

Universidade de Brasília - UnB

Faculdade UnB Gama - FGA

Aplicação de técnicas de XAI em redes
neurais convolucionais na classificação de lesões de pele

Autor: João Vitor Rodrigues Baptista
Orientador: Dr. Nilton Correia da Silva



Sumário

1 | 70

- 1 - Apresentação
- 2 - Introdução
- 3 - Referencial teórico
- 6 - Resultados prévios
- 5 - Próximos passos

Introdução

2 | 70

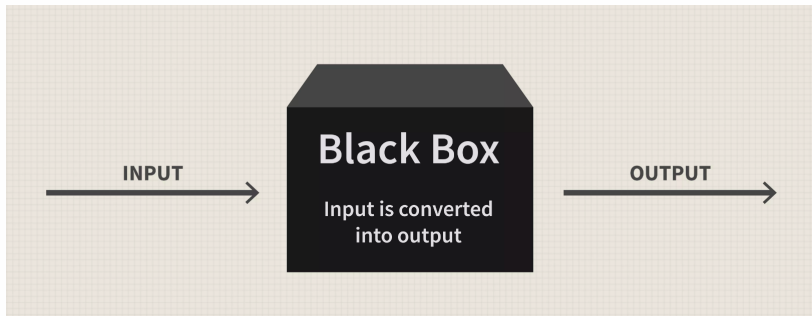


Figure: *Modelo "black box"*

(Julie Bang, Investopedia 2019)

- Processo interno de tomada de decisão do modelo.
- Natureza dos dados.
- O processo de tomada de decisão deve ser baseada em fatos concretos.
- Contribuição com insights.

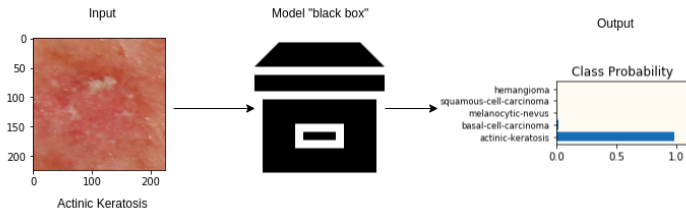


Figure: *Resultado subjetivo do modelo*

(Autor)

- Pequenas características podem contribuir para diferentes resultados.
- Auxiliar médicos a entender quais características o modelo analisou para chegar a conclusão.
- Mostrar novos padrões não usados para identificação das lesões.(insights)

Referencial teórico

6 | 70

Nessa parte será apresentado os processos e técnicas utilizados para a implementação do modelo de classificar lesões de pele.

- Modelo de rede neural convolucional.
- Arquitetura de rede neural artificial.
- Métricas de treino e validação.
- Métricas para teste.
- Base de dados.
- Preparação da base de dados.

Modelo de rede neural convolucional 7 | 70

Essa modelo tem alcançado resultados incríveis no reconhecimento e classificação de imagens.

HUBEL; WIESEL, 1968

Baseada no modo que as conexões dos neurônios de animais estão dispostas de forma dispersa. Portanto, essa característica permite que a rede neural artificial tenha maior capacidade de reconhecer características individuais.

$$S(i,j) = (I * K)(i,j) = \sum_m \sum_n I(i + m, j + n) K(m, n) \quad (1)$$

Arquitetura de rede neural artificial

8 | 70

Como as redes neurais convolucionais são algoritmos muito populares existe uma grande diversidade de pesquisas relacionadas. Contudo, é necessário usar algum tipo de topologia para esse algoritmo. A topologia leve em conta como estão dispostas as camadas da rede.

RUSSAKOVSKY et al., 2014

Foi uma topologia proposta por pesquisadores da Microsoft em 2015, onde ganhou o desafio do ImageNet.

