

Características dos Casos Graves no Brasil sobre Vítimas da Covid-19, com Modelo de *Machine Learning* para Predição de Mortes.

Lucas Pedroso P. da Costa^{1*}, Maria Begoña G. Zapirain², Leonardo R. P. Rauta¹.

¹Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas – Instituto Federal de Santa Catarina (IFSC)
Caixa Postal 89.111-009 – Gaspar – SC – Brasil

²Departamento de Ingeniería – Universidad de Deusto
Bilbao, Spain.

{lucas.pcc@aluno.ifsc.edu.br, mbgarciazapi@deusto.es,
leonardo.rauta@ifsc.edu.br}

Abstract. *This paper has the objective to describe the Brazilian situation about the new pandemic outbreak caused by the coronavirus Sars-Cov-2, detailing which characteristics are more correlated to cases who needed hospitalization and the ones who passed away. These analyses were built with data from seven states, those ones that matched with the requisites described in this project. Beyond the statistic part, was build a model capable to predict deaths based in biological conditions and symptoms, of data available from dataset of the state of Espírito Santo, reaching the accuracy of approach 96%.*

Resumo. *Este artigo tem como objetivo descrever detalhes sobre a situação brasileira na nova pandemia causada pelo coronavírus Sars-Cov-2, detalhando quais são as características mais relacionados com casos que necessitaram de hospitalização e casos que evoluíram a óbito. As análises foram construídas com base nos dados de sete estados brasileiros, dos quais se encaixaram nos requisitos descritos neste trabalho. Além da realização de uma análise estatística, foi construído um modelo capaz de prever óbitos baseado em algumas condições biológicas e sintomatológicas com base no dataset do Espírito Santo, modelo do qual demonstrou uma acurácia de aproximadamente 96%.*

1. Introdução

No mês de março de 2020 a Organização Mundial da Saúde decretou uma nova pandemia mundial causada pelo novo coronavírus Sars-Cov-2 [OMS 2020]. Um vírus com potencial letal, e alta disseminação, espalhou-se através do mundo em poucos meses. A nova corrida mundial é a respeito de quem produz a primeira vacina, sem deixar sua economia parar. Tudo isso vem com o peso das mortes incessantes e constantes especialmente no Brasil, país abordado no trabalho.

* Projeto resultante do intercâmbio estudantil entre o Instituto Federal de Santa Catarina (IFSC) e a Universidade de Deusto, promovido pelo edital N°4/2019/Gabinete da Reitoria PROPICIE 17.

Os estados brasileiros são submetidos aos poderes dos governadores e prefeitos, sendo assim, a autoridade responsável pelas medidas de controle e combate ao surto. Não apenas, também disponibilizam os dados referentes aos próprios com intuito de informar não apenas a imprensa como a população da corrente situação sanitária. Baseado nessas informações e para a realização deste trabalho, foram selecionadas 7 das 27 unidades federativas do Estado brasileiro, sendo elas: Acre, Alagoas, Ceará, Espírito Santo, Mato Grosso do Sul, Rondônia e Santa Catarina. Foram selecionadas pois apresentaram dados estruturados o suficiente para serem utilizados para pesquisas estatísticas sobre as condições mais comuns entre os casos graves.

Não apenas no intuito de registros estatísticos, a elaboração de um modelo capaz de prever óbitos com características biológicas, sintomáticas e morbidades foi efetuada. As séries de dados utilizadas para a elaboração do modelo são referentes ao estado do Espírito Santo, localizado no sudeste brasileiro. A escolha deste deve-se ao bom registro dos dados desde o início da pandemia com as *features* desejadas, das quais são descritas na seção da Coleta de Dados.

Vale ressaltar que os dados obtidos para as análises estatísticas correspondem até o dia 31 de maio de 2020, e os dados utilizados para o modelo de *machine learning* correspondem até o dia 19 de junho de 2020. Para a revisão bibliográfica e o estudo do estado da arte, foram estudados projetos até a data de 24 de junho de 2020.

2. Sars-Cov-2

Em dezembro de 2019 foi noticiado um novo vírus na província de Hubei na China [Huang et al. 2020]. Identificado como um novo coronavírus, o até então nomeado 2019-n Cov causou alarme através de alguns países. Verificou-se também a presente semelhança com o vírus SARS, atualmente nomeado SARS-Cov, vírus responsável pelo primeiro surto epidêmico de coronavírus em humanos no ano de 2002 [Xie et al. 2020].

A velocidade de espalhamento demonstrou-se avassaladora, quando em apenas 26 dias, considerando como dia inicial o dia 11 de janeiro de 2020, somente na China, 31170 pessoas foram diagnosticadas [Wu et al. 2020] [Wei et al. 2020].

O vírus se alastrou de maneira veloz e em 38 dias a Organização Mundial da Saúde (OMS) declarou emergência de saúde pública a nível internacional, quando a infecção já havia alcançado mais de 100 países [OMS 2020]. No dia 5 de maio de 2020, 215 países, áreas ou territórios já haviam registrado ao menos um caso do novo vírus, e a curva só tendia a crescer em diversos países [OMS 2020].

A previsibilidade do espalhamento do vírus é complexa e irregular. Devido ao fato de que variações registradas na literatura sobre a disseminação do vírus de pessoa para pessoa é disforme e inconclusiva. Pan et al. (2020) registrou que no mês de janeiro de 2020 a taxa de espalhamento por milhão na China variou entre 2,0 até 162,6, e [Najafimehr et al. 2020] também sobre a China declarou que nos dois primeiros meses de 2020 a taxa de dispersão foi de 1,4 a 7,23 ambos estudos publicados em abril. Na Itália os níveis médios calculados demonstram cerca de 4,2 [Najafimehr et al. 2020] e os Estados Unidos demonstraram uma variação menor que 5 até aproximadamente 17 segundo [Korolev 2020] publicado em julho de 2020.

A velocidade da propagação se deve ao fato da facilidade de contágio e infecção. Que é transmitido principalmente através das vias aéreas, em aspectos similares aos do vírus da Influenza [Wu et al. 2020]. Também é transmissível pelo contato de superfícies infectadas, já que a sobrevivência do vírus é de até 72 horas em locais como metais ou plásticos [van Doremalen et al. 2020]. Não raro, o descuido de encostar em superfícies não limpas e posteriormente tocar os olhos, boca ou nariz é nocivo e preocupante [OMS 2020] [Minocha et al. 2020].

A propagação é o combustível para a eclosão da periculosidade da nova ameaça. As superlotações dos hospitais devido a necessidade de hospitalizações, sejam de leitos ou Unidades de Terapia Intensiva (UTI), impossibilitam que a rede de saúde receba novos pacientes, vítimas de diversas outras enfermidades [Remuzzi e Remuzzi 2020].

