

Aplicação da rede neural convolucional para determinar e classificar em imagens de raio-X de tórax a presença ou não do vírus da COVID-19*

Wiliam Regone¹, Luciel Henrique de Oliveira¹ e Vitoria Caprioglio Oliveira¹

¹ Pontifícia Universidade Católica de Minas Gerais - Campus Poços de Caldas

Av. Pe. Cletus Francis Cox, 1.661 – Jardim Country Club

37.714-620 – Poços de Caldas – Minas Gerais

{wiliamregone, caprioli.oliveira}@gmail.com

Abstract. Radiographic images, such as chest X-rays, are tools used in disease diagnosis, including types of pneumonia caused by viral infections. Through models based on artificial intelligence, it is possible to develop computational tools that allow the process of image classification from a series of characteristics. Based on this, the present research proposes an algorithm structured as a convolutional neural network to encode and train the classification of chest X-ray images with COVID-19 and regular viruses. The results are obtained using two models. The first model used ResNet50 with an accuracy of 99% for images with COVID-19 and 92% for regular images. The second model used image processing and obtained accuracy of 96% for images with COVID-19 and 92% for regular images.

Resumo. Imagens radiográficas, como a radiografia de tórax, são ferramentas utilizadas no processo de diagnóstico de doenças, incluindo pneumonias causadas por infecções virais. Por meio de modelos baseados em inteligência artificial, é possível desenvolver ferramentas computacionais que permitam o processo de classificação de imagens a partir de uma série de características. Com base nisso, a presente pesquisa propõe um algoritmo estruturado na forma de rede neural convolucional para codificar e treinar a classificação de imagens de raios X de tórax com vírus COVID-19 e normais. Os resultados foram obtidos usando dois modelos. O primeiro modelo usou o ResNet50 com uma precisão de 99% para imagens com COVID-19 e 92% para imagens normais. O segundo modelo usou o processamento de imagens e obteve uma precisão de 96% para imagens com COVID-19 e 92% para imagens normais.

Palavras chaves: Inteligência Artificial; Rede Neural Convolutinal; Imagem de raio X; COVID-19

*Financiado pela PUCMINAS – Campus de Poços de Caldas - FIP - EDITAL Nº 052/2020

1. Introdução

O coronavírus causa uma síndrome respiratória aguda grave. O número de pessoas infectadas pelo vírus aumentou rapidamente. Isso levou a uma grande preocupação de saúde pública na comunidade internacional, como a Organização Mundial da Saúde (OMS). Desta forma, declarou que o surto foi uma emergência de saúde pública e a reconheceu como uma pandemia em 11 de março de 2020 [WORLD HEALTH ORGANIZATION et al. 2020].

A radiografia de tórax é uma das modalidades de imagem comumente usada para visualizar e quantificar a anatomia de uma ou mais estruturas presentes na região do pulmão. Isto se deve ao fato da radiografia proporcionar alta resolução de imagens que indicam se existe alguma anormalidade presente nos pulmões, para assim investigar possíveis patologias. A incidência de pneumonia pode ser confirmada com imagens de raios-X e pode orientar a prevenção e tratamento. Vários artigos da área médica descrevem as manifestações mais comuns e padrões de anormalidade pulmonar na radiografia de tórax com COVID-19. Desta forma, recomendam que a comunidade médica pode confiar frequentemente neste método para a redução dos problemas de controle de infecção [JACOBI et al. 2020].

A visão computacional, uma sub-área da Inteligência Artificial (IA), é uma tecnologia emergente no campo da imagem médica, e apresenta várias técnicas de utilização prática. Assim, podendo contribuir ativamente para atuar contra a COVID-19. Uma delas é a *ResNet50* que é uma técnica de aprendizado de máquina baseado no conceito de reutilização. A maior vantagem de usar este método de aprendizagem por transferência é que permite o treinamento das imagens com um menor conjuntos de dados e requer menos custos de cálculo. Uma outra técnica é a arquitetura de uma Rede Neural Convolucional (RNC) sequencial. Esta é um dos métodos de aprendizado profundo mais populares. O seu principal sucesso é devido à sua capacidade de aprender recursos automaticamente de domínios específicos de imagens [PLATFORM 2017].