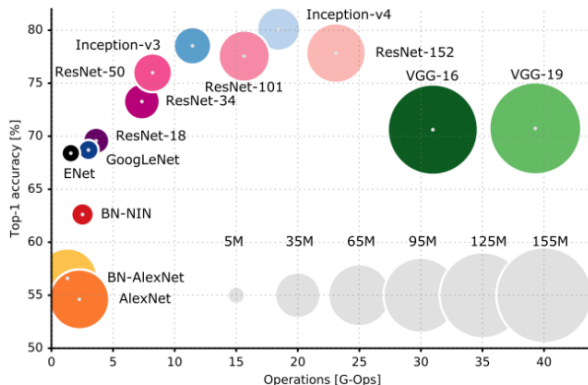


Figure: Comparativo entre as topologias pela acurácia em função da operação em função do tamanho da base de dados

(CANZIANI; PASZKE; CULURCIELLO, 2017)

Métricas de treino e validação

10 | 70

Como foi utilizado treino e validação, as métricas utilizadas foi principalmente a acurácia e a função de perda Cross-Entropy.

$$\textit{Acurácia} = \frac{\sum t_p + \sum t_n}{s} \quad (2)$$

Métricas para teste

11 | 70

Apos o treinamento do modelo, é feita uma etapa final que visa testar o modelo com dados que não foram utilizados no momento do teste e da validação.

$$\textbf{Precisão} = \frac{t_p}{t_p + f_p} \quad (3)$$

$$\textbf{Recall} = \frac{t_p}{t_p + f_n} \quad (4)$$

SWETS,1986

A curva de ROC mapeia a sensitivity (probabilidade de detecção) em função da $1 - \text{specificity}$ (probabilidade de falsos positivos). Tipicamente, essa métrica é implementada em sistemas que a analisa o quão exato é o diagnóstico do estado de um paciente em termos de doenças.

Base de dados

13 | 70

- Dificuldade de encontrar bases completas.
- **Restrição:** Devem ser imagens clínicas.
- Três bases.
- Base criada para complementar.

GIOTIS et al., 2015

A primeira base de dados utilizado foi fornecida pelo Department of Dermatology at the University Medical Center Groningen (UMCG)

Table: *Numero de amostras da base do MED-NODE*

Tipo de lesão	Amostras
Melanoma	70
Melanocytic Nevus	100
Total	170

DERMOFIT IMAGE LIBRARY

Essa é uma base de dados disponível para ser comprada mediante a uma licença de utilização ^a. É a base mais completa encontrada na internet.

São imagens de diagnósticos baseada na avaliação de profissionais coletadas em condições padronizadas.

^aDisponível em:<https://licensing.eri.ed.ac.uk/i/software/dermofit-image-library.html>

Table: *Numero de amostras da base do Edinburgh*

Tipo de lesão	Amostras
Actinic Keratosis	45
Basal Cell Carcinoma	239
Melanocytic Nevus	331
Seborrhoeic Keratosis	257
Squamous Cell Carcinoma	88
Intraepithelial Carcinoma	78
Pyogenic Granuloma	24
Haemangioma	97
Dermatofibroma	65
Malignant Melanoma	76
Total	1300

International Skin Imaging Collaboration (ISIC)

É um projeto que possui um repositório com acesso a imagens clínicas de lesões. Esse projeto é uma parceria entre a indústria e as universidades com o objetivo de reduzir mortes por câncer e biópsias desnecessárias através do avanço nos campos computacionais. Esse repositório de imagens é organizado, imagens possuem *meta-data* que contem o diagnostico, idade aproximada do paciente e o gênero. Em trabalhos futuros, esse tipo de dado pode ser utilizado para a criação de modelos mais sofisticados. A Tabela 3 mostra os tipos de lesões e a quantidade de amostras.

Table: *Numero de amostras da base do ISIC*

Tipo de lesão	Amostras
Actinic Keratosis	132
Basal Cell Carcinoma	480
Seborrhoeic Keratosis	339
Squamous Cell Carcinoma	226
Dermatofibroma	122
Total	1299

Base suplementar

Essa base é um conjunto de imagens com o objetivo de suplementar as bases mais padronizadas. Foi utilizado *scripts* de extração de imagens em diferentes sites dermatológicos usados peças referências utilizadas nesse trabalho. ^a

^aSites utilizados para a base: <http://www.dermatlas.net/>,
<http://www.dermatlas.net/>,
<http://www.dermis.net/dermisroot/en/home/index.htm>,
<http://www.meddean.luc.edu/lumen/MedEd/medicine/dermatology/melton/atlas.htm>, <http://www.dermatoweb.net/>,
<http://www.atlasdermatologico.com.br/>,
<http://www.danderm-pdv.is.kkh.dk/atlas/index.html>,
<http://www.hellenicdermatlas.com/en/?params=en>.

