

# Desenvolvimento de um *software* de auxílio ao diagnóstico clínico em genética

## *Developing a software for supporting clinical diagnosis in genetic*

David Gonçalves Nordon<sup>1</sup>, Marta Wey Vieira<sup>1</sup>

Recebido da Faculdade de Ciências Médicas e da Saúde de Sorocaba, Pontifícia Universidade Católica de São Paulo, Sorocaba, SP, Brasil.

### RESUMO

**JUSTIFICATIVA E OBJETIVO:** O diagnóstico em genética clínica é extremamente complexo, devido à plêiade de apresentações das doenças e à dificuldade de definir uma síndrome a partir dos sinais clínicos observados no paciente. Geneticistas necessitam buscar em livros-textos ou bancos de dados, a partir de cada sinal, a doença de que suspeitam. O objetivo deste estudo foi desenvolver um *software* de inteligência artificial que auxilie no diagnóstico clínico em genética. **MÉTODOS:** Foram listadas as 200 doenças genéticas mais comuns e, para cada uma, foram elencados até 10 dos principais sinais ou sintomas necessários ao diagnóstico. Os dados foram incluídos em um programa com banco de dados, que funciona no Microsoft Office Excel 2007<sup>®</sup>, por meio do preenchimento de uma ficha de anamnese e exame clínico com mais de 800 cancelas de dados. Após o preenchimento, o *software* lista as doenças mais prováveis, em ordem decrescente de semelhança com o caso clínico, com valor apresentado em porcentagem. O *software* foi aplicado em um teste alfa com 58 casos clínicos de diagnósticos concluídos no ambulatório de um serviço. **RESULTADOS:** No teste alfa, o *software* “acertou” os diagnósticos em 74% das vezes e “errou” em 8,6%. **CONCLUSÃO:** O *software* mostrou-se bastante promissor e, aperfeiçoado, deve permitir ao geneticista poupar horas de pesquisa, auxiliando o médico, e não o substituindo. Ainda é necessário o teste beta, preferencialmente interinstitucional.

**Descritores:** Inteligência artificial; Software; Doenças genéticas inatas/diagnóstico; Síndrome; Anormalidades múltiplas/diagnóstico

### ABSTRACT

**BACKGROUND AND OBJECTIVE:** Performing a diagnosis in clinical genetics is a highly complex process due to the wide

range of different presentation of a single genetic disease and to the difficulty of defining a disease based on the clinical signs of a patient. Specialists must look for each sign in textbooks or databases in order to find the disease they suspect. The objective of this study was to develop artificial intelligence software that support the clinical diagnosis in genetics. **METHODS:** The 200 most common genetic diseases were listed and the 10 main signs and/or symptoms necessary for the diagnosis were selected for each one of these disease. Data was included in a database program that runs on Microsoft Office Excel 2007<sup>®</sup>, through the filling of a chart that requires thorough anamnesis and clinical exam with over 800 open spaces for data. After filling, the software lists the most probable diseases in a decreasing order according to its resemblance to the case studied, presenting the findings in a percentage value. The software underwent an alpha testing with 58 clinical cases of defined diagnosed in our service. **RESULTS:** In the alpha testing, the software was “accurate” in 74% of the cases, and “inaccurate” in 8.6%. **CONCLUSION:** The software has shown to be very promising and, once enhanced, it may save the professional several hours in researching, directing the diagnosis and helping the doctor, rather than substitute the professional. It is still necessary to undergo an inter institutional beta-testing.

**Keywords:** Artificial intelligence; Software; Genetic diseases, inborn/diagnosis; Syndrome; Abnormalities, multiple/diagnosis

### INTRODUÇÃO

As doenças genéticas atingem uma parcela considerável da população; projeções apontam que, em países em desenvolvimento, a prevalência possa chegar a 78,6/1.000 habitantes<sup>(1,2)</sup>. Os defeitos congênitos podem representar até 25% de todos os óbitos em crianças e são especialmente importantes quando a taxa de mortalidade infantil é menor do que 50/1.000 nascidos vivos, uma vez que a Organização Mundial de Saúde (OMS) recomenda o desenvolvimento de estratégias para tratamento e prevenção de doenças genéticas a partir dessa taxa<sup>(2,3)</sup>. No Brasil, a mortalidade infantil, em 2000, era de 30,57/1.000 nascidos vivos<sup>(4)</sup>.

