

Data-augmentation de dados de radiografia de tórax no contexto de aprendizagem profunda

Otto Tavares Nascimento
Laboratório de Processamento de Sinais, COPPE/POLI,
Universidade Federal do Rio de Janeiro.
Rio de Janeiro, Brasil
otto.tavares@coppe.ufrj.br

Anete Trajman, MD, PhD
Professora de Medicina Interna, Escola de Medicina,
Universidade Federal do Rio de Janeiro.
Professora Adjunta, McGill University
Rio de Janeiro, Brasil
atrajman@gmail.com

José Manoel de Seixas
Laboratório de Processamento de Sinais, COPPE/POLI,
Universidade Federal do Rio de Janeiro.
Rio de Janeiro, Brasil
seixas@lps.ufrj.br

Resumo—Este artigo discute técnicas de aprendizagem profunda aplicadas ao problema de triagem de tuberculose ativa em radiografias de tórax. O objetivo é contribuir para a prevenção e para o tratamento desta doença infecciosa, propondo uma rede neural convolucional (CNN) classificadora de tuberculose ativa, tendo como dados de entrada um conjunto de radiografias de tórax de indivíduos com ou sem tuberculose. Para melhor ajustar os parâmetros da CNN, técnicas de *data-augmentation*, utilizando modelos generativos adversariais (GAN), são aplicadas e avaliadas.

Index Terms—Aprendizado profundo, Tuberculose, Aprendizado de Máquina na Saúde, *Data-Augmentation*.

I. INTRODUÇÃO

Dentre as técnicas computacionais que se inspiram em fenômenos naturais, uma das mais proeminentes dos últimos anos parte do arcabouço de redes neurais artificiais para gerar aprendizado em contextos complexos. O aprendizado profundo busca representações, em um espaço latente, das variáveis de entrada, a fim de facilitar a tomada de decisão a respeito de um fenômeno complexo [1].

Em outras palavras, redes neurais artificiais dotadas de muitas camadas passaram a ser construídas com o objetivo de identificar representações dos dados de entrada e facilitar a tomada de decisão em tarefas consideradas difíceis até um tempo muito recente. Esta técnica tem despertado o interesse de diversas áreas, como no campo de processamento de linguagem natural, aplicações em sistemas de sonar passivo e em visão computacional [2]. Neste último caso, imagens de radiografia e de ultrassonografia possibilitam a utilização destes modelos no campo da saúde.

O caso da tuberculose é um ótimo exemplo. Como uma doença infecciosa de alta carga em países em desenvolvimento, os quais são responsáveis por ao menos dois terços de sua

incidência mundial [3], é importante que o combate e a prevenção atinjam uma grande quantidade de pessoas a um baixo custo. Utilizar modelos de inteligência computacional na etapa de triagem, tendo como dados radiografias de tórax, pode ser um caminho promissor de contribuição para esse problema de saúde pública.

Isto posto, a proposta deste artigo é trabalhar com modelos de aprendizado profundo no contexto de saúde. Busca-se uma rede neural profunda convolucional (CNN) [4], auxiliada por técnicas de *data-augmentation* [5], que classifique radiografias de tórax e tenha alta sensibilidade - isto é, alta probabilidade de detecção - em identificar pulmões que apresentem tuberculose ativa, sem que a especificidade seja desconsiderada.

Além de ser um problema complexo, um desafio a mais é a escassez de dados no contexto de saúde [6] [7] [8]. Radiografias de qualidade possuem alto custo em sua geração, além de demandarem uma equipe de saúde bem treinada na realização dessa tarefa. Por sua vez, redes neurais profundas exigem que quanto mais camadas uma rede neural artificial possua, maior o número de parâmetros a ser ajustado, para que a aproximação ao fenômeno estudado seja bem-sucedida. Isto é, alta complexidade da arquitetura leva a muitos parâmetros, o que demanda muitos dados.

Para enfrentar esse desafio, este artigo recorre a técnicas de *data-augmentation*. Especificamente, opta-se pela utilização de modelos generativos adversariais (GAN) [9], que, apesar de não serem uma técnica tradicional de *data-augmentation*, estimam a função geradora de eventos, respeitando a distribuição original dos dados, o que possibilita amostragem e, consequentemente, geração de diversidade. Técnicas de *data-augmentation* que se apoiam em operações matemáticas como rotação, translação, reflexão e adição de ruído gaussiano [10] [11] fornecem um aumento dos dados de radiografia, mas

não variedade dos mesmos, fundamental para análises neste contexto.

Logo, como estratégia de *data-augmentation*, dados públicos da base de Shenzhen [12] [13] são utilizados na estimativa de GAN, de modo que a função geradora extraída nesse processo crie dados sintéticos que serão introduzidos no treinamento de uma CNN, que classifique radiografias de tórax dotadas ou não de tuberculose ativa.

A utilização de modelos de aprendizado profundo em problemas de saúde tem recebido uma série de contribuições recentemente. Estratégias que envolvem *data-augmentation* e CNN são abordadas para avaliar radiografias com pneumonia [14]. Técnicas de GAN no contexto de classificação de imagens de câncer de pele também já são discutidas [15]. Modelos generativos utilizados para *data-augmentation* melhoram a performance da CNN em imagens de tomografia de câncer [8].