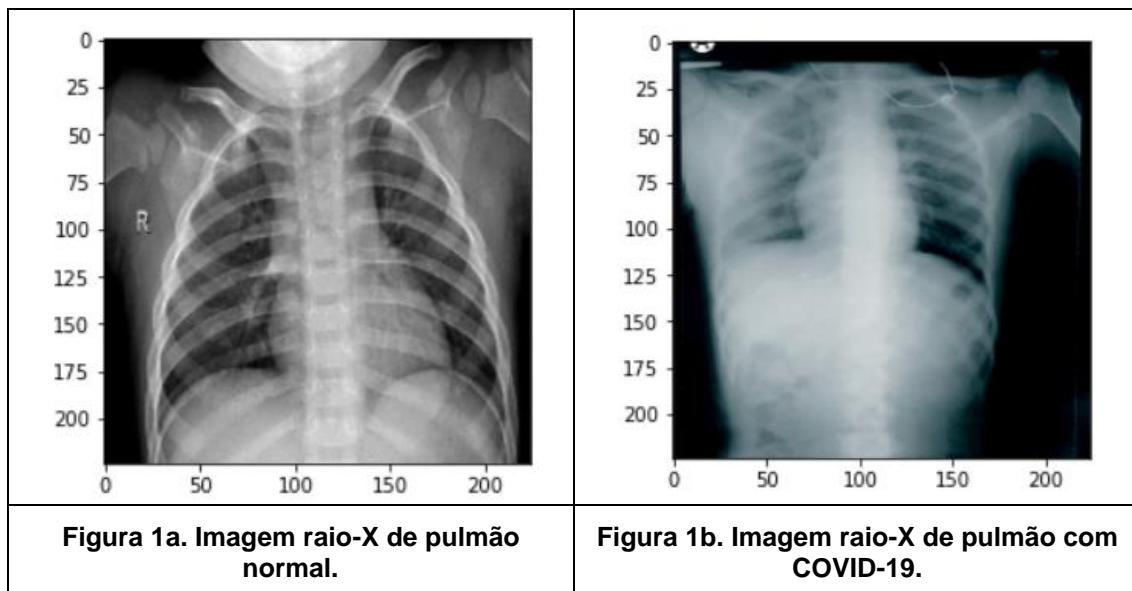
Quando comparada a análise humana com a inteligência artificial por meio da visão computacional, esta última pode permitir uma avaliação mais detalhada das imagens, pois pode processar e interpretar os pixels de forma mais precisa, segura e eficiente. Dada a relevância da IA em todos os campos do saber, a utilização dessa tecnologia para análise de imagens de radiografia com COVID-19 inspira aplicações práticas e metodologias de pesquisa.

Assim, neste projeto codificou-se e identificou-se as diferenças entre as imagens de pulmões normais, e com o vírus da COVID-19. Desta forma, tendo como base na IA, modelos de algoritmos foram escritos em uma técnica de rede neural para modelar imagens e transformá-las em dados estatísticos. Estes modelos foram treinados para classificar as imagens radiográficas de tórax (pulmões) com COVID-19, e imagens normais. Os resultados dos modelos são apresentados no modelo 1 (procedimento dado por *ResNet50*) e no modelo 2 (onde não há aprendizagem por transferência) sendo estes procedimentos de classificação diferentes. Desta forma, espera-se fornecer orientação para pesquisadores e radiologistas dos resultados aqui apresentados [BULLOCK et al. 2020].

2. Metodologia

Neste estudo, usou-se um conjunto de dados de imagens que está disponível no *Kaggle*¹, uma plataforma que hospeda *Dataset*. Nesta há um conjunto de imagens de raios-X de tórax normais, e imagens de pneumonia da COVID-19.

A Figura 1a mostra um exemplo de uma imagem normal, e a Figura 1b exemplo com COVID-19. Pode-se perceber que sem a análise de um especialista é praticamente impossível diagnosticar as imagens. A rede neural pode ser utilizada para melhorar a qualidade de imagens de raio-X de tórax e identificar a presença da doença da COVID-19 nas imagens.



A *ResNet* (do inglês *Residual Network*) é uma rede neural convolucional da *Microsoft* no conjunto de dados *ImageNet* que está evoluindo constantemente. Esta foi construída por blocos residuais. Nesta técnica todas as camadas são excluídas exceto a última, que é treinada para o problema específico. A *ResNet50* é uma rede residual com 50 camadas e usa camadas com pré-ativação de pesos. Esta técnica é útil para aplicativos médicos, uma vez que não requer tantos dados de treinamento, pois pode ser difícil de conseguir em situações médicas [HUANG et al. 2017].

