the guidelines, or to the volume of information to compute and then the time needed to stay informed.

In this section, we enlarge and update this analysis of potential reasons for non-acceptability, applying it more broadly to different aspects of HISs, CDSSs, and DDSSs, with a special focus on customary diagnostic.

3.1 Adverse impacts of HISs, CDSSs and DDSSs

Tsai et al. (2003) studied the impact of wrong diagnostic suggestions given by a DDSS on physicians' performance. They thereby questioned a commonly accepted postulate: if a DDSS does a mistake or a wrong proposal, the physician will detect it. This study was based on 83 simulations adapted from real clinical cases of cardiology. The subjects were 30 internal medicine residents in their second or third years of training and the DDSS was controlled to produce sometimes proposals that did not fit with *"gold standards"*. Tsai et al. (2003) reported that, when the DDSS produced good proposals, the accuracy of subjects increased. By contrast, the subject's accuracy dropped down when the DDSS proposal was incorrect. These authors also reported that subjects followed the DDSS's proposal more often when it was presented with a good confidence index. Povyakalo et al. (2013) developed a similar study on the impact of computer-aided detection of cancer on the performance of 50 radiologists. In this study, they evaluated the discriminating ability of radiologists on 180 mammograms with and without computer support. They reported that computer-aided detection helped less discriminating radiologists, but hindered the more discriminating radiologists by reducing their sensitivity. Bowman (2013), who worked on safety implications of electronic health record (EHR) systems, reported that poor design, improper use, and EHR-related errors, such as bugs or errors in the data, can lead to errors that endanger patients and decrease the quality of care. The risk of poor design and programming errors actually concerns all kinds of HIS, including CDSSs.

Bertillot (2016) studied HISs' attempts at rationalizing and standardizing clinicians daily practices, based on a set of interviews of clinicians (physicians, nurses, etc.) in several hospitals in France. Bertillot (2016) thereby showed that the introduction, in the last decades, of different HISs in hospitals improved the traceability of hospitalized patients and allowed for better transmission of information, but it also set the stage for the introduction of evaluation systems in these hospitals. These evaluation systems allowed comparing performances between hospital services, which led to the introduction of *"competitive managerial practices in public hospital"*. Bertillot (2016) also reported an additional administrative workload for clinicians, who had to enter information in the software. This time spent doing administrative work, though necessary for different reasons, is not a time devoted to patients. Mitchell et al. (2016)'s results highlight the same aspect of the impacts of HISs. They interviewed patient safety experts about their perceptions of works on patient safety incident reporting. This qualitative study highlights that clinicians, mainly due to a lack of time, perceived systematic reviews of patient safety incidents as an additional workload. Hall et al. (2016) reviewed 46 studies on wellbeing and patient safety to determine if there was an association between clinicians' wellbeing, burnout, and patient safety. They reported that clinicians' poor wellbeing was significantly correlated with higher risks of burnout, worse patient safety, and higher risks of medical errors. West et al. (2018) made a similar work on clinicians, burnout, their reasons, and their consequences. They reported the use of HISs as one of the factors leading to clinicians' burnout. One can hence see that, by trying to reduce the risk of medical errors, current HISs increase clinicians' workload. This additional workload reduces the wellbeing of clinicians and, by collateral effect, potentially increase the risk of medical errors in practice.

An associated risk was studied by Cabitza et al. (2017): the unintended consequences of Machine Learning in medicine. They reported that ML systems, due to their efficiency but also their opacity, could amplify the loss of clinicians' skills reported by Tsai et al. (2003) and Povyakalo et al. (2013). They also reported that the intrinsic uncertainty of healthcare contexts affects the performances of ML systems, reducing their accuracy. Similarly, Challen et al. (2019) studied the potential impact of artificial intelligence on clinical safety. They reported potential causes of errors due to AI tools in healthcare contexts. For example, ML systems are generally trained in a specific context and lose their accuracy when the context is changed. The opacity of some ML systems and the automation complacency were also reported as factors increasing the risk of medical errors. Authors also argued that reinforcement-based ML systems for decision support are potentially dangerous in the long run, by making unsafe exploration or reinforcing only short term behaviors.

If physicians are the only ones in charge of detecting potential errors of tools supposed to support them, it simply creates an additional workload and appears counterproductive. Not to mention the fact that, in the case of ML-based DDSS, physicians are supposed to be less "competent" than the DDSS to do the same tasks, and are therefore unlikely to be able to detect if the DDSS has made an error[7].

3.2 Responsibility issues

Itani et al. (2019), who studied the use of data mining algorithms for decision support, showed that social factors, such as patients' and physicians' values, are an important aspect to take into account to understand the acceptability or rejection of DDSS. These values refer to social perceptions and ethical implications of the use of DDSSs, but also to the social pressure on the responsibility of physicians with respect to the consequences of their decisions.

According to Goodman (2016), who surveyed the ethical and legal issues surrounding CDSSs, there is a need to define legal responsibilities in the use of CDSSs. Indeed, if one uses a DDSS and the DDSS is wrong, who is responsible? (De Dombal, 1987) The answer clearly depends on how the DDSS was developed or used.

For example, a technical error in programming could lead to an ill-advised recommendation. In such a case, one might argue that the true responsible is the programmer. But medical errors could also come from mistakes that a physician made when using the DDSS. The method on which the DDSS was based can also be a source of error. In the case of a Guideline-based DDSS using rules defined by experts, these "experts" might have provided rules that can be considered to be "dangerous" or "foolish" by the rest of the medical community.

One might argue that ML-based DDSSs are more trustworthy than Guideline-based DDSSs, due to the high performances of ML algorithms, outperforming physicians (Esteva et al., 2017), and using large amounts of data. However, responsibility issues are not different in the case of ML-based DDSSs: if the ML-based DDSS's recommendation was wrong and led to a medical error, who was responsible? ML-based DDSSs are trained and evaluated on datasets that might fail to encompass all the variety of possible use cases. Even if a trained ML-based DDSS had high sensitivity and specificity on a test dataset, these criteria of performances are not enough when we talk about patient safety in real situations. Moreover, supervised ML algorithms can only reproduce the behaviors they learned. Therefore, just like in the case of Guideline-based DDSSs, if the learning dataset was based on the behaviors of physicians whose behavior can be considered to be "dangerous" or "foolish" by the rest of the medical community, the trained ML algorithm will reproduce, and even amplify, this "dangerous" behavior (Garcia, 2016; Sandvig et al., 2016; Zou and Schiebinger, 2018). The main difference with Guideline-based DDSSs is that it is more difficult for physicians to detect if a ML-based DDSS had an unwanted behavior, especially if the process

[7] The emerging field of Explainable AI (Doran et al., 2017; Gunning, 2017; Rudin and Radin, 2019) holds promises to mitigate this problem.

of the ML-based DDSS is opaque to physicians. This is all the more worrying when physicians have high confidence in the ML-based DDSS's recommendations because the latter outperformed them (Tsai et al., 2003; Povyakalo et al., 2013). In such cases, responsibility problems are all the more worrying.

To summarize, both guideline-based and ML-based DDSSs create problems when physicians are considered to be responsible for patient outcomes. As introduced in subsection 3.1, physicians cannot be the only ones responsible for preventing potential medical errors due to the use of a DDSSs supposed to support them. According to the Asilomar AI Principles[8], developed during a workshop organized by the FutureOfLife Institute and dedicated to guiding institutions and designers to build beneficial Artificial Intelligence (AI), designers of AI systems and institutions must take up their share of responsibility in preventing errors or misuses of AI systems. These principles concern ML-based DDSSs, but also some Guideline-based DDSSs such as expert systems. The emergence of legislative instruments aimed at regulating the use of HISs, and more specifically the use of personal data, and to encourage the transparency of algorithms (e.g. GDPR in Europe (Voigt and Von dem Bussche, 2017)), witnesses the growing public awareness of such problems, pinpointing the fact that current decision support systems fall short of expectations.

3.3 A reality-design gap in customary situations

In addition to the adverse impacts of HISs and to responsibility issues, the literature suggests another reason potentially explaining the non-acceptability of some decision support tools by physicians: the so-called "reality-design gap problem".

This concept was introduced by Heeks et al. (1999) and Heeks (2006) in an attempt to explain why HISs succeed or fail. They argued that the bigger the gap between how a HIS was designed and the reality of daily practices, the higher the risk that the system will fail. To formalize this problem of design-reality gap, Heeks et al. (1999) proposed the ITPOSMO framework, formalizing seven dimensions that could create a gap: Information (Are physicians accustomed to using such kind of information?), Technology (Do the hospital have the technological capacities to run this system?), Processes (How does the system integrate itself into physicians' workflow?), Objectives and values (Do objectives of the system match with physicians' objectives and values?), Staffing and Skills (Does the system necessitate high technical skills to be used?), Management systems (Does the system necessitate additional structures to manage it?) and Other resources (Is it time-costing to use the system? Does the system create any additional workload?).

According to Heeks et al. (1999) and Heeks (2006), if the introduction of a HIS requires too many and/or too profound changes in clinicians' current daily practices, then the risk of non-acceptability is high. However, the goal of the introduction of a HIS is to improve clinical processes and/or healthcare outcomes, and accordingly to induce changes in clinical practices. If a HIS is too close to clinicians' daily practices, no improvement is possible. The difficulty in designing HISs is therefore to find a convenient equilibrium between minimizing the risk of non-acceptability of the HIS and maximizing the potential improvements of clinicians' practices.

[8] `https://futureoflife.org/ai-principles/`

To flesh out the meaning of this reasoning for our investigation, let us detail the content of the seven dimensions of ITPOSMO in the case that we focus on in this article: the one of customary consultations:

– Information: Guideline-based DDSSs generally provide actions/treatment recommendations based on "gold-standard" guidelines adapted to the situation. Physicians and clinicians are accustomed to the use of such "gold-standard" guidelines, part of their work being to be aware of the new "gold-standard" for the cases they treat regularly.
– Technology: Guideline-based DDSSs are generally integrated into already existing HISs and do not necessitate more technological resources than access to a database.
– Processes: Guideline-based DDSSs generally necessitate that physicians enter symptoms of the patient or other additional data asked by the DDSS. In customary situations, the process of Guideline-based DDSSs can be redundant with the physicians' process during a consultation.
– Objectives and values: the objective of Guideline-based DDSSs is generally to improve adherence to "gold-standard" guidelines. In customary situations, this objective is confronted with physicians' values, such as their free will, or the acceptability of "gold standard" guidelines (Cabana et al., 1999). Guideline-based DDSSs can also automatize too many things in physicians' workflow, leading potentially to a sensation of lack of control (Heeks, 2006)
– Staffing and Skills: Guideline-based generally do not necessitate additional skills to be used by physicians.
– Management systems: because "gold-standards" are evolving continuously, guideline-based DDSSs generally need to be managed regularly by an external agent to keep their recommendations up to date with the most recent "gold-standard" guidelines.
– Other resources: the use of Guideline-based DDSSs can be time-consuming for physicians, who can spend more time on the tools than interacting with the patient (Porat et al., 2017). In addition, physicians have to understand the reasons behind recommendations to prevent medical mistakes, creating an additional workload for physicians.

Concerning ML-based DDSSs, the task is a bit more difficult than for Guideline-based DDSSs, mainly because it is still an emerging domain, and ML-based DDSSs are still rarely used in practice (Peiffer-Smadja et al., 2019).

– Information: ML-based DDSSs generally provide recommendations or risk degrees. However, the reasons underlying a given recommendation can be unintelligible for physicians, depending on the ML algorithm used.
– Technology: Some learning algorithms, such as neural networks, can necessitate powerful technological resources. However, the classifier produced by a learning algorithm, such as a decision tree or a trained neural network, does not generally necessitate powerful resources to be used. In the case of online learning or continuous learning, powerful resources might be necessary (Kulikowski, 2019).
– Processes: current ML-based DDSSs are generally based on data already entered in the system and do not necessitate additional actions to be done by physicians. They can provide their support quickly in specific points. They can then easily be integrated into physicians' workflow as the display of an additional piece of information about a patient.

- Objectives and values: the main objective of current ML-based DDSSs is to provide highly performant tools to guide physicians in tasks they are not able to do alone with the same accuracy. In customary situations, for which physicians can be considered to be "competent", this objective may seem superfluous and can arouse their suspicion.
- Staffing and Skills: the use of current ML-based DDSSs might require additional training by physicians at least in terms of know-how to interpret the DDSS's results and to better understand how they work, their strengths and limitations.
- Management systems: It is possible to implement continuous learning by updating regularly the training dataset and rerunning the learning algorithm. This might also require to continuously test the performances of the DDSS, to prevent errors. However, none of these necessarily requires the intervention of an external agent, and everything can be automated.
- Other resources: the understanding of recommendations by physicians, when it is possible, may generate additional workload.

There certainly are exceptions to the general characteristics we explored above in our application of reality/design gaps analysis to Guideline-based DDSSs and ML-based DDSSs. Our goal was simply to highlight general trends in current ways to support physicians during their practices and see if they are applicable or not for customary consultations. Concerning Guideline-based DDSSs, for customary cases, physicians are often already aware of "gold-standard" to follow. Using Guideline-based DDSSs is generally time-consuming for physicians and redundant with their existing workflow. Concerning ML-based DDSSs, the main gap comes from the technology used. If physicians do not understand how the system works and how to interpret its results, the system will be seen as a "black-box", generating distrust. This will be reinforced if the objective of the ML-based DDSS is to outperform physicians in customary situations for which they feel competent.

Besides, for both Guideline-based and ML-based DDSSs, physicians are entrusted with the responsibility to make sure that the DDSS did not mislead her/him with ill-advised recommendations, creating an additional workload.

To sum-up our exploration so far, it appears that current tools used to support physicians are plagued by important drawbacks (adverse impacts, responsibility issues, and reality-design gaps), which are exacerbated in customary consultations. New approaches to support physicians in such situations are hence needed.

4 The way forward: the quest for "the right information"

According to Osheroff et al. (2012), the goal of CDSSs is to improve healthcare decisions and outcomes, including patient safety, by giving physicians the "right information". Osheroff's definition proved successful in the literature because it provides a synthetic formula that looks unquestionable. It also conveniently encompasses the immense diversity of CDSSs. But this successfulness of the formula also lies to a large extent in the indeterminacy of the phrase "the right information". In the case of current DDSSs, the "right information" is embodied by guidelines and/or diagnosis recommendations. In this section, we explore the idea that a crucial reason underlying the lack of acceptability of current DDSSs by physicians in customary consultations might be that this "right information" is not that right after all, and we set out to identify the truly "right information".

4.1 What is "information" in healthcare contexts?

At first glance, one might think that the notion of "information" in our context is unequivocal. A piece of information, one might think, is a raw data formatted to be readable by a physician. The interpretation of a piece of information by a physician gives her/him pieces of knowledge about a situation and allows her/him to make a decision. Collected by EHRs and CPOEs, hospital databases are rich in such raw data on patients, including: weights, ages, symptoms, reviews of hospitalizations, drugs took, allergies, etc. All these raw data can give clinicians a first layer of information.

With the same logic, the evolution of such data through time, their interconnection in patient care processes, gives a second layer of information. The notion of information hence appears more complex after all, since there are several layers of information.

A third layer of information can be found in guidelines summarizing "gold standards" to follow in a specific situation or for a specific operation. As mentioned in section 1, clinical practice guidelines are a list of instructions to follow in a specific situation. Guidelines include various formats such as pathways or algorithms to follow, and/or appropriateness criteria or parameters to check and instructions concerning how to interpret them (Field et al., 1990). But by admitting that "information" can refer to that third kind of entity, one admits that interpretation frameworks thanks to which data are interpreted, such as theories or sets of practices and know-how, are also pieces of "information" in a sense.

We see here that, in healthcare contexts, the term "information" can refer to a large diversity of entities, including raw data, interpreted data, and interpretation frameworks.

This analysis of the notion of "information" in healthcare contexts shows that the current approach, which consists in giving guidelines to physicians, is a particular kind of decision support approach, anchored in a very particular understanding of the notion of "information". This approach reflects a desire to standardize diagnosis processes, based on the presupposed idea that such a standardization will lead to minimizing medical errors. However, as mentioned by Woolf (1993), who studied the impacts of guidelines on patient care, such standardization could harm patients and interfere with the individualization of care. In clinical contexts, physicians' adaptability can, in many cases, be more important than conformism.

This suggests that, instead of clinging to the standard reductive view of "information" aimed at standardizing diagnosis processes, one should strive to identify the kind of information that physicians need when they proceed to make a diagnosis.

4.2 Identifying the constraints binding the decision support process to determine what is the "right information”

We claim that, in order to identify what counts as the "right information" in customary diagnosis decision support, we need to analyze, at a methodological and epistemological level, the true meaning and significance of the activity that consists in providing decision support in this context. Our analysis so far has highlighted the numerous specificities associated with the context of customary diagnosis. As in most medical contexts, this specific context raises responsibility issues, but another marked specificity is that, in this context, physicians are competent, and are not easily outcompeted by sophisticated tools.

These specificities reflect constraints that bind the interaction between decision support providers, developing decision support tools, and physicians, which are decision-makers benefiting from decision support. Decision support providers concerned with providing relevant and *acceptable* decision support have no choice but to take these constraints into account to choose the kind of approach to unfold in their interactions with physicians.

Meinard and Tsoukiàs (2019) showed the pivotal role of an analysis of the constraints binding decision support processes, which is key to choose a relevant decision support approach, which plays, in turn, a decisive role to entrench the validity and the legitimacy of the decision support provided. This framework sheds useful light on our analysis of the various drawbacks plaguing various current DDSSs, developed above.

As explained in section 1, Guideline-based DDSSs include systems providing recommendations based on "gold-standard" guidelines, but also systems providing directly these "gold-standard" guidelines. This approach is relevant when the interaction between the decision-maker and the decision support provider is constrained by a requirement to homogenize decision processes and make them converge towards "gold-standards" that are collectively recognized, by expert institutions and the general public. Such decision support interactions correspond to what Meinard and Tsoukiàs (2019) call situations in which an "irrevocable governance pattern" binds the decision support process. Still according to Meinard and Tsoukiàs (2019), in such situations, decision support providers should endorse a "conformist" approach, striving to identify the tools that will be most acceptable to the members of the governance pattern. In situations in which physicians' skills are deficient due to problems in physician training, as observed in some cases in developing countries, and in which health institutions play a key role in a powerful governance scheme aimed at reinforcing the quality of medical treatment, Guideline-based DDSSs are relevant. In such cases, "the right information" that decision support should provide to physicians really is encapsulated in guidelines.

ML-based DDSSs include all recommendation systems based on supervised ML algorithms. These tools assume that there exists a function linking data on patients to a specific class (e.g., disease, set of treatments, risk degree, etc.), independently of the beliefs and knowledge of the decision-maker, or her/his context of decisions. They also assume that this function can be approximated by machine learning, and more specifically by deep learning because multi-layered neural networks are known to be universal approximators (Hornik et al., 1989). The goal is then to find the best approximation of this function, generally evaluated by its sensitivity and specificity. An approach based on such tools is relevant in situations in which the objectivity and truthfulness of the underlying theories and algorithms can be taken for granted and considered unquestionable. In such situations, in which a given theoretical framework is considered to be unquestionable, Meinard and Tsoukiàs (2019) argue that the relevant approach to decision support is "objectivist". In healthcare contexts, situations bound by this constraint are those in which the data to collect, the efficiency of existing tools to collect them, and their capacity to outperform all other forms of expertise, are clearly established. This is the case, for example, for the detection of ocular diseases (Zhang et al., 2018) or the evaluation of the risk of infection (Peiffer-Smadja et al., 2019) or of treatment reaction (Deig et al., 2019). These are tasks for which we can suppose that a classification function exists, but we cannot suppose that any decision-maker is "competent" enough to approximate it closely.

The decision support context that we are mainly interested in here, customary consultations, is not characterized by patterns similar to those presented above, for which Guideline-based DDSSs and ML-based DDSSs appear relevant, respectively. In customary consultations, physicians are competent: they do not need to be monitored by authorities verifying their compliance with gold-standards, and they do not need tools to replace them. In such contexts, the main constraint is that the conditions should be met for physicians to be able to exercise their responsibilities. This echoes situations that Meinard and Tsoukiàs (2019) refer to by talking about a constraint to respect a "sanctified spirit of initiative" of the decision-maker. This phrase is arguably rather vague, but the case of physicians performing customary diagnosis might be case in point to clarify it. The two distinctive features of customary diagnosis, competence and responsibility, point to the need for decision support providers to leave the decision-maker to make her/his own choices and to take responsibility for them. The role of the decision support provider in such cases is to make all efforts to facilitate, smoothen, and speed-up the processes favored by the decision-maker, and to adjust to her/his needs. This approach to decision support is called "adjustive" by Meinard and Tsoukiàs (2019).

4.2.1 An example of application

In order to flesh out in concrete terms what such an "adjustive" approach consists in, we introduce here a practical example, referring to Richard et al. (2018), a work developed in collaboration with the public hospitals of Lyon to propose a DDSS dedicated to supporting customary consultations. Clinicians of the hospitals of Lyon have a software, called Easily®, at their disposal. This software allows clinicians to access different kinds of HISs. During consultations, physicians have access to the hospital's EHR and to CPOEs, but not to any DDSSs for now.

Richard et al. (2018) proposed an analysis focussed on interactions between physicians and patients, but also between physicians and HISs, during customary consultations of endocrinologists, to identify what kind of tools should provide relevant support in such situations. To do so, Richard et al. (2018) made practical observations and analyzed event logs of customary consultations (more than 12.000 event logs, divided into 2.700 traces).

Based on these analyses, Richard et al. (2018) built a synthetic model of the decision process of physicians during a medical consultation (Fig. 2) This model shows interactions between possible actions of physicians during medical consultations. Richard et al. (2018) highlighted that key points of the decision process, such as the choice of a prescription or the choice to put end to the consultations, are highly dependent on the accumulation of data on patients. Accordingly, the authors concluded that the search for raw data on patients, and then the choice of the raw data to look at, constitute a central key point of consultations.

However, due to the huge quantities of data accumulated on patients in the last decades, physicians are nowadays flooded by medical data (Pivovarov and Elhadad, 2015). Even though most of the data on patients needed by physicians during their medical consultations are available, these data are not always easily accessible.

The model proposed by (Richard et al., 2018) suggests that physicians, during medical consultations, spend more time searching for data about the patient than analyzing them to reach a diagnostic. This analysis shows that, contrary to what most DDSSs currently available assume, what physicians need during customary consultations is not recommendations of diagnoses or diagnostic guidelines. As concluded by Richard

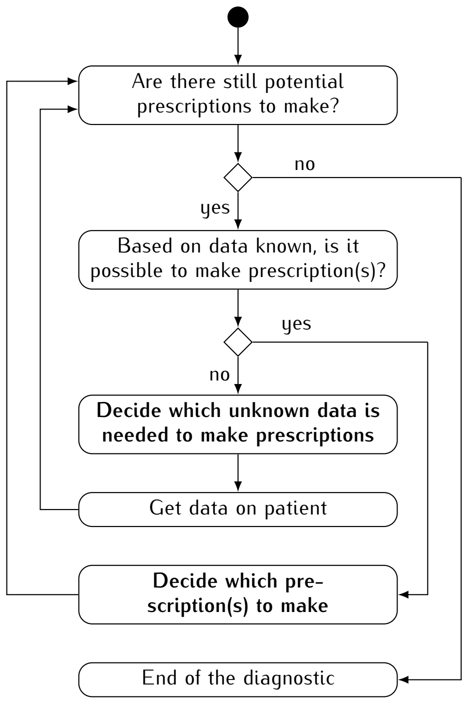


Fig. 2 Graphical representation of the model of physicians' diagnostic decision process proposed by Richard et al. (2018)

et al. (2018), what physicians need are tools that can anticipate, retrieve, and summarize data needed by physicians about patients. A relevant tool is accordingly one that would speed up the search for data. This idea echoes Sittig et al. (2006), who argued that guidelines and/or diagnostic recommendations are useless but for complex cases. It also appears all the more relevant in the light of studies on the summarization of electronic health records (EHRs) such as Pivovarov and Elhadad (2015), showing that there is an increasing need for EHRs summarizers.

Nevertheless, physicians need different data depending on their medical specialty or the pathology of the patient. This can be learned by questioning physicians and by creating a set of rules, but expert systems are generally difficult to build and to maintain through time (Shortliffe, 2012; Miller, 2010). In addition, questioning physicians would be against our aim to prevent any increase in their workload. Richard et al. (2018) therefore set out to learn what data are needed by physicians by analyzing their searches and their entry in the hospital's database, so as to anticipate their needs and provide them with a subset of data about their current patient at the beginning of medical consultation. By doing so, the searching phase of medical consultations should be minimized by handing over to the information system the task for which it is more efficient than human beings: searching data in a large database. In this approach, the aim of decision support is to ensure that physicians have all the data they need on their patients, and the interpretation of these data is then left to physicians.

4.3 Promises and limits

As mentioned before, the conclusions reached by Richard et al. (2018) were based only on observations and analyses of consultations in endocrinology at the HCL. However, the approach proposed in this case-study, illustrating the more largely applicable reasoning developed above, holds promises in light of the above analysis of the reasons underlying the non-acceptability of current DDSSs.

A first strength of the proposed approach is that it draws on the competence and the cumulated experience of the physician. We have seen that current approaches used to support physicians suppose that physicians are not competent enough. Whereas such approaches can be relevant in complex situations, for customary consultations they are inappropriate and they can arouse distrust towards the DDSS among physicians or a feeling of being put aside by the DDSS. With an adjustive approach, physicians keep the leadership of the decision process. Moreover, their competences in drawing diagnosis and interacting with patients are highlighted.

A second important strength of the proposed approach is that it does not increase the workload of physicians, and rather decreases it. We have seen above that the increase of clinicians' workload, due to the introduction of decision support systems, has been reported as a barrier to their acceptability. Current approaches tend to tell physicians how they should work, without taking into account their current decision processes or the impact of the format of decision support. With our approach, the first aim is to understand physicians' current decision processes, to establish on which point of their workflow physicians need support in priority and what kind of decision support is more relevant to provide.

A third strength is that, as compared with numerous other approaches, our approach does not involve a risk to decrease physicians' performances or capacities. According to Povyakalo et al. (2013), the use of current DDSSs tends to decrease the performances of physicians with good diagnosis skills. In addition, Tsai et al. (2003) have reported that wrong recommendations of current DDSSs are less detected by inexperienced physicians. This loss of diagnosis skills is often cited as an important barrier to diagnosis decision support. Focussing on providing data on patients prevents this problem by refusing to prescribe what to do during the diagnosis decision process. The interpretation of data on patients is left to the physicians. The impact of our approach on physicians' diagnosis skills is therefore minimized.

Lastly, and arguably most importantly, a major strength of our approach is that it does not infringe upon the responsibility of the physician. Indeed, the responsibility issues raised by the use of guidelines, described in section 3.2, are no longer a problem if we focus on providing data on patients. As mentioned above, our approach does not prescribe what to do during the diagnosis decision process, it only focusses on providing to physicians with what they need during their decision processes.

5 Conclusion

In this paper, we have developed a reflection on the current approaches to supporting customary diagnostic decisions, which consist mainly of giving guidelines and/or diagnosis recommendations. We have explored the historical reasons that led to the choice of this approach and we have highlighted its drawbacks. In particular, we have stressed the fact that DDSSs tend to put physicians at the background on their own decisions, raise various responsibility issues, and are generally not accepted by physicians.

We have then argued that giving guidelines or recommendations reflects a strong choice on how to support decisions, which ignores the current decision-maker process or the impact of recommendations on this process. In the case of customary medical consultations, the "sanctified spirit of the initiative" of physicians is currently a binding constraint. Current DDSSs are not relevant in such situations. We have argued that DDSSs dedicated to supporting customary consultations must endorse an adjustive approach, which consists in ensuring that physicians have all the data they need about patients to reach a diagnosis. The interpretation and final decisions are then left to the decision-maker and her/his expertise, avoiding responsibility issues raised by Guideline-based and ML-based DDSSs in such situations.

Decision support systems developed in an adjustive approach can be seen as "personal assistants" that provide support during all the decision process and adjust themselves by interacting with decision-makers. However, just like "conformist" and "objectivist" approaches, adjustive approaches are not adapted to all situations. In cases in which conformist and objectivist approaches are relevant, guideline-based and ML-based DDSS undoubtedly have a role to play, and one should certainly not replace them by adjustive approaches. Analyzing the features of decision processes, the constraints binding interactions between decision-makers and decision support providers, and other aspects of the context, is always needed to choose the most relevant approach. Identifying these points during the development of new DDSSs could help designers to have a better understanding of the kind of support needed and to propose more adapted systems to physicians. Works that include physicians or clinicians in the development offer interesting promises in this respect (Giordanengo et al., 2019; Horrocks et al., 2018).

The reasoning developed in this article is focussed on diagnostic decision support for customary medical consultations, tasks for which physicians are considered to be competent and responsible. However, it bears lessons for other contexts in which decision support has to be provided to competent, responsible decision-makers. For example, in the context of the implementation of environmental policies, Meinard and Thébaud (2019) argued that environmental management schemes are currently crippled in France by the lack of a large-scale database on vegetation types, while environmental institutions spend considerable time and money to produce ill-adapted guidelines unusable by experts in the field. Decision support in this area could largely benefit from an analysis developed along the lines of our analysis of DDSSs.

