



# Estudo e aplicação de Redes Neurais para auxiliar na inferência de imagens pulmonares de pacientes com Covid-19

*Study and application of Neural Networks to aid in*

*Covid-19 pulmonary image inference*

Welbe de Souza Gama, Marcella Scoczynski Ribeiro Martins,  
Amanda Mareski Almeida

## RESUMO

A Covid-19 é definida pela Organização Mundial da Saúde como uma doença infecciosa causada pelo vírus SARS-CoV-2, a maioria das pessoas infectadas com o vírus terá uma doença respiratória leve a moderada e se recuperará sem a necessidade de tratamento especial. No entanto, alguns ficarão gravemente doentes e precisarão de atenção médica. Neste contexto, esta pesquisa tem como objetivo colaborar com o enfrentamento da pandemia que se deu nos últimos anos de duas formas: Desenvolvimento de uma rede neural dotada da capacidade de detectar a ação da pneumonia causada pelo coronavírus e levantando dados acerca das medidas de segurança tomadas pelas autoridades nos países da América do Sul. Com este último, seria possível inferir quais as medidas empregadas mais eficazes na mitigação da disseminação do vírus na sociedade. O projeto da rede neural se deu em duas fases, na primeira fase, realizada pelos autores, foi feito o levantamento de bancos de dados propícios a elaboração da rede, sendo no total 2 bancos de dados e a elaboração de um piloto da rede propriamente dita, que obteve uma acurácia de 50,4%. O desenvolvimento do banco de dados acerca do levantamento foi realizado através de pesquisa bibliográfica, o produto final foi um banco com 12 categorias de informações envolvendo tópicos como políticas empregadas e o reflexo no número de casos, mortes e testes.

**Palavras-chave:** Coronavírus, Redes Neurais, Bancos de Dados, Medidas Preventivas.

## ABSTRACT

Covid-19 is defined by the World Health Organization as an infectious disease caused by the SARS-CoV-2 virus, most people infected by the virus will have light to moderate respiratory illness and will be able to recover without the need for any special treatment. However, some will become gravely ill and will have the need for medical attention. In this context, this research aims to collaborate with the fight against the pandemic that has taken place in recent years in two ways: Development of a neural network endowed with the ability to detect the action of pneumonia caused by the coronavirus and raising data on the measures of security taken by authorities in South American countries. With the latter, it would be possible to infer which measures are most effective in mitigating the spread of the virus in society. The neural network project took place in two phases, in the first phase, carried out by the authors, the survey of databases suitable for the elaboration of the network was carried out, with a total of 2 databases and the elaboration of a pilot of the network itself, which obtained an accuracy of 50.4%. The development of the database about the security measures was carried out through bibliographical research. The final product was a database with 12 categories of information involving topics such as policies used and the impact on the number of cases, deaths and tests.

**Keywords:** Coronavirus, Neural Networks, Datasets, Security Measures.

\* Engenharia Elétrica, Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil; [welbegama@alunos.utfpr.edu.br](mailto:welbegama@alunos.utfpr.edu.br)

† Universidade Tecnológica Federal do Paraná, Campus Ponta Grossa; [a.maresski@gmail.com](mailto:a.maresski@gmail.com)

‡ Universidade Tecnológica Federal do Paraná, Ponta Grossa, Paraná, Brasil; [marcella@utfpr.edu.br](mailto:marcella@utfpr.edu.br)



## 1 INTRODUÇÃO

No início do ano de 2020, os primeiros casos de coronavírus (Covid-19) tiveram origem no mercado de frutos do mar da cidade de Wuhan localizada na China. As primeiras ocorrências foram relatadas na virada do ano de 2020 e a incidência aumentou de maneira exponencial nas primeiras semanas.

Segundo Loureiro (2020) pode-se descrever o coronavírus como um vírus RNA capaz de causar infecção em múltiplos sistemas em diversos animais e, em humanos, infecção do trato respiratório. Denominado SARS-CoV-2, este vírus pertence a uma família que já circulava no Brasil antes da pandemia e era responsável por grande parte dos resfriados comuns. Outras espécies foram responsáveis por doenças mais graves, como a Síndrome Aguda Respiratória Severa (SARS-CoV-1) e a Síndrome Respiratória do Oriente Médio (MERS-CoV) que não tiveram casos no Brasil.