Table: *Numero de amostras da base do Atlas*

Tipo de lesão	Amostras
Actinic Keratosis	24
Basal Cell Carcinoma	151
Melanocytic Nevus	71
Seborrhoeic Keratosis	107
Squamous Cell Carcinoma	103
Intraepithelial Carcinoma	81
Pyogenic Granuloma	74
Haemangioma	76
Dermatofibroma	19
Malignant Melanoma	65
Total	771

Lesões de interesse

21 | 70

As lesões de interesse podem ser divididas em dois grandes grupos, o primeiro é composto de lesões que efetivamente são tipos de câncer e são extremamente perigosos para a saúde do paciente, por outro lado, o segundo grupo é composto de lesões que não oferecem riscos mais imediatos a o paciência.

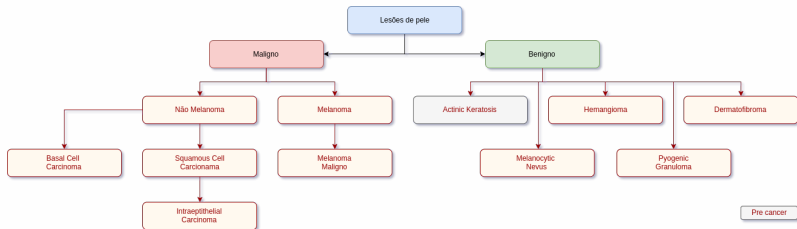


Figure: *Diagrama de lesões de interesse*

(Autor)

Natureza da base de dados

23 | 70

O campo de imagens medicas trás desafios enraizados na natureza das imagens. Por essa razão, é necessário avaliar para traçar as melhores estratégias para trabalhar com essas imagens. Outra desafio é a diferença entre as bases de dados usado no presente trabalho, isso adiciona mais variabilidade para a base e pode acarretar em modelos que não conseguem convergir.

- **Etnia.** (HAN et al., 2018)
- **Idade.** (HAN et al., 2018)
- **Diferentes equipamentos para coleta das imagens.**
- **Posição.**
- **Cabelos.**
- **Tipos de pele.** (elasticidade e a reflexividade) (CUA; WILHELM; MAIBACH, 1990).

Preparação da base de dados

25 | 70

Para o treinamento de modelos de redes neurais artificiais é necessário uma grande quantidade de amostras, contudo devido a falta de amostras de imagens medicas foi necessário a utilização de algumas técnicas para contornar esse problema.

Outro problema encontrado é a criação um novo modelo convolucional para ser usado com a base. Para abordar esse problema, foi utilizado uma técnica que parte do principio que existe um modelo que performa muito bem para problemas de classificação e que a partir desse modelo pode-se construir novos conceitos.

Transfer Learning

Essa estratégia utiliza o modelo pre treinado, porem é alterado a ultima camada da rede, geralmente a camada fully-connected, para o numero de classes do nova base de dados e então atualiza os parâmetros via backpropagation. Normalmente é mais demorado porem resulta em melhores métricas de teste.

Data augmentation

Essa técnica é usada quando não possui uma quantidade infinita de amostras para treinar o modelo. Isso pode ser feito criando dados sintéticos que estão relacionado com as amostras originais da base de dados através de aplicação de transformações.

Transformações

27 | 70

Cada escolha de transformação aplicada é baseada no guia de data augmentation (PEREZ; WANG, 2017; CUBUK et al., 2018) ou na natureza das amostras. As transformações foram organizadas em forma de pipeline para que cada transformação tenha uma probabilidade de ser aplicada na imagem e então salvar a amostra no destino especificado. A Tabela 5 mostra as transformações e as probabilidade do pipeline.

Table: *Transformações aplicadas no processo de data augmentation*

Transformações	Probabilidades
Rotação	0.5
Zoom	0.4
Flip Horizontal	0.7
Flip Vertical	0.5
Distorção	0.8
Variância na Luminosidade	0.5

Pipeline

29 | 70

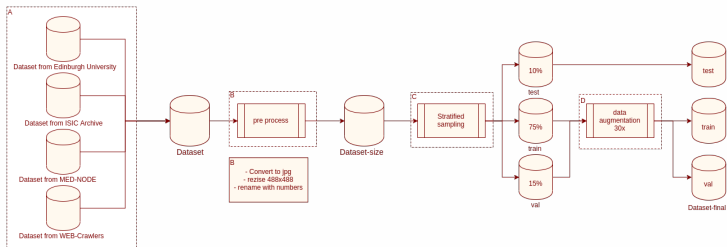


Figure: *Diagrama de processos utilizados para a preparação da base de dados*

(Autor)

