

Um Estudo sobre Doenças Torácicas utilizando Redes Neurais Profundas para o Diagnóstico de Cardiomegalia.

Saulo Enock Rodrigues Fernandes, Ricardo Costa da Silva Marques,
João Dallyson Sousa de Almeida, Anselmo Cardoso de Paiva, Geraldo Braz Junior

¹Núcleo de Computação Aplicada – Vision and Image Processing Lab
Universidade Federal do Maranhão (UFMA) Campus do Bacanga – São Luís – MA

{saulorodrigues, rikardomarques, jdallyson, paiva, geraldo}@enca.ufma.br

Abstract. *Chest X-Ray is one of the most common and vital exams for the observation and diagnosis of thoracic diseases. However, the complexity and variety of chest diseases demand skilled and experienced physicians to avoid errors in the diagnosis of such pathology's. Numerous approaches using Neural Networks have been developed to assist specialists in the task of detecting chest diseases such as pneumonia and pleural effusion. This paper presents an approach still under development aiming to assist health professionals in the excellence of diagnosis of diseases of the chest region. Thus, we propose a methodology based on image quantity balancing and hyperparameter optimization. The results of the initial tests focused on the cardiomegaly class proved to be promising with the best obtained value of 0,879 (AUC), 0,803 (Accuracy), 0,728 (Precision), 0,815 (F1-Score), 0,925 (Sensitivity), 0,600 (Specificity), approaching the best metrics present in the literature.*

Keywords: *Deep Neural Networks; Diagnosis of Cardiomegaly; Hyperparameters.*

Resumo. *A radiografia da região do tórax é um dos exames mais recorrentes e importantes para observação e diagnóstico de doenças torácicas. A complexidade e variedade de patologias do tórax demanda médicos com experiência e habilidade na área para evitar erros no diagnóstico de tais doenças. Inúmeras abordagens utilizando Redes Neurais tem sido elaboradas para o auxílio dos especialistas na tarefa de detecção de patologias torácicas como a pneumonia e o derrame pleural. Este trabalho apresenta uma abordagem em desenvolvimento que objetiva auxiliar os profissionais da saúde na excelência do diagnóstico de doenças da região do tórax. É proposto uma metodologia baseada no balanceamento de quantidades de imagens e em ajustes empíricos dos hiperparâmetros do modelo. Os resultados dos testes iniciais focados na classe de cardiomegalia se provaram promissores com o melhor valor obtido de 0,879 (AUC), 0,803 (Acurácia), 0,728 (Precisão), 0,815 (F1-Score), 0,925 (Sensibilidade) e 0,600 (Especificidade), aproximando-se das melhores métricas presentes na literatura.*

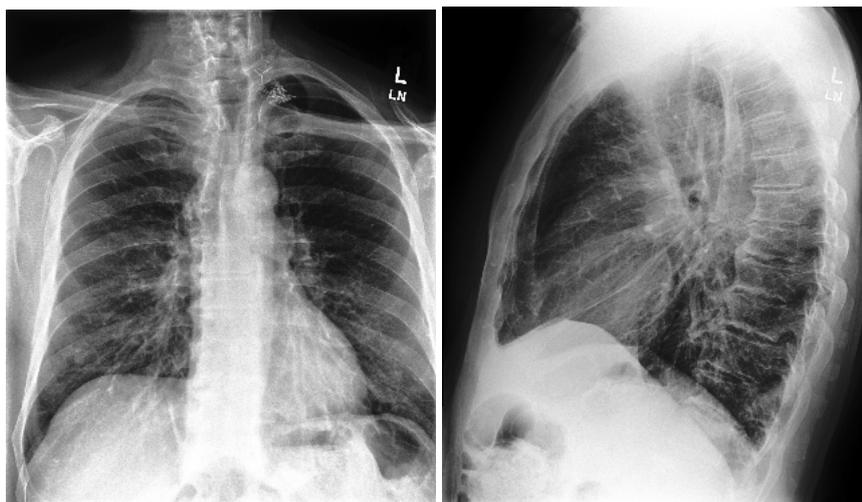
Palavras-chave: *Redes Neurais Profundas; Diagnóstico de Cardiomegalia; Hiperparâmetros.*

1. Introdução

A radiografia do tórax (Figura 1), também denominada de exame de Raio X (*Chest X-ray* ou *CXR*), é um dos tipos mais comuns de exames radiológicos aplicados ao di-

agnóstico de doenças torácicas, especialmente para patologias relacionadas aos pulmões [Anderson et al. 2015] e ao coração [Amin and Siddiqui 2021]. Entre essas patologias, há aquelas particularmente mortais como a cardiomegalia [Agostoni et al. 2000] e o pneumotórax [Bellaviti et al. 2016], ou as que têm consequências debilitantes, como por exemplo a atelectasia [Randtke et al. 2015], que consiste no colapso de um dos pulmões.

