

Pesquisadores usam inteligência artificial para definir causas da anemia aplástica

Estudo prospectivo com a colaboração da USP cria ferramenta para acelerar o diagnóstico e identificar se doença rara é adquirida ou hereditária, o que agiliza a definição do tratamento da condição potencialmente fatal

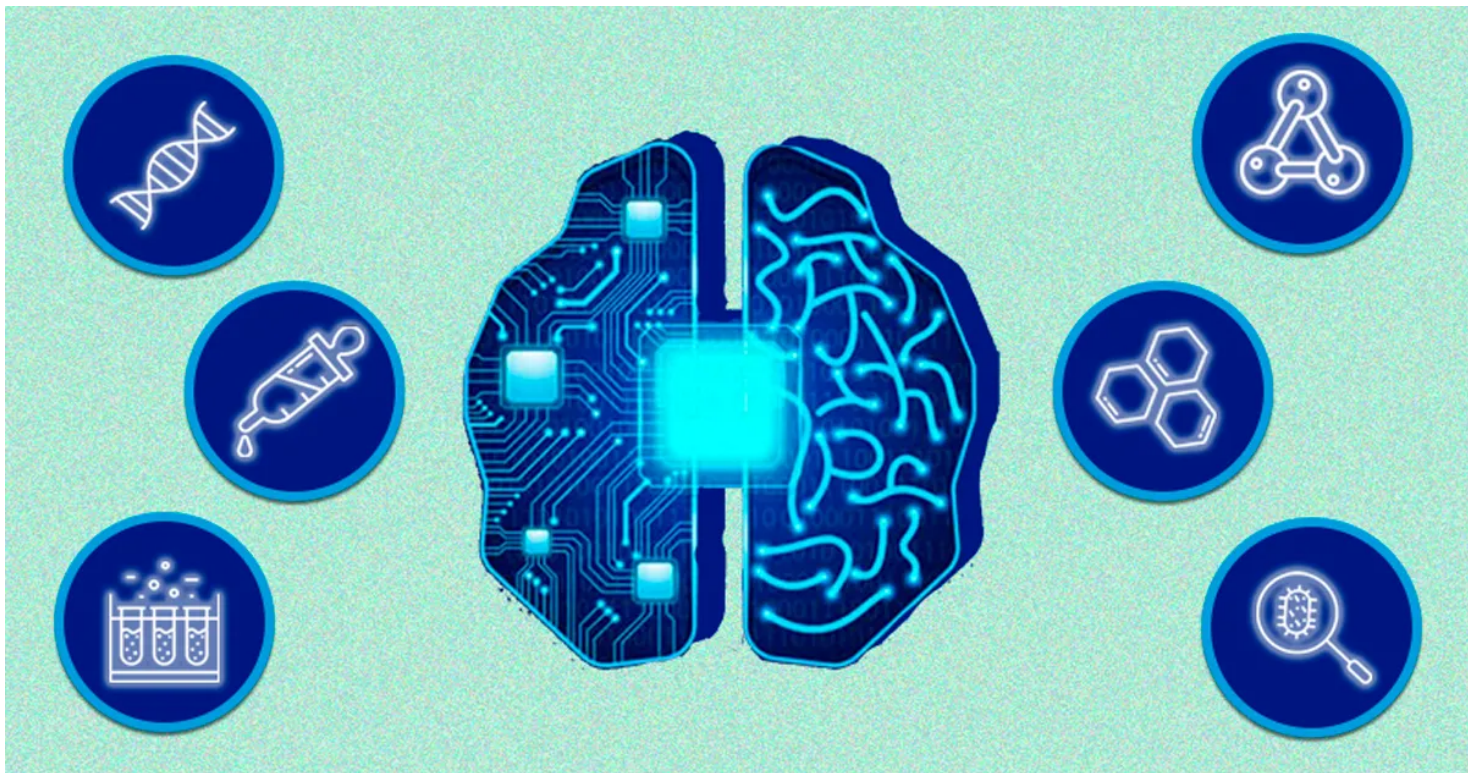
Publicado: 03/03/2023

Texto: Ivan Conterno
Arte: Carolina Borin Garcia

Pesquisadores do Centro de Terapia Celular (CTC) da USP e dos Institutos Nacionais de Saúde (NIH) dos Estados Unidos desenvolveram uma tecnologia capaz de acelerar o diagnóstico da causa da anemia aplástica, um quadro raro que causa a redução do número de todas as células do sangue, como glóbulos vermelhos, glóbulos brancos e plaquetas. Os casos graves podem ser fatais, uma vez que essas células, cada uma à sua maneira, são indispensáveis para o bom funcionamento do corpo humano.

Dentre os pacientes com a doença, são comuns duas situações: por um lado, os casos de origem hereditária e, por outro, uma condição adquirida no percurso da vida, em pessoas cujo sistema imunológico passa a atacar células saudáveis. Os sintomas são muito parecidos, mas as causas são completamente diferentes. Por isso, para iniciar a terapia, é preciso identificar a causa da enfermidade, adquirida ou hereditária, através da avaliação de exames laboratoriais e clínicos.

Para criar o modelo preditivo, os cientistas se basearam em exames clínicos e laboratoriais de rotina e utilizaram, para isso, uma inteligência artificial com aprendizado automático. Essa importante inovação é foco do artigo *Differential diagnosis of bone marrow failure syndromes guided by machine learning*, publicado no periódico *Blood*, da American Society of Hematology, a principal revista no mundo de hematologia, especialidade médica responsável pelo estudo do sangue. O trabalho conta com a colaboração do professor Rodrigo Calado, chefe do Departamento de Imagens Médicas, Hematologia e Oncologia Clínica da Faculdade de Medicina de Ribeirão Preto (FMRP) da USP, presidente executivo do Hemocentro de Ribeirão Preto e principal pesquisador do CTC.



