

Generative adversarial networks: a renewal for data augmentation in lung nodule classification

Bruno H. L. dos Anjos*, Anthony E. A. Jatobá†, Marcelo C. Oliveira†

Programa de Pós-Graduação em Informática (PPGI)

Instituto de Computação, Universidade Federal de Alagoas

Maceió, AL, Brasil

Email: {aeaj,oliveiramc}@ic.ufal.br†, bdosanjos@gmail.com*

Resumo—Obtaining medical images is an ethically restrictive process and still difficult to validate, depending on well-trained professional, being a laborious and time-consuming activity. Therefore, the construction of large databases of structured medical images is one of the major challenges of the deep learning applications in the computerized aid to the diagnosis in medical images. GAN presents itself as an adequate solution to supply the small number of pathological exams that compose the most diverse medical images banks. In this work we intend to develop a method using GAN to balance the data set of Computed Tomography images and improve the performance of an arbitrary classifier of pulmonary nodules. For this, two GAN architectures with the capacity to generate synthetic images of nodal cuts were trained and in a second moment, a Convolutional Neural Network was trained in rooms of different data set. The training of the different data sets were evaluated by AUC-ROC.

I. INTRODUÇÃO

Apesar da Tomografia Computadorizada (do inglês *Computed Tomography* (CT)) ser o principal meio de geração de imagem para detecção do câncer pulmonar, os resultados provenientes dela dependem da interpretação semi qualitativas de um especialista em diagnóstico radiológico. Esta interpretação dos exames é variada, sujeita a interferência intra e interpessoais [1], um problema que se agrava quando levado em consideração o grande volume de dados médicos. Para mitigar esse problema, sistemas de diagnóstico auxiliado por computador são desenvolvidos para melhorar as interpretações [2]. A utilização de ferramentas de diagnóstico auxiliado por computador tem se mostrado eficiente na acurácia dos diagnósticos e detecção de doenças [3].

Um das principais áreas por trás das ferramentas de diagnóstico auxiliado por computador é a inteligência artificial. Esta vive uma nova era diante dos avanços tecnológicos de hardware que possibilitam o processamento de grandes volumes de dados. Neste cenário surgem algoritmos mais sofisticados e mais caros computacionalmente, principalmente aqueles que visam imitar as propriedades de tomadas de decisão dos seres humanos, como o conjunto de algoritmos de Aprendizagem Profunda (do inglês Deep Learning (DP)) [4].

DP consiste em um conjunto de métodos de redes neurais que utiliza uma representação de aprendizagem em múltiplos níveis de representatividade, onde conforme os níveis se aprofundam, a representatividade se torna mais abstrata [5]. A

capacidade da DP de trabalhar com imagens para reconhecimento, análise e processamento obtém resultados próximos aos encontrados por humanos [6]. Uma das áreas mais beneficiadas dos avanços da inteligência artificial é a medicina. Onde, somente em 2017, na área da radiologia médica foram publicados mais de 500 artigos científicos correlacionando os termos *radiotherapy* ou *radiation therapy* ou *radiation oncology* com os termos *deep network* ou *convolutional network* ou *deep learning* [7].

Apesar da crescente quantidade de estudos associando a DP e radiologia médica, existe um problema inerente da área, que é a disponibilidade de grandes volumes de dados em saúde. Isso ocorre devido a obtenção de imagens médicas ser um processo de grande restrição ética, dada pela priorização da privacidade e da confidencialidade do paciente. Ainda, gerar imagens médicas é um processo custoso e a depender do exame, submete o paciente a altas doses de radiação e a procedimentos invasivos. Além disso, as imagens precisam ser diagnosticadas por profissionais bem treinados, atividade laboriosa e que demanda muito tempo do especialista [8][9][10]. Portanto, a construção de grandes bases de imagens médicas estruturadas, bem validadas e abrangentes, é um dos maiores desafios das aplicações de DP no auxílio computadorizado ao diagnóstico em imagens médicas, visto que, pequenas bases de dados podem gerar modelos pouco heterogênicos, levando a resultados enviesados [11].