Table: *Numero de amostras da base agregada da parte A*

Tipo de lesão	Amostras
Actinic Keratosis	185
Basal Cell Carcinoma	832
Melanocytic Nevus	502
Seborrhoeic Keratosis	107
Squamous Cell Carcinoma	417
Intraepithelial Carcinoma	148
Pyogenic Granuloma	98
Haemangioma	173
Dermatofibroma	206
Malignant Melanoma	687
Total	3355

Parte B

31 | 70

Em seguida foi feito um pré processamento nas imagens para converter todas para o mesmo formato, renomear de forma crescente e redimensionar as imagens pois nem todas eram do mesmo tamanho, foi escolhido um tamanho de 448x448 que é o dobro da dimensão das imagens usadas no modelo.

Cuidado !

Um cuidado especial foi feito nessa etapa, pois para não distorcer as imagens que não são quadradas foi feito um complemento de pixels tanto horizontal quanto vertical.

Por quê ?

Na parte de Data augmentation ou problemas de limitações computacionais e de lidar com imagens de diferentes formatos.

Parte C

32 | 70

Em seguida como foi feito a separação da base de dados em três sub bases, para treinamento, validação e teste. Nessa etapa foi utilizado uma biblioteca do python onde esta implementado a separação de imagens de forma estratificada para não misturar amostras entre as sub bases. A proporção escolhida da seguinte forma 75%, 15% e 10%, para respectivamente, treinamento, validação e teste.

Cuidado !

Deve-se certificar que a separação é feita de forma estratificada para garantir que o modelo terá acesso as sub bases no momento correto.

Parte D

33 | 70

Com as bases separadas foi aplicada a Augmentation na base de treinamento e validação usando as transformações citadas na tabela 5 com um fator de multiplicação de 30x.

Cuidado !

Não foi feito Data Augmentation na base de teste, pois adicionaria um custo computacional maior na fase de teste dos experimentos.

Table: *Numero de amostras finais*

Tipo de lesão	treino	validação	teste
Actinic Keratosis	4278	837	20
Basal Cell Carcinoma	18720	3844	84
Melanocytic Nevus	11656	2325	51
Squamous Cell Carcinoma	9672	1922	43
Intraepithelial Carcinoma	3441	682	15
Pyogenic Granuloma	2263	434	11
Haemangioma	3999	775	19
Dermatofibroma	4774	930	22
Malignant Melanoma	15965	3193	69
Total	74768	14942	334

Resultados prévios

35 | 70

Esse espaço esta reservado discussão dos resultados prévios obtidos através das técnicas e referencias teóricas apresentadas.

- Infraestrutura.
- Classificação.
- Resultados.

Infraestrutura

36 | 70

Todos os experimentos quem envolveram treinamento, validação e teste foram realizados na plataforma *open source* da Google que é uma variação do serviço *datalab* que fornece uma infraestrutura que atende as necessidades de processamento do presente trabalho.

Com Linux version 4.14.137+ (chrome-bot@chromeos-legacy-release-us-central1-b-x32-44-v3dn) (gcc version 4.9.x 20150123 (prerelease) rodando o framework PyTorch - 1.3.1 e TorchVision - 0.4.2.

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Resultados prévios

37 | 70

```
[name: "/device:CPU:0"  
device_type: "CPU"  
memory_limit: 268435456  
locality { }  
  
incarnation: 10575003407439260273,  
name: "  
device:XLA_CPU:0"  
device_type: "XLA_CPU"  
memory_limit: 17179869184  
locality { }  
  
incarnation: 12607010922076994446  
physical_device_desc: "  
device: XLA_CPU device",  
name: "  
device:XLA_GPU:0"  
device_type: "XLA_GPU"  
memory_limit: 17179869184  
locality { }  
  
incarnation: 15083069251257245813  
physical_device_desc: "  
device: XLA_GPU device",  
name: "  
device:GPU:0"  
device_type: "GPU"  
memory_limit: 15956161332  
locality { bus_id: 1 links { } }  
incarnation: 7574273019028501897  
physical_device_desc: "  
device: 0,  
name: Tesla P100-PCIE-16GB, pci bus id: 0000:00:04.0, compute capability: 6.0"]
```

Figure: *Detalhes técnicos da infraestrutura*

(Autor)

Classificação

38 | 70

Existem muitas topologias para aplicar a técnica de *transfer learning*, porem como mencionado em seções anteriores a escolha foi a ResNet-152.