Na intenção de controlar a velocidade do avassalador espalhamento do Sars-Cov-2, medidas de contenção foram adotadas, desde alto rigor com os níveis de higiene e limpeza das pessoas, até mesmo com o ambiente, até medidas restritivas de mobilidade como quarentenas e isolamentos.

Considerando que é uma infecção que majoritariamente atinge o sistema respiratório, não apenas o contato físico pode ser transmissor, como também gotículas expelidas através da fala, respiração, espirros ou tosse. Sujeitos a essas condições, o *Center for Disease Control and Prevention* (CDC) recomendou que quando não for possível permanecer em casa, evitar contato próximo, mantendo sempre distância, no contínuo intuito de prevenir a contaminação em massa [CDC 2020].

Todas as intenções são meios de tentar conter um avanço desenfreado de um vírus para o qual, até o presente momento, ainda não há vacina nem uma medicação cientificamente comprovada e eficaz. Causador de uma doença chamada Covid-19 o novo coronavírus não está restrito apenas a danos respiratórios ou pulmonares, revelando-se uma enfermidade multissistêmica [Singhal 2020].

O comportamento do vírus é variável de organismo a organismo, e ainda não se tem qualquer precisão na determinação de seus efeitos no corpo de um determinado hospedeiro. Desde pacientes que não apresentam sintomas, como os assintomáticos, dos quais não há um consenso de quantos casos são, mas acredita-se que em grande proporção [Nishiura et al. 2020] [Casella et al. 2020] [Harries, Martinez e Chakaya 2020], aqueles com sintomas sutis desde a ausência de sabor, e olfato [Hjelmesæth e Skaare 2020], moderados como: febre, tosse, fadiga, diarreia, mialgia, náuseas [Zu et al. 2020], até os casos graves que sofrem danos sérios aos pulmões, com falta de ar, saturação de oxigênio, trombose venosa que possibilitam acidentes vasculares cerebrais (AVCs), embolia pulmonar e pneumonia viral [Tian et al. 2020][Danzi et al. 2020] [Wang et al. 2020] [Filatov et al. 2020].

Ainda que afete diversas partes do corpo humano e não exista uma certeza do quão grave podem ser a infecção ao infectado, há um grupo com maior vulnerabilidade, segundo Onder, Rezza e Brusaferro (2020) ou Wu e McGoogan (2020), que são pessoas com: diabetes, hipertensão, câncer, doenças respiratórias e idosos.

2.1. Covid-19 no Brasil

O primeiro caso registrado no Brasil foi um homem de 61 anos, que realizou uma viagem para a região da Lombardia, Itália, do dia 9 até 20 de fevereiro de 2020, onde encaminhava-se um surto em potencial.

Um estudo realizado pelo Núcleo de Operações e Inteligência em Saúde do Brasil apontou qual seria a real condição dos casos de subnotificação da Covid-19 no Brasil. Os resultados demonstraram que as notificações até o dado momento eram de apenas 8% dos casos positivados e que a taxa de letalidade, razão entre mortes e casos totais, seria de aproximadamente 5,4% [Prado, Bastos e Batista 2020].

Contudo, os 19638 casos confirmados no Brasil, na data de 10 de abril de 2020, representariam uma quantidade 12 vezes menor que a real, segundo Prado, Bastos e Batista (2020). Toda base de cálculo foi estipulada através da taxa de mortalidade observada, que indicou ser 16,3%. Sendo assim, a equiparação com a taxa base, ou esperada, que seria de 1,3% demonstrou o possível volume não notificado de casos positivados ao vírus.

No estudo mencionado [Prado, Bastos e Batista 2020], observou-se também que as situações sobre a taxa de notificação dos estados eram discrepantes e muito dispersas (Figura 1) devido a diferente evolução dos casos. Sendo assim a metodologia aplicada pode gerar distorções devido à série histórica, especialmente nas localidades com poucas notificações.

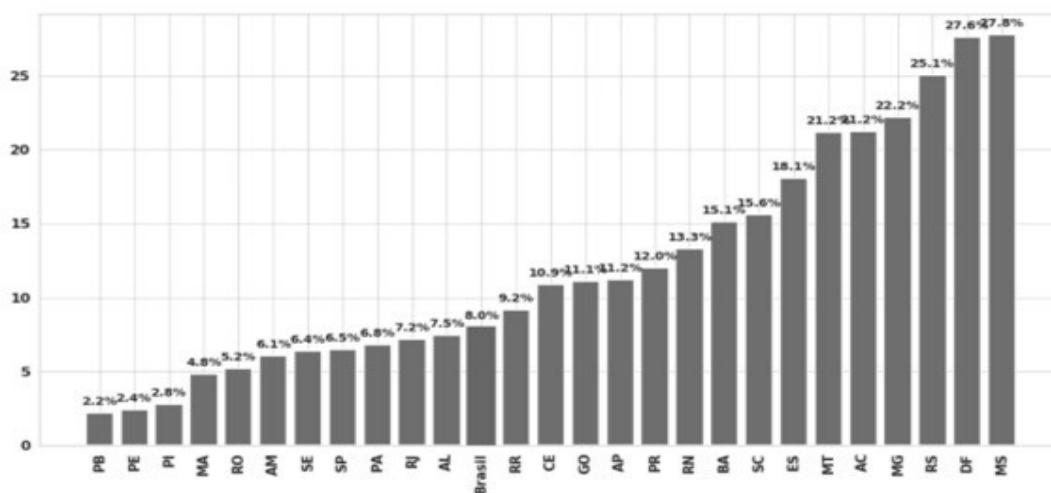


Figura 1: Representação da taxa de notificação do Sars-Cov-2 nos estados brasileiros [Prado, Bastos e Batista 2020].

3. Uso de algoritmos de *Machine Learning* (ML) para previsões no contexto do Covid-19

Uma pesquisa realizada na China, da qual no momento da revisão bibliográfica estava em *preprint*, com dados registrados durante o período de 10 de janeiro de 2020 utilizou-se o prontuário de pacientes da região de Wuhan, mais especificamente do hospital de Tongji, na intenção de obter um modelo de *machine learning* para previsão de óbitos e

identificação do quadro clínico esperado para um paciente, assim tentando antecipar suas necessidades de atendimento [Yan et al. 2020].

Dos dados coletados foram descartadas informações relacionadas a pacientes que se encaixavam nas seguintes condições: gravidez, amamentação, pacientes menores de 18 anos e casos em que 20% ou mais dos dados completos estavam ausentes. Dos 404 pacientes restantes, 191 faleceram e 213 recuperaram-se [Yan et al. 2020].

A separação dos casos caracterizou, dos 404, 375 para o desenvolvimento do algoritmo e 29 para validação. Dos 375 foram configuradas 3 divisões: geral, severa e crítica, das quais representavam 197, 27 e 151 pacientes, respectivamente. Vale ressaltar que todas as divisões obtinham dados de casos sobreviventes e óbitos [Yan et al. 2020].