Tais doenças podem variar desde síndromes leves, com pouco comprometimento do indivíduo, até síndromes graves de grande comprometimento, ou letais; com isso, sua morbidade provoca um alto custo – 10 a 25% das admissões pediátricas em centros urbanos podem se dever a doenças genéticas<sup>(5,6)</sup>. No Brasil do co-

1. Pontifícia Universidade Católica de São Paulo, Sorocaba, SP, Brasil.

Data de submissão: 23/04/2013 – Data de aceite: 08/05/2014  
Conflito de interesses: não há.

#### Endereço para correspondência:

David Gonçalves Nordon  
Rua Karlina Reiman Wandabeg, 106/24 – Jardim Monte Kemel  
CEP: 05634-050 – São Paulo, SP, Brasil  
Tel.: (11) 3895-1806 – E-mail: nordondg@gmail.com

meço deste século, os defeitos congênitos eram responsáveis por 14,9% das mortes infantis, estando em segundo lugar<sup>(7)</sup>.

O diagnóstico precoce de síndromes genéticas permite a prevenção de suas sequelas e de comprometimentos do desenvolvimento da criança, além da orientação adequada quanto à reprodução e à transmissão. O impacto da prevenção em saúde pública e em custos com saúde é inegável.

Contudo, o intervalo entre a suspeita clínica e o diagnóstico de uma doença genética pode ser muito prolongado, até mesmo para geneticistas experientes, dada a variação de apresentações e o número continuamente crescente de síndromes genéticas existentes. Atualmente, para que o geneticista identifique uma síndrome, após o exame clínico, ele necessita buscar qual o diagnóstico mais provável, por meio de livros-texto, artigos ou, mais recentemente, sites especializados, nos quais, por uma ferramenta de busca, com um sinal clínico (como, por exemplo, baixa estatura), ele identifica todas as síndromes que o apresentam – ou seja, trata-se de algo altamente ineficiente<sup>(8-10)</sup>.

Atualmente, vêm se usando cada vez mais tecnologias para facilitar o trabalho médico e atingir amplas populações. No ramo da genética, isso é entendido como genética comunitária, que difere da clínica por ser antecipatória, pró-ativa e com grande abrangência na população<sup>(11,12)</sup>. Investir em tecnologia é o que se deve fazer para a genética, do ponto de vista da saúde pública atualmente, pois um diagnóstico precoce significa menor custo e resultados mais rápidos.

Um diagnóstico mais eficiente em genética, ainda na primeira consulta, pode permitir uma grande diminuição de custos para a saúde pública. O objetivo deste estudo foi criar um *software* que possa ser utilizado por médicos, em especial pediatras e geneticistas, para auxiliar no diagnóstico de doenças genéticas.

## MÉTODOS

A literatura a respeito do tema foi revisada para se listarem as 200 doenças mais prevalentes na população<sup>(8,10)</sup>. Em seguida, para cada doença, foram listados os sinais e sintomas mais frequentes, os essenciais para o diagnóstico, e aqueles que permitam o diagnóstico diferencial com outras doenças. A meta era de dez sinais e sintomas para cada doença; algumas, entretanto, apresentam poucos – às vezes seis ou quatro. O sistema de funcionamento da inteligência artificial do *software* pressupõe a forma como os próprios médicos diagnosticam seus pacientes<sup>(13-15)</sup>.