Inteligência artificial tem se mostrado uma grande aliada em diferentes setores da saúde - Fotomontagem: Jornal da USP/Luana Franzão

A utilização de inteligências artificiais para a interpretação dos exames pode ser uma forma de acelerar o diagnóstico que diferencie a origem da doença nos serviços de saúde, que nem sempre contam com todas as especialidades médicas no atendimento. Isso pode agilizar o tratamento, já que a resposta de um teste genômico para distinguir se a doença é hereditária ou adquirida costuma demorar

de seis a oito semanas, nos Estados Unidos. Em países em desenvolvimento, esse tempo pode chegar a 12 semanas ou mais. A tecnologia permitirá que o encaminhamento correto seja mais rápido e com pouca margem de erro.

Ao **Jornal da USP**, Calado ressaltou que toda uma equipe de hematologia, tanto da USP em Ribeirão Preto quanto em São Paulo, contribuiu com o projeto. “Nós classificamos os pacientes do ponto de vista genômico para saber se tinham doença hereditária (genética) ou doença adquirida e avaliamos cada uma das variáveis que seriam utilizadas no processo de aprendizado de máquina para fazer essa classificação. Foi um trabalho de curadoria das informações.”

Fernanda Gutierrez Rodrigues, que foi orientada em seu mestrado e doutorado por Calado na USP e hoje é pesquisadora no NIH, foi, ao lado de Eric Munger, da Universidade George Mason, a principal autora do artigo. Ela explicou ao **Jornal da USP** a importância da parceria. “Criar modelos de inteligência artificial requer um número muito grande de casos, o que se torna difícil para doenças raras. A colaboração entre diferentes institutos e a USP foi essencial para obter um número maior de diagnósticos para desenvolver o modelo.”

