

**Revista Saúde.Com**

ISSN 1809-0761

<https://periodicos2.uesb.br/index.php/rsc>**CIÊNCIA DE DADOS EM SAÚDE: CONTRIBUIÇÕES E TENDÊNCIAS PARA APLICAÇÕES****HEALTH DATA SCIENCE: CONTRIBUTIONS AND TRENDS FOR APPLICATIONS****Antonio Valerio Netto**

Escola Paulista de Medicina/UNIFESP

**Abstract**

*Objective: To explain the concepts involved with the subject of health data science presenting their fields of action and possible applications focused on user health. Methods: Demonstrate functional and technical differences compared to other areas of knowledge based on the collection, organization, analysis and visualization of medical, hospital management and population health data. Results: Build a knowledge framework to improve the definitions involved with the theme and present through applications their real contribution to health. Conclusion: It is considered that the work of understanding the data promises to transform our current understanding of health and aims to provide new insights, as well as allowing us to innovate in the provision of patient care services.*

**Keywords:** Data science; Medical informatics; Statistics e numerical data.

**Resumo**

*Objetivo: Explicar os conceitos envolvidos com a temática de ciência de dados em saúde apresentando os seus campos de atuação e possíveis aplicações focadas na saúde do usuário. Métodos: Demonstrar as diferenças funcionais e técnicas em comparação com as outras áreas do conhecimento que tem como base a coleta, organização, análise e visualização de dados médicos, de gestão hospitalar e de saúde populacional. Resultados: Construir um arcabouço do conhecimento para aprimorar as definições envolvidas com a temática e apresentar por meio das aplicações a sua real contribuição para área de saúde. Conclusão: Considera-se que o trabalho de entendimento dos dados promete transformar a nossa atual compreensão da saúde e visa fornecer novos insights, além de permitir inovar na prestação de serviços relacionados ao cuidado dos pacientes.*

**Palavras-chave:** Ciência de dados, Informática médica, Estatística e dados numéricos.

## Introdução

Um único paciente crônico gera em torno de 80 *Mbytes* por ano em dados de imagem, *Electronic Medical Record* (EMR) entre outras informações obtidas com exames médicos e testes clínicos. Esses dados tem um importante valor clínico e operacional para assistência médica. As oportunidades de trabalho com esses dados foram estimadas em torno de US\$ 300 bilhões por ano, somente com redução de custos.<sup>(1)</sup> Em 2011, conforme a *National Hospital Ambulatory Medical Care Survey (U.S. Depart. of Health & Human Services)*, ocorreram 125,7 milhões de visitas ambulatoriais sem internação e 136,3 milhões de visitas aos centros de emergências nos EUA.<sup>(2)</sup> Gerando, inclusive, um volume de informações que podem ser aplicadas em estratégias para apoiar ações como as relacionadas com cuidado híbrido.<sup>(3)</sup>

O setor de saúde brasileiro utiliza pouco esses dados. As organizações precisam capturar o valor de suas informações investindo na análise de dados e estabelecendo essa habilidade como uma competência essencial. Neste contexto, é possível fazer uso de soluções baseadas em ciência de dados em saúde (do inglês *Health Data Science*). Ela tem por objetivo gerar soluções por meio da compreensão de problemas reais da área de saúde, empregando o pensamento crítico e a análise para obter conhecimento.

Trata-se de um conhecimento surgido da interseção da bioestatística, ciência da computação e saúde. Ela é dividida em “dados clínicos” e “dados de comportamento e sentimento do paciente”.<sup>(4)</sup> Os “dados clínicos” vêm de anotações médicas, resultados de exames, imagens médicas, dados de medidores fisiológicos entre outros. Esse tipo de dado é coletado durante as rotinas de trabalho do profissional de saúde para diagnosticar e acompanhar o estado de saúde dos pacientes. Os “dados de comportamento e sentimento do paciente” ajudam a maximizar a prevenção. Os dispositivos tecnológicos podem ajudar a captar, além de informações fisiológicas, como: frequência cardíaca, padrões de respiração, dentre outros; o nível de movimentação física, a qualidade de sono entre outras informações de comportamento. Além disso, existem dispositivos para captura de informações oculares (*eye tracking*), de resistência galvânica da pele (MRGP), entre outros. Essas informações auxiliam na detecção de padrões irregulares e pode prever distúrbios baseados no comportamento dessas informações.

