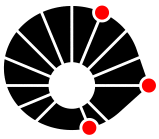
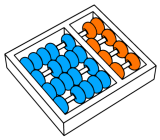




Análise de Vieses em Modelos de Redes Neurais Profundas para Detecção de Pneumonia

Luis Felipe Hamada Serrano *Sandra Eliza Fontes de Avila*

Relatório Técnico - IC-PFG-21-37
Projeto Final de Graduação
2021 - Agosto

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE COMPUTAÇÃO

The contents of this report are the sole responsibility of the authors.
O conteúdo deste relatório é de única responsabilidade dos autores.

Análise de Vieses em Modelos de Redes Neurais Profundas para Detecção de Pneumonia

Luís Felipe Hamada Serrano* Sandra Eliza Fontes de Avila†

Resumo

Conjuntos de dados de imagens médicas são pequenos em relação aos tipicamente utilizados para a aprendizagem profunda, ainda assim as arquiteturas de aprendizado profundo apresentam os melhores resultados em múltiplos conjuntos de dados. Contudo, os resultados de classificadores para tal problema são possivelmente otimistas, em razão da presença de vieses (correlações espúrias) nos dados utilizados para treinamento dos modelos. Neste trabalho, utilizamos uma metodologia de destruição de informação, proposta inicialmente para imagens de lesões de pele, para avaliar a possível presença de vieses em conjunto de dados de detecção de pneumonia. Para isso, criamos seis variações diferentes do conjunto de dados *RSNA Pneumonia Detection Challenge*. Nossos resultados mostraram que as redes neurais profundas foram capazes de aprender correlações espúrias em todos as variações, inclusive no conjunto onde as informações foram quase — totalmente — ocultadas.

1 Introdução

A pneumonia é a principal causa de morte infecciosa de crianças ao redor do mundo. A doença matou 808.694 crianças menores de 5 anos em 2017, totalizando 15% de todas as mortes de crianças menores de 5 anos [1]. O diagnóstico de pneumonia requer a análise de radiografias do tórax por especialistas e confirmação por histórico médico, sinais vitais e exames laboratoriais. Em imagens de raio-x, a pneumonia se manifesta como uma área de opacidade maior na imagem [2], porém múltiplas outras condições podem estar presentes na imagem e dificultam o diagnóstico, por exemplo, atelectasia, infiltrações, câncer de pulmão e alterações pós-cirúrgicas [3].

Em 2018, a Sociedade Norte Americana de Radiologia (*Radiological Society of North America*, RNSA) realizou uma competição pública de detecção de pneumonia em radiografias de tórax. A competição foi criada na plataforma Kaggle [3], onde uma base de dados foi disponibilizada. A base contém imagens de raios-x de pulmões com pneumonia e outras possíveis condições. Cada imagem possui anotações indicando o local das regiões opacas indicadoras de pneumonia, se houver. Em tal competição, redes neurais profundas (*deep neural networks, DNNs*) foram treinadas para localizar essas regiões, ou seja, indicar sua

*luisfelipehsr@gmail.com

†Instituto de Computação, Universidade Estadual de Campinas, 13083-852, Campinas, SP

posição e tamanho. A presença ou não de tais regiões pode ser usada para classificar uma imagem como um caso positivo ou negativo, configurando um prognóstico assistido por computador.

Embora as DNNs tenham grande sucesso em tarefas de visão computacional, essas redes em geral necessitam de grandes bases de dados anotadas. Como imagens médicas, particularmente imagens anotadas, são em geral escassas, a utilização de DNNs na área médica é uma tarefa desafiadora. A utilização de DNNs para imagens médicas pode ser ainda mais problemática dado que as DNNs são vulneráveis a vieses.

Vieses (*bias*) em *Machine Learning* é o nome dado ao fenômeno que ocorre quando um algoritmo produz resultados que são sistematicamente prejudicados devido a suposições errôneas. Neste trabalho, analisamos a presença de vieses (correlações espúrias) em conjuntos de imagens de radiografias de tórax. Este trabalho é uma extensão do trabalho de Bissoto et al. [4], que analisando a presença de correlações espúrias em conjuntos de imagens de lesões de pele, apontou que os modelos desenvolvidos são otimistas e podem não refletir a realidade, ou seja, os resultados podem estar inflados. O principal objetivo do nosso trabalho é avaliar a generalização da metodologia e a possível presença do mesmo problema em outro contexto médico.

Nosso grupo de pesquisa tem abordado o problema de prognóstico de lesões de pele em diferentes frentes — classificação [5–9], segmentação [10], expansão de dados (*data augmentation*) [11] por meio de síntese de imagens [12, 13], análise de vieses [4, 14] — com resultados promissores. Estas frentes são importantes para a construção de bons modelos de aprendizagem de máquina, uma vez que a generalização de DNNs é sensível à qualidade dos dados. Quando um conjunto de dados inadequado não representa suficientemente bem um caso de uso real, o desempenho teórico do modelo não é alcançado quando aplicado num ambiente de produção.

