

Arquiteturas de Redes Neurais Convolucionais para Identificação de Pneumonia e COVID-19 em Raios X de Tórax

Daniel Henrique Comério* Leonardo Xavier Barbosa*
Sabrina Siqueira Panceri** Francisco de Assis Boldt* Filipe Mutz*

* Instituto Federal do Espírito Santo (IFES) - Campus Serra, ES
(e-mails: {danielhcomerio,leoshap}@gmail.com,
{filipe.mutz,franciscoa}@ifes.edu.br)

** Instituto Federal do Espírito Santo (IFES) - Campus Guarapari, ES,
(e-mail: sabrina.panceri@ifes.edu.br)

Abstract: This work compares architectures of convolutional neural networks for the identification of pneumonia and COVID-19 in chest X-rays. The models considered in the evaluation were ResNet-50, MobileNet-v2, Inception-v3, and EfficientNet-B2, besides ensembles built using these neural networks. We also propose and evaluate a preprocessing technique to reduce variations in the images. Experimental results show that MobileNet-v2 achieves the highest individual performance with accuracy of 94.03%, precision of 94.59%, recall of 91.91%, and f1-score of 91.55%. The ensemble with highest performance achieved accuracy of 96.00%, precision of 94.61%, recall of 95.53%, and f1-score of 95.04%. Integrated gradients analyses showed that models focus on regions that make sense when performing the predictions.

Resumo: Este trabalho compara arquiteturas de redes neurais convolucionais para a identificação de pneumonia e COVID-19 em imagens de raios X de tórax. Os modelos considerados na avaliação foram ResNet-50, MobileNet-v2, Inception-v3 e EfficientNet-B2, além de *ensembles* construídos usando estas redes neurais. Nós ainda propomos e avaliamos uma técnica de pré-processamento para reduzir a variação das imagens. Resultados experimentais mostraram que a MobileNet-v2 alcançou a melhor performance individual com acurácia de 94,03%, precisão de 94,59%, revocação de 91,91% e f1-Score de 91,55%. O *ensemble* com maior performance alcançou acurácia de 96,00%, precisão de 94,61%, revocação de 95,53% e f1-score de 95,04%. Análises de gradientes integrados mostraram que modelos se concentram em regiões que fazem sentido para produzir as predições.

Keywords: Convolutional Neural Networks; X-Ray; COVID-19; Pneumonia.

Palavras-chaves: Redes Neurais Convolucionais; Raios X; COVID-19; Pneumonia.

1. INTRODUÇÃO

Após um ano de pandemia de COVID-19, a comunidade científica se mantém em busca de inovações para agilizar o diagnóstico dos pacientes e diminuir a proliferação da doença. De acordo com o relatório de acompanhamento da Organização Mundial da Saúde (OMS), até maio de 2021 mais de 150 milhões de casos e mais de 3 milhões de mortes pelo Coronavírus haviam sido confirmados no mundo. No Brasil, foram registrados mais de 15 milhões de casos e mais de 400 mil mortes (World Health Organization, 2020).

* Agradecemos à FAPES e a CAPES pelo apoio financeiro dado por meio do PDPG (Parcerias Estratégicas nos Estados da CAPES) (PROCESSO: 2021-2S6CD, TO/nº FAPES: 132/2021). Também agradecemos ao Propós (Programa Institucional de Apoio à Pós-graduação Stricto Sensu) do IFES pela apoio financeiro. Filipe Mutz agradece ao Instituto Federal do Espírito Santo (IFES) por incentivar sua pesquisa via o Programa Pesquisador de Produtividade (PPP) - portaria n. 1072 de 21 de maio de 2020.

O vírus causador da COVID-19 é altamente infeccioso e de fácil propagação. Mesmo pessoas que não manifestam os sintomas podem estar contaminadas e atuar como transmissores da doença (World Health Organization, 2020). De acordo com Ai et al. (2020), os principais sintomas relatados são: febre, tosse, falta de ar e dores musculares.

Enquanto os exames de imagem do pulmão, frequentemente, revelam pneumonia bilateral e opacidade em vidro fosco, Wang et al. (2020) explica que a análise dos exames radiográficos, raios X ou tomografia computadorizada (TC) de pulmão, por radiologistas constatou artefatos visuais específicos nos pacientes infectados pelo Coronavirus. Esta observação sugere que, em alguns casos, este tipo de exame pode ser utilizado como ferramenta primária para rastreamento e identificação dos casos positivos, agilizando assim o diagnóstico e início do tratamento.

