

Classificação de Imagens de Raio-x de Torax com Reconhecimento Visual da IBM Cloud para Diagnóstico de Pneumonia

Raianny P. C. de Oliveira, Guilherme R. Sganderla, Claudio R. M. Mauricio, Fabiana F. F. Peres
Universidade Estadual do Oeste do Paraná
Ciência da Computação - Centro e Engenharias e Ciências Exatas
Foz do Iguaçu, Paraná
Email: raiannyproenca1@gmail.com

Resumo—A capacidade de aprender por meio de exemplos e formular previsões são as principais características do *Machine Learning*, uma subárea da inteligência artificial. Existem diversos *frameworks* disponíveis que utilizam *Machine Learning* para solução dos mais variados tipos de problemas, como para reconhecer e classificar objetos em uma imagem. Utilizando os serviços fornecidos por IBM *Watson Visual Recognition* que emprega algoritmos de *deep learning*, uma subárea de *Machine Learning*, um modelo foi criado e aplicado no *dataset Chest X-Ray Images for Classification*. Os resultados obtidos com o modelo criado foram comparados com a classificação geral fornecida pela IBM. Os serviços utilizados da *Watson Visual Recognition* são os disponibilizados para o plano do tipo *Lite*, um plano gratuito. Este trabalho discute como esta limitação afetou os resultados e descreve a eficiência da ferramenta nesta versão. Mesmo com as limitações o modelo obtido reconheceu corretamente um pulmão saudável em 75% das imagens de teste e classificou corretamente 93,34% das imagens de radiografias de tórax que retratam pneumonia.

I. INTRODUÇÃO

Os termos *Machine Learning* e *Deep Learning* compõem uma subárea da inteligência artificial. Uma de suas principais propriedades é a capacidade de aprender através de exemplos e entender a informação recebida gerando um modelo não-linear [1]. Esse modelo possibilita fazer previsões a partir de outros conjuntos de dados e ensinar o computador a encontrar padrões no conjuntos de dados para que então faça previsões baseadas nas variáveis de entrada [2]. Esse campo de pesquisa da inteligência artificial tem por objetivo imitar o processo feito pelo cérebro humano dando ao computador a capacidade de aprender de maneira semelhante, através de experiências e exemplos. A máquina obtém conhecimento a partir destas experiências de tal forma que torna desnecessária a especificação formal desse conhecimento por um operador humano. [3].

Diferentes problemas vêm sendo resolvidos com a aplicação de técnicas de *Deep Learning* disponíveis em *frameworks* que facilitam sua aplicação [4]–[12]. A área da medicina possui exemplos de aplicação para auxílio ao diagnóstico de câncer, pneumonia, nódulos, entre outros [13]–[15]. O *Framework* utilizado para o desenvolvimento deste trabalho é fornecido pela IBM Cloud, identicado como, API *Watson Visual Re-*

cognition [16]. Essa API utiliza algoritmos de *Deep Learning* para analisar imagens e classificá-las, oferece um conjunto de classes prontas para o uso e disponibiliza também meios de o usuário criar seu próprio classificador, customizando-o e treinando-o com a sua própria coleção de imagens.

Para este trabalho foi criado um modelo empregando os serviços dessa API com classes customizadas de acordo com o *dataset* utilizado (*Chest X-Ray Images for Classification* [17]). Foram fornecidos para o modelo criado, centenas de imagens separadas conforme seu conteúdo, como dados de entrada, para ele obter condições de que quando fosse fornecido novas imagens, ele fosse capaz de reconhecê-las e classificá-las. O objetivo foi analisar uma radiografia de tórax e identificar se o paciente retratava um caso de pneumonia ou se estava saudável. O modelo foi treinado e testado e os resultados obtidos foram comparados com a classificação feita pelo modelo geral pronto da IBM fornecido pelo *Watson Visual Recognition* [16].

II. CONCEITOS GERAIS

A. Aprendizagem Supervisionada

Na aprendizagem supervisionada o computador recebe um conjunto de padrões de entrada e seus correspondentes padrões de saída, ou seja, um conjunto de dados pré-estabelecido é fornecido e o computador aprende a partir disto para que receba novos dados e tenha condições de prever a resposta correta [18]. É indicado explicitamente para a rede se a resposta calculada é certa ou não e em seguida define um modelo. É uma técnica comumente utilizada em aplicações nas quais dados históricos podem prever possíveis acontecimentos futuros. Esse tipo de rede, através dos métodos de classificação e regressão, utiliza padrões para prever os valores dos rótulos também conhecidos como *label*. O método de classificação tem como alvo variáveis qualitativas. É o procedimento de receber algum tipo de dado como entrada e atribuir um rótulo a ele. É utilizado quando as previsões são de naturezas distintas e claras, por exemplo, o paciente está ou não doente. Denomina-se como classificação binária quando há apenas duas opções e classificação múltiplas quando há mais opções. Já o método de regressão tem como alvo variáveis

quantitativas. Utilizado quando o valor que está sendo previsto difere de respostas formadas como sim ou não, certo ou errado, mas quando segue um espectro contínuo.” [19].