Figura 1. *Imagens de exames de Raios X Frontal (esquerda) e Lateral (direita) do CheXpert Dataset.*



Fonte: *CheXpert Dataset*, [Irvin et al. 2019].

Na literatura são encontrados trabalhos baseados em Redes Neurais Convolucionais (ConvNet) aplicados ao conjunto de imagens CheXpert. Este trabalho visa apresentar a abordagem desenvolvida e compará-la com os resultados alcançados por outros pesquisadores. Por exemplo, Irvin et al. (2019) introduziram o *CheXpert Dataset* discutindo a construção da base de dados desde o início com o recolhimento de exames, a extração de rótulos e a configuração inicial de treino aplicado.

Allaouzi e Ahmed (2019) propuseram uma abordagem combinando a eficácia das Redes Neurais Convolucionais para a extração de características de imagens e os classificadores supervisionados de múltiplos rótulos na tarefa de detecção de doenças torácicas nos *CXR*, esse método atingiu resultados que superaram o estado da arte no conjunto de dados *ChestX-ray14 Dataset*. Pham et al. (2020) integraram uma metodologia inovadora na qual os rótulos incertos do *CheXpert Dataset* (Tabela 2) e as dependências hierárquicas existentes entre doenças da base de dados são explorados de forma eficiente, superando estudos anteriores com bons resultados.

Yang et al. (2019) consideraram os rótulos incertos como uma classe distinta durante o processo de treinamento, permitindo ao modelo produzir menos previsões de confiança em casos ambíguos, oposto aos modelos treinados com rótulos binários (*U-ones*, etc). Os resultados das abordagens mencionadas são apresentados na Tabela 1.

Este trabalho apresenta uma abordagem que busca estudar o uso de Redes Neurais Profundas aplicadas ao diagnóstico de Cardiomegalia em imagens de radiografia do tórax, do *CheXpert Dataset*. Para alcançar esse propósito, utiliza-se uma metodologia baseada no balanceamento de quantidades de imagens utilizadas no treino e em ajustes empíricos dos hiperparâmetros do modelo.

Tabela 1. Comparação da performance de trabalhos da literatura. Métrica AUC.

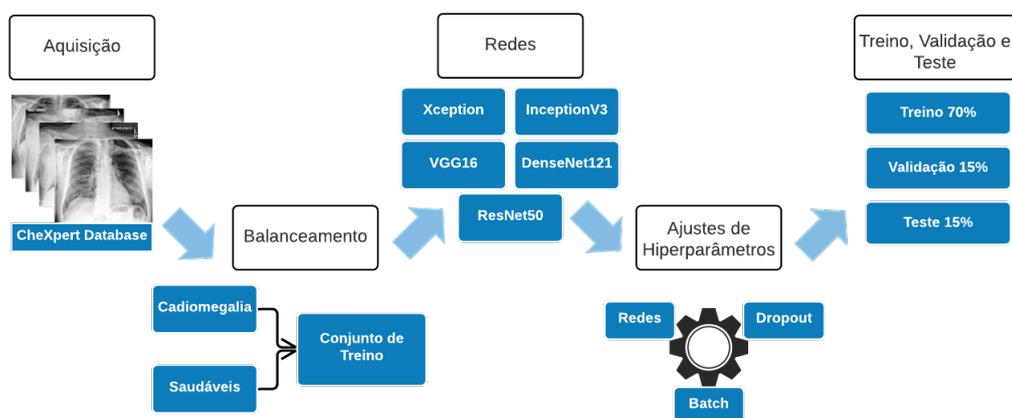
Métodos	Cardiomegalia
U-Ignore [Irvin et al.]	0,818
U-Zeros [Irvin et al.]	0,811
U-Ones [Irvin et al.]	0,858
U-MultiClass [Irvin et al.]	0,821
U-SelfTrained [Irvin et al.]	0,833
U-Ignore+LP [Allaouzi et al.]	0,720
U-Ignore+BR [Allaouzi et al.]	0,720
U-Ignore+CC [Allaouzi et al.]	0,700
U-Ignore+CT [Pham et al.]	0,780
U-Zeros+CT+LSR [Pham et al.]	0,806
U-Ones+CT+LSR [Pham et al.]	0,825
Ensemble with TTA [Pham et al.]	0,909
Nossa Abordagem	0,879

