

tecnologia

USO DE FERRAMENTAS DE *BIG DATA* NA PESQUISA E NO ENSINO ACADÊMICO CONTRIBUEM PARA O AVANÇO NO DIAGNÓSTICO E NO TRATAMENTO DO CÂNCER

Volume, variedade e velocidade

O vertiginoso desenvolvimento que a tecnologia da informação apresentou a partir da segunda metade da década de 2000, aliado à popularização dos computadores e ao crescente acesso à Internet, vêm contribuindo para o uso cada vez mais frequente do *Big Data*. Essa ferramenta proporciona maneiras de extrair informações de forma eficiente, aproveitando as oportunidades oferecidas pelos avanços tecnológicos. Em relação à oncologia, uma das principais características consiste em

permitir análise mais rápida de um alto volume de informações de pacientes. As análises resultantes vêm sendo utilizadas em epidemiologia clínica para produção de conhecimento, em especial àquele relacionado à incidência e à mortalidade por câncer.

De acordo com Luiz Claudio Santos Thuler, pesquisador da Divisão de Pesquisa Clínica do INCA e docente do Programa de Pós-Graduação em Oncologia, o *Big Data* facilita o estabelecimento de características, padrões, diagnósticos e trata-



“Para alguns tipos de tumor, é uma realidade, com a prescrição de terapias mais individualizadas, que evitam, por exemplo, efeitos colaterais da quimioterapia convencional”

MARIANA BORONI, chefe do Laboratório de Bioinformática de Biologia Computacional do INCA

Secomso/INCA



mentos do câncer. “Isso pode refletir na melhoria do ensino, tornando-o mais próximo da realidade. É o que chamamos atualmente de educação baseada em dados. Na área de genética também é muito valioso, pois possibilita procedimentos terapêuticos mais precisos e personalizados.”

Mariana Boroni, chefe do Laboratório de Bioinformática de Biologia Computacional do INCA, concorda. Para ela, o uso de ferramentas de *Big Data* proporcionou a chance de enxergar o câncer de maneira multifatorial. “Elas permitiram a evolução das ciências ômicas [relacionadas a estudos de DNA, RNA, proteínas e metabólitos] e nos deram a capacidade de avaliar a materialidade genética e molecular da doença. Foi uma revolução gigantesca, que deu grande impulso à medicina de precisão [alia o tratamento convencional ao perfil genético de cada indivíduo].”

A médica explica que muitas dessas técnicas ainda estão em fase de pesquisa, embora haja aplicações práticas em tratamentos. “Para alguns tipos de tumor, é uma realidade, com a prescrição de terapias mais individualizadas, que evitam, por exemplo, efeitos colaterais da quimioterapia con-

venção. Hoje já é possível um procedimento mais assertivo. O próprio Sistema Único de Saúde oferece tratamentos customizados.”

Um deles é o trabalho desenvolvido pelo Laboratório de Bioinformática com foco no câncer de ovário, cujos sintomas são similares aos da menstruação, como dores e inchaço abdominais. “Isso resulta, muitas vezes, em um diagnóstico tardio, e nem todas as mulheres respondem bem ao tratamento. A vantagem do *machine learning* [área da Inteligência Artificial (IA) que se concentra no uso de dados para imitar a maneira como os indivíduos aprendem ou executam tarefas específicas, melhorando gradualmente sua precisão] é que os algoritmos conseguem enxergar padrões [em múltiplas dimensões], algo impossível para o ser humano, pois o volume de informação é enorme. Estamos falando da análise de mais de 20 mil genes simultaneamente. Podemos utilizar informações de mulheres que não reagem bem à quimioterapia e pensar em um tratamento específico. A Inteligência Artificial nos ajuda a estabelecer padrões que podem levar a outras abordagens terapêuticas”, afirma.

DEMANDA NO MUNDO ACADÊMICO

Segundo Catherine Moura, médica sanitária e CEO da Associação Brasileira de Linfoma e Leucemia, as ferramentas de *Big Data* e IA vêm sendo utilizadas na Oncologia como forma de aperfeiçoamento técnico em programas de pós-graduação, incluindo mestrado e doutorado. “Elas ajudam a levantar informações do paciente e a oferecer assistência e tratamento de uma maneira geral, contribuindo, assim, para acompanhar desde a efetividade das legislações até o desfecho da jornada de cada um”, diz.

O QUE É O *BIG DATA*?

O Big Data é um gigantesco banco de dados, mas não um mero local de armazenamento. Nele, podem ser depositados volumes massivos de dados estruturados, semiestruturados e não estruturados para posterior análise. Os bancos de dados convencionais trabalham apenas com dados estruturados, tipicamente, em planilhas com linhas e colunas ou com objetos digitais chamados de “entidades” (tabelas) que se relacionam ou não entre si.

