

Tool to Assist in the Prognosis of covid-19 and Pneumonia in Robust or Restricted Computational Environments

João Felipe Barros da Silva¹, Sílvio Roberto Fernandes¹

¹Departamento de Computação
Universidade Federal Rural do Semi-Árido (UFERSA)
Mossoró, RN – Brasil

{joao.silva27990, silvio}@ufersa.edu.br

Abstract. *Effective screening of patients infected with covid-19 plays a crucial role in combating this disease, with chest X-ray examination being one of the main approaches. In this study, we developed four models using Convolutional Neural Network (CNN) trained with 30.000 images, capable of classifying a lung X-ray image as normal, with pneumonia or with covid-19, with an accuracy close to 90%. In addition, we perform a cost-benefit analysis of the models considering the implementation in more restricted systems, such as embedded systems and systems with greater processing power, such as client-server systems.*

Resumo. *A triagem eficaz de pacientes infectados pelo covid-19 desempenha um papel crucial no combate a essa doença, sendo uma das principais abordagens o exame de radiografia do tórax, no entanto seu diagnóstico pode ser confundido com de uma pneumonia. Neste estudo desenvolvemos quatro modelos utilizando Convolutional Neural Network (CNN) treinadas com 30.000 imagens, capazes de classificar uma imagem de raio-x de pulmão em normal, com pneumonia ou com covid-19, com uma acurácia próxima de 90%. Além disso, fazemos uma análise de custo benefício dos modelos considerando a implementação em sistemas mais restritos, como em sistemas embarcados e sistemas com maior poder de processamento, como sistemas em servidores.*

1. Introdução

A covid-19 é uma infecção respiratória aguda causada pelo coronavírus SARS-coV-2. A doença se espalhou rapidamente, apresentando uma crise de saúde global, sendo declarada, em março de 2020 pela Organização Mundial da Saúde (OMS) como uma pandemia global. Desde a descoberta do primeiro caso, na China, houve mais de 700 milhões de infectados com cerca de 6,9 milhões de mortes em todo o mundo [OMS 2023].

Com o advento das vacinas houve uma significativa queda nas mortes pela covid-19, contudo a mesma continua matando diariamente. Assim como em outros casos clínicos o tratamento precoce é essencial para diminuir os índices de mortalidade, atualmente a Reação em Cadeia da Polimerase em Tempo Real (RC-PCR) é o método de diagnóstico padrão para esta doença, entretanto essa técnica tem apresentado altas taxas de falso negativos, além de seu processo ser invasivo e demorado, levando de 2-3 dias para produzir o resultado.

Pesquisas médicas descobriram que exames de imagem como Tomografia Computadorizada (TC) e radiografia de tórax podem ser utilizadas para detectar pessoas infectadas com a covid-19, no entanto, as anormalidades causadas pela covid-19

percebidas nas imagens, são similares às encontradas em pessoas com pneumonia, tornando o diagnóstico mais complicado [Chen et al. 2020], [Guan et al. 2020].

Dessa forma, neste trabalho foram desenvolvidos quatro modelos baseados em aprendizado profundo de máquinas, que permite classificar imagens de raio-x de pulmão em: covid-19, normal e pneumonia. Cada modelo foi desenvolvido com uma arquitetura pré-treinada diferente, utilizando transferência de aprendizagem, sendo elas, VGG16, VGG19, EfficientNetB7 e MobileNet, com o objetivo de adquirir modelos com boa acurácia que possam ser utilizados em sistemas mais robustos do tipo cliente-servidor, quanto em sistemas mais restritos, como sistemas embarcados.

O restante desse artigo é organizado da seguinte maneira: A seção II apresenta um levantamento da literatura de trabalhos semelhantes. A seção III trata do trabalho proposto apresentando as metodologias de desenvolvimento. A seção IV aborda os resultados obtidos do trabalho, assim como, uma comparação do modelos propostos no trabalho

2. Trabalhos Relacionados

Ao longo da pandemia, várias abordagens utilizando aprendizado profundo de máquina foram desenvolvidas para predição de covid-19. Em [Mohammadi et al. 2020] foram propostos quatro modelos utilizando Convolutional Neural Network (CNN) pré-treinadas, VGG16, VGG19, MobileNet e InceptionResNetV2, com apenas 545 imagens de raio-x para treino teste e validação, obtendo assim, os seguintes resultados de acurácia, 93.6%, 90.8%, 99.1%, 96.8%, respectivamente. Para classificação entre covid-19 ou Pneumonia.