Modelos generativos adversariais são discutidos com bastante detalhe em [9]. Em [16] a instabilidade do treinamento de GAN é abordada. GAN condicionais, introduzidas em [17], são a base para o entendimento de GAN *Pix2Pix*. Já em busca de GAN dotada de arquitetura convolucional, pode-se destacar a estratégia de treinamento apresentada por [18].

A GAN de Wasserstein (WGAN) introduz uma métrica que controla o treinamento, de modo que os dados sintéticos criados através da função geradora sejam fidedignos à função de densidade de probabilidade do evento [19]. E para lidar com as instabilidades do treinamento da GAN de Wasserstein, o treinamento com gradiente penalizado é discutido em [20].

GAN condicionais dedicadas à tradução imagem a imagem pareadas são apresentadas em [21]. Por outro lado, uma tradução imagem a imagem realizada de forma não pareada é encontrada em [22]. Por sua vez, o modelo StarGAN propõe uma abordagem para a tradução imagem a imagem em múltiplos domínios, a partir de um modelo único [23]. E o modelo *Pix2Pix* é abordado em radiografias de tórax em [11] com a prerrogativa de gerar regiões sintéticas no entorno de regiões patológicas. Estas são mantidas através de modelos convolucionais de segmentação, evitando um gerador que replique regiões patológicas distorcidas, assegurando assim suas características originais.

Em [24] [25] [3] estratégias para a eliminação da tuberculose, que envolvem prevenção e tratamento, são discutidas. Tuberculose é uma doença infecciosa com a estimativa mundial de 10 milhões de casos e óbitos de aproximadamente 1,4 milhões de pessoas em 2019. Para eliminar a tuberculose, a infecção deve ser tratada mesmo em pacientes que não possuam a manifestação da doença [26]. Modelos que identifiquem tuberculose ativa em radiografias são de grande valia, por aumentarem a escala de diagnósticos e conseqüentemente de efetividade no tratamento¹.

Classificadores de alta sensibilidade servem a sistemas CAD (Computer Aided Diagnosis), que têm por objetivo dar suporte ao processo de triagem de determinada doença baseado em

exames de imagem. Portanto, o modelo classificador abordado neste artigo busca alta sensibilidade e especificidade, de modo a ser um candidato à utilização em sistemas de triagem de tuberculose ativa, fortalecendo assim os esforços para eliminação dessa doença infecciosa.

Além disso, este artigo contribui ao aplicar o conceito de transferência do conhecimento obtido ao se treinar a CNN apenas com dados sintéticos, gerados pela GAN, e utilizar os parâmetros obtidos para inicializar uma CNN de mesma arquitetura treinada com os dados originais das radiografias de tórax dotadas ou não de tuberculose.

O artigo é organizado com a primeira seção sendo a introdução, a segunda seção apresentando o método e as arquiteturas, a terceira explorando os experimentos e resultados e a quarta concluindo o trabalho.

II. MÉTODO E ARQUITETURAS DOS MODELOS

Para avaliar se houve ganho ao utilizar as técnicas de *data-augmentation*, o método toma como *baseline* um modelo de *Machine Learning* básico e um modelo de *Deep Learning* básico, que são comparados com os modelos que utilizam *data-augmentation* em sua estimativa.

O objetivo final é verificar se houve ganho no classificador de tuberculose ativa à medida que se aumentou a complexidade do modelo utilizado. Este ganho é medido através dos indicadores de sensibilidade², especificidade³ e índice soma produto (SP)⁴. O índice SP é utilizado justamente por indicar classificadores que tenham alta sensibilidade sem que a especificidade seja desconsiderada [27] [28].

Dessa forma, o *baseline* consiste na estimativa do modelo One Class SVM [29] e de uma CNN com os dados originais. Dados sintéticos são gerados por uma GAN de Wasserstein e por uma GAN *Pix2Pix*. Esses dados sintéticos são utilizados na estimativa de uma CNN, e verifica-se se há ganho com sua respectiva avaliação.

A. Modelo de Machine Learning básico

O modelo do One Class SVM parte do arcabouço de máquinas de vetores de suporte para abordar o problema de detecção de novidade. A especificação do modelo deste projeto se inspira no livro [29].

Buscam-se os parâmetros que façam do modelo um especialista para uma determinada classe. Ao se tornar especialista em uma das classes, ele identifica classes distintas como sendo novidade. Neste artigo, a classe a se especializar é a de tuberculose ativa, enquanto as radiografias de tórax sem tuberculose ativa passam a representar a novidade.

É essa capacidade de identificar as radiografias que não possuam a classe especialista como novidade que justifica a

²A sensibilidade ou probabilidade de detecção de doentes é dada pela seguinte fórmula: $\frac{\text{Verdadeiro Positivo}}{\text{Verdadeiro Positivo} + \text{Falso Negativo}}$.