Na elaboração do modelo, foi realizada a separação com o modelo de *train and test* com uma proporção de 70% para 30% com uso de validação cruzada de 5 vezes. Além de 84 *features* que foram aplicadas inicialmente no algoritmo de árvore de decisão, XGBoost [Yan et al. 2020].

O modelo de otimização a partir de uma seleção de atributos ranqueados por importância pelo XGBoost, definiram 3 *features* como centrais, sendo elas: Lactato desidrogenase (*LDH*), linfócitos e a ultra sensibilidade à proteína C-reativa (*hs-CRP*). Baseado nos dados fornecidos conseguiu-se estabelecer uma acurácia de 96% utilizando do modelo *single-tree* e 97% no modelo *multi-tree* [Yan et al. 2020]

No projeto a performance de validação demonstrou uma acurácia de 100% sobre os óbitos e 90% aos sobreviventes conforme é apresentado na figura 2. A imagem corresponde a duas tabelas de decisão, que demonstram os casos de previsão realizada pelo modelo e os dados reais, indicados pelas legendas *True Level* e *Predicted Level* que correspondem a construção e guia de valores para verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos. Sendo assim a da esquerda corresponde à validação com 29 casos, onde acerta todos os casos de óbitos, e erra apenas duas previsões de sobrevivência que vieram a óbito, e a direita corresponde ao modelo de treino e teste [Yan et al. 2020].

True Level	Predicted Level	
	Survival	Death
Survival	10	2
Death	0	17

True Level	Predicted Level	
	Survival	Death
Survival	192	9
Death	4	170

Figura 2: Tabelas de decisão correspondendo aos resultados com o algoritmo XGBoost, a Tabela da esquerda corresponde a validação e a direita os dados de treino e teste utilizados [Yan et al. 2020].

Uma pesquisa realizada por Burdick et al. (2020) visou elaborar um modelo de *machine learning* para prever o agravamento dos quadros clínicos, que necessitariam do uso de ventiladores dentro das próximas 24 horas após darem entrada no hospital.

O prognóstico teve o escore medido e comparado com o *Modified Early Warning Score (MEWS)* ou Escore de alerta precoce (Burdick et al. 2020, tradução nossa) para definir qual seria a potencial evolução da doença no organismo do paciente.

Utilizando dados dos pacientes positivos para a Covid-19, coletados entre o dia 24 de março de 2020 até o dia 04 de maio de 2020, 197 foram considerados apropriados para pesquisa ser realizada. As informações obtidas dos prontuários eletrônicos dos pacientes que deram entrada no pronto socorro, sendo estas, coletadas nas primeiras duas horas após admissão no hospital. Ao todo foram 12 informações que compuseram a série de dados para cada instância, sendo elas: Pressão arterial diastólica e sistólica, frequência cardíaca e respiratória, saturação de oxigênio, leucócitos, plaquetas, lactato, nitrogênio ureico no sangue, creatinina e bilirrubina. Todas as informações demográficas foram descartadas [Burdick et al. 2020].

Dos 197 pacientes apenas 10 necessitaram de ventilação mecânica dentro das 24 horas posteriores a admissão no hospital. O modelo, que similarmente foi construído com o algoritmo XGBoost, obteve uma precisão de 16% a mais que o modelo padrão de análise *MEWS*. Melhorando a precisão para ambas avaliações, seja de intubação ou alta, e reduzindo os falsos negativos [Burdick et al. 2020].

Liang et al. (2020) também obteve bons resultados com a um modelo de *machine learning* para prever a possibilidade do surgimento de doenças graves em pacientes da Covid-19. O estudo utilizou de um coorte com 1590 pacientes para o desenvolvimento e 710 para validação.

Construiu-se um escore de risco para o desenvolvimento de doenças graves ou críticas nos infectados. Para a elaboração deste utilizou-se de 10 preditores independentes dos quais foram inclusos: anormalidade na radiografia torácica, idade, dispneia, hemoptise, insciência, quantidade de comorbidades, histórico de câncer, proporção de neutrófilos por linfócitos, lactado desidrogenase e bilirrubina direta [Liang et al. 2020].

Além da estrutura de *machine learning*, o projeto também avaliou em algumas das características mais envolvidas nos casos da Covid-19 em termos estatísticos [Liang et al. 2020].

O modelo utilizou-se de 19 *features*, das quais foram as principais apontadas pela regressão *LASSO*, aplicada para filtrar quais as mais importantes variáveis dentre as 72 aplicadas inicialmente. Com as 19 *features* definidas, aplicou-se o algoritmo de regressão Logística com o método de *Cross Validation* com *k-fold* igual a 10. Definições das quais resultaram em uma precisão de 88% sobre a série de dados de validação. A aplicação também demonstrou que 10 variáveis por si só são estatisticamente significantes para prever doenças graves nos enfermos da Covid-19, das quais foram descritas anteriormente como preditores independentes [Liang et al. 2020].

4. Coleta de dados

A coleta de dados para este projeto foi realizada com base nos casos positivos para o novo coronavírus no Brasil. Durante busca por dados disponibilizados por organizações governamentais, observou-se inicialmente a falta de preparo, a má documentação e registro da evolução da doença em território nacional.

O governo federal, ministério da saúde, apresentou quantidades totais e sobre todos os casos positivados e os óbitos, com a divisão por estados. As secretarias de saúde dos estados conduziram de modo mais assertivo as séries de dados referentes aos casos e as condições vividas pelos pacientes. Sendo dessas secretarias estaduais que foram obtidos, comparados e utilizados os dados.

4.1 Dados estatísticos

A coleta de dados para a elaboração das análises estatísticas tem como mandatório a contenção de informações a respeito dos casos positivos, hospitalizados e todos os casos de óbito dos quais ocorreram no respectivo estado.

Na intenção de reconhecer as condições dos infectados, alguns aspectos foram levados em consideração, por ordem de relevância. Sendo eles: sexo, idade, doenças preexistentes, sintomas e etnia, para todas as 3 condições anteriores.

Dos 26 estados brasileiros mais o distrito federal, 21 unidades federativas não apresentaram as séries de dados desejadas para a elaboração deste projeto.

Os 7 estados selecionados foram: Acre, Alagoas, Ceará, Espírito Santo, Mato Grosso do Sul, Santa Catarina e Rondônia, dos quais foram capturados dados até o dia 31 de maio de 2020. Estados como o Mato Grosso do Sul e o Ceará não possuem informações acerca de etnia, e o estado do Ceará não registra quais os sintomas o enfermo sentiu.

4.2 Dataset para *Machine Learning* (ML)

Para a elaboração do modelo de *machine learning*, o dataset escolhido foi do estado do Espírito Santo. Este apresentou uma série de dados bem estruturada com informações relevantes como as descritas anteriormente, e bem detalhadas, com facilidade de uso pois as *features* foram organizadas de modo classificatório.

