

GAN Pix2Pix e a geração de imagens médicas sintéticas para dados públicos de pacientes com tuberculose

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

João Victor Pinto da Fonseca

Laboratório de Processamento de Sinais, COPPE/POLI,
Universidade Federal do Rio de Janeiro.
Rio de Janeiro, Brasil
jodafons@coppe.ufrj.br

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—A tuberculose (TB) está entre as doenças que mais matam no mundo. A fim de enfrentar o enorme desafio de diminuir sua incidência até alcançar a definitiva eliminação da TB no mundo, a Organização Mundial de Saúde (OMS) recomendou recentemente a utilização de um CAD (Computer-Aided Detection) baseado em imagens radiológicas do tórax. O aprendizado profundo vem sendo utilizado com frequência em diversas tarefas aplicadas em imagens médicas, sendo uma opção valiosa para a operação de tal CAD, notadamente em TB infecção, quando o paciente está infectado, mas não desenvolve a doença. Este artigo discute técnicas de aprendizagem profunda aplicadas ao problema de geração sintética de imagens de radiografia, visando agregar informação para a base de dados de imagens reais e, assim, enfrentar a necessidade de um grande volume de dados para o treinamento de modelos profundos. Para isso, Redes Neurais Adversariais (GAN) são aplicadas em uma base de dados pública, composta por exames de imagem de pacientes com suspeita de tuberculose. Uma GAN pareada de tradução de contexto, GAN Pix2Pix, é explorada neste trabalho, havendo a penalização de uma dada região de interesse da imagem. No caso, a segmentação do pulmão é utilizada como região de interesse. Os resultados avaliam a capacidade de produção de imagens sintéticas neste contexto.

Index Terms—Aprendizado profundo, Tuberculose, Aprendizado de Máquina na Saúde, *Generative Adversarial Networks* (GAN), GAN Pix2Pix.

I. INTRODUÇÃO

Redes neurais generativas adversariais (GAN) [1] são redes neurais profundas que possuem em seu treinamento um jogo. Nesse jogo, a rede geradora busca estimar a função geradora de eventos da distribuição de probabilidade dos dados, tendo nenhum ou pouco conhecimento dos dados experimentais (reais). A rede discriminadora detém acesso aos dados reais e recebe as tentativas do gerador, convencionadas como dados sintéticos (*fake*), em cada iteração de treinamento.

A rede discriminadora avalia a similaridade entre a tentativa do gerador e o dado real que ela observa. A avaliação é feita através de variadas funções custo, dentre as quais destacamos a adversarial clássica (probabilidade) [1], por distância de Wasserstein [2], por norma \mathcal{L}_{L1} combinada com avaliação adversarial [3] e por consistência de ciclo [4].

Há uma diversidade de aplicações envolvendo modelos GAN disponíveis na literatura [3]. Esses modelos são utilizados para sintetizar imagens [5], sons [6], vídeos [7] etc.

Neste trabalho, estamos particularmente interessados em explorar um dos tipos de GAN no problema de geração de imagens médicas, sobretudo radiografias de tórax de pacientes que tenham suspeita de infecção com a bactéria da tuberculose (TB).

A utilização de GAN para geração de imagens médicas, já é técnica consolidada na literatura. Os geradores devidamente treinados podem alimentar modelos com muitos parâmetros, reconstruir imagens com ruído e realizar tradução imagem a imagem em diferentes contextos [8].

Das aplicações citadas, temos o particular interesse nos modelos GAN que realizam tradução imagem a imagem. Nesses modelos, geradores recebem pares de imagens e realizam a tradução entre o contexto da imagem de entrada e a imagem objetivo [3].

Este artigo visa contribuir para o problema de *screening* de pacientes que possuam indicativo de tuberculose ativa em sistemas de detecção automatizados (*Computer Aided Detection - CAD*). Modelos CAD usam a informação de radiografia do tórax e frequentemente lançam mão de redes neurais convolucionais profundas, a fim de classificar e atribuir alta probabilidade à radiografia que apresenta a manifestação de tuberculose em sua forma ativa, para retirar esses casos da

linha de tratamento de pacientes com tuberculose infecção¹ [9], e, dessa forma, tratar ambas as manifestações adequadamente.

A Organização Mundial de Saúde (OMS) possui um plano de eliminação da tuberculose em suas diversas manifestações até 2035 [10]. Neste plano, a entidade incentiva o desenvolvimento de sistemas CAD para auxiliarem o trabalho de radiologistas no processo de triagem e *screening* de pacientes, sobretudo em áreas vulneráveis do ponto de vista social [11].

Além de incentivar o desenvolvimento, a entidade define critérios básicos de qualidade que modelos dessa natureza devem atender. Após treinados e validados, os modelos devem apresentar alta sensibilidade² sem que a especificidade³ seja desconsiderada.

No que diz respeito à alta sensibilidade, a OMS define como sendo úteis os modelos que possuam essa taxa acima de noventa por cento (90%). Para especificidade, é esperada uma taxa acima de setenta por cento (70%). Esses limiares nos permitem identificar uma região na curva ROC⁴ [12], onde estejam presentes as combinações de sensibilidade e especificidade de modelos que seriam recomendados para um CAD.

Modelos classificadores desenvolvidos com redes neurais profundas possuem muitos parâmetros a serem ajustados. Alinha-se a este fato a alta eficiência requerida para a proposição de sistemas CAD. Assim, é necessária a disponibilidade de uma grande quantidade de imagens de radiografias de tórax com qualidade atestada.