³A especificidade ou probabilidade de detecção de saudáveis é dada pela seguinte fórmula: $\frac{\text{Verdadeiro Negativo}}{\text{Verdadeiro Negativo} + \text{Falso Positivo}}$.

⁴O índice soma produto (SP) é dado pela raiz quadrada da média aritmética multiplicada pela média geométrica dos valores de sensibilidade e de especificidade.

¹O tratamento varia de acordo com a manifestação ou não da doença.

utilização deste modelo como *baseline*. Em termos gerais, o *pipeline* deve em sua versão mais simples ser capaz de distinguir as classes, sendo especialista em umas delas e tomando a distinta como novidade.

O modelo se torna especialista quando se encontra a combinação de parâmetros que retorna a maior sensibilidade para tuberculose ativa, sem que a especificidade seja desconsiderada; ou seja, uma parametrização com o mais alto índice SP. A lista de parâmetros varia valores de ν , *kernel* e γ a fim de encontrar a parametrização do modelo que melhor se ajusta aos dados.

É importante destacar o papel do ν neste processo, por dar intuição do quão plástico o modelo especialista se torna [30].

A seguir, é apresentada a lista de parâmetros utilizada:

- 1) ν : [0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9],
- 2) γ : [1.00, 0.01, 0.001, 0.0001, $\frac{1}{n \times \sigma^2}$, $\frac{1}{n}$],
- 3) *kernel*: ['polinomial', 'base radial', 'sigmoide']

A busca em *grid* calcula os indicadores para todas as combinações de parâmetros supracitadas. Com o objetivo de melhor exposição dos resultados, a combinação de parâmetros mais bem-sucedida, [$\nu = 0.4$, $\gamma = 0.0001$, *kernel* = 'base radial'], isto é, com o maior índice SP, será a única reportada na seção de resultados ao ser comparada com os outros modelos.

B. Modelo de Deep Learning básico

Uma CNN é aplicada aos dados originais a fim de gerar um classificador que identifique tuberculose em dados de imagem. Redes convolucionais são um grupo de redes neurais artificiais que processam dados dotados de topologia em *grid* [4].

Ao se trabalhar com imagens de radiografia, a utilização de uma CNN mostra-se adequada para identificar as variáveis representativas das imagens de entrada e identificar quais imagens apresentam tuberculose ativa e quais não apresentam.

Como discutido no One Class SVM, uma CNN bem ajustada deve atingir uma alta sensibilidade sem que a especificidade seja desconsiderada, o que é obtido quando o modelo retorna um alto índice SP.

A arquitetura de uma rede convolucional varia de aplicação a aplicação. A seguir são enumeradas as características da CNN utilizada.

- 1) Quatro convoluções de *kernel* tamanho 3 e função ativação ReLu.
- 2) Todas convoluções são seguidas de uma camada de *MaxPooling*.
- 3) Uma camada de dropout de 0.5 é aplicada antes da camada de saída.
- 4) A saída é do tipo sigmoide.

A arquitetura é customizada para este projeto e obtida através de experimentações e o treinamento é realizado em 20 épocas com *early-stop* definido para encerrar o treinamento quando o índice SP se afasta do máximo, com tolerância de três épocas.

C. Data-Augmentation utilizando uma GAN de Wasserstein

Dentre uma série de modelos generativos adversariais (GAN), um dos utilizados neste artigo é o Wasserstein GAN [19].

Essa abordagem introduz ao jogo adversarial o conceito de *Earth Mover Distance* e sua aproximação. Tal contribuição beneficia o acompanhamento do processo de treinamento.

Dessa forma, utilizamos a WGAN para estimar funções geradoras que aproximem a distribuição dos dados de imagem de Shenzhen. Assim, dois modelos especialistas são estimados:

- 1) Função geradora que aproxime a distribuição das imagens de radiografia de tórax de indivíduos com tuberculose ativa.
- 2) De forma análoga, função geradora que aproxima a distribuição das imagens de radiografia de tórax de indivíduos sem tuberculose ativa.

As imagens são separadas em *batches* de tamanho 32 e o jogo adversarial é iterado em quatro mil épocas, para cada modelo.

A versão da WGAN de gradiente penalizado é implementada, com a parametrização baseada no artigo [20]. Isto é, a regra de otimização é o Adam, com $\beta_1 = 0.5$ e $\beta_2 = 0.999$ e $\lambda = 10$.

A arquitetura convolucional é baseada no artigo [18], com adaptações feitas para este projeto.

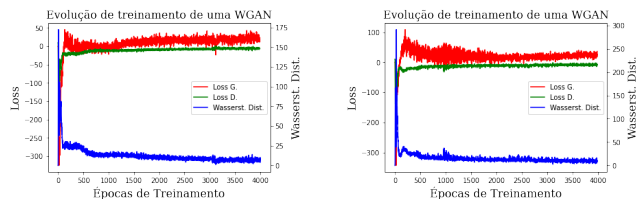
- 1) **Gerador**: três camadas de convolução com *kernel* de tamanho 3, *stride* de tamanho 1 e *padding* de tamanho 1.
- 2) **Discriminador**: quatro camadas de convolução com *kernel* de tamanho 3, *stride* de tamanho 2 e *padding* de tamanho 1.