A escolha do *dataset* não diz respeito apenas por ter informações bem ordenadas, porém, mostrou-se uma organização, e constância da atualização dos dados, sendo desde o princípio dos casos bem descritos e relatados. A série de dados escolhida para integrar este projeto, corresponde até a data de: 19 de junho de 2020.

5. Modelo de *Machine Learning* e Weka

A elaboração do modelo de *machine learning* (ML) se deu utilizando a ferramenta *Waikato environment for knowledge analysis* (Weka). Weka é um software disponibilizado em Java, com o intuito de fornecer uma ferramenta com capacidade de executar algoritmos de machine learning de modo mais simples. Hoje não se limita apenas aos usos de *ML* padrão, como possui algoritmos que podem implementar *Deep Learning* e *data mining* [Weka 2020].

A finalidade da utilização do Weka foi a de elaborar um modelo capaz de efetuar a previsão de óbitos de acordo com o *dataset* do estado do Espírito Santo e as seguintes *features*: Idade, Sexo, Etnia, Febre, Dispneia, Tosse, Coriza, Dor de Garganta, Diarreia, Dor de cabeça, Doenças respiratórias, Doenças cardíacas, Doenças renais, Diabetes, Obesidade, Fumante, Óbito (classe – criada através da coluna evolução).

A seleção dos atributos no *dataset* baseou-se em características com propriedades de origem biológica, sintomatológica ou de caráter relativo à saúde do indivíduo. Com esses critérios, dados como o bairro, cidade, escolaridade ou data de notificação foram descartados. Todos os casos foram utilizados como atributos classificatórios, inclusive a idade, da qual foi registrada utilizando-se de intervalos de aproximadamente 10 anos.

Todas os campos correspondentes as séries de dados vazias que dizem respeito às doenças e sintomas foram considerados negativos. Sendo assim, qualquer valor que não fosse “sim”, tornou-se “não”. O mesmo se fez para a classe de óbito, da qual foi extraída da coluna evolução.

Ao todo foram contabilizados 33.232 casos positivos até o dia 19 de junho de 2020, no estado do Espírito Santo, dos quais, 31.965 não foram registrados como óbitos até aquela data e 1267 casos que evoluíram a óbito.

Alguns algoritmos de classificação foram selecionados na tentativa de obter o melhor resultado na previsão dos falecimentos. Os algoritmos de machine learning implementados foram os seguintes: *Logistic* (Regressão Logística); *Naïve Bayes*; *Simple Logistic* (Regressão Logística Simples); *SVM* (*Support Vector Machine* – Máquina de Vetor de Suporte); *SPEGASOS*; *REPTree* (Árvore de decisão rápida); *Decision Table* (Tabela de Decisão).

Como a classe de previsão encontrava-se desbalanceada, gerando sempre votos para sobreviventes, alguns filtros foram aplicados antes da execução dos algoritmos, no intuito de alcançar melhores previsões.

No intuito de definir tamanhos ou pesos similares, os filtros aplicados no conjunto de dados são caracterizados pela plataforma como filtros supervisionados de instância. Destes utilizou-se: *Class Balancer* (Balanceador de classes) e SMOTE;

Todas as execuções implementadas fizeram uso da validação cruzada ou *Cross-Validation* com *k-fold* estabelecido sendo igual à 10, visando obter uma acurácia ou *F-Measure* maior. A escolha foi baseada no fato de que o modelo de treino não divide estaticamente o *dataset* como geralmente ocorre com o modelo de treino como *Train and Test*, onde ocorre uma divisão estática do conjunto de dados de acordo com a escolha, seja ela, 80% - 20%, ou 70% - 30%.

Além de todas as implementações com os sete algoritmos sem filtros e com a aplicação dos dois filtros, também se utilizou, uma matriz de custo no intuito de pressionar o modelo a não efetuar escolhas incorretas. Contudo, esta não se apresentou como grande ponto de vantagem quando aplicado junto dos filtros *Class Balancer* e SMOTE, demonstrando melhoras apenas em casos onde não houve aplicação de filtros.

6. Resultados

Esta seção apresenta os resultados alcançados no que dizem respeito às análises estatísticas relacionadas a pacientes com Covid-19 em condições graves e em óbitos, nos sete estados brasileiros, que disponibilizaram dados até o dia 31 de maio de 2020 e que se encaixaram nos requisitos desejados. Também são apresentados os resultados alcançados pelo modelo de previsão de óbitos construído sobre o *dataset* do Espírito Santo, com dados até o dia 19 de junho de 2020.

6.1 Estatísticas

Nas análises e verificações e sobre a situação brasileira no que se refere ao Sars-Cov-2, considerou-se os 7 estados avaliados, até a data de 31 de maio de 2020. Nos dados capturados registrou-se algumas condições que são correlatas com os casos graves e a letais do novo coronavírus, e com isso foram formados grupos correspondentes. Os grupos formados foram separados pela condição. Casos de óbito não foram considerados como de hospitalização.

As condições biológicas iniciais sobre os casos graves do novo coronavírus estão expostos na Tabela 1, onde são representados, por sexo, a quantidade e o percentual em relação aos casos graves e óbitos.

Tabela 1. Internações e óbitos por sexo

	Total	Quantidades		Percentual	
		Feminino	Masculino	Feminino	Masculino
Hospitalizados	7678	3266	4410	42,5%	57,5%
Óbitos	4650	1963	2687	42,2%	57,8%

Proporcionalmente, as quantidades são quase idênticas para os óbitos e os casos graves, tendo como maior alvo o sexo masculino.

As relações existentes entre idade e evolução da doença a níveis críticos, são apresentadas na Tabela 2, com total de casos de 7678 hospitalizados e 4650 óbitos.

Tabela 2. Internações e óbitos por idade

	Hospitalizações		Óbitos	
	Quantidade	Percentual	Quantidade	Percentual
0-19	153	2%	25	1%
20-29	299	4%	56	1%
30-39	751	10%	151	3%
40-49	1051	14%	319	7%
50-59	1394	18%	587	13%
60-69	1484	19%	898	19%
70-79	1361	18%	1115	24%
>80	1174	15%	1276	27%
Não identif.	11	0%	223	5%

A senioridade se apresenta como um agravante em potencial. Conforme apresentado (Tabela 2), 51% (24% entre 70-79 somados a 27% acima de 80) das vítimas fatais tinham idade superior a 70 anos, e 70% do total de óbitos, validados, possuíam acima de 60 anos. Os casos não identificados correspondem a casos colhidos pelas secretarias de saúde sem registro de idade.

Em contraste com o caso dos óbitos, não existe grande disparidade. A maior representação de idosos é um efeito da pirâmide etária brasileira e não reflete uma característica do vírus.

No que diz respeito a sintomatologia dos casos avaliados, as quantidades são expressas na Tabela 3. São demonstrados em números absolutos quais as quantidades contabilizadas de cada sintoma. Porém, corresponde ao total de 2938 para hospitalizados e 1458 para óbitos devido ao não registro destas condições pelo estado do Ceará.