Fonte: Adaptado de [Pham et al. 2020]

2. Materiais e Método

A metodologia proposta nesse trabalho está dividida em 5 etapas principais, sendo elas: (1) aquisição e preparação de dados onde ocorre uma etapa adicional de (2) balanceamento de quantidade de imagens; (3) escolha de arquiteturas, subdividida em comentários sobre as redes utilizadas; (4) ajustes de hiperparâmetros e (5) Treino/Validação/Teste. Essas etapas são ilustradas na Figura 2.

Figura 2. Etapas da metodologia utilizada.



Fonte: Autor.

2.1. Aquisição e Preparação dos Dados

O *CheXpert Dataset* foi escolhido para este trabalho, uma base de dados contendo mais de 200,000 imagens de raios-X do tórax. CheXpert [Irvin et al. 2019] é um conjunto de dados disponibilizados ao público para o estudo de radiografias torácicas, ele é composto por 224.316 exames de raio-X do tórax pertencentes a 65.240 pacientes.

Os exames de Raios X, divididos em frontais e laterais, foram realizados no *Stanford Hospital* entre Outubro de 2002 e Julho de 2017, tanto em centros de internamento como em ambulatórios. Cada imagem de raio-X é multi-rotulada dentre catorze desco-

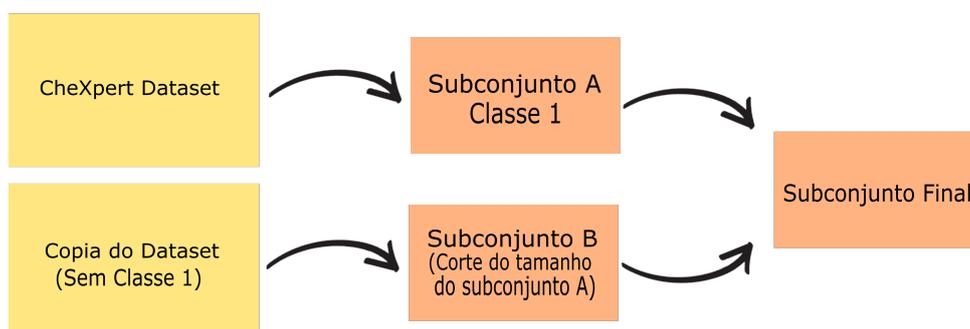
bertas clínicas diferentes, sendo elas Atelectasia, Cardiomegalia, Cardiomediastino Alargado, Consolidação, Pneumonia, Pneumotórax, Edema Pulmonar, Opacidade Pulmonar, Lesão Pulmonar, Derrame Pleural, *Pleural Other*, Fratura, Dispositivos de suporte e Sem descoberta. Essa distribuição de achados clínicos está disposta na Tabela 2.

A preparação dos dados ocorreu na forma de alterações pontuais nas instâncias da base de dados. Foram removidas as imagens de raio-X laterais para que o treino do modelo tivesse foco em apenas exames frontais, os mais comuns e utilizados em trabalhos na literatura. Além disso, as imagens mantidas foram redimensionadas para o tamanho 224×224. Os rótulos incertos foram tratados através da abordagem de *U-ignore* (ignorar as classificações incertas), para que apenas os rótulos positivos e negativos sejam utilizados no treino.

2.1.1. Balanceamento de Quantidade de Imagens

O balanceamento da proporção do número de imagens é promovido através da geração de um subconjunto A composto de casos positivos dos rótulos de interesse, juntamente com um subconjunto B que consiste de imagens rotuladas negativamente na classe em questão com quantidade aproximada ao subconjunto A. O processo é ilustrado na Figura 3.

