

Avaliação de um Software de Diagnóstico Diferencial em Medicina Interna e Cardiologia

Pedro J. Negreiros de Andrade, Henrique V. Bezerra de Menezes, Eduardo L. A. Rocha, Carlos A. C. Teixeira, Ricardo A. Correia, José M. C. Lima, Francisco A. N. Colares, Pedro F. C. Bruin, Emir M. Limoverde, Marta M. C. Medeiros, Carlos C. Ribeiro, João M. Coelho F^o
Fortaleza, CE

a última década, muito se tentou fazer na utilização da informática como auxiliar ao diagnóstico médico^{1,2}. Apesar de esforços nesse sentido, a maioria dos programas apresentados caracteriza-se por abranger uma área limitada do conhecimento médico³⁻⁵, sendo, por conseqüência, de pouca utilidade para a grande maioria dos clínicos. Por outro lado, o grande desenvolvimento da medicina e a clara tendência à sub-especialização têm tornado, cada vez mais difícil ao médico, qualquer que seja a sua área, manter a habilidade de realizar diagnósticos diferenciais eficazes nos mais amplos domínios da medicina interna. Em conseqüência, a consulta a médicos especialistas torna-se freqüentemente necessária. Para muitos profissionais, no entanto, esta consulta é difícil, ou mesmo impossível. Isso torna ainda mais necessário o desenvolvimento de “softwares” aplicativos de auxílio ao diagnóstico diferencial em medicina, que sejam ao mesmo tempo abrangentes, eficazes e de fácil utilização. Algumas poucas tentativas nesse sentido têm sido feitas a nível internacional⁶⁻⁸, cuja aplicação, em nosso meio, parece difícil, devido às diferentes características epidemiológicas, às barreiras de língua e à ênfase, por vezes excessiva, em exames laboratoriais nos referidos programas.

Nosso objetivo é apresentar a avaliação inicial de um programa de auxílio ao diagnóstico diferencial, desenvolvido inicialmente na área de cardiologia⁹ e estendido a toda medicina interna. O programa é o resultado de 5 anos de trabalho de alunos e professores do Departamento de Medicina Clínica da Universidade Federal do Ceará.

Características Básicas do Sistema

O programa, por nós chamado de “Consultor”, foi desenvolvido para micros IBM PC XT ou AT, com pelo menos 540 de RAM e ocupando apenas 1,2 mega de espaço no disco rígido. A base de dados foi desenvolvida,

inicialmente, em linguagem DBase-Fox⁹, tendo sido, posteriormente, modificada para Clipper, o que a torna de utilização extremamente amigável, não exigindo nenhum conhecimento de informática da parte do usuário. A lógica do sistema é baseada no estabelecimento de *scores* para os achados clínicos, de acordo com a sua especificidade e sensibilidade e para as doenças, de acordo com sua prevalência, simulando assim algumas características bayesianas do raciocínio clínico^{10,11}. O sistema funciona através da manipulação de 4 bancos de dados principais: 1) um banco de dados chamado “doença”, contendo 471 patologias; 2) um banco de dados chamado “sintoma”, contendo 1440 itens relacionados a sexo, idade, sintomas, exame físico e dados laboratoriais; 3) um banco de dados chamado “relação”, contendo 6740 correlações entre as “doenças” e os “sintomas”. Na elaboração dessa base de dados (na qual está a maior parte da “lógica” do sistema) para cada correlação doença-sintoma é estabelecida uma pontuação positiva de 1 a 9, de acordo com a especificidade do achado. Pontos negativos foram atribuídos para achados que ajudassem a afastar a patologia. Dessa maneira achados de alta sensibilidade, se ausentes, implicarão em uma pontuação negativa que contribuirá para tornar o diagnóstico menos provável; 4) um banco de dados chamado “multi” que classifica a doença de acordo com a prevalência.

O processo pelo qual o sistema simula o raciocínio médico consiste em, após receber do usuário a lista dos achados clínicos do paciente (“sintomas”), percorrer o banco de dados “relação” somando os valores para cada doença. A seguir, o sistema seleciona as doenças com pontuação mais alta e as multiplica por um valor que leva em conta a prevalência da mesma. As pontuações foram estabelecidas de maneira que: a) pontuações finais superiores a 140 tornam o diagnóstico da doença “altamente provável”, a menos que haja doenças concorrentes com pontuação próxima ou superior; b) pontuações finais entre 90 e 140 fazem com que a doença seja considerada como simplesmente “provável”, sendo a sua exclusão obrigatória, a menos que hajam patologias concorrentes com pontuação muito superior; c) pontuações finais entre 60 e 89 caracterizam a doença como “possível” dentro do diagnóstico diferencial, embora não a tornem de exclusão obrigatória (há menos que inexistam patologias concorrentes com pontuação nitidamente superior); d) pontuações finais entre 30 e 59 pontos fazem com que a doença