Tabela 3. Contabilização dos casos e sintomas para hospitalizados e óbitos

	Hospitalizações	Óbitos
Tosse	1957	822
Dispneia/Falta de ar	1490	817
Febre	1926	792
Dor de garganta	514	154
Outros	306	137
Dor de cabeça	480	126
Coriza	289	115
Diarreia	294	74

Ambas condições, internações e falecimentos, tiveram como principal sintoma a tosse, e o segundo mais comum foi febre e dispneia respectivamente. Na declaração de outros, foram considerados, náuseas, vômitos e a classificação determinada pelas secretarias em específico para classe outros.

A descrição sobre quantidades de casos que dizem respeito a doenças, é expressa na Tabela 4, onde são avaliados todos os 7 estados. Através dela pode-se ver qual a doença mais comum nos casos de hospitalizações e óbitos causados pela Covid-19.

Tabela 4. Doenças em hospitalizados e óbitos

	Hospitalizações	Óbitos
Doenças cardíacas	2327	1480
Diabetes	1833	1005
Doenças respiratórias	417	221
Doenças renais	305	159
Doenças imunológicas	208	111
Doenças neurológicas	200	108
Obesidade	261	105
Outras doenças	237	125

Nos casos da Tabela 4, nota-se que doenças cardíacas são as mais presentes quando analisadas as quantidades sobre pacientes que utilizaram ou deveriam ter tido socorro de apoio médico ambulatorial, sendo a diabetes a segunda colocada.

Durante a contabilização dos dados, alguns grupos foram formados, como os seguintes:

- Doenças cardíacas: Todo tipo de enfermidade cardiovascular, ou que envolva coração, incluiu-se hipertensão arterial;

- Doenças respiratórias: Todos os tipos de pneumopatias e asma;
- Doenças imunológicas: Portadores de doenças imunodepressoras, pacientes com imunidade frágil e doenças autoimunes;
- Doenças neurológicas: Incluem Alzheimer e acidentes vasculares cerebrais;
- Outras: Doenças cromossômicas, tabagismo, hematológicas e hepáticas.

Não foram registradas relações entre a etnia e os casos graves da Covid-19, sendo assim os resultados não são expostos.

6.2 Machine Learning (ML)

Esta seção é destinada a apresentar os melhores resultados obtidos com as aplicações de todos os algoritmos e seus filtros, incluindo as matrizes de custo. São representados os melhores resultados baseados na acurácia, ou *F-Measure*.

6.2.1 Melhor resultado

Nas implementações dos modelos de *machine learning* sobre o *dataset* dos casos positivos para o novo coronavírus no estado do Espírito Santo, o melhor resultado foi obtido utilizando-se o algoritmo de árvore de decisão rápida, *REPTree*, junto da aplicação do filtro SMOTE, sem uso da matriz de custo. Conforme ilustra a Tabela 5, a acurácia, ou *F-Measure*, que é a razão da soma dos acertos totais pelo total de números.

O *F-Measure* obteve um valor considerável, com quase 100% de relevância, tanto para os classificadores positivos quanto aos negativos. Além dos altos resultados obtidos para a *F-Measure*, *Precision*, que é a identificação dos positivos sobre a classificação real, *Recall*, é a proporção dos positivos que foram identificados corretamente e *TP rate (True Positives)* obtiveram altos índices, e o *FP rate (False Positives)* baixos.

Tabela 5. Resultado com *RepTree* e SMOTE

TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,965	0,031	0,969	0,965	0,967	0,934	0,987	0,983	“Não”
0,969	0,035	0,965	0,969	0,967	0,934	0,987	0,984	“Sim”
0,967	0,033	0,967	0,967	0,967	0,934	0,987	0,983	AVG.

O motivo do resultado ter sido superior a 95% foi a utilização do filtro SMOTE, que tem por função balancear as quantidades referentes à classe, criando novas instâncias para o modelo, baseado em casas próximas e semelhanças. No caso aplicado, o filtro foi utilizado até a igualdade das categorias classificatórias, conforme mostra a Tabela 6, correspondente à matriz de confusão resultante da aplicação, que aponta a equidade entre a quantidade numérica sobre os que sobreviveram e os que infelizmente, não.

Tabela 6. Matriz de confusão de REPTree e SMOTE

Expectativa		Classificação	Real
“Não”	“Sim”		
30847	1118	“Não”	
1001	30967	“Sim”	

Nota-se que ambas possibilidades de classificação, seja “Não” ou “Sim”, possuem agora a mesma quantidade de elementos, ou seja, 31965 instâncias para sobreviventes (“Não”) e mesmo valor para falecidos (“Sim”).

6.2.2 Resultado somente com matriz de custo

O mesmo conjunto de dados sem a aplicação de filtros, atingiu uma baixa acurácia, conforme dados mostrados na Tabela 7, já que a quantidade de comparadores de resultado negativos aplicada pelo algoritmo é consideravelmente menor que a positiva.

Contudo, o algoritmo com melhores demonstrações no caso descrito, diferentemente do primeiro caso demonstrado, foi o de regressão logística simples (*Simple Logistic*). Além disso, a utilização de matrizes de custo aprimorou os resultados, ainda que pouquíssimo, quando não houve filtro algum na expectativa de utilizar-se dos dados da forma mais simples possível.

Tabela 7. Resultado com Simple Logistic e matriz de custo

TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,916	0,277	0,988	0,916	0,951	0,394	0,917	0,996	“Não”
0,723	0,084	0,255	0,723	0,377	0,394	0,917	0,365	“Sim”
0,909	0,270	0,960	0,909	0,929	0,394	0,917	0,972	AVG.

Ainda que o classificador tenha atingido bons resultados referentes a média, o resultado referente ao indicador “Sim” foi baixo, sendo apenas 0,377, ao fim que na intenção de prever o óbito de alguém, este modelo seria apenas para prever casos de sobreviventes e não para casos que padecem.

O maior causador da baixa precisão e acurácia na previsão dos falecimentos é o distanciamento de valores referentes aos montantes da classe. Como apontado anteriormente como um fator decisivo, a Tabela 8 demonstra a diferença das quantidades entre os indicadores da classe.

Tabela 8. Matriz de confusão de Simple Logistic com matriz de custo

Expectativa		Classificação	Real
“Não”	“Sim”		
29283	2682	“Não”	
351	916	“Sim”	

A diferença entre as classificações dos falsos positivos (*false positives*) e os verdadeiros positivos (*true positives*), que na Tabela 8 é descrito como “Não - Sim” ou “Sim-Não” para (*false positives*) e “Sim-Sim” ou “Não-Não” (*true positives*) mostra que

a primeira tem uma porcentagem de erro de 8,39% e a segunda de aproximadamente 27,70%, sendo assim, mais de 25% das previsões foram errôneas.