Acknowledgements This work was made in collaboration with employees of the hospitals of Lyon. Thanks to all of them. Special thanks to Pr. Moulin and Dr. Riou for their suggestions and instructive discussions. Special thanks also to Juliette Rouchier, Olivier Cailloux, and Philippe Grill for their advices and comments on earlier versions of this manuscript. We also thank two anonymous reviewers of the journal for their powerful and exacting comments and criticisms.

References

Anand V, Biondich PG, Liu GC, Rosenman MB, Downs SM, et al. (2004) Child health improvement through computer automation: the chica system. In: Medinfo, pp 187–191, DOI 10.3233/978-1-60750-949-3-187

Barnett GO, Cimino JJ, Hupp JA, Hoffer EP (1987) Dxplain: an evolving diagnostic decision-support system. Jama 258(1):67–74, DOI 10.1001/jama.1987.03400010071030

Bernasconi A, Crabbé F, Adedeji AM, Bello A, Schmitz T, Landi M, Rossi R (2019) Results from one-year use of an electronic clinical decision support system in a post-conflict context: An implementation research. PloS one 14(12), DOI 10.1371/journal.pone.0225634

Berner ES (2016) Clinical decision support systems, 3rd edn. Springer

Berner ES, Graber ML (2008) Overconfidence as a cause of diagnostic error in medicine. The American journal of medicine 121(5):S2–S23, DOI 10.1016/j.amjmed.2008.01.001

Bertillot H (2016) Comment l'évaluation de la qualité transforme l'hôpital. les deux visages de la rationalisation par les indicateurs. Cahiers internationaux de sociologie de la gestion 15:11–48

Bessat C, Zonon NA, D'Acremont V (2019) Large-scale implementation of electronic integrated management of childhood illness (eimci) at the primary care level in burkina faso: a qualitative study on health worker perception of its medical content, usability and impact on antibiotic prescription and resistance. BMC public health 19(1):449, DOI 10.1186/s12889-019-6692-6

Beynon-Davies P, Lloyd-Williams M (1998) Health information systems,'safety'and organizational learning. Health Informatics Journal 4(3-4):128–137, DOI 10.1177/146045829800400303

Bowman S (2013) Impact of electronic health record systems on information integrity: quality and safety implications. Perspectives in health information management 10(Fall)

Cabana MD, Rand CS, Powe NR, Wu AW, Wilson MH, Abboud PAC, Rubin HR (1999) Why don't physicians follow clinical practice guidelines?: A framework for improvement. Jama 282(15):1458–1465, DOI 10.1001/jama.282.15.1458

Cabitza F, Rasoini R, Gensini GF (2017) Unintended consequences of machine learning in medicine. Jama 318(6):517–518, DOI 10.1001/jama.2017.7797

Carroll AE, Biondich P, Anand V, Dugan TM, Downs SM (2013) A randomized controlled trial of screening for maternal depression with a clinical decision support system. Journal of the American Medical Informatics Association 20(2):311–316, DOI 10.1136/amiajnl-2011-000682

Chadwick D, Hall C, Rae C, Rayment M, Branch M, Littlewood J, Sullivan A (2017) A feasibility study for a clinical decision support system prompting hiv testing. HIV medicine 18(6):435–439, DOI 10.1111/hiv.12472

Challen R, Denny J, Pitt M, Gompels L, Edwards T, Tsaneva-Atanasova K (2019) Artificial intelligence, bias and clinical safety. BMJ Qual Saf pp bmjqs–2018, DOI 10.1136/bmjqs-2018-008370

Chang PL, Li Yc, Huang ST, Wang TM, Hsien ML (1996) Effects of a medical expert system on differential diagnosis of renal masses: A prospective study. Computerized medical imaging and graphics 20(1):43–48, DOI 10.1016/0895-6111(96)00029-8

Chaudhry B, Wang J, Wu S, Maglione M, Mojica W, Roth E, Morton SC, Shekelle PG (2006) Systematic review: impact of health information technology on quality, efficiency, and costs of medical care. Annals of internal medicine 144(10):742–752, DOI 10.7326/0003-4819-144-10-200605160-00125

Choulak M, Marage D, Gisbert M, Paris M, Meinard Y (2019) A meta-decision-analysis approach to structure operational and legitimate environmental policies–with an application to wetland prioritization. Science of the Total Environment 655:384–394,

DOI 10.1016/j.scitotenv.2018.11.202
Clancy CM (2009) Ten years after to err is human. American Journal of Medical Quality 24(6):525–528, DOI 10.1177/1062860609349728
Dalaba MA, Akweongo P, Williams J, Saronga HP, Tonchev P, Sauerborn R, Mensah N, Blank A, Kaltschmidt J, Loukanova S (2014) Costs associated with implementation of computer-assisted clinical decision support system for antenatal and delivery care: case study of kassena-nankana district of northern ghana. PLoS One 9(9), DOI 10.1371/journal.pone.0106416
De Dombal F (1987) Ethical considerations concerning computers in medicine in the 1980s. Journal of medical ethics 13(4):179–184, DOI 10.1136/jme.13.4.179
De Fauw J, Ledsam JR, Romera-Paredes B, Nikolov S, Tomasev N, Blackwell S, Askham H, Glorot X, O'Donoghue B, Visentin D, et al. (2018) Clinically applicable deep learning for diagnosis and referral in retinal disease. Nature medicine 24(9):1342–1350, DOI 10.1038/s41591-018-0107-6
De Marchi G, Lucertini G, Tsoukiàs A (2016) From evidence-based policy making to policy analytics. Annals of Operations Research 236(1):15–38, DOI 10.1007/s10479-014-1578-6
Deig CR, Kanwar A, Thompson RF (2019) Artificial intelligence in radiation oncology. Hematology/Oncology Clinics 33(6):1095–1104, DOI 10.1016/j.hoc.2019.08.003
Donaldson MS, Corrigan JM, Kohn LT, et al. (2000) To err is human: building a safer health system, vol 6. National Academies Press
Doran D, Schulz S, Besold TR (2017) What does explainable ai really mean? a new conceptualization of perspectives. arXiv preprint arXiv:171000794
Dua S, Acharya UR, Dua P (2014) Machine learning in healthcare informatics, vol 56. Springer
Eccles M, McColl E, Steen N, Rousseau N, Grimshaw J, Parkin D, Purves I (2002) Effect of computerised evidence based guidelines on management of asthma and angina in adults in primary care: cluster randomised controlled trial. Bmj 325(7370):941, DOI 10.1136/bmj.325.7370.941
Elsner P, Bauer A, Diepgen TL, Drexler H, Fartasch M, John SM, Schliemann S, Wehrmann W, Tittelbach J (2018) Position paper: Telemedicine in occupational dermatology–current status and perspectives. JDDG: Journal der Deutschen Dermatologischen Gesellschaft 16(8):969–974, DOI 10.1111/ddg.13605
Elstein AS, Friedman CP, Wolf FM, Murphy G, Miller J, Fine P, Heckerling P, Miller T, Sisson J, Barlas S, et al. (1996) Effects of a decision support system on the diagnostic accuracy of users: a preliminary report. Journal of the American Medical Informatics Association 3(6):422–428, DOI 10.1136/jamia.1996.97084515
Esteva A, Kuprel B, Novoa RA, Ko J, Swetter SM, Blau HM, Thrun S (2017) Dermatologist-level classification of skin cancer with deep neural networks. Nature 542(7639):115–118, DOI 10.1038/nature21056
Etter DJ, McCord A, Ouyang F, Gilbert AL, Williams RL, Hall JA, Tu W, Downs SM, Aalsma MC (2018) Suicide screening in primary care: use of an electronic screener to assess suicidality and improve provider follow-up for adolescents. Journal of Adolescent Health 62(2):191–197, DOI 10.1016/j.jadohealth.2017.08.026
Ferretti V, Pluchinotta I, Tsoukiàs A (2019) Studying the generation of alternatives in public policy making processes. European Journal of Operational Research 273(1):353–363, DOI 10.1016/j.ejor.2018.07.054
Field MJ, Lohr KN, et al. (1990) Clinical practice guidelines: directions for a new program. National Academies Press

Fieschi M (1986) Expert systems for medical consultation. Health Policy 6(2):159–173, DOI 10.1016/0168-8510(86)90005-9

Garcia M (2016) Racist in the machine: The disturbing implications of algorithmic bias. World Policy Journal 33(4):111–117

Garg AX, Adhikari NK, McDonald H, Rosas-Arellano MP, Devereaux P, Beyene J, Sam J, Haynes RB (2005) Effects of computerized clinical decision support systems on practitioner performance and patient outcomes: a systematic review. Jama 293(10):1223–1238, DOI 10.1001/jama.293.10.1223

Giordanengo A, Årsand E, Woldaregay AZ, Bradway M, Grottland A, Hartvigsen G, Granja C, Torsvik T, Hansen AH (2019) Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: Iterative approach and qualitative case study. JMIR diabetes 4(3):e14002, DOI 10.2196/14002

Gonzalvo CM, de Wit HA, van Oijen BP, Hurkens KP, Janknegt R, Schols JM, Mulder WJ, Verhey FR, Winkens B, van der Kuy PHM (2017) Supporting clinical rules engine in the adjustment of medication (scream): protocol of a multicentre, prospective, randomised study. BMC geriatrics 17(1):35, DOI 10.1186/s12877-017-0426-3

Goodman KW (2016) Ethical and legal issues in decision support. In: Clinical decision support systems, Springer, pp 131–146, DOI 10.1007/978-3-319-31913-1_8

Gordon ES, Babu D, Laney DA (2018) The future is now: Technology's impact on the practice of genetic counseling. In: American Journal of Medical Genetics Part C: Seminars in Medical Genetics, Wiley Online Library, vol 178, pp 15–23, DOI 10.1002/ajmg.c.31599

Gräßer F, Beckert S, Küster D, Schmitt J, Abraham S, Malberg H, Zaunseder S (2017) Therapy decision support based on recommender system methods. Journal of healthcare engineering 2017, DOI 10.1155/2017/8659460

Guah M (1998) Evaluation and analysis of multimedia information system design and implementation at the coloscopy unit of st. james university hospital, leeds, uk. PhD thesis, MSc Dissertation, School of Management, UMIST, Manchester

Gunning D (2017) Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web 2

Hall LH, Johnson J, Watt I, Tsipa A, O'Connor DB (2016) Healthcare staff wellbeing, burnout, and patient safety: a systematic review. PloS one 11(7):e0159015, DOI 10.1371/journal.pone.0159015

Heckerling P, Elstein A, Terzian C, Kushner M (1991) The effect of incomplete knowledge on the diagnoses of a computer consultant system. Medical Informatics 16(4):363–370, DOI 10.3109/14639239109067658

Heeks R (2006) Health information systems: Failure, success and improvisation. International journal of medical informatics 75(2):125–137, DOI 10.1016/j.ijmedinf.2005.07.024

Heeks R, Mundy D, Salazar A (1999) Why health care information systems succeed or fail. Information Systems for Public Sector Management DOI 10.2139/ssrn.3540062

Hoffer EP, Feldman MJ, Kim RJ, Famiglietti KT, Barnett GO (2005) Dxplain: patterns of use of a mature expert system. In: AMIA Annual Symposium Proceedings, American Medical Informatics Association, vol 2005, p 321

Hollon TC, Lewis S, Pandian B, Niknafs YS, Garrard MR, Garton H, Maher CO, McFadden K, Snuderl M, Lieberman AP, et al. (2018) Rapid intraoperative diagnosis of pediatric brain tumors using stimulated raman histology. Cancer research 78(1):278–289, DOI 10.1158/0008-5472.CAN-17-1974

Honaker SM, Dugan T, Daftary A, Davis S, Saha C, Baye F, Freeman E, Downs SM (2018) Unexplained practice variation in primary care providers' concern for pediatric obstructive sleep apnea. Academic pediatrics 18(4):418–424, DOI 10.1016/j.acap.2018.01.011

Hornik K, Stinchcombe M, White H, et al. (1989) Multilayer feedforward networks are universal approximators. Neural networks 2(5):359–366, DOI 10.1016/0893-6080(89)90020-8

Horrocks M, Michail M, Aubeeluck A, Wright N, Morriss R (2018) An electronic clinical decision support system for the assessment and management of suicidality in primary care: Protocol for a mixed-methods study. JMIR research protocols 7(12):e11135, DOI 10.2196/11135

Hunt DL, Haynes RB, Hanna SE, Smith K (1998) Effects of computer-based clinical decision support systems on physician performance and patient outcomes: a systematic review. Jama 280(15):1339–1346, DOI 10.1001/jama.280.15.1339

Itani S, Lecron F, Fortemps P (2019) Specifics of medical data mining for diagnosis aid: A survey. Expert systems with applications 118:300–314, DOI 10.1016/j.eswa.2018.09.056

Jaroszewski AC, Morris RR, Nock MK (2019) Randomized controlled trial of an online machine learning-driven risk assessment and intervention platform for increasing the use of crisis services. Journal of consulting and clinical psychology 87(4):370, DOI 10.1037/ccp0000389

Jeanmougin M, Dehais C, Meinard Y (2017) Mismatch between habitat science and habitat directive: Lessons from the french (counter) example. Conservation Letters 10(5):634–644, DOI 10.1111/conl.12330

Johnston ME, Langton KB, Haynes RB, Mathieu A (1994) Effects of computer-based clinical decision support systems on clinician performance and patient outcome: a critical appraisal of research. Annals of internal medicine 120(2):135–142, DOI 10.7326/0003-4819-120-2-199401150-00007

Joo S, Yang YS, Moon WK, Kim HC (2004) Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features. IEEE transactions on medical imaging 23(10):1292–1300, DOI 10.1109/TMI.2004.834617

Kataria S, Ravindran V (2018) Digital health: a new dimension in rheumatology patient care. Rheumatology international 38(11):1949–1957, DOI 10.1007/s00296-018-4037-x

Kaushal R, Shojania KG, Bates DW (2003) Effects of computerized physician order entry and clinical decision support systems on medication safety: a systematic review. Archives of internal medicine 163(12):1409–1416, DOI 10.1001/archinte.163.12.1409

Keen J (1994) Information management in health services. Open University Press, DOI 10.1016/j.jacr.2017.08.033

Kirby AM, Kruger B, Jain R, Daniel P, Granger BB (2018) Using clinical decision support to improve referral rates in severe symptomatic aortic stenosis: a quality improvement initiative. CIN: Computers, Informatics, Nursing 36(11):525–529, DOI 10.1097/CIN.0000000000000471

Kostopoulou O, Porat T, Corrigan D, Mahmoud S, Delaney BC (2017) Diagnostic accuracy of gps when using an early-intervention decision support system: a high-fidelity simulation. Br J Gen Pract 67(656):e201–e208, DOI 10.3399/bjgp16X688417

Kulikowski CA (1988) Artificial intelligence in medical consultation systems: A review. IEEE Engineering in Medicine and Biology Magazine 7(2):34–39, DOI 10.1109/51.

1972
Kulikowski CA (2019) Beginnings of artificial intelligence in medicine (aim): Computational artifice assisting scientific inquiry and clinical art–with reflections on present aim challenges. Yearbook of medical informatics 28(01):249–256, DOI 10.1055/s-0039-1677895
Kuperman GJ, Gibson RF (2003) Computer physician order entry: benefits, costs, and issues. Annals of internal medicine 139(1):31, DOI 10.7326/0003-4819-139-1-200307010-00010
Leape LL (2000) Institute of medicine medical error figures are not exaggerated. Jama 284(1):95–97, DOI 10.1001/jama.284.1.95
Leape LL, Berwick DM (2005) Five years after to err is human: what have we learned? Jama 293(19):2384–2390, DOI 10.1001/jama.293.19.2384
López MM, López MM, de la Torre Díez I, Jimeno JCP, López-Coronado M (2017) Mhealth app for ios to help in diagnostic decision in ophthalmology to primary care physicians. Journal of medical systems 41(5):81, DOI 10.1007/s10916-017-0731-6
Masud R, Al-Rei M, Lokker C (2019) Computer-aided detection for breast cancer screening in clinical settings: Scoping review. JMIR medical informatics 7(3):e12660, DOI 10.2196/12660
Meinard Y, Cailloux O (2020) On justifying the norms underlying decision support. European Journal of Operational Research DOI 10.1016/j.ejor.2020.02.022
Meinard Y, Thébaud G (2019) L'identification syntaxonomique dans les démarches de gestion et/ou de restauration d'espaces naturels en france: pour ou contre? Naturae 2019(6):165–173
Meinard Y, Tsoukiàs A (2019) On the rationality of decision aiding processes. European Journal of Operational Research 273(3):1074–1084, DOI 10.1016/j.ejor.2018.09.009
Miller R, Masarie Jr F (1990) The demise of the "greek oracle" model for medical diagnostic systems. Methods of information in medicine 29(01):1–2
Miller RA (1994) Medical diagnostic decision support systems—past, present, and future: a threaded bibliography and brief commentary. Journal of the American Medical Informatics Association 1(1):8–27, DOI 10.1136/jamia.1994.95236141
Miller RA (2010) A history of the internist-1 and quick medical reference (QMR) computer-assisted diagnosis projects, with lessons learned. Yearbook of medical informatics 19(01):121–136, DOI 10.1055/s-0038-1638702
Miller RA (2016) Diagnostic decision support systems. In: Clinical Decision Support Systems, Springer, pp 181–208, DOI 10.1007/978-3-319-31913-1_11
Miller RA, McNeil MA, Challinor SM, Masarie Jr FE, Myers JD (1986) The internist-1/quick medical reference project—status report. Western Journal of Medicine 145(6):816
Miotto R, Wang F, Wang S, Jiang X, Dudley JT (2017) Deep learning for healthcare: review, opportunities and challenges. Briefings in bioinformatics 19(6):1236–1246, DOI 10.1093/bib/bbx044
Mitchell I, Schuster A, Smith K, Pronovost P, Wu A (2016) Patient safety incident reporting: a qualitative study of thoughts and perceptions of experts 15 years after 'to err is human'. BMJ Qual Saf 25(2):92–99, DOI 10.1136/bmjqs-2015-004405
Mokdad AH, Marks JS, Stroup DF, Gerberding JL (2004) Actual causes of death in the united states, 2000. Jama 291(10):1238–1245, DOI 10.1001/jama.291.10.1238
Morelli RA, Bronzino JD, Goethe JW (1987) Expert systems in psychiatry. Journal of medical systems 11(2-3):157–168, DOI 10.1007/BF00992350

Murphy GC, Friedman CP, Elstein AS, Wolf FM, Miller T, Miller J (1996) The influence of a decision support system on the differential diagnosis of medical practitioners at three levels of training. In: Proceedings of the AMIA Annual Fall Symposium, American Medical Informatics Association, p 219

Musen MA, Middleton B, Greenes RA (2014) Clinical decision-support systems. In: Biomedical informatics, Springer, pp 643–674, DOI 10.1007/978-1-4471-4474-8_22

Nadarzynski T, Bayley J, Llewellyn C, Kidsley S, Graham CA (2020) Acceptability of artificial intelligence (ai)-enabled chatbots, video consultations and live webchats as online platforms for sexual health advice. BMJ Sexual & Reproductive Health DOI 10.1136/bmjsrh-2018-200271

Onega T, Bowles EJA, Miglioretti DL, Carney PA, Geller BM, Yankaskas BC, Kerlikowske K, Sickles EA, Elmore JG (2010) Radiologists' perceptions of computer aided detection versus double reading for mammography interpretation. Academic radiology 17(10):1217–1226, DOI 10.1016/j.acra.2010.05.007

Osheroff JA, Teich JM, Levick D, Saldana FT Luis ans Velasco, Sittig DF, Rogeers KM, Jenders RA (2012) Improving outcomes with clinical decision support: an implementer's guide. HIMSS

Overhage JM, Tierney WM, Zhou XH, McDonald CJ (1997) A randomized trial of "corollary orders" to prevent errors of omission. Journal of the American Medical Informatics Association 4(5):364–375, DOI 10.1136/jamia.1997.0040364

Ozaydin B, Hardin JM, Chhieng DC (2016) Data mining and clinical decision support systems. In: Clinical Decision Support Systems, Springer, pp 45–68, DOI 10.1007/978-3-319-31913-1_3

Pare G, Elam JJ (1998) Introducing information technology in the clinical setting: Lessons learned in a trauma center. International Journal of Technology Assessment in Health Care 14(2):331–343, DOI 10.1017/S0266462300012290

Pasquali P, Sonthalia S, Moreno-Ramirez D, Sharma P, Agrawal M, Gupta S, Kumar D, Arora D, et al. (2020) Teledermatology and its current perspective. Indian Dermatology Online Journal 11(1):12, DOI 10.4103/idoj.IDOJ_241_19

Patel VL, Kushniruk AW, Yang S, Yale JF (2000) Impact of a computer-based patient record system on data collection, knowledge organization, and reasoning. Journal of the American Medical Informatics Association 7(6):569–585, DOI 10.1136/jamia.2000.0070569

Pearce C, McLeod A, Rinehart N, Patrick J, Fragkoudi A, Ferrigi J, Deveny E, Whyte R, Shearer M (2019) Polar diversion: using general practice data to calculate risk of emergency department presentation at the time of consultation. Applied clinical informatics 10(01):151–157, DOI 10.1055/s-0039-1678608

Peiffer-Smadja N, Rawson TM, Ahmad R, Buchard A, Pantelis G, Lescure FX, Birgand G, Holmes AH (2019) Machine learning for clinical decision support in infectious diseases: a narrative review of current applications. Clinical Microbiology and Infection DOI 10.1016/j.cmi.2019.09.009

Pivovarov R, Elhadad N (2015) Automated methods for the summarization of electronic health records. Journal of the American Medical Informatics Association 22(5):938–947, DOI 10.1093/jamia/ocv032

Pluchinotta I, Pagano A, Giordano R, Tsoukiàs A (2018) A system dynamics model for supporting decision-makers in irrigation water management. Journal of environmental management 223:815–824, DOI 10.1016/j.jenvman.2018.06.083

Pluchinotta I, Kazakçi AO, Giordano R, Tsoukiàs A (2019) Design theory for generating alternatives in public decision making processes. Group Decision and Negotiation

28(2):341–375, DOI 10.1007/s10726-018-09610-5

Poels PJ, Schermer TR, Schellekens DP, Akkermans RP, de Vries Robbé PF, Kaplan A, Bottema BJ, Van Weel C (2008) Impact of a spirometry expert system on general practitioners' decision making. European Respiratory Journal 31(1):84–92, DOI 10.1183/09031936.00012007

Porat T, Delaney B, Kostopoulou O (2017) The impact of a diagnostic decision support system on the consultation: perceptions of gps and patients. BMC medical informatics and decision making 17(1):79, DOI 10.1186/s12911-017-0477-6

Povyakalo AA, Alberdi E, Strigini L, Ayton P (2013) How to discriminate between computer-aided and computer-hindered decisions: a case study in mammography. Medical Decision Making 33(1):98–107, DOI 10.1177/0272989X12465490

Reider JM (2016) Impact of national policies on the use of clinical decision support. In: Clinical Decision Support Systems, Springer, pp 111–130, DOI 10.1007/978-3-319-31913-1_7

Richard A, Mayag B, Meinard Y, Talbot F, Tsoukiàs A (2018) How AI could help physicians during their medical consultations: An analysis of physicians' decision process to develop efficient decision support systems for medical consultations. In: PFIA 2018, Nancy, France

Rudin C, Radin J (2019) Why are we using black box models in ai when we don't need to? a lesson from an explainable ai competition. Harvard Data Science Review 1(2), DOI 10.1162/99608f92.5a8a3a3d

Sandvig C, Hamilton K, Karahalios K, Langbort C (2016) When the algorithm itself is a racist: Diagnosing ethical harm in the basic components of software. International Journal of Communication 10:19

Shortliffe E (2012) Computer-based medical consultations: MYCIN, vol 2. Elsevier

Shortliffe EH, Cimino JJ (2014) Biomedical informatics, 4th edn. Springer

Simon G, DiNardo CD, Takahashi K, Cascone T, Powers C, Stevens R, Allen J, Antonoff MB, Gomez D, Keane P, et al. (2019) Applying artificial intelligence to address the knowledge gaps in cancer care. The oncologist 24(6):772–782, DOI 10.1634/theoncologist.2018-0257

Sittig DF, Krall MA, Dykstra RH, Russell A, Chin HL (2006) A survey of factors affecting clinician acceptance of clinical decision support. BMC medical informatics and decision making 6(1):6, DOI 10.1186/1472-6947-6-6

Slain T, Rickard-Aasen S, Pringle JL, Hegde GG, Shang J, Johnjulio W, Venkat A (2014) Incorporating screening, brief intervention, and referral to treatment into emergency nursing workflow using an existing computerized physician order entry/clinical decision support system. Journal of Emergency Nursing 40(6):568–574, DOI 10.1016/j.jen.2013.10.007

Taylor B, Dinh M, Kwok R, Dinh D, Chu M, Tang E (2008) Electronic interface for emergency department management of asthma: a randomized control trial of clinician performance. Emergency Medicine Australasia 20(1):38–44, DOI 10.1111/j.1742-6723.2007.01040.x

Tierney WM, Overhage JM, Murray MD, Harris LE, Zhou XH, Eckert GJ, Smith FE, Nienaber N, McDonald CJ, Wolinsky FD (2003) Effects of computerized guidelines for managing heart disease in primary care. Journal of General Internal Medicine 18(12):967–976, DOI 10.1111/j.1525-1497.2003.30635.x

Titano JJ, Badgeley M, Schefflein J, Pain M, Su A, Cai M, Swinburne N, Zech J, Kim J, Bederson J, et al. (2018) Automated deep-neural-network surveillance of cranial images for acute neurologic events. Nature medicine 24(9):1337–1341, DOI 10.1038/

s41591-018-0147-y
Tsai TL, Fridsma DB, Gatti G (2003) Computer decision support as a source of interpretation error: the case of electrocardiograms. Journal of the American Medical Informatics Association 10(5):478–483, DOI 10.1197/jamia.M1279
Tsoukiàs A, Montibeller G, Lucertini G, Belton V (2013) Policy analytics: an agenda for research and practice. EURO Journal on Decision Processes 1(1-2):115–134, DOI 10.1007/s40070-013-0008-3
Van Der Sijs H, Aarts J, Vulto A, Berg M (2006) Overriding of drug safety alerts in computerized physician order entry. Journal of the American Medical Informatics Association 13(2):138–147, DOI 10.1197/jamia.M1809
Verdoorn S, Kwint HF, Hoogland P, Gussekloo J, Bouvy ML (2018) Drug-related problems identified during medication review before and after the introduction of a clinical decision support system. Journal of clinical pharmacy and therapeutics 43(2):224–231, DOI 10.1111/jcpt.12637
Voigt P, Von dem Bussche A (2017) The EU general data protection regulation (GDPR). Springer, DOI 10.1007/978-3-319-57959-7
Wachter RM (2004) The end of the beginning: Patient safety five years after 'to err is human' amid signs of progress, there is still a long way to go. Health affairs 23(Suppl1):W4–534, DOI 10.1377/hlthaff.W4.534
Warner HR, Haug P, Bouhaddou O, Lincoln M, Warner Jr H, Sorenson D, Williamson JW, Fan C (1988) Iliad as an expert consultant to teach differential diagnosis. In: Proceedings of the Annual Symposium on Computer Application in Medical Care, American Medical Informatics Association, p 371
Warner Jr H (1989) Iliad: moving medical decision-making into new frontiers. Methods of information in medicine 28(04):370–372, DOI 10.1055/s-0038-1636792
Watrous RL, Thompson WR, Ackerman SJ (2008) The impact of computer-assisted auscultation on physician referrals of asymptomatic patients with heart murmurs. Clinical Cardiology: An International Indexed and Peer-Reviewed Journal for Advances in the Treatment of Cardiovascular Disease 31(2):79–83, DOI 10.1002/clc.20185
Weingart SN, Toth M, Sands DZ, Aronson MD, Davis RB, Phillips RS (2003) Physicians' decisions to override computerized drug alerts in primary care. Archives of internal medicine 163(21):2625–2631, DOI 10.1001/archinte.163.21.2625
West CP, Dyrbye LN, Shanafelt TD (2018) Physician burnout: contributors, consequences and solutions. Journal of internal medicine 283(6):516–529, DOI 10.1111/joim.12752
Woolf SH (1993) Practice guidelines: a new reality in medicine: Iii. impact on patient care. Archives of internal medicine 153(23):2646–2655, DOI 10.1001/archinte.1993.00410230060008
World Health Organization, et al. (1992) The ICD-10 classification of mental and behavioural disorders: clinical descriptions and diagnostic guidelines. Geneva: World Health Organization
Yanase J, Triantaphyllou E (2019) A systematic survey of computer-aided diagnosis in medicine: Past and present developments. Expert Systems with Applications p 112821, DOI 10.1016/j.eswa.2019.112821
Yang MS, Choi SI, Song WJ, Kim SH, Cho SH, Min KU, Kim JH, Chang YS (2018) Electronic consultation support system for radiocontrast media hypersensitivity changes clinicians' behavior. Allergy, asthma & immunology research 10(2):165–171, DOI 10.4168/aair.2018.10.2.165

Zhang K, Liu X, Liu F, He L, Zhang L, Yang Y, Li W, Wang S, Liu L, Liu Z, et al. (2018) An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study. Journal of medical Internet research 20(11):e11144, DOI 10.2196/11144

Zier J, Spaulding A, Finch M, Verschaetse T, Tarrago R (2017) Improved time to notification of impending brain death and increased organ donation using an electronic clinical decision support system. American Journal of Transplantation 17(8):2186–2191, DOI 10.1111/ajt.14312

Zou J, Schiebinger L (2018) Ai can be sexist and racist—it's time to make it fair. Nature 559:324–326, DOI 10.1038/d41586-018-05707-8

Part II

Studying current processes of physicians

In addition to our analysis of the different possible approaches to support physicians during their practices, we studied the decision processes of physicians of the HCL during their medical consultations. With this study, we aim to better understand interactions between physicians and their patients, but also interactions between physicians and the Health Information System (HIS) they use at the HCL: Easily®.