Rede neural convolucional é uma arquitetura inspirada em técnicas de aprendizado profundo, representando os padrões das imagens e aprendendo estas características. Usualmente, três tipos de camadas são usados para construir uma RNC: camada convolucional, camada de agrupamento *pooling* e camada totalmente conectada. Esta rede é uma forma de algoritmo e projetada entre dimensões. A arquitetura desta rede possui estruturas de multi-dimensões, sendo atualizado em cada nova dimensão uma melhora no aprendizado desta rede através de estímulos para identificar os padrões de cada imagem [LECUN et al. 2010].

¹<https://www.kaggle.com/tawsifurrahman/covid19-radiography-database>

A convolução é uma operação matemática entre dimensões, representando uma imagem como uma matriz. Uma imagem possui áreas denominadas *pixels* que são separadas como partes da matriz. A estrutura dessas camadas convolucionais são compostas por filtros, também em formato matricial para obter características das imagens. Estas informações da figura original são mantidas, resultando que apenas informações com finalidades classificativas estejam presentes nos filtros. Áreas presentes na imagem são multiplicadas pelos filtros obtendo assim características de partes da imagem [LECUN et al. 2010].

Os resultados são apresentados em dois modelos separados, 1 e 2. Isto se tornou necessário pois na forma de construir a rede neural utilizou-se de duas técnicas. Essa diferença vai averiguar qual delas apresenta melhores resultados.

Os dois modelos apresentados demonstram exemplos de implementações de aprendizado profundo para classificação de imagens. Nestes empregou-se a biblioteca *TensorFlow* e *keras* para construir e treinar uma rede neural com o objetivo de classificar imagens em duas categorias distintas. Os dados de treinamento com 3000 imagens foram pré-processados utilizando o *ImageDataGenerator* do *TensorFlow*. Este foi empregado para configurar o pré-processamento e a intensificação dos dados das imagens. Esta classe aplica uma gama de transformações nas imagens, incluindo rotação de até 50 graus, deslocamento de largura e altura em 20%, zoom de 10% e inversão tanto na horizontal quanto na vertical [PLATFORM 2017]. Desta forma, o conjunto de dados de treinamento constituído de imagens normais e viral com COVID-19 foram expandidos em cada época da compilação do modelo. Para os dados de teste foram usadas 1500 imagens.

A rede *ResNet50* é utilizada como base do modelo 1. Os pesos da rede são inicializados com aqueles pré-treinados na base de dados *ImageNet*, e a camada superior (camada de classificação) é removida. As imagens são pré-processadas para a dimensão de 224x224 pixels, que é a dimensão de entrada padrão para a rede *ResNet50*. Uma camada de agrupamento *GlobalAveragePooling2D* é aplicada para reduzir a dimensão espacial de cada mapa de características na saída do modelo base. A codificação de rótulos é feita em formato *one-hot* e os dados são passados para o modelo [ABADI et al. 2016]. Após isso, implementou-se uma série de camadas ocultas totalmente conectadas *Dense*, cada uma com uma função de ativação *ReLU* e um número variado de neurônios, de 2048, 1885, 1408, 1024, 512, 256 e 128. A camada final é uma camada com duas unidades, correspondendo ao número de classes do problema, e uma função de ativação *softmax*. Por fim, o modelo é compilado com o otimizador *Adam* e utilizou-se a função de perda *categorical_crossentropy* e a métrica *accuracy*. Durante cada uma das 100 épocas de treinamento, a acurácia foi calculada para ambos os conjuntos de treinamento e de teste, fornecendo uma medida contínua do desempenho do modelo.