Ainda que o resultado do modelo da Tabela 7 não tenha atingido resultados considerados relevantes, o uso da matriz de custo influenciou as condições da previsibilidade. Aplicando todos os algoritmos sem filtros e sem matrizes de custo, os resultados obtidos demonstraram-se ainda mais ineficazes no intuito de prever o falecimento.

6.2.3 Resultado sem filtros e sem matriz de custo

A Tabela 9 demonstra os valores atingidos pelo melhor modelo quando se tratou de não fazer uso de balanceadores em geral.

Tabela 9. Resultado com *Logistic* sem filtro e sem matriz de custo

TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,994	0,820	0,968	0,994	0,981	0,293	0,915	0,996	“Não”
0,180	0,006	0,527	0,180	0,268	0,293	0,915	0,368	“Sim”
0,963	0,789	0,951	0,963	0,954	0,293	0,915	0,972	AVG.

O resultado da Tabela 9, alcançado com o algoritmo *Logistic*, foi o melhor obtido sem a utilização de balanceador algum, ou seja, no estado mais puro dos dados, sem interferências bruscas.

A classificação das classes demonstrou uma alta acurácia para “Não”, contudo um baixíssimo valor para “Sim”. Ainda que a média obtida tenha valor alto, não tem real valor para o intuito da elaboração do modelo devida à carência do *F-measure* na condição de “Sim”.

Conforme a Tabela 10, os índices previstos apontados na matriz de confusão, demonstram que aproximadamente apenas 3,13% dos casos sobreviventes foram classificados erroneamente, contudo no que se refere aos óbitos esta marca alcançou o valor de 82%, ou seja, de cada 10 casos cerca de 8 estavam errados.

Tabela 10. Matriz de confusão de *Logistic* sem balanceadores

Expectativa		Classificação	Real
“Não”	“Sim”		
31760	205	“Não”	
1039	228	“Sim”	

Esse fato ocorre devido à previsibilidade do modelo em relação aos casos referentes a classe “Não”. Cerca de apenas 3,81% dos casos padeceram, ou seja, mais de 96% é sobrevivente, enviesando o modelo a majoritariamente classificar os indivíduos como sobreviventes, conforme ilustrou a Tabela 10, onde foram apontados aproximadamente apenas 0,68% de todos os indivíduos como óbitos.

6.2.4 Resultado com *Class Balancer*

Além dos modelos demonstrados, foi aplicado um outro filtro, o *Class Balancer*, que é demonstrado na Tabela 11. Os resultados atingidos são pouco inferiores que os

alcançados com uso do *REPTree* e SMOTE, porém, a estratégia de equilíbrio dos dados é diferente, não se baseando na criação de instâncias, e sim no equilíbrio de pesos com as quantidades inicialmente definidas.

Tabela 11. Resultado com *Logistic* e *Class Balancer*

TP rate	FP rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0,851	0,159	0,842	0,815	0,847	0,692	0,916	0,909	"Não"
0,841	0,149	0,850	0,841	0,845	0,692	0,916	0,908	"Sim"
0,846	0,154	0,846	0,846	0,846	0,692	0,916	0,908	AVG.

Se comparado todo o resultado em relação às médias, a oriunda do *Class Balancer* é o inferior. Contudo, não é a busca somente de atingir um alto valor sobre a média que visa este modelo. Sendo assim, há necessidade de se obter altos classificadores para "Sim", onde este apontou uma precisão de aproximadamente 84,5%.

A matriz de confusão referente ao resultado da Tabela 11 é demonstrada na Tabela 12, onde nota-se a intenção do filtro, que não criou novas instâncias, e sim, equilibrou o peso entre as classes. Sendo assim, não houve uma alteração na quantidade de casos, somente na importância que estes possuem, diferentemente do primeiro filtro demonstrado, SMOTE, do qual pode enviesar o conjunto de dados para obter uma melhor acurácia.

Tabela 12. Matriz de confusão de *Logistic* com *Class Balancer*

Expectativa		Classificação	Real
"Não"	"Sim"		
14144,26	2471,74	"Não"	
2649,12	13966,88	"Sim"	

Dos 80 resultados obtidos, provindos da aplicação dos 7 algoritmos, 3 condições de filtros, e matrizes de custo para cada ocasião, quando passíveis de aplicação, os 4 resultados demonstrados foram os melhores obtidos para cada condição macro, ou seja, sem foco em algoritmo, e sim no resultado obtido pelo *F-measure* da condição "Sim", quando não influenciou gravemente na média. Hierarquizando-os por ordem decrescente de acurácia, temos:

1. *REPTree* com SMOTE e sem matriz de custo;
2. *Logistic* com *class balancer* e sem matriz de custo;
3. *Simple Logistic* sem filtros e com matriz de custo;
4. *Logistic* sem filtros e sem matriz de custo.

Todos os resultados podem ser acessados em: Evidagroup (2020).

7. Conclusão

Ainda que não passível de certezas, pois trata-se de uma doença sem medicação cientificamente eficaz, a Covid-19 deve ser evitada ao máximo. Atualmente não há mais

um alvo preferencial pela doença, ainda que exista um viés, todo cuidado deve ser adotado.

Buscando tabular e registrar algumas condições sobre as características mais presentes em pacientes que necessitaram de atendimento ambulatorial ou evoluíram a óbito, dados capturados até o dia 31 de maio de 2020 condizentes a sete estados brasileiros identificam que a idade foi classificada como um ponto importante, demonstrando que cerca de 70% dos óbitos eram pessoas acima dos 60 anos de idade. Tosse foi registrado como sintoma mais comum, e não foi observada nenhuma correlação nos dados entre a gravidade do quadro clínico e a etnia.

Doenças também foram analisadas e as cardíacas foram identificadas como as mais presentes em casos graves, seguidas de diabetes, que por maior que possam ser as relações são as doenças mais comuns nos brasileiros não permitindo uma conclusão precisa sobre qual doença é mais agravante em pacientes do Covid-19.

Além das condições estatísticas nacionais, foi elaborado um modelo onde seja possível prever os óbitos em um estado brasileiro que foi selecionado por dispor dos dados referentes a idade, sexo, doenças preexistentes e sintomas de todos os casos positivos, hospitalizados e óbitos. A previsibilidade se apresentou satisfatória uma vez que atende uma acurácia média e para os casos positivos, maior que 80% considerando o cenário utilizado, com *dataset* de casos do Espírito Santo, com registros até o dia 19 de junho de 2020, e com a ferramenta *Weka*.

Embora existam incertezas nas previsões do modelo, a capacidade de antever quadros clínicos severos e conseqüentemente dedicar aos pacientes portadores de tais quadros os recursos necessários, não pode ser subestimada.