This work allowed us to propose two models of physicians' decision process during consultations presented in Richard et al. (2018). However, some improvements were made since then, especially on computer analyses of physicians' activities, and further details on the HIS used by physicians can be given.

In Chapter 2, we detail specificities of the HIS Easily®. In Chapter 3, we detail the improvements made on the analyses presented in Richard et al. (2018). Finally, in Chapter 4, we present the paper Richard et al. (2018) which includes the two models we propose, on physicians' decision process.

Chapter 2

Easily®, the information system of the HCL

In this chapter, we present the specificities and functionalities of the HIS used by physicians at the HCL: Easily®. The main objective of this presentation is to better introduce the tools that physicians can access during their consultations, which will be useful to understand the results presented in Chapter 3.

Easily®is a HIS which has been developed, since 2014, by the computer science department of the HCL. Initially dedicated to offering tools to the clinicians of the HCL, Easily®is currently deployed in close to thirty groups of hospitals in France. Figure 2.1 locates the different groups of hospitals using Easily®in 2019. Accordingly, the work proposed in this thesis aims to be applied not only in the HCL, but in all the hospitals using Easily®.

Easily®is based on different "portals" depending on the profession of clinicians using it, such as Nurses, Midwife, Medical Secretary, Pharmacist, Medical Biology, and Physicians, the one that interests us. More specifically, during medical consultations, physicians use a module of Easily®called "*CapMedecin*", giving them access to different functionalities useful for consultations. Figure 2.2 illustrates the user interface of *CapMedecin* used during medical consultations.

The user interface of *CapMedecin* is divided into different parts. The central panel contains two columns, each with several tabs. The left column gives access to the following tab pages:

- *Dossier Spécialités* [Specialty File], giving access to different data on a patient depending on the medical specialty selected;
- *Histoire* [History], the default tab page of *CapMedecin*, giving access to the chronological list of medical documents concerning the patient;
- *Activité* [Activity], giving access to the history of medical acts performed on the patient;
- *Visionneuse Doc* [Document Viewer], displaying medical documents concerning the patient, from the most recent to the oldest.

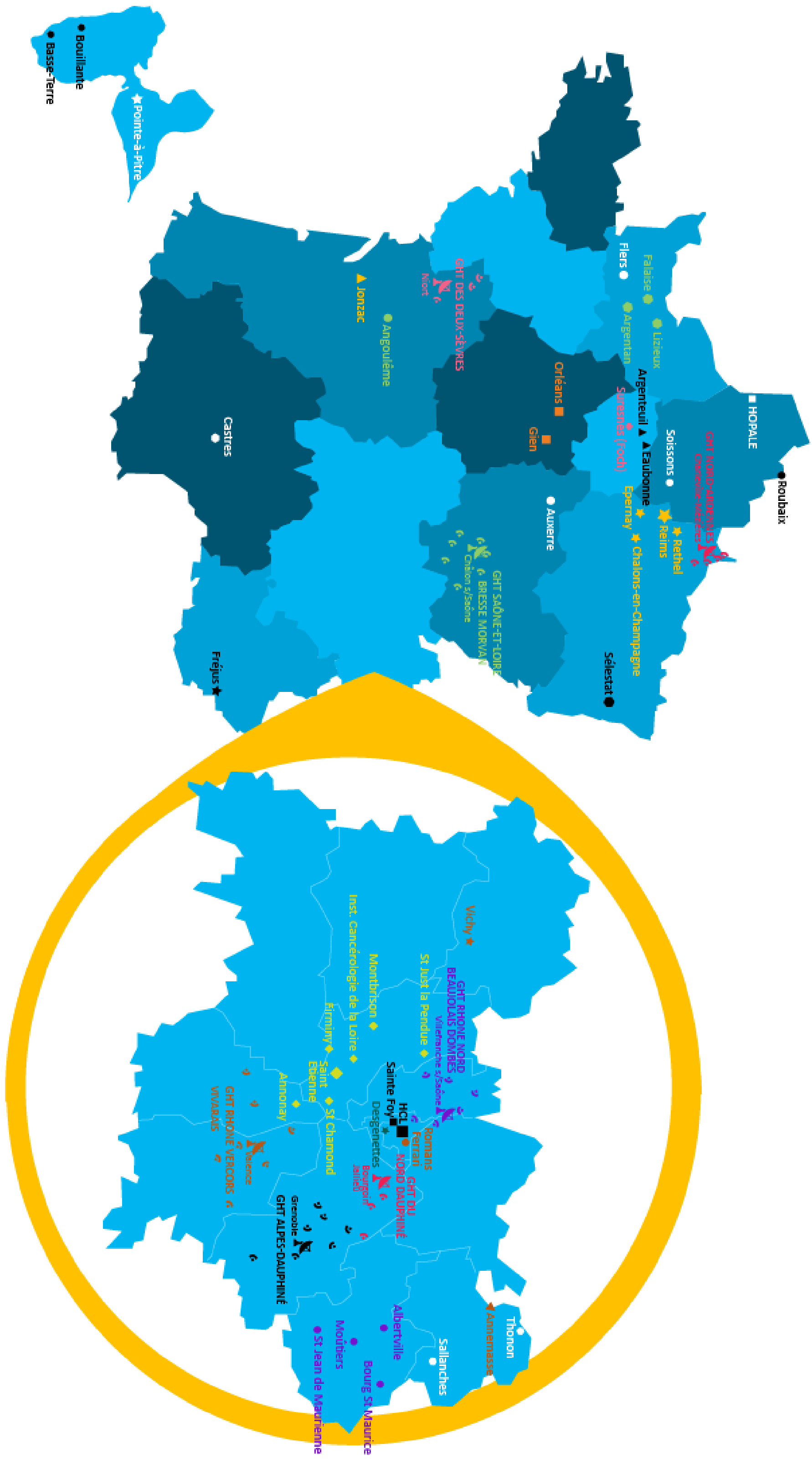


Figure 2.1: Groups of hospitals using Easily® in France in 2019

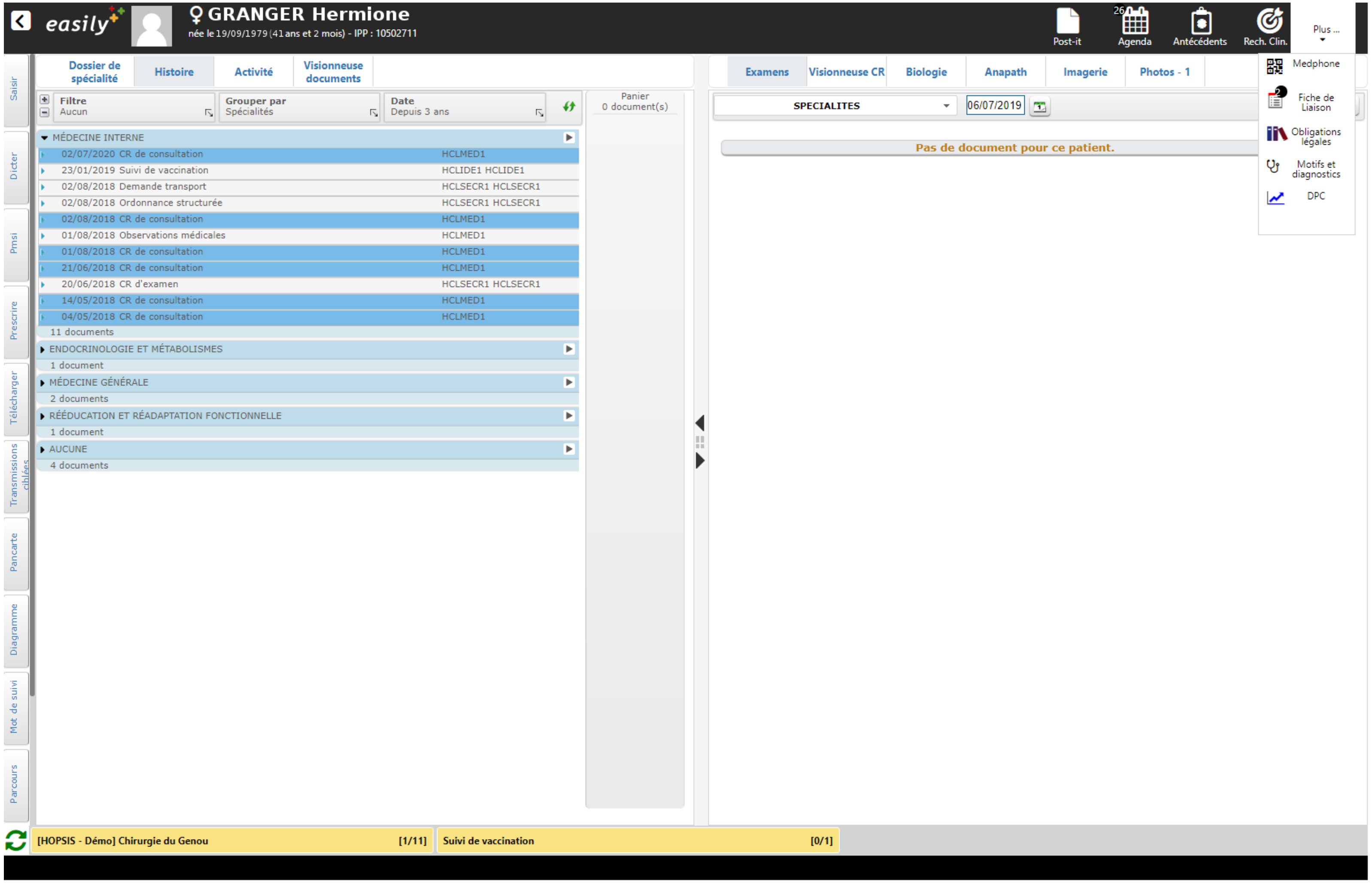


Figure 2.2: Example, for a fictitious patient, of the current user interface of CapMedecin for consultations

The right column is dedicated to displaying results from medical examinations concerning the patient. By default, the right column gives access to the following tab pages:

- *Examens* [Examinations], giving access to the results of medical examinations made in a laboratory;
- *Visionneuse CR* [Report Viewer], giving access to the reports of medical examinations;
- *Biologie* [Biology], giving access to the biological data concerning the patient;
- *Anapath* [Anatomopathology], giving access to the anatomopathological data concerning the patient;
- *Imagerie* [Imagery], giving access to data from medical imagery;
- *Photo* [Pictures], giving access to pictures from the patient's file.

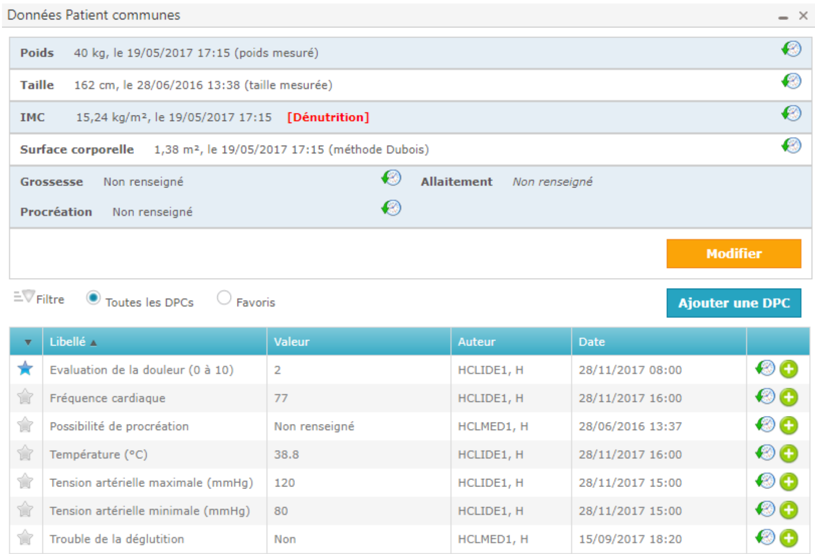
In addition to this central panel, physicians have access to different submodules. Each of these sub-modules gives to physicians accesses to various data on their patients (see Figure 2.3), allows them to record new data on their patients, and/or allows them to write medical documents (see Figure 2.4).

First, we have the sub-modules dedicated to accessing and/or recording data on patients (see Figure 2.2, top-right):

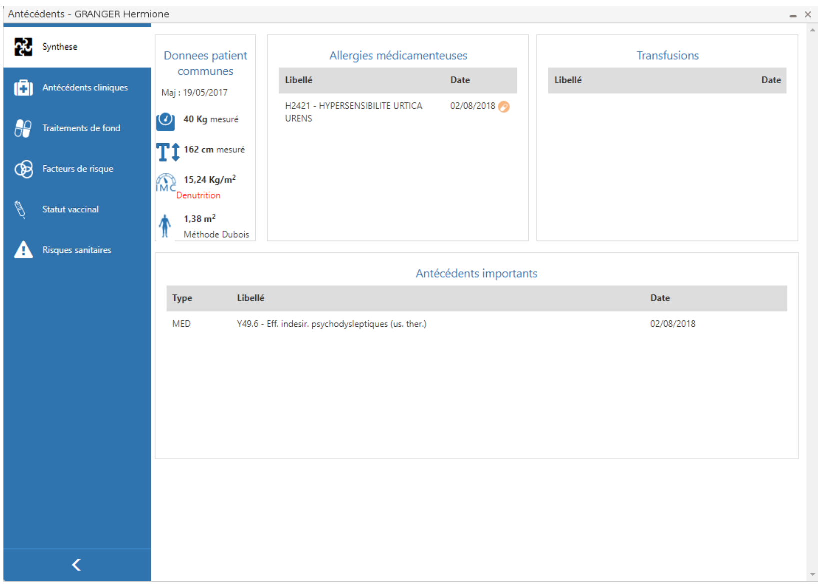
- *Post-it*, allowing to create a note concerning the patient;
- *Agenda*, giving access to all the appointments of the patient;
- *Antécédents* [Medical Background], giving access to all data concerning the medical background of the patient (see Figure 2.3b);
- *Rech. Cli.* [Clinical Research], giving access to all the clinical research concerning the patient;
- *MedPhone*, giving access to the QR code of the patient (useful for mobile application linked to Easily®);
- *Fiches de liaison* [Communication Forms], giving access to messages exchanged between clinicians about clinical pathways followed by the patient;
- *Obligations Légales* [Legal Obligations], giving access to legislative or administrative documents concerning the patient;
- *Motif et Diagnostic* [Reasons and Diagnostics], giving access to the history of reasons why the patient come at the HCL and the history of diagnostics of the patient;
- *DPC*, giving access to common data, or *Données Patient Communes*, on patients (see Figure 2.3a).

Physicians have then access, in just a few clicks, to all the medical data concerning their patients. Lastly, there are sub-modules dedicated to the production of medical documents (see Figure 2.2, left tabs). These modules enable physicians to:

- Produce several kinds of medical documents (see Figure 2.4);
- Make a dictation of medical documents, which are subsequently written down by a medical secretary;
- Register medical acts;
- Create clinical pathways and add the patient to some clinical pathways;
- Access to a computerized physician orders application (see Figure 2.5);
- Access and write data concerning a patient's hospitalization;
- Download documents;
- Transmit data on patients to other departments of the hospital.

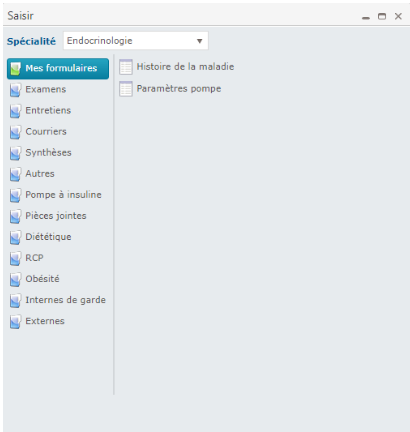


(a) Sub-module DPC

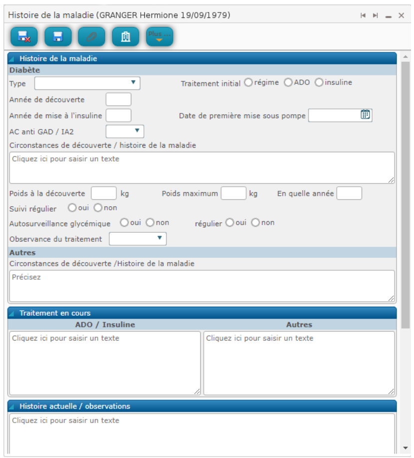


(b) Sub-module Medical Background

Figure 2.3: Example of sub-modules of CapMedecin dedicated to accessing data on a patient



(a) List of writable medical documents



(b) Example of a writing form

Figure 2.4: Example of sub-modules of CapMedecin dedicated to writing of medical documents

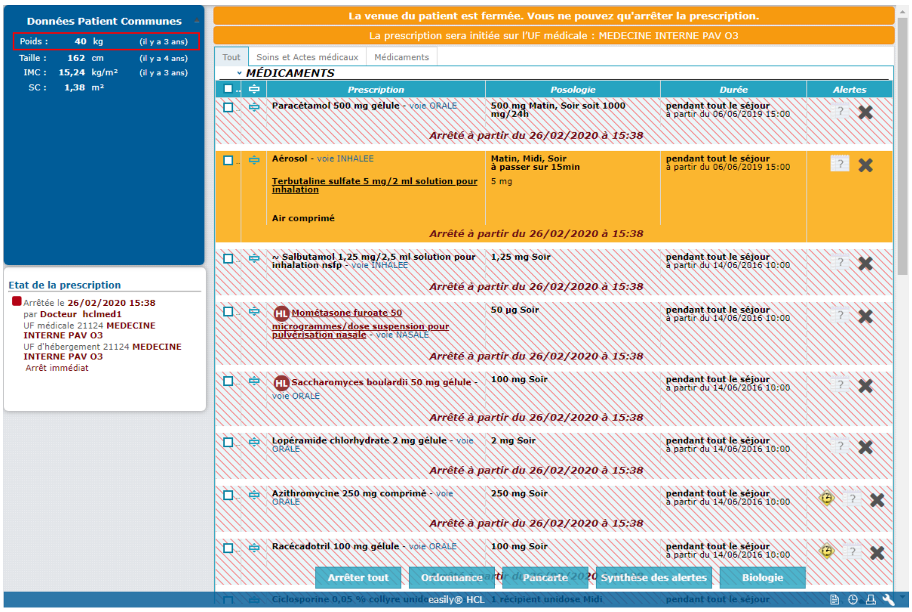


Figure 2.5: User interface of the application "ePrescription" dedicated to writing computerized physicians orders

Chapter 3

Analyzing physicians' workflows

In this chapter, we elaborate on the observations and analyses of physicians' decision processes during medical consultations presented in Richard et al. (2018). These observations and analyses aim to provide us with better understanding of our use case: customary medical consultations at the HCL. We are trying here to understand various aspects of physicians' decision processes during medical consultations:

- What are the different kinds of elementary actions composing the process of medical consultations?
- What are the interconnections between these different actions?
- What are the most frequent actions during consultations?
- Which actions are the most time-consuming to physicians during medical consultations?

Due to the heterogeneity of medical specialties, we focused on consultations made at the HCL's department of endocrinology. In such departments, medical consultations are mainly focused on the follow up of patients with chronic diseases, corresponding to the "customary" aspect we seek to analyze.

Mainly due to the workload of physicians, we limited ourselves to a few field observations, presented in Section 3.1. These field observations allowed us to have an overview of the different kinds of actions that can be made by physicians during medical consultations. However, these few observations are not enough to understand in detail the decision processes of physicians during consultations. In Section 3.2, we use process mining algorithms to extract and analyze the activity of physicians during their medical consultations. Finally, in Section 3.3, we summarize our conclusions from these observations and analyses.

3.1 Field observations of medical consultations

The first step of our analyses was to observe, in real conditions, the decision processes of physicians during medical consultations. As mentioned previously, due to physicians' workload, it was not possible to make large scale field observations. Consequently, we limited ourselves to observing the consultations

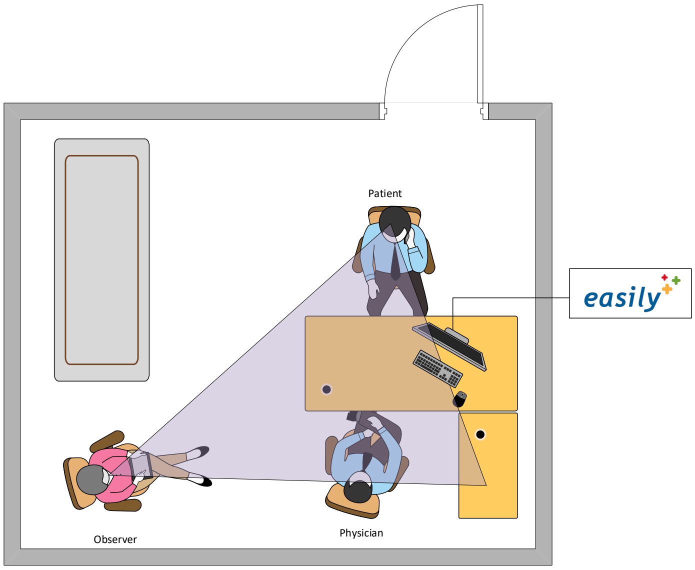


Figure 3.1: Agents' disposition in the consultation room during an observation

of two physicians. Although they are not quantitative, these observations allowed us to catch a glimpse of the actions composing the process of a medical consultation, which will give us a better understanding of the data collected subsequently.

Observation protocol

We consider here that an "observation" corresponds to one, and only one, consultation. An observation starts when a patient arrives in the consultation room and ends when s/he leaves the room. Physicians generally perform several consultations in a row. We call a set of several successive observations "an observation session".

As shown in Figure 3.1, during an observation, the observer is located behind the physician to be able to see her/his interactions with her/his patient and with Easily®, without being in her/his field of vision. The objective is to observe physician-patient interactions, but also physician-HIS interactions.

The following indicators are monitored:

- The consultation's duration (in minutes);
- The number of actions made by the physician;
- The time spent by the physician on Easily®(in minutes).

At the beginning of a new observation, the observer starts by taking note of the hour when the patient comes into the consultation room. Then, during the observation, the observer takes note of each action made by the physician, for example: consulting a medical document, taking a biometric measurement, prescribing a treatment, etc. In parallel, the observer monitors the time spent on Easily® by the physician, considering any interaction with the HIS even if the physician interacted with her/his patient at the same time, for example: searching for data on the patient, recording data about the patient in the system, writing prescriptions of treatments, etc. At the end of the current observation, the observer takes note of the hour when the patient leaves the consultation room and the total time spent by the physician on Easily®. Physicians do not know which data are collected about their activity during observations.

Results

Table 3.1 summarizes, for each observation, the number of pieces of information searched by the physician, the number of prescriptions made, the duration of the consultation, and the time spent on Easily® by the physician.

Observation	Number of data searched	Number of prescriptions	Duration of the consultation	Time spent on Easily®
1	14	1	27’	10’
2	12	3	27’	13’
3	16	1	30’	13’
4	14	3	30’	16’
5	10	1	22’	12’
6	13	4	22’	13’
7	16	0	23’	6’
8	16	10	23’	11’
9	11	5	20’	10’
means	13.6 ± 2.2	3.1 ± 3.1	24.9 ± 3.7	11.6 ± 2.8

(a) Physician 1

Observation	Number of data searched	Number of prescriptions	Duration of the consultation	Time spent on Easily®
1	18	3	20’	10’
2	3	0	10’	7’
3	16	3	16’	8’
4	17	2	25’	8’
5	13	2	17’	9’
6	8	3	19’	9’
7	8	4	18’	12’
8	9	3	12’	7’
means	11.5 ± 5.3	2.5 ± 1.2	17.1 ± 4.7	8.75 ± 1.7

(b) Physician 2

Table 3.1: Summary results of field observations of medical consultations made at the HCL’s department of endocrinology

These observations allow us to better understand some elements of medical consultation processes. Firstly, we identified two kinds of actions that physicians can do during consultations:

- **Searching for a piece of information**: questions asked to the patient, pieces of information collected during the auscultation and data collected from Easily®(this kind of actions is generally associated with the recording of the piece of information into the database of Easily®);

- **Prescribing something**: prescriptions of treatments, prescriptions of analyses in a medical laboratory, and suggestions of appointments with another specialist.

Secondly, the action of searching for a piece of information concerning the current patient appears more frequent than the action of prescribing something. For the first physician, we observe a mean of 13.6 pieces of information searched by consultations for a mean of 3.1 prescriptions by consultations (cf. Table 3.1a). For the second physician (cf. Table 3.1b), we observe a mean of 11.5 pieces of information searched by consultations for a mean of 2.5 prescriptions by consultations.

Thirdly, we can observe that physicians spent close to 50% of the duration of the consultation using Easily®. In the first case (cf. Table 3.1a), we observe a mean duration of consultations of 24.9 minutes, with a mean time spent on Easily®of 11.6 minutes. We can observe, in the second case (cf. Table 3.1b), a mean time spent on Easily®of 8.75 minutes for a mean duration of consultations at 17.1 minutes.

However, let us make two clarifications on this third point:

1. Physicians generally use Easily®while interacting with their current patient, but we cannot make a distinction between the time on Easily®while interacting with the patient and without;

2. Although physicians were not aware of which data were collected about their activity, they knew that they were observed on their use of Easily®, which might have impacted their work process.

Finally, we have observed that prescriptions were generally written at the end of consultations, but that the contents of these prescriptions were specified orally to patients at the moment they were decided by physicians. This last observation echoes the results of Gibson et al. (2006), who showed the central role of "verbal prescriptions" in physician-patient interactions.

3.2 Computer Analysis

As introduced at the beginning of this chapter, the field observations presented in section 3.1 are not enough to understand and to formalize the physicians' decision processes during medical consultations. To improve our understanding of these decision processes, we decided to collect data on physicians' activity during their consultations, to analyze them using process mining algorithms.

As presented by Van Der Aalst et al. (2004, 2007, 2011); Van Der Aalst (2011), process mining algorithms are dedicated to analyzing event logs, collected by an information system of public or private societies, in order to extract processes underlying the functioning of these societies. Therefore, process mining is a subdomain of data mining (Hand, 2007) focused on processes. The data generally used in process mining are not originally stored to be analyzed, but to keep track of users' activities for various reasons.

Rojas et al. (2016) proposed a review of process mining usages in medical contexts. Process mining algorithms can be used to better understand the use of a HIS in a hospital (Rebuge et al., 2013), to discover processes underlying the activity of a hospital (Mans et al., 2008), to evaluate the conformity of a hospital's activities with "gold-standard" guidelines (Kirchner et al., 2012) or to improve work processes by identifying potential bottlenecks (Bose and Van Der Aalst, 2011). According to the authors, processes highlighted by using process mining algorithms can also be used to help the development of future HIS, which is what we aim to do.

Due to regular updates of Easily®, which induce variations in the nature of the data recorded by the HIS, we focused our analyses on event logs corresponding to one month of activity without any notables changes in the nature of the data recorded. We have thereby collected event logs corresponding to the activity of 72 physicians of 5 different HCL departments: endocrinology, dermatology, urology, gynecology, and rheumatology. Close to 40.000 events have been collected, distributed into about 3.500 event logs corresponding to as many medical consultations.

Analyses and results presented in the following sub-sections were performed with the programming language R and using the bupaR library (Janssenswillen, 2020). The source code to reproduce our results is available on the LAMSADE's GitLab[1].

General Analysis

After a first step dedicated to clean irrelevant data from our event logs, we have obtained 59 distinct events. Figure 3.2 illustrates the relative presence, or the frequency, of each event in the various event logs we have collected.