B. Aprendizagem Não-supervisionada

Reconhecida por descobrir novos padrões. Neste caso os dados não possuem rótulos nem históricos, sendo esta a principal diferença em relação a aprendizagem supervisionada. O algoritmo não recebe variáveis de saídas ou possíveis resultados associados as entradas durante seu treinamento, sendo responsabilidade do algoritmo determinar os padrões dos dados. Explora os dados a fim de encontrar alguma estrutura neles para determinar estes padrões. Comumente usados para recomendar itens com base no seu histórico de pesquisa, por exemplo, e identificar valores que sejam discrepantes dentro de um conjunto de dados que segue certo modelo. Seu objetivo é organizar os dados, isso significa, encontrar diferentes maneiras fazer com que dados complexos se pareçam mais simples, ou então realizar a *clusterização*. Sendo este último o processo de agrupar conjuntos de dados similares [19].

III. MATERIAIS E MÉTODOS

A IBM Cloud oferece um amplo conjunto de serviços [16]. Um dos serviços ofertados é o *Watson Visual Recognition* que utiliza algoritmos de *Deep Learning* para analisar imagens e classificá-las com um conjunto de classes prontas para o uso e a opção de o usuário criar seu próprio classificador de imagens customizado. As classes que estão prontas para o uso são as classes que reconhecem: rostos, comidas, conteúdos explícitos e a geral. O modelo geral da IBM foi utilizado neste artigo como base de comparação dos resultados criados no modelo. Entre os vários planos ofertados pela empresa, o plano utilizado neste experimento foi o Lite que permite o treino de até 2 modelos por conta e é possível treinar 1000 imagens por mês apenas [16].

O *dataset* utilizado, Chest X-Ray Images for Classification, contém radiografias de tórax que são descritos e analisados por Kermany e outros [12]. Foram coletadas e identificadas um total de 5.232 radiografias de tórax onde 3.883 retratam pneumonia e 1.349 normais. Neste artigo foi utilizada a segunda versão deste *dataset* que foi disponibilizado no Mendeley Data [17]. A Figura 1 apresenta um exemplo de como as são as imagens desse *dataset*.



Figura 1. Exemplos de imagens presentes no conjunto de imagens do Chest X-Ray.

IV. EXPERIMENTOS

Para a realização do experimento as imagens foram particionadas em dois conjuntos mutuamente exclusivos, o de treinamento contendo 800 imagens e o de teste contendo 50 imagens. Para poder comparar os resultados e verificar de que forma a quantidade de imagens pode interferir no modelo gerado, o treinamento foi realizado da seguinte maneira: Foram realizados 5 treinos e após cada um desses treinos foi testado um conjunto com 50 imagens separadas para teste. A Tabela I demonstra quantas imagens foram usadas em cada treino. No Treino 1 foram treinadas 200 imagens na classe de radiografias de tórax saudáveis e 200 na classe de radiografias que retratam pneumonia, após este treino, o conjunto com as imagens separadas para teste foram passados e os resultados registrados; em seguida, nos demais treinos, foram acrescentadas mais imagens em cada uma das duas classes, conforme a Tabela I, e novamente o mesmo conjunto de imagens testes foi passado, totalizando 800 radiografias em cada uma e por fim foi testado o conjunto de imagens de teste para comparar com os resultados anteriores.

Tabela I
NÚMERO DE IMAGENS TREINADAS EM CADA TREINAMENTO.

	Treino 1	Treino 2	Treino 3	Treino 4	Treino 5	Total Imagens
Saudável	200	200	100	200	100	800
Pneumonia	200	200	100	200	100	800

Foi preparado um conjunto com 20 imagens classificadas como “saudáveis” para testar o modelo, importante ressaltar que as imagens utilizadas para testar o modelo não foram usadas para treiná-lo. Esse conjunto foi utilizado para testar cada um dos modelos gerados nos 5 treinos e os resultados estão demonstrados na Tabela II.

Tabela II
RESULTADO DA CLASSIFICAÇÃO DO CONJUNTO DE RADIOGRAFIAS SAUDÁVEIS NO MODELO CRIADO NO ARTIGO APÓS CADA TREINO.