Dispondo dessas informações, foi elaborado um banco de dados no programa Microsoft Excel 2007<sup>®</sup>. Na primeira pasta de trabalho, foram colocados todos os dados necessários para uma avaliação clínica de genética, e todos os sinais e sintomas encontrados em cada uma das 202 síndromes a segunda pasta, foi deixado um espaço para elaboração do heredograma. A terceira pasta permanecia escondida durante o funcionamento do *software* e consistia no banco de dados, relacionando cada síndrome aos seus sinais e sintomas. Por fim, a quarta pasta de trabalho apresentou os resultados: quando se apertava “ctrl + D”, por meio de um macro, eram listadas as doenças em ordem decrescente de similaridade, conforme ilustrado nas figuras 1 e 2. Como observado nas figuras, as doenças que apresentem similaridade podem ficar agrupadas nas mesmas colocações (por exemplo,

23/01/2013 16:22				Ficha de Avaliação em Genética Clínica			
Todos os direitos reservados para David Gonçalves Nordon, MD							
1. Identificação							
Encaminhado por:							
Nome:		Joana da Silva		Endereço:			
Bairro:		CEP:		Cidade:			
Estado:		País:		Naturalidade:			
Procedência:		Profissão:					
Escolaridade:		Idade:		23/01/2010		3 anos	
Sexo:		Cor:		Branco		x	
				Negro			
				Pardo			
				Amarelo			
Cônjuge:							
Nome:		Profissão:		Cor:			
Idade:		Escolaridade:					
Pai							
Nome:		Profissão:		Cor:			
Idade:		Escolaridade:		43			
Mãe							
Nome:		Profissão:		Cor:			
Idade:		Escolaridade:		41			
2. Motivo do encaminhamento:							
2. Crânio e Face:							
NDN							
Crânio:		Microcefalia		Macrocefalia			
		Braquicefalia		Acrocefalia			
		Dolicocefalia		Escafocefalia			
		Cebocéfalia		Encefalocele			
		Hidrocefalia		Deflexão			
		Limitação de movimento					
Occipital:		Plano		Proeminente			
Assimetria craniana:		Especificar:					
Fontanelas amplas:		Craniossinostose:		Disjunção de sutura:			
Ossificação tardia de suturas cranianas:							
Fronte:		Alta		Estreita		Inclinada	
		Glabella proeminente		Crista metópica		Bossa	
		Defeito do osso frontal		Órbita rasa			
		Bochecha rechonchuda		Hipoplasia malar			
Face:		Achatada		Arredondada		Triangular	
		Estreita		Alongada		Pequena	
		Triste		Em máscara		Imatura	
		Grosseira		Microsomia hemifacial			
		Assobadora		Assimétrica			
		Mlopática		Rubor malar			
		Hipoplasia/malformação da região média facial					
Especificar assimetrias:							
Cabelos:		Implantação na fronte:		Alta		Baixa	
		Implantação na nuca:		Alta		Baixa	
		Claros		Finos		Quebradiços	
		Retorcidos e quebradiços		Áreas de aplasia			
		Até a lateral da face		Widow's peak			

Figura 1. Interface do *software* com dados iniciais (acima) e sinais e sintomas (abaixo).

é possível serem identificadas quatro síndromes com 50% de similaridade, duas com 40%, dez com 37,5% etc.).

Uma vez concluído o *software*, ele foi registrado no Instituto Nacional de Propriedade Industriais (INPI) e testado pelos próprios pesquisadores, no que se chamou de teste alfa, em casos avaliados no ambulatório de doenças genéticas de seu serviço, cujos diagnósticos estavam fechados e que constavam no banco de dados do *software*.

Este projeto foi aprovado pelo Comitê de Ética da Faculdade de Ciências Médicas e da Saúde de Sorocaba, em agosto de 2011.

## RESULTADOS

Durante o teste alfa, 58 casos foram submetidos ao *software*. Em 28 casos, a doença em questão ficou alocada no grupo de