Entende-se que o foco do *Health Data Science* esta na interpretação e entendimento específico de dados do paciente, principalmente, oriundos de sua jornada de tratamento de saúde. *Health Data Science* pode também ser um instrumento válido na pesquisa clínica para compreender e permitir a predição de comportamentos em seres vivos.

É importante não confundir com dados de gestão hospitalar que faz uso de plataformas de BI (*Business Intelligence*)/BA (*Business Analytics*). Neste caso, é aplicado o *Health Analytics* que é um conjunto de metodologias utilizadas para analisar grandes volumes de dados relacionados com o setor de saúde, contribuindo no cuidado clínico geral e na gestão dos recursos de saúde. Engloba tecnologias e habilidades usadas para fornecer *insights* clínicos, programáticos e de negócios sobre as interdependências complexas que geram resultados e custos médicos.<sup>(5)</sup>

Para os gestores de saúde, a contribuição está em reduzir desperdícios e melhorar a produtividade do cuidado com a saúde dos beneficiários, sempre de forma sustentável. *Health Analytics* envolve diversas práticas, incluindo manejo dos dados não padronizados de saúde, clusterização das variáveis a serem estudadas e *machine learning* como mecanismo de predição analítica para tomada de decisão estratégica e precoce.<sup>(6)</sup> Por meio de modelagem, otimização, análise preditiva, as organizações de saúde podem obter informações estratégicas para fortalecer o desempenho financeiro e orçamentário, aprofundar os relacionamentos centrados no consumidor e melhorar a forma como a assistência à saúde é concebida e entregue para gerar melhores resultados.<sup>(7)</sup>

## Metodologia

*Health Data Science* e *Health Analytics* utilizam *Data Analytics*. Trata-se de um conhecimento cujo objetivo é examinar dados brutos para encontrar padrões e tirar conclusões sobre essa informação, aplicando um processo algorítmico ou mecânico para obter informações<sup>(8)</sup>, isto é, refere-se à prática de coletar dados agregados e analisá-los para obter *insights* e informações importantes contidos nele. Esses sistemas transformam, organizam e modelam os dados para tirar conclusões e identificar padrões. *Data Analytics* é um campo amplo e existem quatro tipos de análise de dados: *Descriptive analytics*, *Diagnostic analytics*, *Predictive analytics* e *Prescriptive analytics*.<sup>(9)</sup>

A análise descritiva (*Descriptive analytics*) ajuda a responder as perguntas sobre o que

aconteceu. Ao desenvolver indicadores-chave de desempenho (KPIs), a análise dessa informação pode ajudar a rastrear sucessos ou fracassos. Métricas especializadas são desenvolvidas para acompanhar o desempenho em setores específicos. Este processo requer uma coleta de dados relevantes, processamento, análise e visualização desses dados. Esse processo fornece uma visão essencial do desempenho passado.

A análise de diagnóstico (*Diagnostic analytics*) ajuda a responder perguntas sobre por que as coisas aconteceram. Este tipo de verificação complementa as análises descritivas básicas, pois se aprofundam para encontrar a causa. Os indicadores de desempenho são investigados para descobrir por que eles melhoraram ou pioraram. Isso geralmente ocorre em três etapas: identificação de anomalias nos dados (mudanças inesperadas em uma métrica específica); coleta desses dados relacionados a essas anomalias; e por fim aplicação de técnicas estatísticas para encontrar relacionamentos e tendências que expliquem essas anomalias.