	Treino 1	Treino 2	Treino 3	Treino 4	Treino 5
Saudável	15	11	8	9	15
Pneumonia	5	9	12	11	5

Observa-se que as 20 imagens deveriam ser classificadas como “saudáveis”, mas o modelo encontrou dificuldades em classificá-las. O melhor resultado foi após o treino 1 e o 5, quando ambas classes continham 200 e 800 imagens, respectivamente. Após ambos os treinos o modelo classificou corretamente 15 das 20 imagens, ou seja, 75%. A Figura 2 apresenta a classificação da imagem IM-0041-0001.JPG que foi testada após cada treino. A Figura 2 (a) foi testada após o primeiro treino, a Figura 2 (b) após o segundo, e assim respectivamente até a Figura 2 (e) ser passada como teste após o quinto treino. Observando os resultados verifica-se que o modelo classificou a imagem de forma correta após o

primeiro treino, ou seja, quando continha a menor quantidade de imagens. A partir do segundo treino o resultado foi invertido e após o quinto treino, quando o modelo continha 800 imagens, ele voltou classificar corretamente. O resultado da classificação da imagem IM-0041-0001.JPEG utilizando o modelo geral da IBM está ilustrada na Figura 3. Por ser algo tão específico, não era esperado que o modelo geral da IBM conseguisse distinguir se a radiografia retrata um caso de pneumonia ou não. Contudo, as classificações indicaram que a imagem apresenta um radiograma com 99% de certeza, cor cinza com 84%, cor de cinzas 81% e equipamento de fotografia com 80%.

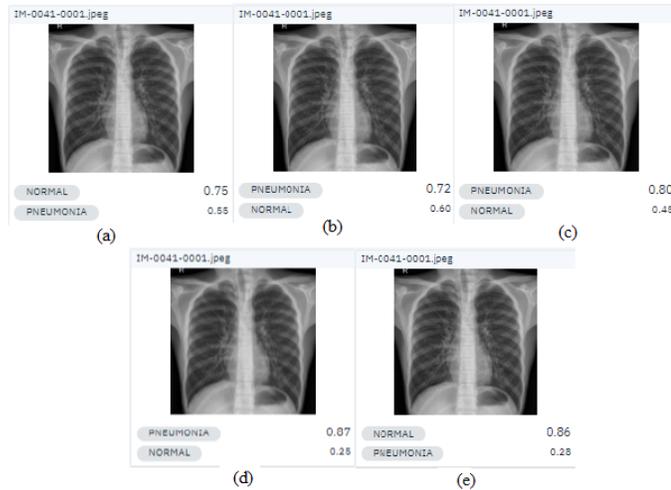


Figura 2. Classificação da imagem IM-0041-0001.jpeg após cada um dos cinco treinos.

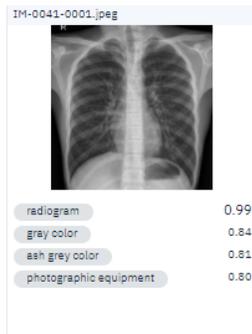


Figura 3. Resultado da classificação da imagem-IM-0041-0001 utilizando o modelo Padrão da IBM.

Assim como foi selecionado um conjunto com imagens de radiografia retratando um pulmão saudável também foram selecionadas um conjunto de 30 imagens para testes retratando um caso de pneumonia. Os resultados da classificação das 30 imagens selecionadas para o teste, após cada treinamento, estão expostos na tabela III

Nota-se através dos números expostos na tabela III que o modelo apresentou o mesmo resultado após o primeiro e o último treino, isto é, reconheceu 28 radiografias corretamente

Tabela III

RESULTADO DA CLASSIFICAÇÃO DO CONJUNTO DE RADIOGRAFIAS QUE RETRATAM PNEUMONIA NO MODELO DO ARTIGO APÓS CADA TREINO.

	Treino 1	Treino 2	Treino 3	Treino 4	Treino 5
Pneumonia	28	29	28	29	28
Saudável	2	1	2	1	2

de 30 no total. Os valores oscilaram a cada treino periodicamente. Os melhores resultados foram após o segundo e quarto treino em que a taxa de acerto foi de 96,67%. Porém, ao considerar a última classificação como correta, após o quinto treino o modelo apresentou uma taxa de acerto de 93,34%.

Os resultados alcançados e as classificações das imagens obtidas estão expostas e comentadas abaixo. A Figura 4 apresenta o resultado da classificação da imagem person1610_virus_2793.jpeg após cada um dos seus cinco treinos.