É importante destacar que, inicialmente, este trabalho expandiria estudos sobre a qualidade das imagens sintéticas geradas, sendo esta sessão do projeto reportada no Apêndice 1. Como os resultados obtidos não foram promissores, decidiu-se mudar a direção para a investigação da qualidade de dados em conjuntos de dados de diagnóstico por imagens de outras doenças. O problema escolhido foi o de investigar a qualidade dos dados disponíveis para a diagnóstico de pneumonia.

As seções seguintes deste trabalho estão organizadas da seguinte forma: na Seção 2 são apresentados trabalhos relacionados a este trabalho; na Seção 3 é explicada a metodologia com que foi selecionado o conjunto de dados de pneumonia e como foram avaliadas as correlações; na Seção 4 são apresentamos os resultados dos experimentos; na Seção 5, conclusões e trabalho futuros.

2 Trabalhos Relacionados

Nos últimos anos as DNNs têm demonstrado ganhos significativos sobre as abordagens tradicionais de Visão Computacional. Estas têm se tornado progressivamente mais acuradas e/ou eficientes, e aplicadas a uma gama maior de problemas. Alguns exemplos notáveis são a ResNet [15], Inception [16], EfficientNet [17] e Visual Transformers [18]. As DNNs têm sido

aplicadas com sucesso por empresas como Google, Microsoft, Facebook e Baidu [19], também sendo utilizadas com sucesso em diversas frentes de análise de imagens médicas [5, 12, 20].

DNNs baseiam seu funcionamento no reconhecimento de padrões presentes no conjunto de dados fornecido durante seu treinamento. Um grande desafio na utilização de DNNs é treiná-las de modo que estas não aprendam correlações espúrias dos conjuntos de dados. Tais correlações podem inflar a performance da DNN e levar o modelo a dar ênfase em fatores não representativos para sua tomada de decisão, podendo até mesmo ignorar fatores relevantes. Estas correlações exploradas pelas redes são chamadas de tendências, viés ou *bias*. Quanto menor um conjunto de dados, mais provável é a presença de correlações espúrias que levem os modelos a serem enviesados. Vale notar que isto não implica que conjuntos grandes sejam livres de correlações indesejadas. DNNs são capazes de aprender vieses mesmo em conjuntos vastos e diversos como o ImageNet [21], que possui milhões de imagens, de forma que é de se esperar que conjuntos reduzidos e obtidos de poucas fontes, tal como os de imagens médicas, são sujeitos a este mesmo problema em maior escala.

No caso de bases de dados de imagens de lesões de pele, por exemplo, podemos perceber que os dados ISIC Challenge [22] apresentam diversos artefatos que podem dar origem a padrões indesejados, como a presença de bordas escuras, réguas, gel e pelos (Figura 1) [23].

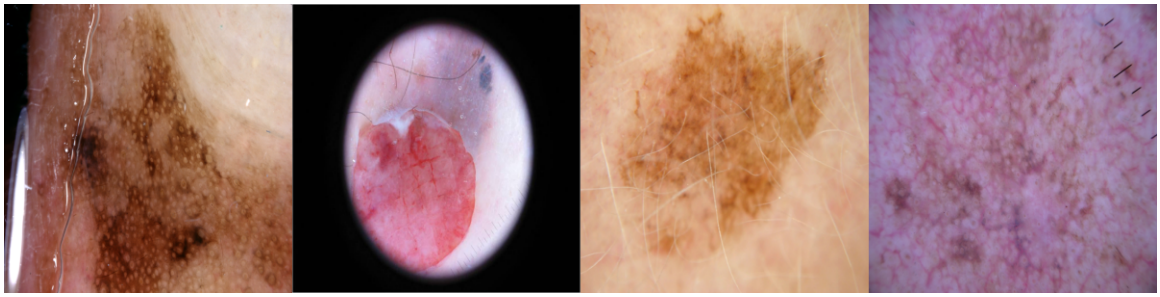


Figura 1: Imagens de lesões de pele do conjunto ISIC Challenge [22]. Da esquerda para direita, temos exemplos de imagens com gel, bordas escuras, pelos e régua.

2.1 "(De)Constructing Bias on Skin Lesions Datasets"

Bissoto et al. [4], no seu artigo "(De)Constructing Bias on Skin Lesions Datasets", analisaram a presença de correlações espúrias nos conjuntos de dados mais comuns de classificação de lesões de pele. Estas correlações podem distorcer a performance de um modelo e suprimir o aprendizado de correlações de fato significantes num caso real.

Para o diagnóstico de lesões de pele, um dos fatores significativos na identificação da doença pelos médicos é a análise de padrões de imagens dermatoscópicas da lesão. A presença de certos padrões (rede pigmentar, glóbulos ou pontos) são característicos das lesões melanocíticas. Tais atributos podem ser entendidos como texturas de regiões da lesão, e são utilizados pelos médicos para identificar os diferentes tipos de doença causadora [24,25]. É esperado que tal relação seja também importante para a tomada de decisão das DNNs. Embora a pele saudável ao redor da lesão tenha alguma relevância, ela não é um fator considerado importante na identificação da doença.