A análise preditiva (*Predictive analytics*) ajuda a responder as perguntas sobre o que acontecerá no futuro. Essas técnicas usam dados históricos para identificar tendências e determinar se podem recorrer. Ferramentas analíticas preditivas fornecem informações valiosas sobre o que pode acontecer no futuro e suas técnicas incluem uma variedade de técnicas estatísticas e de aprendizado de máquina (*machine learning*), tais como: redes neurais, árvores de decisão e regressão.

A análise prescritiva (*Prescriptive analytics*) ajuda a responder perguntas sobre o que deve ser realizado mediante o entendimento do que irá acontecer. Usando *insights* da análise preditiva, as decisões baseadas em dados podem ser realizadas. Isso permite que sejam tomadas decisões em face da incerteza. As técnicas de análise prescritiva dependem de estratégias de aprendizado de máquina que podem encontrar padrões em grandes conjuntos de dados. Ao analisar decisões e eventos passados, a probabilidade de resultados diferentes pode ser estimada. Esse tipo de análise fornece a percepção da necessidade de tomar decisões eficazes e eficientes.

Para ambos, *Health Data Science* e *Health Analytics*, as aplicações podem envolver os conceitos de *Big Data* para construir um processo de análise de informações de saúde e tomada de decisão baseado em Volume, Velocidade e Variedade (três “Vs”).<sup>(10)</sup> Além disso, existem mais três características relacionadas com Variabilidade (consistência dos dados ao longo

do tempo), Veracidade (confiabilidade dos dados obtidos) e Valor (dados clinicamente relevantes). No caso da Veracidade, a mesma está relacionada aos registros de saúde que podem conter erros tipográficos, abreviações e notas enigmáticas. Às vezes, as medidas ambulatoriais são captadas em ambientes menos confiáveis e não controlados, em comparação com dados clínicos, coletados por profissionais treinados ou por equipamentos de medição de forma automática.<sup>(11)(12)</sup>

*Big data* na área de saúde se refere a conjuntos de dados eletrônicos de saúde tão grandes e complexos que são difíceis de gerenciar com *software* e/ou *hardware* tradicionais; nem podem ser facilmente manipulados com ferramentas e métodos tradicionais ou comuns de gerenciamento de dados.<sup>(13)</sup> Entre os conteúdos do chamado *Big Data Analytics* (BDA) estão dados oriundos de processamento de imagens, de sinais e genômica<sup>(14)</sup>. No caso do processamento de sinais, a biotelemetria e os dispositivos fisiológicos de monitoramento de sinais tem sido atualmente mais utilizado, contudo esses dados gerados pelos sistemas não costumam ser armazenados por longos períodos de tempo, negligenciando assim uma extensa investigação sobre esses dados. Já existem exemplos da utilização do monitoramento fisiológico contínuo para melhorar o atendimento e o gerenciamento de pacientes. Como exemplos de fontes de dados para aplicações em BDA, são possíveis citar: administrativo, biomarcadores, biométrico, registro clínico, registro eletrônico de saúde (do inglês, *electronic health record* - EHR), Internet, imagens médicas e relatórios de paciente.

### Aplicações envolvendo a área

As oportunidades para aplicação de *Health Data Science* permite um largo espectro para soluções envolvendo *medical and health informatics*, *translational bioinformatics*, *sensor informatics* e *imaging informatics*.<sup>(11)</sup> No caso de *medical and health informatics*, a proposta é lidar com dados estruturados e não estruturados de diferentes fontes, com o objetivo de gerar estudos longitudinais em larga escala, bem como capturar tendências e propor modelos preditivos para dados gerados a partir de *Electronic Health Records* (EHRs). Uma oportunidade reside na integração da informática médica tradicional com *mhealth* e *social health*, abordando doenças agudas e crônicas de uma maneira até então não realizada.