Para o modelo 2 o processamento das imagens é realizado e incorpora várias camadas. As duas primeiras são camadas de convolução 2D. Estas camadas recebem imagens de dimensões 224x224 *pixels* [LECUN et al. 2015]. Para cada camada de convolução 2D aplicou-se 32 filtros com um kernel de tamanho 3x3 e usando *padding same*, garantindo que a altura e a largura da saída correspondam às da entrada. As camadas convolucionais foram seguidas por uma camada de *MaxPooling* com *pool size* e *strides* de 2x2 para reduzir a dimensionalidade e a quantidade de parâmetros do modelo. Depois os dados gerados formaram uma camada *Flatten* para poderem ser alimentados para as camadas totalmente conectadas *Dense*. Há duas camadas ocultas com 128

neurônios cada e usando a função de ativação *ReLU*. A camada de saída é uma camada *Dense* com dois neurônios, correspondendo às duas categorias que o modelo está tentando classificar, e usa a função de ativação *softmax*. O modelo é então compilado com o otimizador *Adam* e a função de perda *categorical_crossentropy*. A métrica de desempenho usada para avaliar o modelo durante o treinamento é a acurácia. O modelo é treinado por 100 épocas, e após o treinamento, obteve-se uma acurácia de treinamento e teste que indica a eficiência na classificação das imagens nas duas categorias propostas.

Para executar e criar códigos na linguagem *Python* usou-se o *Google Colab*. Esta é uma ferramenta de serviço de armazenamento em nuvem de *notebooks*, diretamente em um navegador, sem a necessidade de nenhum tipo de instalação de *software* em uma máquina. Hospedado pelo próprio *Google* para incentivar a pesquisa de inteligência artificial. *Google Colab* executa os códigos das bibliotecas *Numpy*, *Pandas*, *Matplotlib*, *Tensor Flow* e *Keras*, e outras bibliotecas *Python*, que são pacotes usados para criar uma rede neural e desenvolver a estrutura de aprendizado profundo.

3. Resultados e Discussões

O sistema de saúde não deve ser estacionário, mas deve aprender com suas próprias experiências e se esforçar para implementar melhorias contínuas de processos. Este processo envolve a construção ou participação em uma organização, que usa inteligência artificial para alcançar um progresso significativo.

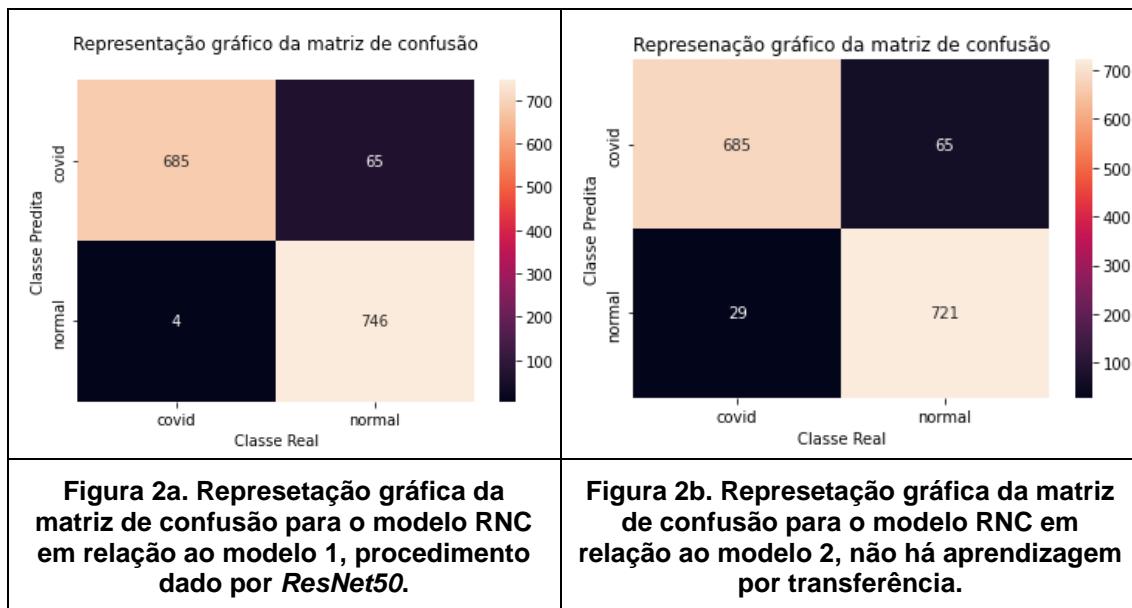
O método utilizado para medir e analisar os resultados dos modelos 1 e 2 consistem em obter matrizes de confusões e, a partir delas, extrair as métricas de acurácia, precisão, *recall* e *F1 score*.