Em seus experimentos, Bissoto et al. propuseram experimentos contrafactuais, que progressivamente destroem a informação relevante dos dados e medem como a performance do modelo diminui. O objetivo é detectar se a performance dos modelos é inflacionada pela presença de correlações espúrias. Nestes experimentos, as DNNs foram avaliadas quando treinadas sobre quatro versões diferentes dos conjuntos de imagens de lesão de pele. Inicialmente, o treinamento envolve somente as imagens originais, de forma a se obter um desempenho base, enquanto que nos seguintes as imagens são modificadas com ocultamentos de regiões. Nas versões seguintes, as imagens tiveram primeiramente a região da lesão ocultada (usando a máscara de segmentação da lesão), de forma a se eliminar qualquer informação sobre quais são atributos. Em seguida, a região é ocultada com uma caixa delimitadora (*bounding box*) que tem tamanho da máscara de segmentação da lesão, de forma a eliminar a informação da forma da borda das lesões. E, por fim, na última versão, a caixa delimitadora ocupa no mínimo 70% da região da imagem, eliminando também a informação do tamanho de tais lesões.

Para todos os experimentos, as DNNs continuaram a ter um desempenho relativamente alto mesmo quando a região inteira da lesão era ocultada, sendo até maior do que o encontrado quando somente os atributos dermatoscópicos eram considerados.

3 Metodologia

Para aplicar a metodologia utilizada por Bissoto et al., é necessário ter as regiões de interesse anotadas para que seja possível removê-las (da mesma forma como as lesões foram removidas). Assim, utilizamos as regiões indicativas de pneumonia em radiografias de pulmões da base de dados da competição *RSNA Pneumonia Detection Challenge*.

Para todos os experimentos, utilizamos o código de Bissoto et al., disponível em https://github.com/alceubissoto/deconstructing-bias-skin-lesion. Nenhum hiperparâmetro da rede ou aumentações de dados foram alterados. A rede utilizada é uma Inception-v4 [26] pré-treinada na ImageNet.

Para a avaliação dos modelos, foi utilizado o protocolo *k-fold cross-validation*, onde o conjunto de dados é dividido em k subconjuntos mutuamente exclusivos do mesmo tamanho, sendo que um subconjunto é utilizado para teste e os $k-1$ restantes são utilizados para o treinamento. Este processo é realizado k vezes. Neste trabalho, foram utilizados 10 *folds*. Para cada treinamento, foi medido o valor da curva ROC AUC (*Area Under the Curve*) sobre o conjunto de teste correspondente, sendo a métrica final a média aritmética sobre todos os conjuntos.

3.1 Base de Dados

Para investigar os possíveis vieses, utilizamos como base de dados o conjunto de treinamento do *RSNA Pneumonia Detection Challenge* (2018) [3], disponível publicamente. A partir daqui, deve-se assumir que referências à base RSNA dizem respeito somente ao conjunto de treino original de tal base. A base contém imagens de radiografias de tórax de pacientes, as quais podem ser de pacientes com sinais de pneumonia ou não. Como é uma base para a tarefa de detecção de objetos, as imagens positivas possuem caixas delimitadoras (*bounding*

box) que determinam a presença de evidências de pneumonia. Tais evidências são regiões mais opacas nos pulmões (Figura 2).

As imagens possuem artefatos notáveis como a presença de ícones (Figura 3a), objetos, fios e bordas pretas em diversos exemplos (Figura 3b). Possivelmente, tais artefatos podem dar origem a correlações indesejadas, por exemplo, é possível que a maioria das imagens que possuem fios sejam de uma mesma classe, levando a DNN a associar fios com a doença.

A base utilizada (conjunto de treino do RSNA) possui 26.684 imagens, sendo 20.672 casos negativos, ou seja, sem a presença da doença anotada, e 6.012 casos positivos, ou seja, possuem um ou mais objetos anotados, como ilustrado na Figura 2.

Para aplicar a proposta de Bissoto et al. [4], em vez de utilizar o problema original de detecção de objetos, o problema foi convertido num de classificação binária entre casos positivos e negativos.

Os casos positivos são aqueles onde há a presença de ao menos uma região escurecida no pulmão sintomática de pneumonia, enquanto os negativos são os que não há tal região. A região de interesse das imagens positivas foi considerada a região anotada como tendo presença do sintoma de pneumonia (posição do *bounding box*). As imagens são representadas com um único canal de cor (tons de cinza) de tamanho 1024×1024 pixels, convertido para 299×299 pixels nos nossos experimentos.

Para a análise destrutiva de informação, é importante não adicionar novas informações aos dados. De forma a aproveitar as anotações de *bounding box* e evitar complexidades desnecessárias, a estratégia de ocultamento escolhida foi de ocultar regiões de interesse com um retângulo preto, como feito por Bissoto et al. [4]. Como somente as imagens positivas possuem anotação, ocultar somente as regiões de interesse adicionaria um novo padrão aos dados: imagens com ocultamento são positivas. Assim, é necessário que sejam adicionados ocultamentos às imagens negativas também. Tais ocultamentos de imagens negativas devem ser escolhidos de forma a não introduzir novas formas de correlação óbvias.