O trabalho de Yang et al. (2021) também utilizou quatro modelos de CNN pré-treinadas, VGG16, DenseNet121, ResNet50 e ResNet152. Neste caso foi utilizado um conjunto de dados próprio obtido da junção de outros dois conjuntos de dados públicos, alcançando assim 8.461 imagens, separadas em três classes: covid-19, normal e pneumonia. No mesmo trabalho foram feitas três abordagens de classificação diferentes, primeiramente utilizou uma predição binária, ou seja, a CNN classifica em duas classes, covid-19 ou Pneumonia, e na segunda abordagem, também binária, em covid-19 ou Normal, e por último, a predição em múltiplas classes: covid-19, Normal ou Pneumonia. Em todas as abordagens a precisão passou dos 90%.

Hernandez et al. (2020) desenvolveram quatro modelos, um deles uma arquitetura customizada e os outros utilizando as redes pré-treinadas ResNet50, VGG16 e DenseNet121. Todos os modelos utilizaram o mesmo conjunto de dados, contendo mais de 30.000 imagens, para treino, validação e teste, e obtiveram as seguintes acurácias: 68%, 90%, 82% e 83%, respectivamente.

Outros trabalhos focaram especificamente em construir modelos próprios para a classificação das imagens de tórax. Em [Ozturk et al. 2020] foi proposto um modelo de aprendizado profundo chamado DarkCovidNet, inspirado na arquitetura DarkNet-19, o modelo proposto no trabalho possui 17 camadas convolucionais, assim como, várias filtragens em cada camada. Sendo capaz de executar tarefas binárias e de múltiplas classes com acurácia de 98.8% e 87.02% respectivamente.

Wang et al. (2020) apresentaram o Covid-Net, um modelo de CNN para a detecção de covid-19, de código aberto composto por 13.975 imagens, que obteve 92.2% de acurácia na classificação das classes normal, pneumonia e covid-19.

Ao analisar os trabalhos relacionados foi identificado que a maioria deles apresenta apenas a métrica de acurácia para avaliação da qualidade dos seus modelos. Aqueles que apresentam maiores taxas de acurácia utilizam *datasets* relativamente pequenos ou para classificação binárias (COVID/Normal ou Pneumonia/Normal). Além disso, tais trabalhos não apresentam quaisquer restrições de inferência ao ambiente computacional onde tais soluções poderiam ser empregadas. Portanto, a contribuição do trabalho proposto diz respeito ao preenchimento das lacunas identificadas nos trabalhos relacionados.

3. Trabalho Proposto

Neste trabalho foram desenvolvidos quatro modelos para a classificação de imagens de radiografia do tórax. Para isso, utilizamos quatro arquiteturas de aprendizado profundo disponíveis na API Keras [ONEIROS 2015]. Para todos os modelos, foram utilizados os mesmos parâmetros para treino. Primeiramente, as camadas das arquiteturas escolhidas, foram congeladas evitando a alteração dos seus pesos e garantindo a transferência de aprendizado, em seguida adicionamos três camadas treináveis para compor a rede, uma camada “*Flatten*” responsável por converter o tensor em um vetor unidimensional, em seguida, uma camada densa com 64 neurônios e a função de ativação “*Relu*”, logo após, incluímos o *Dropout* com taxa de 0,2 para evitar *overfitting* e por último, adicionamos uma camada densa para classificação com três neurônios, correspondente às três classes.

Para as configurações específicas de treinamento dos modelos utilizamos o algoritmo Adam para ajustar os pesos, “*SparseCategoricalCrossentropy*” para calcular função de perda, a métrica “*Accuracy*” para avaliar o desempenho do modelo e treinamos por 30 épocas. Além disso, também utilizamos a técnica de “*Earlystopping*” com paciência de dez, monitorando o valor de perda. Isso evita o *overfitting* e interrompe os treinamentos quando a rede não obtém melhorias. A máquina utilizada para todos os testes possui um processador Rayzen 7 de 3.7 GHz de 8 núcleos e 16 GB de memória RAM.