Outro campo de aplicação é a *translational bioinformatics* que surgiu após o primeiro mapeamento do genoma humano. Seu foco está em conectar a biologia molecular, a bioestatística e a genética estatística com a informática clínica. Este conhecimento está evoluindo e outros campos foram propostos. Entre eles, a farmacogenômica que um ramo da genômica preocupada com as variações dos indivíduos na resposta aos medicamentos devido às diferenças genéticas. Esse campo é importante para projetar medicamentos. Novas descobertas, resultantes do projeto Genoma Humano, são frequentemente aplicadas para desenvolver diagnósticos, prognósticos e terapias para doenças complexas, conhecidas como *translational genomics*.

O *sensor informatics* é dividido em três campos e apoiam, em grande parte, aplicações relacionadas à biotelemetria. O primeiro envolve sensores vestíveis (*wearable*), implantáveis (*implantable*) e ambientais. Três fatores contribuíram para a popularização desse tipo de tecnologia: maior poder de processamento de dados, comunicações sem fio (*wireless*) mais rápidas com maior largura de banda e *design* aprimorado de dispositivos microeletrônicos. O segundo provém de sensores de dados para gerenciamento estratificado de pacientes. A detecção fisiológica por esses dispositivos inteligentes pode ser de longo prazo e contínua, impondo novos desafios para a interpretação de sua relevância clínica. E por fim, dados oriundos de tecnologia baseada em *MHealth* que fazem uso de *smartphones*, *tablets*, entre outros dispositivos para captura de dados que possuem como característica a necessidade da mobilidade do paciente ou do profissional de saúde.

Com relação à *imaging informatics* enquanto as modalidades de imagem já estabelecidas estão se disseminando como é o caso do MRI (*magnetic resonance imaging*), novas modalidades de imagem em todo o espectro de frequência estão surgindo. Como são os casos do: *Optical imaging*, X-Ray, PET (*Positron-emission tomography*), SPECT (*Single photon emission computed tomography*). Muitas dessas técnicas estão voltadas para aplicações em tempo real *in situ* ou *in vivo*. Os novos equipamentos têm permitido que uma ressonância magnética cardiovascular, por exemplo, possa capturar detalhes sobre as paredes dos vasos, perfusão e difusão do miocárdio, além de padrões de fluxo complexos *in vivo*. Isto tem servido como um auxílio terapêutico e intervencionista, em vez de apenas uma modalidade de diagnóstico. Além disso,

esforços extensos na combinação de diferentes modalidades de imagens abrem uma série de novas oportunidades, particularmente para imagens oncológicas e terapias direcionadas.

Existem diversos exemplos de aplicações que já chegaram ao mercado consumidor final. É o caso da empresa *Omada Health* que desenvolveu uma tecnologia utilizando *Health Data Science* para redução de riscos. Eles mapearam que os EUA têm uma taxa de obesidade de 30% em mais de 25 estados, que estão causando doenças crônicas como diabetes e hipertensão. Propuseram uma solução que utiliza balanças e pedômetros para processar os dados comportamentais dos pacientes e desenvolver um programa personalizado com base nesses resultados. Atualmente a empresa descreve este processo como uma terapia digital (*digital therapeutics*) para doenças crônicas. Outra aplicação foi desenvolvida pela empresa Enlitic que utiliza *Health Data Science* para aumentar a precisão e a eficiência do diagnóstico. De acordo com a *National Academics of sciences and engineering*, cerca de 12 milhões de pacientes adultos são mal diagnosticados a cada ano nos EUA. A empresa usa algoritmos de aprendizagem de máquina para produzir dados precisos a partir de tomografias computadorizadas, raios-X, dentre outros.

### Considerações finais

Em 2018 foi publicado o primeiro plano estratégico para *Data Science* onde foi fornecido um *roadmap* para modernização do *NIH-funded biomedical data science ecosystem*.<sup>(15)</sup> É importante constatar que a ciência de dados, cada vez mais, tem sido discutida nas organizações de saúde com o objetivo de promover uma visão estratégica no médio e longo prazo.