Atualmente, de acordo com Ponte (2020) o diagnóstico da doença é realizado através de dois tipos de exames principais: exame de detecção do vírus por PCR (Reação em cadeia da Polimerase) e os testes sorológicos (anticorpos presente no sangue). O primeiro exame RT-PCR é descrito pela OMS como o isolamento e purificação do RNA viral das amostras do trato respiratório superior e inferior, o qual é transcrito reversamente para cDNA e, em seguida, amplificado no Instrumento de PCR em tempo real. Essas amostras podem ser obtidas por intermédio de swabs nasofaríngeos ou orofaríngeos, escarro, aspirados do trato respiratório, lavagem broncoalveolar e lavagem/aspirado nasal ou aspirado nasofaríngeo. O segundo tipo de teste, chamados de exames sorológicos, identifica anticorpos IgM e IgG, possuem vantajoso tempo de resposta (mais rápido), menor carga de trabalho e alto rendimento, uma vez que são menos complexos que os testes moleculares (PCR).

Entretanto só os testes RT-PCR e os testes rápidos (soro), se percebe uma sensibilidade não totalmente satisfatória. Quando a amostra é obtida do trato respiratório superior, falsos negativos podem ocorrer pelo momento inadequado da coleta da amostra em relação ao início da doença e por deficiência na técnica da coleta, no caso do RT-PCR. Estudos realizados pela Unimed mostram que a positividade do teste de RT-PCR foi de cerca de 86% com especificidade de 96%. Já para o caso de testes sorológicos, Santos (2021) em uma matéria publicada no canal de medicina PubMed explica que testes de IgM apresentam em média sensibilidade de 82% com especificidade de 97%, e com o teste IgG apresentando entre 72 e 81% de sensibilidade. Outras literaturas foram consultadas, e os dados referentes a diversos laboratórios indicam que a sensibilidade destes testes podem ir desde 70 a 86%.

Deste modo, levanta-se a dúvida, seria possível o desenvolvimento de uma ferramenta auxiliadora no diagnóstico da doença, que fosse capaz de gerar ainda maior confiabilidade na hora dos diagnósticos? A fim de resolver esta problemática, buscou-se nesta pesquisa utilizar modelos de aprendizado de máquinas (redes neurais convolucionais) no reconhecimento de casos da doença por intermédio de Raios-X de pulmões, diferenciando casos positivos para a doença de casos negativos, procurando agilizar o processo de diagnóstico. Assim, com uma série de testes conversando entre si e cruzando-se essas informações, almejar melhora no processo de diagnóstico.

A pesquisa foi realizada com a colaboração de alunos e professores, tanto na busca por bancos de dados com imagens de Raios-X positivos para a doença, como na elaboração do código para a análise dessas imagens (rede neural). O intuito geral seria após elaborado o respectivo código, a implementação deste em aplicativo de fácil manuseio para entidades da saúde poderem utilizá-lo. Deste modo, a pesquisa pode ser segmentada em duas etapas: Busca por bancos de dados e os primeiros testes da rede desenvolvida (2020) e o aprimoramento da rede, buscando maior acurácia, somado com o desenvolvimento do aplicativo (2021).

Os autores participaram da primeira etapa, auxiliando na busca por bancos de dados e na formulação do código. Paralelamente a esta pesquisa, em 2021 através de uma pesquisa bibliográfica, foi elaborado um banco



de dados que relacionava o número de mortes, testes realizados e casos da doença em países da América do Sul às medidas de segurança empregadas por cada Estado, objetivando futuramente avaliar com dados o impacto das políticas públicas no número de óbitos.

## 2 MÉTODO

A busca pelos bancos de dados para o desenvolvimento do aplicativo foi feita inicialmente através de canais oficiais de divulgação de casos de covid-19, geralmente hospitais como o John Hopkins dos Estados Unidos, alguns portais do próprio governo brasileiro, etc. Os que mais se adequavam para compor o base de dados da rede na pesquisa foram o *NIH Chest X-Ray Dataset* e o *HM Hospitales*.

O banco de dados NIH é um conjunto de dados de raios-X de tórax, composto por 112.120 imagens de raios-X com rótulos de doenças de 30805 pacientes. Ele não possuía imagens referentes à covid, no entanto teve grande importância para o desenvolvimento do projeto, uma vez que, para fim de treinamento da rede, já estava pré-processado e ainda apresentava um conjunto de imagens de pulmões saudáveis.