As redes são inicializadas devidamente e toma-se o cuidado de atualizar os gradientes do gerador a cada cinco passos de atualização do discriminador, seguindo as recomendações do artigo [19]. Na Figura 1, é possível ver os valores de perda do gerador (Loss G), de perda do discriminador (Loss D), bem como a medida de distância de Wasserstein (Wasserst. Dist.) durante o treinamento.

É possível verificar que a distância de Wasserstein atinge valor aproximadamente zero de forma estável a partir de mil iterações de atualização do gerador. Continua-se com o jogo adversarial de modo a buscar uma maior nitidez da imagem, porém sabendo que regiões do pulmão podem apresentar ruído introduzido pelo gerador, o que, em consequência, pode influenciar negativamente na performance do classificador.

A Figura 2 ilustra o quão diverso o gerador de imagens sintéticas se torna após ser treinado. Como citado na introdução, a diversidade é a principal motivação de se trabalhar com modelos generativos adversariais na tarefa de *data-augmentation*, embora essa etapa torne o *pipeline* mais complexo se comparado a operações matemáticas algébricas sobre as imagens.

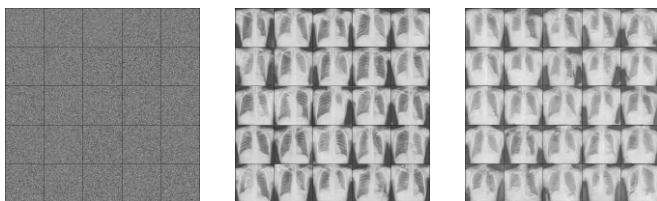
Figura 1



A - Modelo especialista em radiografias com tuberculose ativa

B - Modelo especialista em radiografias sem tuberculose ativa

Figura 2



A - Exemplo de imagens no passo inicial do jogo adversarial. Gerador parte de ruídos gaussianos.

B - Modelo especialista em radiografias com tuberculose ativa após 4 mil iterações de jogo adversarial.

C - Modelo especialista em radiografias sem tuberculose ativa após 4 mil iterações de jogo adversarial.

D. GAN Pix2Pix no contexto de Data-Augmentation e detalhes de seu treinamento

Espera-se que a CNN tome decisão baseada na região patológica, que no caso da tuberculose concentra-se majoritariamente na região do pulmão. Dessa forma, este projeto busca na GAN *Pix2Pix* uma alternativa que possibilite a criação de radiografias sintéticas, sem que a região do pulmão esteja suscetível a distorções, como pode ocorrer com a GAN de Wasserstein. Esta é a razão pela qual se recorre à amostragem através da GAN *Pix2Pix*.

O modelo *Pix2Pix* requer imagens pareadas para aprender o mapeamento entre entrada, em um dado contexto, e saída, para um contexto distinto [21].

Em imagens de face, por exemplo, o problema seria traduzir diferentes expressões faciais, em que a imagem de entrada do modelo seria de uma pessoa sorrindo e sua respectiva saída seria a face dessa pessoa chorando. No caso em questão, temos por objetivo manter a região do pulmão sem alterações e traduzirmos o entorno baseado nas diferentes imagens disponíveis para treinamento.

Como há escassez de variadas imagens de pulmão com perfeito alinhamento, mas em diferentes contextos, uma forma de lidar com essa restrição é o pareamento artificial das imagens entre a radiografia de tórax e máscaras de pulmão [11], como é visto na Figura 3.

A necessidade de uma grande quantidade de pareamentos dessa natureza para o treinamento nos leva à implementação

Figura 3



Pareamento artificial para GAN *Pix2Pix*

deste modelo em sua versão não especialista, na qual as duas classes são consideradas. Ao final do processo, o gerador será capaz de criar imagens sintéticas para ambas as classes, sejam elas com tuberculose ou sem tuberculose.

A partir daí, faz-se o *Pix2Pix* aprender como recuperar as regiões faltantes no entorno. Finalmente o gerador será capaz de criar imagens sintéticas sem que a região do pulmão esteja distorcida [11].

Diferentemente da GAN de Wasserstein, a GAN *Pix2Pix* apresenta desafios na medida de qualidade de seu treinamento. Quando o modelo busca gerar imagens sintéticas de objetos, faces ou mapas, a tarefa de avaliação é facilitada, pois há modelos pré-treinados de reconhecimento de imagem ou mesmo sites como Amazon Turk podem facilitar testes de Turing [31] [21] [22].

Como os dados de interesse são radiografias de tórax, houve dificuldade em replicar essas técnicas de avaliação do modelo generativo. Dessa forma, o treinamento separa as imagens em *batches* de tamanho 64 e o jogo adversarial é iterado em 2 mil épocas até encerrar, quando as redes são consideradas treinadas. A Figura 4 compara após treinamento uma imagem original e uma imagem sintética geradas para um mesmo pulmão.

