



Comparando Deep e Transfer Learning na Classificação de Imagens da Membrana Timpânica

Comparing Deep and Transfer Learning in Tympanic Membrane Image Classification

Comparación del Deep y Transfer Learning en la Clasificación de Imágenes de la Membrana Timpánica

Maurício Koenigkam Santos¹, Carine Geltrudes Webber²

RESUMO

Descritores: Aprendizado de Máquina; Redes Neurais; Processamento de Imagens Assistida por Computador

Objetivos: As redes neurais de aprendizado profundo compreendem técnicas que produzem os melhores resultados na tarefa de classificação de imagens. Em especialidades como a radiologia e a dermatologia, as redes neurais têm contribuído na identificação, por meio de imagens, de tumores e melanomas respectivamente. **Método:** No contexto deste artigo, explora-se o uso das redes neurais para diagnóstico de otites, por meio de imagens da membrana timpânica. Considerando esta tarefa, o método seguido compreendeu a aquisição de imagens do conduto auditivo externo humano de pacientes saudáveis, com otite e com cerumen. **Resultado:** Posteriormente, constituiu-se dois modelos, sendo o primeiro baseado em uma arquitetura deep learning e o segundo utilizando transfer learning. Como resultado, identificou-se que o modelo produzido por meio de transfer learning apresentou melhores resultados. **Conclusão:** A acurácia obtida na tarefa de classificação foi de 96%, revelando a aptidão do método para a classificação e diagnóstico esperados.

ABSTRACT

Keywords: Machine Learning; Neural Networks; Image Processing

Objectives: Deep learning neural networks comprise techniques that produce the best results in the task of classifying images. In specialties such as radiology and dermatology, neural networks have contributed to the identification, through images, of tumors and melanomas, respectively. In the context of this article, the use of neural networks for the diagnosis of otitis is explored, by means of images of the tympanic membrane. **Method:** Considering this task, the method followed comprised the acquisition of images of the human external auditory duct of healthy patients, with otitis and with cerumen. Two models were later built, the first being based on a deep learning architecture and the second using transfer learning. **Results:** As a result, it was identified that the model produced by means of transfer learning presented better results. **Conclusion:** The accuracy obtained in the classification task was 96%, revealing the suitability of the method for the classification and diagnosis expected.

RESUMEN

Descriptores: Aprendizaje Automático; Redes Neuronales; Procesamiento de Imagen Asistido por Computador

Objetivos: Las redes neuronales de aprendizaje profundo comprenden técnicas que producen los mejores resultados en la tarea de clasificar imágenes. En especialidades como la radiología y la dermatología, las redes neuronales han contribuido a la identificación, mediante imágenes, de tumores y melanomas, respectivamente. **Métodos:** En el contexto de este artículo, se explora el uso de redes neuronales para el diagnóstico de otitis, a través de imágenes de la membrana timpánica. Teniendo en cuenta esta tarea, el método seguido consistió en la adquisición de imágenes de la conducción auditiva externa humana de pacientes sanos, con otitis y cerumen. Posteriormente se crearon dos modelos, el primero basado en una arquitectura de aprendizaje profundo y el segundo en el aprendizaje de transferencia. **Resultados:** En consecuencia, se ha comprobado que el modelo producido mediante la transferencia de conocimientos ha producido mejores resultados. **Conclusión:** La precisión obtenida en la tarea de clasificación fue del 96%, lo que reveló la idoneidad del método para la clasificación y el diagnóstico esperados.

¹ Bacharel em Tecnologias Digitais, Médico Otorinolaringologista, Área de Exatas e Engenharias, Universidade de Caxias do Sul - UCS, Caxias do Sul (RS), Brasil.

² Professor Titular, Área de Exatas e Engenharias, Universidade de Caxias do Sul - UCS, Caxias do Sul (RS), Brasil.

INTRODUÇÃO

O aprendizado de máquina (AM) é um campo de estudo da IA que se ocupa do desenvolvimento de métodos que permitam que as máquinas aprendam por meio da experiência⁽¹⁾. Além de ajudar na compreensão da inteligência humana, o AM pode trazer vários benefícios ao cotidiano humano, possibilitando a criação de artefatos adaptáveis, evolutivos e cognitivos. Como exemplos cita-se veículos autoguiados, detectores de fraudes, assistentes pessoais e sistemas preditivos. Dentre os métodos mais usados, destacam-se as Redes Neurais.