23/01/2013 16:22		Ficha de Avaliação em Genética Clínica	
Todos os direitos reservados para David Gonçalves Nordon, MD			
1. Identificação			
Encaminhado por:			
Nome:	Joana da Silva	Endereço:	
Bairro:		CEP:	
Estado:		Pais:	
Procedência:		Naturalidade:	
Escolaridade:		Profissão:	
Data de nascimento:		Idade:	3 anos
Sexo:	<input type="checkbox"/> Masculino	Cor:	<input checked="" type="checkbox"/> Branco
	<input checked="" type="checkbox"/> Feminino		<input type="checkbox"/> Negro
			<input type="checkbox"/> Pardo
			<input type="checkbox"/> Amarelo
Associação VACTERL-H		30,00	
Síndrome de Klippel-Feil		22,22	
Associação MURCS		20,00	
Associação CHARGE		12,50	
Embriopatia pela Talidomida		12,50	
Pseudocondroplasia (Displasia Espôndilo Epifisária Pseudocondroplásica)		11,11	
Síndrome de Down		10,00	
Síndrome de Edwards		10,00	
Síndrome de Patau		10,00	
Síndrome de Warkany (trissomia do 8)		10,00	
Síndrome de Kabuki		10,00	
Síndrome de Williams		10,00	
Síndrome cárdio-facio-cutânea		10,00	
Síndrome de Beckwith-Wiedemann		10,00	
Síndrome cérebro-costal-mandibular		10,00	
Síndrome de Pallister-Hall		10,00	
Síndrome Velo-Cárdio-Facial (de Shprintzen)		10,00	
Síndrome de Larsen		10,00	
Síndrome de Catel-Manzke		10,00	
Sequência de Poland		10,00	
Síndrome de Holt-Oram		10,00	

**Figura 2.** Diagnósticos listados em ordem decrescente de similaridade com o paciente. As associações VACTERL-H, MURCS e CHARGE são geralmente utilizadas assim na literatura e na prática diária.

primeira colocação; em 9, no de segunda colocação; e, em 6, no de terceira colocação. Desta forma, o *software* “acertou” o diagnóstico em 74,13% dos casos. Cinco casos não foram solucionados (8,6%), e os restantes ficaram em colocações mais distantes.

## DISCUSSÃO

O objetivo do *software* não foi o diagnóstico de certeza, mas direcionar o diagnóstico. Por meio dele, torna-se mais fácil ao geneticista buscar as doenças possivelmente relacionadas ao paciente em questão. Trata-se de uma forma de inteligência artificial, mas que necessita da *expertise* do profissional que a utiliza, especialmente para saber diferenciar quando há mais de uma doença no grupo em primeira colocação.

Ele segue os princípios definidos pela inteligência artificial, ou seja: ele é um software que imita a forma como um médico pensa e pode desta forma realizar os cálculos em seu lugar<sup>(13-15)</sup>.

A maior dificuldade se deu com relação a síndromes com poucos dados. Algumas vezes, por apresentar poucos sinais, em pacientes que também apresentavam poucas alterações, essas síndromes, embora erradas, apareciam nas primeiras colocações.

Outra dificuldade se relacionou aos casos de mosaïcismo e penetrância incompleta. As síndromes e seus sinais clássicos não foram observados, ou foram muito leves, comprometendo a acurácia do *software*. O mesmo ocorreu com pacientes com poucas alterações, casos em que a *expertise* do profissional é essencial.

Os casos não solucionados muitas vezes incorreram nessa dificuldade.

Ademais, os dados utilizados para a elaboração do banco de dados incluíam livros relativamente antigos, que não necessariamente apresentam as alterações mais frequentes (embora o banco de dados tenha sido elaborado focando especificamente as alterações mais comuns). Esta dificuldade deve ser contornada pela fase beta de testes.

Com a implantação da fase beta, que será realizada por profissionais de outras instituições, espera-se ampliar e adequar o banco de dados e a interface de dados do paciente. Os autores esperam também o apoio de instituições de ensino e pesquisa para que possam melhorar o *software*, visando operar em um programa próprio e aperfeiçoar funcionamento de seu método de inteligência artificial.

Não há relato na literatura de um software que tenha sido desenvolvido para direcionar o diagnóstico médico sindrômico a partir de sinais identificados em pacientes. Uma iniciativa recente é do PhenoTips,<sup>(16)</sup> um banco de dados que promete associar os achados fenotípicos dos pacientes, sendo alimentado por médicos. No entanto, ele ainda não permite dar o diagnóstico a partir dos achados e funciona da mesma forma que outros sites com este intuito<sup>(10)</sup>.