3.1. Modelo com VGG16

A arquitetura VGG16 emprega 16 camadas de convolução 3x3 e pooling 2x2, resultando em modelo que ocupa 528 megabytes, de acordo com ONEIROS (2015). Utilizando essa arquitetura obtivemos 16.320.891 milhões de parâmetros, sendo que 1.605.891 são treináveis. Para treinar esse modelo foram necessárias 15 épocas, com cada época durando em média 3.410 segundos, o que acarretou em cerca de 14 horas.

3.2. Modelo com VGG19

O VGG19 é uma extensão do VGG16 com mais camadas de convolução mantendo a mesma estrutura básica, a mesma emprega 19 camadas, possuindo um tamanho de 549 megabytes em disco, segundo ONEIROS (2015). O treinamento com essa arquitetura exigiu 7 épocas para ser concluído. Cada época teve uma duração média de 4.387

segundos totalizando aproximadamente 8 horas para o treinamento completo. Durante esse processo foram utilizados 21.630.275 milhões de parâmetros, sendo que 1.605.891 eram treináveis.

3.3. Modelo com MobilNet

O MobileNet, uma arquitetura indicada para aplicações em dispositivos móveis, consiste em 55 camadas e possui um tamanho de aproximadamente 16 megabytes [ONEIROS 2015]. Por ser um modelo mais compacto, adequado para uso em dispositivos embarcados, o tempo de treinamento foi de aproximadamente duas horas, com um total de 14 épocas. Cada época teve em média uma duração de 602 segundos. Possuindo 6.272.323 milhões de parâmetros com 4.014.339 treináveis

3.4. Modelo com EfficientNetB7

Por último, o EfficientNetB7, que utiliza 438 camadas, possuindo um tamanho de 256 megabytes [ONEIROS 2015]. Esse modelo possui 72.126.106 milhões de parâmetros, sendo que 8.028.419 são treináveis, o que acarretou no maior tempo de treino dentre os modelos propostos com cerca de 28 horas em 16 épocas, ou seja, 6.208 segundos por época.

Na Tabela 1 é apresentado um comparativo das características das dimensões dos quatro modelos utilizados, a partir do treinamento deles.

Tabela 1. Comparativo entre os modelos propostos

Modelo	Tamanho (MB)	Nº de Parâmetros	Nº de Épocas (<i>early stopping</i>)	Tempo para treino (h)	Nº de camadas
VGG16	528	16.320.821	15	14	20
VGG19	549	21.630.275	7	8	23
MobileNet	16	6.272.323	14	2	59
EfficientNetB7	256	72.126.106	16	28	442

4. Resultados

4.1. Conjunto de Dados

Para a realização deste trabalho foi desenvolvido um conjunto de dados a partir de três *datasets* públicos para treinar, validar e testar os modelos propostos. O primeiro conjunto de dados utilizado foi RSNA Pneumonia Detection Challenge [Radiological Society of North America 2020], disponibilizado pela plataforma de competição Kaggle. Ele consiste em 29.687 imagens de radiografias torácicas, no formato DICOM, divididas em duas classes, normal e pneumonia. O segundo conjunto de dados foi o COVIDx CXR-3-classification [Hosseini et al. 2022], também adquirido na plataforma Kaggle, tendo 29.986 imagens no formato PNG, classificadas em duas classes: positivo e negativo para a covid-19. O último conjunto de dados utilizado foi o Chest X-Ray

Images [Kermany et al. 2018], contendo 5.857 imagens no formato JPEG, separadas em duas classes: normal e pneumonia.

O conjunto de dados desenvolvido consiste em 30.600 imagens divididas em três partes: treino (80%), teste (10%) e validação (10%) conforme mostra a Tabela 2.