Proposta por I. Goodfellow [12], a técnica de Redes Neurais Generativas Adversárias (do inglês *Generative Adversarial Networks* (GAN)) é baseada na teoria dos jogos e vem revolucionando a área de visão computacional com a sua capacidade de gerar imagens sintéticas de alta qualidade. GANs consistem em um modelo para geração de dados probabilisticamente heterogêneos, onde o modelo combina uma rede geradora responsável por gerar dados com características reais a partir de um vetor de ruído e de uma rede discriminatória que tenta distinguir os dados reais dos dados sintéticos gerados, tendo ambas um aprimoramento simultâneo. Essa relação contraditória abriu grandes possibilidades para as mais diversas formas de utilização, como na pura geração de imagens sintéticas [13] ou na criação de modelos de transformações, como transformar texto em imagem. [14].

Imagens médicas sintéticas proveniente de uma GAN se apresentam como uma solução adequada para suprir o redu-

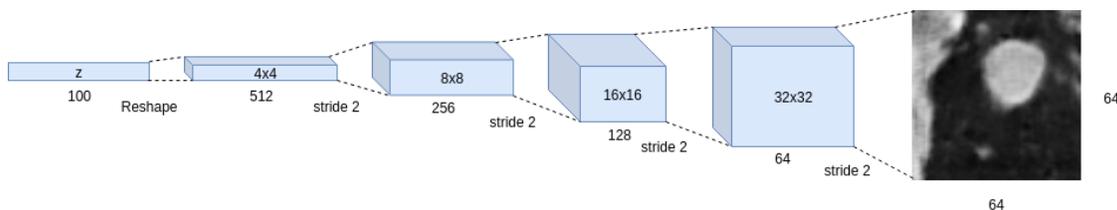


Figura 1. Gerador das redes GANs. O *input* do gerador é o vetor ruído representado por *z*. Todas as camadas, após o *reshape* do vetor ruído, são de convolução transposta.

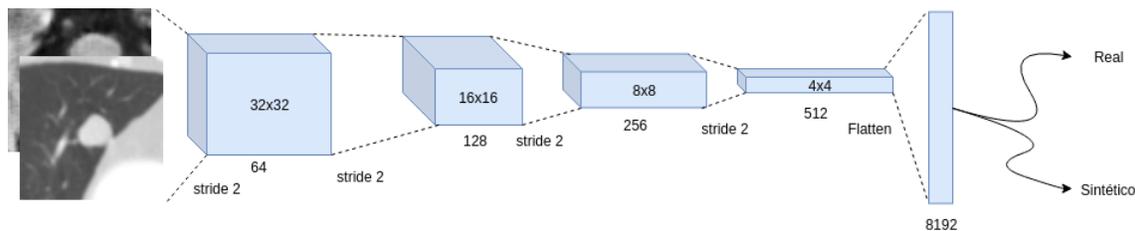


Figura 2. Discriminante das redes GANs

zido número de exames patológicos que compõem os mais diversos bancos de imagens médicas [15][16]. Diferentemente das técnicas clássicas de aumento de base de imagens em DP, que adotam técnicas de transformação, como rotação, inversão, tradução e escala, os dados sintéticos gerados pela GAN aumentam a variabilidade dos dados, logo melhorando o treinamento dos sistemas [17][10][18][19].

O objetivo principal deste trabalho é avaliar a qualidade de imagens sintéticas produzidas pela GAN no processo de classificação de nódulos pulmonares usando uma rede neural convolucional (do inglês *Convolutional Neural Networks* (CNN)) arbitrária.

A. Trabalhos Relacionados

O trabalho de Frid-Adar M. et al. [17] gerou imagens sintéticas de CT de lesões hepáticas de alta qualidade com intuito de melhorar o desempenho de um classificador baseado em CNN sobre a tarefa de classificação de lesões hepáticas. Os autores utilizaram uma *Deep convolutional GAN* (DCGAN) para a geração das imagens sintéticas. Com a utilização das novas imagens sintéticas houve um ganho significativo na acurácia da classificação quando comparado com a utilização de aumento de base clássico, sendo de 85,7% com aumento de base por imagens sintéticas GAN e 78,6% com aumento de base clássico [17]. Apesar do objeto de interesse de Frid-Adar M. et al. ser diferente do apresentado nesse trabalho, a utilização do mesmo tipo de exame encoraja a geração de imagens sintéticas para diferentes segmentos provenientes de CT.