Processo de treinamento

Foram conduzido 8 experimentos com diferentes combinações de parâmetros livres com o objetivo de conseguir o melhor modelo possível. Cada experimento teve um tempo de duração de aproximadamente 12 horas ininterruptas o que resultou em aproximadamente 10 - 14 epochs.

Resultados

39 | 70

A acurácia total do modelo foi de 78.44% para a base de teste. Contudo, a acuraria não é a única métrica para avaliar o modelo, pois possuem o viés da base não esta balanceada

Matriz de confusão

A lesão Basal Cell Carcinoma é confundida com Squamous Cell Carcinoma, bem como a dificuldade de predizer Intraepithelial Cell Carcinoma com relação ao Basal Cell Carcinoma. Essas predições são de se esperar visto a natureza maligna da lesão e a semelhança superficial.

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Resultados prévios

40 | 70

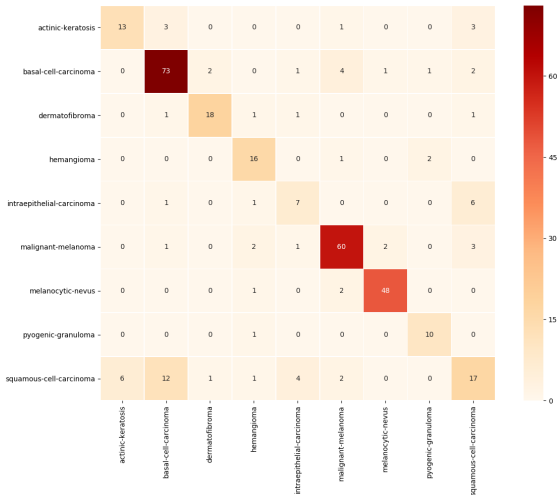


Figure: Matriz de confusão do modelo para as 9 lesões

Report de métricas

Além disso, foi calculado um reporte de classificação que fornece o recall, precisão e F1 score para as classes individuais assim como as medias.

ROC curve

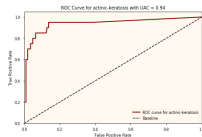
Por ultimo foi plotado a ROC curve para cada lesão e sua respectiva UAC as curvas. Essa métrica é uma das mais populares para a avaliação de modelos de machine learning.

Table: *Reporte das métricas para cada lesão*

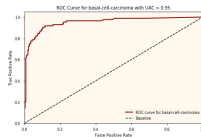
Tipo de lesão	Precisão	Recall	F1 Score	Support
Actinic Keratosis	0.68	0.65	0.67	20
Basal Cell Carcinoma	0.80	0.87	0.83	84
Melanocytic Nevus	0.94	0.94	0.94	51
Squamous Cell Carcinoma	0.53	0.40	0.45	43
Intraepithelial Carcinoma	0.50	0.47	0.48	15
Pyogenic Granuloma	0.77	0.91	0.83	11
Haemangioma	0.70	0.84	0.76	19
Dermatofibroma	0.86	0.82	0.84	22
Malignant Melanoma	0.86	0.87	0.86	69
Media/total	0.78	0.78	0.78	334

Universidade de Brasília - UnB
Faculdade UnB Gama - FGA
Resultados prévios

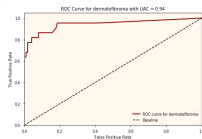
43 | 70



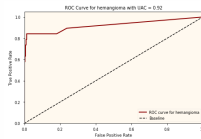
(a) *Curva ROC para Actinic Keratosis*



(b) *Curva ROC para Basal Cell Carcinoma*



(c) *Curva ROC para Dermatofibroma*

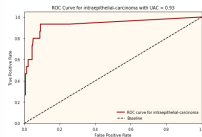


(d) *Curva ROC para Hemangioma*

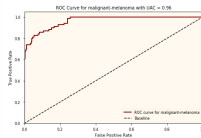
Figure: *Curva ROC para o restante das lesões. Continuação 1/3.*