Figura 3. Processo de balanceamento de quantidades.



Fonte: Autor.

Deste modo, o modelo é treinado com um conjunto de tamanho condizente com a quantidade de imagens do rótulo de interesse, visando evitar que a aprendizagem do modelo em tal classe seja alienada para os casos negativos devido aos seus números. A discrepância entre as quantidades de rótulos positivos e negativos pode ser vista na Tabela 2.

2.2. Escolha de Arquiteturas

As redes VGG16, DenseNet121 e ResNet50 foram escolhidas devido a sua aplicação anterior nos trabalhos relacionados a essa pesquisa, por outro lado, as arquiteturas InceptionV3 e Xception selecionadas em razão de haver uma melhora na métrica de especificidade, na qual as três redes anteriores apresentam certa inferioridade em comparação (Tabela 3). Além disso, as camadas das arquiteturas utilizadas nos testes foram congeladas, para evitar *overfitting* e gerar o *fine-tuning* nas imagens de raio-X de tórax do *CheXpert Dataset*.

Tabela 2. Distribuição da quantidade de cada rótulo do *Chexpert Dataset*.

Patologia	Positivo (%)	Incerto (%)	Negativo (%)
Sem achados	16627 (8,86)	0 (0,0)	171014 (91,14)
Cardiomeiastino Alargado	9020 (4,81)	10148 (5,41)	168473 (89,78)
Cardiomegalia	23002 (12,26)	6597 (3,52)	158042 (84,23)
Lesão Pulmonar	6856 (3,65)	1071 (0,57)	179714 (95,78)
Opacidade Pulmonar	92669 (49,39)	4341 (2,31)	90631 (48,3)
Edema Pulmonar	48905 (26,06)	11571 (6,17)	127165 (67,77)
Consolidação	12730 (6,78)	23976 (12,78)	150935 (80,44)
Pneumonia	4576 (2,44)	15658 (8,34)	167407 (89,22)
Atelectasia	29333 (15,63)	29377 (15,66)	128931 (68,71)
Pneumotórax	17313 (9,23)	2663 (1,42)	167665 (89,35)
Derrame Pleural	75696 (40,34)	9419 (5,02)	102526 (54,64)
Pleural Other	2441 (1,3)	1771 (0,94)	183429 (97,76)
Fratura	7270 (3,87)	484 (0,26)	179887 (95,87)
Dispositivo de Suporte	105831 (56,4)	898 (0,48)	80912 (43,12)

Fonte: [Irvin et al. 2019]

2.2.1. VGGNet

VGGNet, ou VGG-16, é uma arquitetura de rede neural convolucional proposta por Karen Simonyan e Andrew Zisserman da *University of Oxford* no ILSVRC (*ImageNet Large Scale Visual Recognition Challenge*) 2014 [Simonyan and Zisserman 2014].

É uma rede caracterizada pela sua simplicidade e arquitetura uniforme, com 5 etapas de camadas convolucionais 3x3 uma após a outra em profundidade crescente, lidando com a redução de tamanho do volume através de camadas de *MaxPooling*, sua saída ocorre por meio de duas camadas completamente conectadas de 4.096 nódulos cada, seguido de uma função *Softmax*.

Portanto, a rede VGG16 possui boa eficiência na extração de características de imagens. Sua desvantagem é o grande número de parâmetros, mais de 130 milhões, o que a torna a arquitetura computacionalmente pesada com um longo tempo de treino necessário caso for treinada inicializando os pesos aleatoriamente.

2.2.2. InceptionV3

Inception network, ou GoogLeNet, com 42 camadas e a menor taxa de erro da competição foi a arquitetura classificada em segundo lugar no ILSVRC (*ImageNet Large Scale Visual Recognition Competition*) 2015 [Szegedy et al. 2015].

Em contraste com a VGGNet, a rede Inception torna a arquitetura mais larga, em vez de mais profunda, com camadas convolucionais de diferentes tamanhos em um mesmo nível, concatenando a saída. A arquitetura foi atualizada com melhoras iterativas, a segunda versão renovou o conceito das camadas da rede com novos princípios, fatorando as camadas de tamanho 5x5 em duas de 3x3, as quais são 2,78 vezes mais eficientes que as anteriores. E fatorou camadas de filtro com tamanho NxN em combinações de 1xN e Nx1, aumentando a profundidade, mas diminuindo o custo. Tornando a saída ainda mais larga com combinações de camadas de tamanho 1x1 e 3x3.