No entanto, mesmo havendo bases públicas disponíveis, como em [13] e [14], elas ou não possuem qualidade atestada ou dispõem de baixa estatística, dentre outras limitações [15].

Dessa forma, a geração de imagens sintéticas surge como recurso de contribuição para modelos neurais profundos, que possuem alta capacidade de impacto na cadeia de tratamento de pacientes dessa doença altamente infecciosa [16], conhecida por atingir populações em áreas densas e mais afetadas pela desigualdade social e pela pobreza.

Com isso, propomos a utilização da GAN Pix2Pix [3] para a geração de imagens sintéticas a partir de dados de radiografias da base pública de Shenzhen.

Utilizamos esta abordagem, pois temos o objetivo de preservar uma região de interesse, no caso o recorte de pulmões, visando à menor interferência possível nos achados radiológicos, bem como de gerar variabilidade no entorno dos pulmões, de modo a contribuir com a introdução de diversidade nas bases de dados de estudo.

¹Nos casos de Tuberculose Infecção, o paciente não apresenta sinais da doença em seu pulmão, como ocorre com os pacientes que possuem a tuberculose ativa.

²A sensibilidade de um modelo está relacionada à alta probabilidade de detecção de uma classe de interesse.

³A especificidade de um modelo captura uma estimativa da sua taxa de acerto em pacientes que não tenham a doença durante a realização de *screening*, sendo o complemento da probabilidade de falso alarme.

⁴Curvas denominadas *Receiver Operating Characteristics* (ROC) são úteis para organizar classificadores e visualizar sua performance [12].

A GAN Pix2Pix é um caso particular dos modelos GAN condicionais [17]. GAN do tipo condicional aplicada a imagens médicas é encontrada em [18], [19], [20], [21], [22] e [23]. Por sua vez, modelos GAN aplicados a radiografias são encontrados em [24] e [25], o que advoga para a solidez na literatura do método empregado neste trabalho.

A utilização de GAN Pix2Pix em exames de imagem em radiografias é encontrada no artigo [26], de tomografia computadorizada é encontrada em [27] e MRI em [28]. No entanto, geração de imagens sintéticas a partir deste método para pacientes com suspeita de tuberculose ainda é aplicação a ser explorada pela literatura.

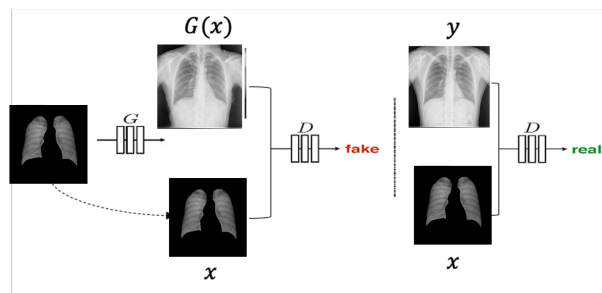
Portanto, podemos resumir a contribuição deste trabalho em duas frentes: aplicação da GAN Pix2Pix para geração de imagens sintéticas de pacientes com suspeita de tuberculose; geração de imagens sintéticas que preservem o achado radiológico, neste contexto.

II. MÉTODO E ARQUITETURAS DOS MODELOS

Modelos generativos adversariais de tradução de domínios são úteis para o enriquecimento de bases de dados, pois são capazes de preservar regiões de interesse das imagens [29] [15]. A GAN Pix2Pix, exemplo de GAN de tradução de contexto, permite o treinamento de um gerador que prioriza determinadas regiões da imagem para criar informação sintética (*fake*) [15].

Os dados sintéticos são gerados, $G(x)$, como correspondências pareadas dos dados originais. Isto é, o jogo adversarial recebe um par de imagens em que o gerador tem uma região de interesse, x , como entrada e o discriminador tem acesso à imagem experimental completa, y . Neste artigo, trabalhamos com radiografias de tórax como imagens a ser sintetizadas, tendo os pulmões como região de interesse. Este processo é ilustrado na figura 1.

Figura 1: Esquema treinamento da GAN Pix2Pix. Adaptado de [3].



O gerador treina para introduzir informação sintética livremente na imagem que recebeu como entrada. Já o discriminador avalia as imagens geradas sinteticamente e as compara com as imagens experimentais, retornando probabilidades de as imagens sintéticas serem oriundas da mesma distribuição que as imagens experimentais.

Como o gerador é condicional à região de interesse de entrada, o discriminador exerce papel de penalizador na região

de interesse, uma vez que atribui maior probabilidade de pertencimento ao conjunto de imagens experimentais, para essa região da imagem.

Assim, utilizamos esta técnica para criar imagens sintéticas que tenham a menor interferência possível na região dos pulmões, pois, embora o gerador esteja livre para fazer alterações nesta região, o discriminador penaliza essas intervenções.

Por ser limitada à segmentação da região de interesse, a produção sintética é capaz de, no máximo, dobrar o tamanho da amostra dos dados. Essa baixa capacidade de geração de eventos é consequência do condicionamento das regiões de interesse ao processo de geração de imagens.

O gerador da GAN Pix2Pix tem arquitetura Unet-256⁵ [30], como se apresenta na figura 2-A. A Unet é uma rede neural convolucional que a cada camada concentra a informação da imagem em representações latentes até a camada intermediária denominada de *bottleneck*, de dimensão reduzida e informação concentrada, de modo a possibilitar a reconstrução da imagem.

A partir da camada intermediária, ocorre o processo de reconstrução da imagem em uma sequência de camadas convolucionais transpostas, até chegar a dimensão 256x256 da imagem de entrada. É importante destacar que, a cada camada convolucional, a Unet realiza operações de *skip connections*⁶ entre as camadas de compressão (*downsampling*) e reconstrução (*upsampling*), de modo a evitar perda exacerbada de informação durante a compressão e reconstrução.