Tabela 2. Distribuição do conjunto de dados obtido

Categoria	Dados de Treinamento	Dados de Validação	Dados de Teste
covid-19	8160	1020	1020
Normal	8160	1020	1020
Pneumonia	8160	1020	1020

4.2. Pré-processamento de dados

Para lidar com a falta de uniformidade dos dados, redimensionamos todas as imagens para a resolução 224 x 224 e o formato PNG com 3 canais de cores RGB. Essa abordagem é comumente utilizada para padronizar as dimensões e formato das imagens, tornando-as adequadas para a alimentação em modelos de aprendizado profundo.

4.3. Resultado dos modelos

Os resultados obtidos nesta pesquisa demonstram o desempenho dos quatro modelos de classificação de imagens de radiografia torácica desenvolvidos. Esses modelos foram treinados utilizando o conjunto de dados descrito na Seção 4.1 e os parâmetros abordados na Seção 3. A avaliação do desempenho dos modelos foi realizada a partir da acurácia, precisão, recall e f1-score para as três classes: covid-19, normal e pneumonia. Além disso, a matriz de confusão foi utilizada para fornecer mais detalhamento do desempenho dos modelos em relação às classes previstas e reais.

As Figuras 1, 2 e 3 mostram os gráficos das métricas de desempenho, precisão, recall e F1-score, das classes: covid-19, normal e pneumonia, para cada um dos modelos propostos. Observa-se que em todos os modelos a classe covid-19 obtém resultados acima dos 90% e com baixa variação em todas as métricas. Isso indica um desempenho consistente e preciso na detecção de casos dessa doença, minimizando os erros de predição para essa classe específica. No entanto, para as outras duas classes normal e pneumonia, constata-se a ocorrência de mais oscilações nas métricas.