A terceira versão continuou a diminuição de custo da arquitetura fatorando camadas de tamanho 7x7 em camadas menores, acompanhado da assimilação de *Label*

Smoothing, do otimizador RMSprop e da normalização de *Batch* nos classificadores auxiliares da primeira versão. Deste modo, a rede Inception é eficiente com baixo custo computacional.

2.2.3. Xception

A Xception foi proposta por Francois Chollet sendo uma versão adaptada da arquitetura Inception [Szegedy et al. 2015], substituindo os módulos padrões da rede Inception com uma convolução modificada separável em profundidade, criando uma arquitetura de desempenho ainda melhor que a Inception-v3 em ambos os conjuntos de dados ImageNet ILSVRC e JFT [Chollet 2017].

A convolução modificada é a convolução 1×1 seguida de uma convolução em profundidade sem ativação intermediária da função ReLU. Esta modificação é motivada pelo módulo inicial da Inception-v3 em que a convolução 1×1 é feita primeiro antes de qualquer convolução espacial $N \times N$, logo, é ligeiramente diferente da original.

A estrutura da rede Xception é composta de camadas de convolução separáveis tratadas como os blocos da Inception, sendo posicionadas ao longo de toda a estrutura da arquitetura. Além de haver conexões residuais (tal qual proposto na ResNet [He et al. 2015]) em todos os fluxos, beneficiando a acurácia do modelo e sendo uma parte essencial da rede Xception.

2.2.4. DenseNet

A arquitetura da DenseNet foi proposta no artigo *Densely Convolutional Networks* [Huang et al. 2016]. Esta rede foi criada com o intuito de mitigar o problema da perda informação original que pode ocorrer quando uma rede possui muitas camadas.

Cada camada obtém entradas adicionais de todas as camadas anteriores e passa os seus próprios mapas de características para todas as camadas subsequentes por meio da concatenação, cada camada recebe o resultado atual e uma cópia da entrada da camada anterior. Logo, as camadas recebem os mapas de características das camadas precedentes, a rede se torna mais robusta na atualização dos mapas de características, ou seja, o número de canais pode ser menor. Desse modo, a arquitetura DenseNet assegura eficiência computacional e eficiência de memória, tendo um tamanho consideravelmente menor que a ResNet [He et al. 2015].

A DenseNet tem como vantagens, um forte fluxo de gradiente, o sinal de erro pode ser facilmente propagado para camadas anteriores mais diretamente, sendo uma espécie de supervisão profunda implícita, dado que as camadas anteriores podem obter supervisão direta a partir da camada de classificação final.

Ademais, possui características mais diversificadas, uma vez que cada camada na DenseNet recebe todas as camadas anteriores como entrada, assegurando características mais variadas e tende a obter padrões com maior poder discriminatório. Mantém, características de baixa complexidade, o classificador utiliza atributos de todos os níveis de complexidade, inclinada a dar limites de decisão mais suaves. Isso justifica o desempenho da DenseNet com os dados desbalanceados.

2.2.5. ResNet

A ResNet, também conhecida como rede neural residual [He et al. 2015], venceu a competição ImageNet 2015 com a ideia de adicionar aprendizagem residual ao conceito das redes neurais convolucionais tradicionalmente usadas. Resolvendo então o problema da dispersão de gradientes e degradação de precisão (no conjunto de treino) em redes neurais profundas, de modo que a rede possa obter cada vez mais profundidade, mantendo o controle sobre a precisão e reduzindo o custo de treinamento.

A ResNet introduz uma ligação de atalho de identidade, a qual salta uma ou mais camadas. Dessa forma, é mais fácil deixar as camadas empilhadas encaixarem num mapeamento residual do que deixá-las se encaixar diretamente no mapeamento subjacente desejado.

2.3. Ajuste de Hiperparâmetros

Nesta seção são discutidos os ajustes aplicados nos hiperparâmetros do modelo, de forma empírica, no qual 5 diferentes redes inicialmente, seguidos de valores de *Dropout* e *Batch Size* aplicados a duas destas redes.