O banco de dados do HM Hospitales EHR possui informações desde o início da pandemia. Composto por registros de 2,310 pacientes que foram testados para Covid-19, tanto positivos como negativos. O contato com a instituição espanhola foi realizado via e-mail pela orientadora, rogando pelo acesso às informações (imagens e metadados) pertinentes à pesquisa.

Dando continuidade, a seguir foi feita a elaboração do banco de dados que relacionava o número de mortes por covid-19 com as medidas de segurança empregadas por cada país. A fim de manter-se o mais próximo da realidade, a parte que referencia o número de mortes foi obtido pela mesma fonte utilizada pelo Google, o site *Our World in Data*, sob controle da Universidade de Oxford.

Os dados referentes às medidas preventivas foram obtidos através do site *acaps.org*, se tratando de uma iniciativa global criada com o intuito de examinar o impacto da covid-19 ao redor do mundo. Cada medida empregada por cada Estado encontrada neste site pode ser verificada através de publicações nos veículos de imprensa consultados.

## 3 RESULTADOS

O primeiro tópico da pesquisa, isto é, a rede convolucional desenvolvida para análise dos raios-x teve sua performance avaliada através da metodologia dos positivos e negativos. São quatro categorias distintas possíveis de serem obtidas para o modelo proposto: Positivo verdadeiro, positivo negativo, falso positivo e falso negativo. Caso o raio-x de um paciente saudável seja categorizado pela rede como afetado pela Covid-19, essa classificação pode ser chamada do tipo falso positivo, sendo uma análise semelhante para as demais classificações.

Com estes dados, podem ser estabelecidas algumas métricas para avaliar o desempenho da rede. Dentre as mais usuais, foram utilizadas: Acurácia, precisão, *recall* e *F1-Score*. A primeira métrica promove uma análise geral do modelo, a segunda métrica avalia se uma dada classificação foi de fato feita corretamente, a terceira verifica a frequência com que o classificador encontra os exemplos de uma classe e, por fim, a quarta métrica combina as últimas duas para uma análise final.

A seguir podem ser vista as fórmulas para o cálculo dessas métricas:

$$acuracia = \frac{TP + TF}{TP + TF + FP + FN} \quad (1)$$

$$precisao = \frac{TP}{TP + FP} \quad (2)$$



$$recall = \frac{TF}{TF + FP} \quad (3)$$

$$F1Score = \frac{2 * precisao * recall}{precisao + recall} \quad (4)$$

Legenda: TP: *True Positive* (verdadeiro positivo), TF: *True False* (falso positivo), TN: *True Negative* (verdadeiro negativo), FN: *False Negative* (falso negativo).

Os dados obtidos resultaram no que pode ser visto na tabela a seguir:

**Tabela 1 – Resultados**

Diagnóstico	Precisão	Recall	F1-Score	Amostras
Não Covid-19	0,49	0,85	0,62	380
Covid-19	0,54	0,17	0,26	401

**Fonte: Autoria Própria (2020)**

Em relação ao desenvolvimento de um banco de dados, a partir do *Our World in Data* foi possível extrair as datas, dia após dia desde meados de fevereiro até dezembro de 2020, o número total de casos, de novos casos, total de mortes, novas mortes, novos testes e total de testes e, do segundo banco de dados, foi possível extrair as medidas empregadas, as categorias em que elas se encaixavam, comentários descrevendo resumidamente as medidas adotadas e os endereços eletrônicos (*links*) das notícias vinculadas aos meios de imprensa consultados.