Resultado para precisão

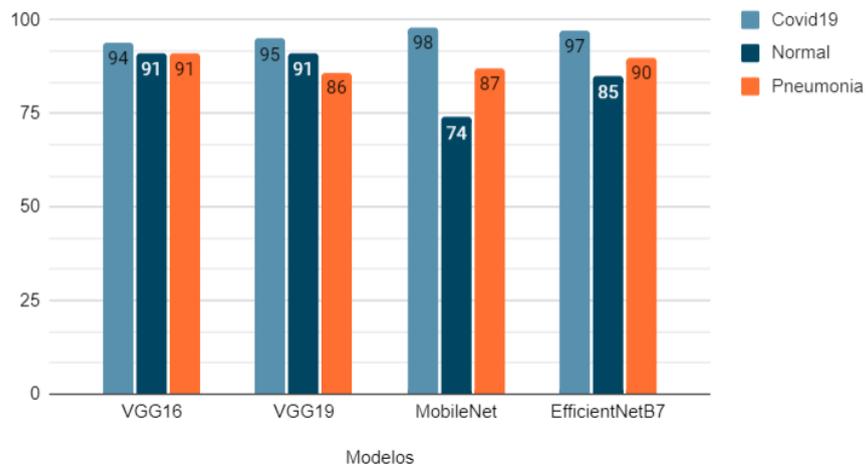


Figura 1. Gráfico da métrica de precisão para os dados de teste

Resultados para recall

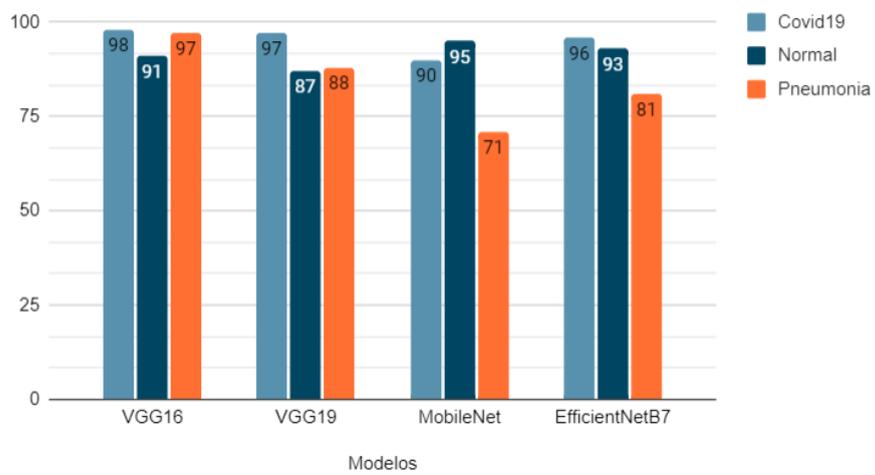


Figura 2. Gráfico da métrica de recall para os dados de teste

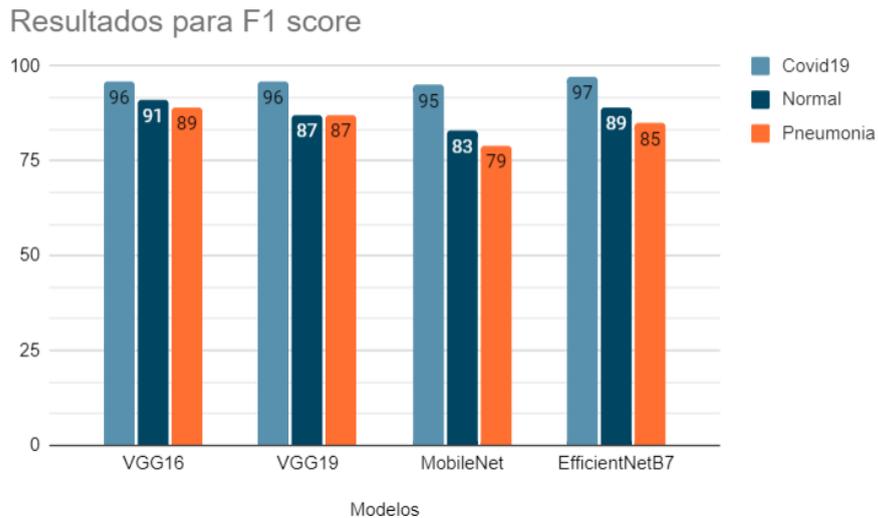


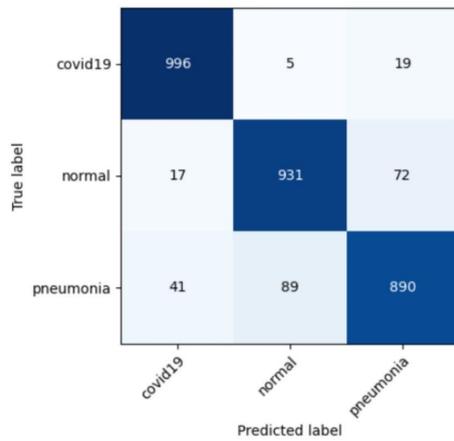
Figura 3. Gráfico da métrica de F1 score para os dados de teste

A Figura 4 apresenta as matrizes de confusão, que permitem visualizar de forma mais detalhada os acertos e erros de classificação para cada classe nos modelos propostos. Os resultados demonstrados nas matrizes corroboram com as métricas anteriormente mencionadas. Observa-se que em todos os modelos, a classe covid-19 apresenta poucas confusões, o que indica um alto nível de acerto nas previsões desses casos, entrando em acordo com os resultados das métricas de desempenho acima.

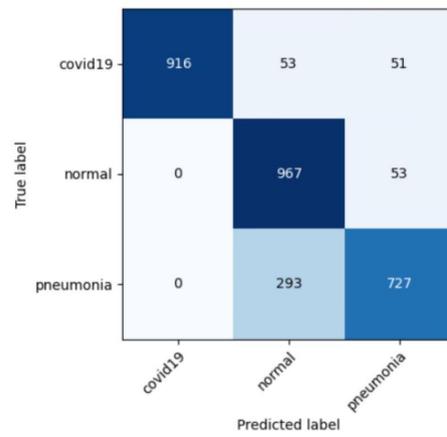
Por outro lado, é notável que a classe pneumonia apresenta bastante confusão com a classe normal em todos os modelos, especialmente no modelo que utiliza MobileNet. Indicando que os modelos apresentam mais dificuldades em discriminar pneumonia de casos normais.

