

Análise inteligente de dados para o combate da Covid-19 na cidade de Recife

Raydonal Ospina

Departamento de Estatística, CASTLab
Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.
e-mail: raydonal@castlab.org

Andre Leite

Departamento de Estatística, CASTLab
Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.
e-mail: leite@castlab.org

André Magalhães

Departamento de Economia,
Universidade Federal de Pernambuco, Recife, Pernambuco, Brasil.
e-mail: andre.magalhaes@ufpe.br

RESUMO

A epidemia do Covid-19 é uma das piores crises de saúde pública que o Brasil já enfrentou. A grande dificuldade dos estados e seus sistemas de controle sanitário para a tomada de decisões é que protocolos testados em outras epidemias não são garantia de sucesso para seu controle dada sua complexidade. Nesse contexto, uma resposta eficaz a fim de orientar as autoridades competentes na adoção de políticas públicas para o enfrentamento da epidemia do Covid-19 depende de uma análise e visualização esperta, idealmente baseada em diferentes fontes de informação. Neste trabalho, abordamos algumas ferramentas que podem ser adotadas numa tentativa de avaliar a propagação do vírus e o sucesso dos esforços de contenção da saúde pública para diminuir a velocidade e grau de propagação da doença na cidade de Recife em Brasil.

PALAVRAS CHAVE. Covid-19, dados, modelagem, visualização

ABSTRACT

The Covid-19 epidemic is one of the worst public health crises that Brazil has ever faced. The major difficulty for states and their health control systems for decision making is that protocols tested in other epidemics are no guarantee of success for their control given their complexity. In this context, an effective response in order to guide the competent authorities in the adoption of public policies to face the Covid-19 epidemic depends on smart analysis and visualization, ideally based on different sources of information. In this work, we discuss some tools that can be adopted in an attempt to assess the spread of the virus and the success of efforts to contain public health to decrease the speed and degree of spread of the disease in the city of Recife in Brazil.

KEYWORDS. Covid-19, data, modeling, visualization

1. Introdução

A epidemia da Covid-19, causada pelo coronavírus (SARS-CoV-2) se alastrou rapidamente por todo o mundo, a partir de 16 de fevereiro de 2020, tinham sido relatados casos em mais de 26 países tornando-se desta forma em uma das maiores crises de saúde pública mundial [CSSE 2020]. No Brasil, o primeiro caso da doença foi confirmado na cidade de São Paulo em 26 de fevereiro de 2020. Desde então, a epidemia se alastrou pelo país, o que tem obrigado a estados e municípios brasileiros a adotarem medidas de distanciamento social e medidas restritivas, para conter a propagação do vírus. Até a presente data (29/07/2020) o país já acumula um total de mais 88 mil mortes pela Covid-19 sendo que mais 2 mil óbitos provém da cidade de Recife. Contudo, a taxa de crescimento da epidemia ainda não apresenta sinais evidentes de arrefecimento e a doença ainda continua em franca expansão no país, embora em um ritmo mais lento [Brum et al. 2020]. Alguns estados e municípios do Brasil já lograram um certo controle sobre a epidemia, o qual se manifesta, por exemplo, em uma desaceleração das respectivas curvas de óbitos [Duarte-Filho et al. 2020].

No contexto da resposta a esta pandemia, observamos que é fundamental construir respostas baseadas em conhecimento aberto e compartilhado a fim de que a circulação da informação seja mais rápida e a iteração de pesquisas emergentes relacionadas a nova doença permita uma maior integração de múltiplas fontes de dados para mapear e antecipar a propagação do vírus. Também é fundamental a publicação de recursos abertos para comunicação com o público, treinamento contínuo e a abertura geral de conhecimentos especializados para facilitar a continuidade dos serviços e da atividade econômica, especialmente durante o período em que muitas pessoas estão em quarentena ou em condições precárias e restritas como as impostas pela pandemia da Covid-19.

2. Resposta a epidemia do Covid-19

A complexidade e mudanças de uma epidemia ou surto como a do Covid-19 implica numa resposta dinâmica das políticas de saúde públicas, protocolos de controle e análise de informações. De forma geral podemos identificar quatro estágios de resposta ao surto. O estágio de *detecção* que começa com o primeiro caso e termina com a primeira intervenção de atividades (por exemplo, isolamento do paciente, rastreamento de contatos, vacinação) e envolve sistemas de vigilância e principalmente medidas qualitativas de avaliação de risco. Em seguida, a *resposta inicial* é a parte inicial da intervenção durante a qual a primeira análise simples pode ocorrer, essencialmente centrado em torno da dinâmica da transmissão da infecção. Isso se mistura ao estágio de *intervenção*, em que análises mais complexas podem estar envolvidas para informar (por exemplo, estratégias de vacinação, lockdown intermitente), que termina uma vez o último caso relatado se recupera ou morre. A *pós-intervenção* é o estágio em que lições aprendidas, podem ajudar e melhorar protocolos e preparação para uma próxima epidemia.