**Figura 1 – Banco de Dados Covid-19 na América do Sul**

	A	B	C	D	E	F	G	H	I	J	K	L
	LOCATION	DATE	CATEGORY	MEASURE	COMMENTS	LINK	TOTAL CASES	NEW CASES	TOTAL DEATHS	NEW DEATHS	NEW TESTS	TOTAL TESTS
74	Argentina	12/03/2021	Public health mea	Isolation and quarantin	From Europe, UK, USA,	<a href="https://www">https://www</a>	19	0	1	0	47	726
75	Argentina	12/03/2022	Social distancing	Limit public gatherings	Restrict public gatherin	<a href="https://www">https://www</a>	19	0	1	0	47	726
77	Argentina	14/03/2020					34	3	2	0	49	862
78	Argentina	15/03/2020	Movement restric	Visa restrictions	For non-residents and t	<a href="https://www">https://www</a>	45	11	2	0	81	943
79	Argentina	16/03/2020	Governance and s	Economic measures	Social protection plan t	<a href="https://www">https://www</a>	56	11	2	0	106	1049
80	Argentina	16/03/2021	Social distancing	Closure of businesses a	All relevant public servi	<a href="https://www">https://www</a>	56	11	2	0	106	1049
81	Argentina	16/03/2022	Social distancing	Schools closure	For 2 weeks.	<a href="https://www">https://www</a>	56	11	2	0	106	1049
82	Argentina	17/03/2020	Governance and s	Economic measures	Plan of investment to p	<a href="https://www">https://www</a>	68	12	2	0	128	1177
83	Argentina	18/03/2020	Public health mea	Strengthening the publi	Construction of 8 temp	<a href="https://www">https://www</a>	79	11	2	0	203	1380
84	Argentina	19/03/2020	Social distancing	Limit public gatherings	Forbidding standing roo	<a href="https://ar.u">https://ar.u</a>	97	18	3	1	152	1532
85	Argentina	19/03/2021	Movement restric	Domestic travel restrict	Restriction on intercity	<a href="https://ar.u">https://ar.u</a>	97	18	3	1	152	1532
86	Argentina	19/03/2022	Lockdown	Partial lockdown	Until 31st of March incl	<a href="https://www">https://www</a>	97	18	3	1	152	1532
87	Argentina	19/03/2023	Governance and s	State of emergency dec	National state of emerg	<a href="https://ar.u">https://ar.u</a>	97	18	3	1	152	1532
88	Argentina	20/03/2020	Governance and s	Economic measures	Price control on basic g	<a href="https://www">https://www</a>	128	31	3	0	232	1764
89	Argentina	20/03/2021	Governance and s	Economic measures	Toll collection is suspen	<a href="https://www">https://www</a>	128	31	3	0	232	1764
90	Argentina	21/03/2020					158	30	4	1	323	2087
91	Argentina	22/03/2020					266	108	4	0	247	2334
92	Argentina	23/03/2020	Governance and s	Economic measures	One-time payment of 1	<a href="https://www">https://www</a>	301	35	4	0	288	2622
93	Argentina	24/03/2020	Governance and s	Economic measures	Government suspends t	<a href="https://www">https://www</a>	387	86	6	2	445	3067
94	Argentina	25/03/2020	Movement restric	Surveillance and monit	Every person entering t	<a href="https://www">https://www</a>	387	0	8	2	550	3617
95	Argentina	26/03/2020	Governance and s	Economic measures	Every health worker will	<a href="https://www">https://www</a>	502	115	9	1	586	4203
96	Argentina	26/03/2021	Movement restric	Border closure	Country's borders an	<a href="https://www">https://www</a>	502	115	9	1	586	4203
97	Argentina	26/03/2022	Movement restric	Border closure	Border closure is exten	<a href="https://www">https://www</a>	502	115	9	1	586	4203
98	Argentina	27/03/2020					589	87	13	4	457	4660
99	Argentina	28/03/2020	Governance and s	Economic measures	Every security and defe	<a href="https://www">https://www</a>	690	101	18	5	482	5142
100	Argentina	29/03/2020	Governance and s	Economic measures	Rent payments and evic	<a href="https://www">https://www</a>	745	55	19	1	718	5860

**Fonte: Autoria própria (2021)**

A planilha conta com 4431 linhas, relacionando informações sobre 12 países: Argentina, Bolívia, Brasil, Chile, Colômbia, Equador, Guiana Francesa, Paraguai, Peru, Suriname, Uruguai e Venezuela. As fontes dos dados vão desde os sites oficiais dos países estudados até mídias com impacto considerável mundialmente.

#### 4 CONCLUSÃO

Este trabalho visou contribuir para o enfrentamento da pandemia do coronavírus em dois frentes, auxiliando no diagnóstico da doença por intermédio do desenvolvimento de redes convolucionais e de um



banco de dados das medidas de segurança empregadas, buscando provar a validade delas como medidas de contenção da propagação do vírus.