A classe normal mantém uma situação intermediária, entre a classe covid-19 e a classe pneumonia em termos de erros de classificação, exceto no modelo MobileNet onde a mesma apresenta uma taxa de acertos maior que a classe covid-19.

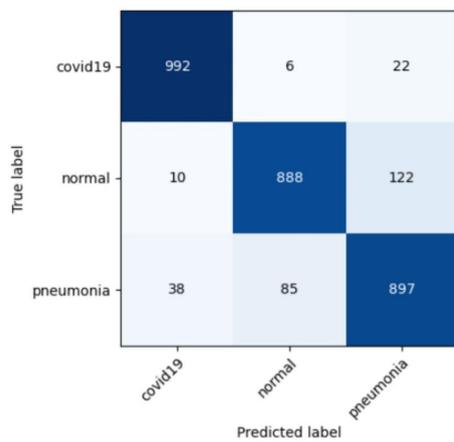
a) VGG16



e) MobileNet



b) VGG19



d) EfficientNetB7

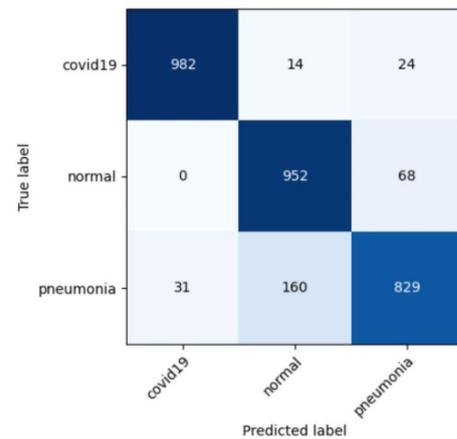


Figura 4. Matriz de confusão para os dados de teste, a) Modelo com VGG16; b) Modelo com VGG19; c) Modelo com MobileNet; d) Modelo com EfficientNetB7

Além das métricas já destacadas é importante considerar outra medida de desempenho, a acurácia, que fornece uma média geral de acertos do modelo, sendo relevante para avaliar a eficiência do mesmo. A Figura 5, exibe os resultados de acurácia de cada modelo testado. Onde é possível observar que o modelo que utiliza o VGG16 alcançou melhor acurácia em comparação com os outros modelos e MobileNet a menor acurácia. Contudo, nenhum modelo teve acurácia abaixo de 85%, indicando uma validação da metodologia de treinamento/teste e qualidade da solução independente do modelo.