We can find the two types of actions that we observed in real conditions: the search for pieces of information about the current patient ("*Accès Onglet Visionneuse*" [Access to document viewer], "*Accès Onglet Anapath*" [Access to Anapath's tab], "*Lecture des résultats de biologie de ville d'un patient*" [Reading results of biology from city's laboratories for a patient], etc.) and the production of prescriptions ("*Production Document Ordonnance*" [Production of treatment prescription], "*Acces ePrescription externe*" [External access to the ePrescription tool], "*Recherche de tous les rendez-vous d'un patient*" [Search for all the appointments of a patient], etc). Within the "searching for information" category, we can also include events corresponding to data recorded by physicians about their patients. These events are recognizable by the presence of the term "Saisie" at the beginning of their names. Finally, we can also observe events linked to the production of documents other than prescriptions, such as "*CRC*" (the report of the consultation), "*Mot de synthèse*" [A summary word] (a message summarizing the consultation and sent to the general practitioner following the patient), "*Observ. Exam. Clin.*" (the report of a clinical examination), etc. These documents help patients follow-up because they give to physicians a summary of previous consultations concerning the

[1]`https://git.lamsade.fr/a_richard/consultation-process-analysis`

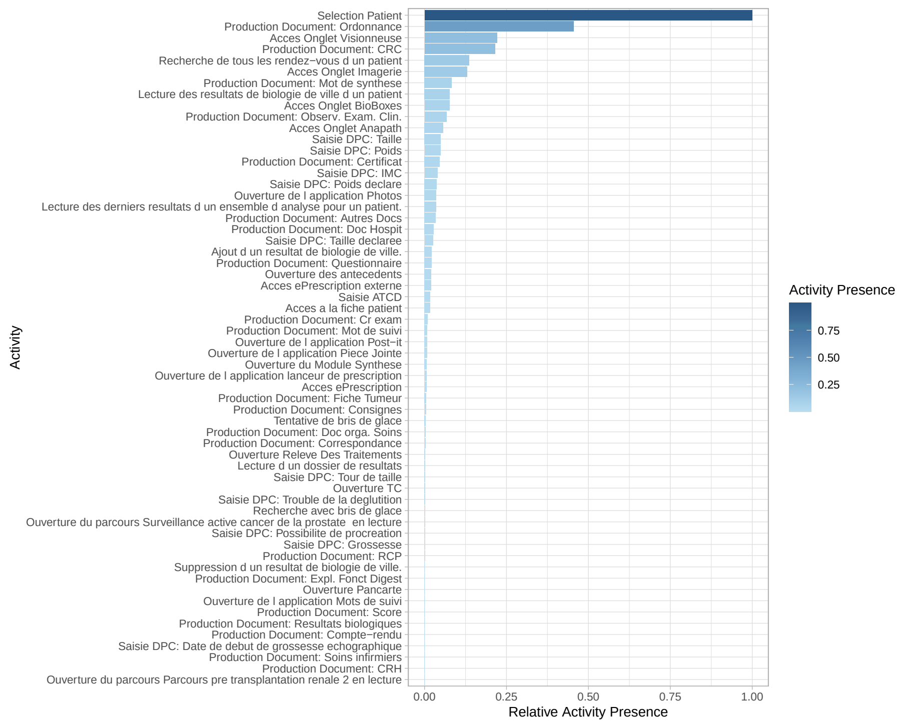


Figure 3.2: Relative presence of events in collected event logs

current patient. This last group of events is hence closer to "searching for information" about a patient than to "prescribing something" to this patient.

Concerning the relative presence of these events in collected event logs, the event "*Sélection Patient*" [Patient Selection] is present in 100% of event logs. This event corresponds to the opening, in Easily®, of a patient file by the physician. It starts more than 99% of collected event logs, but also ends about 38% of event logs, which suggests that physicians only opened the file of the current patient for these consultations.

For events with a relative presence higher than 12.5%, we have:

- "*Production Document: Ordonnance*", present in close to half of the collected event logs;
- "*Accès Onglet Visionneuse*", indicating that the physician opened at least one medical document concerning the current patient;
- "*Production Document: CRC*", corresponding to the production of a consultation report;
- "*Recherche de tous les rendez-vous d'un patient*", indicating that the physician had access to the patient's agenda, suggesting the addition of an appointment for medical follow-up;
- "*Accès Onglet Imagerie*"[Access to medical imagery's tab], indicating that the physician opened at least one report of medical imagery concerning the current patient.

The production of treatment prescriptions appears here more frequently than the search for information. However, if we combine events corresponding to the search for information about the current patient, such as "*Accès Onglet Visionneuse*", "*Accès Onglet Imagerie*" ou "*Accès Onglet BioBoxes*" [Access to biological results], we obtain a relative presence similar to the relative presence of the event "*Production Document: Ordonnance*". Besides, the collected event logs do not report precisely which piece of information physicians look at when they opened these tabs, nor in what quantity. The relative presence of such events hence probably largely higher than observed.

We can also approximate the mean duration of consultations from the time-lag between the start and the end of each event log. However, some logs showed a duration higher than one hour, which suggests that physicians accessed the patient's file several hours after the consultation. Figure 3.3 illustrates the distribution of duration of event logs shorter than 60 minutes.

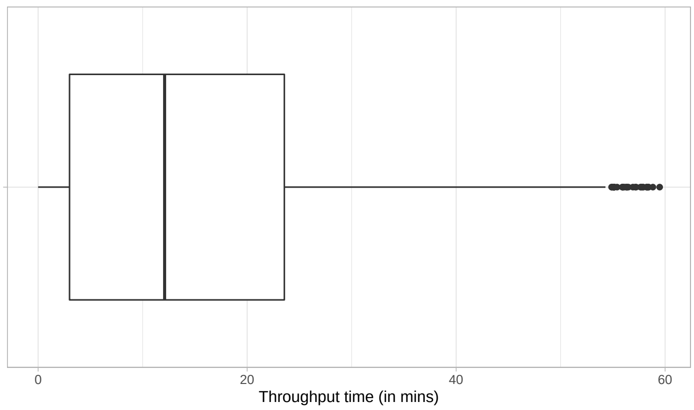


Figure 3.3: Distribution of duration of event logs shorter than 60 minutes

On this basis, the mean duration of consultations can be estimated between 10 and 25 minutes, assuming a margin of at least 5 minutes between the event log's duration and the real duration of the consultation. A consultation ends when the patient leaves the consultation room, and not when the physicians interacted for the last time with Easily®for this consultation.

Finally, we can compute the matrix of connections between the collected events to represent the process of medical consultations. A matrix of connections is built as follow: each event is associated to a row and a column of the matrix. Each cell of the matrix is associated with the number of time the row event is followed by the column event. This matrix allows then to know whether an event A has been followed by an event B, and if so, how many times. It also allows building a graph of connections between the events. Figure 3.4 shows the graph of connections obtained for our collected event logs.

We can observe that the graph obtained contains many intermixing connections. This kind of result corresponds to what Van Der Aalst (2011) called "spaghetti" processes. The study of such processes implies the use of process mining algorithms in order to extract the most relevant processes hidden by all this noise.

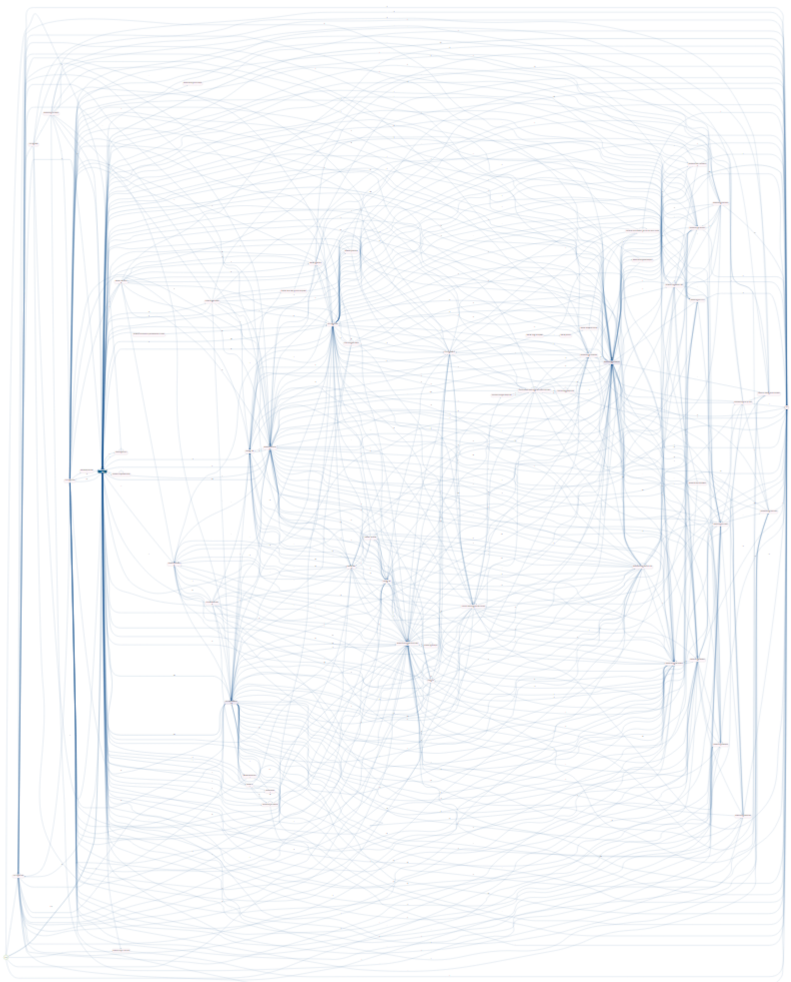


Figure 3.4: Graph of connections between the different events collected

Heuristic Miner

The Heuristic Miner algorithm, proposed by Weijters et al. (2006), is based on the heuristic that the most co-dependent events form the most relevant processes underlying a more general process. The algorithm computes the degree of dependency $a \Rightarrow_W b$, for each couple of events a and b from a set of event logs W, as described in Equation (3.1) (with $|a >_W b|$ the number of times the event a was followed by the event b in W).

$$a \Rightarrow_W b = \frac{|a >_W b| - |b >_W a|}{|a >_W b| + |b >_W a| + 1} \tag{3.1}$$

Once these degrees of dependency are computed, we can build a matrix of dependency connecting all the events, which allows to build a dependency graph. A threshold λ is used to show only the connections associated with a degree of dependency higher than this threshold. Figure 3.5 displays the dependency graph obtained by using the Heuristic Miner algorithm on our collected event logs, with a threshold $\lambda = 0.9$ and by removing events that are not included in a path between the "start" node and the "end" node.

Based on this graph, we can observe that the process of a medical consultation is divided into two distinct phases:

1. A phase dedicated to the search for pieces of information about the current patient, illustrated by the events "*Accès Onglet BioBoxes*", "*Lecture des résultats de biologie de ville d'un patient*", "*Saisie DPC: Poids déclaré*" [Record data: declared weight] and "*Ouverture des antécédents*" [Openning patient's medical background];

2. A phase dedicated to the production of documents, such as prescriptions and reports of consultation, which generally close the consultation's process.

In the second phase, between the access to the "ePrescription" tool and the production of prescription documents, one can observe the event "*Saisie ATCD*" [Record Medical Background], corresponding to the recording of information concerning the medical background of the current patient. It may correspond, as observed in real situations, to the action of asking to the patient if s/he has already taken a similar treatment or if s/he has any contraindications concerning a specific treatment, for examples due to allergies. The presence of this event at the end of the process may reflect a last search for information concerning the patient when the physician writes down her/his final decision concerning the treatment to prescribe.

Some events are co-dependent on themselves, suggesting that actions linked to these events are done repeatedly. This is the case for the events: "*Lecture des résultats de biologie de ville d'un patient*", "*Recherche de tous les rendez-vous d'un patient*", "*Saisie ATCD*", "*Production Document: Ordonnance*", "*Production Document: CRC*".

Lastly, several events can lead to the end of the process. These are events linked to the search for information as well as events linked to the production of documents. If we focus on events with a dependency degree with the "end" node equal to 1, we obtain:

- "*Sélection Patient*", suggesting that, as mentioned in section 3.2, physicians only opened and closed the patient's file during the consultation;

- "*Recherche de tous les rendez-vous d'un patient*", suggesting that physicians ended their consultations by proposing a new appointment to the current patient;

- "*Production Document: Ordonnance*", suggesting that the consultation ended by the production of prescriptions by the physician;

- "*Production Document: CRC*", suggesting that physicians ended the consultation by the redaction of a consultation report, summarizing points discussed and decisions taken during the consultation.

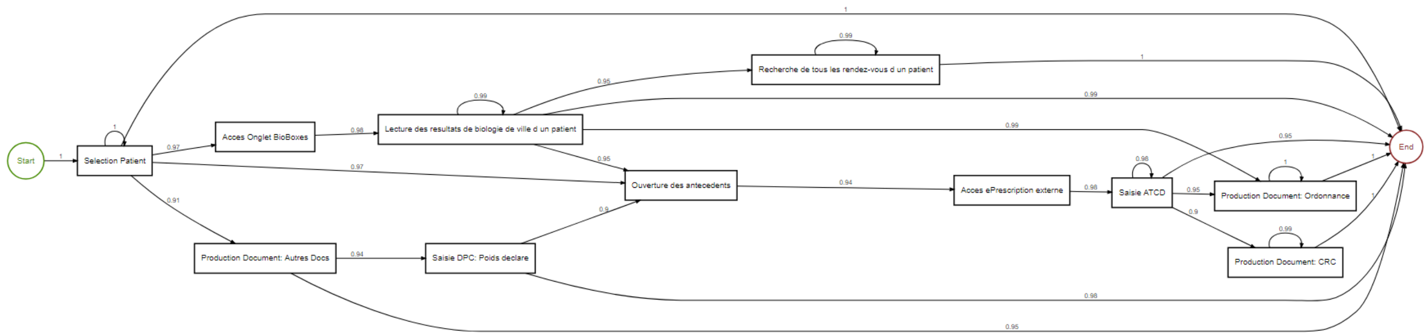


Figure 3.5: The dependency graph obtained by using the Heuristic Miner algorithm on event logs corresponding to 3439 consultations performed during one month by 75 physicians of 5 different HCL's departments

While the ending of a consultation by the event "*Sélection Patient*" is not very informative, this is not the case for the three other situations. The fact that a consultation regularly ends by the scheduling of a new appointment or the production of a summary document, without any signs that physicians searched for further information about the patient after these events, suggests that physicians decided to put an end to their consultations when they though they had covered all possible decisions for the current situation. In other terms, physicians end their consultations when the accumulation of pieces of information about the current patient is no longer possible or no longer useful to choose a prescription to give to the patient.

Fuzzy Miner

To analyze our collected event logs from another perspective, we decided to use another well-known process mining algorithm: Fuzzy Miner. To do so we used the library *fuzzymineR*[2]. This library is based on ProM 6.4.1 (Van Dongen et al., 2005; Verbeek et al., 2010) and allows the use of the Fuzzy Miner algorithm with the programming language R.

The Fuzzy Miner algorithm, proposed by Günther and Van Der Aalst (2007), was initially designed to analyze "spaghetti" processes, such as the one we study. To do so, Fuzzy Miner uses various metrics to measure the "significance" degree of each node and each edge of the graph, but also degrees of connections between the nodes. Based on these metrics and different thresholds, the Fuzzy Miner algorithm clusters or removes nodes and edges as follows:

- Nodes and edges highly "significant" are preserved;
- Nodes and edges not "significant" enough, but strongly "correlated", are clustered in a new node;
- Nodes and edges not "significant" enough and not "correlated" are simply removed from the graph.

Therefore, the higher the signifiance thresholds of significance, the more the nodes and edges are clusterized or removed. As a consequence, the studied graph is highly simplified. Unfortunately, the *fuzzymineR* library does not indicate which nodes are clustered together, which will severely limit our analysis and conclusions.

Figure 3.6 shows the graph obtained by using the Fuzzy Miner algorithm on our collected event logs, with thresholds of significance set at 0.75 for both nodes and edges.

Due to the use of high thresholds, many nodes and edges have been clustered into "cluster" nodes. However, we can still observe a few similarities and differences with the graph obtained by using the Heuristic Miner algorithm on the same event logs. Firstly, consultations' processes start with the selection of the patient's file and end with the production of a consultation's report. This suggests, as mentioned at the end of section 3.2, that physicians end their consultations when they have covered all the possible prescriptions to give to their current patient.

[2]https://github.com/nirmalpatel/fuzzymineR

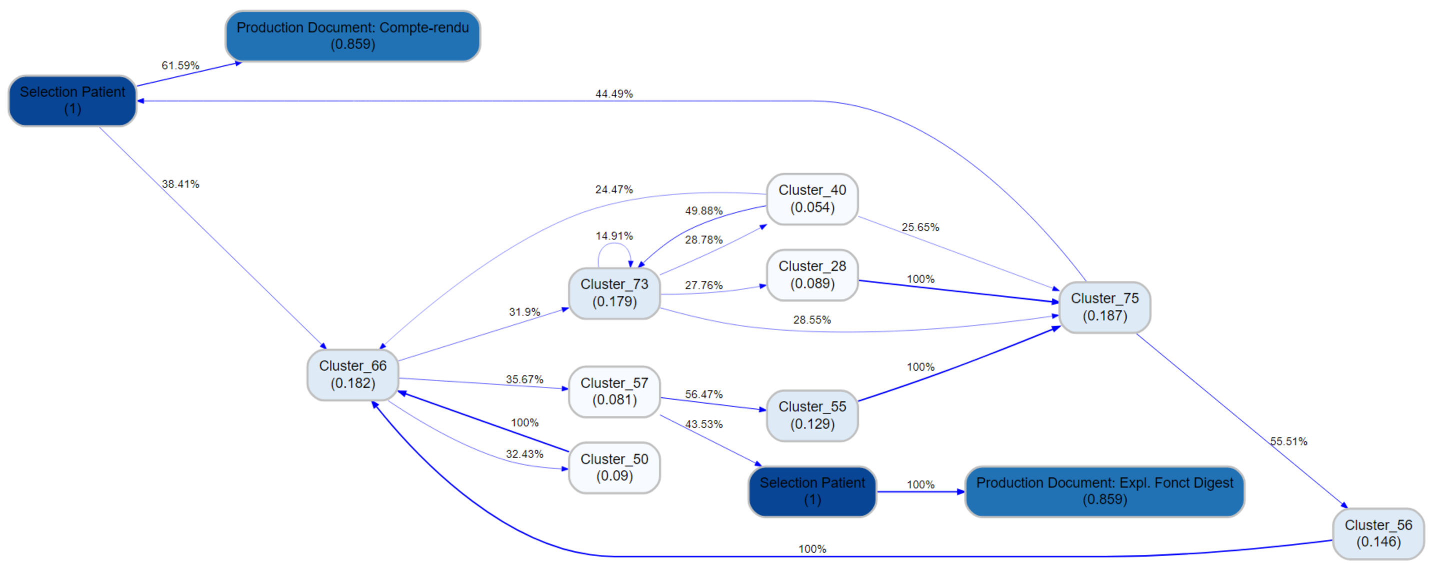


Figure 3.6: The graph obtained by using the Fuzzy Miner algorithm, with thresholds of significance set at 0.75 for both nodes and edges, on our collected event logs

Secondly, as opposed to Heuristic Miner, the Fuzzy Miner algorithm highlighted many cyclical sub-processes. Because the production of prescription documents is unique and the production of report documents is already identifiable, the events gathered in "cluster" nodes correspond to events linked to the search for information about the current patient. This suggests that the process of a medical consultation is based on various internal sub-processes, dedicated to the search for information and varying depending on the type of information researched. For example, the search for biological data doesn't follow the same process as the search for a piece of information concerning the family medical history of the patient. These sub-processes are linked to each other, sometimes in a cyclical way, suggesting that physicians do not search at the same time for all the pieces of the same type of information, but search for pieces of information corresponding to their needs as the consultation unfolds. For example, a physician can search for a biological data, then check if the patient has a family history of her/his disease, then search for another biological data, then search for the result of the last medical scan made by the patient, then check the allergies of the patient, and so on.

3.3 Synthesis

Based on field observations and computer analyses, we are now able to answer the questions raised at the beginning of this chapter. This synthesis of our results will allow us to propose models of the physicians' decision process during their consultations.

Firstly, we have identified, in our field observations as well as in our computer analyses, the elementary actions composing the process of medical consultations. We proposed to regroup these actions into two types: the search for information and the production of prescriptions. The search for information concerning patients is, according to our observations and analyses, the most reccurent type of actions performed by physicians during their consultations.

As analyzed in event logs of physicians performing medical consultations, the production of prescriptions generally puts an end to consultations. On the contrary, the search for pieces of information constitutes the starting part (cf. Figure 3.5) and the central part (cf. Figure 3.6) of medical consultations. These observations suggest that the search for information about the current patient is essential to physicians in order to determine which prescriptions to give to their patients at the end of consultations.

The search for information is generally linked to the entry of these information into the system and the production of documents summarizing the consultation (ex. consultation reports). This type of documents is necessary for physicians to remind themselves of decisions taken during previous consultations. The production of summary documents ends consultation as well as the production of prescriptions. Based on this observation, we surmise that physicians put an end to their consultations when they think they have covered all the possible prescriptions to give to their patients. In other terms, physicians end their consultations when the search for information is no longer useful to decide which prescription to give to the current patient. This highlights the essential aspect of collecting information for physicians during their consultations.

The physicians' working process, during medical consultations, may be compared to an investigating work to determine which decision to take concerning the health of their patients. Physicians accumulate pieces of information, they prune the domain of possible prescriptions and they stop when there is no more ambiguity concerning prescriptions to give for the current consultation.

Chapter 4

Modelling physicians' decision process

How AI could help physicians during their medical consultations: An analysis of physicians' decision process to develop efficient decision support systems for medical consultations

Antoine RICHARD[1,2], Brice MAYAG[1], Yves MEINARD[1], François TALBOT[2], Alexis TSOUKIAS[1]

[1] LAMSADE UMR 7243, Université Paris-Dauphine, France

[2] Hospices Civils de Lyon, France

Résumé : Physicians of the Hospitals of Lyon (HCL) use a Health Information System (HIS) during their medical consultations that gives them access to several functionalities. With the aim to develop Clinical Decision Support Systems (CDSSs) as new functionalities of the HIS, we analyse what kind of tools could be the most useful to help physicians during their day-to-day medical consultations. This paper presents the methods we used to collect data to analyse physicians' decision process during medical consultations. We also present two models of these decision processes and a discussion about the nature of tools that physicians really need during a medical consultation.

Mots-clés : Clinical Decsion Support Systems, Health Information System, Decision Process Analysis

1 Introduction

Physicians at the Hospitals of Lyon (HCL) use a Health Information System (HIS) during their medical consultations. This HIS allows them to access to several functionalities such as : access to information concerning patients, possibility to write down drug prescriptions, enter information about a patient, etc. In order to develop Clinical Decision Support Systems (CDSSs) as new functionalities of this HIS, we want to determine what kind of tools could be the most useful to help physicians during their daily medical consultations. Indeed, following (Heeks, 2006), we admit that a gap between the design of our systems and the reality of physicians' practices can produce the reject of our systems by physicians. To avoid that risk, we have had the opportunity to discuss with physicians of the HCL about their practices, observe some medical consultations and have access to the HCL database.

The paper is organized as follows. In section 2, we review previous analyses of physicians' decision process during medical consultations. In section 3, we explain th several medical consultations, and we explain how we analyse data collected by the HIS during medical consultations by using process mining. In section 4, we present results of process mining and we propose two models to describe physicians' decision process during medical consultations. The first model is a rule-based model that describes the decision process that physicians follow during a specific medical consultation. The second one is a more general model that describes the decision process followed by physicians during medical consultations in general. In section 5, we discuss the revelance of our proposed models and what they show on the physicians' decision process during medical consultations. We then conclude this paper in section 6, by presenting our future work.

2 Related Works

Even through numerous studies have analysed reasoning of physicians during diagnoses, there are only a limited number of papers analyzing the decision process of physicians during medical consultations. According, as an example, to (Rojas *et al.*, 2016), process mining in healthcare context is mostly used to analyse the evolution of patients' treatments or to analyse workflows in hospitals. Medical consultations, in those types of analysis, are often considerate to be a part of a larger clinical process, and not as a decision process itself.

An early analysis was made by (Leaper *et al.*, 1973). They observe 1 307 diagnoses made by 28 physicians, on patients suffering from abdominal pain. One major result of this study is the fact that physicians, for a same patient and a same diagnostic, do not follow the same decision path. They conclude that an absolute decision process of a diagnosis does not exist, and that a diagnostic support system must be adapted to the personal diagnostic process of each physician.

Earlier still, (Taylor *et al.*, 1971) proposed a comparison between physicians' diagnostic process and a computer diagnostic process, and how their diagnosis could diverge. This study was made on 6 physicians who were asked to solve 20 cases of non-toxic goitre. It was not a direct analysis of the diagnostic decision process, but they show in this paper that physicians follow a cyclical decision process.

A recent analysis was proposed by (Rebuge *et al.*, 2013). They used process mining to analyse data collected by the HIS of the Hospital of São João. This study shows the various uses that can be made of the HIS. Unfortunately, an HIS can't log everything that happens during a medical act, and the results are focused on how the HIS is used, not on the decision process of the physician.

3 Methods

In this section, we present how we have collected data for our analysis of the decison process of physicians during medical consultations. We start by some observations of medical consultations, following the protocol described in section 3.1. We then analyse data corresponding to a large set of medical consultations using process mining, as described in section 3.2.

3.1 Observations

Our aim during these observations was not only to understand how physicians interact with patients during a consultation, but also to understand how s/he interacts with the HIS. To that end, we have used indicators such as : the number of information obtained by interacting with the patient (by question or measurement), the number of information obtained by using the HIS, the duration of the medical consultation and the time passed to use the HIS.

For this study we observed eight medical consultations of a specialist in endocrinology. For each medical consultation, a single observer was positioned behind the physician to have a clear view on how s/he used the HIS. The observer took notes each time the physician got an information, registering the type of information and how the physician got it (via HIS or not). The observer also took notes when the physician decided of a prescription (drug prescription, medical laboratory analysis, etc.).

Due to the small number of medical consultations observed, we cannot present meaningful statistics in this paper. However, the HCL logs the activity of physicians in their database. Analysing these logs, as described in section 3.2, allows us to better understand the decision process of physicians during medical consultations.

3.2 Process Mining

The HIS used by physicians collects a lot of data on patients and physicians' activities. These data allow us to better understand how physicians use the HIS. Observations of medical consultations (Cf. section 3.1) allow us to better understand the relevance of data collected by the HIS, and then, help us during the cleaning step of the data analysis. To analyse the data collected by the HIS, we used the HeuristicMiner algorithm, developed by (Weijters *et al.*, 2006). HeuristicMiner is useful to extract an event graph from real data with noise, by using heuristic such as frequency of an event or dependency between several events.

4 Results

We present, in section 4.1, results of process mining. We then propose two models of the physicians' decision process during medical consultations. The first one, developed in section 4.2, is a local rule-based model of the decision process followed by a physician during a specific medical consultation. The second one, developed in section 4.3, is a more general model of the decision process followed by physicians during medical consultation in general.

4.1 Exported process

Figure 1 shows, as an example, the event graph extracted using HeuristicMiner from data collected by the HIS corresponding to 57 medical consultations made by the same physician during a month of work. Each node is associated with its frequency and each edge is associated with the degree of dependency between two events.

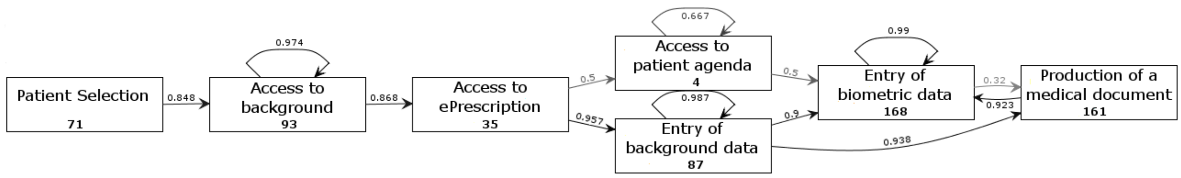


FIGURE *1 – Event graph obtained with HeuristicMiner by analysing 619 events, shared between 57 event logs*

We can see in figure 1 that the physician starts by selecting the patient directory, getting access to her/his medical background and getting access to the application *ePrescription* (the physician uses this application to write her/his drug prescriptions). Then, the physician generally enters some background information about the patient into the system or, sometimes, s/he gets access to the patient's agenda. Then, the physician enters several biometric information about the patient and finishes her/his consultations by producing several medical documents (generally corresponding to drug prescriptions or analysis prescriptions). We can also see that, after producing medical documents, the physician could enter again biometric information about the patient. This could be considered to be a preview of the cyclical decision process that we develop in section 4.3.

4.2 Local rule-based model

A medical consultation could be seen as a set of moments $\mathcal{T} = \{t_0, t_1, \cdots, t_n\}$, where t_i is a moment of the medical consultation when the physician makes a decision $d(t_i)$, and n is the number of decisions taken by the physician. The physician could make two types of decision : request an information (by asking a specific question to the patient, by consulting the HIS or by doing some measurements on the patient) or determine a prescription (drug, medical laboratory analysis, etc.). In fact, a physician could determine a prescription during the diagnostic process and write it only at the end of the consultation (as shown in figure 1). In general, the physician has an access to a set $\mathcal{C}$ of information about the patient, such as height, blood pressure or glycated haemoglobin. $\mathcal{C}_{t_i} \subset \mathcal{C}$ is defined as the set of information about the patient, known by the physician at a moment t_i of the consultation process. An element $c_j(t_i)$ is defined as the content of a piece of information $c_j \in \mathcal{C}$, known by the physician at t_i (ex. $Weight(t_5) = 66.5kg$).