Os primeiros testes foram aplicados na classe *No Finding* com as 5 arquiteturas citadas na seção anterior, treinados por 100 épocas cada, com divisão de 80%, 10% e 10% para treino, validação e teste, respectivamente, e otimizador Adam. Os resultados dos testes apontam um melhor desempenho das arquiteturas InceptionV3 e Xception em comparação as outras, por esse motivo estas duas redes foram utilizadas com mais ênfase nesse trabalho.

Utilizando apenas a rede InceptionV3 nos testes subsequentes de tamanho do *Batch* e *Dropout*, devido a sua equivalência a Xception, foram experimentados diferentes tamanhos de *Batch* (16, 32, 64, 128, 256). Os resultados mais satisfatórios se mostraram com *Batch* de tamanho 32, o padrão geralmente utilizado na literatura. Continuando os testes, foram experimentados diferentes valores de *Dropout*, entre 0,15, 0,20, 0,25 e 0,35, onde o *Dropout* 0,25 foi resultado mais satisfatório dentre eles.

A configuração de Hiperparâmetros escolhida empiricamente a serem utilizados no treino das arquiteturas em foco, Inceptionv3 e Xception, chegou à configuração atual com o otimizador Nadam, valores de *Dropout* 0,25 e de tamanho de *Batch* 32. O Processo de seleção em cada uma dessas etapas é ilustrado na Figura 4.

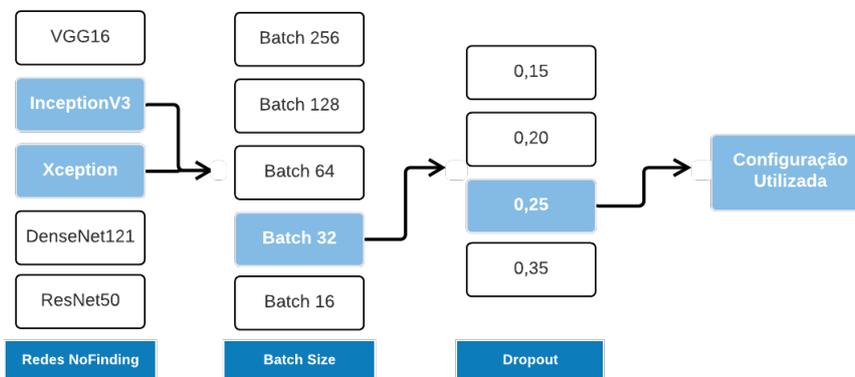
2.4. Treino, Validação e Teste

Após o ajuste de hiperparâmetros, os conjuntos de Treino, Validação e Teste foram divididos respectivamente em 70%, 15% e 15%. Os resultados estão sendo apresentados para experimentos realizados com 100 e 200 épocas. Os treinos dos modelos focaram-se na classe de Cardiomegalia para ser observada a capacidade de detecção da rede entre exames de uma patologia específica e de pacientes saudáveis.

3. Resultados Preliminares e Discussão

Nesta seção são apresentados os resultados iniciais obtidos com a metodologia aplicada no diagnóstico de cardiomegalia. A Tabela 3 apresenta os resultados dos experimentos com a metodologia na base aplicada nas classes normal e cardiomegalia.

Figura 4. Fluxo de etapas dos ajustes empíricos de Hiperparâmetros.



Fonte: Autor.

Tabela 3. Resultados iniciais da metodologia de balanceamento de quantidades e ajustes empíricos de de parâmetros aplicada na classe de Cardiomegalia. ¹Acurácia, ²Precisão, ³F1-Score, ⁴Sensibilidade, ⁵Especificidade

Arquitetura	Épocas	AUC	Acc. ¹	Loss	Prec. ²	F1. ³	Sens. ⁴	Spec. ⁵
Xception	100	0,879	0,803	0,607	0,728	0,815	0,925	0,600
InceptionV3	100	0,877	0,809	0,650	0,725	0,818	0,939	0,596
VGG16	100	0,893	0,813	0,412	0,601	0,747	0,986	0,387
ResNet50	100	0,852	0,773	0,506	0,561	0,713	0,980	0,250
DenseNet121	100	0,903	0,825	0,398	0,650	0,782	0,981	0,469
Xception	200	0,867	0,803	0,855	0,752	0,812	0,881	0,701
InceptionV3	200	0,857	0,796	0,898	0,740	0,799	0,857	0,708
VGG16	200	0,904	0,817	0,404	0,565	0,720	0,991	0,270
ResNet50	200	0,873	0,802	0,462	0,537	0,697	0,993	0,190
DenseNet121	200	0,886	0,805	0,446	0,613	0,754	0,980	0,391

