

I SEMANA ACADÊMICA INTEGRADA
Cocriar, Viver, Inovar, Despertar (COVID)
Temáticas para Superar a Pandemia

BIG DATA NA SAÚDE

Barreto ME^{1,2,3,4}

1. Professor Associado, Departamento de Ciência da Computação, Instituto de Matemática e Estatística, Universidade Federal da Bahia - UFBA, Salvador, BA, Brasil
2. Pesquisador Associado, Centro de Integração de Dados e de Conhecimento para Saúde (CIDACS), Fundação Oswaldo Cruz, Salvador, BA, Brasil
3. Associated Researcher, Institute of Health Informatics, University College London, London, UK
4. Teaching Professor, Department of Statistics, London School of Economics and Political Science, London, UK

E-mail: marcosb@ufba.br

O termo *big data* tornou-se popular na década de 90 para denotar grandes quantidades de dados produzidos diariamente, advindos de diferentes fontes e em diferentes formatos [1]. Neste contexto, adotou-se uma nomenclatura baseada em “Vs” para representar o **volume** (quantidade) de dados, a **velocidade** na qual os dados são gerados, a **variedade** de formatos (dados estruturados e semiestruturados, imagens, textos), a **variabilidade** em termos de qualidade e disponibilidade, a **veracidade** das fontes de dados e o **valor** associado a estes dados. Ao longo do tempo, outros Vs vem sendo acrescentados ao contexto de *big data*, resultando em definições contendo até 42 termos [2], o que demonstra a complexidade associada a esta área de pesquisa e desenvolvimento. Na área da Saúde, *big data* é associado à quantidade de registros eletrônicos de saúde, exames clínicos e laboratoriais, imagens e diversos outros tipos de dados capturados através de sensores e aplicativos para saúde (*mobile health*), bem como dados genômicos, que podem ser integrados para possibilitar análises complexas e tomada de decisões envolvendo dados quantitativos e qualitativos [3]. A integração de grandes volumes de dados ditos “multimodais” exige o emprego de infraestruturas e ferramentas computacionais capazes de suportar os requisitos associados aos diferentes Vs do *big data*. Métodos estatísticos e computacionais, baseados em conceitos e modelos de Inteligência Artificial [4], são constantemente empregados em aplicações de **saúde pública** (análises baseadas em coortes populacionais ou estudos envolvendo exposição e desfecho para avaliação de políticas públicas), **processamento de imagens médicas** (suporte ao diagnóstico baseado em imagens), **medicina de precisão** (combinação e análise de dados multimodais para o tratamento individualizado de pacientes), estudo prospectivos de **saúde única** (*one health*), na qual abordagens para medicina humana e animal são combinadas para o combate a zoonoses [5], **telemedicina** e **saúde móvel**, bem como para otimização da **gestão hospitalar**. Esta palestra introduz conceitos referentes a *big data* e sua inserção na área da Saúde, discute o uso de ferramentas computacionais para o processamento de grandes volumes de dados e análise de cenários complexos envolvendo dados multimodais e desfechos específicos (por exemplo, Covid-19), bem como apresenta desafios atuais e próximas fronteiras referentes à exploração de *big data* em saúde.

Referências

[1] Gil Press. **A very short history of big data**. Forbes. 2013. Disponível em <https://www.forbes.com/sites/gilpress/2013/05/09/a-very-short-history-of-big-data/>

[2] Juan Pablo Boeira. **Conheça os 42 Vs do big data**. Época Negócios. 2020. Disponível em <https://epocanegocios.globo.com/colunas/Changemaker/noticia/2020/08/conheca-os-42-vs-do-big-data.html>

[3] Po-Yen Wu et al. **-Omic and electronic health record big data analytics for precision medicine**. IEEE Transactions on Biomedical Engineering. 64(2), 2017. 263-273 pp. DOI: [10.1109/TBME.2016.2573285](https://doi.org/10.1109/TBME.2016.2573285)

[4] Michael Matheny et al (Eds). **Artificial intelligence in health care: the hope, the hype, the promise, the peril**. The Learning Health System Series. Washington, DC: National Academy of Medicine, 2019. 245 pp.

[5] Jakob Zinsstag et al (Eds). **One health: the theory and practice of integrated health approaches**. 2 ed. Wallingford, UK: CAB International, 2021. 425 pp.

Suporte Financeiro: Newton International Fellowships Programme (The Royal Society, UK), NVIDIA Large-scale Applied Data Science Grant (NVIDIA Corporation), Covid-19 AI for Social Good Programme (Google.org), MRC / GCRF Global Multimorbidity (UK), Bill & Melinda Gates Foundation – Data Science Approaches to Improve Maternal and Child Health in Brazil.