A matriz de confusão oferece uma medida efetiva dos resultados na forma de classificação em visualização de tabela e ou formato gráfico. Esta mostra o número de classificações reais versus as classificações preditas para cada classe, sobre um conjunto de dados. A matriz de confusão é dada em termos de verdadeiro positivo e falso negativo. Destaca-se que o termo positivo é quando a imagem de raio-X de tórax tem o vírus da COVID-19, e negativo quando na imagem não há o vírus, ou seja, imagem normal. Verdadeiro positivo é observação cujo valor real é positivo e o valor previsto é positivo, assim, o modelo acertou. Verdadeiro negativo é observação cujo valor real é negativo e o valor previsto é negativo, isto é, o modelo também acertou. Desta forma, temos uma a visualização do desempenho do algoritmo de classificação [KOHAVI 1998].

Os resultados dados pelas matrizes de confusões para os procedimentos dos modelos 1 e 2 são mostrados nas Figuras 2a e 2b. Na diagonal principal temos o acerto do modelo e secundária o erro. Um total de 1500 imagens de teste foram usadas em cada modelo.

O resultado do modelo 1 é mostrado na Figura 2a. Este previu corretamente 685 imagens com COVID-19 (verdadeiro positivo) e 746 imagens normais (verdadeiro negativo). Na diagonal secundária temos que a previsão não consegue acertar o valor real. Assim, ele previu 4 imagens como normal (falso negativo), quando na realidade eram de COVID-19. Tem-se também que houve 65 erros de imagem normais (falso positivo) e foram previstas como COVID-19.

A Figura 2b, apresenta os resultados para o modelo 2. Neste caso previu-se corretamente 685 imagens como COVID-19 (verdadeiro positivo) e 721 imagens normais (verdadeiro negativo). Lembrando que na diagonal secundária temos os erros do modelo. Assim, foram previstas 29 imagens como normais (falso negativo), mas na realidade eram de COVID-19. Houve 65 erros de imagem normais (falso positivo) que foram prevista como COVID-19.



Acurácia significa exatidão e a medida acurada é aquela cujo valor se aproxima do valor correto. No caso em estudo considere como exemplo um classificador para avaliar os resultados de um exame de raio-X de tórax e indicar a probabilidade (fazer previsões de classificação) de uma imagem apresentar o COVID-19 ou não. Considerase que a base de dados para classificação são os dados de teste utilizados nos modelos 1 e 2 em um total de 1500 imagens.

O conceito de precisão está associado a ideia de medida e de repetição. Quanto mais preciso é uma medição, menor será a variação entre os valores medidos. A métrica de *recall* é utilizada para indicar a relação entre as previsões positivas realizadas corretamente e todas as previsões que realmente são positivas. O *recall* é bastante útil quando precisamos minimizar os falsos negativos. Sempre que precisarmos minimizar os falsos negativos devemos buscar maior percentual no *recall*. *F1 Score* é a média harmônica da precisão e do *recall*. Quando os valores são próximos, o resultado é semelhante à média aritmética [KOHAVI 1998].

A métrica dada pela acurácia (acerto), ou seja, quanto o valor se aproxima do valor correto para o modelo 1, é em média de aproximadamente 95% com desvio padrão de 0,05 para os dados de teste, um total de 1500 imagens. A Tabela 1 descreve os valores de precisão, *recall* e *F1 Score* para o modelo 1 em porcentagem de acerto com imagem que contém o vírus da COVID-19 e também imagem normal, ou seja, sem o vírus.

Tabela 1. Métricas para o modelo utilizando a ResNet50.

modelo 1	precisão	recall	F1 Score
imagem com COVID-19	99%	91%	95%
imagem normal	92%	99%	96%