Neste trabalho, foram experimentadas cinco estratégias de ocultamento: 1) ocultamento aleatório baseado na distribuição dos *bounding box* das imagens positivas, 2) ocultamento aleatório baseado na distribuição dos *bounding box* das imagens positivas, porém unindo múltiplos *bounding box* em um único ocultamento, 3) ocultamento padrão 70, onde um único ocultamento é adicionado, como o anterior, porém com tamanho padronizado de 70% da imagem, de forma a remover correlações ligadas ao tamanho do ocultamento, 4) ocultamento fixo de 70%, onde um único ocultamento é adicionado ao centro da imagem e com tamanho fixo de 70% da imagem, de forma a remover correlações ligadas à posição do ocultamento e 5) ocultamento fixo ocupando 90% da imagem.

Os *folds* foram determinados antes da geração de imagens para as diferentes estratégias. Assim, todas as estratégias são executadas para os mesmos *folds*.

A seguir, descrevemos cada uma das estratégias.

3.2 Ocultamento Aleatório

Nesta estratégia as imagens positivas possuem ocultamentos que correspondem perfeitamente a cada *bounding box* anotado (Figura 4a). As imagens positivas possuem um ou dois ocultamentos. Por sua vez, as imagens negativas precisam receber ocultamentos que

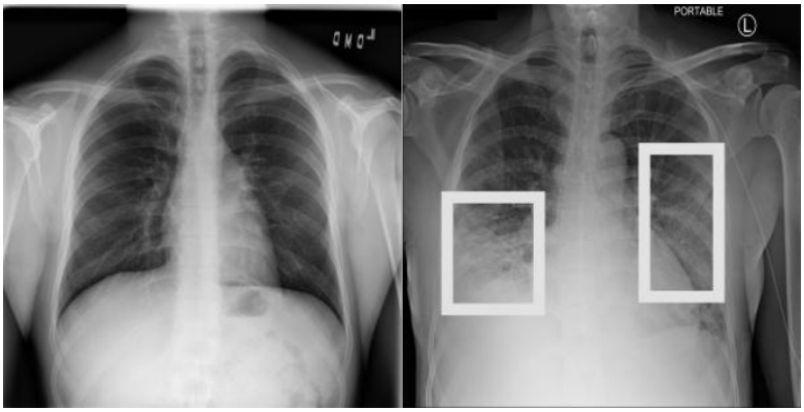
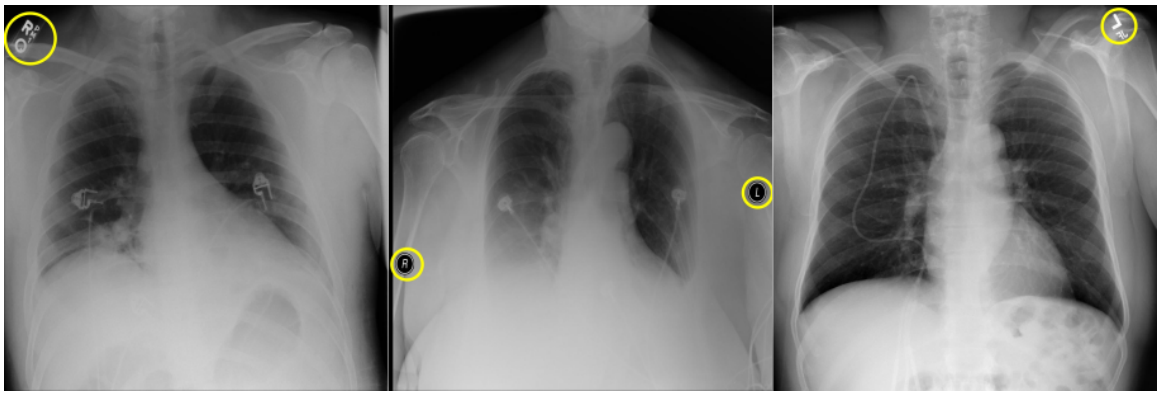
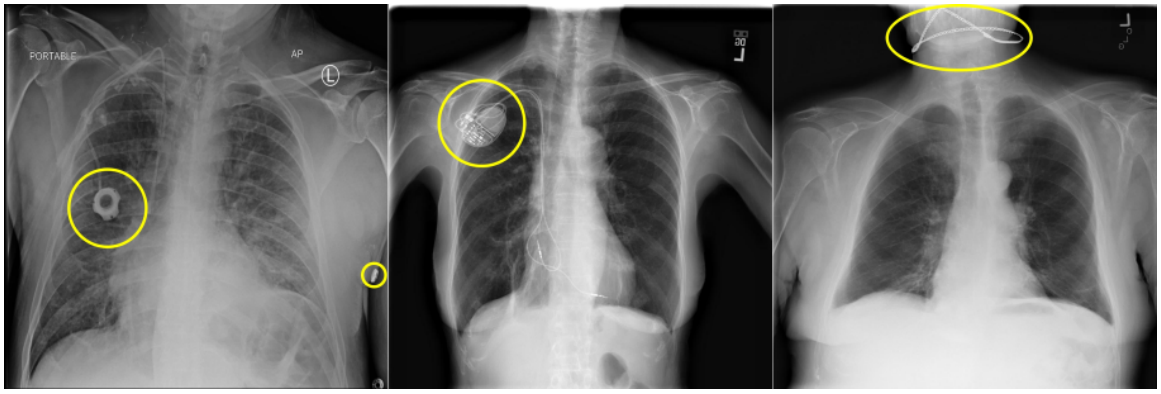


Figura 2: Exemplos da base *RSNA Pneumonia Detection Challenge* (2018) [3]. À esquerda pulmão sem sinais de pneumonia, à direita pulmão com regiões opacas que indicam pneumonia (retângulos indicam as regiões anotadas).