Ao término da fase 1 do projeto da rede convolucional podemos identificar resultados pouco satisfatórios, dada a obtenção de uma acurácia de 50%. No entanto, é importante lembrar que este foi somente o piloto do modelo a ser empregado na versão final. Projeta-se que com as melhorias do modelo a partir de consultas a bibliografia pertinente seja possível de elevar a acurácia de 50% até 85-95%, se equiparando aos modelos de teste empregados atualmente. Por ser uma forma de teste que necessitaria somente do celular/computador com o aplicativo e de um raio-x do pulmão do paciente, agilizaria o processo de diagnóstico e diminuiria a demanda pela compra de testes e também da necessidade de pessoas especializadas como no caso do teste RT-PCR. É importante citar que o intuito não é substituí-los, somente contribuir com uma nova alternativa.

Além disso, com o banco de dados elaborado em 2021 será possível avaliar, em pesquisas futuras, a influência das medidas de segurança com base em dados e estatísticas, sendo possível verificar as medidas mais eficientes. Cerca de 4,55 milhões de pessoas morreram em decorrência da doença Covid-19 em todo o mundo desde que a pandemia teve seu início, sendo 219 milhões em número de casos. Deste modo, é de extrema importância que sejam conduzidas pesquisas que ajudem a pôr um fim a esse quadro de mortes.

## REFERÊNCIAS

- LOUREIRO, Camila Melo Coelho et al. Alterações pulmonares na COVID-19. **Revista Científica Hospital Santa Izabel**, v. 4, n. 2, p. 89-99, 2020.
- RITCHIE, Hannah et al. Coronavírus Pandemic (Covid-19). Publicado online em OurWorldInData.org. Disponível em: <[https:// ourworldindata.org/coronavirus](https://ourworldindata.org/coronavirus)> Acesso em 15 mai 2021.
- ZHANG, Tao; WU, Qunfu; ZHANG, Zhigang. **Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak**. Current biology, v. 30, n. 7, p. 1346-1351. e2, 2020.
- KRIZHEVSKY, Alex; SUTSKEVER, Ilya; HINTON, Geoffrey E. **Imagenet classification with deep convolutional neural networks**. Advances in neural information processing systems, v. 25, p. 1097-1105, 2012.
- NARIN, Ali; KAYA, Ceren; PAMUK, Ziyet. **Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks**. Pattern Analysis and Applications, p. 1-14, 2021.
- BALTRUSCHAT, Ivo M. et al. **Comparison of deep learning approaches for multi-label chest X-ray classification**. Scientific reports, v. 9, n. 1, p. 1-10, 2019.
- COHEN, Joseph Paul et al. **On the limits of cross-domain generalization in automated X-ray prediction**. In: Medical Imaging with Deep Learning. PMLR, 2020. p. 136-155.
- WANG, Xiaosong et al. Chestx-ray8: **Hospital-scale chest x-ray database and benchmarks on weakly-supervised classification and localization of common thorax diseases**. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 2097-2106.
- BALTRUŠAITIS, Tadas; AHUJA, Chaitanya; MORENCY, Louis-Philippe. **Multimodal machine learning: A survey and taxonomy**. IEEE transactions on pattern analysis and machine intelligence, v. 41, n. 2, p. 423-443, 2018.



SEI-SICITE 2021

Pesquisa e Extensão para um  
mundo em transformação

XI Seminário de Extensão e Inovação  
XXVI Seminário de Iniciação Científica e Tecnológica  
08 a 12 de Novembro - Guarapuava/PR



- HUANG, Guang-Bin et al. **Extreme learning machine for regression and multiclass classification.** IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), v. 42, n. 2, p. 513-529, 2011.
- SANTOS, Roberto. **Testes diagnósticos para Covid-19: quais as evidências disponíveis?** Revista eletrônica PebMed. Disponível em: < <https://pebmed.com.br/testes-diagnosticos-para-covid-19-quais-as-evidencias-disponiveis/>>. Acesso em: 28 set 2021.
- FLORIANO, Idevaldo; SILVINATO, Antonio. **Testes diagnóstico para novo coronavírus 2019 (COVID-19).** Núcleo de Medicina Baseada em Evidência - Unimed Regional da Baixa Mogiana. 2020.
- PONTE, Gabriella. **Quais exames são usados para o diagnóstico da COVID-19?** FioCruz. Disponível em: < <https://www.bio.fiocruz.br/index.php/br/noticias/1785-quais-exames-sao-usados-para-o-diagnostico-da-covid-19>>. Acesso em: 28 set 2021.