O discriminador, conhecido como patchGAN [31], figura 2-B, tem por característica passar por toda imagem através de janelas, *patches*, de acordo com um tamanho pré-definido⁷, calculando a probabilidade de, dada a mesma região de interesse de origem, a janela sintética pertencer ao conjunto de janelas experimentais⁸.

Ao final do processo de caminhada por toda a imagem, o discriminador toma a média amostral dessas probabilidades, com igual peso para cada janela, definindo a estatística de avaliação da verossimilhança entre a imagem gerada sinteticamente e a imagem real.

É através dessa característica do discriminador que introduzimos a penalização, pois podemos segmentar a imagem de raio-x de tórax em uma região de interesse, no caso, onde está posicionado o pulmão na imagem, e definirmos a região segmentada como entrada do gerador. Condicional ao pulmão, o gerador produz uma imagem sintética e o discriminador penaliza a introdução de informação na região de interesse.

De posse do gerador e do discriminador, a GAN Pix2Pix é treinada por mil (1000) épocas, sendo quinhentas (500)

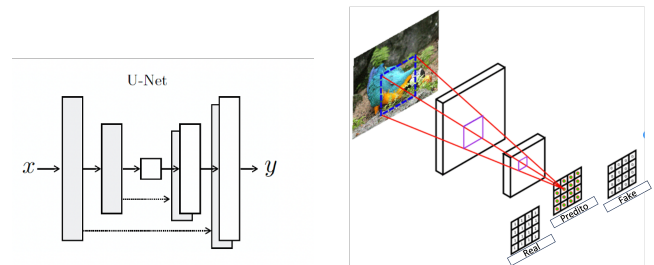
com o coeficiente de aprendizagem (*learning rate*)⁹ fixo e nas quinhentas (500) épocas restantes com caimento linear. O otimizador é do tipo *Adam*, com seus gradientes sendo atualizados de acordo com a função custo do jogo adversarial acrescido de uma regularização de norma \mathcal{L}_{L1} ¹⁰. O jogo adversarial é descrito pela seguinte equação:

$$\mathcal{L}_{cGAN} = E_{x,y} [\log \mathcal{D}(x, y)] + E_{x,z} [1 - \log \mathcal{D}(x, \mathcal{G}(x, z))] \quad (1)$$

$$\mathcal{L}_{L1} = E_{x,y,z} \|y - \mathcal{G}(x, z)\|_1 \quad (2)$$

$$\mathcal{G}^* = \arg \min_{\mathcal{G}} \arg \max_{\mathcal{D}} \mathcal{L}_{cGAN}(\mathcal{G}, \mathcal{D}) + \lambda \mathcal{L}_{L1}(\mathcal{G}) \quad (3)$$

Figura 2: Arquiteturas do gerador (A) e do discriminador (B).



A - Arquitetura da rede Unet. Imagem adaptada do artigo base da GAN Pix2Pix [3]

B - Arquitetura da *patchGAN*. Imagem adaptada do artigo base da *patchGAN* [31]

III. TREINAMENTO E PRODUÇÃO SINTÉTICA

Os treinamentos consistem na estimativa da GAN Pix2Pix para um conjunto de dados, levando em consideração o método de flutuação estatística, para, finalmente, gerar uma produção de dados sintéticos. São treinados modelos classe especialistas.

A abordagem classe especialista é a escolhida, pois evitamos que artefatos presentes nas imagens da classe TB incorporem o ajuste de parâmetros do gerador para imagens da classe NTB¹¹. Isto é, o modelo classe especialista é uma escolha para termos a segurança de que os geradores não sejam contaminados por características presentes em outras classes.

O modelo classe especialista evita a introdução de viés, porém ao custo de utilizar baixa estatística no treinamento de cada modelo e aumentando o custo computacional. Para lidar com a baixa estatística, realizamos validação cruzada, através do método *k-fold*, com estimativa de diversos geradores e com acompanhamento de suas características. A observação

⁵As radiografias são importadas com dimensão 256x256, conseqüentemente a rede Unet tem por objetivo comprimir e reconstruir uma imagem com essa dimensão.

⁶São conexões intermediárias entre as camadas para evitar o problema de saturação dos gradientes no processo de compressão e reconstrução.

⁷Seguimos o padrão de 70x70 recomendado pelos autores do artigo base da GAN Pix2Pix [3].

⁸Supõe-se que os pixels que pertencem a um determinado *patch* são independentes entre si, mas possuem relação estrutural com os pixels pertencentes aos outros *patch* da imagem.

⁹Tamanho do passo de atualização dos gradientes das redes pertencentes à arquitetura.

¹⁰A constante λ tem valor de 100, como é sugerido pelos autores do artigo base da GAN Pix2Pix [3].

¹¹De forma análoga, evitamos que artefatos presentes nas imagens da classe NTB incorporem o ajuste de parâmetros do gerador para imagens da classe TB.

de convergência, capacidade de geração de imagens em alta qualidade e diversas entre si são atestadas nos experimentos.

Optamos por aumento do custo computacional para controlar qualquer possibilidade de introdução de viés aos geradores, pois a perda de controle dos artefatos presentes nas imagens pode contaminar os classificadores que essas imagens alimentarão, levando à consequência adversa de diagnósticos passíveis de dúvida.

A. Dados

A base de dados utilizada neste projeto é pública, coletada em Shenzhen, na China [13], com exames de radiografia de tórax de pessoas com suspeita de contaminação por tuberculose. A base de dados de Shenzhen possui seiscentas e sessenta e duas (662) imagens.