Com objetivo de acelerar a identificação de pacientes contaminados, foram conduzidas diversas pesquisas sobre o uso de inteligência artificial para identificação de COVID-

19 em exames de raios X (Roberts et al., 2021) e TC (Ulhaq et al., 2020).

Este trabalho contribui com a pesquisa nesta área comparando técnicas de pré-processamento, além de arquiteturas e ensembles de redes neurais convolucionais para a identificação de pneumonia e COVID-19 em raios X de tórax. Foi realizado *transfer learning* de modelos pré-treinados na base de dados ImageNet (Deng et al., 2009) para a tarefa objetivo e todos os pesos foram ajustados durante o treinamento.

Experimentos mostraram que a configuração de modelo único com maior performance é a rede MobileNet-v2 sem pré-processamento que alcançou acurácia de 94,03%, precisão de 94,59%, revocação de 91,91% e *f1-Score* de 91,55%. O *ensemble* com maior performance usou os modelos alcançou acurácia de 96,00%, precisão de 94,61%, revocação de 95,53% e *f1-score* de 95,04%. Análises de gradientes integrados mostraram que a região dos pulmões tem a maior influência no cálculo das saídas das redes neurais. Este resultado mostra que as redes aprenderam a detectar características dos exames que fazem sentido para a tarefa.

O restante do trabalho está organizado da seguinte forma: a Seção 2 apresenta trabalhos anteriores relacionados ao tema de pesquisa. A Seção 3 descreve a construção e pré-processamento da base de dados, assim como as arquiteturas de redes neurais utilizadas. A Seção 4 traz a metodologia experimental e discute os resultados obtidos. Finalmente, a Seção 5 apresenta as principais conclusões obtidas no trabalho e direções de trabalhos futuros.

2. TRABALHOS RELACIONADOS

Wang et al. (2020) usaram a técnica de *generative synthesis* para produzir uma rede neural capaz de classificar raios x do tórax como saudáveis, com pneumonia ou com COVID-19 e que obtivesse alta sensibilidade e valor preditivo positivo na identificação de COVID-19. Eles utilizaram uma versão da mesma base de dados usada neste trabalho. Os autores alcançaram acurácia e sensibilidade de 93.3% e 91%, respectivamente.

Ozturk et al. (2020) propôs um modelo que usou a DarkNet como um classificador e obteve uma acurácia de 98.08% para classificação binária (COVID-19 ou não) e 87.02% para classificação ternária com as mesmas classes deste trabalho.

Khan et al. (2020) propôs um modelo baseado na Xception para detectar COVID-19 que alcançou acurácia de 89.6% e revocação de 93% para COVID-19, sendo que o modelo foi treinado para classificar raios x do tórax em saudável, com COVID-19, com pneumonia viral ou pneumonia bacteriana.

Narin et al. (2021) avaliou três arquiteturas de redes neurais (ResNet50, InceptionV3 and Inception-ResNetV2) e obteve acurácia de 98% com uma ResNet50 treinada para classificação em normal ou com COVID-19. Os autores não incluíram casos de pneumonia nos experimentos.

Wang et al. (2021) propuseram um sistema composto por duas redes neurais convolucionais para resolver a mesma tarefa estudada neste trabalho. O primeiro extrai caracte-

rísticas dos pulmões e classifica os raios x, enquanto o segundo localiza em qual(is) pulmão(ões) está(ão) as lesões. O sistema alcançou acurácia de 93.65%, sensibilidade de 90.92% e especificidade de 92.62%. Esta performance foi superior ao alcançado em média por radiologistas.

As contribuições deste trabalho em relação à literatura são a avaliação de arquiteturas de redes neurais diferentes daquelas usadas em trabalhos anteriores, os efeitos de usar a técnica de pré-processamento e a análise dos potenciais benefícios de usar *ensembles* para na tarefa estudada.

3. DESENVOLVIMENTO

As redes neurais convolucionais recebem como entrada uma imagem de raios X de tórax e produzem como saída a probabilidade daquela imagem ser de um indivíduo saudável, com pneumonia ou com COVID-19. As próximas seções descrevem as metodologias para construção da base de dados, treinamento e avaliação dos modelos.