## CONCLUSÃO

O *software* se mostrou bastante promissor em sua fase de teste alfa. Entretanto, para o seu aperfeiçoamento e uso cotidiano, ainda é necessário o teste beta, para o qual os autores gostariam de contar com o apoio interinstitucional.

## REFERÊNCIAS

- Christianson RE, Van den Bergh BJ, Milkovich L, Oechsli FW. Incidence of congenital anomalies among white and black live births with long-term follow-up. *Am J Public Health.* 1981; 71(12):1333-41.
- World Health Organization (WHO). Services for the prevention and management of genetic disorders and birth defects in developing countries: report of a joint WHO/WAOPBD meeting [Internet]. The Hague, 5-7 January 1999 [cited 2013 Jun 21]. Available from: [http://www.bvs.org.ar/genetica/pdf/Hague\\_report.pdf](http://www.bvs.org.ar/genetica/pdf/Hague_report.pdf)
- Pan American Health Organization. Consultation on the development of PAHO strategy on congenital disorders. Washington, DC: PAHO; 7-8 October, 1998.
- Instituto Brasileiro de Geografia e Estatística (IBGE). Censo demográfico 2000: características da população e dos domicílios: resultados do universo [Internet]. Brasília, DF: IBGE; 2000 [citado 2013 Mar 21]. Disponível em: <http://www.ibge.gov.br/home/estatistica/populacao/censo2000/>
- Barreiro CZ, Negrotti T, Penchaszadeh VB. Prevalence of genetic disease in a paediatric referral hospital. In: International Congress of Genetics, Mexico, 1976. *Excerpta Med Intl Congr Series.* 1976;397:60.
- Yoon PW, Olney RS, Khoury MJ, Sappenfiled Wm, Chavez GF, Taylor D. Contribution of birth defects and genetic diseases to pediatric hospitalizations. A population-based study. *Arch Pediatr Adolesc Med.* 1997;151(11):1096-103. Comment in: *Arch Pediatr Adolesc Med.* 1997;151(11):1082-3.
- Brasil. Ministério da Saúde. Sistemas de informações sobre mortalidade (SIM) e Nascidos Vivos (Sinasc) para profissionais do Programa

- Saúde da Família [Internet]. 2ª ed rev atual. Brasília, DF; 2004 [citado 2013 Mar 21]. Disponível em: [http://dtr2001.saude.gov.br/editora/produtos/livros/pdf/04\\_0049\\_M.pdf](http://dtr2001.saude.gov.br/editora/produtos/livros/pdf/04_0049_M.pdf)
8. Jones KL. Smith – Padrões reconhecidos de malformações congênitas. 5ª ed. São Paulo: Manole; 1998.
  9. Wiedermann HR, Kunze J, Dibbern H. Atlas de síndromes clínicas dismórficas. 2ª ed. São Paulo: Manole; 1992.
  10. OMIM. Online Mendelian Inheritance in Man [Internet]. McKusick-Nathans Institute of Genetic Medicine, The Johns Hopkins University School Medicine. [cited 2011 May 19]. Available from: <http://www.ncbi.nlm.nih.gov/omim>
  11. Brisson D. Analysis and Integration of definitions of community genetics. *Community Genet.* 2000;3:99-101.
  12. Ten Kate LP, Beemer FA, Broertjes JJ. Community genetics in the Netherlands. *Community Genet.* 2000;3:131-3.
  13. Fattu JM, Patrick EA. Computer-aided diagnosis and decision making. In: Javitt J, editor. *Computers in Medicine. Applications and possibilities.* Philadelphia, PA: W.B. Saunders; 1986. p. 201-33.
  14. Szolovits P. Artificial intelligence and medicine. In: Szolovits P, editor. *Artificial intelligence in Medicine.* Boulder, CO: Westview Press; 1982. p. 21-55.
  15. Sabattini RM. Uso do computador no apoio ao diagnóstico médico. *Rev Informéica.* 1993;1(1):5-11.
  16. Girdea M, Dumitriu S, Fiume M, Bowdin S, Boycott KM, Chénier S, et al. PhenoTips: patient phenotyping software for clinical and research use. *Hum Mutat.* 2013;34(8):1057-65.