A representação dos metadados de laudo é consolidada em uma variável que designa se o paciente apresenta radiografia indicando tuberculose em sua forma ativa (TB) ou se o exame aponta radiografia do tipo normal (NTB).

Este padrão de referência é obtido através do processo de leitura humana, realizado por três radiologistas [13].

Utilizamos as máscaras de pulmão para definir as regiões de interesse das imagens. Essas máscaras são disponibilizadas publicamente no trabalho de [32].

Como nem todas as imagens da base possuem suas máscaras disponibilizadas, levamos em consideração no treinamento apenas aquelas dotadas de máscara. Isto é, duzentas e oitenta e sete (287) máscaras de pulmões que apresentam tuberculose em sua forma ativa e duzentas e setenta e nove (279) máscaras de pulmões que estão em sua forma normal.

B. GAN com flutuação estatística

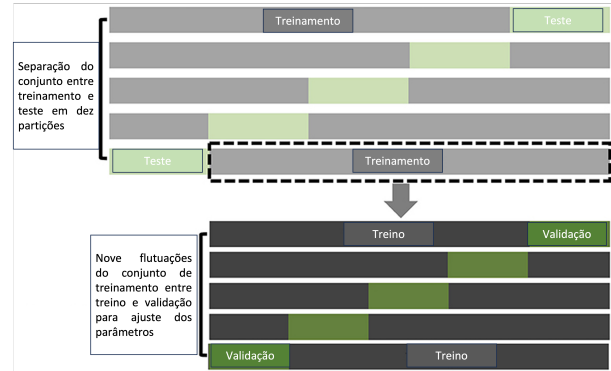
A produção sintética aqui trabalhada visa alimentar modelos de muitos parâmetros. Dessa forma, a escassez de dados, isto é, menos de mil imagens, para ajustar modelos com milhares ou até milhões de parâmetros, justifica e motiva esta pesquisa para geração sintética de imagens.

A baixa quantidade de imagens faz com que a flutuação estatística seja fundamental para o treinamento das variações de GAN deste artigo, exigindo que o processo de treinamento da GAN e da geração de dados sintéticos seja feito através de validação cruzada.

O método *k-fold*, com $k = 10$, é o método de validação cruzada realizado neste projeto. Dessa forma, são treinados cento e oitenta (180) modelos generativos, noventa (90) para cada classe especialista TB e NTB. Nesse sentido, separam-se os dados em dez caixas, uma por vez, para teste, nove caixas, uma por vez, como validação e as demais são utilizadas para treinamento. A figura 3 ilustra o processo descrito.

É importante destacar que a flutuação estatística é extremamente custosa do ponto de vista computacional, mas é fundamental para mitigar viés e aumentar a confiabilidade nas imagens geradas.

Figura 3: Flutuação estatística: Validação cruzada, método *k-fold*, com flutuação dos *folds* restantes para treinamento e validação.



C. Treinamento e avaliação do modelo treinado

A GAN Pix2Pix é ajustada ao problema de radiografias de tórax, como classe especialista para TB e NTB. Os geradores são salvos em diferentes épocas, de modo a coletar estatísticas da qualidade das imagens geradas por toda trajetória dos modelos treinados e avaliados.

As estatísticas que utilizamos para acompanhar o treinamento são a divergência de Kullback-Leibler (D_{KL}) [33], na equação 4, e a norma \mathcal{L}_{L1} , já apresentada na equação 2.

$$D_{KL} = (\mathcal{G}(x, z) || y) \log \left(\frac{y}{\mathcal{G}(x, z)} \right) \quad (4)$$

Para aferir a qualidade do treinamento, definiu-se como referência a flutuação estatística da divergência D_{KL} e da norma \mathcal{L}_{L1} entre as imagens reais. Ou seja, selecionam-se todas as imagens reais por partição¹², calculam-se as estatísticas e comparam-se as imagens uma a uma, registrando a dispersão dessas estatísticas.

A partir daí, para avaliar as imagens sintéticas, tomam-se todas as imagens reais contra todas as sintéticas e acompanha-se a evolução dessa flutuação estatística nas épocas de treinamento, fazendo a análise no conjunto de treinamento e no conjunto de validação, como em [6].

Nas figuras 4 e 5 é perceptível que o gerador especialista em TB, na última época, apresenta D_{KL} e norma \mathcal{L}_{L1} com dispersão muito semelhante ao observado no padrão de referência calculado, apontando para aproximação razoável à função geradora de eventos. Adicionalmente, os patamares são distintos, o que advoga para a diversidade criada.

Além disso, é observado um padrão de evolução das épocas em conformidade com a análise da última época, apontando para uma estabilidade do treinamento. A mesma análise é feita para os geradores classe-especialistas em NTB e os resultados são análogos com seus gráficos apresentados no apêndice.

¹²Apresentamos a evolução do treinamento para partição zero (*fold 0*).

Figura 4: Evolução do treinamento - Avaliação do treinamento para o modelo classe especialista em TB nos conjuntos de treinamento da *fold 0*.