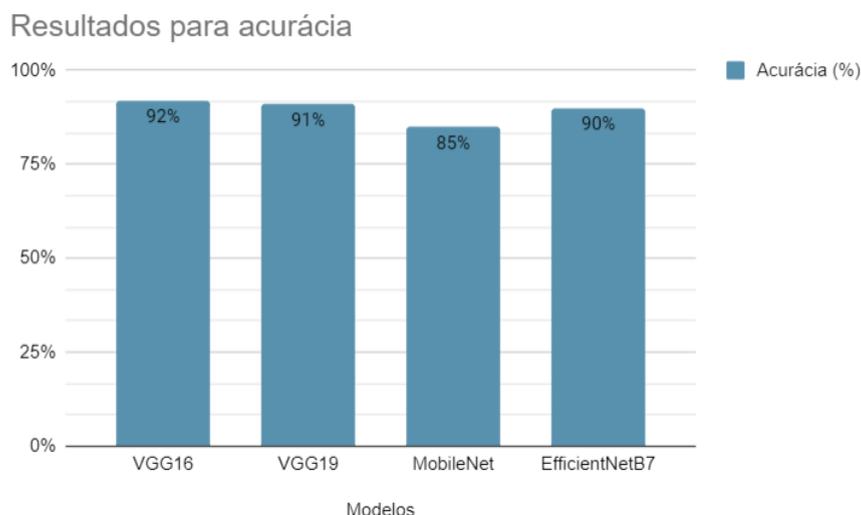


Figura 5. Gráfico de acurácia média dos quatro modelos

4.8. Comparação entre os modelos

Considerando outros aspectos é possível estabelecer uma relação de custo-benefício para determinar a eficiência e viabilidade dos modelos em cenários práticos, uma vez que, o cenário de aplicação tem necessidades específicas considerando os recursos computacionais disponíveis e as restrições de tempo. Para tanto, listamos na Tabela 3 informações acerca do tamanho em megabytes, n° de parâmetros, acurácia e tempo médio de predição/inferência para cada modelo, com desvio padrão 0,01902 para o VGG16, 0,01877 para o VGG19, 0,01871 para o MobileNet e 0,01775 para o EfficientNetB7. Ressaltando que o tempo de inferência foi medido utilizando a mesma máquina para todos os modelos, conforme configuração apresentada na Seção 3.

Tabela 3. Custo-benefício para os modelos

Modelo	Tamanho (mb)	N° de Parâmetros	Acurácia (%)	Tempo médio de Predição (ms)
VGG16	528	16.320.821	92	123
VGG19	549	21.630.275	91	155
MobileNet	16	6.272.323	85	20
EfficientNetB7	256	72.126.106	90	223

Com base na análise da tabela percebe-se que o MobileNet se destaca pela sua baixa quantidade de parâmetros, o que resulta em um tamanho compacto do modelo. Além disso, este modelo apresenta uma acurácia razoável e tempo de predição/inferência significativamente mais rápido do que o modelo com melhor acurácia (VGG16). Essa combinação torna o MobileNet mais adequado para sistemas que possuem recursos computacionais limitados, como sistemas embarcados. A

importância de ter um tempo de inferência mais baixo possível pode ser determinante para o desenvolvimento de um sistema embarcado. Para exemplificar, considere os tempos médios de predição para os modelos VGG16 e MobileNet da Tabela 3. Esses valores foram obtidos por meio dos testes em uma máquina cuja frequência de operação é 3,7 GHz (3,7 bilhões de ciclos por segundo). Logo, o modelo VGG16 que executou em 123 ms (0,123 segundos) equivale a 455.100.000 ciclos nessa máquina, enquanto o MobileNet, que executou em 20 ms (0,020 segundos), utilizou 74.000.000 ciclos dessa mesma máquina. Considerando a possibilidade de desenvolver uma aplicação embarcada com esses modelos em uma plataforma como Arduino Nano 33 BLE [Arduino 2023] a diferença de tempo entre os modelos é bem mais impactante. Tal plataforma possui frequência de operação de 64 MHz, dessa forma, se for considerado que os modelos implementados executem a mesma quantidade de ciclos, o tempo de execução desses modelos no sistema embarcado seria de 1,15 e 7,11 segundos, respectivamente no MobileNet e VGG16. Portanto, o modelo MobileNet parece ser uma escolha mais apropriada para este ambiente de computação na borda.

Considerando a presença de recursos robustos com alto poder de processamento, é possível priorizar o desempenho máximo utilizando o modelo com VGG16 que oferece a melhor acurácia entre os modelos testados, porém, é importante considerar que ele possui o maior tamanho em MB entre todos os modelos. Também é importante destacar que as medidas de tempo aqui foram úteis para efeito de comparação entre os modelos, de modo que em um ambiente mais restrito, como sistemas embarcados, o tempo de inferência deve ser maior e em um ambiente computacional mais robusto poderia reduzir todos esses tempos.

Ainda nessa perspectiva é possível notar que o modelo que faz uso do EfficientNetB7 possui acurácia muito próxima do VGG16, ao mesmo tempo em que apresenta um tempo de predição e número de parâmetros bem maior. Por último temos o VGG19 que possui uma acurácia de 91%, com o maior tamanho e segundo maior tempo de predição em comparação com os quatro modelos, nesse sentido comparado esse modelo aos outros ele seria inviável para uma arquitetura com baixo poder computacional e também não oferece um bom custo benefício para uma arquitetura mais robustas, uma vez que, o VGG16 obteve melhores resultados utilizando menos recursos.

5. Conclusões

Nessa pesquisa, foram apresentados quatro modelos CNN pré-treinados, VGG16, VGG19, MobileNet, EfficientNetB7, para classificação de imagens de raio-x de pulmão em três categorias: normal, com pneumonia e com covid-19, utilizando um conjunto de dados de 30.600 imagens obtidos da junção de três conjunto de dados públicos.