Já o Big Data se refere a conjuntos de dados que são tão grandes, complexos e diversos que os métodos tradicionais de processamento e gerenciamento não são adequados para lidar com eles de forma eficiente. É centrado nos chamados “3Vs”: volume (quantidade), velocidade (taxa de entrada/saída) e variedade (diversidade de tipos de dados). Também está voltado para análise de padrões, tendências, insights e tomada de decisões a partir de dados complexos e massivos. Os dados não estruturados armazenados pelo Big Data podem incluir texto livre, imagens, áudio, vídeo etc.

É projetado para escalabilidade (capacidade de um sistema se expandir para lidar com aumento na demanda) horizontal extrema, distribuindo dados e tarefas em vários nós (ou seja, em cada servidor ou computador que faz parte de um conjunto espalhado de máquinas) de processamento (manipulação de dados). Na prática, o Big Data é muito usado para análise de redes sociais, processamento de dados científicos, monitoramento de tráfego, avaliação de logs (registros que capturam informações sobre atividades, eventos, transações e comportamentos que ocorrem em um sistema).

Alexandre Dias Porto Chiavegatto Filho, professor da Faculdade de Saúde Pública da Universidade de São Paulo e diretor do Laboratório de Big Data e Análise Preditiva em Saúde (Labdaps), afirma que a demanda para que mais disciplinas sobre ciência de dados e Inteligência Artificial sejam incluídas no ensino acadêmico tem partido, principalmente, dos alunos. “Muitos departamentos estão atualmente ampliando sua atuação nesse universo tecnológico, porém isso ainda está longe de ser uma realidade no Brasil.”

Mas levar o pragmatismo das ciências exatas para o setor saúde nem sempre é tarefa fácil. De acordo com Mariana Boroni, tudo o que é novidade acaba gerando um natural estranhamento. “Em especial no campo das ciências da vida, falar de modelos matemáticos gera certa resistência. As pessoas não entendem bem como aquele padrão foi identificado. No entanto, observo que os profissionais da ciência da computação vêm desenvolvendo ferramentas para tornar o *machine learning* mais explicável, uma maneira de demonstrar como a máquina chegou àquele raciocínio.” A pesquisadora também afirma que já é possível identificar uma mudança de comportamento nos profissionais de saúde. “Tenho notado que médicos estão buscando ampliar seus conhecimentos, especializando-se nesse aspecto tecnológico.”

Para Luiz Thuler, a resistência a mudanças é comum, sobretudo quando há desconhecimento e incerteza em relação ao novo. “Qualquer inovação pode so-

frer oposição, principalmente quando percebida como ameaça. Especificamente na área médica, muito se tem enfatizado que o uso do *Big Data* não substituirá a experiência clínica e o julgamento dos profissionais, mas vai operar como um complemento ou ferramenta de apoio à prática clínica e ao ensino”, afirma.

SUBSÍDIO PARA TOMADA DE DECISÃO

Não é o que pensa Catherine Moura. Segundo ela, de maneira geral não há resistência. Professores e alunos têm utilizado o *Big Data* para o aprendizado, não apenas no campo médico. “Bancos de dados têm proporcionado muitas descobertas interessantes do ponto de vista prático para a epidemiologia e a tomada de decisão sobre terapias. Para o futuro, a tendência é que haja crescimento contínuo dessa tecnologia no tratamento de câncer, contemplando toda a linha de cuidado.”

Mesmo já sendo usado no campo acadêmico, o *Big Data* ainda não provocou uma transformação que incorpore as novas tecnologias à grade curricular. “Desconheço alterações formais”, diz Thuler. “Mas, na prática, é impossível não incluir os avanços que têm ocorrido nos anos recentes.” Para Catherine Moura, ainda há passos a serem dados. “Não existe previsão de mudança no conteúdo educacional, ao menos no Brasil, que contemple o uso dessas inovações. Infelizmente, há coisas básicas



“Quando há inovação, automaticamente seu risco precisa ser avaliado. Do ponto de vista acadêmico, essas questões são discutidas há tempos”

CATHERINE MOURA, médica sanitária e CEO da Associação Brasileira de Linfoma e Leucemia

Divulgação



nas quais ainda precisamos evoluir, como a inclusão da disciplina Oncologia tanto na graduação do médico como na de outros profissionais envolvidos no cuidado da doença”, argumenta.