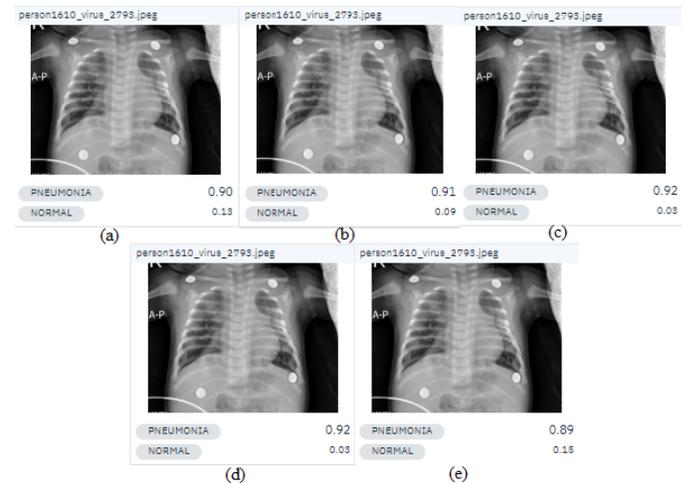


Figura 4. Resultado da classificação da imagem person1610_virus_2793.jpeg após cada um dos cinco treinos

O resultado da classificação da imagem person1610_virus_2793.jpeg, ilustrada na Figura 4 utilizando o modelo geral da IBM é demonstrado na Figura 5.



Figura 5. Resultado da classificação da imagem person1610_virus_2793.jpeg utilizando o modelo Padrão da IBM.

O resultado não difere muito do que foi apresentado anteriormente. O modelo geral da IBM reconheceu com 99% de

certeza que a imagem se trata de um radiograma e afirmou apresentar cor cinza. Importante levar em consideração o que foi dito anterior sobre o modelo geral da IBM, em que este está capacitado para reconhecer o conteúdo de uma imagem de forma geral e não está preparado para uma classificação tão específica como a de reconhecer se a imagem refere-se a uma radiografia que retrata pneumonia ou não. Logo, as comparações foram apenas para demonstrar a eficiência do modelo do artigo e como o modelo geral da IBM classifica essas imagens. O resultado da classificação das 50 imagens utilizadas para testar o modelo criado no artigo estão expostas na Tabela IV.

Tabela IV
RESULTADO DA CLASSIFICAÇÃO DAS 50 IMAGENS TESTADAS NO MODELO

	Imagens Acertadas/ Imagens Testadas	Média (%)
Saudável	15 de 20	75
Pneumonia	28 de 30	93,34

V. CONCLUSÃO

Para demonstrar a eficiência dos modelos criados, foram utilizados os resultados da classificação do modelo geral da IBM para comparar com o modelo gerado. É importante observar que modelo geral da IBM, como o próprio nome diz, tem a capacidade de reconhecer o conteúdo geral de imagens, mas não reconhece propriedades específicas como é o caso do modelo gerado neste trabalho, que tem como objetivo específico identificar se a radiografia do tórax retrata o caso de pneumonia ou não.

A experiência com a ferramenta IBM demonstrou que não foi possível classificar as imagens manualmente circulando cada objeto, como o conceito de *bounding box*; é necessário criar as classes e incluir as imagens do objeto que se quer reconhecer. Por isso é interessante incluir em cada classe, imagens do objeto abrangendo o máximo das diversidade ou casos que este possa apresentar. Devido a essa característica foram criadas classes dos itens que desejamos reconhecer nas imagens.

É fundamental levar em consideração alguns pontos relevantes sobre esse experimento como um todo que definem o modelo e interferem diretamente nos resultados. Em função das limitações da conta *lite*, foram utilizadas apenas 1600 imagens do *dataset* Chest X-ray, e em processos de *deep learning* o número de imagens no treinamento é relevante. O modelo classificou corretamente 75% das imagens da classe Saudável e 93,34% da classe Pneumonia, conforme pode ser visto na Tabela IV

Portanto, mesmo com as limitações do plano utilizado, pode-se considerar que os resultados das classificações foram bons, pois o modelo reconheceu corretamente 86% das 50 imagens do conjunto de testes. Acredita-se que utilizando todas as imagens do *dataset* os resultados podem ser melhorados.