Fonte: Autor

A partir dos resultados apresentados, notou-se que no teste de 100 épocas a rede Xception alcançou o melhor resultado de AUC (**0,879**), enquanto a InceptionV3 obteve melhor acurácia (**0,809**), sensibilidade (**0,939**) e F1-Score (**0,818**). Por outro lado, no treino de 200 épocas, verificou-se que a especificidade continua inferior em comparação com as outras métricas, todavia, os resultados demonstram uma melhora significativa na métrica de ambas as arquiteturas comparativamente aos testes anteriores.

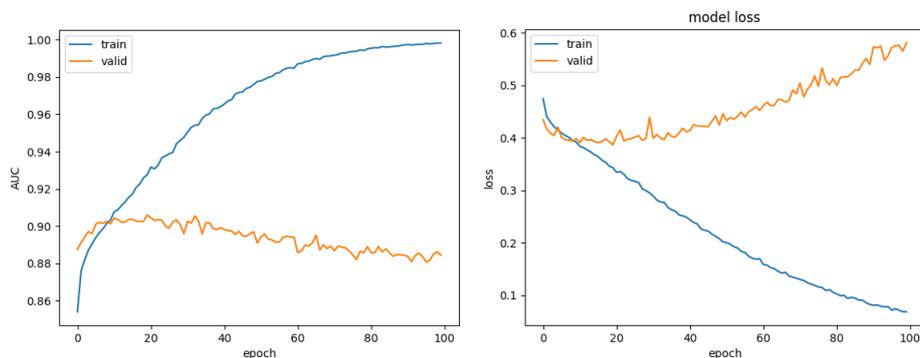
Na Figura 5 observa-se que o modelo Xception treinado por 100 épocas alcançou bons resultados durante o treino, entretanto, na etapa de validação constatou-se uma decaída dessas métricas juntamente com a estabilidade do gráfico da métrica AUC.

Na Figura 6 há um exemplo de *Heat Map* de um caso positivo de cardiomegalia, referente ao modelo Xception de 100 Épocas após o processo de treino, o que resultou em uma ativação precisa na região do coração, onde ocorre a patologia.

4. Conclusão e Trabalhos Futuros

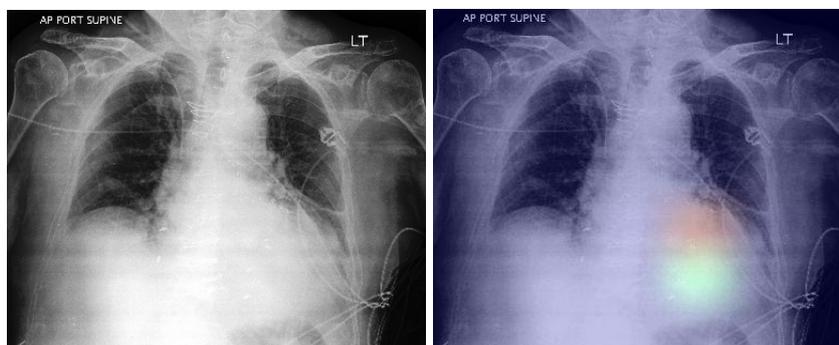
A metodologia apresentada neste trabalho em desenvolvimento promove o uso do balanceamento de quantidades de imagens e ajustes empíricos de parâmetros para a tarefa de detecção de doenças em imagens de raio-X da região do tórax. Os testes iniciais realizados obtiveram resultados promissores sob a métrica AUC, utilizada como principal

Figura 5. Plot do Gráfico de AUC (Esquerda) e Loss (Direita) do modelo Xception 100 Épocas.



Fonte: Autor.

Figura 6. Original (Esquerda) e Heat Map (Direita). Modelo Xception 100 épocas.



Fonte: Autor.

parâmetro de validação pelos trabalhos presentes na literatura. Os valores obtidos foram de 0,879 com a arquitetura Xception e 0,877 para InceptionV3.