Já Onishi et al. [9] e Zhao et al. [10], a proximidade é mais tangível. Onishi et al. [9] utiliza uma *Wasserstein GAN* (WGAN) para gerar cortes de imagens sintéticas de nódulos de CT usadas para pré-treinamento de uma CNN profunda que classifica nódulos pulmonares e tendo o ajuste fino feito com as imagens originais dos nódulos. Essa abordagem gerou

um ganho de quase 30% na área sobre a curva ROC em relação ao caso sem pré-treinamento [9]. Zhao et al. [10] utiliza sua própria abordagem: a *Forward and Backward GAN* (F&BGAN). O modelo proposto pelos autores gera cortes de imagens de CT de nódulos malignos e benignos demonstrando a capacidade das imagens sintéticas na melhora da performance de uma classificador M-VGG16. Para isso, foram gerados sete conjuntos de dados misturando técnicas de aumento de base distintas incluindo a F&BGAN. Cada base e algumas combinações entre si foram utilizadas no treinamento para um classificador de nódulos M-VGG16. Foi obtido um total de 0,954 de área sobre a curva ROC com a utilização de aumento de base clássico, com o aumento feito apenas pela *Forward GAN* o valor foi superior, chegando a 0,966 [10].

II. MÉTODO

A construção do ambiente para a realização dos experimentos foi feita com a utilização da biblioteca *Keras*, sendo o processamento executado em uma GPU NVIDIA GeForce GTX 1060. A base utilizada para o treinamento da CNN e das GANs é um subconjunto da base pública de exames de CT LIDC-IDRI [20]. O subconjunto utilizado do LIDC-IDRI contém 98 nódulos benignos provendo um total de 679 cortes de nódulos rotulados como benignos e 212 nódulos malignos provendo um total de 2013 cortes de nódulos rotulados como malignos, onde cada corte possui o tamanho 64x64 *pixels*, com o nódulo centralizado na imagem com tamanho variando entre 10mm a 30mm e as imagens estão em um janelamento da escala de cinza do pulmão usado em CT.

O trabalho se divide em duas etapas, a primeira consiste em avaliar empiricamente o comportamento das imagens sintéticas geradas por duas arquiteturas GANs sobre cortes de nódulos pulmonares benignos. E a segunda etapa consiste em treinar uma CNN arbitrária para classificar nódulos.