Uma característica importante da resposta moderna às epidemias é o foco crescente na exploração de todos os dados disponíveis para o monitoramento dos estágios de resposta a epidemia para fornecer respostas rápidas que permitam a tomada de decisão baseada em evidências [Cauchemez et al. 2014], [Cori et al. 2017]. O uso de dados para melhorar a percepção da situação é complexo, envolvendo uma gama de funções e técnicas interligadas, desde a coleta de dados nos pontos de atendimento até a geração de relatórios situacionais informativos (*dashboards*, *tweets*,

etc) [Yozwiak et al. 2015]. No caso do Recife podemos citar o esforço colaborativo do D.A.D.O. (Dados e Análises para Decisões e Operações) disponível em <https://dado.recife.br/> e do CASTLab (Computational Agricultural Statistics Laboratory) que fornecem diferentes índices, métricas e visualizações para informar a população sobre o estágio da epidemia.

A ciência subjacente a esses ambientes de dados aplicados por exemplo no D.A.D.O. envolve uma ampla gama de abordagens, incluindo design de banco de dados e tecnologia móvel [Drew et al. 2020], estatísticas frequentista e estimções via máxima verossimilhança [Zhuang et al. 2020], [Plohl e Musil 2020] visualização interativa de dados [Dong et al. 2020], [Dey et al 2020], [Hamzah et al. 2020], geoestatística [Cheshmehzangi 2020], [Fronterre et al. 2020] teoria dos grafos [Nagraj et al. 2018] [Block et al. 2020], [Liu et al. 2020], estatística bayesiana [Parag e Donnelly 2020], [Chowell, 2020], modelagem matemática e computacional [Liang 2020], [Van den Broeck et al 2011], [Vasconcelos et al. 2020a], [Moore e Rogers 2020] [Chowell 2017a,], análises genéticas [Heymann e Shindo 2020], [Sun et al. 2020] e abordagens de síntese de evidências [Birrell et al. 2018], [Kucharski et al. 2020]. [Alahmadi et al. 2020].

Essa interdisciplinaridade de áreas pode ser apresentada como *análise inteligente para o combate de surtos (A.I.C.S)*, e forma um domínio emergente da ciência de dados no sentido de fazer uso de métodos, processos, algoritmos e sistemas científicos para extrair conhecimento útil e insights de dados a fim de dar uma resposta a epidemia. A *A.I.C.S.* interseja o conhecimento e planejamento de saúde e políticas públicas, a epidemiologia o desenvolvimento metodológico e tecnologias da informação para atender às necessidades de uma resposta à epidemia. Neste trabalho, nos concentramos em algumas ferramentas que podem ajudar desde a perspectiva de análise de dados para dar resposta a surto do Covid-19 na cidade de Recife.

3. Analítica inteligente da epidemia do Covid-19 no Recife

Podemos pensarmos a *A.I.C.S.* como a variedade de ferramentas e métodos usados para coletar, selecionar, visualizar, analisar, modelar, simular, otimizar e relatar dados sobre surtos de forma inteligente e interligada.

Para uma analítica inteligente é preciso usar e otimizar uma variedade de fontes de informação de livre acesso para atualizar e selecionar bancos de dados [Ienca e Vayen 2020], [Moorthy et al. 2020]. Nosso caso do Recife as fontes de dados governamentais e oficiais provém do Centro de Informações Estratégicas em Vigilância em Saúde <https://cievsrecife.wordpress.com>, Secretaria de Planejamento e Gestão de Pernambuco <https://www.seplag.ce.gov.br/>, Community Mobility Reports da Google para índices de mobilidade <https://www.google.com/covid19/mobility/> e Ministério da Saúde <https://covid.saude.gov.br>, além de informações de artigos científicos revisados por pares que relatam dados primários como o padrão-ouro para inclusão de dados. Outras fontes do governo incluem comunicados de imprensa nos sites oficiais dos Ministérios da Saúde. Para encontrar detalhes adicionais para cada caso ou paciente, os dados podem ser aumentados com laudos médicos, informações de cartórios e relatórios capturados (*Web scraping*) principalmente através de sites de notícias.