3.1 Base de Dados

A base de dados foi construída unindo as cinco bases de dados públicas *COVID-19 Image Data Collection* (Cohen et al., 2020), *Figure 1 COVID-19 Chest X-ray Dataset Initiative* (COVID-Net Team, 2021b), *Actualmed COVID-19 Chest X-ray Dataset Initiative* (COVID-Net Team, 2021a), *RSNA Pneumonia Detection Challenge Dataset* (Radiological Society of North America and Kangle, 2018) e *COVID-19 Radiography database* (Haghanifar et al., 2020). A mesma estratégia de construção da base de dados foi usada por Wang et al. (2020). Infelizmente, dados demográficos e de estados clínicos dos pacientes estão incompletos nas bases de dados e, portanto, não é possível apresentar a distribuição de amostras considerando estas características.

Algumas operações de filtragem foram aplicadas para limpar a base de dados. Imagens duplicadas foram removidas e não foram utilizadas imagens com dados de anotação incompletos, de outros tipos de exames (e.g., TC), ou com diagnóstico diferente de normal, com pneumonia ou com COVID-19. Outras doenças não foram consideradas pois, o número de imagens para estas doenças era pequeno (menor que uma centena). Após esta filtragem, a base de dados resultante contém 10.340 (52,39 %) casos normais, 7.601 (38,51 %) casos de pneumonia e 1.797 (09,10 %) casos de COVID-19. A base foi dividida de forma estratificada em conjuntos de treino e teste (90% para treino e 10% para teste). Posteriormente, o conjunto de treino foi subdividido de forma estratificada em treino e validação na mesma proporção.

A primeira linha da Figura 1 exemplifica as amostras existentes na base de dados. Como pode ser observado, as imagens possuem diferentes níveis de contraste e brilho. Além disso, posição, orientação, tamanho e perspectiva dos pulmões variam significativamente entre as radiografias.

3.2 Pré-Processamento

A grande variação entre as imagens poderia tornar a tarefa de aprendizado desnecessariamente difícil para as redes neurais. Assim, foi avaliado se o uso de processamento

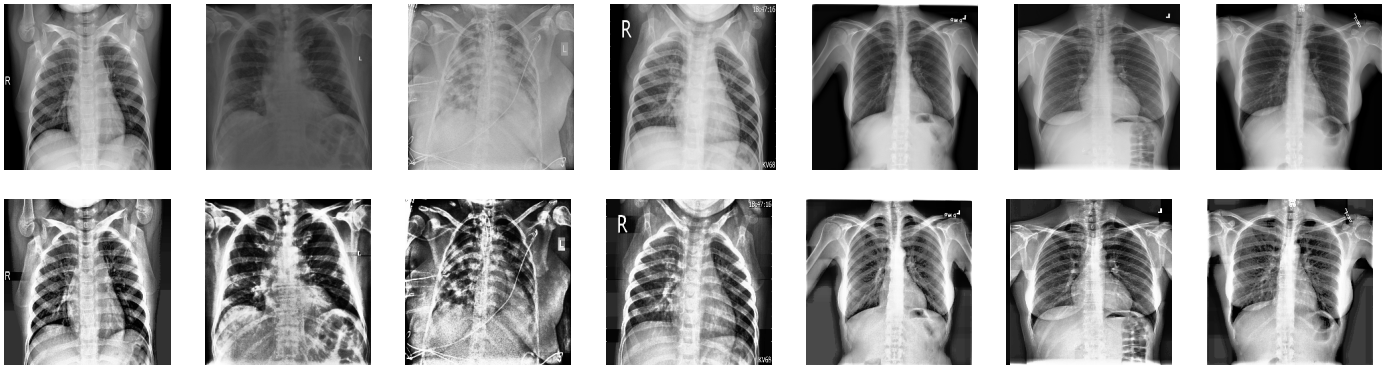


Figura 1. Amostras de exames de raios X da base de dados utilizada no trabalho. A primeira linha contém as imagens cruas e a segunda linha mostra as mesmas imagens após o pré-processamento.

de imagens para reduzir esta variação levaria a ganhos de performance. A primeira operação foi o realce de detalhes proposto por Mehrotra et al. (2014). Dada uma imagem de entrada I , a imagem de saída I' é calculada por:

$$I' = I + TopHat(I) - BlackHat(I) \quad (1)$$

na qual, $TopHat$ e $BlackHat$ são operações morfológicas tais que $TopHat$ é a diferença entre a imagem e sua abertura e $BlackHat$ é a diferença entre o fechamento da imagem e a imagem original (Serra and Soille, 2012). Estas operações foram realizadas usando *kernels* de 32×32 pixels.