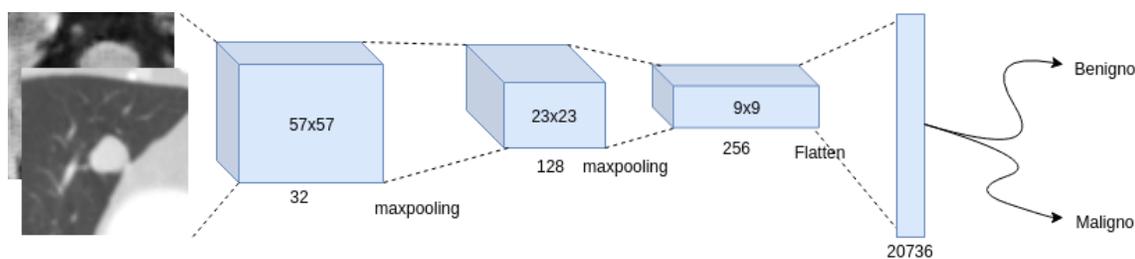


Figura 3. CNN para classificação

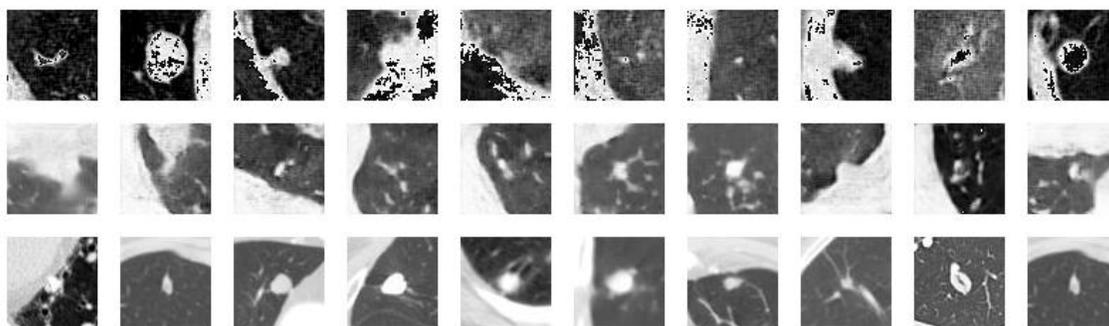


Figura 4. Na primeira linha nódulos gerados pela WGAN, na segunda linha nódulos gerados pela DCGAN e a terceira nódulos originais

A. Arquiteturas

Para a primeira etapa foram escolhidas duas arquitetura: DCGAN e WGAN. A DCGAN é uma arquitetura composta por uma GAN onde suas camadas são do tipo convolucionais tanto na parte da rede discriminante quanto na parte da rede geradora. Pela natureza das camadas de convolução, essa abordagem é utilizada com ênfase em imagens [21]. Mesmo com a grande capacidade para trabalhar com imagens fornecidas pela arquitetura DCGAN, ela ainda sofre com o problema de diversidade nas imagens geradas, assim avaliamos também a WGAN. Sendo esta uma modificação na DCGAN, onde a métrica para função de perda é definida pela distância de *Wasserstein*, substituindo a mais utilizada divergência de Jensen–Shannon [22]. Os parâmetros utilizados na construção das GANs são inspirados nos trabalhos da seção I-A, havendo modificações para adaptar ao conjunto de imagens usado neste trabalho.

B. Geradores e discriminadores

Nos geradores das GANs todas as camadas, após o *reshape* do vetor ruído, são de convolução transposta contendo filtros de *kernel* de tamanho 5 e ativação *LeakyRelu*. Em todas as camadas foi utilizado o *batch normalization*, com exceção da última camada. A Figura 1 apresenta a organização das camadas da rede geradora para ambas as redes GAN. Para os discriminantes, todas as camadas são de convolução contendo filtros de *kernel* de tamanho 5 e ativação *LeakyRelu*. Com exceção da camada de *flatten*, em todas as outras camadas é utilizado o *batch normalization*. A Figura 2 apresenta a organização das camadas da rede geradora para ambas as redes GAN.

A WGAN utiliza em seu discriminante e seu gerador o *loss* baseado na distância de *wasserstein* e o otimizador utilizado é RMSprop com um *learning rate* de 0,00005. Para a DCGAN a perda é sobre a *binary cross entropy* e o otimizador utilizado é RMSprop com um *learning rate* de 0,0001. Ambas, DCGAN e WGAN foram treinadas durante 6000 épocas.

C. Classificador

Na segunda etapa treinamos uma CNN arbitrária para a classificação da malignidade dos nódulos em quatro conjuntos de dados. O primeiro conjunto (G1) é desbalanceado contendo majoritariamente cortes malignos, sendo 545 cortes benignos e 1879 cortes malignos utilizados no treinamento da CNN. O segundo conjunto (G2), utiliza o mesmo número de cortes, sendo 545 cortes benignos e 545 cortes malignos. O terceiro (G3), realizamos um aumento da base de nódulos benignos aplicando técnicas clássicas, totalizando 1879 cortes malignos e 1845 cortes benignos e o quarto (G4), com aumento dos cortes benignos utilizando a GAN, sendo 1879 cortes malignos e 545 cortes benignos e mais 1300 cortes benignos sintéticos.