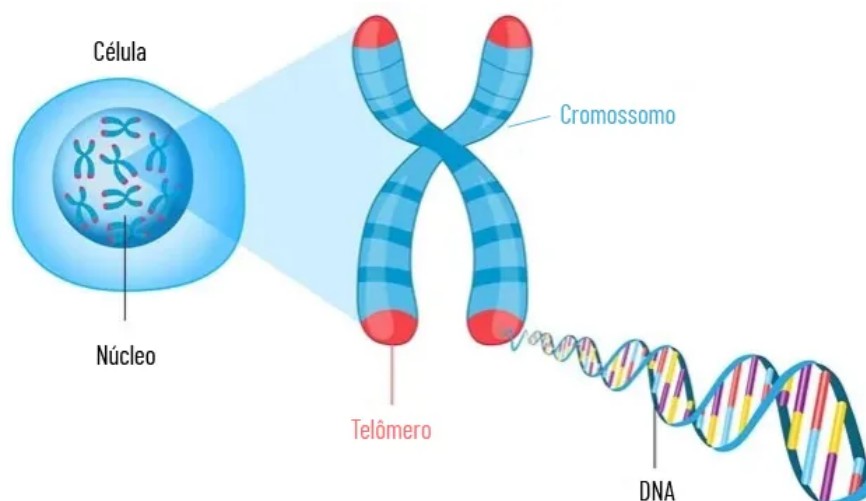


Rodrigo Calado - Foto: Lattes

Inteligência artificial

O diagnóstico errado, muitas vezes, leva à administração de medicamentos caros e sem eficácia e à inclusão do paciente na fila pelo transplante de medula óssea desnecessariamente. Foi por isso que a equipe criou uma **calculadora que facilita a investigação clínica**. Ela servirá como um guia para auxiliar profissionais da saúde no encaminhamento correto dos doentes e pode antecipar o tratamento emergencial de quadros graves da disfunção.

A partir dos resultados dos primeiros testes, constatou-se que todas as 25 variáveis coletadas no estudo foram importantes, porém a medida dos telômeros foi o principal fator da equação para determinar a origem da disfunção. A explicação é que a forma hereditária da doença pode vir acompanhada de telômeros muito curtos, enquanto que, nos casos adquiridos, o comprimento do telômero costuma ser normal. Os telômeros são estruturas presentes nas extremidades dos cromossomos (presentes no núcleo das células, que abrigam o material genético). O teste para medição dos telômeros, embora seja uma técnica recente, é muito mais barato que os exames genéticos, mas ainda sem grande acesso em muitos países.



Estrutura da célula com enfoque no cromossomo e telômero - Imagem: Cedida por Diego Villa Clé

O aprendizado de máquina já é utilizado no diagnóstico de doenças comuns. Trata-se de um mecanismo de inteligência artificial que adequa os dados a modelos probabilísticos e estatísticos automaticamente, aprendendo conforme mais informações são adicionadas aos arquivos.

Nesse estudo em específico, foi utilizado um método de agrupamento no qual a maioria dos casos, cujos exames se assemelhavam, foram separados dos demais. A razão para isso é que a população de pessoas com a doença é pequena e os casos atípicos poderiam interferir negativamente nos resultados. Essas ocorrências incomuns serão usadas quando a amostra for grande o suficiente para que eles também possam se aproximar de um padrão de dados correlacionados. Como explica Fernanda: “Quando tentamos fazer diretamente um modelo de classificação com os nossos casos consecutivos, não conseguimos um modelo de acurácia muito alta. Por isso a gente utilizou o que a gente chama de *two step modeling* [modelagem em duas etapas]. Nesse método, os pacientes são classificados em dois grupos antes de desenvolver um modelo preditivo: o de casos em que são identificadas características comuns e os casos atípicos.”



Fernanda Gutierrez Rodrigues - Foto: Arquivo Pessoal

A amostra inicial do estudo foi composta de 441 pacientes dos Estados Unidos, cujos exames seriam usados para treinar o algoritmo, e 165 da USP, que validaram o modelo criado. Nesse caso, o diagnóstico genético precisaria ser conhecido para que o algoritmo fosse treinado. Portanto, todos os voluntários se submeteram a exames de DNA previamente.

Dos 441 americanos selecionados, 82 foram descartados da amostra, parte por possuírem alguma variante genética cujo significado é desconhecido e parte por conta da falta de dados na coleta. Dos 359 restantes desse grupo, 59 apresentaram observações discrepantes e foram separados em um novo agrupamento, sobrando 300 na amostra principal.