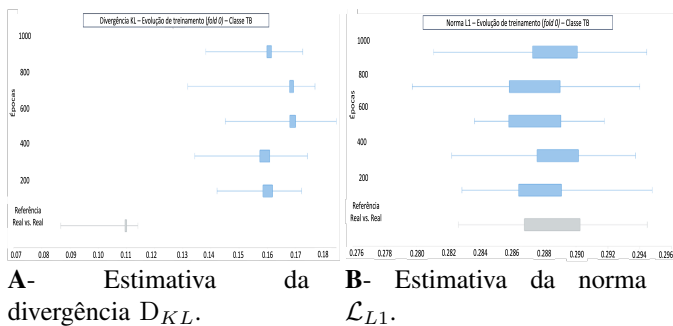
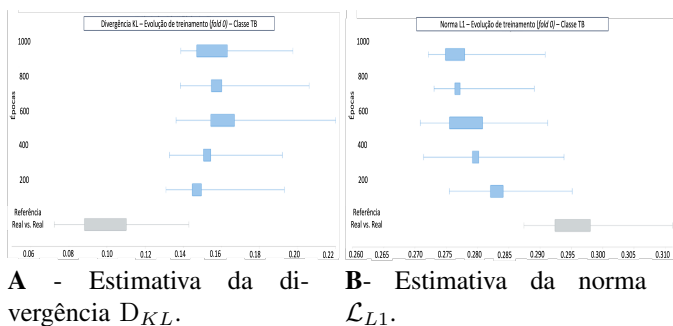


Figura 5: Evolução do treinamento - Avaliação do treinamento para o modelo classe especialista em TB nos conjuntos de validação da *fold 0*.



D. Produção sintética

Após o treinamento dos modelos generativos para as 180 flutuações, é possível gerar imagens para cada flutuação da partição, selecionando as máscaras de pulmão para serem entrada do gerador.

Como estamos trabalhando com a base pública de Shenzhen, temos as máscaras dos pulmões disponibilizadas em [32]. Assim como no treinamento, a produção é limitada ao múltiplo da quantidade de máscaras de pulmão.

Portanto, a partir dos pares imagens e máscaras somos capazes de gerar uma imagem sintética nova. Como possuímos duzentos e oitenta e sete (287) máscaras de pulmões da classe TB e duzentos e setenta e nove máscaras da classe NTB (279), geramos essas quantidades de imagens sintéticas para cada modelo treinado.

O número de modelos treinados é de noventa (90) para cada classe, logo são geradas vinte e cinco mil oitocentos e trinta (25830)¹³ imagens da classe TB e de vinte e cinco mil cento e dez (25110)¹⁴ imagens da classe NTB.

¹³São 287 imagens sintéticas por cada modelo treinado em uma dada flutuação da partição. Logo, para se chegar ao número de 25830, basta fazer 90 vezes 287.

¹⁴São 279 imagens sintéticas por cada modelo treinado em uma dada flutuação da partição. Logo, para se chegar ao número de 25110, basta fazer 90 vezes 279.

Esses milhares de imagens alimentam modelos profundos de alta quantidade de parâmetros que também venham a ser estimados, respeitando a flutuação estatística aqui especificada.

Em seguida, avalia-se a qualidade das novas imagens, de forma semelhante ao que é feito no acompanhamento do treinamento.

Para isso, utilizamos a flutuação da partição para estimar a divergência D_{KL} e a norma \mathcal{L}_{L1} entre as imagens experimentais e as sintéticas. Na avaliação da produção, apresentamos a flutuação em todas as *fold*s da partição.

E. Modelos classe especialistas em TB

Fica nítido que a produção sintética, tanto para a norma \mathcal{L}_{L1} quanto para a divergência D_{KL} , apresenta patamares semelhantes ao calculado na referência extraída dos dados reais para o modelo classe especialista em TB. Percebe-se também, a maior variabilidade da produção sintética se comparada à referência nas figuras 6 e 7.

Figura 6: D_{KL} para a produção de imagens do modelo especialista na classe TB. Comparação da distribuição de divergências Reais vs. Fakes, tomando a distribuição de Reais vs. Reais como padrão.

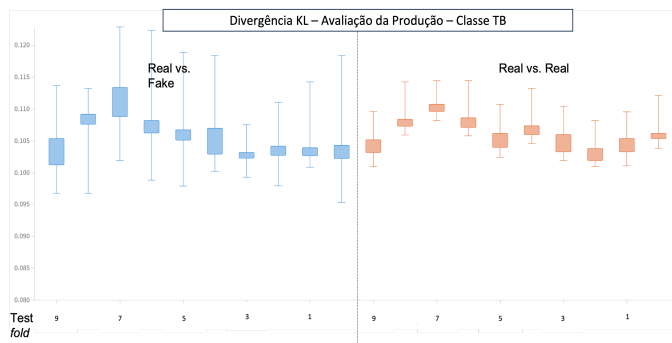
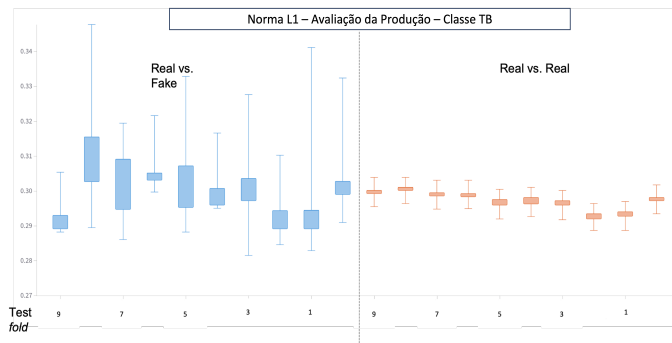
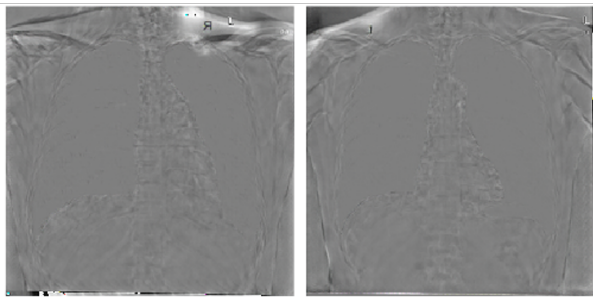


Figura 7: Norma \mathcal{L}_{L1} para a produção de imagens do modelo especialista na classe TB. Comparação da distribuição de divergências Reais vs. Fakes, tomando a distribuição de Reais vs. Reais como padrão.