Em seguida, os valores dos pixels da imagem resultante foram normalizados e multiplicados por 255, de forma que todos os pixels estivessem no intervalo $[0, 255]$. A normalização foi feita subtraindo de todos pixels o valor mínimo e, depois, dividindo todos os valores pelo máximo. As imagens foram então convertidas para escala de cinza e passaram por uma equalização de histograma. Por fim, elas foram redimensionadas para os tamanhos de entrada esperados pelas redes neurais.

O resultado do pré-processamento é ilustrado na segunda linha da Figura 1. Como pode ser observado, as operações reduziram a diferença de contraste e brilho nas imagens. Além disso, elas tornaram os ossos mais evidentes o que permite a identificação mais clara de manchas sobre a região dos pulmões.

3.3 Arquiteturas de Redes Neurais

As redes neurais convolucionais consideradas no trabalho foram ResNet-50 (He et al., 2016), MobileNet-v2 (Sandler et al., 2018), EfficientNet-B2 (Tan and Le, 2019) e Inception-v3 (Szegedy et al., 2016). A ResNet-50 foi escolhida por ser uma rede de referência tipicamente utilizada como *baseline* na literatura. A MobileNet-v2 é eficiente do ponto de vista de uso de recursos computacionais. Ela é uma alternativa economicamente vantajosa para organizações com recursos computacionais limitados. Por fim, os modelos EfficientNet-B2 e Inception-v3 foram selecionados devido à sua alta performance na base de dados ImageNet (Deng et al., 2009). Estudos apontam que existe correlação positiva entre a performance nesta base de dados e a

performance em aprendizado por transferência (*transfer learning*) para tarefas objetivo (Kornblith et al., 2019).

Para evitar o custo de treinar redes neurais do zero, foram utilizadas redes pré-treinadas usando a base de dados ImageNet. O treinamento foi realizado por 40 épocas usando o algoritmo Adam (Goodfellow et al., 2016) com taxa de aprendizado de 10^{-4} e todos os parâmetros das redes foram adaptados.

Buscando aumentar a generalização dos modelos, a técnica de aumento de dados (*data augmentation*) foi utilizada para produzir novas amostras a partir da base de dados inicial. As transformações aplicadas foram espelhamento horizontal, rotação aleatória escolhida de forma uniforme no intervalo $[-36, 36]$ graus (10% de 360 graus), ajuste de contraste com fator de 0% – 20%, translação aleatória de 0% – 10% em cada direção e zoom aleatório com fator de 0% – 20%.

3.4 Ensembles

Além da avaliação de redes neurais individuais, foram realizados testes com *ensembles*. Dada uma imagem de raios X, as predições das redes que compõem o *ensembles* são calculadas e estratégias de sumarização são usadas para produzir uma saída única a partir das respostas individuais. As redes foram treinadas individualmente e então usadas para compor o *ensemble* sem treinamento conjunto.

Foram avaliadas três estratégias de sumarização, cada uma com um princípio de funcionamento próprio. A estratégia mais usada na literatura é retornar a classe mais votada (chamaremos de MaxVotos) pelos classificadores. Ela assume que o *ensemble* opera como um comitê de pares com aproximadamente a mesma competência na classificação das amostras. Classificadores que discordam da maioria são considerados *outliers* e ignorados.

A segunda estratégia considerada é calcular a média das saídas dos modelos (chamaremos de Média). Ela assume que modelos podem ser especialistas em alguns tipos de amostras. Modelos que possuem confiança nas predições atribuem probabilidade próxima ao máximo para uma das classes, enquanto modelos com pouca confiança tendem a distribuir a probabilidade entre duas ou mais classes. Esta

estratégia é similar à uma votação com pesos dados pela confiança das redes.

Por fim, a terceira estratégia é retornar a classe cuja probabilidade é máxima (chamaremos de MaxProb) dentre as saídas de todos os classificadores. Esta estratégia possui uma hipótese de operação similar à anterior. Ela assume que os modelos são especialistas e que a resposta do classificador com maior confiança deve ser a vencedora.

Para todos os métodos de sumarização, se houver empate entre duas ou mais classes, a saída é escolhida aleatoriamente (de forma uniforme) dentre as classes que empataram.

4. EXPERIMENTOS

Esta seção descreve os experimentos realizados e discute os resultados alcançados. O código-fonte foi desenvolvido usando a linguagem Python e as bibliotecas Tensorflow (versão 2.4.1) e Scikit-Learn (versão 0.24.2). O código é *open source* e para garantir a reprodutibilidade do trabalho pode ser acessado em https://github.com/danielcomerio/2021-SBAI-COVID19_em_RaioX/.