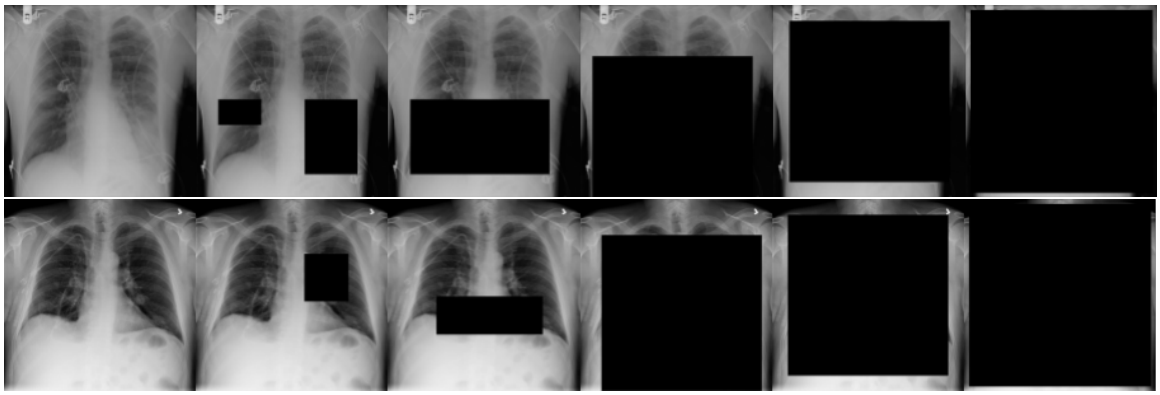


(a) Ícones (circulados em amarelo para ilustração) presentes nos dados.



(b) Objetos (circulados em amarelo para ilustração) presentes nos dados.

Figura 3: Exemplos de artefatos, como ícones, objetos, fios e bordas pretas.



(a) (b) (c) (d) (e) (f)

Figura 4: Exemplos de imagens das classes positiva (primeira linha) e negativa (segunda linha). Os pares verticais pertencem aos seguintes conjuntos de dados, da esquerda para direita: (a) imagem original do conjunto RSNA, (b) Pneumonia Aleatório, (c) Pneumonia Aleatório Combinado, (d) Pneumonia Padronizado, (e) Pneumonia Fixo 70, (f) Pneumonia Fixo 90.

simulem os ocultamentos das imagens positivas. A fim de emular os ocultamentos de imagens positivas, foram adicionados entre um e dois ocultamentos em cada imagem negativa (Figura 4b).

Inicialmente, uma estratégia simples foi utilizada: os ocultamentos foram gerados utilizando a distribuição de tamanho e posição geral dos *bounding box* das imagens positivas. Os resultados (não reportados aqui) apontavam grande facilidade da DNN em determinar quais imagens eram negativas. Isto se dá em razão da correlação entre as posições de ocultamentos em imagens positivas que possuem múltiplos ocultamentos. É raro que múltiplos ocultamentos estejam presentes somente em um dos pulmões.

Desta forma foi necessário extrair duas distribuições distintas para se tentar simular a posição dos ocultamentos: uma baseada nos casos positivos onde só há um ocultamento, e outra baseada nos casos positivos com múltiplos ocultamentos.

Foi obtida a proporção entre imagens positivas com um único ou múltiplos ocultamentos, e foram adicionados ocultamentos às imagens negativas de forma a se manter a mesma distribuição.

O conjunto de dados desta estratégia foi chamado de **Pneumonia Aleatório**.

3.3 Ocultamento Aleatório Combinado

Nesta estratégia, os casos em que múltiplos ocultamentos estão presentes na mesma imagem são combinados num único ocultamento, seja positiva ou negativa. Os ocultamentos originais são combinados de forma a criar um único e menor ocultamento retangular que oculte todos os ocultamentos originais.

As imagens positivas têm este ocultamento combinado derivado das anotações originais de *bounding box* (Figura 4b). Nas imagens negativas, foram utilizados os mesmos critérios

de geração de ocultamento do conjunto Pneumonia Aleatório, com a única mudança de combinar ocultamentos múltiplos num único ocultamento (Figura 4c).

O conjunto de dados desta estratégia foi chamado de **Pneumonia Aleatório Combinado**.

3.4 Ocultamento Padronizado

Um dos problemas com as estratégias anteriores é a dificuldade de se simular o tamanho e formato que os ocultamentos deveriam ter. Falhas nesta geração podem levar criação ou manutenção de correlações indesejadas nos dados.

Para se resolver tais problemas, uma alternativa é fixar o tamanho e formato da região ocultada. Foi escolhido o formato de um quadrado por simplicidade, similar ao que foi realizado por Bissoto et al. [4]. Tal ocultamente precisa ser grande o bastante para conter todas as regiões de interesse das imagens positivas. O tamanho escolhido foi igual a 70% do tamanho da imagem, de forma a ser consistente com o trabalho de Bissoto et al..