seja citada apenas por razões acadêmicas. Mais dados precisariam ser fornecidos para torná-la “provável”; e) pontuações entre 0 e 29 tornam a doença improvável à luz dos dados fornecidos, levando o sistema a excluí-la da lista. O usuário é também informado que, para um diagnóstico de certeza, deve a doença figurar, não apenas com uma elevada pontuação, mas, apresentar uma diferença de pontos superior a 50 em relação às patologias concorrentes. Após listar os diagnósticos possíveis, o sistema oferece ainda uma “análise complementar”, através da qual o usuário é informado dos achados dos casos que encontram-se presentes em cada uma das doenças citadas, assim como dos achados de cada doença que ajudaria a confirmar o diagnóstico e que não foram citados nas informações sobre o paciente.

O programa oferece ainda os seguintes módulos: 1) módulo pesquisa-doença: dada uma determinada doença, o sistema lista os achados clínicos com ela correlacionados. O CID também é fornecido; 2) módulo pesquisa-sintoma: dado um determinado achado clínico, o sistema fornece a lista das doenças com ela correlacionadas; 3) módulo simulação: o sistema escolhe uma determinada doença e lista os seus achados (sinais e sintomas apenas). A seguir, roda o programa de diagnóstico, listando a doença e seus diagnósticos concorrentes; 4) módulo terapêutica: dada uma determinada doença, o sistema lista brevemente a conduta terapêutica na mesma. O usuário pode, além disso, modificar o texto da terapêutica, acrescentando ou retirando o que quiser, de acordo com sua conveniência.

Avaliação do Sistema

Um *software* de diagnóstico diferencial em medicina interna precisa ser intensamente avaliado na sua capacidade de auxiliar, tanto na resolução de casos simples quanto complexos.

Numa tentativa de avaliar a capacidade diagnóstica do sistema, solicitou-se inicialmente aos professores do Depto de Medicina Clínica da UFCE que elaborassem 25 casos sumários, de baixo nível de complexidade, para a prova de residência em Clínica Médica. Tais casos foram submetidos a 20 médicos recém-formados, já aprovados na prova classificatória para residência de Clínica Médica. Solicitou-se aos mesmos que listassem o diagnóstico correto, assim como os possíveis diagnósticos diferenciais. A seguir, os mesmos casos foram submetidos à análise do programa. Considerou-se que o sistema acertava o diagnóstico quando a patologia com maior número de pontos correspondia ao diagnóstico considerado como correto. Considerou-se que o sistema lembrava corretamente o diagnóstico quando, entre as 6 patologias com maior pontuação, estava o diagnóstico considerado como correto. O desempenho do programa e sua comparação com a média dos resultados obtidos pelos médicos recémformados

estão sintetizados na tabela I. Nela vemos que o sistema foi superior, embora não significativamente, aos médicos recém-formados na capacidade de fazer o diagnóstico (18 acertos contra 14 $p>0,05$) sendo significativamente superior aos mesmos na capacidade de lembrar o diagnóstico (25 acertos contra 18, $p<0,01$).

Tabela I - Comparação entre o desempenho do programa e o de médicos recém-formados no diagnóstico de 25 casos simples.

	Acertos	Citações
Programas	18	25
Recém-formados	14	18
Significância	$P>0,05$	$P<0,01$
Obs: Significância calculada através do teste do “qui quadrado”		

Tabela II - Comparação entre o desempenho do programa e o de médicos residentes de clínica médica no diagnóstico de 25 casos complexos.

	Acertos	Citações
Programa	10	22
Residentes	6	16
Significância	$P>0,05$	$P<0,05$
Obs: Significância calculada através do teste do “qui quadrado”		

Em uma segunda etapa, testou-se a capacidade do sistema de resolver casos relativamente complexos, extraídos de sessões Clínico-Patológicas do Massachusetts General Hospital, da Washington University e da John Hopkins School of Medicine. A fonte foi o livro “Differential Diagnosis”, de Harvey & Bordley¹². A escolha deveu-se ao fato do livro conter casos de raridade não excessiva e de certa forma representativos das patologias encontradas em um hospital terciário brasileiro. Da referida publicação escolheram-se, aleatoriamente, 25 casos que foram entregues a residentes de Clínica Médica do HUWC da UFCE. Residentes que conheçam o livro foram excluídos da investigação. Cada residente recebeu em torno de 2 casos, dispondo de cerca de 2h para levantar o seu diagnóstico e os possíveis diagnósticos diferenciais. Posteriormente, os mesmos casos foram entregues à análise do programa dentro dos mesmos critérios anteriormente estabelecidos. Os resultados estão sintetizados na tabela II que compara os resultados do programa com o desempenho dos residentes. Nela vemos que o sistema teve um desempenho superior, embora não significativamente, ao dos pós-graduandos, no que toca à capacidade de “acertar” o diagnóstico (10 acertos contra 6, $p>0,05$). No que concerne à capacidade de lembrar o diagnóstico, o desempenho do sistema foi significativamente superior ao dos médicos residentes (22 acertos contra 16, $p<0,05$). Numa terceira etapa, comparou-se o desempenho do programa com o dos responsáveis pela discussão dos casos nas referidas sessões anátomo-clínicas (tab. III). Neste caso o desempenho do sistema foi