A utilização do modelo de árvore de decisão, o *REPTree*, demonstrou-se um forte aliado na classificação de fatores para a previsão de falecimentos em decorrência do novo coronavírus. Não obstante, a interferência causada pelos algoritmos de balanceamento sobre a árvore de decisão, potencializaram os resultados de maneira decisiva, através do filtro SMOTE.

Sendo assim, valida-se que a acurácia pode ter grandes variações quando aplicados balanceadores, no intuito de remover desbalanceamentos de conjuntos de dados, favorecendo, assim, o resultado alcançado, removendo vieses e projetando um modelo mais preciso.

8. Interface

Esta seção apresenta partes da interface construída, em formato de website no idioma inglês, para expor os resultados obtidos nesta pesquisa, tanto estatísticos quanto os referentes ao *machine learning*.

Na figura 3 é apresentado o layout inicial do website que descreve brevemente a doença, e algumas informações sobre o projeto. Além disso são expostos 3 gráficos para melhor visualização dos casos no mundo.



Figura 3. Página inicial do website criado para expor os resultados [Autor, 2020].

Na aba *Statistics* (figura 4) são apresentados os estados utilizados para a pesquisa estatística, do mesmo modo, todos os resultados obtidos das análises sobre os óbitos e hospitalizados pela Covid-19 são expostos graficamente.

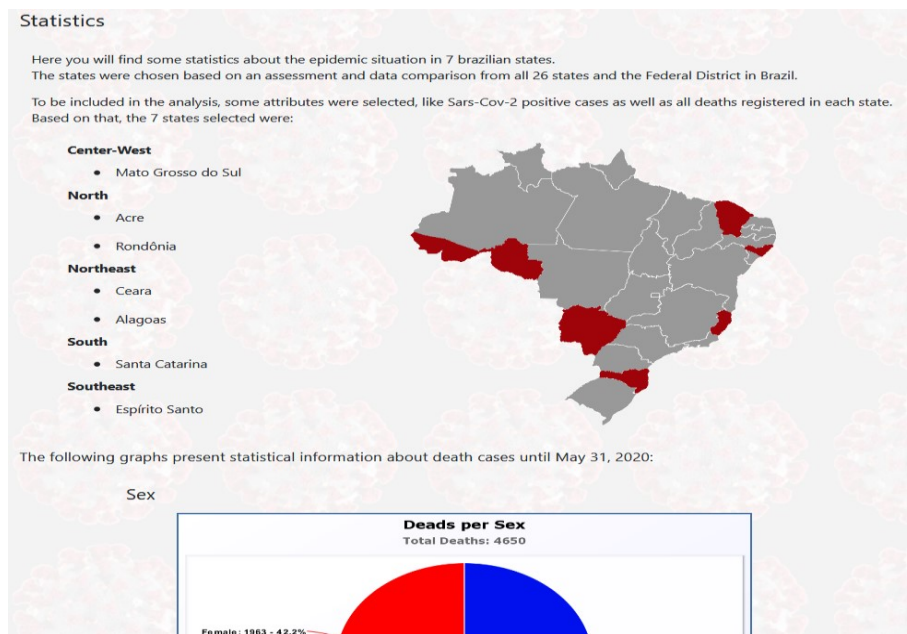


Figura 4. Pagina *Statistics* do website, onde são ilustrados os resultados estatísticos [Autor, 2020].

A aba *Machine Learning in Deaths*, descreve quais foram as *features* utilizadas, quais os intervalos de tempo, e qual *dataset* (Espírito Santo). Também são descritos os

métodos de aplicação com filtros, algoritmos, matrizes de custo e seus valores e por fim uma seleção de atributos para escolha.

A seleção apresenta 5 campos para escolha sendo eles, data, algoritmo, filtro, matriz de custo e tipo de apresentação, seja ela em modo resumido de sumário ou avançado. Após escolher as características desejadas, ao pressionar o botão *Select*, será exibido, no campo abaixo, o resultado obtido através da ferramenta WEKA com as informações desejadas conforme é exposto na figura 5.

Além do resultado exposto há opção de efetuar o *Download* deste resultado, que será executado com a permissão do usuário na máquina de uso em formato de texto.

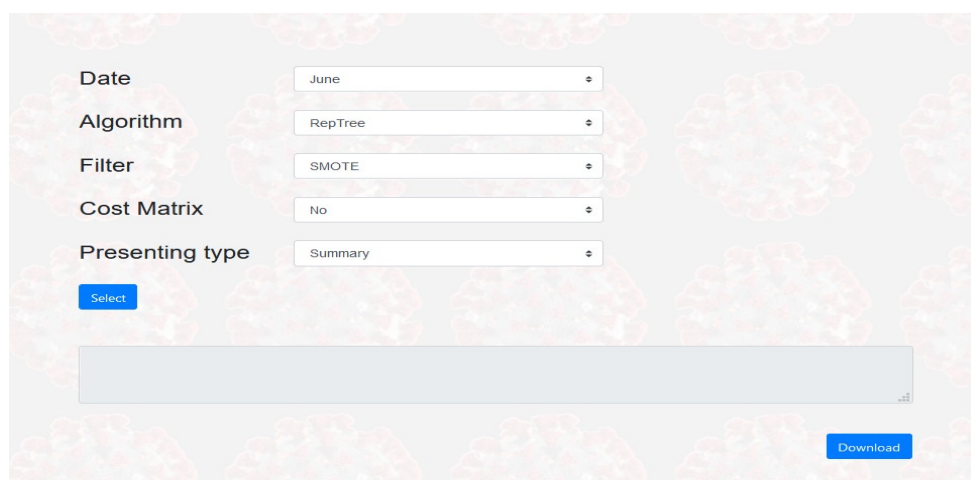


Figura 5. Página *Machine Learning in Deaths* do website, onde se apresenta os atributos que podem ser selecionados e capturar o resultado desejado [Autor, 2020].

9. Referências

- Burdick et al. (2020) “Prediction of respiratory decompensation in Covid-19 patients using machine learning: The READY trial”. *Computer in Biology and Medicine*, 124. Doi: 10.1016/j.combiomed.2020.103949
- Cascella, M. et al. (August 10, 2020). “Features, Evaluation, and Treatment of Coronavirus (COVID-19)” In: StatPearls. Treasure Island (FL): StatPearls Publishing. Disponível em: <https://www.ncbi.nlm.nih.gov/books/NBK554776/>
- Centers for Disease Control and Prevention (2020) “Coronavirus Disease 2019 (COVID-19)”. Disponível em: <https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/social-distancing.html>, May.
- Evidagroup (2020) “Covid-19’s evolution in Brazil: Statistics from 7 states, including Espírito Santo”. Disponível em: <http://covid.evidagroup.es/>, July.
- Filatov A. et al. (March 21, 2020) “Neurological Complications of Coronavirus Disease (COVID-19): Encephalopathy”. *Cureus*, 12(3), e7352. doi:10.7759/cureus.7352
- Danzi, G. B. et al. (March 30, 2020) “Acute pulmonary embolism and COVID-19 pneumonia: a random association?”, *European Heart Journal*, 41(19), 1858. doi.org/10.1093/eurheartj/ehaa254