A maior variabilidade é esperada, uma vez que a GAN Pix2Pix introduz informação nova ao entorno dos pulmões.

Figura 8: Entorno modificado para imagem gerada a partir do modelo especialista da classe TB. Imagens extraídas após subtração entre a imagem sintética (*fake*), $G(X)$, e a imagem experimental (real), y .



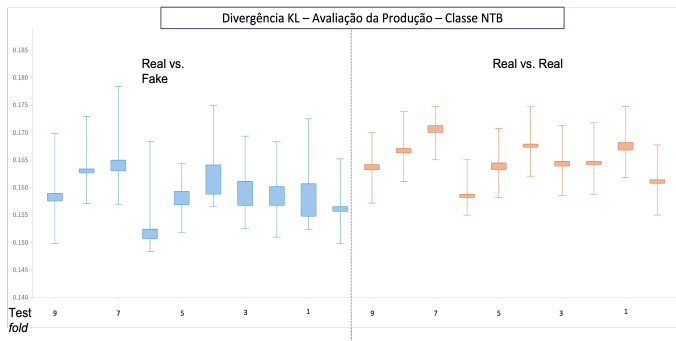
A região dos pulmões pode sofrer alterações, mas, com a aplicação de penalização pelo discriminador, esperamos que a interferência seja mínima.

Acompanhamos se o controle da região de interesse está sendo efetivo através da construção de figuras que façam a diferença entre a imagem sintética e o par real correspondente. Separamos na figura 8 um exemplo desse acompanhamento.

F. Modelos classe especialistas em NTB

Apresentamos nesta seção a avaliação da qualidade das imagens geradas a partir do modelo especialista para a classe NTB nas Figuras 9 e 10. É perceptível que as imagens sintéticas, assim como no modelo especialista para classe TB, apresentam distribuição com padrão semelhante aos dados experimentais, porém com a variabilidade esperada no entorno das regiões de interesse.

Figura 9: D_{KL} para a produção de imagens do modelo especialista na classe NTB. Comparação da distribuição de divergências Reais vs. Fakes, tomando a distribuição de Reais vs. Reais como padrão.



Finalmente, é apresentado na Figura 11 o entorno modificado em duas imagens geradas sinteticamente tomadas como exemplo para o modelo especialista na classe NTB.

IV. CONCLUSÃO E TRABALHOS FUTUROS

Em resumo, este trabalho propõe a aplicação da GAN Pix2Pix como instrumento de geração de imagens sintéticas

Figura 10: Norma \mathcal{L}_{L1} para a produção de imagens do modelo especialista na classe NTB. Comparação da distribuição de divergências Reais vs. Fakes, tomando a distribuição de Reais vs. Reais como padrão.

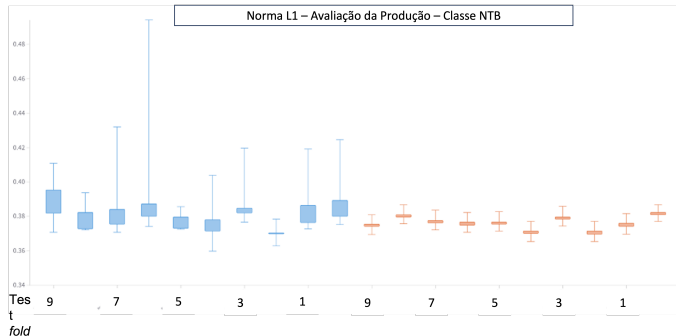
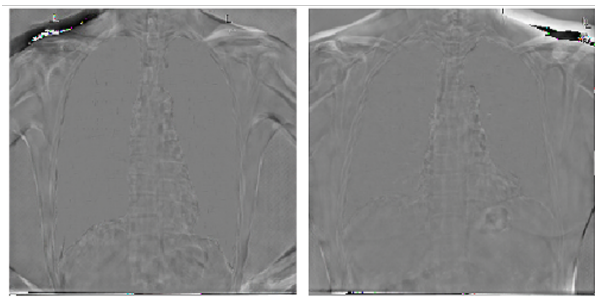


Figura 11: Entorno modificado para imagem gerada a partir do modelo especialista da classe NTB. Imagens extraídas após subtração entre a imagem sintética (*fake*), $G(X)$, e a imagem experimental (real), y .



com capacidade de preservar uma determinada região de interesse da imagem.

Objetivou-se a geração de imagens médicas, em especial radiografias de tórax, que sejam utilizadas no treinamento dos modelos que serão consumidos por sistemas CAD inseridos no contexto de *screening* de tuberculose.

A utilização das imagens sintéticas geradas com o método abordado neste artigo são exploradas a partir de classificadores com arquitetura CNN, como apresentado pelos mesmos autores em [29] e inspirado no trabalho [6]. Além da arquitetura da CNN, classificadores construídos a partir do paradigma de aprendizado contínuo são apresentados pelos autores em [16].

Dado o grau de exigência da OMS para sistemas dessa natureza, é fundamental que os dados expostos aos modelos classificadores tenham alta qualidade. Além da alta qualidade, é esperada a introdução de diversidade para essa produção sintética.