4.1 Performance de Redes Neurais Individuais

Este experimento teve como objetivo comparar as redes neurais convolucionais individualmente na tarefa de classificar imagens de raios X como sendo de indivíduos saudáveis, acometidas por pneumonia ou portadoras de COVID-19. O objetivo da comparação foi identificar qual arquitetura seria mais adequada para ser utilizada em um sistema de diagnóstico assistido por computador. Um objetivo secundário do experimento foi analisar se o pré-processamento das imagens (seção 3.2) levava a modelos com melhor performance. Para isto, foram avaliadas as saídas dos modelos com e sem o uso do pré-processamento. Um grupo de modelos foi treinado e testado com imagens sem pré-processamento e um segundo grupo treinado e testado com imagens pré-processadas. As métricas de performance utilizadas foram acurácia (ACC) e as médias macro entre classes de precisão (PRC), revocação (RVC) e *f1-score* (F1). Os resultados alcançados são apresentados na Tabela 1. As matrizes de confusão foram omitidas por limitações de espaço, mas elas podem ser visualizadas no repositório do projeto.

Como pode ser observado, a rede que obteve a melhor performance foi a MobileNet-v2 sem pré-processamento. A rede alcançou acurácia, precisão, revocação e *f1-score* de 94,03%, 94,59%, 91,91% e 93,16%, respectivamente. A boa performance da MobileNet-v2 associada ao fato de que ela é a rede mais computacionalmente eficiente dentre as avaliadas indica que a MobileNet-v2 é o melhor modelo individual para implantação.

Os resultados experimentais mostraram ainda que o pré-processamento das imagens não contribuiu para melhoria de performance dos modelos. Ao contrário, houve redução na taxa de acerto de todos os modelos, com uma redução mais expressiva no modelo ResNet-50 onde a acurácia caiu de 92,36% para 70,60%.

Apesar da redução na acurácia de todos os modelos que utilizaram o pré-processamento, dois casos mere-

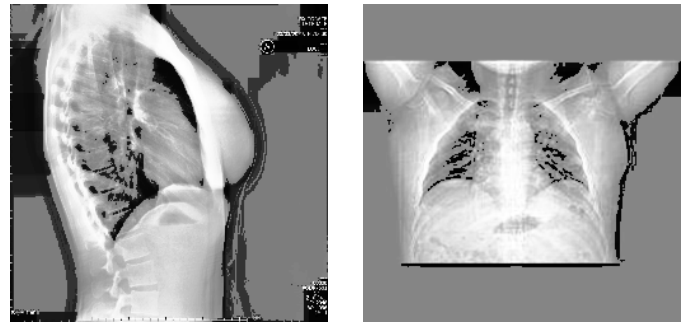


Figura 2. Amostras de COVID-19 não identificadas pelo modelo ResNet-50 com pré-processamento.

Tabela 1. Performance de Redes Neurais

Modelo	ACC	PRC	RVC	F1
sem pré-processamento				
ResNet-50	92,36%	93,99%	87,99%	90,54%
MobileNet-v2	94,03%	94,59%	91,91%	93,16%
EfficientNet-B2	93,32%	94,69%	89,19%	91,55%
Inception-v3	92,91%	92,50%	89,58%	90,89%
com pré-processamento				
ResNet-50	70,60%	71,36%	78,73%	66,51%
MobileNet-v2	92,21%	91,38%	87,98%	89,52%
EfficientNet-B2	82,74%	76,26%	86,67%	77,31%
Inception-v3	92,26%	93,82%	86,34%	89,37%

cem destaque. A revocação da EfficientNet-B2 com pré-processamento na classe COVID-19 foi de 96,68%. Dos 181 casos de COVID-19, apenas 6 erros foram cometidos, sendo 2 preditos como normais e 4 preditos como pneumonia. O modelo ResNet-50 com pré-processamento alcançou revocação para classe COVID-19 de 98,89%. Das 181 imagens da classe COVID-19, apenas as duas apresentadas na Figura 2 foram classificadas incorretamente, uma como normal e outra como pneumonia. Como pode ser observado, ambas as imagens representam visões não usuais dos pulmões e que foram afetadas negativamente pelo pré-processamento. Estes dois modelos alcançaram a maior revocação para a classe COVID-19 dentre todos os modelos individuais avaliados. Portanto, eles seriam os mais adequados para uso em sistemas de triagem nos quais pessoas classificadas como tendo COVID-19 seriam submetidas à exames mais detalhados.