Recentemente, com o aparecimento das redes Deep Learning ou redes neurais profundas (RNP), o processo de aprendizado se tornou mais eficiente podendo ocorrer a partir de imagens. As RNPs são formadas por uma arquitetura em camadas, englobando extração de características (camadas convolucionais) e classificação (camadas densamente conectadas). Cada camada convolucional realiza uma etapa do processamento de imagens, extraíndo objetos ou figuras em imagens⁽²⁾. As camadas densas se apoiam no comportamento de neurônios, interconectados, que mapeiam padrões de entrada para saídas esperadas. Na área da saúde, uma extensa pesquisa analisou 308 trabalhos que utilizaram arquiteturas de Deep Learning de forma automática em imagens médicas⁽³⁾. Os autores relatam o uso dessa técnica em diversas áreas da medicina como neurologia, oftalmologia (retina), pneumologia, patologia digital, tórax, cardiologia, abdômen e musculoesquelético. Posteriormente a este estudo, diversos outros seguiram, dando continuidade e constituindo um repertório válido de trabalho na área médica, validando o uso das RNP na tarefa de classificação de imagens. Tais análises podem auxiliar em tarefas de diagnóstico de doenças e tomada de decisões na área da saúde^{3,4}. Pode-se seguir uma abordagem original de criar ou inventar Redes Neurais convolucionais, ou pode-se usar modelos de redes pré-treinadas com dados similares, mas de muito maior volume. Usualmente, as redes das famílias VGG e Inception são exemplares de arquiteturas que obtêm resultados promissores⁽⁴⁻⁵⁾ em processos de aprendizado por transferência.

A otite média representa uma condição inflamatória dos espaços da orelha média e da mastoide, sem referência para sua etiologia ou patogênese⁽⁶⁾. Essa doença é o principal motivo de visita aos médicos por crianças na idade pré-escolar. Estudos realizados nos Estados Unidos mostram que 71% das crianças com até 3 anos de idade tiveram pelo menos uma otite média aguda, sendo que 33% tiveram três ou mais episódios⁽⁷⁾. Nesse mesmo país, o custo anual para o tratamento de otite média em crianças menores que 13 anos de idade é de 4 bilhões de dólares⁽⁸⁾. No mundo, acredita-se que aproximadamente 740 milhões de pessoas por ano serão afetadas por otite média aguda ou otite média crônica supurada, e aos 10 anos de idade pelo menos 90% das crianças terão sido acometidas por otite média com efusão⁽⁸⁾. O diagnóstico é muito importante em crianças menores que 1 ano de idade, pois a chance de ocorrer

otites de repetição nessas crianças é grande. Um estudo realizado por Kuruvilla et al. propôs uma classificação automática de dois tipos de otite (otite média aguda e otite média com efusão) baseada na captura de imagens através da otoscopia⁽⁹⁾. Os autores utilizaram algumas técnicas para trabalhar as imagens, com o objetivo de eliminar obstáculos do conduto auditivo, eliminar problemas de luminosidade e identificar a membrana timpânica. Desse modo, puderam obter várias características da imagem da membrana (luz, opacidade, concavidade, posição do martelo, etc.) e usaram um algoritmo de vocabulário para fazer a classificação. Obtiveram uma acurácia de 89.9%, maior que a dos médicos não especialistas, segundo o estudo.

Este trabalho teve como objetivo avaliar e comparar modelos de classificação de imagens da membrana timpânica humana, podendo assim ajudar no diagnóstico da otite média. Buscou-se avaliar um modelo gerado por meio de uma rede neural *deep learning*, originalmente concebida para a tarefa, comparando-o com um segundo modelo obtido por meio de *transfer learning*. Os métodos empregados, bem como os resultados obtidos pelos modelos, são descritos nas seções seguintes.

MÉTODOS

O método seguido compreendeu etapas de coleta de imagens, preparação e transformação das imagens, treinamento dos dois modelos e testes e ajustes para aprimoramento dos resultados. Os dois modelos definidos foram denominados de aprendizado profundo (uso do método tradicional de treinamento completo da rede via *deep learning*) e aprendizado por *transfer learning* (uso de um modelo pré-treinado).