A posição dos ocultamentos foi determinada a partir do conjunto Pneumonia Aleatório Combinado. Os novos ocultamentos (Figura 4d) são centralizados sobre a posição onde está o centro do ocultamento correspondente no conjunto Pneumonia Aleatório Combinado (Figura 4c). Para esta estratégia, ocultamentos muito próximos à borda das imagens as toquem ou ultrapassem foram permitidos.

O conjunto criado por esta estratégia foi chamado de **Pneumonia Padronizado**.

3.5 Ocultamento Fixo 70

Esta estratégia remove também a necessidade de determinação da posição de ocultamentos nas imagens negativas. Todos os ocultamentos são centralizados, tanto em imagens positivas quanto negativas (Figura 4e).

A região ocultada é grande o bastante para abranger a maioria dos *bounding box* anotados nas imagens positivas. Descartamos 10% das imagens positivas em que há um ou mais *bounding box* que não estariam totalmente cobertos pelo ocultamento. Não foi realizada uma redistribuição das imagens entre os *10 folds*, somente foram removidas das divisões já presentes.

Como acreditamos que não existe a introdução de nenhum viés por parte de posição, tamanho ou formato de ocultamento, ela é a estratégia mais agressiva. Porém, como a região ocultada é muito grande em relação ao tamanho da imagem, na maioria dos casos o ocultamento obstrui completamente a visão dos pulmões, que são o local onde os indícios de pneumonia estão presentes.

Neste conjunto, esperamos que o desempenho da DNN seja baixo dado que não há correlações indesejadas presentas em regiões fora das regiões de interesse. Note que há ainda a presença de vários dos artefatos mencionados: fios, ícones, alguns objetos e bordas.

Como este conjunto de dados possui um número de imagens diferente dos anteriores, ele foi analisado e apresentado separadamente na Seção 4.

O conjunto criado por esta estratégia foi chamado de **Pneumonia Fixo 70**.

3.6 Ocultamento Fixo 90

Esta estratégia é similar à estratégia anterior, porém com uma região de ocultamento de tamanho de 90% da imagem. Somente poucos pixels das bordas das imagens estão visíveis. Neste conjunto, os artefatos se resumem aos fios e bordas.

Foram descartadas as mesmas imagens descartadas pelo conjunto Pneumonia Fixo 70, de forma que ambos sejam comparáveis. Estes conjuntos foram analisados conjuntamente na Seção 4.

O conjunto criado por esta estratégia foi chamado de **Pneumonia Fixo 90**.

4 Resultados

A seguir, são apresentados os resultados para as diferentes estratégias apresentadas na Seção 3. Primeiro, são apresentados os resultados para os conjuntos Original (sem qualquer ocultamento), *Pneumonia Aleatório*, *Pneumonia Aleatório Combinado* e *Pneumonia Padronizado*, pois todos têm as mesmas imagens em todos os *folds*. Os conjuntos *Pneumonia Fixo 70* e *Pneumonia Fixo 90* tiveram imagens excluídas, conforme explicado nas Seções 3.3 e 3.4.

Caso as DNNs não sejam capazes de encontrar correlações nas imagens, seus resultados serão ruins, se aproximando de 0.500. Como todos os conjuntos, com exceção do original, ocultam as regiões de interesse no diagnóstico de pneumonia, é esperado que todos os resultados, com exceção daquele, sejam ruins.

Na Tabela 1, são apresentados os resultados. Podemos perceber que houve a introdução de novas correlações indesejadas nos conjuntos Pneumonia Aleatório e Pneumonia Aleatório Combinado. Os resultados de tais conjuntos são significativamente maiores do que o conjunto Original, o que significa que se tornou mais fácil para a rede aprender a diferenciar entre imagens positivas e negativas. Tal comportamento não ocorreu no conjunto Pneumonia Padronizado, de forma que foi concluído que as estratégias de geração de formato e tamanho dos ocultamentos não foram capazes de gerar ocultamentos similares o suficiente aos das imagens positivas.

Tabela 1: Média de ROC AUC para um 10-fold de conjuntos derivados do *RSNA Pneumonia Detection Challenge*.

Conjunto de Dados	AUC
Original	0.884 ± 0.006
Pneumonia Aleatório	0.987 ± 0.002
Pneumonia Aleatório Combinado	0.971 ± 0.003
Pneumonia Padronizado	0.853 ± 0.009

Além disso, notamos que no conjunto Pneumonia Padronizado, mesmo que a informação de tamanho das regiões de interesse tenha sido removida, os resultados se mantiveram elevados e próximos aos do conjunto Original. Isto indica que ainda há fontes de correlação

presentes. Para determinar se tais correlações vêm da estratégia de geração de ocultamentos para as imagens negativas ou se há correlações ainda presentes na própria imagem, é necessário avaliar as duas estratégias restantes: Pneumonia Fixo 70 e 90.

Na Tabela 2, são apresentados os resultados dos conjuntos Pneumonia Fixo 70 e 90, que possuem resultados similares. Estes conjuntos não são diretamente comparáveis aos anteriores em razão da exclusão de imagens do conjunto de dados.