Table 1 shows an example of a decision process followed by a physician, summarized from one of our observations. Each row of table 1 corresponds to a moment t_i (the left column) of

$\mathcal{T}$	Gender	Age	Weight	Height	BMI	HChol	HDL	LDL	TG	$\mathcal{D}$
t_0	♂	55	$\emptyset$	$\emptyset$	$\emptyset$	true	$\emptyset$	$\emptyset$	$\emptyset$	Look at HDL
t_1	♂	55	$\emptyset$	$\emptyset$	$\emptyset$	true	1.1	$\emptyset$	$\emptyset$	Look at LDL
t_2	♂	55	$\emptyset$	$\emptyset$	$\emptyset$	true	1.1	5.53	$\emptyset$	Look at TG
t_3	♂	55	$\emptyset$	$\emptyset$	$\emptyset$	true	1.1	5.53	1.98	Prescribe Ezetrol
t_4	♂	55	$\emptyset$	$\emptyset$	$\emptyset$	true	1.1	5.53	1.98	Look at Weight
t_5	♂	55	66.5	$\emptyset$	$\emptyset$	true	1.1	5.53	1.98	Look at Height
t_6	♂	55	66.5	165	24.43	true	1.1	5.53	1.98	End of Consultation

TABLE *1 – Example of a summarized physician's decision process during a medical consultation, for a patient with hypercholesterolemia*

the process and to the decision $d(t_i)$ made by the physician at this moment (the right column). The other columns correspond to a subset of $\mathcal{C}$. We can see in table 1 that the physician, for each moment t_i, bases her/his decision $d(t_i)$ on the set $\mathcal{C}_{t_i}$. We assume that a physician must follow a set of rules $\mathcal{R}$ to make a decision $d(t_i)$ at t_i, based on the set of information $\mathcal{C}_{t_i}$, as described in equation 1 (with v and w standing the values of $c_j(t_i)$ and $d(t_i)$).

$$\bigwedge_{c_j \in \mathcal{C}_{t_i}} (c_j(t_i) = v) \Rightarrow (d(t_i) = w) \tag{1}$$

According to (Leaper *et al.*, 1973), for the same patient and the same diagnosis, each physician follows her/his own idiosyncratic path. Then, we may assume that each physician has her/his own personal set of rules $\mathcal{R}$, her/his own decision process constructed from experience. However, we can suppose that, even if they do not follow the same path, physicians walk through the same steps, but not in the same order.

4.3 General model

In previous section, we saw that, in practice, physicians base their decisions on an idiosyncratic set of decision rules. This statement is confirmed by (Leaper *et al.*, 1973) : *"the diagnostic process - viewed as a monolithic structure - does not exist. Each clinician has his own pathway to diagnosis"*. However, if we do not take into account the medical specialty of the physician, the type of questions asked, the type of prescriptions made and the pathology of the patient, it seems that a general model of a diagnostic decision process, followed by physicians, exists.

As introduced by (Taylor *et al.*, 1971), physicians generally follow a cyclical decision process. We propose in this paper a general model of this cyclical decision process that physicians seem to follow. Figure 2 shows the general model of the decision process followed by physicians during medical consultations in general. Here, we will make the disctinction between medical diagnoses, about the pathology of the patient, and the prescriptions that the physician gives to the patient (drug prescription, medical laboratory analysis, etc.).

The decision process of medical consultation can be described as follows. According to the current set of information known about the patient : (1) Establish a set of possible medical diagnoses, if the set is empty go to (6), else go to (2). (2) Establish a set of certain medical diagnoses, if the set is empty go to (3), else go to (5). (3) Determine which information could be relevant to establish a diagnostic. (4) Try to obtain information, by asking the patient or making a measurement, add it to set of known information, then go back to (1) (if information

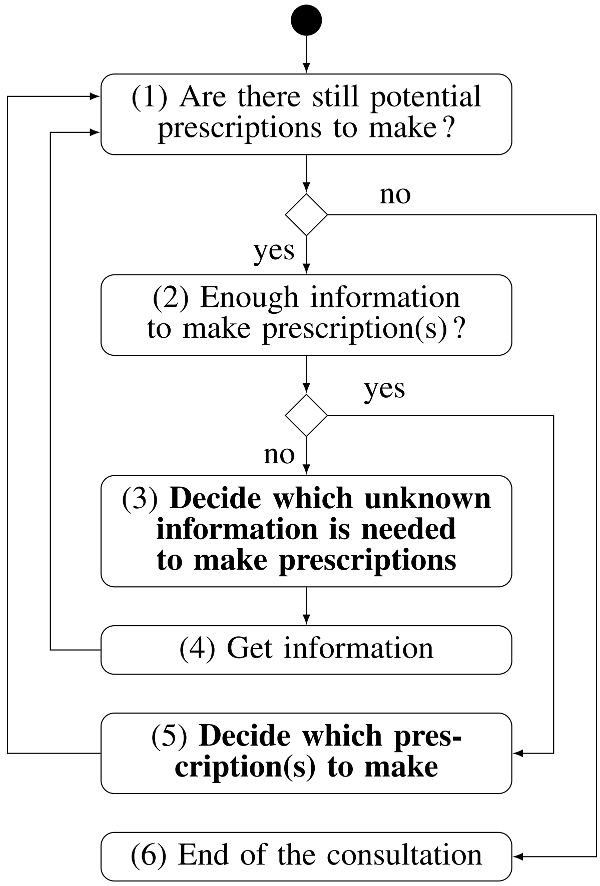


FIGURE 2 – *General model of the decision process of physicians during medical consultations. A graphical representation in the form of a UML activity diagram*

are still unknown, the physician can prescribe a medical laboratory analysis). (5) Determine which prescriptions to make according to the set of certain medical diagnoses, then go to (1). (6) The physician writes down determinate prescriptions and ends the consultation.

5 Discussion

Clinical Decision Support Systems (CDSSs) are healthcare tools that could help decision in various areas, in a whole hospital. (Osheroff *et al.*, 2012) define a CDSS as a system that must : *"provide the right information, to the right person, in the right intervention format, through the right channel, at the right time in workflow to improve health and healthcare decisions and outcomes"*. Unfortunately, as introduced in section 2, medical consultations and diagnostics are often seen as but a part of a larger clinical process, where patients just come with symptoms and physicians give drug prescriptions. Consequently, CDSSs used during medical consultations are often Alert Systems or Diagnostic Decision Support Systems made for a specific domain. (For an overview of DDSSs, see (Miller, 2016)).

However, as we have seen in section 4, a medical consultation is akin to a decision process. It is an investigation process, where the physician searches information about her/his patient to decide which prescription(s) s/he must make. The physician more often decides which piece of information s/he needs to reach quickly a diagnostic, and then formulates a prescription. We can also say that the decision of a prescription is secondary in comparison with the decision of which piece of information is needed. This gap between physicians' practices and CDSSs put at their disposal may explain the *"[. . .] disturbingly high percen-*

tage (i.e., 54-91%) of real-time clinical decision support suggestions are being over-ridden, or ignored, by clinicians" observed by (Sittig *et al.*, 2006), who made a literature review on acceptance of CDSSs by physicians.

Consequently, we assume that *if providing suggestions of prescriptions could be useful in certain circumstances, providing to the physician a selection of information about the patient at the beginning and during a medical consultation could be more helpful for day-to-day practices*. Currently, HISs used by physicians have generally a plethora of information about patients in their databases, and physicians have an access to all the information about the patient, generally at the same time when they open a patient directory. Those information must be summarized and adapted for each medical consultation.

6 Conclusion & Future Work

To conclude, we surmise that developing summarizers of electronic health records (EHR), such as those presented by (Pivovarov & Elhadad, 2015), combined with learning systems and adaptive algorithms, could provide a good daily assistance to physicians during their medical consultations. We also assume that, if we manage to develop AI tools to help physicians, these tools must be as transparent and understandable as possible. Indeed, physicians need to know how the system works to accept it. Also, physicians need to know how to help the system to help them, to reach a cooperative work between physicians and machines, *"to improve health and healthcare decisions and outcomes"*.

In the future, we aim at developing an EHR Summarizer that could be able to learn which information is needed by physicians, and able to adapt its results to information known about the patient at the beginning of a medical consultation (ex. symptoms, background, etc.).

Acknowledgment

This paper was made in collaboration with physicians and employees of the Hospitals of Lyon. Special thanks to Dr. Riou for his advices (and long inspiring discussions) on our work and to Pr. Moulin who allowed us to observe his practices.

Références

HEEKS R. (2006). Health information systems : Failure, success and improvisation. *International journal of medical informatics*, **75**(2), 125–137.

LEAPER D., GILL P., STANILAND J., HORROCKS J. C. & DE DOMBAL F. (1973). Clinical diagnostic process : an analysis. *British medical journal*, **3**(5880), 569.

MILLER R. A. (2016). Diagnostic decision support systems. In *Clinical decision support systems*, p. 181–208. Springer.

OSHEROFF J. A., TEICH J. M., LEVICK D., SALDANA, LUIS ANS VELASCO F. T., SITTIG D. F., ROGEERS K. M. & JENDERS R. A. (2012). *Improving outcomes with clinical decision support : an implementer's guide*. HIMSS.

PIVOVAROV R. & ELHADAD N. (2015). Automated methods for the summarization of electronic health records. *Journal of the American Medical Informatics Association*, **22**(5), 938–947.

REBUGE Á., LAPÃO L. V., FREITAS A. & CRUZ-CORREIA R. (2013). A process mining analysis on a virtual electronic patient record system. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*, p. 554–555 : IEEE.

ROJAS E., MUNOZ-GAMA J., SEPÚLVEDA M. & CAPURRO D. (2016). Process mining in healthcare : A literature review. *Journal of biomedical informatics*, **61**, 224–236.

SITTIG D. F., KRALL M. A., DYKSTRA R. H., RUSSELL A. & CHIN H. L. (2006). A survey of factors affecting clinician acceptance of clinical decision support. *BMC medical informatics and decision making*, **6**(1), 6.

TAYLOR T., AITCHISON J. & MCGIRR E. M. (1971). Doctors as decision-makers : a computer-assisted study of diagnosis as a cognitive skill. *British Medical Journal*, **3**(5765), 35.

WEIJTERS A., VAN DER AALST W. M. & DE MEDEIROS A. A. (2006). Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, **166**, 1–34.

Part III

Proposing a decision support system adapted to physicians' needs

Based on our analysis of physicians' processes at the HCL during their medical consultations and our reflections on current approaches used to support physicians in practice, we decided to propose a decision support system designed to learn and anticipate the data on patients needed by physicians at the beginning of their consultations. This kind of tool should be able to decrease their workload and then perform better healthcare.

In addition, to improve the acceptability of our tools, which is based on a classification system using a machine learning algorithm, we decided to investigate the notion of "transparency" in the literature.

In Chapter 5, we present our work on the notion of transparency for classification systems (Richard et al., 2020a). In this work, we detail requirements in terms of "transparency", and the different criteria associated, that we imposed on ourselves to select the algorithm used in our decision support system. Let us specify that the criterion of "linearity", associated with requirements of "interpretability", have been updated to the notion of "easily reproducible". We also made some experiments to evaluate whether the choice of a "transparent" classification system implies a loss of performance. In Richard et al. (2020a) we present only the results for a global metric, but the results for other metrics leads to the same results: "transparency" does not appear to be correlated with performance. Besides, the classification system we choose, an adapted version of the well-known Naive Bayes algorithm, also produced in Richard et al. (2020a), presented suitable results for simple cases such as ours. The source code to reproduce our experiments is available on LAMSADE's GitLab[1].

In Chapter 6, we present our proposal for a decision support system dedicated to physicians during their medical consultations (Richard et al., 2021). We also present, in this chapter, clinical trials of our decision support system that were made at HCL's department of endocrinology. Although our results should certainly be strengthened by further experiments, they show that our tool is acceptable to physicians.

[1]`https://git.lamsade.fr/a_richard/transparent-performances`

Chapter 5

Improving acceptability by using a "transparent" classification system

Transparency of classification systems for clinical decision support

Antoine Richard[1,3][0000−0001−8677−8910], Brice Mayag[1], François Talbot[2], Alexis Tsoukias[1][0000−0001−5772−3988], and Yves Meinard[1][0000−0001−5928−8959]

[1] Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France
[2] DSII Bron, Hospices Civils de Lyon, 61 Boulevard Pinel, 69672 Bron, France
[3] GIE Hopsis Lyon, 3 Quai de Célestins, 69229 Lyon, France
antoine.richard0@dauphine.eu

Abstract. In collaboration with the Civil Hospitals of Lyon, we aim to develop a "transparent" classification system for medical purposes. To do so, we need clear definitions and operational criteria to determine what is a "transparent" classification system in our context. However, the term "transparency" is often left undefined in the literature, and there is a lack of operational criteria allowing to check whether a given algorithm deserves to be called "transparent" or not. Therefore, in this paper, we propose a definition of "transparency" for classification systems in medical contexts. We also propose several operational criteria to evaluate whether a classification system can be considered "transparent". We apply these operational criteria to evaluate the "transparency" of several well-known classification systems.

Keywords: Explainable AI · Transparency of Algorithms · Health Information Systems · Multi-label Classification

1 Introduction

In collaboration with the Civil Hospitals of Lyon (HCL), in France, we aimed to develop and to propose decision support systems corresponding to the clinicians' needs. In 2018, the HCL received more than one million patients for medical consultations. Therefore, the decision has been made to build a decision support system focused on supporting physicians during their medical consultations. After some observations and analyses of medical consultations in the endocrinology department of the HCL [31], we drew two conclusions: physicians mainly need data on patients to reach diagnoses, and getting these data from their information system is quite time-consuming for physicians during consultations. To reduce physicians' workload, we decided to support them by using a classification system learning which data on patients physicians need in which circumstance. By doing this, we should be able to anticipate and provide the data that physicians will need at the beginning of their future consultations.

This can be formalized as a multi-label classification problem, as presented in Table 1 with fictitious data.

In this paper, a ”classification system” refers to the combination of a ”learning algorithm” and the ”type of classifier” produced by this learning algorithm. For example, a classification system based on decision trees can use a learning system such as C4.5 [30], the type of classifier produced by this learning system being a decision-tree. This distinction is necessary because a learning algorithm and a classifier produced by this learning algorithm are not used in the same way and do not perform the same functions.

X: data known on patient				Y: data on patient needed by physician					
Sex	Age	BMI	Disease	HbA1c	Blood Sugar	HDL	LDL	Creatinine	Microalbumin
♀	42	34.23	DT2	1	1	0	0	0	0
♂	52	27.15	HChol	0	0	1	1	0	0
♂	24	21.12	DT1	1	1	0	0	1	1
♀	67	26.22	HChol	0	0	1	1	0	0

Table 1. Example of multi-label dataset based on our practical case

However, in the case of clinical decision support systems (CDSSs), a well-known problem is the lack of acceptability of support systems by clinicians [5, 19]. More than being performant, a CDSS has first to be accepted by clinicians, and ”transparent” support systems are arguably more accepted by clinicians [22, 33]. Mainly because ”transparency” allows clinicians to better understand the proposals of CDSSs and minimize the risk of misinterpretation. Following these results, we posit that the ”transparency” of support systems is a way to improve the ”acceptability” of CDSSs by clinicians.

In the literature, one can find several definition of the concept of ”transparency”: *”giving explanations of results”* [9, 10, 15, 20, 26, 28, 33, 36], *”having a reasoning process comprehensible and interpretable by users”* [1, 11, 12, 24, 27, 34], *”being able to trace-back all data used in the process”* [2–4, 16, 40], but also *”being able to take into account feedbacks of users”* [7, 40]. Individually, each of the above definitions highlights an aspect of the concept of ”transparency” of classification systems, but do not capture all aspects of ”transparent” classification systems in our context. In addition, definitions are abstract descriptions of concepts and there is a lack of operational criteria, in the sense of concrete properties one can verify in practice, to determine whether a given algorithm deserves to be called ”transparent” or not.

The main objective of this paper is to propose a definition of transparency, and a set of operational criteria, applicable to classification systems in a medical context. These operational criteria should allow us to determine which classification system is ”transparent” for users in our use case. Let us specify that, in this paper, the term ”users” refers to physicians.

In section 2 we detail the definition and operational criteria we propose to evaluate the transparency of classification systems. In section 3, based to our definition of transparency, we explain why we choose a version of the naive bayes algorithm to handle our practical case. We briefly conclude in section 4, with a discussion on the use of an evaluation of "transparency" for practical use cases.

2 Definition of a "transparent" classification system

Even though the concept of algorithm "transparency" is as old as recommendation systems, the emergence and the ubiquity of "black-box" learning algorithms nowadays, such as neural networks, put "transparency" of algorithms back in the limelight [14]. As detailed in section 1, numerous definitions have been given to the concept of "transparency" of classification systems, and there is a lack of operational criteria to determine whether a given algorithm deserves to be called "transparent" or not.

In this paper, we propose the definition below, based on definitions of "transparency" in the literature. Let us recall that our aim here is to propose a definition, and operational criteria, of what we called a "transparent" classification system in a medical context with a user-centered point-of-view.

Definition 1. *A classification system is considered to be "transparent" if, and only if:*

- *the classification system is* ***understandable***
- *the type of classifier and learning system used are* ***interpretable***
- *results produced are* ***traceable***
- *classifiers used are* ***revisable***

2.1 Understandability of the classification system

Although transparency is often defined as *"giving explanations of results"*, several authors have highlighted that these explanations must be "understandable", or "comprehensible", by users [12, 26, 33]. As proposed by Montavon [28], the fact that something is "understandable" by users can be defined as its belonging to a domain that human beings can make sense of.

However, we need an operational criterion to be sure that users can make sense of what we will provide them. In our case, users being physicians, we can consider that users can make sense of anything they have studied during their medical training. Therefore, we define as "understandable" anything based on notions/concepts included in the school curriculum of all potential users. Based on this operational criterion, we propose the definition below of what we call an "understandable" classification systems.

Definition 2. *A classification system is considered to be understandable by users if, and only if, each of its aspects is based on notions/concepts included in the school curriculum of all potential users.*

Let us consider a classification system based on a set C of notions/concepts, and a set S of notions/concepts included in the school curriculum of all potential users, such than $S \cap C$ can be empty. Defined like this, the "understandability" of a classification system is a continuum extending from $S \cap C = \emptyset$ to $S \cap C = C$.

2.2 Interpretability of classifiers and learning system

According to Spagnolli [34], the aim of being "transparent" is to ensure that users are in a position to make informed decisions, without bias, based on the results of the system. A classification system only "understandable" does not prevent misinterpretations of its results or misinformed decisions by users. Therefore, to be considered "transparent" a classification system must also be "interpretable" by users. The criterion of "interpretability" is even more important when applied to sensitive issues like those involved in medical matters. But what could be operational criteria to establish whether a classification system is "interpretable" or not by users?

Let us look at the standard example of a classification system dedicated to picture classification [17]. In practice, the user will use the classifier produced by the learning algorithm and not directly the learning algorithm. Therefore, if the user gives a picture of an animal to the classifier and the classifier says "it's a human", then the user can legitimately ask "Why did you give me this result?" [33]. Here, we have two possibilities: the classifier provides a good classification and the user wants to better understand the reasons underlying this classification, or the classifier provides a wrong classification and the user wants to understand why the classifier didn't provide the right classification.

In the first case, the user can expect "understandable" explanations on the reasoning process that conducted to a specific result. Depending on the classifier used, explanations can take different forms such as "because it has clothes, hair and no claws" or "because the picture is similar to these others pictures of humans". In addition, to prevent misinterpretations, the user can also legitimately wonder "To what extent can I trust this classification?" and expect the classifier to give the risk of error of this result.

In the second case, the user needs to have access to an understandable version of the general process of the classifier and not only the reasoning process that conducts to the classification. This allows the user to understand under which conditions the classifier can produce wrong classifications. In addition, the user can legitimately wonder "To what extent can I trust this classifier in general?". To answer this question, the classifier must be able to provide general performances rates such as its error rate, its precision, its sensitivity and its specificity.

Based on all the above aspects, we are now able to propose the following definition of the "interpretability" of the type of classifier used in the classification system.

Definition 3. *A type of classifier is considered to be "interpretable" by users if, and only if, it is able to provide to users:*

- *understandable explanations of results, including :*
  - *the reasoning process that conducts to results*
  - *the risk of error of results*
- *an understandable version of its general process*
- *its global error, precision, sensitivity and specificity rates*

Nevertheless, although the classifier can answer the question "Why this result?", it will not be able to answer if the user asks, still to prevent a potential misinterpretation, "How the process of classification have been built? Where does it come from?". Only the learning algorithm used by the classification system can be able to bring elements of a response to users because the function of the learning system is to build classifiers, whereas the function of classifiers is to classify.

Therefore, a "transparent" classification system must be based on a type of classifier "interpretable", as defined in Definition 3, but it must also use an "interpretable" learning algorithm, still to ensure that users are in a position to make informed decisions. A first way to establish whether a learning algorithm is "interpretable" could be to evaluate if users can easily reproduce the process of the algorithm. However, evaluating "interpretability" in this way would be tedious for users. We have then to establish operational criteria of learning algorithms that can contribute to its "interpretability" by users.

First, the more linear it is, the more reproducible it is by users. However, linearity alone is not enough to allow "interpretability". For example, this is the case if the various steps of the algorithm fail to be understandable by users or if branching and ending conditions are not understandable by users. Accordingly, we proposed the following definition of the "interpretability" of a learning algorithm.

Definition 4. *A learning algorithm is considered to be "interpretable" by users if, and only if it has:*

- *a process as linear as possible*
- *understandable steps*
- *understandable branching and ending conditions*

The use of concept such as "possibility" of the algorithm implies that we cannot tell that a learning algorithm is absolutely "interpretable". By corollary, the assessment algorithm's "interpretability" is quite subjective and dependent on what we consider as "possible" in terms of linearity for an learning algorithm.

2.3 Traceability of results

Another aspect we have to take into account is the capacity to traceback data used to produce a specific classification. As introduced by Hedbom [18], a user has the right to know which of her/his personal data are used in a classification system, but also how and why. This is all the more true in medical contexts, where the data used are sensitive.

The "understandability" and "interpretability" criteria alone are not enough to ensure the ability to traceback the operations and data used to produce a given result. For example, let us suppose we have a perfectly understandable and interpretable classification system, if this system does some operations randomly, it becomes difficult to traceback operations made from a given result.

By contrast, if a classification system is totally "understandable" and "interpretable", the determinism of classifiers and the learning system is a necessary and sufficient condition to allow "traceability". We can then propose the following definition of the traceability of results.

Definition 5. *The results of a classification system are considered to be "traceable" if, and only if, the learning system and the type of classifier used have a non-stochastic process.*

2.4 Revisability of classifiers

Lastly, the concept of "transparency" can be associated with the possibility for users to make feedbacks to the classification system to improve future results [40]. When a classification system allows users to make feedbacks that are taken into account, this classification system appears less as a "black-box" system to users.

For example, in the medical context, Caruana et al. [7] have reported that physicians had a better appreciation of a rule-based classifier than of a neural network, in the case of predicting pneumonia risk and hospital readmission. This is despite the fact that neural network had better results than the rule-based classifier. According to the authors, the possibility to modify directly wrong rules of the classifier played a crucial role in the preference of physicians.

However, not all classifiers can be directly modified by users. Another way to take account of users' feedbacks is to use continuous learning algorithms (or online learning). The majority of learning algorithms are offline algorithms, but all can be modified, more or less easily, to become online learning algorithms. In that case, the classifier is considered to be partly "revisable". We then obtain the following definition of "revisability" of the type of classifier used by a classification system.

Definition 6. *A type of classifier used by a classification system is considered to be "revisable" by users if, and only if, users can directly modify the classifier's process or, at least, the learning algorithm can easily become an online learning algorithm.*

3 Evaluation of different classification systems

In this section, we use the operational criteria we have established in section 2 to evaluate the degree of "transparency" of several well-known classification systems. With this evaluation, we aim to determine whether one of these classification systems can be used in our use case, from a "transparency" point of view.

We also evaluate the performances of these algorithms on datasets similar to our use case, to evaluate the cost of using a "transparent" alogrithm in terms of performances.

3.1 "Transparency" evaluation

Our evaluation of "transparency" has been made on six different classification systems. The BPMLL algorithm (based on artificial neural networks) [42], the MLkNN algorithm (based on k-Nearest Neighbors) [41], the Naive Bayes algorithm (producing probability-based classifiers) [23], the C4.5 algorithm (producing decision-tree classifiers) [30], the RIPPER algorithm (producing rule-based classifiers) [8] and the SMO algorithm (producing SVM classifiers) [29, 25].

Fig. 1 displays a summary of the following evaluation of our different classification systems. Due to their similarities in terms of "transparency", C4.5 and RIPPER algorithms have been considered as the same entity.

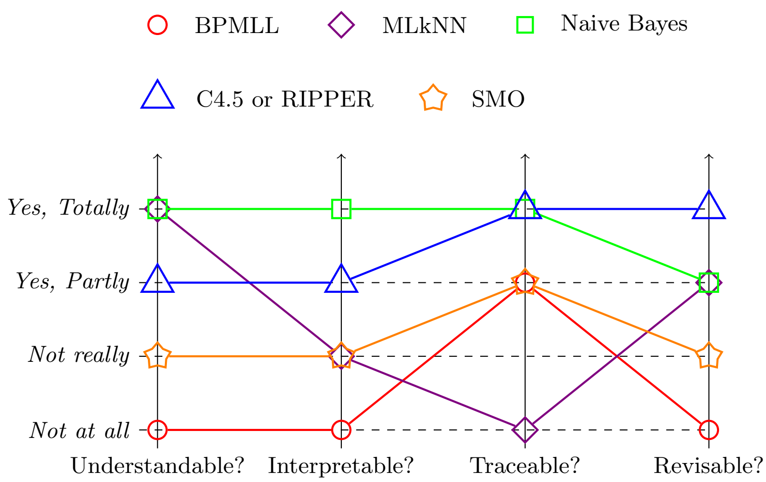


Fig. 1. Graphical representation of the potential "transparency" of different classification systems according to our operational criteria.

Let us start with the evaluation of a classification system based on the BPMLL algorithm [42] (red circles in Fig. 1). The BPMLL algorithm is based on a neural network and neural networks are based on notions/concepts that are not included in the school curriculum of users such as back-propagation and activation functions. Therefore, the steps of the BPMLL algorithm, as well as its branching/ending conditions, cannot be considered to be "understandable" by users. In addition, the learning process of neural networks is not what might be called a linear process. Accordingly, we cannot consider this classification system to be "understandable" and "interpretable" by users. However, neural networks are generally determinist but, due to their low "understandability", they can only be considered to be partly "traceable". Finally, concerning the "revisability" of such a classification system, users cannot directly modify a wrong part of the classifier process and neural networks are not really adapted to continuous learning due to the vanishing gradient problem [21].

The ML-KNN algorithm [41] (violet diamonds in Fig. 1) is considered to be fully "understandable" because it is based on notions like distances and probabilities. Classifiers produced by the ML-KNN algorithm can produce explanations such as "x is similar to this other example". However, due to nested loops and advanced use of probabilities, the learning algorithm does not fit our criteria of "interpretable". In addition, the k-Nearest Neighbors algorithm [13], used by ML-KNN, is generally not determinist which makes the classification system not "traceable". Nevertheless, although classifiers produced by the ML-KNN algorithm cannot be directly modified by users, ML-KNN can easily be modified to become online learning. Consequently, it is partly "revisable".

The Naive Bayes algorithm [23] (green squares in Fig. 1) is considered to be fully "understandable" because, in our context, probabilities and the Bayes theorem are included in the school curriculum of all potential users. The Naive Bayes algorithm is also quite linear and all its steps, as well as its branching/ending conditions, are "understandable". Accordingly, the Naive Bayes algorithm is considered to be fully "interpretable" by users. In addition, the Naive Bayes algorithm is fully determinist, so considered to be fully "traceable". Lastly, users cannot easily modify the classifier, because its a set of probabilities, but the Naive Bayes algorithm can update these probabilities with users' feedbacks, becoming an online learning algorithm. The Naive Bayes algorithm is then considered to be partly "revisable".

The C4.5 and RIPPER algorithms are considered to be partly "understandable" because, even though decision trees or rulesets are notions fully "understandable" by users, these two learning algorithms are based on the notion of Shannon's entropy [32], a notion that is not included into the school curriculum of all potential users. With the same logic, even though decision trees or rulesets are fully "interpretable" classifiers, these learning algorithms are quite linear but their steps and branching/ending are not "understandable" by users because based on Shannon's entropy. The only difference between C4.5 and RIPPER could be on the linearity of their learning algorithm, because RIPPER may be considered to be less linear than C4.5, so less "interpretable". Accordingly, C4.5 and RIPPER are considered to be partly "interpretable" by users. In addition, the C4.5 and RIPPER algorithms are determinists, so fully traceable, and they are considered to be fully "revisable", because users can modify directly classifiers such as decision trees or rulesets.

Lastly, concerning the SMO algorithm, it is mainly based on mathematical notions, such as a combination of functions, that are not necessarily included in the school curriculum of all potential users. The SMO algorithm is not considered to be really "understandable" and "interpretable" by users. The SMO algorithm is determinist but, due to its low "interpertability" it could be more diffcult to traceback its results. It is then considered to be partly "traceable". In addition, the SMO algorithm can become online [35], but not as easily as ML-kNN or Naive Bayes algorithms (for example), it is not considered to be really "revisable".

Consequently, if we start from the classification system with less operational criteria of "transparency" checked, to the classification system with a majority of operational criteria checked, we obtain: BPMLL, SMO, MLkNN, RIPPER, C4.5 and Naive Bayes. Accordingly, a classification system based on the Naive Bayes algorithm can be considered as the best alternative, from a "transparency" perspective, to treat our medical use case.

3.2 Naive Bayes algorithm for multi-label classification

As developed in section 3.1, the Naive Bayes algorithm can be considered to be "transparent" according to our operational criteria. A common way to apply a one-label classification system to a multi-label classification problem, like in our case, is to use the meta-learning algorithm RAkEL [37]. However, the use of RAkEL, which is stochastic and combine several classifiers, makes classification systems less "interpretable" and "traceable". We proposed then a version of the Naive Bayes algorithm, developed in Algorithm 1, to treat directly multi-label classification problems staying as "transparent" as possible.