Da mesma forma, entre os 165 voluntários selecionados pela USP, 38 não foram incluídos na validação, 11 por conta dos exames genéticos e 27 por não possuírem todas as informações clínicas coletadas. Os dados discrepantes foram observados em 17 dos 127 pacientes restantes, sobrando 110 no grupo que validaria o modelo.

O código foi capaz de encontrar padrões para o diagnóstico com 92% de precisão para casos da doença adquirida. Porém, se forem reavaliados os pacientes cujos testes genéticos divergiram do diagnóstico clínico, a acurácia do modelo sobe para 98%. Isso ocorre porque nem todas as causas genéticas para a doença são conhecidas.

Por outro lado, 20% dos casos de doença genética foram mal previstos pelo modelo. Isso decorre da alta heterogeneidade das mutações genéticas responsáveis pela síndrome e demandaria uma maior quantidade de afetados com cada uma delas para que os padrões fossem encontrados pelo algoritmo.

A inteligência artificial consegue se aperfeiçoar com pouca interferência humana, conforme mais pacientes sejam adicionados ao banco de dados. Quanto maior a amostra, mais ela será representativa da população com a doença, como explica Fernanda Rodrigues. “O nosso planejamento é continuar acrescentando mais casos, principalmente esses que a gente não tem tanta confiança, como os casos pediátricos e esses casos hereditários, para que a gente consiga refinar o modelo”.

Esse tipo de código precisa de um grande número de entradas para identificar algumas associações entre as variáveis, o que não foi possível fazer com os dados classificados como atípicos. Ou seja, com amostras maiores, será possível trabalhar também com os dados que, por enquanto, são destoantes e que foram separados do grupo principal responsável pelo treinamento da ferramenta. Mesmo com esses pequenos empecilhos, no entanto, o índice de acerto atingido foi bastante alto para uma doença que é rara.

+ Mais



**“Machine learning”
pode ser aplicado até
mesmo na medicina**



**USP
lança
curso
on-line
gratuito
sobre**

inteligência artificial na saúde



**Revista traz estudo
sobre transformação
gelatinosa da medula
óssea**

Quem sofre com esse quadro?

A falência medular pode ser adquirida devido à ativação de um componente imunológico, que ataca células saudáveis da medula, ou pode ser resultado de uma alteração genética. Ambas as causas são capazes de desencadear quadros conhecidos como anemia aplástica ou aplasia da medula. Os sintomas são quase idênticos, de diminuição das células sanguíneas, como hemácias (anemia), leucócitos (leucopenia) e plaquetas (trombocitopenia). Por ser uma doença rara, alguns dos profissionais podem não conhecer todos os dados que precisam ser levantados para um encaminhamento correto.

Casos de origem genética necessitam de transplante de medula óssea com características específicas, enquanto que ocorrências autoimunes podem ser tratadas com imunossupressão, sem efeito para o outro grupo.

O método de diagnóstico mais eficaz ainda é a detecção ou não da mutação responsável pela doença não adquirida em um teste genético. No entanto, esse teste pode ser muito caro ou estar indisponível na maioria dos serviços, especialmente nos países não desenvolvidos. Os pesquisadores acreditavam que o conjunto de outros dados correlacionados poderia permitir uma triagem onde o teste genético ainda fosse inviável.



Ações durante o acidente com o céσιο-137 em Goiânia, em 1987 - Foto: Secretaria Estadual de Saúde de Goiás

Existem também outras causas para a doença. Há 35 anos, 14 das pessoas contaminadas pelo acidente com o céσιο-137 em Goiânia, capital de Goiás, apresentaram anemia aplástica. O mesmo quadro foi identificado nas vítimas do acidente nuclear de Chernobil, um ano antes. A aquisição da doença por meio de um acidente com radiação, porém, atinge pouquíssimas pessoas.

Mais informações: e-mail rtcalado@fmrp.usp.br, com Rodrigo Calado, e e-mail fernanda.rodrigues@nih.gov, com Fernanda Gutierrez Rodrigues