Diante do exposto, podemos concluir que o pré-processamento em geral causa redução da performance preditiva dos modelos, mas em alguns casos leva a modelos mais seguros (no sentido de alta revocação) para a identificação da COVID-19.

4.2 Performance de Ensembles de Redes Neurais

Este experimento comparou a performance de *ensembles* construídos usando os modelos individuais como classificadores base. As mesmas métricas apresentadas na seção anterior foram usadas para comparar os modelos. Foram construídos quatro tipos de *ensembles*. O primeiro é composto apenas pelas redes treinadas com imagens sem pré-processamento e foi chamado de S/PProc. O segundo é composto apenas pelas redes treinadas com pré-processamento e foram chamados de C/PProc. O terceiro utilizou ambos os grupos e, por isso, foi chamado de S/+C/. Todas as redes são usadas para construir os *en-*

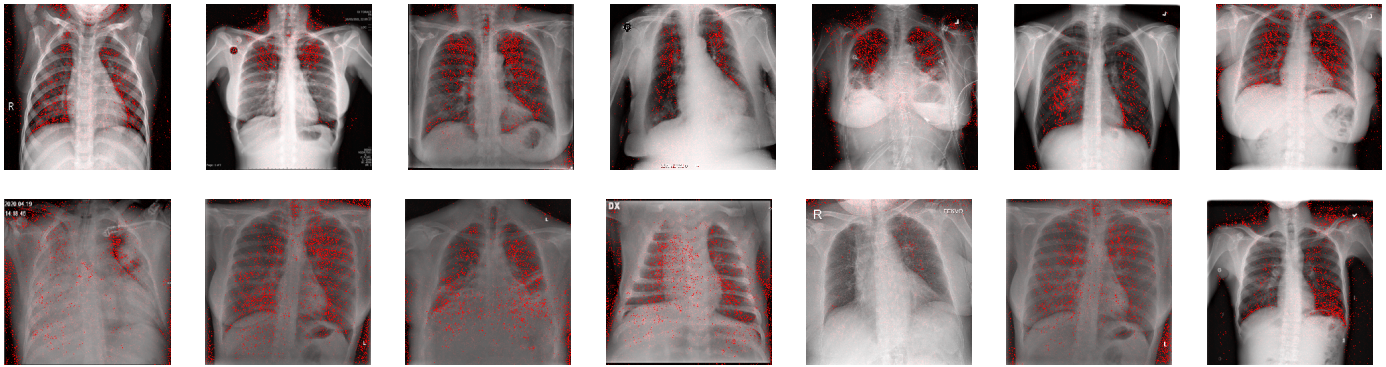


Figura 3. Mapas de calor representando a influência dos pixels na classificação produzida pela rede neural para subconjunto de imagens de teste. A primeira linha contém imagens classificadas corretamente, enquanto a segunda linha contém imagens classificadas de forma incorreta.

Tabela 2. Performance dos *ensembles*

TIPO	SUM	ACC	PRC	RVC	F1
S/ PProc	MaxProb	94,53%	95,66%	91,79%	93,55%
S/ PProc	MaxVotos	94,08%	94,66%	91,18%	92,77%
S/ PProc	Média	94,58%	95,68%	91,59%	93,44%
C/ PProc	MaxProb	91,75%	85,28%	93,41%	87,99%
C/ PProc	MaxVotos	90,49%	83,89%	90,48%	86,17%
C/ PProc	Média	93,83%	89,77%	94,13%	91,63%
S/ + C/	MaxProb	94,53%	91,54%	94,50%	92,87%
S/ + C/	MaxVotos	94,84%	95,25%	92,96%	94,03%
S/ + C/	Média	95,34%	95,94%	93,95%	94,89%
Busca	MaxProb	94,74%	90,23%	95,33%	92,35%
Busca	MaxVotos	95,14%	93,47%	93,65%	93,54%
Busca	Média	96,00%	94,61%	95,53%	95,04%

sembles deste tipo. As imagens originais são usadas como entrada para o grupo S/PProc e imagens pré-processadas são usadas como entrada para o grupo C/PProc.

Por fim, para identificar um limite superior para a performance do *ensemble* foi realizado um procedimento de busca no qual foram avaliados *ensembles* com todas as combinações possíveis de modelos e com a possibilidade de atribuir pesos 0, 1 ou 2 para cada modelo. A configuração com melhor acurácia usou os modelos MobileNet-v2 sem pré-processamento e peso 2, Inception-v3 sem pré-processamento e peso 1, e EfficientNet-B2 com pré-processamento e peso 2.