Algorithm 1: A Naive Bayes algorithm for multi-label classification

Data: a learning dataset $\mathcal{I}$, a set of variables $\mathcal{X}$ and a set of labels $\mathcal{Y}$
Result: sets of approximated probabilities $P_{\mathcal{Y}}$ and $P_{\mathcal{X}|\mathcal{Y}}$

`// Computing subsets of numerical variables`
1 **foreach** *variable* $X \in \mathcal{X}$ **do**
2   Discretize domain of X according to its values in $\mathcal{I}$

`// Counting occurences of` $\mathcal{Y}$ `and` $\mathcal{X} \cap \mathcal{Y}$
3 **foreach** *instance* $I \in \mathcal{I}$ **do**
4   **foreach** *label* $Y \in \mathcal{Y}$ **do**
5         $y_I \leftarrow$ value of Y for instance I
6   Increment by one the number of occurences of $Y = y_I$
7   **foreach** *variable* $X \in \mathcal{X}$ **do**
8             $t_X^I \leftarrow$ the subset of X corresponding to its value in instance I
9   Increment by one the number of occurences of $Y = y_I \cap X = t_X^I$

10 Compute probabilities $P_{\mathcal{Y}}$ and $P_{\mathcal{X}|\mathcal{Y}}$ from computed number of occurences
11 **return** $P_{\mathcal{Y}}$ *and* $P_{\mathcal{X}|\mathcal{Y}}$

To treat numerical variables, the first step of our algorithm is to discretize these numerical variables into several subsets (Algorithm 1, line 2). Discretizing numerical variables allows us to treat them as nominal variables. For each instance of the learning dataset, we get the subset corresponding to the value of each variable for the instance (Algorithm 1, line 8). Then, our algorithm counts occurences of each value of label and variables, and computes their frequency of occurence.

To discretize numerical variables, we first decided to use the fuzzy c-means clustering algorithm [6]. The fuzzy c-means allows to determine an "interpretable" set of subsets T_X of a variable X based on the distribution of observed values in this variable domain. Therefore, the subset t corresponding to a new value $x \in X$ is the subset $t \in T_X$ with the highest membership degree $\mu_t(x)$ (Equation 1).

$$t_X \leftarrow \underset{t \in T_X}{\arg\max}\ \mu_t(x) \tag{1}$$

However, we see here that the use of the fuzzy c-means algorithm requires introducing new concepts such as fuzzy sets, membership functions and membership degrees [39]. These concepts are not included into the school curriculum of users, reducing the "transparency" of the classification systems.

Therefore, we propose to use another discretizing method, more "transparent". This method, inspired by histograms, consists in splitting the variable domain into n subsets of equal size. Therefore, the subset t corresponding to a new value $x \in X$ is the subset $t \in T_X$ such as $min(t) \leq x < max(t)$. This method was preferred due to its simplicity and its potential better "transparency".

3.3 The search for a right balance between performances and transparency

Now that we have evaluated the "transparency" of several classifier systems, and we have identified the Naive Bayes algorithm as the most "transparent" alternative in our context, a question still remains: Does "transparency" have a cost in terms of performances?

To answer this question we evaluated classifiers presented at the beginning of this section on performance criteria for different well-known multi-label datasets and a dataset named *consultations* corresponding to our use case. Table 1 is an example based on this dataset. Currently, our dataset contains 50 instances with 4 features (patients' age, sex, BMI and disease) and 18 labels corresponding to data potentially needed by endocrinologists during consultations.

Our aim in this sub-section is to determine if the use of our version of the Naive Bayes algorithm offers suitable performances in our use case. If this is not the case, we won't have the choice but to envisage using a less "transparent" algorithm if it offers better performances.

These evaluations were made by using the Java library Mulan [38], which allowed to use several learning systems and cross-validation metrics. The program to reproduce these evaluations can be found on the GitLab of the LAMSADE[4].

Fig. 2 shows the distribution of macro-averaged F-measures of classifier systems computed for different multi-label datasets. The F-measure is a harmonic mean of the precision and the recall of evaluated classification systems. These

[4] https://git.lamsade.fr/a_richard/transparent-performances

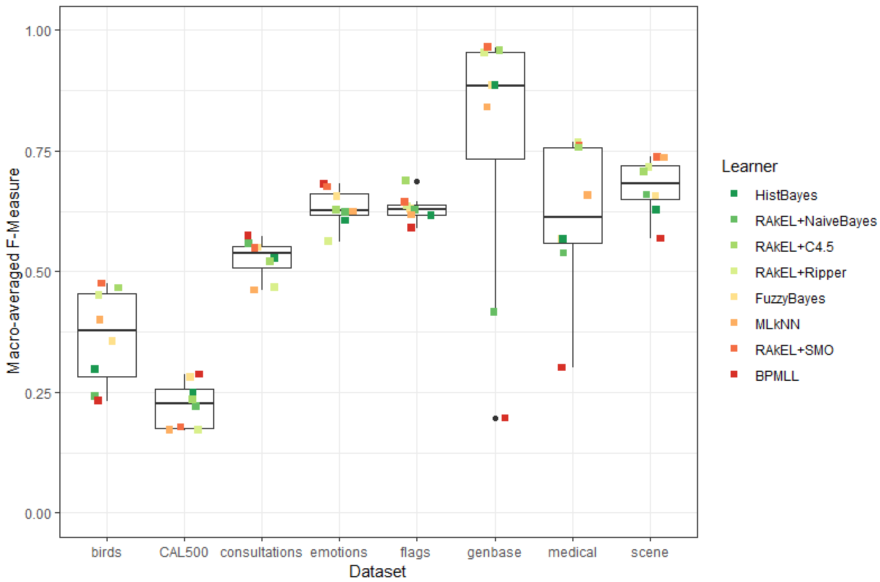


Fig. 2. Distribution of macro-averaged F-measures of several multi-label classification systems for different datasets. Results obtained by cross-validation.

results have been obtained by cross-validation. Classification systems have been ordered by their degree of "transparency" according to the definition developed in section 2. Green for the most "transparent", red for the less "transparent". Although a macro-averaged F-measure alone does not allow a precise evaluation, it allows us to have an overview of classification systems' performances.

We can see that the most "transparent" classification systems (greenest squares in Fig. 2) are not necessarily offering the worst performances. We can also see that, in some cases, "transparent" classification systems can offer performances close to the performances of the less "transparent" ones. In our case, represented by the *consultations* dataset, although the BPMLL algorithm offers the best F-Measure with 0.57, we can see that our version of the Naive Bayes algorithm (HistBayes) offers a quite close F-Measure with 0.53. Note that these results have to be nuanced by the small size of our dataset.

4 Discussion

As introduced in section 2, the definition and operational criteria of "transparency" we proposed are centered on our use case: classification systems in medical contexts. Because this context is sensitive, we had to establish clear operational criteria of what we called a "transparent" classification system. Based on these definitions we have been able to determine what kind of classification system we must use in priority. Besides, we can suppose that the operational criteria we proposed can be used to evaluate the "transparency" of healthcare information systems in general. It would also be interesting to establish operational criteria of "transparent" systems in other contexts than medicine and to compare these operational criteria.

However, these definitions and operational criteria have their limitations. First, they are mainly based on our definitions of "transparency" and on our understanding of the medical context(as computer scientist and engineers). Consequently, they are not exhaustive and can be improved. And secondly, operational criteria were chosen to be easily evaluated without creating additional workload to clinicians, but it could be interesting to integrate them in the evaluation process. For example, the "understandability" of provided explanations could be evaluated directly in practice by clinicians.

Nevertheless, we claim that establishing clear operational criteria of "transparency" can be useful for decision-makers to determine which systems or algorithm is more relevant in which context. These operational criteria of "transparency" must be balanced with performance criteria. Depending on the use case, performances could be more important than "transparency". In our case, the medical context requires to be as "transparent" as possible. Fortunately, as developed in sub-section 3.3, in our case being "transparent" had not a lot of impact on performances and did not implies the use of a less "transparent" classification system with better performances.

Acknowlegment

This paper was made in collaboration with employees of the Civil Hospitals of Lyon (France). Thanks to all of them. Special thanks to Pr. Moulin and Dr. Riou for their suggestions and instructive discussions.

References

1. Abdollahi, B., Nasraoui, O.: Transparency in fair machine learning: The case of explainable recommender systems. In: Human and Machine Learning, pp. 21–35. Springer (2018). https://doi.org/10.1007/978-3-319-90403-0_2
2. Akkermans, H., Bogerd, P., Van Doremalen, J.: Travail, transparency and trust: A case study of computer-supported collaborative supply chain planning in high-tech electronics. European Journal of Operational Research **153**(2), 445–456 (2004). https://doi.org/10.1016/S0377-2217(03)00164-4
3. Amiribesheli, M., Hosseini, M., Bouchachia, H.: A principle-based transparency framework for intelligent environments. In: Proceedings of the 30th International BCS Human Computer Interaction Conference: Fusion! p. 49. BCS Learning & Development Ltd. (2016). https://doi.org/10.14236/ewic/HCI2016.68
4. Ananny, M., Crawford, K.: Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability. new media & society **20**(3), 973–989 (2018). https://doi.org/10.1177/1461444816676645
5. Berner, E.S.: Clinical decision support systems. Springer, 3rd edn. (2016)

6. Cannon, R.L., Dave, J.V., Bezdek, J.C.: Efficient implementation of the fuzzy c-means clustering algorithms. IEEE transactions on pattern analysis and machine intelligence (2), 248–255 (1986). https://doi.org/10.1109/TPAMI.1986.4767778
7. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N.: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 1721–1730. ACM (2015)
8. Cohen, W.W.: Fast effective rule induction. In: Twelfth International Conference on Machine Learning. pp. 115–123. Morgan Kaufmann (1995)
9. Cramer, H., Evers, V., Ramlal, S., Van Someren, M., Rutledge, L., Stash, N., Aroyo, L., Wielinga, B.: The effects of transparency on trust in and acceptance of a content-based art recommender. User Modeling and User-Adapted Interaction **18**(5), 455 (2008). https://doi.org/10.1007/s11257-008-9051-3
10. Datta, A., Sen, S., Zick, Y.: Algorithmic transparency via quantitative input influence: Theory and experiments with learning systems. In: 2016 IEEE symposium on security and privacy (SP). pp. 598–617. IEEE (2016)
11. Dinka, D., Nyce, J.M., Timpka, T.: The need for transparency and rationale in automated systems. Interacting with computers **18**(5), 1070–1083 (2006)
12. Doran, D., Schulz, S., Besold, T.R.: What does explainable ai really mean? a new conceptualization of perspectives. arXiv preprint arXiv:1710.00794 (2017)
13. Dudani, S.A.: The distance-weighted k-nearest-neighbor rule. IEEE Transactions on Systems, Man, and Cybernetics (4), 325–327 (1976)
14. Goebel, R., Chander, A., Holzinger, K., Lecue, F., Akata, Z., Stumpf, S., Kieseberg, P., Holzinger, A.: Explainable ai: the new 42? In: International Cross-Domain Conference for Machine Learning and Knowledge Extraction. pp. 295–303. Springer (2018). https://doi.org/10.1007/978-3-319-99740-7_21
15. Göritzlehner, R., Borst, C., Ellerbroek, J., Westin, C., van Paassen, M.M., Mulder, M.: Effects of transparency on the acceptance of automated resolution advisories. In: 2014 IEEE International Conference on Systems, Man, and Cybernetics (SMC). pp. 2965–2970. IEEE (2014). https://doi.org/10.1109/SMC.2014.6974381
16. Groth, P.: Transparency and reliability in the data supply chain. IEEE Internet Computing **17**(2), 69–71 (2013). https://doi.org/10.1109/MIC.2013.41
17. Gunning, D.: Explainable artificial intelligence (xai). Defense Advanced Research Projects Agency (DARPA), nd Web **2** (2017)
18. Hedbom, H.: A survey on transparency tools for enhancing privacy. In: IFIP Summer School on the Future of Identity in the Information Society. pp. 67–82. Springer (2008). https://doi.org/10.1007/978-3-642-03315-5_5
19. Heeks, R., Mundy, D., Salazar, A.: Why health care information systems succeed or fail. Information Systems for Public Sector Management (1999)
20. Herlocker, J.L., Konstan, J.A., Riedl, J.: Explaining collaborative filtering recommendations. In: Proceedings of the 2000 ACM conference on Computer supported cooperative work. pp. 241–250. ACM (2000)
21. Hochreiter, S.: The vanishing gradient problem during learning recurrent neural nets and problem solutions. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems **6**(02), 107–116 (1998)
22. Holzinger, A., Biemann, C., Pattichis, C.S., Kell, D.B.: What do we need to build explainable ai systems for the medical domain? arXiv preprint arXiv:1712.09923 (2017)
23. John, G.H., Langley, P.: Estimating continuous distributions in bayesian classifiers. In: Eleventh Conference on Uncertainty in Artificial Intelligence. pp. 338–345. Morgan Kaufmann, San Mateo (1995)

24. Karsenty, L., Botherel, V.: Transparency strategies to help users handle system errors. Speech Communication **45**(3), 305–324 (2005)
25. Keerthi, S., Shevade, S., Bhattacharyya, C., Murthy, K.: Improvements to platt's smo algorithm for svm classifier design. Neural Computation **13**(3), 637–649 (2001). https://doi.org/10.1162/089976601300014493
26. Kim, T., Hinds, P.: Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction. In: ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication. pp. 80–85. IEEE (2006). https://doi.org/10.1109/ROMAN.2006.314398
27. Michener, G., Bersch, K.: Identifying transparency. Information Polity **18**(3), 233–242 (2013). https://doi.org/10.3233/IP-130299
28. Montavon, G., Samek, W., Müller, K.R.: Methods for interpreting and understanding deep neural networks. Digital Signal Processing **73**, 1–15 (2018)
29. Platt, J.: Fast training of support vector machines using sequential minimal optimization. In: Schoelkopf, B., Burges, C., Smola, A. (eds.) Advances in Kernel Methods - Support Vector Learning. MIT Press (1998)
30. Quinlan, R.: C4.5: Programs for Machine Learning. Morgan Kaufmann Publishers, San Mateo, CA (1993)
31. Richard, A., Mayag, B., Meinard, Y., Talbot, F., Tsoukiàs, A.: How AI could help physicians during their medical consultations: An analysis of physicians' decision process to develop efficient decision support systems for medical consultations. In: PFIA 2018. Nancy, France (2018)
32. Shannon, C.E.: A mathematical theory of communication. Bell system technical journal **27**(3), 379–423 (1948)
33. Sinha, R., Swearingen, K.: The role of transparency in recommender systems. In: CHI'02 extended abstracts on Human factors in computing systems. pp. 830–831. ACM (2002). https://doi.org/10.1145/506443.506619
34. Spagnolli, A., Frank, L.E., Haselager, P., Kirsh, D.: Transparency as an ethical safeguard. In: International Workshop on Symbiotic Interaction. pp. 1–6. Springer (2017). https://doi.org/10.1007/978-3-319-91593-7_1
35. Tax, D.M., Laskov, P.: Online svm learning: from classification to data description and back. In: 2003 IEEE XIII Workshop on Neural Networks for Signal Processing (IEEE Cat. No. 03TH8718). pp. 499–508. IEEE (2003)
36. Tintarev, N., Masthoff, J.: Effective explanations of recommendations: user-centered design. In: Proceedings of the 2007 ACM conference on Recommender systems. pp. 153–156. ACM (2007). https://doi.org/10.1145/1297231.1297259
37. Tsoumakas, G., Katakis, I., Vlahavas, I.: Random k-labelsets for multi-label classification. IEEE Transactions on Knowledge and Data Engineering **23**(7), 1079–1089 (2011). https://doi.org/10.1109/TKDE.2010.164
38. Tsoumakas, G., Spyromitros-Xioufis, E., Vilcek, J., Vlahavas, I.: Mulan: A java library for multi-label learning. Journal of Machine Learning Research **12**, 2411–2414 (2011)
39. Zadeh, L.A., Klir, G.J., Yuan, B.: Fuzzy sets, fuzzy logic, and fuzzy systems: selected papers, vol. 6. World Scientific (1996)
40. Zarsky, T.: Transparency in data mining: From theory to practice. In: Discrimination and Privacy in the Information Society, pp. 301–324. Springer (2013)
41. Zhang, M.L., Zhou, Z.H.: Ml-knn: A lazy learning approach to multi-label learning. Pattern Recogn. **40**(7), 2038–2048 (2007)
42. Zhang, M., Zhou, Z.: Multi-label neural networks with applications to functional genomics and text categorization. IEEE Transactions on Knowledge and Data Engineering **18**, 1338–1351 (2006). https://doi.org/10.1109/TKDE.2006.162

Chapter 6

CoBoy: a virtual assistant for decision support during medical consultations

A virtual assistant dedicated to supporting day-to-day medical consultations

Antoine Richard*‡, Brice Mayag*, François Talbot†, Alexis Tsoukias* and Yves Meinard*
*Université Paris-Dauphine, PSL Research University, CNRS, LAMSADE, UMR 7243, 75016 Paris, France
†DSII Bron, Hospices Civils de Lyon, 61 Boulevard Pinel, 69672 Bron, France
‡GIE Hopsis Lyon, 3 Quai de Célestins, 69229 Lyon, France

***Abstract*—In this paper, we present a virtual assistant developed in collaboration with the civil hospitals of Lyon ("Hospices Civils de Lyon" or HCL in french), a group of hospitals in the area of Lyon (France). This virtual assistant is dedicated to supporting physicians during day-to-day medical consultations. It aims to anticipate which pieces of information physicians will need, given the information already known on patients, and provide them with these needed pieces of information. According to clinical trials made at the HCL, physicians appreciate the help that the proposed system provides for day-to-day practices.**

***Index Terms*—Clinical Decision Support System, Virtual Assistant, Health Informatics**

I. Introduction

In collaboration with the civil hospitals of Lyon ("Hospices Civils de Lyon" or HCL in French), a group of 14 hospitals in the area of Lyon (France), we have developed a virtual assistant, called CoBoy, dedicated to supporting day-to-day medical consultations. This virtual assistant was developed as a module of the Health Information System (HIS) currently used by HCL's clinicians: Easily®.

Consultations are a large part of the HCL's physicians' activities (they performed 1 million of them in 2020). During these consultations, physicians make decisions repeatedly and we aim to support physicians during these customary activities. Diagnostic Decision Support Systems (DDSSs) have been developed as an attempt to support physicians during diagnosis decisions [1], [2].

Famous examples of DDSSs include MYCIN [3], [4], an expert system dedicated to antimicrobial therapy, or INTERNIST-1/QMR [5], [6], an expert system designed to support decisions in generalist medicine. The CHICA system, used at Wishard Memorial Hospital of Indianapolis, is a more recent example dedicated to supporting child health in primary care [7]–[9]. The eIMCI and ALMANACH projects are also recent examples of DDSSs dedicated to supporting child health in developing countries [10], [11].

Some systems based on machine learning algorithms, also called ML-based DDSSs, have recently been developed to support diagnoses. These ML-based DDSSs include, for example, systems dedicated to detecting ocular diseases from a picture of a patient's eye [12]–[14] or systems dedicated to detecting cancer nodules on radiographies [15]–[18].

However, a well-known problem of DDSSs is their lack of acceptability by physicians during day-to-day practices and non-complex situations, which leads physicians to overriding or ignoring current DDSSs' recommendations [19]–[21]. These recommendations are generally "gold-standard" guidelines or suggestions of diagnoses. In previous works [22], we have argued that this lack of acceptability is due to the fact that the support provided by current DDSSs is at odds with the constraints implied by supporting decisions during day-to-day practices and non-complex situations, like customary medical consultations. The main constraint that a DDSS have to deal with, in such situations, is the strong will of the physicians to stay in charge of their decision processes and to stay responsible of the safety of their patient. We concluded that an adapted approach for decision support during customary medical consultations must respect the know-how of physicians and leave them the responsibility of the decisions taken during these consultations, by limiting itself to providing what physicians need most during consultations: pieces of information concerning their patients [22], [23].

Accordingly, we have developed a decision support system, in the form of a virtual assistant, able to anticipate and provide pieces of information that the physician needs to reach diagnoses concerning a specific patient. This system aims to reduce the workload of physicians and to allow them to make informed decisions during medical consultations. In section II, we present the system's capabilities: how it determines which pieces of information to provide to physicians using both a rule-based classifier and a Naive Bayes classifier, how it searches for raw data associated to these pieces of information in the database of Easily®, and how it displays these pieces of information to physicians through its user interface. In section III, we present a simple use case scenario to illustrate the process of the system during a medical consultation. In section IV, we present the clinical trials we made with physicians of the HCL's department of endocrinology. As a result of these clinical trials, despite some practical limitations, physicians showed high acceptability and much enthusiasm for the proposed system. Lastly, in section V, we discuss limitations and perspectives, before briefly concluding in section VI.

II. The process of the proposed system

As introduced in section I, the goal of CoBoy is to determine which pieces of information on a patient could be needed by a physician for a specific consultation, to search for raw data associated to each of these pieces of information in the database of Easily®, and to display these pieces of information to physicians in an interpretable way. Fig. 1 summarizes the process of CoBoy at the beginning of a medical consultation.

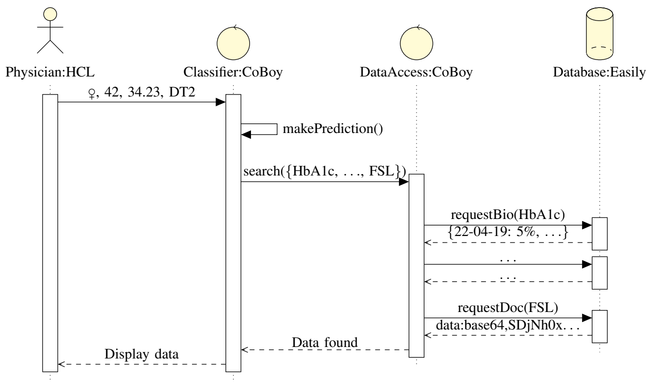


Fig. 1. Summary of the CoBoy's process when providing pieces of information on a patient to a physician at the beginning of a medical consultation

In this section, we detail the different steps of this process. Sub-section II-A details how the system determines, based on two classification systems, which pieces of information are needed by a physician for a specific consultation. Sub-section II-B then explains how the system searches for raw data associated to each piece of information needed, given their data type. Lastly, sub-section II-C details how pieces of information collected by the system are displayed to users.

A. Determining which pieces of information on patients are needed

The use case of determining which pieces of information physicians need for specific consultations can be formalized as a multi-label classification problem [24]. Sex, age, BMI, and the pathology of the patient are common pieces of information about patients that physicians always have at the beginning of consultations. The learning algorithm will use these pieces of information as a base $\mathcal{X}$ to learn which other pieces of information on a patient $\mathcal{Y}$ are needed and which pieces of information are not. Each label corresponds to each piece of information on a patient that could be needed by a physician during a medical consultation. Table I shows an example, with fictitious data, of our use case.

TABLE I
EXAMPLE OF A MULTI-LABEL DATASET BASED ON OUR USE CASE

$\mathcal{X}$: pieces of information known on patients				$\mathcal{Y}$: pieces of information on patients needed by physicians					
Sex	Age	BMI	Disease	HbA1c	Blood Sugar	HDL	LDL	Creatinine	Microalbumin
♀	42	34.23	DT2	1	1	0	0	0	0
♂	52	27.15	HChol	0	0	1	1	0	0
♂	24	21.12	DT1	1	1	0	0	1	1
♀	67	26.22	HChol	0	0	1	1	0	0

To tackle this problem, we've developed a version of the well-known Naive Bayes algorithm [25] adapted to multi-label problems. According to previous works on algorithmic transparency [26], the Naive Bayes algorithm is one of the easiest learning algorithms understandable by physicians. The choice to search for transparency is motivated by the results of studies that have shown that transparent support systems are more accepted in practices by physicians, including those that are more recalcitrant to support systems [27], [28]. Transparency also potentially allows a better understanding of a DDSS and then a better use of this DDSS. Besides, for use cases without complex correlations to learn, like ours, it is possible to use a transparent algorithm without fearing losing performance [26].

To determine which pieces of information are probably needed by a physician for a specific consultation with a patient, CoBoy computes an estimation of the probability $P(Y = 1|X_p)$ that a piece of information Y could be needed by the physician, with $X_p \in \mathcal{X}$ a set of pieces of information known on the patient at the beginning of the consultation. To do so, CoBoy uses a set of probabilities learned from logs of pieces of information searched or recorded by physicians during previous consultations, and applies a naive version of the Bayes theorem (1) which assumes that each variable $X \in \mathcal{X}$ are independent of each other.

$$P(Y = 1|X_p) = \frac{P(Y{=}1) \prod\limits_{X \in \mathcal{X}} P(X{=}x|Y{=}1)}{\sum\limits_{y=0}^{1} \left(P(Y{=}y) \prod\limits_{X \in \mathcal{X}} P(X{=}x|Y{=}y) \right)} \tag{1}$$

Therefore, if $P(Y = 1|X_p) \geq \lambda$ the label $Y \in \mathcal{Y}$ is set to 1, meaning that the physician will probably need the piece of information Y on the current patient p. The λ parameter is used as a threshold to define from which minimum probability a piece of information on a patient may be needed by a physician. By default, this threshold λ is set at 0.5.

However, the current granularity of logs recorded by Easily® for some pieces of information searched by physicians during their consultations limits the construction of a precise learning dataset. We have therefore decided to complete the Naive Bayes classification algorithm with a classification system based on rules defined by physicians of the HCL themselves. A rule-based system is still transparent for physicians [26], [29], especially if rules are defined by physicians themselves and not determined by a learning algorithm. These rules take the form of "IF ... THEN ..." implications, for example:

IF disease = diabetes THEN SEARCH FOR: HbA1c

Thereby obtained a mixed classification system to determine which pieces of information are needed by a physician for a specific situation. When a physician asks CoBoy for pieces of information on a given patient, the system will use the set of rules defined by physicians to determine a first set of pieces of information to provide, and then it will use the Naive Bayes classification system to complete this set.

Currently, the CoBoy system is able to search for, depending on physicians' needs, close to 50 types of biological analyses and close to 30 types of medical documents. Besides, CoBoy is currently able to support consultations for around twenty different kinds of diseases.

B. Searching for raw data in the database

Once the pieces of information to provide to the physician are determined for a specific medical consultation, the system searches for raw data associated to these pieces of information into the database Easily®. For each data type, CoBoy will request the HCL's database in a specific way to obtain the targeted data.

For biological analyses made in medical laboratories, this is the result of the last analysis and the results history that are requested. Because biological analyses are often linked to minimum and maximum thresholds, the system also searches for these thresholds if they are available. Let us specify that results of analyses made by patients in medical laboratories outside the HCL are not available into the database of Easily®if the laboratory has not transmitted these results to the HCL. Due to this practical limitation, CoBoy is unable to provide these results to physicians. For performance reasons, only the hundred most recent results are requested.

For medical documents, such as reports of medical imaging (scanners, MRI, radiographies, etc.), reports from other specialists following the current patient (ophtalmology, dietetic, cancerology, etc.), or drug prescriptions, CoBoy searches for the most recent document in the database.

For "general" pieces of information, such as background history, family medical history, or patient's allergies, CoBoy searches for all data about it into the database of Easily®.

Therefore, CoBoy collects into the database of Easily®all the raw data associated to pieces of information that could be needed by a physician for the current patient, according to the subset determined before (see sub-section II-A). However, it could happen that CoBoy does not find any data concerning the current patient for some pieces of information requested. In such cases, CoBoy simply returns the fact that no data was found in the database for these pieces of information, because the absence of data on the current patient can also be relevant to the decision process of physicians [23].

C. Displaying pieces of information to physicians

CoBoy then displays all the targeted pieces of information on the patient to the physician according to their type. Fig. 2 displays, with fictitious data, the current user interface of CoBoy.

This user interface is divided into two columns. The left column is dedicated to displaying the results of biological analyses and "general" pieces of information. For each piece of information corresponding to biological analyses, CoBoy plots the results history and highlights the latest result obtained. If biological analyses were associated with thresholds, CoBoy also plots these thresholds with red horizontal lines. Concerning "general" pieces of information, CoBoy simply displays them textually.

The right column is dedicated to displaying medical documents concerning the current patient. These documents are displayed using a PDF viewer and are accessible through a tab system.

Pieces of information determined using the Naive Bayes classifier (see sub-section II-A) are sorted according to the estimated probability that the physician will need them during the consultation. The more probable is the higher, but pieces of information determined by the rule-based system come first. In the case in which no data was found in the HCL's database for a piece of information, CoBoy says that it has searched for raw data for these piece of information but didn't found anything, with a message such as "I didn't find any data about HbA1c of this patient".

1) Results' explanations: Each provided piece of information is associated with explanations of how CoBoy determined that the physician could need this piece of information during the current consultation. These explanations are available by clicking on the information point next to the associated name of the piece of information (see Fig. 2).

For pieces of information determined using the Naive Bayes classifier, two levels of explanations are provided. First, CoBoy provides the probability that the physician may need this piece of information on the patient given the patient's age, weight, BMI, and disease. This first level of explanation takes the form of a sentence such as: "Based on your previous consultations and given the sex, age, BMI, and disease of the patient, the probability that you may need to know her/his HbA1c is 72%". The second level of explanation, accessible by clicking on a "Details" button, contains the details of the calculations made to produce this result. This second level of explanation allows the physician to know which pieces of information were crucial in the prediction of her/his needs.