Por fim, as arquiteturas das redes no jogo adversarial possuem as características a seguir. Adicionalmente, seguindo o arcabouço de GAN *Pix2Pix*, uma restrição L1 é acrescentada ao jogo adversarial, de modo a garantir a proximidade das imagens sintéticas às imagens originais:

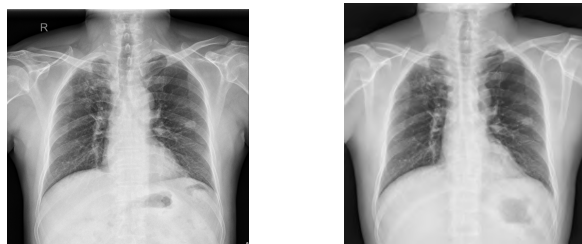
- 1) **Gerador:** Uma *Unet* que recebe imagens de dimensão 256x256. São realizados 8 *downsamplings* na imagem até o *bottleneck*.
- 2) **Discriminador:** Uma rede com cinco camadas convolucionais, tendo *kernel* de tamanho 3, *stride* de tamanho 2 e *padding* de tamanho 1.

III. EXPERIMENTOS E RESULTADOS

Apresentam-se nesta seção os dados e os experimentos aplicados para atender ao método proposto pelo artigo. Em seguida, resultados são debatidos.

Destaca-se que os modelos e os experimentos foram implementados na linguagem de programação Python, em especial, utilizando as bibliotecas *Pytorch*, *Scikit-Learn* e *Keras Tensorflow* para obter os resultados. As visualizações são implementadas a partir da biblioteca *Matplotlib*. Em especial,

Figura 4



A - Radiografia de pulmão com tuberculose ativa extraída dos dados de Shenzhen

B - Radiografia sintética gerada a partir de uma GAN *Pix2Pix* para o mesmo pulmão de A.

o modelo *Pix2Pix* utiliza a biblioteca desenvolvida pelos autores do artigo [11].

A. Dados

Para os experimentos, os dados utilizados são as radiografias de Shenzhen [12] [13], das quais 336 são radiografias de tórax com tuberculose ativa e 326 sem a presença de tuberculose ativa.

Imagens em *png* são importadas, convertidas para escala de cinza e ajustadas para terem dimensão 256x256.

As imagens são tratadas como tensores em todos os modelos, com exceção do modelo One Class SVM, para o qual as imagens são linearizadas e tratadas como vetores⁵.

A GAN *Pix2Pix* exige *bounding box* das radiografias apontando regiões patológicas. Neste artigo, no lugar das áreas de patologia, levam-se em consideração máscaras de pulmão. Máscaras de pulmão para os dados de Shenzhen são apresentadas no trabalho [32]. Este artigo utiliza tais máscaras na estimativa da função geradora da GAN *Pix2Pix*⁶.

B. Experimentos abordados

A fim de avaliar se a utilização dos modelos generativos adversariais na tarefa de *data-augmentation* melhoraram a performance do classificador, o experimento abordado neste artigo foi separado em três etapas.

Primeiramente, os modelos que compõem a referência de aprendizagem são estimados com base nos dados originais, com os resultados apresentados na Tabela I.

Posteriormente, os geradores da GAN de Wasserstein e da GAN *Pix2Pix* são acionados para gerar duas mil imagens sintéticas balanceadas entre as classes, de modo que uma nova CNN é estimada nesses dados sintéticos. Os resultados dessa etapa são considerados intermediários e estão presentes na Tabela II.

⁵O conjunto de dados é o mesmo para todos os modelos e a vetorização em linha das imagens é um pré-processamento necessário para adequação ao modelo One Class SVM. No entanto, outros tratamentos, além da vetorização em linha, podem ser dados às imagens, como vetorização em coluna, serpentina etc. Essas abordagens estarão em trabalhos futuros.

⁶Futuramente a segmentação será integrada como pré-processamento, de modo a não necessitar de máscaras previamente geradas por outros projetos.

Finalmente, na terceira e última etapa, os pesos da CNN obtidos na fase intermediária são utilizados para inicializar uma nova CNN, que será estimada com os dados originais. Dessa forma, transfere-se o conhecimento adquirido com os dados sintéticos para a estimativa com dados originais. Isso é possível porque os dados sintéticos são gerados por GAN, que procuram aproximar a função de densidade de probabilidade dos dados, mas sem buscar a obtenção de cópias dos mesmos.

Assim, os resultados presentes na Tabela III devem ser comparados com os valores obtidos no *baseline*, para verificar-se se a utilização de dados sintéticos aumenta a performance da CNN classificadora. Busca-se o modelo que tenha índice SP médio máximo acompanhado de estimativas de barra de erro para a tarefa de identificação de tuberculose ativa.

Em todas as abordagens supracitadas, o treinamento da CNN e a calibragem do One Class SVM são obtidos através de validação cruzada [33], pela qual se separam os dados em cinco *folds*, com teste flutuante, e obtendo os indicadores de sensibilidade, de especificidade e de SP.

A validação cruzada com teste flutuante se dá através da separação dos dados em treino, validação e teste. O conjunto de dados de treinamento é utilizado no ajuste de parâmetros, enquanto o conjunto de validação é utilizado para o critério de parada do treinamento impedir *overfitting*. Já o conjunto de teste não participa de nenhuma etapa do treinamento, sendo utilizado apenas para medir a generalização do modelo treinado.