Coleta de Dados

Uma etapa importante deste trabalho foi a aquisição das imagens, para constituição do *dataset*, bem como a sua rotulação prévia. O processo de aquisição de imagens do conduto auditivo externo humano foi realizado por um pesquisador, participante do estudo, médico especialista em otorrinolaringologia pela ABORL-CCF (Associação Brasileira de Otorrinolaringologia e Cirurgia-Cérvico Facial). A aquisição das imagens do conduto auditivo externo humano foi realizada por meio do dispositivo *Oto for Clinicians (Cellscope)*⁽¹⁰⁾. O dispositivo acompanha um aplicativo denominado *CellScopeLite*⁽¹¹⁾. Os arquivos são gerados em formato JPG. O dispositivo utiliza a câmera e iluminação de um *smartphone* compatível, capturando a imagem por meio de um espelho de ouvido, e salvando-a no próprio *smartphone*.

Foram identificadas imagens, posteriormente rotuladas conforme três classes: “sem doença”, “com doença” e “com cerumen”. Os arquivos gerados variaram entre 625 kB e 1.48 MB. As imagens mantiveram as dimensões de 3024 x 3024 pixels, a resolução de 72 x 72 DPI (pontos por polegada) e a representação de cores em sRGB. Previamente à coleta das imagens foi realizada uma otoscopia com otoscópio pneumático da marca Welch Allyn. A partir das informações clínicas do paciente foi

realizado o diagnóstico em relação a orelha externa e média. Esta etapa produziu como resultados um conjunto de imagens classificadas. Foram obtidas 214 imagens do conduto auditivo externo humano, sendo que 91 foram classificadas como sendo “sem doença”, 61 como sendo “com doença” e 62 como sendo “com cerumen”.

Pré-processamento dos Dados para Aprendizado Profundo

Inicialmente foi realizada a diminuição das dimensões da imagem através do corte de pixels periféricos. As imagens coletadas apresentam uma região de coloração preta, que correspondem ao espelho do dispositivo que capta as imagens da otoscopia. Essa região correspondia a quase 50% das dimensões da imagem e poderia gerar um custo computacional desnecessário durante o treinamento da rede neural. Portanto, optou-se por utilizar um quadrado central da imagem, que correspondia a 50% dos pixels, importante para a classificação proposta neste trabalho. Para este processo, foi utilizada uma biblioteca para Python chamada Pil (criada por Fredrik Lundh e colaboradores), que é utilizada na manipulação de imagens. Esta biblioteca apresenta uma função chamada “crop”, na qual recebe um conjunto de quatro números que delimitam um quadrado para ser retirado de uma imagem. Os 2 primeiros números são as coordenadas do canto superior esquerdo, e os outros 2 números são as coordenadas do canto inferior direito. Através desta biblioteca, foi possível representar as imagens originais de 3024 x 3024 pixels, reduzindo para 1512 x 1512 (50% do original), o que corresponde a uma matriz com 6.858.432 valores (uma redução de 75% nestes valores).

Em seguida, foi aumentado o número de imagens totais através das operações de rotação e espelhamento. Cada imagem originou um total de 6 novas imagens: rotações de 90, 180, 270 graus; espelhamento direita-esquerda; espelhamento em cima-embaixo e espelhamento diagonal. Deste modo, foram obtidas 931.

Treinamento para Aprendizado Profundo

Uma rede neural convolucional (CNN) possibilita a construção de modelos aprendidos automaticamente a partir de sinais de entrada, como imagens médicas. Um modelo de CNN compreende um conjunto de filtros convolucionais cuja função principal é a extração de características úteis e eficientes para entendimento de imagens médicas⁽²⁾. Nesta parte do estudo desenvolveu-se uma arquitetura de rede neural profunda partindo-se de modelos existentes e adaptando-os.

Foram testados vários modelos de CNNs para se obter um resultado satisfatório. Dentre os modelos consagrados, foram testados adaptações do AlexNet e Inception V1. Foram feitas adaptações para que essas redes pudessem ser construídas de modo sequencial na linguagem Python, utilizando o TensorFlow. A AlexNet, por exemplo, tem sua arquitetura original feita para ser treinada de modo paralelo em dois processadores⁽¹²⁾. Originalmente o modelo da AlexNet apresenta um total de 28.081.754 parâmetros, contendo grande número de camadas e filtros. O modelo adaptado que obteve os melhores resultados é o representado na figura 1, e apresenta um total de 3.340.875 parâmetros, além da seguinte seqüência de camadas: 2 camadas convolucionais, 1 camada de agregação, 2 camadas convolucionais, 1

```
model.summary()
```