Os resultados mostraram que os modelos propostos foram capazes de realizar a classificação com acurácia próxima dos 90%. Entre os quatro modelos, o VGG16 obteve o melhor desempenho, alcançando uma taxa de acurácia de 92%. No entanto, esse modelo também apresentou o maior tamanho em MB além de possuir o segundo maior tempo de inferência em comparação aos outros modelos testados. Para utilização em

uma arquitetura cliente-servidor, onde não há restrição severa de recursos, o modelo VGG16 seria o mais indicado devido ao seu melhor desempenho, de acurácia.

No entanto, para sistemas com recursos limitados (sistemas embarcados), o modelo MobileNet, embora tenha a menor taxa de acurácia entre os modelos propostos com 85%, pode ser uma opção mais adequada, devido à sua menor quantidade de parâmetros e conseqüentemente menor quantidade de memória requerida, além do menor tempo de inferência. Para direções futuras, a detecção de covid-19 em conjunto com pneumonia surge como uma direção promissora, uma vez que, a pneumonia, uma doença oportunista, frequentemente é observada em pacientes com baixa imunidade resultante da covid-19.

Referências

- Arduino (2023) “Especificação do produto nRF52840 v.1.1”, https://content.arduino.cc/assets/Nano_BLE_MCU-nRF52840_PS_v1.1.pdf?_gl=1*16psgzk*_ga*MTMzOTczNzM5Ny4xNjkyMDQ3NTU0*_ga_NEXN8H46L5*MTY5MjA0NzU1My4xLjEuMTY5MjA0ODAwMC4wLjAuMA.., Junho.
- Chen N, Zhou M, Dong X, et al (2020). Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *Lancet* 2020;395:507-13. 10.1016/S0140-6736(20)30211-7.
- Guan WJ, Ni ZY, Liang WH, et al (2020). Clinical Characteristics of Coronavirus Disease 2019 in China. *N Engl J Med* 2020;382:1708-20. 10.1056/NEJMoa2002032
- Hernandez, D., Pereira, R. and Georgevia, P. (2020). COVID-19 detection through X-Ray chest images. *International Conference Automatics and Informatics (ICAI)*.
- Hossein, Wong, A., Gunraj, H., et al. (2022). “COVIDx CXR-3”, <https://www.kaggle.com/datasets/andyczhao/covidx-cxr2>. Fevereiro.
- Kermany, D., Zhang, K. and Goldbaum, M. (2018). Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification. v. 2.
- Mohammadi, R., Salehi M., Ghaffari, H., et al. (2020). Transfer Learning-Based Automatic Detection of Coronavirus Disease 2019 (COVID-19) from Chest X-ray Images. *Journal of Biomedical Physics and Engineering*, p.559-568.
- ONEIROS (2015). “Keras API Reference – Applications”. 2015. <https://keras.io/api/>
- Organização Mundial da Saúde (2023) “COVID-19 Weekly Epidemiological Update”, <https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19--8-june-2023>, Junho.
- Ozturk, T., Talo, M., Yildirim, E., Baloglu U., Yildirim. et al. (2020) Automated detection of COVID-19 cases using deep neural networks with X-ray images. *Computers in Biology and Medicine*, v. 121.
- Radiological Society of North America (2018) “RSNA Pneumonia Detection Challenge”, <https://www.kaggle.com/competitions/rsna-pneumonia-detection-challenge/data>. Agosto.

- Tangudu, V., Kakarla, J. and Venkateswarlu, I. (2022) "COVID-19 detection from chest x-ray using MobileNet and residual separable convolution block", *Soft Computer*, p. 2197-2208
- Wang, L., Lin, Z. and Wong, A., (2020) "COVID-Net: a tailored deep convolutional neural network design for detection of COVID-19 cases from chest X-ray images", *Sci Rep.* v. 10.
- Yang, D., Martinez, C., Visuña, L. et al. (2021) "Detection and analysis of COVID-19 in medical images using deep learning techniques", *Sci Rep.* v. 11.