significativamente inferior ao dos *Case Discussants* no que toca à capacidade de “fazer” o diagnóstico (10 acertos contra 18, $p < 0,01$). Seu desempenho foi, no entanto, comparável ao dos *Case Discussants* no que concerne à capacidade de lembrar o diagnóstico (22 acertos contra 23, $p > 0,05$). É interessante observar que o programa foi capaz de acertar o diagnóstico em 68% dos casos cardiológicos complexos e em 88% dos casos simples.

Tabela III - Comparação entre o desempenho do programa e o dos <i>case discussants</i> das sessões anátomo-clínicas norte-americanas, no diagnóstico de 25 casos complexos		
	Acertos	Citações
Programa	10	22
“Case Discussants”	18	23
Significância	$P < 0,01$	$P > 0,05$
Obs: Significância calculada através do teste do “qui quadrado”		

Comentários

Qualquer programa de auxílio ao diagnóstico diferencial em medicina deve ter algumas características básicas¹³. A mais importante é ter um banco de dados suficientemente extenso para cobrir o grande número de patologias na área¹⁴, além de apresentar um conteúdo lógico, que seja referendado por especialistas de reconhecida competência nas várias sub-especialidades médicas. Deve, além disso, ser de fácil uso e não levar a desperdício de tempo na entrada ou na saída dos dados.

Essas características iniciais parecem ser preenchidas pelo nosso sistema, cuja base de dados foi laboriosamente trabalhada, tendo sido revista por professores de reconhecida competência em cada uma das subespecialidades da medicina interna. Além disso, o usuário experiente levará pouco mais de um minuto na entrada dos dados, levando o sistema tempo equivalente para fornecer as principais hipóteses diagnósticas com a respectiva pontuação. Uma outra característica desejável é que o sistema seja “inteligente” e tenha uma interação com o usuário. Sob esse aspecto, embora o sistema não tenha rigorosamente “inteligência artificial”, porque não se modifica com o uso, ele é de, certa forma, “inteligente”, na medida em que simula as características bayesianas do raciocínio clínico^{10,11}, levando em conta na feitura do diagnóstico, a prevalência da doença e a sensibilidade e especificidade dos achados (ainda que o modelo utilizado na análise dos casos não seja rigorosamente bayesiano, da mesma forma que o raciocínio clínico habitual não o é). Além disso o módulo “análise complementar” permite também uma interação com o usuário, de grande utilidade por simular a abordagem de diagnóstico diferencial comum a muitos quadros clínicos.

Por fim, o sistema deve ter a sua eficiência testada

tanto na resolução de casos simples quanto de casos complexos. Os presentes dados sugerem que o programa teve um desempenho superior ao dos candidatos à residência médica na resolução de casos simples e ao dos próprios residentes de clínica médica no diagnóstico de casos complexos. Mas, o estresse de um ambiente de prova e o pouco tempo disponível para as respostas (os candidatos à residência dispunham de apenas alguns minutos para a solução dos casos simples e uma hora parece um tempo relativamente curto para os pós-graduandos concluírem a análise de um caso de sessão anátomo-clínica) podem ter favorecido o sistema na análise comparativa. Além disso, o fato do desempenho dos *case discussants* ter sido superior ao do sistema na análise de casos complexos, parece indicar que o programa está longe ainda de superar o desempenho de especialistas renomados que disponham de longo tempo para a análise do caso (como é habitual nas sessões anátomo-clínicas). É interessante lembrar que o sistema teve um desempenho comparável ao dos *case discussants*, no que toca à capacidade de “lembrar” o diagnóstico o que se torna particularmente relevante se lembrarmos a rapidez com que o faz e o fato de tais especialistas, geralmente, não estarem disponíveis para a grande maioria dos médicos. Aliás, a rapidez na análise e a facilidade na entrada dos dados é uma característica do sistema que o diferenciam de programas semelhantes e mais próximos da chamada inteligência artificial^{6,7,15,16}. Os quais têm no laborioso processo de entrada dos dados uma das suas maiores limitações⁶⁻⁸.