Neste trabalho, iteramos cinco vezes, separando um conjunto teste por vez. Após separar o conjunto teste, os dados restantes são separados entre treino e validação iterativamente também por cinco vezes, o que significa dizer que, a cada conjunto teste separado, cinco modelos são treinados e validados. Ao final da separação dos cinco *folds* teste, 25 modelos são treinados, validados e testados. Assim, são reportados nas três tabelas de resultados experimentais os valores médios destes indicadores e suas respectivas barras de erro calculadas a partir dos *folds*.

Tabela I: Modelos de referência de aprendizagem (*baseline*).

Indic.	Conj. Dados	One Class SVM	CNN
SP	Treino	0.62 +- 0.01	0.84 +- 0.03
	Val.	0.63 +- 0.02	0.85 +- 0.02
	Teste	0.63 +- 0.04	0.83 +- 0.04
Sensib.	Treino	0.60 +- 0.001	0.81 +- 0.04
	Val.	0.58 +- 0.07	0.81 +- 0.06
	Teste	0.59 +- 0.07	0.79 +- 0.06
Especif.	Treino	0.65 +- 0.03	0.87 +- 0.04
	Val.	0.69 +- 0.05	0.88 +- 0.05
	Teste	0.67 +- 0.07	0.86 +- 0.04

Grande assertividade da CNN, quando aplicada apenas a dados sintéticos gerados pela WGAN e pela GAN *Pix2Pix*, sinaliza que a CNN é bem-sucedida em separar as imagens em classes distintas. Por sua vez, esse resultado não é garantia de boa performance da CNN quando operando com imagens reais, pois os geradores sintéticos incorrem em especificidades

Tabela II: Modelos de aprendizado profundo treinados apenas com imagens sintéticas.

Indic.	Conj. Dados	CNN WGAN	CNN Pix2Pix
SP	Treino	0.99 +- 0.001	0.90 +- 0.05
	Val.	0.99 +- 0.005	0.88 +- 0.05
	Teste	0.99 +- 0.004	0.87 +- 0.04
Sensib.	Treino	0.99 +- 0.002	0.87 +- 0.07
	Val.	0.98 +- 0.007	0.82 +- 0.08
	Teste	0.98 +- 0.006	0.83 +- 0.07
Especif.	Treino	0.99 +- 0.002	0.92 +- 0.04
	Val.	0.99 +- 0.004	0.92 +- 0.04
	Teste	0.99 +- 0.004	0.91 +- 0.04

Tabela III: Modelos de aprendizagem transferida dos dados sintéticos para os dados originais.

Indic.	Conj. Dados	CNN WGAN	CNN Pix2Pix
SP	Treino	0.83 +- 0.02	0.88 +- 0.03
	Val.	0.84 +- 0.03	0.87 +- 0.04
	Teste	0.82 +- 0.03	0.86 +- 0.04
Sensib.	Treino	0.80 +- 0.04	0.85 +- 0.05
	Val.	0.79 +- 0.05	0.82 +- 0.05
	Teste	0.76 +- 0.05	0.80 +- 0.09
Especif.	Treino	0.85 +- 0.04	0.90 +- 0.04
	Val.	0.87 +- 0.05	0.91 +- 0.04
	Teste	0.87 +- 0.04	0.90 +- 0.05

no treinamento que podem atribuir ruídos às duas classes inexistentes na distribuição dos dados originais.

Isto posto, ao inicializarmos uma CNN que será estimada baseada nos dados originais com os pesos da CNN que já foi ajustada aos dados sintéticos, espera-se um aumento de performance no treinamento, na validação do treinamento e por fim na generalização com a flutuação medida nos cinco conjuntos de teste.

Fica visível, quando se leem os resultados da Tabela I e da Tabela III, que todas as abordagens que envolvem CNN têm melhor performance se comparadas ao One Class SVM. Por outro lado, não observamos ganho em transferir conhecimento dos dados sintéticos gerados a partir da GAN de Wasserstein para CNN, uma vez que a CNN aplicada apenas nos dados originais tem melhor performance em todos os cenários.

Por fim, quando os dados sintéticos são gerados pela GAN *Pix2Pix*, observamos ganho no processo de treinamento, de validação e generalização se comparado ao modelo aplicado apenas aos dados originais.

Dessa forma, a estratégia de gerar dados sintéticos através de uma GAN *Pix2Pix* se mostra próspera para auxiliar na estimativa de um classificador especialista em identificar tuberculose ativa.

IV. CONCLUSÃO E TRABALHOS FUTUROS

Em resumo, este artigo aplica técnicas de aprendizado profundo em dados de saúde pública, com o interesse em contribuir para estratégias de prevenção e combate à tuberculose.

Para isso, dados públicos de radiografia de tórax são utilizados na estimativa de uma rede neural convolucional profunda com papel de classificadora de imagens que apresentem tuberculose ativa.