REFERÊNCIAS

- [1] A. L. Christiane Spurl, Emiliano G. Castro, "Aplicação de redes neurais artificiais na construção de modelos de fragilidade ambiental," *Revista do Departamento de Geografia - USP*, vol. 21, pp. 113–135, 2011.
- [2] S. S. Damaceno and R. O. Vasconcelos, "Inteligência artificial: uma breve abordagem sobre seu conceito real e o conhecimento popular," *Revista periódica Caderno de Graduação das Ciências Exatas e Tecnológicas – CGCET*, vol. 5, no. 1, pp. 11–16, 2018.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016, <http://www.deeplearningbook.org>.
- [4] V. Kovalev, A. Kalinovsky, and S. Kovalev, "Deep learning with theano, torch, caffe, tensorflow, and deeplearning4j: Which one is the best in speed and accuracy?" in *XIII International Conference on Pattern Recognition and Information Processing*. Minsk: Publishing Center of BSU, 2016.
- [5] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: a Novel Image Dataset for Benchmarking Machine Learning Algorithms," aug 2017. [Online]. Available: <http://arxiv.org/abs/1708.07747>
- [6] E. Santos, "O uso da ferramenta Watson na definição dos protocolos quimioterápicos."
- [7] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," *Computers and Electronics in Agriculture*, vol. 147, pp. 70–90, apr 2018. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0168169917308803>
- [8] Q. Zhang, L. T. Yang, Z. Chen, and P. Li, "A survey on deep learning for big data," *Information Fusion*, vol. 42, pp. 146–157, jul 2018. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S1566253517305328>
- [9] M. Strumpe and M. Craig, "Applying Deep Learning to Metastatic Breast Cancer Detection," *Google AI Blog*, pp. 2018–2020, 2018. [Online]. Available: <https://ai.googleblog.com/2018/10/applying-deep-learning-to-metastatic.html>
- [10] W. Zhang, L. Yu, Y. Zhao, T. Bai, S. Song, F. Ma, and J. Su, "Chromosome Classification with Convolutional Neural Network Based Deep Learning," *2018 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)*, pp. 1–5, 2019.
- [11] P. Kiruthika, K. B. Jayanthi, and M. Nirmala, "Classification of metaphase chromosomes using deep learning neural network," *2018 4th International Conference on Frontiers of Signal Processing, ICFSP 2018*, no. c, pp. 110–114, 2018.
- [12] D. S. Kermany, M. Goldbaum, W. Cai, C. C. Valentim, H. Liang, S. L. Baxter, A. McKeown, G. Yang, X. Wu, F. Yan, J. Dong, M. K. Prasadha, J. Pei, M. Ting, J. Zhu, C. Li, S. Hewett, J. Dong, I. Ziyar, A. Shi, R. Zhang, L. Zheng, R. Hou, W. Shi, X. Fu, Y. Duan, V. A. Huu, C. Wen, E. D. Zhang, C. L. Zhang, O. Li, X. Wang, M. A. Singer, X. Sun, J. Xu, A. Tafreshi, M. A. Lewis, H. Xia, and K. Zhang, "Identifying Medical Diagnoses and Treatable Diseases by Image-Based Deep Learning," *Cell*, vol. 172, no. 5, pp. 1122–1131.e9, feb 2018. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0092867418301545>
- [13] L. S. Marques, R. R. Magalhães, and D. D. Ferreira, "Inteligência computacional aplicada à detecção de câncer de mama," *Revista Brasileira de Computação Aplicada*, vol. 11, no. 1, pp. 28–35, apr 2019. [Online]. Available: <http://seer.upf.br/index.php/rbca/article/view/8727>
- [14] W. E. Fathy and A. S., "A Deep Learning Approach for Breast Cancer Mass Detection," *International Journal of Advanced Computer Science and Applications*, vol. 10, no. 1, 2019.
- [15] M. Stumpe and L. Peng, "Assisting Pathologists in Detecting Cancer with Deep Learning," *Google AI Blog*, pp. 1–11, 2017. [Online]. Available: <https://research.googleblog.com/2017/03/assisting-pathologists-in-detecting.html>
- [16] I. W. Studio. Ibm watson visual recognition. [Online]. Available: <https://www.ibm.com/br-pt/cloud/watson-visual-recognition>
- [17] D. Kermany, K. Zhang, and M. Goldbaum, "Labeled Optical Coherence Tomography (OCT) and Chest X-Ray Images for Classification." [Online]. Available: <https://data.mendeley.com/datasets/rsbjbr9sj/2>
- [18] E. Ferneda, "Redes neurais e sua aplicação em sistemas de recuperação de informação," *Ci. Inf. [online]*, vol. 35, no. 1, pp. 20–35, 2006.
- [19] R. L. Stange, "Adaptatividade em aprendizagem de máquina: Conceitos e estudos de caso," Mestrado em Engenharia Elétrica, Adaptatividade em Aprendizagem de Máquina: Conceitos e Estudos de Caso, São Paulo, 2011.