Os resultados alcançados pelos *ensembles* são apresentados na Tabela 2. A coluna TIPO indica o tipo de *ensemble*, enquanto a coluna SUM representa a estratégia de sumarização utilizada. O valor “Busca” na coluna TIPO foi usado para indicar que aqueles resultados se referem ao melhor modelo obtido usando a estratégia de busca descrita acima.

Pode-se verificar que *ensembles* alcançam performance superior aos modelos individuais. Vale observar, contudo, que eles também possuem maior custo computacional uma vez que cada imagem de raios X precisa passar por todas as redes neurais que compõem o *ensemble*. Quanto às estratégias de sumarização, pôde-se verificar que a estratégia de média obteve resultados superiores às demais.

Como esperado, *ensembles* compostos de modelos com pré-processamento (C/PProc) obtiveram resultados inferiores em relação aqueles sem pré-processamento (S/PProc). Ao analisar as matrizes de confusão dos *ensembles* de

modelos com pré-processamento, percebemos que estes foram os que alcançaram maior revocação para a classe COVID-19. Outra questão a se notar, é que os *ensembles* S/PProc e C/PProc alcançaram performance igual ou inferior aos mistos (S/+C/ e Busca). Isto indica que as imagens com pré-processamento e sem são, de certa forma, complementares e levam à modelos que aprendem a identificar características distintas nos exames.

Os *ensembles* obtidos pelo processo de busca alcançaram a maior performance dentre as configurações avaliadas atingindo valores de acurácia, precisão, revocação e f1-score de 96,00%, 94,61%, 95,53% e 95,04%, respectivamente.

4.3 Análise de Gradientes Integrados

Redes neurais podem alcançar boa performance preditiva encontrando padrões nos dados que não necessariamente representam regras de interesse para a tarefa em questão (Zhang and Zhu, 2018). Uma forma tipicamente usada na literatura (Hooker et al., 2018) para ganhar intuição sobre os mecanismos usados pelas redes neurais é analisar quais regiões da entrada causam maior impacto na saída.

Assim, este experimento teve como objetivo avaliar se as redes neurais treinadas estão considerando partes relevantes das radiografias no processo de tomada de decisão. O método dos gradientes integrados (Hooker et al., 2018) foi usado para produzir mapas de calor representando as regiões da entrada que mais influenciam as saídas das redes neurais.

Na Figura 3 são apresentados mapas de calor produzidos usando a MobileNet-v2, o modelo individual com melhor performance, para um subconjunto das imagens de teste. A primeira linha da figura contém imagens que a rede neural classificou corretamente, enquanto a segunda linha contém imagens classificadas incorretamente. Como pode ser observado nas imagens classificadas de maneira correta, o foco da rede está em ambos os pulmões, já nas imagens classificadas de maneira incorreta, a região de interesse inclui em algumas situações os ombros e a lateral do abdômen. Isto pode indicar que no processo de treinamento, a rede aprendeu um padrão que permite prever a doença a partir destas regiões. Como estes padrões são espúrios, as predições feitas usando eles no conjunto de teste levam à classificações incorretas.

5. CONCLUSÃO

Este trabalho comparou arquiteturas de redes neurais convolucionais na tarefa de identificar pneumonia e COVID-19 em imagens de raios X de tórax. Resultados experimentais mostraram que o modelo com melhor performance foi a MobileNet-v2 alcançando valores de acurácia, precisão, revocação e *f1-score* de 94,03%, 94,59%, 91,91% e 93,16%, respectivamente. Contrariando as perspectivas, o pré-processamento das imagens de treino e teste não implicou em melhoria da performance em todas as redes analisadas. O *ensemble* com melhor performance alcançou acurácia, precisão, revocação e *f1-score* de 96,00%, 94,61%, 95,53% de 95,04%, respectivamente.

A avaliação dos modelos usando uma divisão simples em treino, validação e teste pode em princípio levar à resultados enviesados, tanto positivamente (otimistas) quanto negativamente (pessimistas). Este efeito tende a ser menos significativo em grandes bases de dados como a base utilizada neste trabalho uma vez que o conjunto de teste aproxima a população total para grandes quantidades de amostras. Mesmo assim, para confirmar os resultados alcançados, em trabalhos futuros será realizada a avaliação usando a técnica de *k-fold cross-validation*.