Sendo assim, controlamos o treinamento e avaliamos a produção das imagens, verificando se aquelas sinteticamente geradas se aproximam da distribuição dos dados reais, enquanto sua distribuição apresenta grau de variabilidade razoável. Esse acompanhamento é realizado através da divergência D_{KL} e da norma \mathcal{L}_{L1} .

Os treinamentos dos modelos GAN foram realizados através de validação cruzada em dez *folde*s. Ao final do processo, foram gerados noventa modelos por classe especialista. A partir dos modelos treinados, gerou-se um total de cinquenta mil novecentos e quarenta (50940) imagens, sendo vinte e cinco mil oitocentos e trinta (25830) da classe TB e vinte e cinco mil cento e dez (25110) da classe NTB.

Como passo futuro, visamos incorporar novas bases de dados à produção, sobretudo brasileiras e de pacientes no grupo de tratamento de TB infecção, a fim de contribuir para sistemas CAD que sejam desenvolvidos e implementados para *screening* de tuberculose no país. Além disso, as bases que não forem incorporadas à produção serão utilizadas para validação *out-of-sample* de todo o *pipeline* de análise, i.e., imagens sintéticas junto aos classificadores.

Além de novas bases de dados, novas estratégias de GAN de tradução de contextos estão sendo trabalhadas em análise comparativa à técnica abordada neste trabalho, de modo a posicionar as produções sintéticas do grupo em relação ao estado da arte.

AGRADECIMENTOS

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

Um agradecimento também é reservado aos institutos National Library of Medicine, National Institutes of Health, Bethesda, MD, USA and Shenzhen No.3 People’s Hospital, Guangdong Medical College, Shenzhen, China, pela disponibilização dos dados utilizados neste projeto.

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

APÊNDICE

Para os modelos classe especialistas em NTB, apresentamos o acompanhamento do treinamento. Observa-se que, na última época, as imagens geradas alcançam o patamar da D_{KL} e da norma \mathcal{L}_{L1} em conformidade com os níveis apontados pelo padrão dispersivo das imagens reais. Esse aspecto também advoga tanto para a qualidade quanto para a diversidade das imagens geradas nesse processo de treinamento.

Figura 12: Evolução do treinamento - Avaliação do treinamento para o modelo classe especialista em NTB nos conjuntos de treinamento da *fold 0*.

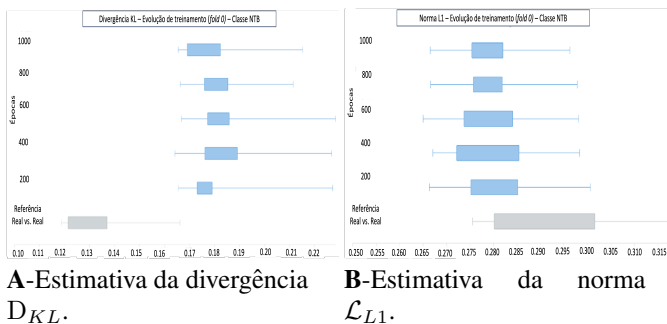
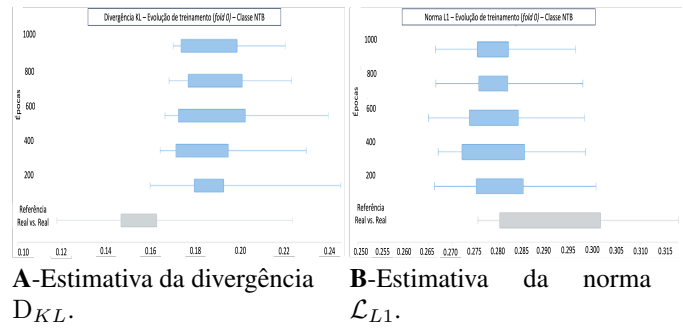


Figura 13: Evolução do treinamento - Avaliação do treinamento para o modelo classe especialista em NTB nos conjuntos de validação da *fold 0*.



REFERÊNCIAS

- [1] Goodfellow I., Bengio Y. e Courville A. - Deep Learning Book. MIT Press <http://www.deeplearningbook.org> (2016)
- [2] 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)
- [3] 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)
- [4] 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)
- [5] Ledig C., Theis L., Huszar F., Caballero J., Aitken A., Tejani A., Totz J., Wang Z., Shi W. Photo-realistic single image super-resolution using a generative adversarial network. In CVPR, (2017)
- [6] Fernandes J. C. V., Junior N. N., Seixas J. M. - Deep Learning Models for Passive Sonar Signal Classification of Military Data, Remote Sensing, Volume 14, Number 11, Article-Number 2648, <https://www.mdpi.com/2072-4292/14/11/2648>, ISSN 2072-4292, doi: 10.3390/rs14112648 (2022)
- [7] Zhou Y., Berg T. L. - Learning temporal transformations from time-lapse videos. In ECCV, (2016)
- [8] Singh N. K., Khalid R. - Medical Image Generation using Generative Adversarial Networks, arXiv:2005.10687 (2020)
- [9] Latent tuberculosis infection, Updated and consolidated guidelines for programmatic management, <https://www.who.int/publications/i/item/9789241550239> (2018)
- [10] WHO’s global tuberculosis report <https://apps.who.int/iris/bitstream/handle/10665/336069/9789240013131-eng.pdf> (2020)
- [11] Frija G., Blažić I., Frush D. P., Hierath M., Kawooya M., and Donoso-Bach L., Brkljačić B. - How to improve access to medical imaging in low- and middle-income countries?, EclinicalMedicine, Elsevier BV, volume 38, number 101034, pages 101034, copyright <http://creativecommons.org/licenses/by-nc-nd/4.0/> (2021)
- [12] Fawcett T. - An introduction to ROC analysis, Pattern Recognition Letters, volume 27, number 8, pages 961-874, 0167-8655, <https://doi.org/10.1016/j.patrec.2005.10.010> (2006)
- [13] National Library of Medicine, National Institutes of Health, Bethesda, MD, USA and Shenzhen No.3 People’s Hospital, Guangdong Medical College, Shenzhen, China
- [14] Chauhan A., Chauhan D., Rout C. - Role of Gist and PHOG Features in Computer-Aided Diagnosis of Tuberculosis without Segmentation. PLoS ONE 9(11): e112980. <https://doi.org/10.1371/journal.pone.0112980> (2014)
- [15] Y. Xing, Z. Ge, R. Zeng, D. Mahapatra, J. Seah, M. Law, T. Drummond - 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)