Dessa forma, cabe ressaltar que os resultados preliminares apresentados nesse trabalho se provaram robustos e competitivos ao se aproximarem do segundo melhor resultado aplicado à classe de Cardiomegalia presente na literatura.

Portanto, com o objetivo de amadurecer a metodologia apresentada, como trabalhos futuros, pretende-se aplicar técnicas automáticas de otimização de hiperparâmetros com o HyperOpt [Bergstra et al. 2015], assim como adicionar uma etapa de pré processamento de imagens e utilizar estratégias como *Data Augmentation* para ampliar as quantidades de dados relevantes para as classes com poucos indivíduos. Por fim, continuar com a ampliação da pesquisa, contendo as 14 classes do *CheXpert Dataset*, para fins de validação da metodologia utilizada em um grande volume de dados.

5. Agradecimentos

Os autores desejam expressar seus agradecimentos ao apoio financeiro concedido pela CAPES (Finance code 001), CNPQ e FAPEMA para o desenvolvimento desta pesquisa e ao Núcleo de Computação Aplicada (NCA-UFMA) pelo suporte técnico e científico.

Referências

Agostoni, P., Cattadori, G., Guazzi, M., Palermo, P., Bussotti, M., and Marenzi, G. (2000). Cardiomegaly as a possible cause of lung dysfunction in patients with heart failure.

- American Heart Journal*, 140(5):A17–A21.
- Allaouzi, I. and Ahmed, M. B. (2019). A novel approach for multi-label chest x-ray classification of common thorax diseases. *IEEE Access*, 7:64279–64288.
- Amin, H. and Siddiqui, W. J. (2021). Cardiomegaly.
- Anderson, L., Dean, A., Falzon, D., Floyd, K., Baena, I., Gilpin, C., Glaziou, P., Hamada, Y., Hiatt, T., and A. Char, e. a. (2015). *Global tuberculosis report 2015*. World Health Organization, 20th edition.
- Bellaviti, N., Bini, F., Pennacchi, L., Pepe, G., Bodini, B., Ceriani, R., D’Urbano, C., and Vaghi, A. (2016). Increased incidence of spontaneous pneumothorax in very young people: Observations and treatment. *Chest Journal*, 150(4):560A.
- Bergstra, J., Komer, B., Eliasmith, C., Yamins, D., and Cox, D. D. (2015). Hyperopt: a python library for model selection and hyperparameter optimization. *Computational Science & Discovery*, 8(1):014008.
- Chollet, F. (2017). Xception: Deep learning with depthwise separable convolutions. pages 1800–1807.
- He, K., Zhang, X., Ren, S., and Sun, J. (2015). Deep residual learning for image recognition. *CoRR*, abs/1512.03385.
- Huang, G., Liu, Z., and Weinberger, K. Q. (2016). Densely connected convolutional networks. *CoRR*, abs/1608.06993.
- Irvin, J., Rajpurkar, P., Ko, M., Yu, Y., Ciurea-Ilcus, S., Chute, C., Marklund, H., Haggoo, B., Ball, R., Shpanskaya, K., Seekins, J., Mong, D. A., Halabi, S. S., Sandberg, J. K., Jones, R., Larson, D. B., Langlotz, C. P., Patel, B. N., Lungren, M. P., and Ng, A. Y. (2019). Chexpert: A large chest radiograph dataset with uncertainty labels and expert comparison. *AAAI-19/IAAI-19/EAAI-20 Proceedings*, 33(1):590–597.
- Pham, H. H., Le, T. T., Tran, D. Q., Ngo, D. T., and Nguyen, H. Q. (2020). Interpreting chest x-rays via cnns that exploit hierarchical disease dependencies and uncertainty labels.
- Randtke, M. A., Andrews, B. P., and Mach, W. J. (2015). Pathophysiology and prevention of intraoperative atelectasis: A review of the literature. *Journal of PeriAnesthesia Nursing*, 30(6):516–527.
- Simonyan, K. and Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., and Wojna, Z. (2015). Rethinking the inception architecture for computer vision. *CoRR*, abs/1512.00567.
- Yang, H.-Y., Yang, J., Pan, Y., Cao, K., Song, Q., Gao, F., and Yin, Y. (2019). Learn to be uncertain: Leveraging uncertain labels in chest x-rays with bayesian neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*.