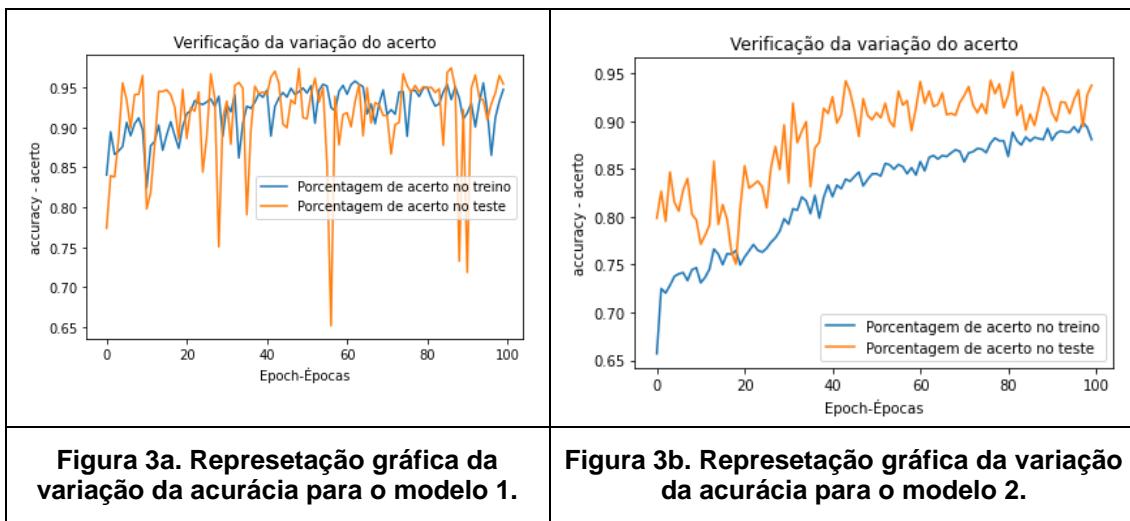
Para o modelo 2 a métrica dada pela acurácia, é em média de aproximadamente 94% com desvio padrão de 0,05 para os dados de teste, ou seja, 1500 imagens. A Tabela 2 descreve os valores de precisão, *recall* e *F1 Score* em porcentagem de acerto com imagem com COVID-19 e normal.

Tabela 2. Métricas para o modelo que não utilizou a aprendizagem por transferência.

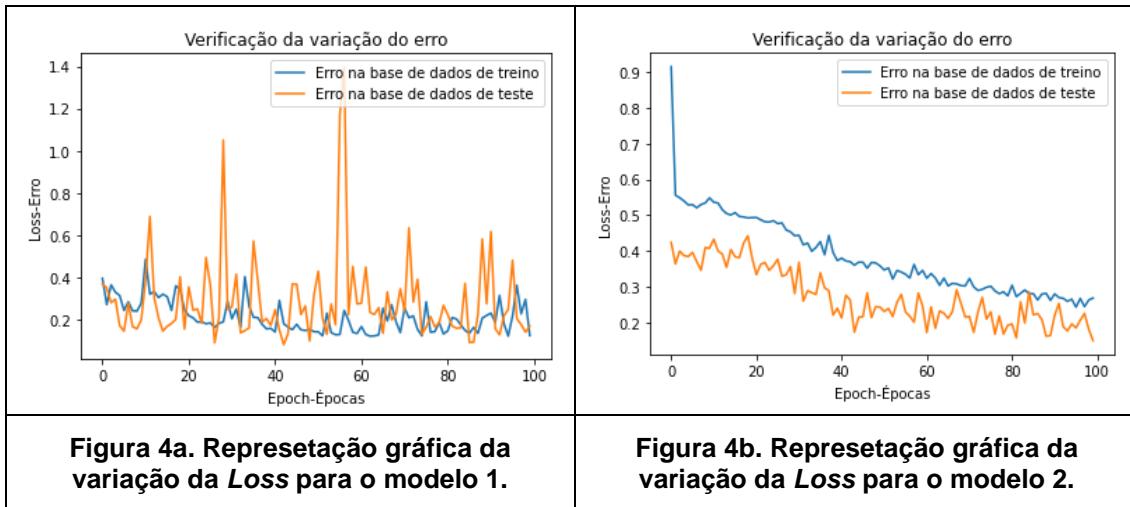
modelo 2	precisão	recall	F1 Score
imagem com COVID-19	96%	91%	94%
imagem normal	92%	96%	94%

Comparando as tabelas, temos que a Tabela 1, que apresenta o modelo 1 dado pela *ResNet50* obteve resultados melhores. Os dados de precisão das imagens com COVID-19 e no *recall* com imagens normal, como o *F1 Score* comprovam essa afirmação. A *ResNet50* é mais precisa em classificar imagens de pacientes em condições normais e pacientes com o vírus da COVID-19. Os dados mostrados nas Figura 2a e 2b (matriz de confusão), quando comparados, indicam que o valor do *recall* para o modelo 1 tem a capacidade maior de encontrar todas as amostras positivas e negativas.