A arquitetura da CNN utilizada pode ser vista na Figura 3, após cada camada de convolução é aplicado um *maxpooling* de tamanho dois e foi utilizado a *binary cross entropy* como função de perda. O otimizador utilizado foi o *Adam* com um *learning rate* de 0,0001 e a CNN foi treinada durante 50 épocas. Para o treinamento da CNN, foram utilizados 134 cortes para validação e 134 cortes para testes, onde 67 cortes de cada tipo de nódulo para o treinamento em todos os conjuntos. Em todos os conjuntos de dados de treinamento foi utilizado o modelo de validação *holdout*. As imagens usada para o treinamento das GANs são imagens que não

pertencem ao grupo de teste da CNN. No grupo de teste da CNN não contém imagem sintéticas GANs e também não contém imagens do aumento de base clássico.

III. RESULTADOS E DISCUSSÃO

Apesar do tempo de treinamento da WGAN e da DCGAN usadas terem sido o mesmo, a convergências para uma boa imagem em cada arquitetura é diferente, sendo influenciada também pela quantidade de imagens usadas para treino, logo é esperado um comportamento diferente para ambas arquiteturas. A Figura 4 ilustra essa diferença, onde na primeira linha são imagens geradas pela WGAN que visivelmente aprende a forma dos cortes, mas não converge em sua plenitude. Na segunda linha estão as imagens provenientes da DCGAN que apresentam uma grande semelhança com as imagens da linha três, onde se encontram imagens originais. Apesar de semelhantes, é visível borrões nas imagens da DCGAN. Apesar disso elas foram mais próximas das originais e foram escolhidas para serem usadas no balanceamento.

No que tange a classificação, a Tabela I mostra a matriz de confusão dos resultados obtidos utilizando o grupo de imagens G1. Com a maior quantidade de cortes malignos o comportamento do classificador fica dentro do esperado, apresentando um enviesamento para classificação dos cortes de nódulo como maligno e obtendo um total de 0,686 na área sobre a curva ROC, demonstrada na Figura 5.

Tabela I
MATRIZ DE CONFUSÃO COM O TREINAMENTO G1

	Benigno	Maligno
Benigno	21	8
Maligno	46	58

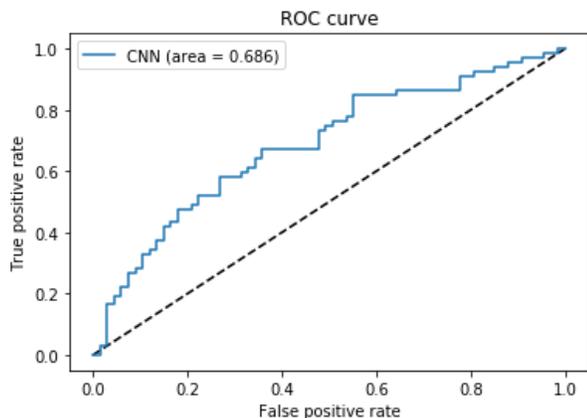


Figura 5. Curva ROC de G1

A matriz de confusão do treinamento com o conjunto de imagens de G2 é apresentada na Tabela II. Esta acaba por ter um comportamento similar aos resultados encontrados com o G1, mas por motivos distintos. O balanceamento mesmo que para baixo torna o grupo mais bem dividido tendendo a melhorar

a aprendizagem, isso se reflete no melhor equilíbrio dos verdadeiros positivos e negativos. Entretanto, da margem para o problema de *overfitting* que acaba por tornar o classificador pouco genérico e mais propenso a erros, a Figura 6 apresenta o comportamento do *loss* e da acurácia tanto no treino quanto na validação, deixando evidente o surgimento do *overfitting*. Consequentemente, obtemos o menor valor de curva ROC entre os quatros conjuntos de treinamento com um total de 0,631 (FIG. 7).