Técnicas avançadas de modelos generativos adversariais foram utilizadas em uma estratégia de introdução de novos dados amostrados de uma WGAN ou gerados a partir de uma GAN *Pix2Pix*, a fim de auxiliar a rede neural profunda a ajustar sua vasta quantidade de parâmetros. A abordagem escolhida se deu através de transferência de aprendizagem entre um treinamento apenas em dados sintéticos para um treinamento com dados originais. Foram verificados ganhos no processo de treinamento quando os dados sintéticos são gerados pela GAN *Pix2Pix*.

Este trabalho está em andamento, logo há espaço para abordarmos os diferentes passos da pesquisa com maiores detalhes.

O primeiro deles está na vetorização das imagens pelo método serpentina [34], que pode ser implementado para o *baseline* do One Class SVM como alternativa à vetorização em linha implementada neste projeto.

Ainda sobre o One Class SVM, será interessante utilizar dados sintéticos no seu treinamento, de forma análoga ao explorado com a CNN neste artigo.

Outro passo futuro será a introdução das GAN na validação cruzada das CNN, fazendo com que o treinamento da GAN seja imediatamente sucedido pelo treinamento da CNN para cada separação de treino, validação e teste. A abordagem escolhida para este projeto foi tratar a etapa generativa adversarial como independente do processo de estimativa da CNN, o que permite avaliar o incremento da utilização dos dados sintéticos através da transferência do conhecimento, mas impede que validações mais cuidadosas das GAN sejam realizadas.

Essas validações das GAN podem ser feitas por especialistas ou por técnicas de inferência aproximada, que também serão tópicos a serem investigados em passos futuros.

AGRADECIMENTOS

Os autores gostariam de agradecer à FAPERJ e CNPq pelo apoio a este trabalho.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

REFERÊNCIAS

- [1] Goodfellow I., Bengio Y. e Courville A. - Deep Learning Book. MIT Press <http://www.deeplearningbook.org> (2016)
- [2] Zhang A., Lipton Z. C., Li M., Smola A. J. - Dive into Deep Learning, arXiv preprint arXiv:2106.11342, (2021)
- [3] WHO's global tuberculosis report <https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf> (2020)
- [4] LeCun Y. - Generalization and Network Design Strategies, in Pfeifer, R. e Schreter, Z. e Fogelman, F. e Steels, L. (Eds), Connectionism in Perspective, Elsevier, Zurich, Switzerland, an extended version was published as a technical report of the University of Toronto, 1989
- [5] Shorten C., Khoshgoftaar T.M. - A survey on Image Data Augmentation for Deep Learning. J Big Data 6, 60 (2019). <https://doi.org/10.1186/s40537-019-0197-0>