Informações geográficas podem ser usadas para avaliar espalhamento da doença sendo que informações georeferenciadas a assentamentos específicos ou áreas, tipicamente de unidades administrativas como bairros ou setores censitários podem ajudar numa visão macro do casos.

Todos os metadados geográficos devem ser padronizados a fim de tornar possível comparações de localidades e geração de mapas e indicadores, como por exemplo os apresentados na Figura 1.

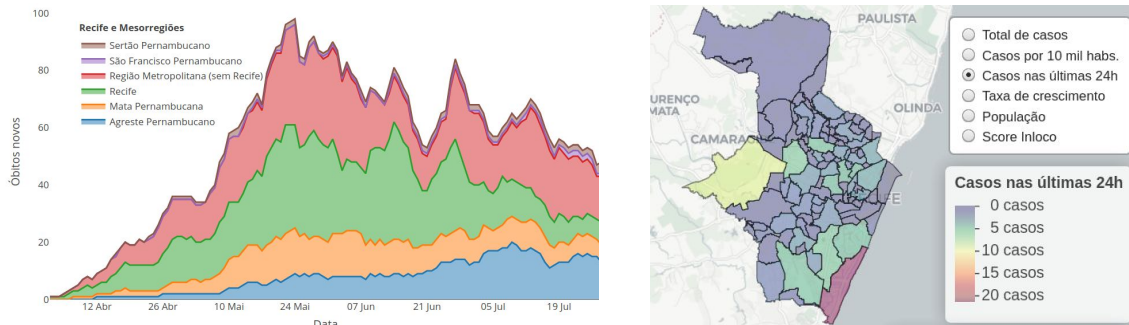


Figura 1: Representação gráfica de indicadores com informação geográfica. (a) Número de óbitos novos por mesorregião e Recife. (b) Mapa coroplético por bairros da Cidade de Recife com Número de casos nas últimas 24 hs

Após a entrada inicial dos dados, o banco de dados deve ser submetido a uma verificação e análise de consistência usando metodologias complementares para identificar possíveis registros duplicados, valores faltantes, etc. Dentre elas podemos mencionar a verificação manual feita pelos curadores de dados ou o uso de ferramentas estatísticas e de AI (Artificial Intelligence - Inteligências artificial) como *Feature Hashing* e análise de componentes principais entre outros [Duboue 2020], [Khashan et al. 2020].

Com os dados verificados pode ser investigado o surto epidemiológico de COVID-19 no Recife incluído mapeamento descritivo das ocorrências ao longo do tempo e estimativa dos principais parâmetros epidemiológicos usando modelos matemáticos. Especificamente podemos obter inteligentes insights das curvas epidêmicas (C.E.) que representam o número de novos casos ou óbitos por unidade de tempo com base na data ou hora do início dos sintomas [Thibaut, 2020]. As C.E. são úteis em muitos aspectos elas fornecem um esboço visual simples de dinâmica demográfica, que pode ser usada para avaliar o crescimento ou declínio de um surto [Matrajt e Leung 2020] além de permitir avaliar o efeito de medidas de intervenções não farmacológicas como nos gráficos da Figura 2.

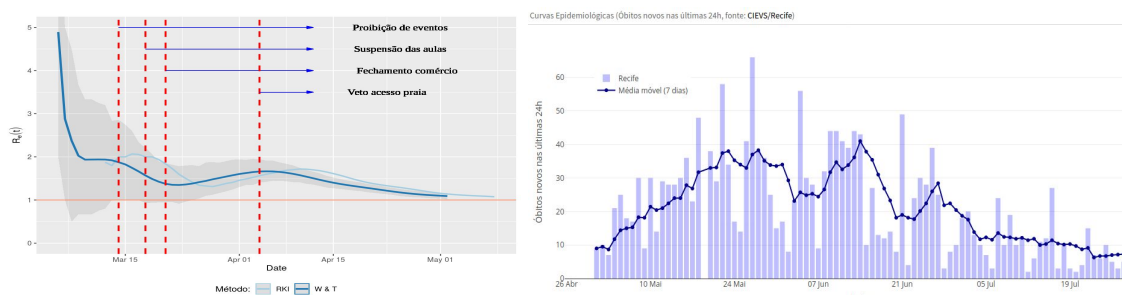


Figura 2: Curvas Epidêmicas para Recife. (a) Evolução temporal da taxa de Reprodução Efetiva e datas de intervenção. (b) Evolução temporal do número de casos com monitoramento de tendência por média móvel de 7 dias.