Tabela II
MATRIZ DE CONFUSÃO COM O TREINAMENTO G2

	Benigno	Maligno
Benigno	30	23
Maligno	37	44

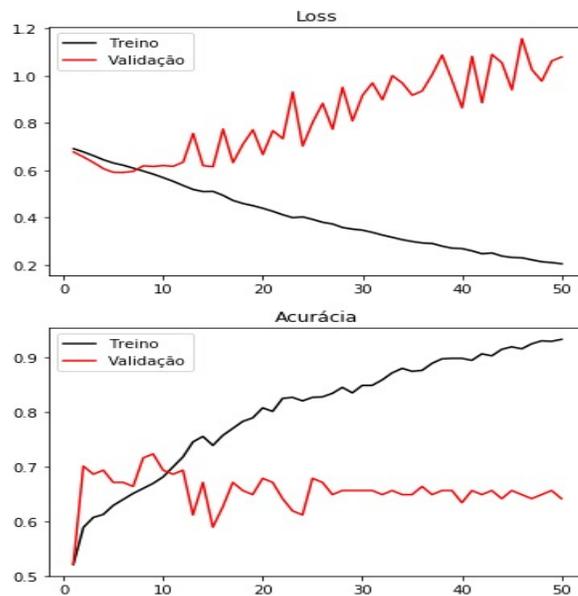


Figura 6. Gráficos da *loss* e da acuracia ao longo do treinamento de G2

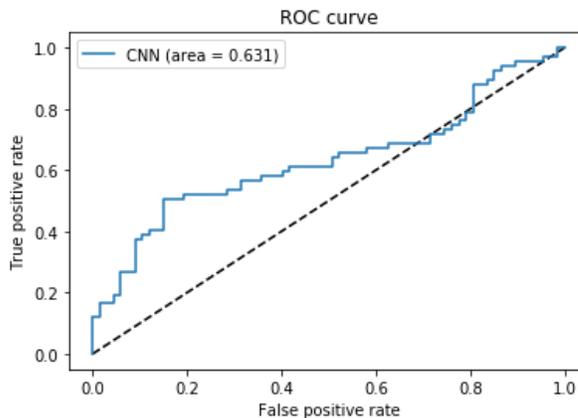


Figura 7. Curva ROC de G2

O conjunto G3 não apresenta melhora em relação ao grupo desbalanceado. Sua matriz de confusão é apresentada na Tabela III, onde é possível perceber uma melhor distribuição entre os verdadeiros-positivos e falsos-positivos. Entretanto, os erros ainda são significativos e as melhoras apresentadas são apenas em relação ao G1. Esse fato se reflete seu no resultado de 0,647 de AUC-ROC que pode ser vista na Figura 8.

Tabela III
MATRIZ DE CONFUSÃO COM O TREINAMENTO G3

	Benigno	Maligno
Benigno	30	17
Maligno	37	51

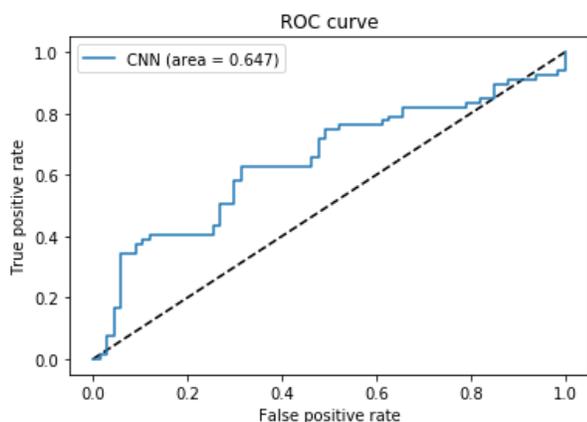


Figura 8. Curva ROC de G3

No grupo de G4 é evidenciado uma melhora significativa tanto na matriz de confusão quanto curva ROC. A matriz de confusão apresentada pela Tabela IV demonstra o maior número de acertos tanto no caso dos cortes benignos quanto nos cortes malignos, indicando uma melhora proporcionada pela utilização das imagens sintéticas para o balanceamento. Além de um resultado melhor evidenciado na matriz de confusão, esse padrão se repete quando observado a curva ROC, ilustrada na Figura 9, onde a área obtida é de 0,803.