Universidade de Brasília – UnB
Faculdade UnB Gama – FGA
Resultados prévios

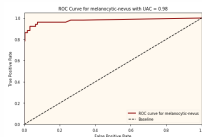
44 | 70



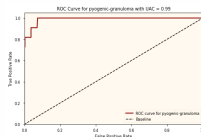
(a) *Curva ROC para Intrapithelial Carcinoma*



(b) *Curva ROC para Malignant Melanoma*



(c) *Curva ROC para Melanocytic nevus*



(d) *Curva ROC para Pyogenic Granuloma*

Figure: *Curva ROC para o restante das lesões. Continuação 2/3.*

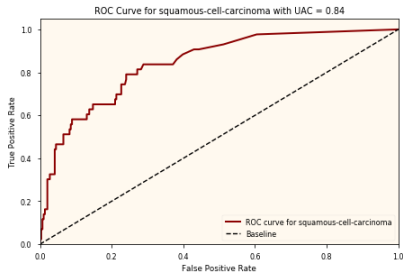


Figure: *Curva ROC para Squamous Carcinoma*

Figure: *Curva ROC para o restante das lesões. Continuação 3/3.*

(Autor)

Table: *Comparativo entre trabalhos correlatos as AUC*

Tipo de lesão	(ESTEVA et al., 2017)	(HAN et al., 2018)	(MENDES; SILVA, 2018)	Atual
Actinic Keratosis	-	0.83	0.96	0.94
Basal Cell Carcinoma	-	0.90	0.91	0.95
Melanocytic Nevus	-	0.94	0.95	0.98
Squamous Cell Carcinoma	-	0.91	0.95	0.84
Intraepithelial Carcinoma	-	0.83	0.99	0.93
Pyogenic Granuloma	-	0.97	0.99	0.99
Haemangioma	-	0.83	0.99	0.92
Dermatofibroma	-	0.90	0.90	0.94
Malignant Melanoma	0.96	0.88	0.96	0.96

Próximos passos

47 | 70

Essa seção esta reservada para discutir os próximos passos de implementação e pesquisa deste trabalho. Com aplicação de técnicas de XAI.

- Interpretabilidade e Explicabilidade.
- Métodos de interpretabilidade.
- Cronograma de atividades.
- Discursão de possíveis resultados.

Interpretabilidade e Explicabilidade

48 | 70

- Necessidade e Impacto da explicação.
- Explicações produzidas pode não ser consoante com o estilo de decisões humanas.

Essa ideia não é nova.

A busca das respostas do "Por quê ?" não é recente, pela literatura existem pesquisas datada da década de 80 (CLANCEY; SHORT-LIFFE, 1984; CLANCEY, 1981; CHANDRASEKARAN; TANNER; JOSEPHSON, 1989).

Conceitos

49 | 70

BIRAN; COTTON, 2017

Define que um sistema é interpretável se o ser humano pode entender suas operações, tanto por inspeção ou uma explicação produzida pelo modelo.

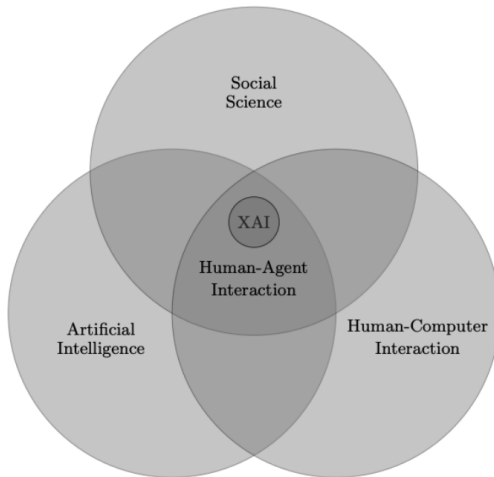


Figure: *Escopo da área XAI.* (MILLER, 2017)

Métodos de interpretabilidade

51 | 70

- Métodos de avaliação locais e globais.

Avaliações locais

As predições únicas levam em consideração uma única entrada e explica quais foram os fatores que levaram o modelo a determinada decisão baseado na entrada.

Avaliações globais

Avaliações globais são mais complexas pelo fato de ser necessário conhecer o modelo como um todo de uma única vez (LIPTON, 2016).