Tabela 2: Média de ROC AUC para um 10-fold de conjuntos derivados do *RSNA Pneumonia Detection Challenge*. Estes conjuntos excluem imagens com regiões de interesse que não estejam contidas na área central de 70% (Pneumonia Fixo 70) ou 90% (Pneumonia Fixo 90) da imagem.

Conjunto de Dados	AUC
Pneumonia Fixo 70	0.774 ± 0.012
Pneumonia Fixo 90	0.741 ± 0.014

Mesmo com a exclusão completa das regiões de interesse de imagens positivas, e sem a introdução de qualquer possível correlação de posição, tamanho ou formato de ocultamento, as redes foram capazes de determinar quais imagens são positivas ou negativas com considerável desempenho. Em razão da surpresa deste resultado, foram realizados testes de sanidade com imagens totalmente ocultadas para se avaliar a validade do código e dos experimentos. O desempenho das redes de fato tiveram desempenho de 0.5 de ROC AUC.

Nas imagens do conjunto Pneumonia Fixo 90, mais ainda que no Pneumonia Fixo 70, os pulmões estão completamente ocultados na grande maioria das imagens. Os únicos objetos visíveis nos poucos pixels presentas da borda, são os ossos, que podem ter correlações ligadas a sua posição, e alguns dos artefatos, como fios e bordas pretas. Não é possível por esta metodologia determinar se são ossos, artefatos, ou ambos, que são os responsáveis pela forte correlação ainda presente.

5 Conclusão e Trabalhos Futuros

A avaliação dos resultados no conjunto de dados do *RSNA Pneumonia Detection Challenge* sugere a presença de correlações espúrias indesejadas relacionado às informações presentes nas bordas das imagens, uma vez que a DNN é capaz de determinar a presença de pneumonia, mesmo em conjuntos em que a imagem foi quase — totalmente — ocultada, como o Pneumonia Fixo 90.

A habilidade das DNNs de detectarem facilmente correlações indesejadas introduzidas neste conjunto é demonstrada também pelos resultados superiores ao de uma imagem original nos conjuntos Pneumonia Aleatório e Pneumonia Aleatório Combinado. O método de geração de ocultamentos para as imagens negativas acabou por facilitar a classificação das imagens. No conjunto Pneumonia Padronizado a destruição da informação de tamanho do ocultamento diminuiu o desempenho da rede como esperado, porém pouco abaixo do conjunto Original, indicando que ainda há fontes de correlação espúrias ligadas à geração

dos ocultamentos ou às regiões de não interesse da imagem, como bordas.

Nos conjuntos Pneumonia Fixo 70 e 90, não há correlações introduzidas pela geração de ocultamentos para as imagens negativas, uma vez que todas as imagens possuem o mesmo ocultamento. Mesmo assim, a rede apresentou um desempenho elevado em ambos os conjuntos. O aumento do ocultamento de 70% para 90% da imagem pouco diminuiu tal desempenho. Como a única informação restante para a classificação destas imagens são as bordas, conclui-se que há fortes fontes de correlações espúrias nesta região.

Os prováveis motivos para a presença de tais correlações nas bordas são padrões na presença, posição e ângulos de ossos e artefatos. A possibilidade de que há regiões de interesse não anotadas nas bordas das imagens foi considerada, contudo isto requer que grande parte das imagens tenham pulmões visíveis nas bordas, algo que não ocorre.

Este estudo pode ser ainda expandido com a análise da importância de detalhes da imagem para a classificação, seguindo a mesma metodologia, mas borrando a região visível da imagem. Tal procedimento permitiria medir a contribuição da região central da imagem para a identificação de pneumonia, através da diferença de desempenho entre as imagens ocultadas e borradas.

Outra alternativa é o treinamento de um detector de tórax em radiografias, de forma a se ter com mais precisão os limites da região de interesse que deveria ser observada. Isto permitiria a eliminação por completo do tórax (e portanto das regiões de pneumonia), de forma que as únicas informações "relevantes" seriam os artefatos e ossos, eliminando a possibilidade de erro de anotação por alguma região de pneumonia não anotada.

Por fim, a metodologia de estudo apresentada no trabalho referência pôde ser expandida para outro conjunto de dados de imagens médicas, evidenciando a presença de correlações espúrias nos dados. Ressaltamos que, mesmo confirmando a metodologia para lesões de pele e radiografias de tórax, é interessante — e urgente — a aplicação da metodologia para outras bases de dados e, principalmente, outros contextos médicos.