For pieces of information determined using the rule-based system, CoBoy provides the activation condition of the rule. This explanation takes the form of a sentence such as: "The HbA1c of the current patient is provided because: disease = diabetes".

There are also pieces of information on patients that are requested and displayed by default for all the patients, such as the report of the last consultation or her/his last drug prescription. In this case, as an explanation, CoBoy simply says that this piece of information on the patient is requested by default.

2) User's feedbacks: In addition, each piece of information provided by CoBoy is associated with a switch button (see Fig. 2). This button allows the physician to indicate to CoBoy if s/he didn't need the associated piece of information for the current patient. This feedback feature allows us to have an estimation of CoBoy's precision in practical use cases.

The feedback button is also available for pieces of information requested by CoBoy but not found in databases. The goal is to know whether or not CoBoy was right to request a specific piece of information for a specific medical consultation.

In the case of pieces of information on patients displayed by default, this feature is not available, because default pieces of information do not depend on CoBoy's predictions.

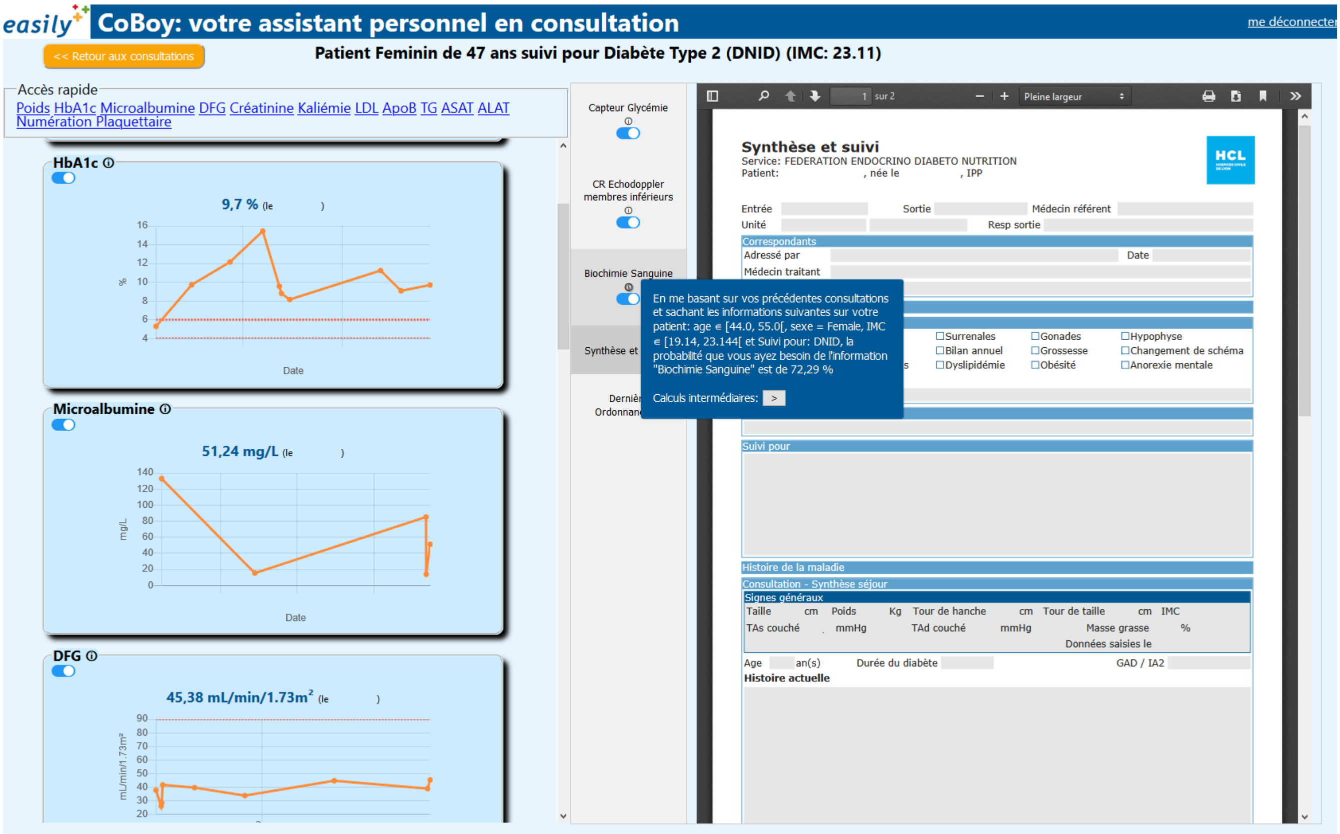


Fig. 2. The current user interface of CoBoy (with fictitious data)

III. Use Case

In this section, we use a simplified example of a medical consultation to illustrate the process of CoBoy. Let us take a fictitious patient p characterized by the set $X_p = \{sex : ♀,\ age : 42, bmi : 34.23, disease : DT2\}$, with $DT2$ corresponding to type 2 diabetes. At the beginning of the consultation, the physician selects the patient in her/his list of patients for the day.

When the patient is selected, CoBoy uses its rule-based classifier to determine a first subset of pieces of information to provide to the physician. Let us suppose that this only rule is activated: “IF disease=diabetes THEN SEARCH FOR HbA1c”, because the patient p is followed up for her/his type 2 diabetes.

Then CoBoy computes the probabilities that the physician may need additional pieces of information on the patient during the consultation. As developed in sub-section II-A, these probabilities are computed by using the Naive Bayes classifier of CoBoy, given the set X_p of common information on the patient. For each piece of information $Y \in \mathcal{Y}$ potentially needed by physicians, CoBoy computes the probability $P(Y = 1|X)$ by applying Bayes' theorem (1). Let us suppose that pieces of information potentially needed by physicians are glycated hemoglobin (HbA1c), blood sugar, HDL cholesterol, LDL cholesterol, creatinine, and microalbumin. Because “HbA1c” is already determined as a piece of information to provide to the physician, CoBoy will not compute the probability that the physician may need “HbA1c” for the current patient p. Let us suppose that CoBoy computes the following probabilities (2).

$$P(\text{Blood Sugar} = 1 \mid X_p) \simeq 0.66 \tag{2a}$$

$$P(\text{HDL} = 1 \mid X_p) \simeq 0.01 \tag{2b}$$

$$P(\text{LDL} = 1 \mid X_p) \simeq 0.01 \tag{2c}$$

$$P(\text{Creatinine} = 1 \mid X_p) \simeq 0.34 \tag{2d}$$

$$P(\text{Microalbumin} = 1 \mid X_p) \simeq 0.42 \tag{2e}$$

In this example, if we use a threshold $\lambda = 0.5$, CoBoy determines that the physician probably needs to obtain the patient's blood sugar history. We can see in this example that, with a more permissive threshold, e.g. $\lambda = 0.4$, CoBoy would also determine that the physician probably needs to know the microalbumin history of the patient. However, using a more permissive threshold may overwhelm physicians with useless pieces of information on the patient.

Once CoBoy has determined which pieces of information to provide to the physician, it requests associated raw data in the HCL's database. In our example, CoBoy has to request raw data for glycated hemoglobin (HbA1c) and blood glucose history concerning the current patient.

The HbA1c is a piece of information obtained through laboratory analyses. It is hence a biological analysis on the patient. CoBoy then requests the hundred most recent results concerning the patient's HbA1c from the HCL's database. Once it has obtained these data, CoBoy plots the results history and displays it to the physician (see Fig. 2).

Concerning the patient's blood sugar history, this piece of information can also be considered to be a biological analysis. However, blood sugar is generally collected by patients themselves using a blood sugar meter. CoBoy then requests the document produced by the patient's blood sugar meter and displays it to the physician (see Fig. 2). If the patient doesn't have a blood sugar meter, nothing is displayed to the physician, but CoBoy tells the physician that the piece of information has been searched but not found.

IV. Clinical Trials

In this section, we detail the clinical trials we have conducted at the HCL to evaluate the viability and the applicability of our decision support system during real consultations. Sub-section IV-A presents our observation protocol and our hypotheses. Sub-section IV-B then details the results obtained from our clinical trials.

A. Protocol

We use here the term "observation" to refer to the study of one, and only one, consultation starting when the patient enters the consultation room and ending when s/he leaves it. We use the term "observation session" to denote a set of consecutive observations. Fig. 3 schematizes the setting during observations. A second screen has been used to allow physicians to access to both Easily® and CoBoy at the same time.

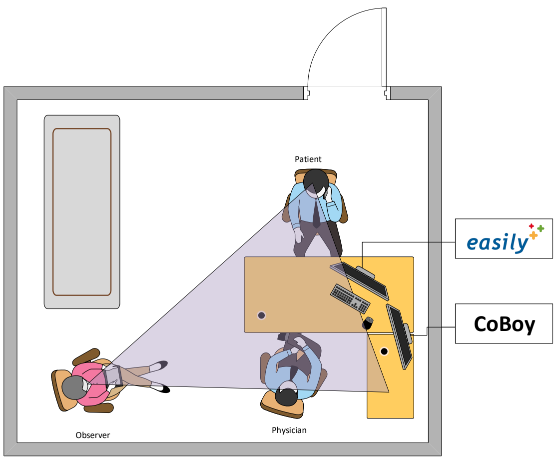


Fig. 3. Agents' disposition during clinical trials of CoBoy

The first part of this study aims to evaluate whether the number of pieces of information provided by the systems had an impact, beneficial or not, on physicians' decision processes. During observations, the following criteria have been monitored by the observer:

- The number of pieces of information provided by CoBoy, in order to quantify the support provided by the system;
- The number of pieces of information searched by the physician, explained orally to the patient, in order to quantify the need for information of the physician;
- The number of mouse clicks made by the physician when using Easily®, used to quantify physician-HIS interactions;
- The duration of the consultation.

We hypothesize that, if pieces of information provided by our decision support system are useful to physicians, this will decrease the need for physicians to look for other pieces of information on their own. In other words, the higher the number of pieces of information provided, the lower the number of pieces of information searched by physicians on their own. We have chosen to use the number of mouse clicks made by physicians, when they use Easily®, as an indicator of their interactions with their HIS when they search for or record pieces of information. Hence, according to our assumptions, the higher the number of pieces of information provided, the lower the number of mouse clicks. Concerning the duration of consultations, physicians generally try to stay on schedule as much as possible. Therefore, we assume that the number of pieces of information provided by our system will not impact, either positively or negatively, the duration of consultations.

However, the introduction of a new tool in physicians' work processes unavoidably entails an entry cost [22], [30]. Therefore, we expect to observe, in many cases, a rise in the number of pieces of information searched, the number of mouse clicks, and the duration of consultations, due to the introduction of CoBoy in the work processes of physicians. If the correlation between the rise of these three criteria and the rise of the number of pieces of information provided by the tool is weak, overall this will support the hypothesis that the support provided by CoBoy has been able to compensate this inevitable entry cost.

The second part of our study aims to assess the potential acceptability of the developed decision support system. To that end, we asked physicians to fill in a brief questionnaire at the end of observation sessions, once all the scheduled consultations were finished. To minimize the impact of the presence of the observer on physicians' answers, questionnaires were completed anonymously and were then shuffled. As a consequence, the two parts of our study cannot be connected. This questionnaire is composed of three main questions, aiming to assess respectively the useability of the system, the perceived utility of the system, and the intention to use the system in day-to-day practices:

1) "*Would you say that getting started with CoBoy is*":
   - "*Very easy*"
   - "*Rather easy*"
   - "*Neither easy nor difficult*"
   - "*Rather difficult*"
   - "*Very difficult*"
2) "*Would you say that, during consultations, CoBoy is*":

- "*Very useful*"
- "*Rather useful*"
- "*Neither useful nor useless*"
- "*Rather useless*"
- "*Totally useless*"

3) "*If the HCL integrated CoBoy into Easily®, would you use CoBoy during your consultations?*"
   - "*Certainly*"
   - "*Rather yes*"
   - "*I don't know*"
   - "*Rather no*"
   - "*Certainly not*"

For each question, physicians were given the opportunity to complete their answers with "additional comments". Finally, we asked two general questions to physicians concerning their opinion about the decision support system and potential improvements:

- "*Do you have general comments concerning CoBoy?*"
- "*Do you have suggestions for improvements?*"

B. Results

Overall, we have observed 49 consultations performed by 7 physicians working at the HCL's department of endocrinology and using Easily®every day. 4 to 11 consultations were observed for each physician, with a mean of 7 observations per physician. Table II summarizes, for each observed physician, the means and standard deviations of results for each criterion observed during clinical trials.

TABLE II
SUMMARY OF RESULTS OBTAINED FOR EACH PHYSICIAN OBSERVED DURING CLINICAL TRIALS OF COBOY

Physician	Number of observations	Mean number of pieces of information provided	Mean number of pieces of information searched	Mean number of mouse clicks	Mean duration of consultations (in minutes)
1	8	14 ± 3	14 ± 2	82 ± 26	23 ± 2
2	6	11 ± 5	14 ± 5	126 ± 44	26 ± 7
3	4	14 ± 1	15 ± 2	188 ± 78	35 ± 11
4	11	8 ± 6	8 ± 2	96 ± 32	18 ± 4
5	8	12 ± 6	19 ± 5	97 ± 39	23 ± 6
6	6	15 ± 1	16 ± 1	208 ± 69	36 ± 10
7	6	4 ± 4	15 ± 4	100 ± 57	23 ± 8

For all the physicians, except for physician number 7, the mean number of pieces of information provided by the system and the mean number of pieces of information searched by the physician are similar. This observation can be explained by several phenomena that we were able to observe:

- In some cases, although pieces of information provided were corresponding to what physicians needed, they asked their patients to confirm these pieces of information;
- In other cases, pieces of information provided were corresponding to what physicians needed, but they were not up to date and physicians had to search for updated information on their own. These cases occurred regularly during the clinical trials, particularly for pieces of information from laboratories external to the HCL, since the information was not available through Easily®;
- In still other cases, pieces of information provided were corresponding to what physicians needed, but they also needed further pieces of information about the patient;
- Finally, in some other cases, pieces of information provided did not correspond to physicians' needs, and the latter therefore found themselves in a classical consultation situation.

Table II also shows that interactions with Easily®, measured by the number of mouse clicks, differ greatly depending on the physician observed. Some physicians have developed work processes optimized according to their needs, while others follow more exploratory processes. Physicians who work in this second way are the ones who interact the most with Easily®. Concerning the duration of the observed consultations, except for a few cases, they appear to be fairly stable from one consultation to another, because physicians tried to stay on schedule as much as possible.

To identify a general tendency concerning the impact of our system on the work processes of observed physicians, we have computed the correlation coefficients between the different criteria monitored. Pearson's correlation coefficient $\rho_{X,Y} \in [-1, 1]$ of two variables X and Y is computed as detailed in (3). The closer $\rho_{X,Y}$ is to 1, the more the variables X and Y are positively correlated. The closer $\rho_{X,Y}$ is to -1, the more the variables X and Y are negatively correlated. And the closer $\rho_{X,Y}$ is to 0, the less the two variables are correlated.

$$\rho_{X,Y} = \frac{\mathbb{E}[(X-\mu_X)(Y-\mu_Y)]}{\sigma_X \sigma_Y} \quad (3)$$

By computing these correlation coefficients for each pair of criteria, we obtained the correlation matrix presented in Fig. 4. The duration of consultation and the number of mouse clicks are highly correlated, with a correlation coefficient of 0.88. To a lesser extent, the duration of a consultation is also positively correlated with the number of pieces of information searched by physicians. These two points make sense because the more a physician searches for pieces of information, or the more s/he needs to interact with Easily®, the more the duration of the consultation is extended. On the other hand, the number of pieces of information provided by the decision support system seems to have a low degree of correlation with the duration of consultations, which validates our hypothesis concerning this point. The reason is that physicians try to stay on schedule as much as possible, with a fixed time for all consultations.

The number of pieces of information provided by CoBoy has a low degree of correlation with the number of pieces of information searched by physicians and to the number of mouse clicks. Although the coefficients computed between these three criteria are not negative, they are relatively close to 0. These results are in line with our hypotheses, because the rise of the number of pieces of information provided does not appear to have a high degree of correlation with the rise of the number of pieces of information searched by physicians, or with the rise of the number of mouse clicks. The support

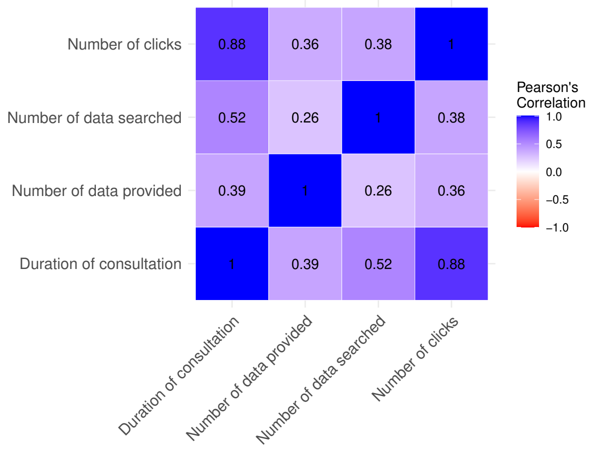


Fig. 4. Correlation matrix between each criterion observed during clinical trials of CoBoy

provided by CoBoy hence appears to have compensated for its entry cost.

Concerning the results of our questionnaire on the acceptability of our decision support system, they are positive. Fig. 5 shows the distribution of physicians' answers to the three main questions of our questionnaire, concerning respectively the useability of the system, its perceived utility, and the intention to use the system in practice. Because the five possible answers were different for the different questions, we have established a scale from -2 to +2: -2 and -1 corresponding to negatives answers, 0 correspondings to the neutral answer, and +1 and +2 corresponding to positives answers. We have also decided to accept intermediate answers, because some physicians were not able to decide between two adjacent proposals.

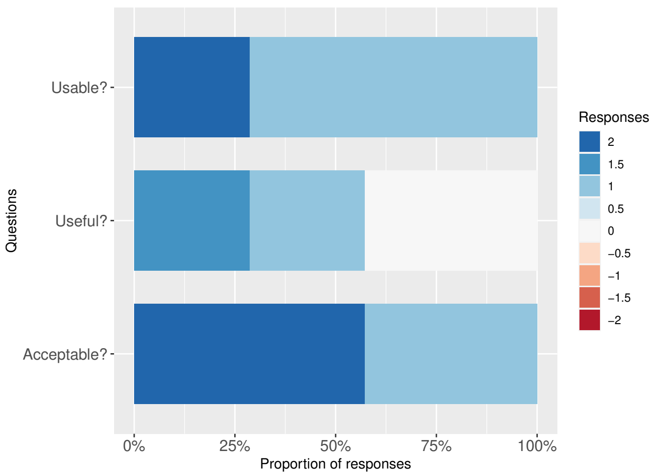


Fig. 5. Distribution of answers to the main questions of the questionnaire

75% of physicians (5 physicians out of 7) found that getting started with CoBoy was "rather easy". The two remaining physicians found that getting started with CoBoy was "very easy". The inconvenience caused by the use of a second screen, but also by a fairly rudimentary user interface, were the main comments concerning the useability of CoBoy.

Concerning the perceived utility of the tool, results are more mixed but overall they are positive. More than half of the observed physicians (4 physicians out of 7) found CoBoy "rather useful" or "very useful". The other three physicians found CoBoy "Neither useful nor useless". According to comments made by this second group of physicians, they were not able to estimate the usefulness of the support system due to the limitations encountered during clinical trials, but they did not find the proposed approach "useless".

Lastly, concerning the intention to use CoBoy in practice, more than half the physicians (4 physicians out of 7) answered that they will "certainly" use it, if it is well integrated into Easily®and if the limitations encountered are overcome. The other three physicians answered "rather yes" to the same question under the same requirements. Despite the current limitations, the observed physicians were particularly enthusiastic about the possibilities offered by the decision support system. The involvement of physicians during the conception and development of the decision support system can have played a role in enhancing this high acceptability.

The most recurrent general comments were, for the most negative ones, about the current technical and ergonomic limitations of the system, and, for the most positive ones, about the working comfort offered by having access to a summary of information concerning their patient for the whole duration of the consultation. The suggested improvements were mainly about possible ways to have access to missing biological pieces of information, but also about various possible improvements for the user interface.

V. Limitations and perspectives

As explained previously, one of the main limitations to the viability of the support provided by our system is its impossibility to provide data coming from laboratories external to the HCL. However, the reasons for these limitations are not due to our decision support system but come from the context of the use of the system. Indeed, although large amounts of data on patients are available and accessible in the HCL's databases, if specific pieces of information necessary to take decisions are not available, the support provided by the system is crippled. A decision support system such as CoBoy must hence be built on a robust and viable structure.

Because our system is a prototype not integrated into the current work processes of physicians, it remains difficult to foresee its real impact in practical situations. Besides, the simple introduction of a second screen in the physicians' workspace is not without consequences, for two reasons. The first reason is that physicians are currently accustomed to using only one screen and having to navigate between two screens, or simply to think about looking at the second screen, can cause discomfort. The second reason is the physical space that a second screen takes in a consultation room which can create an additional physical barrier between the physician and her/his patient. However, some physicians noted the comfort allowed by the use of two screens, especially thanks to the possibility that it offers to record information on patients and write documents on one screen, while having access to a summary of information about the current patient on the other screen. An analysis of the physical space that computers take in a consultation room would be relevant to find a balance between comfort and physician-patient interactions.

Despite these limitations, the decision support system CoBoy have aroused interest among the physicians who have participated in the clinical trials. They articulated various suggestions to improve the decision support system and to overcome the problems encountered during the clinical trials. More specifically, physicians saw the potential utility of the system if these problems could be solved. Concerning the external biological analyses, for example, physicians proposed that these pieces of information should be recorded by themselves or their secretaries, before consultations or during consultations, using the user interface of Easily®or CoBoy. However, this could generate an additional workload. The relevance of engaging such an additional workload depends on the real usefulness of the support provided. The ideal situation would be to be able to extract, without risks of errors, results of biological analyses from reports of external laboratories, or to have a common structure between the HCL and external laboratories, allowing the communication of results to physicians.

Physicians also made various suggestions about the user interface of CoBoy. The main idea that emerged from discussions with physicians is that the user interface should provide, at a glance, a summary of the most important information on patients according to their diseases, and quick access to secondary information. The user interface proposed by [31], for the follow-up of patients with diabetes diseases, could be a basis for the future user interface of CoBoy, for all kinds of diseases. We also intend to give to physicians the possibility to indicate to the system which pieces of information to display under which conditions, through a system allowing physicians to configure the user interface to meet their needs. This could improve not only the working comfort for the physicians, but also their appreciation of the system. However, this could also generate an additional workload. Once again, a balance has to be found between the modularity of the user interface and this additional workload.

To summarize, although it is currently difficult to assess whether CoBoy impacts positively or negatively the work processes of physicians, the proposed approach is appreciated and accepted by the physicians who have participated in clinical trials. The many suggestions made by physicians for a better integration of CoBoy in their work processes highlight an interest in the proposed decision support system. The introduction of a decision support system such as CoBoy, aiming to provide physicians with a set of targeted pieces of information, appears to be adapted to the needs of physicians during customary medical consultations, and seems to be more acceptable than the approaches materialized by the current DDSSs [22]. However, other clinical trials should be performed once CoBoy will be integrated into Easily®and into the work processes of physicians, to have a more reliable assessment of its impact and its real acceptability.

Besides, we have developed the proposed system in the specific context of medical consultations concerning diseases treated in endocrinology. Although we aim to propose a support system as generic as possible, the proposed system might well fail to correspond to physicians' needs during consultations in other medical specialties. It would therefore be interesting to study, through other clinical trials, the applicability of the proposed decision support system in other hospital departments.

VI. Conclusion

In this paper, we have presented a decision support system taking the form of a virtual assistant dedicated to supporting physicians during their day-to-day medical consultations. This system is the result of several years of works made in collaboration with the employees of the HCL on how to support physicians during customary situations such as medical consultations [22], [23], [26]. Our first goal was to propose a decision support system acceptable for physicians and adapted to the constraints and the challenges of supporting customary medical consultations.

Named "CoBoy", this decision support system can anticipate and provide pieces of information needed by physicians for their consultations, given common pieces of information on the patient: age, sex, BMI, and the disease for which s/he is followed-up. This system has been developed in collaboration with physicians working at the HCL's department of endocrinology. It is hence calibrated for diseases treated in this domain. Currently, CoBoy is able to treat around twenty distinct diseases and can search for, according to physicians' needs, among a hundred different pieces of information about patients.

We have conducted a set of clinical trials to evaluate the feasibility of the introduction of CoBoy in the work process of HCL's physicians. Although the impact of our system on physicians' work processes is not entirely assessable, our first results are positive. In addition, physicians who used CoBoy during these clinical trials showed a certain interest in it and showed an interest in being involved in the process of improving the decision support system. Although much work remains to be done, the decision support system we proposed corresponds to physicians' needs during medical consultations and to the constraints underlying decision support in such situations.

Further works are needed to evaluate more precisely the impact of such a decision support system on decision processes and physicians' workload. Besides, further works are also needed to propose a better user interface and to adapt our decision support system in several services of the HCL. However, the high acceptability showed by physicians during the clinical trials towards CoBoy, compared with the low acceptability of current DDSSs [19]–[21], supports the validity of our choice of a self-adaptive virtual assistant for supporting decisions during medical consultations.

Acknowledgments

Works presented in this paper were made during the pandemic wave of SARS-CoV-2. This particularly impacted the clinical trials of the decision support system. We want to address a special thanks to Pr. Moulin, Pr. Raverot, Dr. Abeillon, Dr. Bouzehouane, Dr. Brac De La Perriere, Dr. Charriere, Dr. Lasolle, Dr. Moret, Dr. Renault, and Dr. Villar-Fimbel for their disposal and for allowing this work to be done. This paper is also the result of a long collaboration with the employees of the Civil Hospitals of Lyon (France). Thanks to all of them.

References

[1] R. A. Miller, "Diagnostic decision support systems," in *Clinical Decision Support Systems*. Springer, 2016, pp. 181–208.

[2] J. M. Reider, "Impact of national policies on the use of clinical decision support," in *Clinical Decision Support Systems*. Springer, 2016, pp. 111–130.

[3] E. H. Shortliffe, S. G. Axline, B. G. Buchanan, T. C. Merigan, and S. N. Cohen, "An artificial intelligence program to advise physicians regarding antimicrobial therapy," *Computers and Biomedical Research*, vol. 6, no. 6, pp. 544–560, 1973.

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

[5] R. A. Miller, M. A. McNeil, S. M. Challinor, F. E. Masarie Jr, and J. D. Myers, "The internist-1/quick medical reference project—status report," *Western Journal of Medicine*, vol. 145, no. 6, p. 816, 1986.

[6] R. A. Miller, "A history of the internist-1 and quick medical reference (qmr) computer-assisted diagnosis projects, with lessons learned," *Yearbook of medical informatics*, vol. 19, no. 01, pp. 121–136, 2010.

[7] V. Anand, P. G. Biondich, G. C. Liu, M. B. Rosenman, S. M. Downs *et al.*, "Child health improvement through computer automation: the chica system." in *Medinfo*, 2004, pp. 187–191.

[8] D. J. Etter, A. McCord, F. Ouyang, A. L. Gilbert, R. L. Williams, J. A. Hall, W. Tu, S. M. Downs, and M. C. Aalsma, "Suicide screening in primary care: use of an electronic screener to assess suicidality and improve provider follow-up for adolescents," *Journal of Adolescent Health*, vol. 62, no. 2, pp. 191–197, 2018.

[9] S. M. Honaker, T. Dugan, A. Daftary, S. Davis, C. Saha, F. Baye, E. Freeman, and S. M. Downs, "Unexplained practice variation in primary care providers' concern for pediatric obstructive sleep apnea," *Academic pediatrics*, vol. 18, no. 4, pp. 418–424, 2018.

[10] A. Bernasconi, F. Crabbé, A. M. Adedeji, A. Bello, T. Schmitz, M. Landi, and R. Rossi, "Results from one-year use of an electronic clinical decision support system in a post-conflict context: An implementation research," *PloS one*, vol. 14, no. 12, 2019.

[11] C. Bessat, N. A. Zonon, and V. D'Acremont, "Large-scale implementation of electronic integrated management of childhood illness (eimci) at the primary care level in burkina faso: a qualitative study on health worker perception of its medical content, usability and impact on antibiotic prescription and resistance," *BMC public health*, vol. 19, no. 1, p. 449, 2019.

[12] N. Asiri, M. Hussain, F. Al Adel, and N. Alzaidi, "Deep learning based computer-aided diagnosis systems for diabetic retinopathy: A survey," *Artificial Intelligence in Medicine*, 2019.

[13] J. De Fauw, J. R. Ledsam, B. Romera-Paredes, S. Nikolov, N. Tomasev, S. Blackwell, H. Askham, X. Glorot, B. O'Donoghue, D. Visentin *et al.*, "Clinically applicable deep learning for diagnosis and referral in retinal disease," *Nature medicine*, vol. 24, no. 9, pp. 1342–1350, 2018.