Método Model-agnostic

52 | 70

- **Flexibilidade do modelo:** O método de interpretação pode funcionar com diversos modelos, desde *random forest* até redes neurais profundas.
- **Flexibilidade da explicação:** Não possui limitação quanto ao modo de explicação. Tem situações que é necessário uma formula linear outros casos gráficos com características importantes.
- **Flexibilidade da representação:** O sistema de explicação deve ser capaz de usa diferentes características de representação em quanto o modelo esta sendo explicado.

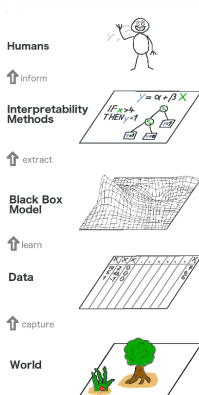


Figure: *O panorama geral de interpretabilidade para modelos de machine learning (MOLNAR, 2019)*

Métodos de input-perturbations

Usa entradas com pequenas perturbações para testar se a decisão final do modelo muda em relação a entrada sem perturbação

Métodos baseados em cálculos do gradiente

A abordagem baseada em nos gradientes usa o calculo dos gradientes de saída do neurônio em relação a entrada

Perturbation-based

55 | 70

A abordagem adiciona uma perturbação na entrada e avalia as consequências dessa perturbação na decisão final do modelo. Essa perturbação consiste na remoção de pequenas porções específicas de informação da entrada aplicando ruído. Comparado com a abordagem baseada na computação dos gradientes esse método é computacionalmente mais custoso.

Outro problema enfrentado nessa abordagem é a dificuldade de escolha da perturbação que era aplicada na amostra. Pois a intenção não é mudar a natureza da amostra portanto a remoção de pedaços da imagem é feito apenas substituindo regiões de interesse por pixels cinza.

LIME

56 | 70

Ou Local Interpretable Model-agnostic Explanations é uma metodologia apresentada por (RIBEIRO; SINGH; GUESTRIN, 2016) que implementa um modelo local que explica previsões individuais. LIME usa o conceito de perturbação apresentado na subseção anterior

Outro problema enfrentado nessa abordagem é a dificuldade de escolha da perturbação que era aplicada na amostra. Pois a intenção não é mudar a natureza da amostra portanto a remoção de pedaços da imagem é feito apenas substituindo regiões de interesse por pixels cinza.

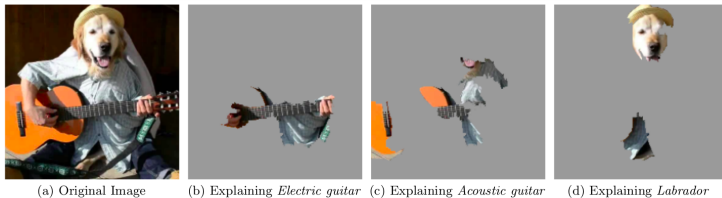


Figure: *Explicação usando LIME para as top 3 classes na classificação de uma imagem feita pelo modelo da Google Inception.*(RIBEIRO; SINGH; GUESTLIN, 2016)

Gradient-based

58 | 70

É a abordagem mais popular dos métodos de explicação local para classificação de imagens (ERHAN et al., 2009; SMILKOV et al., 2017; SUNDARARAJAN; TALY; YAN, 2017). Essa metodologia é baseada na importância de ativação de cada pixel da imagem de entrada em relação a predição de saída.

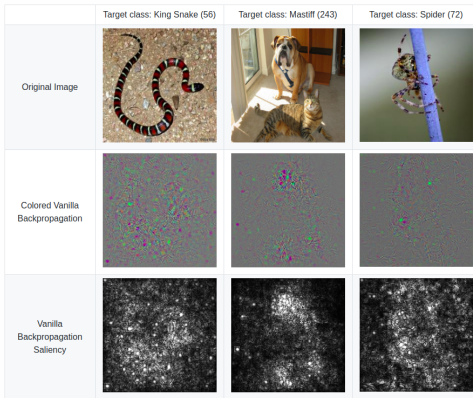


Figure: *Exemplos de visualização do gradiente das imagens (OZBULAK, 2019)*

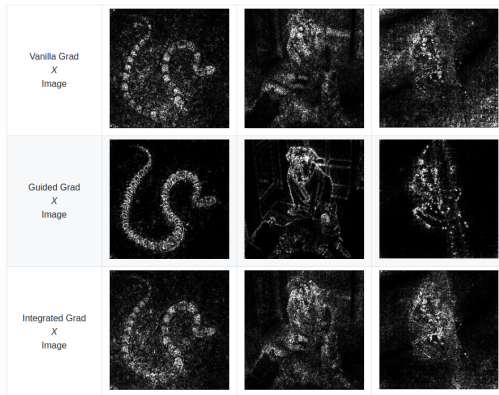


Figure: *Multiplicação da imagem de entrada com o gradiente de ativação respectivo (OZBULAK, 2019)*