Tabela IV
MATRIZ DE CONFUSÃO COM O TREINAMENTO G4

	Benigno	Maligno
Benigno	34	6
Maligno	33	61

Esses valores são próximos com o trabalho de Onishi Y. et al. [9], apesar da distinção na forma de treinamento do classificador, o que colabora para confirmar a eficiência da GAN para geração de imagens de corte de nódulos pulmonares no aprimoramento de classificadores.

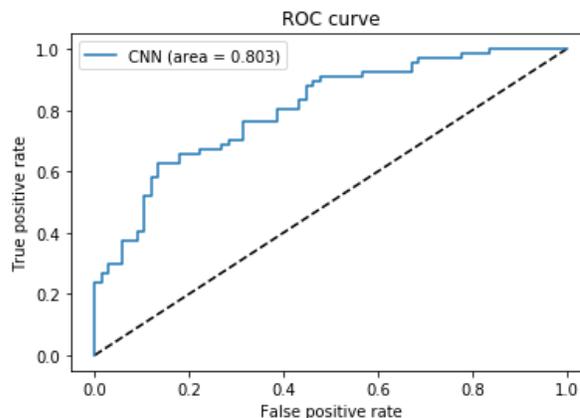


Figura 9. Curva ROC de G4

IV. CONCLUSÃO

Neste trabalho avaliamos a qualidade de imagens sintéticas de nódulos pulmonares produzidas por GANs com o propósito de aumento de base de dados. A utilização das imagens sintéticas gerou um aumento de 17,06% na área sob a curva ROC em relação as imagens desbalanceadas, 27,24% em relação as imagens balanceadas para baixo e 24,11% em relação ao conjunto com aumento de base clássica usando transformações de rotação, flip e translação.

Para trabalhos futuro iremos avaliar os resultados do algoritmo com radiologistas para a validação das imagens através de teste de Turing Visual, assim como a construção de uma nova arquitetura GAN capaz de melhorar a geração das imagens de nódulos de CT.

REFERÊNCIAS

- [1] P. Azevedo-Marques, "Diagnóstico auxiliado por computador na radiologia," *Radiologia Brasileira*, vol. 34, 10 2001.
- [2] J. R. Ferreira, M. C. Oliveira, and A. L. Freitas, "Performance evaluation of medical image similarity analysis in a heterogeneous architecture," in *2014 IEEE 27th International Symposium on Computer-Based Medical Systems*, May 2014, pp. 159–164.
- [3] K. Doi, "Computer-aided diagnosis in medical imaging: Historical review, current status and future potential," *Computerized Medical Imaging And Graphics*, vol. 31, no. 4-5, pp. 198–211, 2007.
- [4] M. P. McBee, O. A. Awan, A. T. Colucci, C. W. Ghobadi, N. Kadom, A. P. Kansagra, S. Tridandapani, and W. F. Auffermann, "Deep learning in radiology," *Academic Radiology*, vol. 25, no. 11, pp. 1472–1480, Nov 2018. [Online]. Available: <https://doi.org/10.1016/j.acra.2018.02.018>
- [5] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, pp. 436 EP –, May 2015. [Online]. Available: <https://doi.org/10.1038/nature14539>
- [6] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghahfoorian, J. A. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60 – 88, 2017. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S1361841517301135>
- [7] P. Meyer, V. Noblet, C. Mazzara, and A. Lallemand, "Survey on deep learning for radiotherapy," *Computers in Biology and Medicine*, vol. 98, 05 2018.
- [8] G. Chartrand, P. M. Cheng, E. Vorontsov, M. Drozdal, S. Turcotte, C. J. Pal, S. Kadoury, and A. Tang, "Deep learning: A primer for radiologists," *RadioGraphics*, vol. 37, no. 7, pp. 2113–2131, 2017, pMID: 29131760. [Online]. Available: <https://doi.org/10.1148/rg.2017170077>