- [16] Nascimento O. T., Alves R., Seixas J. M., Jandre F., Trajman A. - An innovative artificial intelligence approach for fighting against tuberculosis, 2nd BRICS Postgraduate Forum, EDITOR Fulufhelo Netswera, Renitha Rampersad, PUBLISHER BRICS Research Institute, ADDRESS Cape Town, South Africa, (2023)
- [17] Mirza M. e Osindero S. - Conditional Generative Adversarial Nets - arxiv:1411.1784. (2014)
- [18] Lau, F., Hendriks, T., Lieman-sifry, J., Sall, S. (2017). ScarGAN: Chained Generative Adversarial Networks to Simulate Pathological Tissue on Cardiovascular MR Scans. Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support. <https://doi.org/10.1007/978-3-319-67558-9>
- [19] Mok, T. C. W., Chung, A. C. S. (2019). Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-11723-8_7
- [20] Tom, F., Sheet, D. (2018). Simulating patho-realistic ultrasound images using deep generative networks with adversarial learning. Proceedings - International Symposium on Biomedical Imaging. <https://doi.org/10.1109/ISBI.2018.8363780>
- [21] Zhao, H., Li, H., Maurer-Stroh, S., Cheng, L. (2018). Synthesizing retinal and neuronal images with generative adversarial nets. Medical Image Analysis. <https://doi.org/10.1016/j.media.2018.07.001>
- [22] Costa, P., Galdran, A., Meyer, M. I., Niemeijer, M., Abràmoff, M., Mendonça, A. M., Campilho, A. (2018). End-to-End Adversarial Retinal Image Synthesis. IEEE Transactions on Medical Imaging. <https://doi.org/10.1109/TMI.2017.2759102>
- [23] Iqbal, T., Ali, H. (2018). Generative Adversarial Network for Medical Images (MI-GAN). Journal of Medical Systems. <https://doi.org/10.1007/s10916-018-1072-9>
- [24] Salehinejad, H., Valaee, S., Dowdell, T., Colak, E., Barfett, J. (2018). Generalization of Deep Neural Networks for Chest Pathology Classification in X-Rays Using Generative Adversarial Networks. ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings, 2018-April, 990-994. <https://doi.org/10.1109/ICASSP.2018.8461430>
- [25] Madani, A., Moradi, M., Karargyris, A., Syeda-mahmood, T. (2018). SEMI-SUPERVISED LEARNING WITH GENERATIVE ADVERSARIAL NETWORKS FOR CHEST X-RAY CLASSIFICATION WITH ABILITY OF DATA DOMAIN ADAPTATION. Isbi, 1038-1042.
- [26] Mahapatra D., Bozorgtabar B., Thiran J. P., Reyes M. - Efficient active learning for image classification and segmentation using a sample selection and conditional generative adversarial network. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-00934-2_65 (2018)
- [27] Jin D., Xu Z., Tang Y., Harrison A. P., Mollura D. J. - CT-realistic lung nodule simulation from 3D conditional generative adversarial networks for robust lung segmentation. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030009342_81 (2018)
- [28] Shin, H. C., Tenenholtz, N. A., Rogers, J. K., Schwarz, C. G., Senjem, M. L., Gunter, J. L., Andriole, K. P., Michalski, M. (2018). Medical image synthesis for data augmentation and anonymization using generative adversarial networks. Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics). https://doi.org/10.1007/978-3-030-00536-8_1
- [29] Nascimento T. O., Seixas J. M., Trajman A. - Data-augmentation de dados de radiografia de torax no contexto de aprendizagem profunda, PAGES 1-7, Anais do 15 Congresso Brasileiro de Inteligência Computacional, EDITOR Carmelo José Albanes Bastos Filho, Hugo Valadares Siqueira, Danton Diego Ferreira, Douglas Wildgrube Bertol, Roberto Célio Limão de Oliveira, PUBLISHER SBIC, ADDRESS Joinville SC, (2021)
- [30] Ronneberger O., Fischer P., Brox T. - U-Net: Convolutional Networks for Biomedical Image Segmentation, eprint 1505.04597 (2015)
- [31] Demir U., Gozde U. - Patch-Based Image Inpainting with Generative Adversarial Networks, arXiv:1803.07422 (2018)
- [32] Pandey N. - Pulmonary Chest X-Ray Defect Detection - Kaggle (2019)
- [33] Kullback S., Leibler R. A. - On Information and Sufficiency - journal Ann. Math. Statist, pages 79-86, volume 22, https://www.bibsonomy.org/bibtex/2560a5719c537c5c4a496bfebd4a21603/lee_peck (1951)