Referências

[1] "Pneumonia fact sheet," Available: https://www.who.int/news-room/fact-sheets/detail/pneumonia, 2019. 1

[2] T. Franquet, "Imaging of community-acquired pneumonia," *Journal of Thoracic Imaging*, vol. 33, p. 1, 07 2018. 1

[3] "RSNA Pneumonia Detection Challenge," Available: https://www.kaggle.com/c/rsna-pneumonia-detection-challenge/overview, 2018. 1, 4, 6

[4] A. Bissoto, M. Fornaciali, E. Valle, and S. Avila, "(De)Constructing Bias on Skin Lesion Datasets," in *ISIC Skin Image Analysis Workshop, IEEE CVPR Workshops*, 2019. 2, 3, 5, 8

[5] E. Valle, M. Fornaciali, A. Menegola, J. Tavares, F. V. Bittencourt, L. T. Li, and S. Avila, "Data, depth, and design: Learning reliable models for skin lesion analysis," *Neurocomputing*, vol. 383, pp. 303–313, 2020. 2, 3

[6] A. Menegola, M. Fornaciali, R. Pires, F. V. Bittencourt, S. Avila, and E. Valle, "Knowledge transfer for melanoma screening with deep learning," in *IEEE International Symposium on Biomedical Imaging*, 2017. 2

[7] A. Menegola, J. Tavares, M. Fornaciali, L. T. Li, S. Avila, and E. Valle, "RECOD titans at ISIC challenge 2017," *CoRR*, vol. abs/1703.04819, 2017. [Online]. Available: http://arxiv.org/abs/1703.04819 2

[8] A. Bissoto, F. Perez, V. Ribeiro, M. Fornaciali, S. Avila, and E. Valle, "Deep-learning ensembles for skin-lesion segmentation, analysis, classification: RECOD titans at ISIC challenge 2018," *CoRR*, vol. abs/1808.08480, 2018. [Online]. Available: http://arxiv.org/abs/1808.08480 2

[9] L. Chaves, A. Bissoto, E. Valle, and S. Avila, "An evaluation of self-supervised pre-training for skin-lesion analysis," *arXiv preprint arXiv:2106.09229*, 2021. 2

[10] V. Ribeiro, S. Avila, and E. Valle, "Less is more: Sample selection and label conditioning improve skin lesion segmentation," in *ISIC Skin Image Analysis Workshop, IEEE CVPR Workshops*, 2020. 2

[11] F. Perez, C. Vasconcelos, S. Avila, and E. Valle, "Data augmentation for skin lesion analysis," in *ISIC Skin Image Analysis Workshop, MICCAI*, 2018. 2

[12] A. Bissoto, F. Perez, E. Valle, and S. Avila, "Skin lesion synthesis with generative adversarial networks," in *ISIC Skin Image Analysis Workshop, MICCAI*, 2018. 2, 3

[13] A. Bissoto, E. Valle, and S. Avila, "Gan-based data augmentation and anonymization for skin-lesion analysis: A critical review," in *ISIC Skin Image Analysis Workshop, IEEE CVPR Workshops*, 2021. 2

[14] ——, "Debiasing skin lesion datasets and models? Not so fast," in *ISIC Skin Image Analysis Workshop, IEEE CVPR Workshops*, 2020. 2

[15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *CoRR*, vol. abs/1512.03385, 2015. [Online]. Available: http://arxiv.org/abs/1512.03385 2

[16] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. E. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *CoRR*, vol. abs/1409.4842, 2014. [Online]. Available: http://arxiv.org/abs/1409.4842 2

[17] M. Tan and Q. V. Le, "Efficientnet: Rethinking model scaling for convolutional neural networks," *CoRR*, vol. abs/1905.11946, 2019. [Online]. Available: http://arxiv.org/abs/1905.11946 2

[18] B. Wu, C. Xu, X. Dai, A. Wan, P. Zhang, M. Tomizuka, K. Keutzer, and P. Vajda, "Visual transformers: Token-based image representation and processing for computer vision," *CoRR*, vol. abs/2006.03677, 2020. [Online]. Available: https://arxiv.org/abs/2006.03677 2

[19] "Why deep learning is suddenly changing your life," Available: http://fortune.com/ai-artificial-intelligence-deep-machine-learning/, Roger Parloff, Sep 28^{th}, 2016. 3

[20] A. Maier, C. Syben, T. Lasser, and C. Riess, "A gentle introduction to deep learning in medical image processing," *Zeitschrift für Medizinische Physik*, vol. 29, no. 2, pp. 86 – 101, 2019, special Issue: Deep Learning in Medical Physics. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S093938891830120X 3

[21] R. Geirhos, P. Rubisch, C. Michaelis, M. Bethge, F. A. Wichmann, and W. Brendel, ""imagenet-trained cnns are biased towards texture; increasing shape bias improves accuracy and robustness," *International Conference on Machine Learning*, 2018. 3

[22] "ISIC Challenge 2019," Available: https://challenge2019.isic-archive.com/leaderboard.html, 2019. 3

[23] N. K. Mishra and M. E. Celebi, "An overview of melanoma detection in dermoscopy images using image processing and machine learning," *CoRR*, vol. abs/1601.07843, 2016. [Online]. Available: http://arxiv.org/abs/1601.07843 3

[24] R. I. N. Gisele Gargantini Rezze, Bianca Costa Soares de Sá, "Dermatoscopia: o método de análise de padrões. anais brasileiros de dermatologia," *Anais Brasileiros de Dermatologia*, vol. 81, pp. 261–268, 2006. 3

[25] G. Argenziano, G. Fabbrocini, P. Carli, V. De Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the abcd rule of dermatoscopy and a new 7-point checklist based on pattern analysis," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, 1998. 3

[26] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," *CoRR*, vol. abs/1512.00567, 2015. [Online]. Available: http://arxiv.org/abs/1512.00567 4