- [9] Y. Onishi, A. Teramoto, M. Tsujimoto, T. Tsukamoto, K. Saito, H. Toyama, K. Imaizumi, and H. Fujita, "Automated pulmonary nodule classification in computed tomography images using a deep convolutional neural network trained by generative adversarial networks," *BioMed Research International*, vol. 2019, pp. 1–9, 01 2019.
- [10] D. Zhao, D. Zhu, J. Lu, Y. Luo, and G. Zhang, "Synthetic medical images using fbgan for improved lung nodules classification by multi-scale vgg16," *Symmetry*, vol. 10, p. 519, 2018.
- [11] Y. Liu, Y. Zhou, X. Liu, F. Dong, C. Wang, and Z. Wang, "Wasserstein gan-based small-sample augmentation for new-generation artificial intelligence: A case study of cancer-staging data in biology," *Engineering*, vol. 5, no. 1, pp. 156 – 163, 2019. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S2095809918301127>
- [12] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 2672–2680. [Online]. Available: <http://papers.nips.cc/paper/5423-generative-adversarial-nets.pdf>
- [13] M. J. M. Chuquicusma, S. Hussein, J. R. Burt, and U. Bagci, "How to fool radiologists with generative adversarial networks? a visual turing test for lung cancer diagnosis," *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pp. 240–244, 2018.
- [14] S. E. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *ICML*, 2016.
- [15] S. G. M. Loew, "Breast cancer detection using synthetic mammograms from generative adversarial networks in convolutional neural networks," *SPIE Proceedings*, vol. 10718, 2018.
- [16] F. Galbusera, F. Niemeyer, M. Seyfried, T. Bassani, G. Casaroli, A. Kienle, and H. Wilke, "Exploring the potential of generative adversarial networks for synthesizing radiological images of the spine to be used in in silico trials," *Frontiers in Bioengineering and Biotechnology*, vol. 6, no. MAY, 5 2018.
- [17] M. Frid-Adar, I. Diamant, E. Klang, M. Amitai, J. Goldberger, and H. Greenspan, "Gan-based synthetic medical image augmentation for increased cnn performance in liver lesion classification," *Neurocomputing*, vol. 321, pp. 321 – 331, 2018. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0925231218310749>
- [18] "A novel method for enhancing the classification of pulmonary data sets using generative adversarial networks," *Biomedical Research*, vol. 29, no. 14, pp. 3022–3027, 2018.
- [19] A. K. T. S.-M. Ali Madani, Mehdi Moradi, "Chest x-ray generation and data augmentation for cardiovascular abnormality classification," pp. 10574 – 10574 – 6, 2018. [Online]. Available: <https://doi.org/10.1117/12.2293971>
- [20] S. G. Armato, G. McLennan, L. Bidaut, M. F. McNitt-Gray, C. R. Meyer, A. P. Reeves, B. Zhao, D. R. Aberle, C. I. Henschke, E. A. Hoffman, E. A. Kazerooni, H. MacMahon, E. J. R. Beek, D. Yankelevitz, A. M. Biancardi, P. H. Bland, M. S. Brown, R. M. Engelmann, G. E. Laderach, D. Max, R. C. Pais, D. P.-Y. Qing, R. Y. Roberts, A. R. Smith, A. Starkey, P. Batra, P. Caligiuri, A. Farooqi, G. W. Gladish, C. M. Jude, R. F. Munden, I. Petkovska, L. E. Quint, L. H. Schwartz, B. Sundaram, L. E. Dodd, C. Fenimore, D. Gur, N. Petrick, J. Freymann, J. Kirby, B. Hughes, A. Vande Casteele, S. Gupte, M. Sallam, M. D. Heath, M. H. Kuhn, E. Dharaiya, R. Burns, D. S. Fryd, M. Salganicoff, V. Anand, U. Shreter, S. Vastagh, B. Y. Croft, and L. P. Clarke, "The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans," *Medical Physics*, vol. 38, no. 2, pp. 915–931. [Online]. Available: <https://aapm.onlinelibrary.wiley.com/doi/abs/10.1118/1.3528204>
- [21] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *CoRR*, vol. abs/1511.06434, 2015. [Online]. Available: <http://arxiv.org/abs/1511.06434>
- [22] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *arXiv e-prints*, p. arXiv:1701.07875, Jan 2017.