[14] K. Zhang, X. Liu, F. Liu, L. He, L. Zhang, Y. Yang, W. Li, S. Wang, L. Liu, Z. Liu *et al.*, "An interpretable and expandable deep learning diagnostic system for multiple ocular diseases: qualitative study," *Journal of medical Internet research*, vol. 20, no. 11, p. e11144, 2018.

[15] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, 2017.

[16] T. C. Hollon, S. Lewis, B. Pandian, Y. S. Niknafs, M. R. Garrard, H. Garton, C. O. Maher, K. McFadden, M. Snuderl, A. P. Lieberman *et al.*, "Rapid intraoperative diagnosis of pediatric brain tumors using stimulated raman histology," *Cancer research*, vol. 78, no. 1, pp. 278–289, 2018.

[17] S. Joo, Y. S. Yang, W. K. Moon, and H. C. Kim, "Computer-aided diagnosis of solid breast nodules: use of an artificial neural network based on multiple sonographic features," *IEEE transactions on medical imaging*, vol. 23, no. 10, pp. 1292–1300, 2004.

[18] G. H. B. Miranda and J. C. Felipe, "Computer-aided diagnosis system based on fuzzy logic for breast cancer categorization," *Computers in biology and medicine*, vol. 64, pp. 334–346, 2015.

[19] D. F. Sittig, M. A. Krall, R. H. Dykstra, A. Russell, and H. L. Chin, "A survey of factors affecting clinician acceptance of clinical decision support," *BMC medical informatics and decision making*, vol. 6, no. 1, p. 6, 2006.

[20] T. Onega, E. J. A. Bowles, D. L. Miglioretti, P. A. Carney, B. M. Geller, B. C. Yankaskas, K. Kerlikowske, E. A. Sickles, and J. G. Elmore, "Radiologists' perceptions of computer aided detection versus double reading for mammography interpretation," *Academic radiology*, vol. 17, no. 10, pp. 1217–1226, 2010.

[21] R. Masud, M. Al-Rei, and C. Lokker, "Computer-aided detection for breast cancer screening in clinical settings: Scoping review," *JMIR medical informatics*, vol. 7, no. 3, p. e12660, 2019.

[22] A. Richard, B. Mayag, F. Talbot, A. Tsoukiàs, and Y. Meinard, "What does it mean to provide decision support to a responsible and competent expert?" *EURO journal on decision processes*, Aug. 2020. [Online]. Available: https://hal.archives-ouvertes.fr/hal-02916089

[23] A. Richard, B. Mayag, Y. Meinard, F. Talbot, and A. Tsoukiàs, "How AI could help physicians during their medical consultations: An analysis of physicians' decision process to develop efficient decision support systems for medical consultations," in *PFIA 2018*, Nancy, France, 2018.

[24] G. Tsoumakas and I. Katakis, "Multi-label classification: An overview," *International Journal of Data Warehousing and Mining (IJDWM)*, vol. 3, no. 3, pp. 1–13, 2007.

[25] G. H. John and P. Langley, "Estimating continuous distributions in bayesian classifiers," in *Eleventh Conference on Uncertainty in Artificial Intelligence*. San Mateo: Morgan Kaufmann, 1995, pp. 338–345.

[26] A. Richard, B. Mayag, F. Talbot, A. Tsoukiàs, and Y. Meinard, "Transparency of classification systems for clinical decision support," in *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, M.-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, and R. R. Yager, Eds. Cham: Springer International Publishing, 2020, pp. 99–113.

[27] R. Sinha and K. Swearingen, "The role of transparency in recommender systems," in *CHI'02 extended abstracts on Human factors in computing systems*. ACM, 2002, pp. 830–831.

[28] A. Holzinger, C. Biemann, C. S. Pattichis, and D. B. Kell, "What do we need to build explainable ai systems for the medical domain?" *arXiv preprint arXiv:1712.09923*, 2017.

[29] R. Caruana, Y. Lou, J. Gehrke, P. Koch, M. Sturm, and N. Elhadad, "Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission," in *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2015, pp. 1721–1730.

[30] R. Heeks, "Health information systems: Failure, success and improvisation," *International journal of medical informatics*, vol. 75, no. 2, pp. 125–137, 2006.

[31] A. Giordanengo, E. Årsand, A. Z. Woldaregay, M. Bradway, A. Grottland, G. Hartvigsen, C. Granja, T. Torsvik, and A. H. Hansen, "Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: Iterative approach and qualitative case study," *JMIR diabetes*, vol. 4, no. 3, p. e14002, 2019.

Chapter 7

Conclusion

The works presented in this thesis were focused on the subject of decision support for customary situations in medical contexts. More specifically, we have studied how to support the decisions of physicians, who are considered responsible medical experts, during their medical consultations. This subject led us to work on various issues pertaining to the decision support of medical consultations: the different conceivable approaches to support physicians' decisions, the decision processes and the needs of physicians during medical consultations, and the impact of systems' "transparency" on their acceptability by physicians.

To bring a conclusion to this manuscript, Section 7.1 synthesizes our different contributions. Section 7.2 summarizes the various limitations of our works, which one should keep in mind when interpreting our results. Lastly, Section 7.3 offers several perspectives on possible future works on subjects close to the ones we treated in this thesis.

7.1 Contributions

During this thesis, we have defended that it is more adapted and acceptable to respect the know-how of physicians and to leave them the responsibility of the decisions taken during consultations, by limiting decision support to providing them information on their patients which are necessary for their decision-making. We have then tackled a series of issues linked to the decision support for medical consultations.

- What tools are currently used to support clinical decisions?
- How do physicians work during their consultations and what are their needs?
- How to propose relevant and acceptable support for physicians?

In this section, we propose a summary of our different contributions concerning these issues.

A critical analysis of clinical decision support systems

In the first part of this thesis, based on works published in Richard et al. (2020b), we have studied the various approaches and systems currently used to support decisions in medical contexts. We aimed to determine whether the current, or past, decision support systems could be adapted to the constraints of decision support for customary medical situations, such as medical consultations. To do so, we have retraced the history of various trends underlying the development of clinical decision support systems, by focusing on diagnostic decision support systems. We then focused on the impact of these tools on physicians' performances and patient safety, but also on the acceptability of these tools by physicians. We highlighted that, despite a potential beneficial impact on physicians' performances, perceived by them, the current decision support systems are poorly accepted in practice by physicians.

To understand this paradoxical observation, we investigated the various barriers that could explain the non-acceptance of the current decision support systems. We showed that, in practice, the introduction of new software in medical contexts is not necessarily to the advantage of clinicians. Indeed, they have to adapt themselves to the use of new tools, generating an additional workload, stress, and potential mistakes in other aspects of their work. We also highlighted that the use of decision support systems raised various responsibility issues. Particularly in the case of medical errors, if physicians remain responsible, then they have to stay on alert for errors that may come from their decision support system, generating an additional workload that could lead to more medical errors. This observation is even more deplorable considering that proposed systems are supposed to support users. Accordingly, it is important, during the development of decision support systems, to properly evaluate the effort asked to physicians to adapt themselves in comparison to the support provided by the tools.

Finally, we highlighted that the main barrier to the use of current decision support systems comes from the choice of approaches not adapted to the constraints of our context of interest. The current systems are based on approaches that we have termed "conformist" or "objectivist", according to the taxonomy proposed by Meinard and Tsoukiàs (2019), which reflects strong ideological choices, sometimes not assumed, about the aim of decision support. As we have shown, these ideological choices do not correspond neither to the objectives, the needs, nor the constraints of physicians during their customary medical consultations. In such situations, this is the physicians' "spirit of initiative", but also their responsibility, that is highly engaged. We concluded that developing a decision support system based on an "adjustive" approach of decision support (Meinard and Tsoukiàs, 2019), adjusting itself to the needs of decision-makers, is more adapted to the constraints underlying the decision support of customary medical consultations.

Modelization of physicians' decision processes during medical consultations

In addition to our analysis of current decision support systems, we have conducted field observations and analyses of the physicians' decision processes (cf. Chapter 3). These analyses extended previous works presented in Richard et al. (2018). We highlighted the different types of actions that physicians can do during a consultation. The task to search for information about the current patient was identified as an essential activity to physicians in order to be able to decide which prescriptions to give to their patients. Based on these analyses, we proposed two models of physicians' decision processes during medical consultations in order to better understand them.

The first model is dedicated to formalizing the process of a specific consultation, by detailing each decision made by physicians as their consultations unfolds. This model allowed us to highlight that physicians' decisions are mainly based on information known about the current patient. At each step of the consultation, physicians can then decide to search for an unknown piece of information about their patient or decide to prescribe something to the current patient. The consultation ends when physicians think they have covered all the possible prescriptions.

We proposed a second model, more generic, allowing to highlight links between the different types of actions and decisions that physicians can perform during their consultations. This second model allowed us to highlight the cyclical structure of consultation processes. In this model, the accumulation of pieces of information concerning the current patient appeared as a central sub-process, essential to the functioning of the general decision process of physicians.

Based on these two models, we highlighted the need for physicians to accumulate information about their patients. However, access to patients' information can be tedious for physicians, even if patients' information are available and accessible through their information systems. We concluded that a decision support system adapted to physicians' needs should be a system able to learn and anticipate which subset of information is needed by physicians according to the current patient's diseases. This kind of system would correspond to the "adjustive" approach we aimed to develop.

Lastly, we highlighted that the kind of problem that such a system has to solve can be formalized as a multi-label classification problem, where each label corresponds to a piece of information potentially useful for physicians. These classification problems can then be treated by computers through classification systems based on machine learning algorithms.

Proposition of operational criteria to evaluate classification systems' "transparency"

In the first part of this thesis, we highlighted a non-acceptance of decision support systems, partly due to the distrust toward algorithms used by these systems which could be opaque to users. To propose a decision support system accepted in practice by physicians, we have investigated what could be the criteria maximizing the "transparency" of the system we aimed to develop. This direction has been taken because several studies have shown the beneficial impact the "transparency" of an algorithm could have on the acceptability of this algorithm by users. Our work on these questions has also been published in Richard et al. (2020a).

First, we proposed a general definition gathering the different requirements, in terms of "transparency", that we imposed on ourselves. Among these requirements, we have listed: the "understandability" of the system, the "interpretability" of the algorithm used and its results, the "traceability" of the system, and the "revisability" of the system. For each requirement, we proposed a set of operational criteria that allowed us to evaluate the "transparency" of a classification system.

We studied then different kinds of well-known classification systems, each based on distinct approaches, and we evaluated their "transparency" according to our operational criteria. Our aim was to select the classification system which respected all the requirements that we imposed on ourselves in terms of "transparency". The Naive Bayes classification system has been evaluated as respecting all our requirements, partly due to its overall "simplicity" and due to the fact that the theory of probabilities, on which Naive Bayes is based, is well-known by physicians.

We also made series of experiments in order to evaluate whether the choice of a "transparent" system implied a significant loss of performance. Had that been the case, we would have needed to modify the system or to select a system less "transparent" but more performant. At the end of our experiments, we showed that the "transparency" of a system is not necessarily correlated to a decrease in performance. In addition, although the Naive Bayes system is not the most performant, in simple use cases like our own, its performances remain decent.

Development of a decision support system dedicated to medical consultations

After having positioned ourselves on the approach which is the most relevant in terms of decision support, after having studied the physicians' decision process during their consultations and their needs in terms of support, and after having selected a classification system respecting all our requirements in terms of "transparency" to maximize the acceptability of the decision support system to develop, we have developed and proposed a decision support system dedicated to customary medical consultations.

Named "CoBoy", this decision support system takes the form of a virtual personal assistant which is able to anticipate and to provide pieces of information needed by physicians for their consultations, according to common information on the patient: age, sex, BMI and the disease for which s/he is followed-up. This system has been developed in collaboration with physicians working at the HCL's department of endocrinology. It is hence calibrated for diseases treated in such departments. Currently, CoBoy is able to treat around twenty distinct diseases and is able to search, according to physicians' needs, among a hundred different pieces of information about patients.

Finally, we conducted a set of clinical trials to evaluate the feasibility of the introduction of CoBoy in the work process of HCL's physicians. Although the impact of our system on physicians' work processes is not entirely assessable, our first results are positive. In addition, physicians who used CoBoy during these clinical trials showed a certain interest in it and they were involved in the improvements of the system. Although a lot of works remain to be done, the decision support system we proposed corresponds to the physicians' needs during medical consultations and to the constraints underlying decision support in such situations.

7.2 Limitations

To have a better understanding of our contributions, it is important to take into account the various limitations of our work. In this section, we propose to summarize the limitations we noted in the different chapters of this thesis. These limitations encompass limitations due to our application case, technical limitations from our application context, and limitations linked to the analytical frameworks and paradigms used.

As detailed in many parts of this thesis, particularly in Part I, we focused our work on the proposal of a decision support system adapted to customary medical consultations. In such situations, physicians are considered to be experts, and their responsibility is highly engaged. It is not necessarily the case for all the situations where medical decisions are made. The main part of our conclusions, particularly those concerning the approach we have chosen, is then not necessarily relevant for application cases other than medical consultations or, more generally, for application cases where the responsibility of the decision-makers is not highly engaged.

Similarly, we have studied various information systems through the prism of the constraints and challenges linked to a medical context. Although this allowed us to highlight some requirements about the software we can use and develop, these requirements are potentially specifics to a healthcare context. In addition, we focused our studies on certain types of systems, such as decision support systems and classification systems, applied to the use case of medical consultations. Our conclusions concerning the use of such systems are then not necessarily relevant for classification systems used in other contexts than medical consultations, or for other types of information systems used during medical consultations.

The hospital context in which this thesis took place was not without constraints either. One of the main barriers with which we had to work was the limited availability of physicians. Indeed, physicians have a busy schedule that we had to deal with. This constraint has particularly limited us during or field observations, presented in Chapter 3 and Chapter 6, for which a larger panel would have allowed us to draw more robust conclusions. We have been able to draw general trends, such as the models presented in Chapter 4, but the latter would benefit from being verified with larger clinical trials.

Finally, during the development of our system, we have worked mainly with physicians specialized in endocrinology. The decision support system is then, by construction, calibrated to support consultations of this medical specialty. This is particularly the case for the classification system we used, because it learned only through consultations performed by endocrinologists. However, we have built this system with a view to genericity, in order to be able to use it in consultations of other medical specialties. In addition, our analysis highlighted various similarities in the work process of many physicians, independently of their medical specialty (cf. Chapter 3), which suggests that our approach could easily be adapted for medical specialties other than endocrinology.

7.3 Perspectives

The works presented in this thesis allowed us to open interesting avenues for future works. In this section, we propose to summarize the perspectives that may have been mentioned in the different chapters composing this thesis.

Adapting the proposed support to other hospital departments

In Chapter 6, we have detailed a set of possible improvements for the decision support system we developed during this thesis. Except for some limitations due to the lack of communications between the HCL and medical laboratories, we have noted that there is still a lot of work to be done on the user interface. In addition, the system we proposed is specifically dedicated to medical consultations, particularly consultations concerning diseases treated in endocrinology. Although our aim was to propose a support system as generic as possible, the proposed system might well fail to correspond to physicians' needs in consultations of other medical specialties. It would therefore be interesting to study, through other clinical trials, the applicability of the proposed system in consultations of other hospital departments.

The constraints underlying the decision support of customary medical consultation, which are the spirit of autonomy and the responsibility of decision-makers, might be found in other medical situations. Besides, we can assume that clinicians' needs in these other situations are not necessarily a need for information. Therefore, it could be interesting to study the applicability of an adjustive approach of decision support in medical situations other than the one studied in this thesis. More precisely, it could be interesting to study the constraints underlying potential decision support in various medical situations in order to determine what would be the more adapted approach for each situation studied.

A medical situation that could be interesting to study in future works is patient care in emergency services. In such situations, clinicians have to make many decisions quickly and repeatedly, to prioritize patients to treat as well as to allocate of medical resources. Besides, the transfer of information concerning patients within the emergency department, between the emergency department and the emergency vehicles, but also between the emergency department and other hospital departments, is of crucial importance for patient care. Scheduling algorithms or resource managers could be conceivable, however it's a safe bet that these kinds of tools would generate distrust and would barely be used in practice. Therefore, studying the different work processes and decision processes of emergency departments could be a line of research extremely challenging.

Rethinking the role of information systems in hospital contexts

In the previous section, we have developed different perspectives linked to the creation of new decision support systems in medical contexts. However, it is also important to consider the possible improvements of information systems used by clinicians in general and to think about the role of these systems in patient care processes.

In Part I, we have mentioned the importance of involving users in the elaboration process of the systems that they will use, to improve the acceptability of these systems, but also to prevent potential barriers and limitations to their use. User involvement to improve the system we have developed, mentioned in Chapter 6, highlighted that physicians appreciate to be integrated into the conception process that they will potentially use during their daily practices. Giordanengo et al. (2019) offered another illustration of the importance of involving physicians in the conception process of information systems. According to the authors, physicians highlighted various possible improvements and potentials barriers. Therefore, an interesting line of research could be to elaborate methods facilitating user involvement in the conception process of new systems or the improvement of existing systems. For example, proposing user interfaces highly configurable could allow a better working comfort. It could also be interesting to analyze, with users, the way they use their current information systems to highlight common practices or potential improvements.

During our clinical trials (cf. Chapter 6), we saw that the simple introduction of a second screen can generate discomfort in physicians' work processes. Indeed, the introduction of this second screen into the consultation room created a physical barrier between physicians and their patients, which was detrimental to their interactions. This observation echoes what we have noted, in Part I, concerning the barriers to the acceptance of information systems by physicians. If the introduction of a new system generates more constraints than benefits, this system risk not to be used in practice. In this part of the thesis, we focused on the "software" aspect of information systems but, as we have seen during our clinical trials, the "hardware" aspect of information systems must also be taken into account. The study of the physical place taken by information systems in various hospital processes could be an interesting line of research for future works. To take the example of medical consultations, the conception of specific hardware, such as modular screens, could be interesting to investigate.

To go further, the communication of information between the different information systems used in a hospital, software or hardware, could also constitute an interesting line of research. For example, we can think about the domain of the Internet of Things (IoT), which is focused on connected objects and their communication protocols. The use of such tools could allow the creation of an "ambient intelligence" that would work in synergy with clinicians. In medical consultations, for example, we could imagine a consultation room with various connected objects (a weighing machine, a tensiometer, etc.) that would send collected data to physicians and to their HIS, which will display automatically these pieces of information. Obviously enough, the use of such tools must be constrained from an ethical point of view, mainly because data used would be extremely sensitive and because, from a practical point of view, we cannot anticipate the impact that these tools would be on physicians' performances and patient safety.

Lastly, as detailed in Part I, the impact of information systems, and their ubiquity in hospitals, on patient safety remains understudied. The introduction of new information systems in clinicians' work processes can be a double-edged sword if their use implies an additional workload for clinicians. The first and main aim of health information systems is to facilitate and to improve patient care processes. As highlighted and argued throughout this thesis, a way to achieve this is to reduce clinicians' workload, through innovative tools, so that clinicians can focus on patient care. It is accordingly relevant to study and to improve information systems currently used by clinicians, always with this aim to allow clinicians to focus on their domain of expertise: the quality of care provided to their patients.

Investigating the relevance of the adjustive approach in other domains

Similar constraints to those underlying decision support of medical consultations can certainly be found in other situations where the decision process is close to that of a diagnosis, or situations where decision-makers' responsibility is highly engaged. Such situations could provide a field of experimentation for an adjustive approach of decision support.

In the context of the implementation of environmental policies, for example, Meinard and Thébaud (2019) argued that environmental management schemes are currently crippled in France by the lack of a large-scale database on vegetation types, while environmental institutions spend considerable time and money to produce ill-adapted guidelines intended for experts in the field. Decision support, in such contexts, could largely benefit from adjustive decision support systems based on the study of decision-makers' needs.

More generally, an adjustive decision support can be relevant for any decision for which the responsibility of decision-makers is highly engaged. An interesting line of research could be the development of decision support systems dedicated to collective decisions processes. For example, we could imagine a "smart" city with tools collecting and displaying information that could be used as a basis for discussions during citizens' conventions, or systems facilitating deliberations and summarizing elements of the current debate.

Lastly, the ethical issues raised by the use of decision support systems in such situations could be particularly interesting to investigate. The various requirements that we imposed on ourselves in terms of "transparency", detailed in Chapter 5, could serve as a basis for future works on similar use cases, outside the medical field. Also, although we focused on the specific concept of "transparency" due to our application context, it appears relevant to study other concepts, just as fundametal, linked to the ethics of algorithms and to proposed operational criteria to evaluate algorithms according to these concepts. For example, we can think to the concepts of "justice", "equity" or "fairness", formalized by Beauchamp et al. (2009) and Clément et al. (2008), particularly relevant for systems used to support decisions about public policies.

Appendix A

Observation form

Fiche d'observation

Heure d'arrivée du patient (hh:mm)

Informations recherchées

Nombres de clics de souris

Heure de départ du patient (hh:mm)

Remarques du médecin (optionnel)

Appendix B

Questionnaire

Questionnaire

Diriez-vous que la prise en main de CoBoy est :

Très Facile	Plutôt Facile	Ni Facile, Ni Difficile	Plutôt Difficile	Très Difficile

Commentaires additionnels : ....
....

Diriez-vous que, en consultation, CoBoy est :

Vraiment Utile	Plutôt Utile	Ni Utile, Ni Inutile	Plutôt Inutile	Totalement Inutile

Commentaires additionnels : ....
....

Si les HCL intégraient CoBoy dans Easily, utiliseriez-vous CoBoy durant vos consultations ?

Certainement	Plutôt Oui	Ne sais pas	Plutôt Non	Pas du tout

Commentaires additionnels : ....
....

Avez-vous des commentaires généraux à propos de CoBoy ?

....
....

Avez-vous des suggestions d'amélioration ?

....
....

Bibliography

Beauchamp, T. L., J. F. Childress, et al.
2009. *Principles of biomedical ethics*, 7th edition. Oxford University Press, USA.

Bose, R. J. C. and W. M. Van Der Aalst
2011. Analysis of patient treatment procedures. In *Business Process Management Workshops (1)*, volume 99, Pp. 165–166.

Clément, V., C. Leclainche, and D. Serra
2008. *Economie de la justice et de l'équité*.

Ducray, V., A. Vlachomitrou, M. Bouscambert-Duchamp, S. Si-Mohamed, S. Gouttard, A. Mansuy, F. Wickert, A. Sigal, A. Gaymard, F. Talbot, C. Michel, T. Perpoint, J. Pialat, O. Rouvière, L. Milot, F. Cotton, P. Douek, M. Rabilloud, and L. Boussel
2020. Chest CT for rapid triage of patients in multiple emergency departments during covid-19 epidemic: experience report from a large French university hospital. *European Radiology*, Pp. 1432–1084.

Gibson, M., K. N. Jenkings, R. Wilson, and I. Purves
2006. Verbal prescribing in general practice consultations. *Social Science & Medicine*, 63(6):1684–1698.

Giordanengo, A., E. Årsand, A. Z. Woldaregay, M. Bradway, A. Grottland, G. Hartvigsen, C. Granja, T. Torsvik, and A. H. Hansen
2019. Design and prestudy assessment of a dashboard for presenting self-collected health data of patients with diabetes to clinicians: Iterative approach and qualitative case study. *JMIR diabetes*, 4(3):e14002.

Günther, C. W. and W. M. Van Der Aalst
2007. Fuzzy mining–adaptive process simplification based on multi-perspective metrics. In *International conference on business process management*, Pp. 328–343. Springer.

Hand, D. J.
2007. Principles of data mining. *Drug safety*, 30(7):621–622.

Janssenswillen, G.
2020. *processmapR: Construct Process Maps Using Event Data*. https://www.bupar.net, https://github.com/bupaverse/processmapr.

Kirchner, K., N. Herzberg, A. Rogge-Solti, and M. Weske
2012. Embedding conformance checking in a process intelligence system in hospital environments. In *Process Support and Knowledge Representation in Health Care*, Pp. 126–139. Springer.

Mans, R. S., M. Schonenberg, M. Song, W. M. Van Der Aalst, and P. J. Bakker
2008. Application of process mining in healthcare–a case study in a Dutch hospital. In *International joint conference on biomedical engineering systems and technologies*, Pp. 425–438. Springer.

Meinard, Y. and G. Thébaud
2019. L'identification syntaxonomique dans les démarches de gestion et/ou de restauration d'espaces naturels en France: pour ou contre? *Naturae*, 2019(6):165–173.

Meinard, Y. and A. Tsoukiàs
2019. On the rationality of decision aiding processes. *European Journal of Operational Research*, 273(3):1074–1084.

Rebuge, Á., L. V. Lapão, A. Freitas, and R. Cruz-Correia
2013. A process mining analysis on a virtual electronic patient record system. In *Computer-Based Medical Systems (CBMS), 2013 IEEE 26th International Symposium on*, Pp. 554–555. IEEE.

Richard, A., B. Mayag, Y. Meinard, F. Talbot, and A. Tsoukiàs
2018. How AI could help physicians during their medical consultations: An analysis of physicians' decision process to develop efficient decision support systems for medical consultations. In *PFIA 2018*, Nancy, France.

Richard, A., B. Mayag, F. Talbot, A. Tsoukias, and Y. Meinard
2020a. Transparency of classification systems for clinical decision support. In *Information Processing and Management of Uncertainty in Knowledge-Based Systems*, M.-J. Lesot, S. Vieira, M. Z. Reformat, J. P. Carvalho, A. Wilbik, B. Bouchon-Meunier, and R. R. Yager, eds., Pp. 99–113, Cham. Springer International Publishing.

Richard, A., B. Mayag, F. Talbot, A. Tsoukiàs, and Y. Meinard
2020b. What does it mean to provide decision support to a responsible and competent expert? *EURO journal on decision processes.*

Richard, A., B. Mayag, F. Talbot, A. Tsoukias, and Y. Meinard
2021. A virtual assistant dedicated to supporting day-to-day medical consultations. In *2021 IEEE 9th International Conference on Healthcare Informatics (ICHI)*, Pp. 330–338.

Rojas, E., J. Munoz-Gama, M. Sepúlveda, and D. Capurro
2016. Process mining in healthcare: A literature review. *Journal of biomedical informatics*, 61:224–236.

Van Der Aalst, W.
2011. *Process mining: discovery, conformance and enhancement of business processes*, volume 2. Springer.

Van Der Aalst, W., A. Adriansyah, A. K. A. De Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. Van Den Brand, R. Brandtjen, J. Buijs, et al.
2011. Process mining manifesto. In *International Conference on Business Process Management*, Pp. 169–194. Springer.

Van Der Aalst, W., A. Weijters, and L. Maruster
2004. Workflow mining: Discovering process models from event logs. *IEEE Transactions on Knowledge and Data Engineering*, 16(9):1128–1142.

Van Der Aalst, W. M., H. A. Reijers, A. Weijters, B. F. van Dongen, A. A. De Medeiros, M. Song, and H. Verbeek
2007. Business process mining: An industrial application. *Information Systems*, 32(5):713–732.

Van Dongen, B. F., A. K. A. de Medeiros, H. Verbeek, A. Weijters, and W. M. Van Der Aalst
2005. The prom framework: A new era in process mining tool support. In *International conference on application and theory of petri nets*, Pp. 444–454. Springer.

Verbeek, H., J. Buijs, B. Van Dongen, and W. M. Van Der Aalst
2010. Prom 6: The process mining toolkit. *Proc. of BPM Demonstration Track*, 615:34–39.

Weijters, A., W. M. Van Der Aalst, and A. A. De Medeiros
2006. Process mining with the heuristics miner-algorithm. *Technische Universiteit Eindhoven, Tech. Rep. WP*, 166:1–34.

RÉSUMÉ

Réalisée en collaboration avec le GIE Hopsis et les employés des Hospices Civils de Lyon (HCL), cette thèse a pour objectif de proposer une réflexion sur les contraintes et les enjeux liés à l'introduction d'outils d'aide à la décision dans le processus de travail de médecins lors de consultations médicales. Nos travaux se sont organisés autour de trois axes principaux. Une étude des outils actuellement employés pour soutenir le personnel soignant dans leurs processus de décision, qui nous a permis de mettre en évidence les limites des approches actuelles pour l'aide aux décisions médicales. En second lieu, une analyse du processus de décision de médecins travaillant aux HCL, qui nous a permis de mettre en évidence le besoin en informations des médecins afin de prendre des décisions concernant leurs patients. Et enfin, la proposition d'un outil d'aide à la décision, qui vise à l'apprentissage et à l'anticipation des besoins en informations des médecins durant leurs consultations médicales coutumières.

MOTS CLÉS

Aide à la décision, Systèmes d'information hospitaliers, Analyse du processus de décision, Systèmes de classification multi-labels, Transparence des algorithmes

ABSTRACT

Conducted in partnership with the GIE Hopsis and the employees of the Hospitals of Lyon (HCL), this thesis proposes a reflection about the constraints and the challenges linked to the introduction of decision support systems in the workflow of physicians during medical consultations. Our work is organized into three main axes. Firstly, the study of current decision support systems used in healthcare contexts to support clinicians during their decision processes, which allowed us to highlight the limits of current approaches used to support decisions in customary clinical situations. Secondly, the analysis of HCL's physicians' decision process, which allowed us to highlight the physicians' need for patients' information to be able to take relevant decisions. And lastly, the proposal of a decision support system, which aims to learn and to anticipate the patients' information needed by physicians during their customary medical consultations.

KEYWORDS

Decision support, Health Information Systems, Decision Analysis, Multi-label Classification Systems, Algorithmic Transparency