Soluções envolvendo ciência de dados promete transformar a nossa compreensão dos processos de saúde e fornecer *insights* sobre o desenvolvimento de novas aplicações. Além disso, pode promover significativas mudanças na prestação dos serviços de cuidado. Para isto, são necessárias habilidades para aquisição, organização e exploração dos dados; como também domínio da modelagem estatística, algoritmos de predição, além de técnicas de visualização para uma adequada tomada de decisão. Essas soluções irão permitir a inovação tanto na indústria de medicamentos, como também, em operadoras de saúde. É capaz de apoiar, desde o descobrimento de novos *insights*

até promover, de forma segura, a transição do cuidado.

## Referências

1. Huesch MD, Mosher TJ. Using It or Losing It? The Case for Data Scientists Inside Health Care. *NEJM Catalyst* [Internet]. [cited 2019 August 19]. Available from: <https://catalyst.nejm.org/case-data-scientists-inside-health-care/>
2. Mushtaq Z. When Data Science met Medicine! Team Research Nest [Internet]. [cited 2019 August 19]. Available from: <https://medium.com/the-research-nest/when-data-science-met-medicine-8d3971a0ade9>
3. Valerio Netto A, Tateyama AGP. Avaliação de tecnologia de telemonitoramento e biotelemetria para o cuidado híbrido para o idoso com condição crônica. *Journal of Health Informatics*, 2018; 10(4).
4. Centre for Big Data research in Health. What is Health Data Science? UNSW Sydney [Internet]. [cited 2019 August 07]. Available from: <https://cbrh.med.unsw.edu.au/what-health-data-science>
5. SAS. Health Analytics – What it is and why it matters. [Internet]. [cited 2019 August 20]. Available from: [https://www.sas.com/en\\_us/insights/health-analytics.html](https://www.sas.com/en_us/insights/health-analytics.html)
6. Raghupathi W, Raghupathi V. An overview of health analytics. *J Health Med Informat*, 2013; 4(132), 2.
7. Kohn MS et al. IBM's health analytics and clinical decision support. *Yearbook of medical informatics*, 2014; 23(01), 154-162.
8. What is Data Analytics? Masters in Data Science [Internet]. [cited 2019 August 19]. Available from: <https://www.mastersindatascience.org/resources/what-is-data-analytics/>
9. Banerjee A, Bandyopadhyay T, Acharya, P. Data analytics: Hyped up aspirations or true potential? *Vikalpa*, 38(4), 2013; 1-12.
10. Laney D. 3D data management: Controlling data volume, velocity and variety. Gartner. [Internet]. [cited 2019 August 25]. Available from: <https://blogs.gartner.com/doug-laney/files/2012/01/ad949-3D-Data-Management-Controlling-Data-Volume-Velocity-and-Variety.pdf>
11. Andreu-Perez J, Poon CC, Merrifield RD, Wong, ST, Yang GZ. Big data for health. *IEEE journal of biomedical and health informatics*, 2015; 19(4), 1193-1208.
12. Qin SJ. Process data analytics in the era of big data. *AIChE Journal*, 2014; 60(9), 3092-3100.
13. Raghupathi W, Raghupathi V. Big data analytics in healthcare: promise and potential. *Health information science and systems*, 2014; 2(1), 3.
14. Belle A, Thiagarajan R, Soroushmehr SM, Navidi F, Beard DA, Najarian K. Big data analytics in healthcare. *BioMed research international*, Hindawi Publishing Corporation 2015.
15. NIH strategic plan for data science. [Internet]. [cited 2019 August 21]. Available from: [https://datascience.nih.gov/sites/default/files/NIH Strategic Plan for Data Science Final 5 08.pdf](https://datascience.nih.gov/sites/default/files/NIH%20Strategic%20Plan%20for%20Data%20Science%20Final%2008.pdf).

Endereço para Correspondência

Antonio Valerio Netto

Rua Alcatrazes, 189 – apto. 33. São Paulo - SP

CEP 04144-100

E-mail: [avalerionetto@gmail.com](mailto:avalerionetto@gmail.com)

---

Recebido em 27/02/2020

Aprovado em 12/08/2021

Publicado em 27/09/2021