Além disso, as C.E. também formam a matéria-prima usada por uma variedade de técnicas de modelagem para monitoramento e previsão [Vasconcelos et al. 2020b], [Wu 2020]. Por exemplo, se usamos a curva epidêmica para o número acumulado de óbitos, em função do tempo, ela pode apresentar três regiões distintas de comportamento. Uma fase inicial, em que tipicamente verifica-se um crescimento exponencial da curva epidêmica, a partir do número de óbitos observados no primeiro dia em que um óbito aconteceu. Essa subida rápida é caracterizada por uma taxa r de crescimento. Contudo, a depender da complexidade da dinâmica de contágio, que pode resultar em uma rede complexa de contatos humanos, ou quando medidas de mitigação são adotadas no início da epidemia, é possível que esse crescimento dê-se de forma mais lenta que o crescimento exponencial (*subexponencial*) caracterizado através de um parâmetro $q \in (0, 1)$, que interpola entre o regime linear ($q = 0$) e o regime exponencial ($q = 1$).

Na sua região intermediária, a curva epidêmica apresenta um *ponto de inflexão*, aqui denotado por t_c (vide Figura 3(a)), que corresponde ao instante de tempo em que a curva acumulada muda sua concavidade, ou seja, a taxa de crescimento atinge seu valor máximo e passa a decrescer a partir de então. A parte final da curva, após o ponto de inflexão, pode ser caracterizada por um parâmetro α que controla a rapidez com que a curva epidêmica se afasta da tendência linear e “se dobra” em direção do platô. O valor do platô, aqui denotado pelo parâmetro K , representa o número total de óbitos no final da epidemia. Com esta informação em mente podemos modelar a C.E. através de um modelo generalizado de Richards (MGR) [Chowell 2016], [Bürger et al. 2019].

No caso de Recife é possível constatar que a C.E. se encontra na fase de saturação até o dia 30/07/2020 evidenciando que medidas de mitigação da Covid-19 adotadas no Recife tiveram efeito expressivo no controle da pandemia (vide Figura 3(b)).

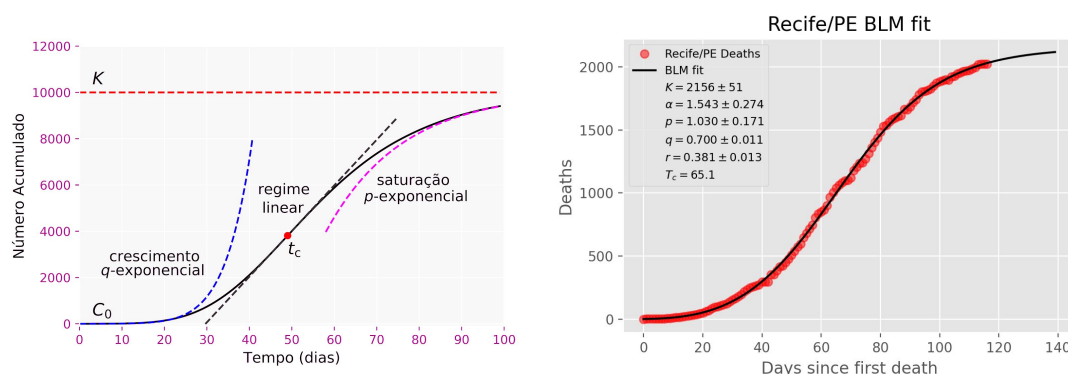


Figura 3. Painel esquerdo (a) mostra o esboço qualitativo de uma curva epidêmica (preto) para o número acumulado de óbitos, com indicação de suas diferentes ‘fases’. O círculo vermelho indica o ponto de inflexão da curva. Painel da direita (b) representa o ajuste pelo modelo generalizado de Richards do número cumulativo de óbitos pela Covid-19 no Recife. Os círculos vermelhos representam os dados e a curva preta é o ajuste pelo modelo.

Os dados disponíveis abertamente deverão ser selecionados continuamente e usar informações via modelagem e cenários de simulação a fim de retroalimentar e manter o banco de dados atualizado. No entanto, se a epidemia continuar a crescer, as agências de saúde pública tendem a omitir os casos em nível individual e passam a relatar apenas o número total (ou estimativas) de casos confirmados ou suspeitos, como ocorreu em grandes surtos anteriores, como o da gripe pandêmica H1N1 [Khan et al. 2020]. Quando dados detalhados se tornam cada vez menos disponíveis à medida que a epidemia cresce, podemos fazer a transição para um estrutura de banco de dados aumentada que relata apenas o total de novos casos por local. De qualquer forma, os dados disponíveis principalmente numa doença nova como a Covid-19 geralmente não são suficientes para investigar o impacto de pressupostos em projeções futuras relacionadas com o planejamento de políticas de saúde pública.