SEGURANÇA E ÉTICA EM FOCO

Ao lidar com dados sensíveis, as ferramentas exigem não somente bom senso e cuidado da parte de desenvolvedores, cientistas e médicos, mas também uma abordagem honesta, segundo Thuler. “A privacidade e a segurança dos dados devem seguir sendo garantidas por sua anonimização [técnica de tratamento automático de informação], já prevista nos códigos de ética médica e pela Comissão Nacional de Ética em Pesquisa”, afirma. Para Catherine Moura, a Lei Geral de Proteção de Dados Pessoais é uma aliada nesse campo. “Quando há inovação, automaticamente seu risco precisa ser avaliado. Do ponto de vista acadêmico, essas questões são discutidas há tempos. Um exemplo são os estudos clínicos amparados pelos comitês de ética em pesquisa, nos quais o investigador, desde a graduação, aprende sobre essas temáticas. Com o *Big Data* não será diferente.”

Mariana Boroni salienta a grande preocupação com a segurança dos dados, pois informações genéticas podem ser usadas tanto eticamente quanto para finalidades indesejadas. “Existe o temor de que operadoras de planos de saúde possam prever algumas



Shutterstock

doenças e venham a recusar clientes ou que, então, cobrem valores muito elevados para aceitá-los. Há também o risco de que empresas usem esses dados em seleções para vagas de emprego, eliminando candidatos que possam apresentar futuros problemas de saúde. Só que o câncer está muito mais relacionado às consequências de maus hábitos do que pela predisposição de cada um. É uma discussão calorosa, que merece mais atenção”, acredita. Ela considera que existem outros aspectos importantes além da privacidade, como a equidade. “Nesse ponto, ainda há muito a avançar. As ferramentas dependem dos *inputs* [as entradas, as informações iniciais] que damos a elas.”

Alexandre Chiavegatto, por sua vez, chama a atenção para os dados que serão inseridos nos algoritmos. “Se forem treinados com um perfil específico de paciente, por exemplo de regiões mais ricas, corre-se o risco de que tomem boas decisões para esse grupo e piores para os que residem em localidades mais remotas. Será importante adicionar debates éticos nas novas disciplinas de tecnologia”, opina.

MODELOS NACIONAIS

Na área médica, os recortes geográficos, comportamentais e étnicos são relevantes. Entretanto, a maior parte dos modelos usados pelas ferramentas de *Big Data* é baseada em informações internacionais, o que exige adaptação à realidade brasileira. “Isso não é novidade, pois, no ensino, a prática é antiga. Há muita limitação de dados epidemiológicos, de incidência e de prevalência de diversas patologias no País, o que limita os estudos de tendência e análises preditivas. No entanto, existem iniciativas de instituições educacionais no Brasil para montar painéis de controle analíticos, com informações nacionais em sua totalidade”, diz Catherine Moura.

Esse aspecto também é apontado por Alexandre Chiavegatto, que considera que a transposição de algoritmos para realidades além daquelas em que foram programados será um desafio. “Resultados recentes do nosso laboratório indicam que os softwares treinados nos grandes centros urbanos brasileiros perdem muito de sua precisão quando são usados em outras regiões. A tendência é que essa queda na qualidade seja ainda maior se forem usados algoritmos alimentados com dados de pacientes de outros países”, afirma o diretor do Labdaps.

Na opinião de Luiz Thuler, é importante que as ferramentas de *Big Data* utilizadas no tratamento do câncer sejam cada vez mais abrangentes e atualizadas, aprimorando o conhecimento de como o câncer ocorre no Brasil e aprofundando o aprendizado de profissionais e estudantes. “Embora existam inúmeras publicações internacionais sobre a importância da etnia na incidência e na mortalidade por tumor do colo do útero, por exemplo, por meio de três grandes bases de dados nacionais, nosso grupo mostrou, pela primeira vez no País, que as taxas são maiores em mulheres negras [neste estudo incluindo pretas e pardas, conforme classificação do IBGE] e indígenas, sendo possível conhecer sua vulnerabilidade e apontar caminhos para reduzir as disparidades”, conta.

Para funcionar, a tecnologia depende de dados, mas o custo para gerá-los é alto. Mariana Boroni esclarece que a população brasileira é miscigenada e, conseqüentemente, possui uma heterogeneidade molecular bem maior do que a de outros países. O problema, de acordo com ela, é que muitos estudos internacionais utilizados aqui são baseados em populações caucasianas. “Estamos começando a gerar nossos próprios dados. O Governo Federal lançou o projeto Genoma Brasil, que tem como objetivo criar um banco de informações genômicas de vários tumores. E o INCA participa desse trabalho. Isso vai nos permitir desenvolver ferramentas mais assertivas para os tratamentos”, acredita. Para Luiz Thuler, o ideal é desenvolver modelos próprios voltados para o ensino em Oncologia. “Meu grupo já vem retratando os padrões de ocorrência de diferentes tipos de câncer no Brasil. Isso vem permitindo o reconhecimento de características até então desconhecidas, o que pode contribuir para a melhora da prática clínica e do ensino.”