As Figuras 3a e 3b demonstram o comportamento da acurácia (acerto) das arquiteturas durante a etapa de treinamento e teste. Na Figura 3a, que representa o modelo 1, onde utilizou-se a *ResNet50*, há pico de valores na etapa de teste e obteve uma convergência mais rápida e superior a 85% depois da época 20. Para a Figura 3b esse procedimento não ocorre, atingindo uma tendência para valores constantes e maior que 80% a partir da época 50 para os dados de teste. Para os dados de treino tem-se uma inclinação constante dos valores a partir da época 40.



Em relação ao *loss* (erro) são observadas nas Figuras 4a e 4b. Essas figuras mostram aspectos diferentes. Percebe-se que na Figura 4a na etapa de teste, a função *ResNet50* apresenta picos com valores altos. Já na Figura 4b apresentam valores próximos para teste e treinamento e com a mesma tendência decrescente.



Os altos picos no comportamento médio das arquiteturas apresentados nas Figuras 3a e 4a são provavelmente por causa da taxa de aprendizado que é otimizada e diminui no decorrer das épocas. Outro fato deve-se ao otimizador *Adam*, que é mais lento.

Analizando as Figuras 3b e 4b, observar-se que os modelos não apresentam um comportamento de *OverFitting* (quando modelo estatístico se ajusta muito bem ao conjunto de dados). Desta forma, as curvas identificadas tanto para os dados de treino quanto para os dados de teste estão relativamente separadas com o passar das épocas, indicando que o modelo não se ajustou muito fortemente na etapa de teste e treinamento.

É importante considerar como a inteligência artificial é uma das tecnologias emergentes e servirá para o desenvolvimento de nossa saúde. A quantidade de dados

precedentes disponíveis, combinados com avanços em algoritmos de processamento da rede neural, e a consciência social com o uso de aplicativos da IA serão cada vez mais úteis para os consumidores. A partir das informações obtidas, o médico pode fazer inferências sobre o que aflige o paciente.

4. Conclusão

Portanto, o objetivo de usar a inteligência artificial para descrever estatisticamente como uma técnica médica para ajudar no combate da COVID-19 mostra-se bastante promissor. Pois considerando as 1500 imagens o algoritmo acertou entre 96 a 99% imagens que contém o vírus da COVID-19 e 92% imagens normais. Desta forma, processar e classificar as imagens de raios-X de tórax que contém ou não o vírus de COVID-19 por rede neural convolucional pode ajudar os médicos no diagnóstico da doença. Portanto, o uso das técnicas de inteligência artificial em radiografias do pulmão pode-se tornar tão popular quanto fácil. Assim, podendo contribuir para o combate desta pandemia e outras futuras.

5. Referência Bibliográfica

- ABADI, Martín et al. Tensorflow: A system for large-scale machine learning. In: 12th {USENIX} symposium on operating systems design and implementation ({OSDI} 16). 2016. p. 265-283.
- BULLOCK, Joseph et al. Mapping the landscape of artificial intelligence applications against COVID-19. *Journal of Artificial Intelligence Research*, v. 69, p. 807-845, 2020.
- HUANG, Gao et al. Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2017. p. 4700-4708.
- JACOBI, Adam et al. Portable chest X-ray in coronavirus disease-19 (COVID-19): A pictorial review. *Clinical imaging*, v. 64, p. 35-42, 2020.
- KOHAVI, Ron. Glossary of terms. Special issue on applications of machine learning and the knowledge discovery process, v. 30, n. 271, p. 127-132, 1998.
- LECUN, Yann; BENGIO, Yoshua; HINTON, Geoffrey. Deep learning. *nature*, v. 521, n. 7553, p. 436-444, 2015.
- LECUN, Yann; KAVUKCUOGLU, Koray; FARABET, Clément. Convolutional networks and applications in vision. In: Proceedings of 2010 IEEE international symposium on circuits and systems. IEEE, 2010. p. 253-256.
- PLATFORM, G. C. Image Classification Transfer Learning with Inception v3. 2017.
- WANG, Wenling et al. Detection of SARS-CoV-2 in different types of clinical specimens. *Jama*, v. 323, n. 18, p. 1843-1844, 2020.
- WORLD HEALTH ORGANIZATION et al. Coronavirus disease 2019 (COVID-19): situation report, 73. 2020.