4. Análise automático da epidemia: Desafios e oportunidades

Um passo fundamental que o Recife está dando na criação de uma cultura de construção de cenários para desenho e evolução de estratégias para a cidades incluem o uso Aprendizado de máquina (ML), inteligência artificial (IA) e outros métodos estatísticos modernos. Estas novas ferramentas oferecem novas oportunidades para operacionalizar fontes de dados anteriormente inexploradas e em rápido crescimento (Big Data) além de ajudar a sintetizar informações para entender padrões epidemiológicos o de risco da cidade além de ajudar a desenvolver evidências quantitativa para a tomada de decisões em saúde e política pública. Nesse sentido, abordagens analisadas por [Gupta e Katarya 2020], [Lanza et al. 2020], [Vollmer et al. 2020] podem ser de potencial ajuda para a implementação de metodologias mais aprimoradas na cidade para o controle de calamidades. Os usos potenciais de ML/AI incluem melhorar a precisão do diagnóstico da Covid-19 (ou de alguma nova doença ou calamidade), prognóstico mais confiável, tratamentos direcionados, e aumento da eficiência operacional dos sistemas de saúde, entre outros [Badie-Modiri et al. 2020], [Topirceanu et al. (2020)], [Alamo et al. 2020], [Van den Broeck et al. 2020], [Nikolaou e Dimitriou 2020], [Ruiz-Estrada e Koutronas 2020].

Exemplos de tecnologia potencialmente disruptiva com promessa antecipada incluem aplicativos de diagnóstico de ML/AI com base em imagem, que mostraram a promessa clínica mais precoce (por exemplo, algoritmos baseados em aprendizado profundo que melhoram a precisão no diagnóstico de patologias em comparação com os médicos especialistas [Wang et al 2020], [Pereira et al. 2020], ou processamento de linguagem natural usado como uma ferramenta para extrair informações de texto estruturado e não estruturado (ou seja, gratuito) incorporado nos registros eletrônicos de saúde [Son et al. 2019], [Ayyoubzadeh et al. 2020], [Jahanbin e Rahmanian 2020].

Embora estejamos apenas começando a entender a riqueza de oportunidades oferecidas por esses métodos, há uma crescente preocupação na comunidade acadêmica e o setor público porque os resultados baseados em análise automática não são percebidos da mesma maneira que outras intervenções médicas (por exemplo, farmacológicas), [Bukowski et al. 2020], [Foraker et al. 2018] eles não têm diretrizes bem definidas para seu desenvolvimento e raramente sofrem o mesmo grau de escrutínio [Hollingsworth e Medley 2017]. Várias publicações de alto padrão demonstraram falta de transparência, replicabilidade, ética, e eficácia no relato e avaliação de

modelos de previsão baseados em ML/AI. Esse crescente conjunto de evidências sugere que, embora muitas recomendações de melhores práticas para projetos, conduta, análise, relatório, avaliação de impacto e implementação clínica, possam ser emprestadas da literatura tradicional de economia, política pública, sistemas de saúde e estatística médica, elas não são suficientes para orientar o uso de ML/IA em pesquisa. A produção de tais orientações é um empreendimento importante devido à crescente bateria de algoritmos ML/AI e à natureza multifacetada da avaliação de desempenho e impacto clínico político e social. Não tomar medidas é inaceitável e, se esperarmos por uma solução mais definitiva, arriscamos desperdiçar um trabalho valioso [Király, et al. 2018], [Collins et al. 2015], [Salman et al. 2014] , [Cortes et al. 1995].

Referências

Alamo, T., Reina, D. G., & Millán, P. (2020). Data-Driven Methods to Monitor, Model, Forecast and Control Covid-19 Pandemic: Leveraging Data Science, Epidemiology and Control Theory. *arXiv preprint arXiv:2006.01731*.

Alahmadi, A., Belet, S., Black, A., Cromer, D., Flegg, J. A., House, T., ... & Ross, J. V. (2020). Influencing public health policy with data-informed mathematical models of infectious diseases: Recent developments and new challenges. *Epidemics*, 100393.

Ayyoubzadeh, S. M., Ayyoubzadeh, S. M., Zahedi, H., Ahmadi, M., & Kalhori, S. R. N. (2020). Predicting COVID-19 incidence through analysis of google trends data in iran: data mining and deep learning pilot study. *JMIR Public Health and Surveillance*, 6(2), e18828.

Badie-Modiri, A., Karsai, M., & Kivelä, M. (2020). Efficient limited-time reachability estimation in temporal networks. *Physical Review E*, 101(5), 052303.