- Harries, A. D., Martinez, L., e Chakaya, J. M. (2020) “Monitoring the COVID-19 pandemic in sub-Saharan Africa: focusing on health facility admissions and deaths”. *The International Journal of Tuberculosis and Lung Disease*, 24(5), 550–552. doi:10.5588/ijtld.20.0176
- Hjelmesæth, J, e Skaare, D. (May 5, 2020) “Loss of smell or taste as the only symptom of COVID-19”. *Tidsskrift for den Norske Laegeforening*, 140(7). doi: 10.4045/tidsskr.20.0287.
- Hu, V. H. et al. (2020) “Protecting yourself and your patients from COVID-19 in eye care”. *Community eye health journal*, 33(108), S1-S6. <https://cehjournal.org/wp-content/uploads/COVID-19-1.pdf>, September.
- Huang, C., et al. (February 15, 2020) “Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China”. *The Lancet*, 395(10223), 497-506. doi:10.1016/s0140-6736(20)30183-5
- Korolev I. (2020). “Identification and estimation of the SEIRD epidemic model for COVID-19”. *Journal of econometrics*. doi: 10.1016/j.jeconom.2020.07.038
- Liang et al. (May 12, 2020) “Development and validation of a clinical risk score to predict the occurrence of critical illness in hospitalized patients with Covid-19”. *Journal of the American Medical Association*, 180(8), 1081-1089. doi:10.1001/jamainternmed.2020.2033
- Minocha, A. et al. (April 7, 2020). “Survey of ophthalmology practitioners in A&E on current COVID-19 guidance at three Major UK Eye Hospitals”. *Eye*, 34, 1243-1245. <https://doi.org/10.1038/s41433-020-0857-5>.
- Najafimehr, H., et al. (August, 2020). “Estimation of basic reproduction number for COVID-19 and the reasons for its differences”. *International Journal of Clinical Practice*, 74(8), e13518. doi:10.1111/ijcp.13518.
- Nishiura, H., et al. (2020). “Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19)”. *International Journal of Infectious Diseases*, 94, 154-155. doi: 10.1016/j.ijid.2020.03.020.
- Onder, G., Rezza, G., e Brusaferro, S. (2020). “Case-Fatality Rate and Characteristics of Patients Dying in Relation to COVID-19 in Italy”. *Journal of the American Medical Association*, 323(18), 1775-1776. doi:10.1001/jama.2020.4683
- Organização Mundial da Saúde (2020) “Q&A: Influenza and COVID-19 - similarities and differences”. Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/question-and-answers-hub/q-a-detail/q-a-similarities-and-differences-covid-19-and-influenza>, September.
- Organização Mundial da Saúde (2020). “Coronavirus disease 2019 (COVID-19) Situation Report - 48”. Disponível em: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200308-sitrep-48-covid-19.pdf?sfvrsn=16f7ccef_4, September.
- Organização Mundial da Saúde (2020). “Coronavirus disease 2019 (COVID-19) Situation Report - 52”. Disponível em: https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200312-sitrep-52-covid-19.pdf?sfvrsn=16f7ccef_4, September.

- source/coronaviruse/situation-reports/20200312-sitrep-52-covid-19.pdf?sfvrsn=e2bfc9c0_4, September.
- Organização Mundial da Saúde (2020). “Numbers at a glance”. Disponível em: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>, September.
- Pan, A., et al. (2020) “Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China”. *Journal of the American Medical Association*, 323(19), 1915-1923. doi:10.1001/jama.2020.6130.
- Prado, M., Bastos, L. e Batista, A. (2020). “Análise de subnotificação do número de casos confirmados da COVID-19 no Brasil. Núcleo de Operações e Inteligência em Saúde (NOIS)”. Disponível em: http://www.supersuporte.com/myRpubs/NT7_Subnotificacao_notDia11-abr-2020.pdf, July.
- Remuzzi, A. e Remuzzi, G. (2020) “COVID-19 and Italy: what next?”. *The Lancet*, 395(10231), 1225-1228. doi:10.1016/S0140-6736(20)30627-9.
- Singhal, T. (2020). “A Review of Coronavirus Disease-2019 (COVID-19)”. *The Indian Journal of Pediatrics*, 87, 281–286. doi:10.1007/s12098-020-03263-6.
- Tian, S. et al. (2020). “Characteristics of COVID-19 infection in Beijing”. *Journal of Infection*, 80(4), 401-406. doi: 10.1016/j.jinf.2020.02.018
- van Doremalen, N. et al. (2020). “Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1”. *The New England journal of medicine*, 382(16), 1564–1567. doi:10.1056/NEJMc2004973.
- Wang, T., et al. (2020). “Attention should be paid to venous thromboembolism prophylaxis in the management of COVID-19”. *The Lancet Haematology*, 7(5), 362-363. doi:10.1016/s2352-3026(20)30109-5.
- Wei, M. et al. (2020). “Novel Coronavirus Infection in Hospitalized Infants Under 1 Year of Age in China”. *JAMA*. 323(13), 1313–1314. doi:10.1001/jama.2020.2131.
- Weka (2020). “The Workbench for machine learning”. Disponível em: <https://www.cs.waikato.ac.nz/ml/weka/>, September.
- Wu, D. et al. (2020). “Positive effects of COVID-19 control measures on influenza prevention”. *International Journal of Infectious Diseases*, 95, 345-346. doi: 10.1016/j.ijid.2020.04.009.
- Wu, P. et al (2020). “Real-time tentative assessment of the epidemiological characteristics of novel coronavirus infections in Wuhan, China, as at 22 January 2020”. *Eurosurveillance*, 25(3). <https://doi.org/10.2807/1560-7917.ES.2020.25.3.2000044>.
- Wu. Z e McGoogan JM. (2020). “Characteristics of and important lessons from the coronavirus disease 2019 (COVID-19) outbreak in China: summary of a report of 72314 cases from the Chinese Center for Disease Control and Prevention”. *Journal of the American Medical Association*, 323(13), 1239-1242. doi: 10.1001/jama.2020.2648.

- Xie, X. et al. (2020). "Chest CT for Typical 2019-nCoV Pneumonia: Relationship to Negative RT-PCR Testing". *Radiological Society of North America*, 296(2), 41-45. doi:10.1148/radiol.2020200343.
- Yan, L. et al. (2020). "Prediction of criticality in patients with severe COVID-19 infection using three clinical features: a machine learning-based prognostic model with clinical data in Wuhan". *MedRxiv*. doi: 10.1101/2020.02.27.20028027v3.
- Zu, Z. et al. (2020). "Coronavirus Disease 2019 (COVID-19): A Perspective from China". *Radiological Society of North America*, 296(2), 15-25. doi: 10.1148/radiol.2020200490.