Mas essa realidade não vai se concretizar se não houver investimentos em pesquisa, avalia Mariana Boroni. “Precisamos destinar mais recursos



Divulgação

“Isso pode refletir na melhoria do ensino, tornando-o mais próximo da realidade. É o que chamamos atualmente de educação baseada em dados. Na área de genética também é muito valioso”

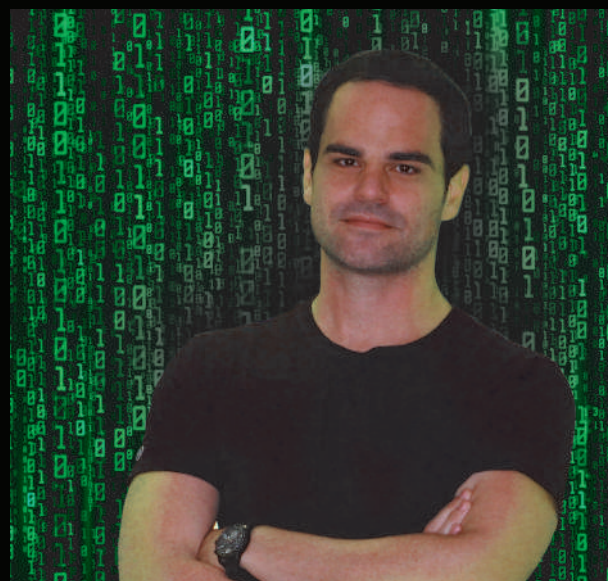
LUIZ CLAUDIO SANTOS THULER, pesquisador da Divisão de Pesquisa Clínica do INCA e docente do Programa de Pós-Graduação em Oncologia

à ciência e à pesquisa, pois a geração dos dados é muito cara. Em um futuro próximo, teremos um banco com muitas informações genéticas. Isso nos permitirá fazer diferentes modelos, que preencherão essas lacunas, tanto no diagnóstico quanto no tratamento.”

NOVA FORMA DE LINGUAGEM

O ChatGPT, algoritmo generativo de Inteligência Artificial que interage com humanos por meio do oferecimento de soluções de textos e que vem ganhando notoriedade nos últimos dois anos, já começa a ser usado como ferramenta de apoio para professores e alunos de Oncologia. “Ele tem servido para obter exemplos a serem usados em sala de aula e em questões para avaliação dos estudantes”, conta Luiz Thuler. “Como desconheço seu uso para análise de dados, resolvi perguntar à própria plataforma e obtive a seguinte resposta: ‘Ele não é uma ferramenta específica para análise de *Big Data*, mas um modelo de linguagem’. Ele segue informando que pode ser usado para extrair informações relevantes, classificar ou categorizar, resumir ou sumarizar dados textuais, como registros médicos e relatórios de pacientes. Testei e funciona. Penso que sua utilização nesse campo é promissora, mas requer, como em todas as suas funcionalidades, julgamento crítico do usuário.”

A ressalva é corroborada por Alexandre Chiavegatto, para quem é fundamental esclarecer que o ChatGPT não foi desenvolvido nem validado para finalidades médicas. “Vários profissionais usam essa ferramenta para auxiliá-los em dúvidas clínicas. No entanto, é importante que eles façam perguntas em áreas que



Divulgação

Segundo Chiavegatto, alunos têm pedido que mais disciplinas sobre ciência de dados sejam incluídas no ensino acadêmico

já dominam e utilizem a plataforma apenas para consultas, nunca para tomar uma decisão final”, ressalta.

Catherine Moura pondera que a plataforma, assim como foi o *Big Data*, é uma nova tecnologia de inovação e pode contribuir com a aprendizagem de forma mais assertiva e coordenada. Porém, considera que sua disponibilização é muito precoce e precisa ser mais explorada. “Há um tempo de adoção plena com efeitos controlados sobre os impactos promovidos. Mas a questão é que, seja *Big Data*, IA ou ChatGPT, é inegável a contribuição que a tecnologia oferece em termos de ampliação de geração de dados, informação e conhecimento, além do que ainda trará para a oncologia, para a saúde e para a sociedade em geral.” ■