Birrell, P. J., De Angelis, D., & Presanis, A. M. (2018). Evidence synthesis for stochastic epidemic models. *Statistical science: a review journal of the Institute of Mathematical Statistics*, 33(1), 34.

Block, P., Hoffman, M., Raabe, I. J., Dowd, J. B., Rahal, C., Kashyap, R., & Mills, M. C. (2020). Social network-based distancing strategies to flatten the COVID-19 curve in a post-lockdown world. *Nature Human Behaviour*, 1-9.

Brum, A. A., Duarte-Filho, G. C., & Vasconcelos, G. L. (2020). Application Modinterv Covid-19. *Disponível em: fisica.ufpr.br/modinterv*.

Bukowski, M., Farkas, R., Beyan, O., Moll, L., Hahn, H., Kiessling, F., & Schmitz-Rode, T. (2020). Implementation of eHealth and AI integrated diagnostics with multidisciplinary digitized data: are we ready from an international perspective?. *Eur Radiol*. <https://doi.org/10.1007/s00330-020-06874-x>.

Bürger, R., Chowell, G., and Lara-Díaz, L.Y. (2019). Comparative analysis of phenomenological growth models applied to epidemic outbreaks. *Mathematical Biosciences and Engineering: MBE* 16, 4250–4273.

Cauchemez, S., Fraser, C., Van Kerkhove, M. D., Donnelly, C. A., Riley, S., Rambaut, A., ... & Ferguson, N. M. (2014). Middle East respiratory syndrome coronavirus: quantification of the extent of the epidemic, surveillance biases, and transmissibility. *The Lancet infectious diseases*, 14(1), 50-56.

Cheshmehzangi, A. (2020). How Cities Cope in Outbreak Events?. In *The City in Need* (pp. 17-39). Springer, Singapore.

Chowell, G., Hincapié-Palacio, D., Ospina, J., Pell, B., Tariq, A., Dahal, S., Moghadas, S., Smirnova, A., Simonsen, L., and Viboud, C. (2016). Using phenomenological models to characterize transmissibility and forecast patterns and final burden of zika epidemics. *PLOS Currents Outbreaks*.

Chowell, G. (2017). Fitting dynamic models to epidemic outbreaks with quantified uncertainty: A primer for parameter uncertainty, identifiability, and forecasts. *Infectious Disease Modelling*, 2(3), 379-398.

Chowell, G., Luo, R., Sun, K., Roosa, K., Tariq, A., & Viboud, C. (2020). Real-time forecasting of epidemic trajectories using computational dynamic ensembles. *Epidemics*, 30, 100379.

Collins, G. S., Reitsma, J. B., Altman, D. G., & Moons, K. G. (2015). Transparent Reporting of a Multivariable Prediction Model for Individual Prognosis or Diagnosis (TRIPOD) The TRIPOD Statement. *Circulation*, 131(2), 211-219.

Cori, A., Donnelly, C. A., Dorigatti, I., Ferguson, N. M., Fraser, C., Garske, T., ... & Van Kerkhove, M. D. (2017). Key data for outbreak evaluation: building on the Ebola experience. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372(1721), 20160371.

Cortes, C., Jackel, L. D., & Chiang, W. P. (1995). Limits on learning machine accuracy imposed by data quality. In *Advances in Neural Information Processing Systems* (pp. 239-246).

CSSE, J. (2020). Coronavirus COVID-19 Global Cases by the Center for Systems Science and Engineering (CSSE) at Johns Hopkins University (JHU). 2020-03-15]. <https://gisanddata.maps.arcgis.com/apps/opsdashboard/index.html#/bda7594740fd40299423467b48e9ecf>

Dey, S. K., Rahman, M. M., Siddiqi, U. R., & Howlader, A. (2020). Analyzing the epidemiological outbreak of COVID-19: A visual exploratory data analysis approach. *Journal of medical virology*, 92(6), 632-638.

Dong, E., Du, H., & Gardner, L. (2020). An interactive web-based dashboard to track COVID-19 in real time. *The Lancet infectious diseases*, 20(5), 533-534.

Drew, D. A., Nguyen, L. H., Steves, C. J., Menni, C., Freydin, M., Varsavsky, T., ... & Spector, T. D. (2020). Rapid implementation of mobile technology for real-time epidemiology of COVID-19. *Science*.

Duarte-Filho, G. C., Brum, A. A., Ospina, R., Almeida, F. A., Macêdo, A. M., & Vasconcelos, G. L. (2020). Recife e Belém são atualmente as únicas capitais que já estão na fase de saturação da Covid-19 no Brasil. *SciELO Preprints*.