Layer (type)	Output Shape	Param #
conv2d_1 (Conv2D)	(None, 749, 749, 8)	6152
conv2d_2 (Conv2D)	(None, 369, 369, 16)	18448
max_pooling2d_1 (MaxPooling2D)	(None, 184, 184, 16)	0
conv2d_3 (Conv2D)	(None, 89, 89, 32)	32800
conv2d_4 (Conv2D)	(None, 43, 43, 48)	24624
max_pooling2d_2 (MaxPooling2D)	(None, 21, 21, 48)	0
conv2d_5 (Conv2D)	(None, 10, 10, 64)	12352
max_pooling2d_3 (MaxPooling2D)	(None, 5, 5, 64)	0
dropout_1 (Dropout)	(None, 5, 5, 64)	0
flatten_1 (Flatten)	(None, 1600)	0
dense_1 (Dense)	(None, 2024)	3240424
dropout_2 (Dropout)	(None, 2024)	0
dense_2 (Dense)	(None, 3)	6075
Total params: 3,340,875		
Trainable params: 3,340,875		
Non-trainable params: 0		

Figura 1 - Arquitetura da Rede do Modelo 1

camada de agregação, 1 camada convolucional, 1 camada de agregação e 2 camadas densamente conectadas. Observou-se que, em relação a AlexNet, neste estudo foram obtidos melhores resultados trabalhando imagens com maior resolução e utilizando-se uma menor quantidade de características destas imagens.

Pré-processamento dos Dados para *Transfer Learning*

Nesta etapa, o número de imagens foi aumentado com o auxílio de uma classe da biblioteca Keras, o ImageDataGenerator⁽¹³⁾. Para cada imagem foram produzidas mais outras 5, com distorções do tipo deslocamento de largura e altura, cortes, rotação e zoom. A figura 2 ilustra um exemplo de uma imagem original e uma produzida pela ferramenta. Estas distorções são produzidas de modo aleatório, mas dentro de uma faixa determinada pelo programador. Portanto, o treinamento da rede neural foi realizado com um total de 1284 imagens, mantendo-se a proporção dentro das classes. As imagens foram redimensionadas para o tamanho de 224 x 224 pixels com o auxílio da biblioteca Python Imaging Library⁽¹⁴⁾, isto para se adequar ao formato de entrada na rede neural escolhida: VGG19.

Além disso, foi utilizado um método do Keras de pré-processamento específico para rede VGG19⁽¹⁵⁾. Este método converte as imagens de RGB para BGR, centralizando em zero cada canal de cor em relação ao conjunto de dados ImageNet, sem escala. Ao término do pré-processamento obteve-se um conjunto de 1284 imagens no formato BGR, representadas na forma de matrizes tridimensionais (224,224,3).

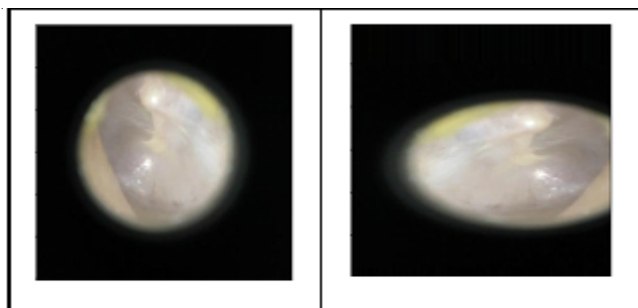


Figura 2 - Distorção de uma imagem de membrana timpânica

Treinamento para *Transfer Learning*

Neste estudo foi utilizada a técnica do *Transfer Learning*, no qual são aproveitados os pesos de uma rede neural previamente treinada a partir de um *dataset* volumoso. A etapa de *Transfer Learning* é indicada em treinamentos com *datasets* menores, como é o caso do estudo publicado por Sarvamangala e Kulkarni⁽⁴⁾. A rede escolhida neste contexto foi a denominada VGG19⁽¹⁶⁾. Ela possui um ótimo desempenho, graças a sua grande profundidade, mas baixo custo computacional por utilizar filtros pequenos do tamanho 3x3⁽¹⁷⁾. A arquitetura da VGG19 pode ser vista na figura 3.

