

# A CIÊNCIA DOS DADOS E A APRENDIZAGEM COMPUTACIONAL AO SERVIÇO DA DECISÃO CLÍNICA ONCOLÓGICA\*

As instituições de saúde, e os hospitais em particular, lidam diariamente com um objecto de valor incalculável: dados. Se esses dados puderem ser extraídos dos registos clínicos e originar modelos de machine learning de apoio à decisão clínica, abrem-se novas possibilidades para a prática clínica, para a investigação, mas sobretudo para os pacientes.

**CATARINA SANTOS<sup>(1,2)</sup>,  
MÁRIO AMORIM LOPES<sup>(1,2)</sup>**

<sup>(1)</sup> INESC TEC;

<sup>(2)</sup> FACULDADE DE ENGENHARIA DA UNIVERSIDADE DO PORTO

catarina.s.santos@inesctec.pt

mario.a.lopes@inesctec.pt

\*Artigo escrito de acordo com a antiga ortografia

Apesar da evolução constante em oncologia, demonstrada através da disponibilidade de novas opções terapêuticas, o cancro continua a ser uma das doenças com maior prevalência (afectava, em 2017, cerca de 233 milhões de pessoas em todo o mundo) e uma das principais causas de morte nos países desenvolvidos, o que gera um enorme fardo social e económico. De facto, com um custo anual de mais de 199 biliões de Euros só na Europa – e com uma porção significativa dessa quantia a ser gasta exclusivamente em medicação –, encontrar uma cura para o cancro tem-se mostrado um processo extremamente complexo, moroso e dispendioso. Como parte dos procedimentos para diagnóstico e tratamento de pacientes é necessária a recolha diária, por parte dos profissionais de saúde, de enormes quantidades de dados. Esta informação é armazenada no processo clínico do paciente, e engloba os indicadores do seu estado geral de saúde, histórico, exames e diagnósticos, notas de acompanhamento médico, entre muitos outros. Esta informação, que aporta um enorme valor clínico, mas também científico, é geralmente guardada em texto livre, o que torna difícil a sua utilização para os mais diversos fins, incluindo tratamento estatístico ou tomada de decisão clínica. Mais grave ainda, dificulta a tarefa do médico, pois tem de analisar enormes volumes de texto, e leva a erros como duplicação de exames.

O projeto de investigação Mine4Health tem o objetivo de ajudar a resolver ou, pelo menos, minimizar os problemas descritos acima, através de dois principais contributos. Primeiro, através de uma abordagem de processamento de linguagem natural, visa converter o texto livre presente nos registos clínicos em blocos estruturados e organizados cronologicamente, guardados numa base de dados, para que se transformem na narrativa clínica de cada doente e possam, posteriormente, ser aproveitados para fins clínicos, de investigação e de gestão. Já o segundo contributo é mais ambicioso. O objectivo é utilizar os resultados sistematizados da etapa anterior, desenvolver e aprimorar modelos de Aprendizagem Computacional e de Inteligência Artificial para que possam ser usados para o apoio à decisão clínica, em particular para prever a resposta a tratamentos (modelos preditivos) ou sugerir procedimentos e acções (modelos prescritivos), tendo em conta as características individuais de cada paciente, incluindo a sua faixa etária, género, etnia, comorbilidades, condições prévias e perfil biológico. Neste âmbito, o desenvolvimento de novos modelos pode tornar-se num importante aliado no combate contra o cancro — desde que os dados existam e possam ser usados. A Figura 1 ilustra esta ideia.