Como apontado em trabalhos recentes (Tartaglione et al., 2020; Maguolo and Nanni, 2021), mesmo que as redes neurais possuam valores significativos para as métricas avaliadas, não é possível afirmar que elas possuem performance diagnóstica sem um estudo clínico de sua utilização. Esperamos que este trabalho possa motivar tais estudos futuros por grupos de pesquisa da área médica.

REFERÊNCIAS

- Ai, T. et al. (2020). Correlation of chest ct and rt-pcr testing for coronavirus disease 2019 (covid-19) in china: a report of 1014 cases. *Radiology*, 296(2), E32–E40.
- Cohen, J.P. et al. (2020). Covid-19 image data collection: Prospective predictions are the future. *arXiv 2006.11988*. URL <https://github.com/ieee8023/covid-chestxray-dataset>.
- COVID-Net Team (2021a). Actualmed covid-19 chest x-ray dataset initiative. URL <https://github.com/agchung/Actualmed-COVID-chestxray-dataset>.
- COVID-Net Team (2021b). Figure 1 covid-19 chest x-ray dataset initiative. URL <https://github.com/agchung/Figure1-COVID-chestxray-dataset>.
- Deng, J. et al. (2009). Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, 248–255. Ieee.
- Goodfellow, I. et al. (2016). *Deep learning*, volume 1. MIT press Cambridge.
- Haghanifar, A. et al. (2020). Covid-cxnet: Detecting covid-19 in frontal chest x-ray images using deep learning. URL <https://github.com/armiro/COVID-CXNet>.
- He, K. et al. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.
- Hooker, S. et al. (2018). A benchmark for interpretability methods in deep neural networks. *arXiv preprint arXiv:1806.10758*.
- Khan, A.I. et al. (2020). Coronet: A deep neural network for detection and diagnosis of covid-19 from chest x-ray images. *Computer Methods and Programs in Biomedicine*, 196, 105581.
- Kornblith, S. et al. (2019). Do better imagenet models transfer better? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2661–2671.
- Maguolo, G. and Nanni, L. (2021). A critic evaluation of methods for covid-19 automatic detection from x-ray images. *Information Fusion*.
- Mehrotra, A. et al. (2014). Blood vessel extraction for retinal images using morphological operator and kcn clustering. In *2014 IEEE International Advance Computing Conference (IACC)*, 1142–1146. IEEE.
- Narin, A. et al. (2021). Automatic detection of coronavirus disease (covid-19) using x-ray images and deep convolutional neural networks. *Pattern Analysis and Applications*, 1–14.
- Ozturk, T. et al. (2020). Automated detection of covid-19 cases using deep neural networks with x-ray images. *Computers in biology and medicine*, 121, 103792.
- Radiological Society of North America and Kangle (2018). Rsn pneumonia detection challenge. URL <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/data>.
- Roberts, M. et al. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for covid-19 using chest radiographs and ct scans. *Nature Machine Intelligence*, 3(3), 199–217.
- Sandler, M. et al. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 4510–4520.
- Serra, J. and Soille, P. (2012). *Mathematical morphology and its applications to image processing*, volume 2. Springer Science & Business Media.
- Szegedy, C. et al. (2016). Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2818–2826.
- Tan, M. and Le, Q. (2019). Efficientnet: Rethinking model scaling for convolutional neural networks. In *International Conference on Machine Learning*, 6105–6114. PMLR.
- Tartaglione, E. et al. (2020). Unveiling covid-19 from chest x-ray with deep learning: a hurdles race with small data. *International Journal of Environmental Research and Public Health*, 17(18), 6933.
- Ulhaq, A. et al. (2020). Computer vision for covid-19 control: A survey. *arXiv preprint arXiv:2004.09420*.
- Wang, L. et al. (2020). Covid-net: a tailored deep convolutional neural network design for detection of covid-19 cases from chest x-ray images. *Scientific Reports*, 10(1), 19549. doi:10.1038/s41598-020-76550-z. URL <https://doi.org/10.1038/s41598-020-76550-z>.
- Wang, Z. et al. (2021). Automatically discriminating and localizing covid-19 from community-acquired pneumonia on chest x-rays. *Pattern recognition*, 110, 107613.
- World Health Organization, W. (2020). Coronavirus disease (covid-19) pandemic. URL <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>.
- Zhang, Q.s. and Zhu, S.C. (2018). Visual interpretability for deep learning: a survey. *Frontiers of Information Technology & Electronic Engineering*, 19(1), 27–39.