Apesar disso, o sistema, na sua versão atual, apresenta algumas limitações que necessitam ser ressaltadas. A primeira delas é a de que, analogamente a programas semelhantes^{6,7} não “raciocina” temporalmente, nem fisiopatologicamente, como é hábito saudável de muitos clínicos. Além disso, o sistema não “faz” diagnósticos, limitando-se a listar as patologias possíveis por ordem decrescente de probabilidade (embora uma pontuação acima de 170 pontos torne o diagnóstico quase indiscutível). O programa tem, como última limitação, a incapacidade de dizer se existe apenas um ou mais de um diagnóstico para o caso, assim como de estabelecer relações de causa e efeito entre patologias aparentemente concorrentes, como, por exemplo, miocardiopatia dilatada e insuficiência cardíaca. Essas conclusões, assim como a definição do diagnóstico, são deixadas com o usuário, atuando o sistema como mero auxiliar.

Em face das vantagens e limitações do sistema, aqui expostas, podemos concluir que “Consultor”, embora ainda um programa experimental com base de dados incompleta, já se mostra capaz de “analisar” com alguma proficiência casos clínicos de natureza teórica a ele apresentados. Sua maior utilidade parece ser, no momento, a de lembrar ao médico um grande número de patologias que freqüentemente não vêm à mente dos envolvidos no cuidado de pacientes complexos. Parece-nos ter além disso,

o mérito de tratar os achados clínicos de

maneira não tendenciosa, o que pode ajudar a abalar convicções prematuras em casos ainda não suficientemente definidos. Apesar do seu grande potencial, como auxiliar ao ensino de clínica médica, seu uso nessa área, assim como sua utilidade no auxílio ao diagnóstico de pacientes reais, merece futuras investigações.

Referências

1. Barnett GO - The computer and clinical judgement. (Editorial) N Eng J Med. 1982; 307: 493-4
2. Leão BF, Luchese MA, Rocha AF - Apoio à decisão em Medicina. Arq Bras Cardiol, 1989; 52: 239-41.
3. Moura MRS, Luchese MA, - Software especialista aplicativo com rede lógica de decisões para laboratórios de hemodinâmica e angiocardiografia. Arq Bras Cardiol, 1991; 57: 275-80.
4. Goldman L, Weinberg M - A computer derived protocol to aid in the diagnosis of emergency room patients with chest pain. N Eng J Med. 1982; 307: 588-62.
5. Leaper DJ, Horrock JC, Staniland IR, De Dombal FT - Computer assisted diagnosis of abdominal pain using stimates provided by clinicians. Br Med J, 1972; 4: 350.
6. Miller RA, Pople Jr HE, Myers JD - Internist I, an experimental computer based diagnostic consultant in general internal medicine. N Eng J Med; 1989; 110: 824-32.
7. Bankowitz RA, McNeil MA, Challinor SM, Parker RC, Kapoor WN, Miller RA - A computer-assisted medical diagnostic consultation service. Ann Int Med, 1989; 110: 824-32
8. Hayden SP - Computer assisted diagnosis. Ann Int Med. 1989; 111: 544-5.
9. Menezes HVAB, Andrade PJN et al - Projeto Hipócrates-sistema de auxílio ao diagnóstico diferencial em cardiologia e medicina interna. Sessões de Temas Livres VIII Congresso Norte-Nordeste de Cardiologia, 1988.
10. Diamond GA, Forrester IS - Analysis of probability as an aid in the clinical diagnosis of coronary artery disease. N Eng J Med. 1979; 300: 1350-8.
11. Rembold CM, Watson D - Posttest probability calculation by weights. A simple form of Bayes theorem. Ann Intern Med. 1988; 108: 115-20.
12. Harvey AM, Bordley J - Diferential Diagnosis, 2nd ed. Philadelphia, WB Saunders Co, 1972.
13. Szolovits P, Ramesh SP, Schwartz WB - Artificial inteligency in medical diagnosis. Ann Intern Med. 1988; 108: 80-7.
14. Miller RA, McNeil MA, Challinor SM, Masarie FE, Myers JD - The internist-1/quick medical reference project-status report. West J Med. 1986; 145: 816-22.
15. Masarie FE, Miller RA, Myers JD - Internist I properties: representing common sense and good medical knowledge in a computerized medical knowledge base. Comp Biomed Res, 1985; 18: 458-79.
16. Barnett GO, Cimino JJ, Hupp JA, Hoffer EP. - DXplain: An evolving diagnostic-decision system. JAMA, 1987; 258: 64-7.