Duboue, P. (2020). *The Art of Feature Engineering: Essentials for Machine Learning*. Cambridge University Press.

Foraker, R., Mann, D. L., & Payne, P. R. (2018). Are synthetic data derivatives the future of translational medicine?. *JACC: Basic to Translational Science*, 3(5), 716.

Fronterre, C., Read, J. M., Rowlingson, B., Bridgen, J., Alderton, S., Diggle, P. J., & Jewell, C. P. (2020). COVID-19 in England: spatial patterns and regional outbreaks. *medRxiv*.

Gupta, A., & Katarya, R. (2020). Social Media based Surveillance Systems for Healthcare using Machine Learning: A Systematic Review. *Journal of Biomedical Informatics*, 103500.

Hamzah, F. B., Lau, C., Nazri, H., Ligot, D. V., Lee, G., & Tan, C. L. (2020). CoronaTracker: worldwide COVID-19 outbreak data analysis and prediction. *Bull World Health Organ*, 1, 32.

Heymann, D. L., & Shindo, N. (2020). COVID-19: what is next for public health?. *The Lancet*, 395(10224), 542-545.

Hollingsworth, T. D., & Medley, G. F. (2017). Learning from multi-model comparisons: collaboration leads to insights, but limitations remain. *Epidemics*, 18.

Ienca, M., & Vayena, E. (2020). On the responsible use of digital data to tackle the COVID-19 pandemic. *Nature medicine*, 26(4), 463-464.

Jahanbin, K., & Rahmanian, V. (2020). Using Twitter and web news mining to predict COVID-19 outbreak. *Asian Pacific Journal of Tropical Medicine*, 13.

Khan, U., Mehta, R., Arif, M. A., & Lakhani, O. J. (2020). Pandemics of the past: A narrative review. *Public Health*, 70(5).

Khashan, E. A., Eldesouky, A. I., Fadel, M., & Elghamrawy, S. M. (2020). A Big Data Based Framework for Executing Complex Query Over COVID-19 Datasets (COVID-QF). *arXiv preprint arXiv:2005.12271*.

Király, F. J., Mateen, B., & Sonabend, R. (2018). NIPS-Not Even Wrong? A Systematic Review of Empirically Complete Demonstrations of Algorithmic Effectiveness in the Machine Learning and Artificial Intelligence Literature. *arXiv preprint arXiv:1812.07519*.

Kucharski, A. J., Russell, T. W., Diamond, C., Liu, Y., Edmunds, J., Funk, S., ... & Davies, N. (2020). Early dynamics of transmission and control of COVID-19: a mathematical modelling study. *The lancet infectious diseases*.

Lanza, F., Seidita, V., & Chella, A. (2020). Agents and Robots for Collaborating and Supporting Physicians in Healthcare Scenarios. *Journal of Biomedical Informatics*, 103483.

Liang, K. (2020). Mathematical model of infection kinetics and its analysis for COVID-19, SARS and MERS. *Infection, Genetics and Evolution*, 104306.

Liu, C., Wu, X., Niu, R., Wu, X., & Fan, R. (2020). A new SAIR model on complex networks for analysing the 2019 novel coronavirus (COVID-19). *Nonlinear Dynamics*, 1-11

Matrajt, L., & Leung, T. (2020). Evaluating the Effectiveness of Social Distancing Interventions to Delay or Flatten the Epidemic Curve of Coronavirus Disease. *Emerging infectious diseases*, 26(8).

Moore, S., & Rogers, T. (2020). Predicting the Speed of Epidemics Spreading in Networks. *Physical Review Letters*, 124(6), 068301.

Moorthy, V., Restrepo, A. M. H., Preziosi, M. P., & Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19). *Bulletin of the World Health Organization*, 98(3), 150.

Nagraj, V. P., Randhawa, N., Campbell, F., Crellen, T., Sudre, B., & Jombart, T. (2018). epicontacts: Handling, visualisation and analysis of epidemiological contacts. *F1000Research*, 7.

Nikolaou, P., & Dimitriou, L. (2020). Identification of critical airports for controlling global infectious disease outbreaks: Stress-tests focusing in Europe. *Journal of Air Transport Management*, 101819.

Ozturk, T., Talo, M., Yildirim, E. A., Baloglu, U. B., Yildirim, O., & Acharya, U. R. (2020). Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, 103792.

Parag, K. V., & Donnelly, C. A. (2020). Using information theory to optimise epidemic models for real-time prediction and estimation. *PLOS Computational Biology*, 16(7), e1007990.

