

DATA SCIENCE AND MACHINE LEARNING AT THE SERVICE OF CLINICAL DECISION-MAKING IN ONCOLOGY

Health institutions, and hospitals in particular, deal on a daily basis with an object of immense value: data. If these data could be extracted out of the textual clinical records and fed into machine learning models developed for assisting physicians in clinical decision-making, new opportunities would arise to clinical practice, clinical research, but above all to patients.

**CATARINA SANTOS^(1,2),
MÁRIO AMORIM LOPES^(1,2)**

⁽¹⁾INESC TEC;

⁽²⁾FACULTY OF ENGINEERING OF THE UNIVERSITY OF PORTO

catarina.s.santos@inesctec.pt

mario.a.lopes@inesctec.pt

Despite the constant evolution in oncology, shown through the emergence and availability of new therapeutic options, cancer remains one of the most prevalent diseases (affecting, in 2017, around 233 million people worldwide) and one of the leading causes of death in developed countries, which creates enormous social and economic burdens. In fact, with an annual cost of more than 199 billion Euros in Europe alone – and with a significant portion of that amount being spent exclusively on medication –, finding a cure for cancer has proved to be an extremely complex, time-consuming and expensive process.

As part of the procedures to diagnose and treat patients with cancer, massive amounts of data need to be collected by clinicians. This information is stored in the patient's clinical process, and it includes indicators of his/her general health status, history, exams and diagnoses, follow-up notes, among many others. This information, which has an enormous clinical – but also scientific – value, is generally kept in free text, hampering its use in diverse purposes, including statistical treatment or clinical decision making. Even worse, it is an addition to the physician's tasks, as he/she has to analyze huge volumes of text, and leads to errors, such as exam duplications.

The objective of the Mine4Health investigation project is to help solve, or at least ease, the aforementioned problems, through two main contributions. First, using a Natural Language Processing approach, it aims to convert the free text present in clinical records into structured and chronologically organized blocks stored in a database, so that they can become each patient's clinical narrative and later be retrieved for clinical, research and management purposes. Figure 1 illustrates this idea.

The second contribution is more ambitious. The goal is to use the systematized results from the previous step, to develop and improve machine learning and artificially intelligent models so that they can be used to support clinical decision-making, particularly to predict treatment responses (predictive models) or to suggest procedures and actions (prescriptive models), considering each patient's individual characteristics, including their age, gender, ethnicity, comorbidities, previous conditions and biological profile. In this context, the development of new models can become an important ally in the fight against cancer – provided data exists and can be used.

Although previous scientific work devoted to this topic already exists, it is scarce at a national level, and it tends to resort to small datasets, limiting the applicability of its results. Hence, the main advantage of this project is the partnership with IPO Porto – Portugal's largest oncological hospital, and one of the largest in Europe –, which allowed obtaining over 10 years of duly protected medical records, containing information about over 795 808 unique patients and 7 791 918 clinical episodes, and with 2000 records created or updated on a daily basis. As a contribution to society, we expect the methods developed in the context of this project to be able to help healthcare professionals in the decision-making process, thus exploring the potential of the clinical experience of IPO Porto, and assisting decision making with specific recommendations and guidelines for each patient. Among many other possible applications, this tool can assist in performing patient stratification according to the risk of relapsing, developing metastases or undergoing a certain treatment or intervention, and reduce the need for invasive exploratory procedures.

By being based on all relevant information and providing up-to-date recommendations – that is, according to the most recent and scientifically validated oncological techniques – this tool will have the potential to avoid unnecessary expenses by significantly reducing the number of misdiagnoses and misprescriptions, to reduce the workload suffered by clinicians and even detect subtle markers that might not typically be considered by physicians. The products of this work can also be transferred to other oncological facilities, as well as to generalist hospitals with medical oncology services, national and international, which will encourage sharing practices between centers and facilitate future cancer research.

Finally, and despite all the efficiency gains, this tool can be translated into faster and more accurate diagnoses, personalized treatments according to the patient's biological characteristics and his/her specific cancer, a better understanding of his/her diagnosis and therapeutic options, and better health care for the population as a whole.

ESPECIAL EDITION - DATA SCIENCE, ARTIFICIAL INTELLIGENCE AND HEALTH

ARTIFICIAL INTELLIGENCE AND THE SCREENING OF DIABETIC RETINOPATHY

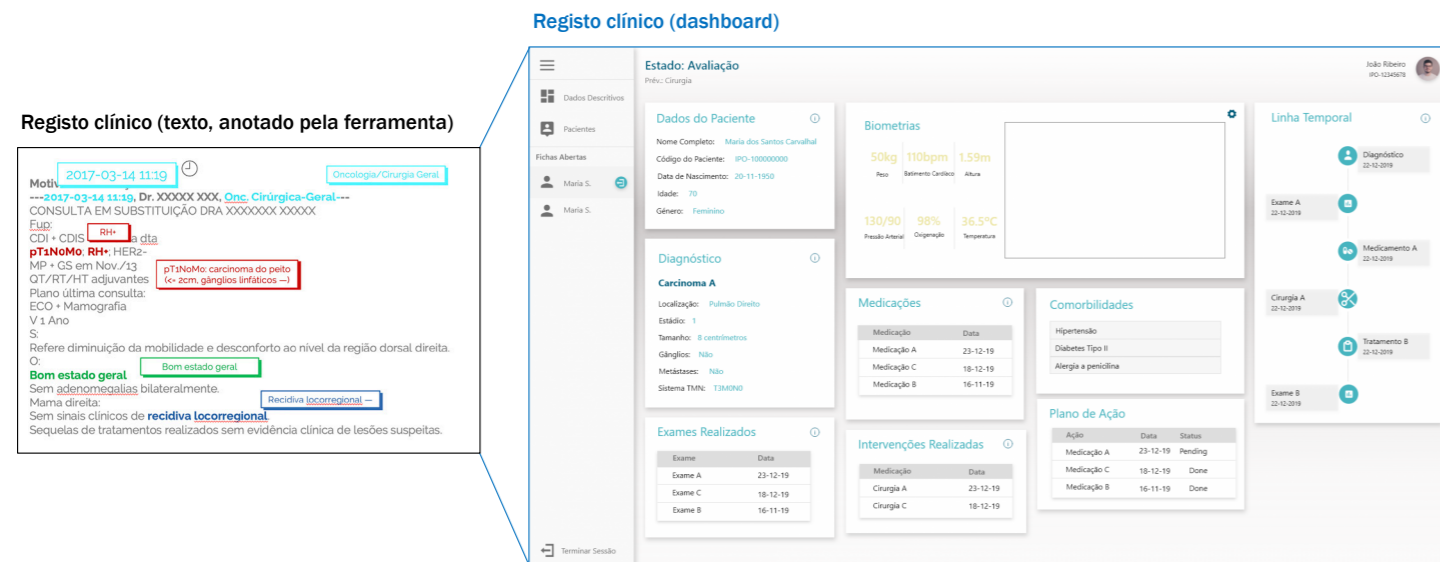


Figure 1. Example of the extraction of structured data to be used for building dashboards.

Artificial Intelligence can be useful in scenarios where skill labour fails to meet needs. This article describes the challenges and solutions found to face the diagnosis of diabetic retinopathy considering the increasing prevalence of diabetes in the Portuguese population.

TERESA ARAÚJO⁽¹⁾,
 PEDRO COSTA⁽¹⁾,
 CATERINA CARVALHO⁽¹⁾

⁽¹⁾ INESC TEC;
 teresa.f.araujo@inesctec.pt
 pedro.v.costa@inesctec.pt
 catarina.b.carvalho@inesctec.pt