- [6] Eaton-Rosen Z., Bragman F. J. S., Ourselin S. e Cardoso M. J. - Improving Data Augmentation for Medical Image Segmentation *In International Conference on Medical Imaging with Deep Learning* (2018)
- [7] Esteva A., Kuprel B., Novoa R.A., Ko J., Swetter S.M., Blau H.M. e Thrun S. - Dermatologist-level classification of skin cancer with deep neural networks. *Nature* 542 (7639) (2017)
- [8] Frid-Adar M., E. Klang M., Amitai M. e Goldberger J. - Synthetic data augmentation using gan for improved liver lesion classification - Conference: 2018 IEEE 15th International Symposium on Biomedical Imaging (DOI:10.1109/ISBI.2018.8363576) (2018)
- [9] Goodfellow I. J. , Pouget-Abadie J. , Mirza M., Xu B., WardeFarley D., Ozair S., Courville A. e Bengio Y. - Generative adversarial nets. *In Advances in Neural Information Processing Systems* 27, pages 2672–2680. Curran Associates, Inc., (2014)
- [10] Krizhevsky A., Sutskever I., Hinton G. E. - ImageNet classification with deep convolutional neural networks - *Advances in neural information processing systems*, Volume 25, Pages 1097-1105 (2012)
- [11] Xing Y., Ge Z., Zeng R., Mahapatra D., Seah J., Law M. e Drummond T. - Adversarial Pulmonary Pathology Translation for Pairwise Chest X-ray Data Augmentation - *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Pages 757-765. Springer, Cham (2019)
- [12] National Library of Medicine, National Institutes of Health, Bethesda, MD, USA and Shenzhen No.3 People's Hospital, Guangdong Medical College, Shenzhen, China
- [13] Jaeger S., et al. - Automatic tuberculosis screening using chest radiographs. - *IEEE Trans Med Imaging* 2014 Feb;33(2):233-45. doi: 10.1109/TMI.2013.2284099 (2014)
- [14] Yadav S.S. e Jadhav S.M. - Deep convolutional neural network based medical image classification for disease diagnosis. *J Big Data* 6, 113. <https://doi.org/10.1186/s40537-019-0276-2> (2019)
- [15] Oliveira D. A. B. - Controllable Skin Lesion Synthesis Using Texture Patches, Bézier Curves and Conditional GANs, 2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI), pp. 1798-1802, doi: 10.1109/ISBI45749.2020.9098676. (2020)
- [16] Salimans T., Goodfellow I., Zaremba W., Cheung V., Radford A. e Chen X. - Improved Techniques for Training GANs - NIPS'16: Proceedings of the 30th International Conference on Neural Information Processing Systems (2016)
- [17] Mirza M. e Osindero S. - Conditional Generative Adversarial Nets - arxiv:1411.1784. (2014)
- [18] Radford A., et. al. - Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks - arxiv:1511.06434 (2015)
- [19] Arjovsky M., et. al - Wasserstein GAN - Proceedings of the 34th International Conference on Machine Learning, pages 214-223, editor Precup, Doina and Teh, Yee Whye, volume 70, series Proceedings of Machine Learning Research, publisher PMLR, <http://proceedings.mlr.press/v70/arjovsky17a.html> (2017)
- [20] Gulrajani I., et. al. - Improved Training of Wasserstein GANs - NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems December 2017 Pages 5769–5779 (2017)
- [21] Isola P., Zhu J., Zhou T. e Efros A. A. - Image-to-Image Translation with Conditional Adversarial Networks - 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017, pp. 5967-5976, doi: 10.1109/CVPR.2017.632. (2017)
- [22] Zhu J., Park T., Isola P. e Efros A. A. - Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks, 2017 IEEE International Conference on Computer Vision (ICCV), 2017, pp. 2242-2251, doi: 10.1109/ICCV.2017.244. (2017)
- [23] Choi Y., Choi M., Kim M., Jung-Woo Ha, Kim S. e Choo J. - StarGAN: Unified Generative Adversarial Networks for Multi-Domain Image-to-Image Translation - *Computer Vision and Pattern Recognition (CVPR)* (2017)
- [24] Reid A., Grant A. D., White R. G., Dye C., Vynnycky E., Fielding K., Churchyard G. e Pillay Y. - Accelerating progress towards tuberculosis elimination: the need for combination treatment and prevention - *The International Journal of Tuberculosis and Lung Disease*, Volume 19, Number 1, 1 January 2015, pp. 5-9(5) - International Union Against Tuberculosis and Lung Disease (DOI: <https://doi.org/10.5588/ijtld.14.0078>) (2015)
- [25] Uplekar et. al. - WHO's new end TB strategy (2015) - *The Lancet*, VOLUME 385, ISSUE 9979, P1799-1801, MAY 02, 2015, Published:March 23, 2015 DOI:[https://doi.org/10.1016/S0140-6736\(15\)60570-0](https://doi.org/10.1016/S0140-6736(15)60570-0)
- [26] Menzies et. al. - Four Months of Rifampin or Nine Months of Isoniazid for Latent Tuberculosis in Adults - *N Engl J Med* 2018;379:440-453 DOI: 10.1056/NEJMoa1714283 (2018)
- [27] Torres R. C., De Lima D. E. F., De Simas Filho E. F. e De Seixas J. M. - Neural online filtering based on preprocessed calorimeter data, in 2009 IEEE Nuclear Science Symposium Conference Record (NSS/MIC), pp. 530–536, 2009.
- [28] Simas Filho E. F., De Seixas J. M., e Calôba L. P. - Modified postnon-linear ica model for online neural discrimination, *Neurocomputing*, vol. 73, no. 16, pp. 2820–2828, 2010. 10th Brazilian Symposium on Neural Networks (SBRN2008)
- [29] Schölkopf B., Smola A. J. - *Learning with Kernels - Support Vector Machines, Regularization, Optimization, and Beyond* - MIT Press, 55 Hayward St. Cambridge, MA, United States - ISBN:978-0-262-19475-4 (1998)
- [30] N. Junior - Detecção de novidade para sistemas de sonar passivo. Tese de Doutorado apresentada ao Programa de Pós-graduação em Engenharia Elétrica, COPPE, da Universidade Federal do Rio de Janeiro, como parte dos requisitos necessários à obtenção do título de Doutor em Engenharia Elétrica. (2018)
- [31] Zhang R., Isola P., Efros A.A. - Colorful Image Colorization. In: Leibe B., Matas J., Sebe N., Welling M. (eds) *Computer Vision – ECCV 2016*. ECCV 2016. Lecture Notes in Computer Science, vol 9907. Springer, Cham. https://doi.org/10.1007/978-3-319-46487-9_40 (2016)
- [32] Pandey N. - Pulmonary Chest X-Ray Defect Detection - Kaggle (2019)
- [33] M. Stone - Cross-Validatory Choice and Assessment of Statistical Predictions. *Journal of the Royal Statistical Society, Series B (Methodological)*. 36 (2): 111–147. (1974)
- [34] Gonzalez R. C., Woods R. E. - *Digital Image Processing (3rd Edition)* Prentice-Hall, Inc. Division of Simon and Schuster One Lake Street Upper Saddle River, NJ United States ISBN:978-0-13-168728-8 (2006)
- [35] Candemir S., et al. - Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration - *IEEE Trans Med Imaging* 2014 Feb;33(2):577-90. doi: 10.1109/TMI.2013.2290491 (2014)
- [36] Filho R. T. A. - Aplicação de redes neurais no controle de tuberculose bovina. Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.) (2016)