Pereira, R. M., Bertolini, D., Teixeira, L. O., Silla Jr, C. N., & Costa, Y. M. (2020). COVID-19 identification in chest X-ray images on flat and hierarchical classification scenarios. *Computer Methods and Programs in Biomedicine*, 105532.

Plohl, N., & Musil, B. (2020). Modeling compliance with COVID-19 prevention guidelines: The critical role of trust in science. *Psychology, Health & Medicine*, 1-12.

Ruiz-Estrada, M. A., & Koutronas, E. (2020). The Application of the 2019-nCoV Global Economic Impact Simulator (the 2019-nCoV-GEI-Simulator) in China. *Available at SSRN 3542817*.

Salman, R. A. S., Beller, E., Kagan, J., Hemminki, E., Phillips, R. S., Savulescu, J., ... & Chalmers, I. (2014). Increasing value and reducing waste in biomedical research regulation and management. *The Lancet*, 383(9912), 176-185.

Shi, F., Wang, J., Shi, J., Wu, Z., Wang, Q., Tang, Z., ... & Shen, D. (2020). Review of artificial intelligence techniques in imaging data acquisition, segmentation and diagnosis for covid-19. *IEEE reviews in biomedical engineering*.

Son, H., Hyun, C., Phan, D., & Hwang, H. J. (2019). Data analytic approach for bankruptcy prediction. *Expert Systems with Applications*, 138, 112816.

Sun, J., He, W. T., Wang, L., Lai, A., Ji, X., Zhai, X., ... & Veit, M. (2020). COVID-19: epidemiology, evolution, and cross-disciplinary perspectives. *Trends in Molecular Medicine*.

Thibaut Jombart, Zhian N. Kamvar, Rich FitzJohn, Jun Cai, Sangeeta Bhatia, Jakob Schumacher and Juliet R.C. Pulliam (2020). incidence: Compute, Handle, Plot and Model Incidence of Dated Events. *R package version 1.7.2*. URL <https://doi.org/10.5281/zenodo.2584018>

Topirceanu, A., Udrescu, M., & Marculescu, R. (2020). Centralized and decentralized isolation strategies and their impact on the COVID-19 pandemic dynamics. *arXiv preprint arXiv:2004.04222*.

Van den Broeck, W., Gioannini, C., Gonçalves, B., Quaggiotto, M., Colizza, V., & Vespignani, A. (2011). The GLEaMviz computational tool, a publicly available software to explore realistic epidemic spreading scenarios at the global scale. *BMC infectious diseases*, 11(1), 1-14.

Vasconcelos, G. L., Macêdo, A. M., Ospina, R., Almeida, F. A., Duarte-Filho, G. C., Brum, A. A., & Souza, I. C. (2020a). Modelling fatality curves of COVID-19 and the effectiveness of intervention strategies. *PeerJ*, 8, e9421.

Vasconcelos, G.L., Macêdo, A.M.S., Duarte-Filho, G.C., Brum, A.A., Ospina, R., Almeida, F.A., (2020b). Complexity signatures in the COVID-19 epidemic: power law behaviour in the saturation regime of fatality curves”, medRxiv, <https://doi.org/10.1101/2020.07.12.20152140>.

Vollmer, S., Mateen, B. A., Bohner, G., Király, F. J., Ghani, R., Jonsson, P., ... & Granger, D. (2020). Machine learning and artificial intelligence research for patient benefit: 20 critical questions on transparency, replicability, ethics, and effectiveness. *bmj*, 368.

Wang, S., Kang, B., Ma, J., Zeng, X., Xiao, M., Guo, J., ... & Xu, B. (2020). A deep learning algorithm using CT images to screen for CoronaVirus Disease (COVID-19). *MedRxiv*.

Wu, K., Darcet, D., Wang, Q., and Sornette, D. (2020). Generalized logistic growth modeling of the covid-19 outbreak in 29 provinces in china and in the rest of the world. medRxiv, preprint *medRxiv:2020.03.11.20034363*.

Yozwiak, N. L., Schaffner, S. F., & Sabeti, P. C. (2015). Data sharing: Make outbreak research open access. *Nature*, 518(7540), 477-479.

Zhang, Dayong, Min Hu, and Qiang Ji. "Financial markets under the global pandemic of COVID-19." *Finance Research Letters* (2020): 101528.

Zhuang, Z., Zhao, S., Lin, Q., Cao, P., Lou, Y., Yang, L., & He, D. (2020). Preliminary estimation of the novel coronavirus disease (COVID-19) cases in Iran: A modelling analysis based on overseas cases and air travel data. *International Journal of Infectious Diseases*, 94, 29-31.