Para a etapa de *Transfer Learning* foram removidas as três camadas inteiramente conectadas e substituídas por

apenas uma com três neurônios de saída. Os pesos das camadas convolucionais foram congelados, por isso somente a última camada foi treinada. O método EarlyStopping⁽¹⁸⁾, implementado com a biblioteca Keras, foi utilizado para interromper o treinamento caso ocorresse aumento na taxa de erro para os dados de validação. Por fim, foi realizado o processo de *fine-tuning*⁽¹⁹⁾, no qual todos os pesos foram descongelados e a rede foi inteiramente treinada por dez épocas.

```
Model: "vgg19"
```

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 150, 150, 3)]	0
block1_conv1 (Conv2D)	(None, 150, 150, 64)	1792
block1_conv2 (Conv2D)	(None, 150, 150, 64)	36928
block1_pool (MaxPooling2D)	(None, 75, 75, 64)	0
block2_conv1 (Conv2D)	(None, 75, 75, 128)	73856
block2_conv2 (Conv2D)	(None, 75, 75, 128)	147584
block2_pool (MaxPooling2D)	(None, 37, 37, 128)	0
block3_conv1 (Conv2D)	(None, 37, 37, 256)	295168
block3_conv2 (Conv2D)	(None, 37, 37, 256)	590080
block3_conv3 (Conv2D)	(None, 37, 37, 256)	590080
block3_conv4 (Conv2D)	(None, 37, 37, 256)	590080
block3_pool (MaxPooling2D)	(None, 18, 18, 256)	0
block4_conv1 (Conv2D)	(None, 18, 18, 512)	1180160
block4_conv2 (Conv2D)	(None, 18, 18, 512)	2359808
block4_conv3 (Conv2D)	(None, 18, 18, 512)	2359808
block4_conv4 (Conv2D)	(None, 18, 18, 512)	2359808
block4_pool (MaxPooling2D)	(None, 9, 9, 512)	0
block5_conv1 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv2 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv3 (Conv2D)	(None, 9, 9, 512)	2359808
block5_conv4 (Conv2D)	(None, 9, 9, 512)	2359808
block5_pool (MaxPooling2D)	(None, 4, 4, 512)	0
Total params: 20,024,384		
Trainable params: 20,024,384		
Non-trainable params: 0		

Figura 3 - Arquitetura da CNN VGG19.

RESULTADOS E DISCUSSÃO

Resultados do Aprendizado Profundo

Para avaliação do modelo utilizou-se o método de validação cruzada (10-fold). Obteve-se os resultados apresentados na matriz de confusão (tabela 1). A acurácia obtida foi de 0.81. Embora este valor tenha sido obtido, a classe “doente” não apresentou bons indicadores. Pode-se observar na matriz de confusão (tabela 1) que ocorreram diversos erros do classificador.

Analisando-se as métricas de precisão, *recall* e *f1-score* (tabela 2) observa-se que a classe “com cerumen” apresentou os melhores resultados (*f1-score*). O modelo não apresentou, contudo, bons resultados para a classe “doente”.

Resultados do *Transfer Learning*

Para avaliação do modelo utilizou-se o método de

Tabela 1 - Matriz de confusão com aprendizado profundo

		Classe Prevista		
		cerumen	doente	normal
Classe Real	cerumen	311	25	15
	doente	42	113	52
	normal	15	12	243

Tabela 2 - Resultados obtidos com aprendizado profundo

	cerumen	doente	normal
precisão	0,85	0,75	0,78
recall	0,88	0,55	0,90
f1-score	0,86	0,63	0,84
support	351	207	270

Tabela 3 - Matriz de confusão com transfer learning

		Classe Prevista		
		cerumen	doente	normal
Classe Real	cerumen	364	5	3
	doente	4	338	24
	normal	5	11	530

Tabela 4 - Resultados obtidos com transfer learning

	cerumen	doente	normal
precisão	0.98	0.95	0.95
recall	0.98	0.92	0,97
f1-score	0.98	0.94	0.96
support	372	366	546

validação cruzada (10-fold). Obteve-se os resultados apresentados na matriz de confusão (tabela 3). A acurácia obtida foi de 0.96. Pode-se observar na matriz de confusão (tabela 3) que todas as classes obtiveram bons resultados do classificador.

Analisando-se as métricas de precisão, *recall* e *f1-score* (tabela 4) observa-se que a classe “com cerumen” apresentou os melhores resultados. Contudo, houve boa diferenciação entre as classes de imagens, o que produziu baixa quantidade de falsos positivos e falsos negativos.

CONCLUSÕES

Na clínica médica, diversos são os desafios enfrentados pelos profissionais. O correto diagnóstico de problemas frequentes, tais como a otite, é de extrema importância. O processamento da visão computacional, por meio das CNNs, tem se revelado um mecanismo importante, útil

e cujos resultados atestam a possibilidade de uso em casos médicos. Neste trabalho, reconhecer as características da membrana timpânica foi um desafio. Diferentemente dos problemas de classificação tratados pelas CNN, onde existem bordas evidentes, os tecidos estudados são homogêneos e translúcidos. Os contornos ou traços da membrana timpânica não são tão importantes para o diagnóstico de otite média. Já características como coloração, brilho e presença de secreção geralmente amarelada são de extrema importância. A capacidade da rede neural em extrair e reconhecer tais características foi fundamental para a correta classificação das imagens da membrana timpânica. Os resultados evidenciam que o modelo VGG19 é apropriado para tratar imagens médicas, convergindo com outros trabalhos da área^{2,3}. Como trabalhos futuros pretende-se ampliar o estudo e desenvolver novas arquiteturas para diagnóstico por imagens na área da saúde.

REFERÊNCIAS

1. Mitchell, T. M. Machine learning. [S.l.]: McGraw-Hill, 1997.
2. Goodfellow, I.; Bengio, Y.; Courville, A. Deep learning. [S.l.]: MIT Press, 2016.
3. Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., van der Laak, J.A.W.M, van Ginneken, B., Sánchez, C.I. A survey on deep learning in medical image analysis. *Medical Image Analysis*, [S.l.], v. 42, p. 60 – 88, 2017.
4. Sarvamangala D.R., Kulkarni R.V. Convolutional neural networks in medical image understanding: a survey. *Evol. Intelligence*. (2021). <https://doi.org/10.1007/s12065-020-00540-3>
5. Yadav S.S., Jadhav S.M. Deep convolutional neural network based medical image classification for disease diagnosis. *J Big Data* 6, 113 (2019). <https://doi.org/10.1186/s40537-019-0276-2>
6. Bailey, B.; Johnson, J. T. Head and neck surgery otolaryngology. [S.l.]: Lippincott Williams & Wilkins, 2006. v. 2.
7. Teele, D. W.; Klein, J. O.; Rosner, B. A. Epidemiology of otitis media in children. *The Annals of otology, rhinology & laryngology. Supplement.*, [S.l.], v. 89, p. 5–6, 1980.
8. Lundberg T. et al. Diagnostic accuracy of a general practitioner with video-otoscopy collected by a health care facilitator compared to traditional otoscopy. *International Journal of Pediatric Otorhinolaryngology*, [S.l.], v. 99, p. 49 – 53, 2017.
9. Kuruvilla, A.; Shaikh, N.; Kovacevic, J. Automated diagnosis of otitis media: vocabulary and grammar. *International Journal of Biomedical Imaging*, [S.l.], 2013.
10. Cellscope [homepage na internet]. Cellscope US Berkeley [acesso em 10 out 2021] Disponível em: <https://cellscope.berkeley.edu/>
11. AppAdvice [homepage na internet]. AppAdvice [acesso em 10 out 2021] Disponível em: <https://appadvice.com/app/cellscopelite/>
12. Krizhevsky, A.; Sutskever, I.; Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in Neural Information Processing Systems*, [S.l.], p. 1097 – 1105, 2012.
13. Keras Image Data Processing [homepage na internet]. API Keras Image Data Processing [acesso em 20 jan 2021] Disponível em: <https://keras.io/api/preprocessing/image/>
14. Python Imaging Library [homepage na internet]. API Python Imaging Library [acesso em 20 jan 2022] Disponível em: <https://python-pillow.org/>
15. VGG19 [homepage na internet]. Keras API reference VGG16 and VGG19 [acesso em 20 mar 2022] Disponível em: <https://keras.io/api/applications/vgg/>
16. Visual Geometry Group [homepage na internet]. VGG Homepage [acesso em 20 mar 2021] Disponível em: <https://www.robots.ox.ac.uk/~vgg/>
17. Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. In: *International Conference On Learning Representations*. San Diego, USA: 201
18. EarlyStopping [homepage na internet]. Keras API reference EarlyStopping [acesso em 20 mar 2021] Disponível em: https://keras.io/api/callbacks/early_stopping
19. Transfer learning & fine-tuning [homepage na internet]. Keras API Transfer learning & fine-tuning [acesso em 20 mar 2021] Disponível em: https://keras.io/guides/transfer_learning/