GradCAM

61 | 70

Abreviação para *Gradient-weighted Class Activation Mapping* proposta por (SELVARAJU et al., 2016). Com essa abordagem é possível localizar as principais áreas de interesse da cada classe no modelo. A ideia geral dessa abordagem é tentar direcionar os mapas de ativação da última camada para inferir a relevância dos pixels. E o resultado final é parecido com o mapa de calor

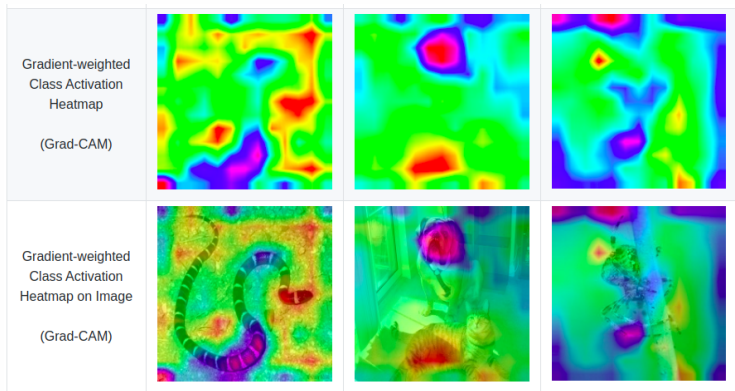


Figure: *Exemplo de mapa de calor usando a técnica do GradCAM (OZBULAK, 2019)*

Proposta e cronograma de atividades 63 | 70

Table: *Cronograma de atividades para o desenvolvimento das próximas etapas*

Atividades	Meses					
	Fev	Mar	Abr	Mai	Jun	Jul
Revisão bibliográfica e embasamento teórico sobre o LIME	X	X				
Implementação e Avaliação no modelo proposto usando LIME		X				
Escrita da seção relacionada ao LIME no TCC	X	X				
Revisão bibliográfica e embasamento teórico sobre Gradient-based			X			
Implementação e Avaliação no modelo proposto usando Gradient-based			X	X	X	
Escrita da seção relacionada ao Gradient-based no TCC			X	X	X	
Finalização da escrita do TCC						X
Ajustes finais no texto e correções do TCC						X

Resultados esperados

64 | 70

O fundamento principal de aplicação de XAI é ter uma visão melhor de como são feitas as decisões do modelo. Quais são as principais características de determinísticas de cada classe.

Para tanto, as lesões que tiveram as melhores métricas no reporte provavelmente serão as classes que irão produzir os resultados mais interpretáveis.

Portanto acredita-se que as lesões Basal Cell Carcinoma, Malignant Melanoma e Dermatofibroma produzirão os resultados mais expressivos desse trabalho, contudo será avaliado o possível *overfitting* da classe Melanocytic Nevus pois a métrica está muito alta em relação às outras classes. O restante das classes serão utilizados para verificar possíveis *underfitting* das decisões.

Obrigado

Referências

66 | 70

- HUBEL, D. H.; WIESEL, T. N. Receptive fields and functional architecture of monkey striate cortex. *The Journal of physiology*, v. 195 1, p. 215–43, 1968.
- RUSSAKOVSKY, O. et al. Imagenet large scale visual recognition challenge. *International Journal of Computer Vision*, v. 115, p. 211–252, 2014.
- CANZIANI, A.; PASZKE, A.; CULURCIELLO, E. An analysis of deep neural network models for practical applications. *ArXiv*, abs/1605.07678, 2017.
- SWETS, J. A. Indices of discrimination or diagnostic accuracy: their rocs and implied models. *Psychological bulletin*, v. 99 1, p. 100–17, 1986.

Referências

67 | 70

- GIOTIS, I. et al. Med-node: A computer-assisted melanoma diagnosis system using non-dermoscopic images. *Expert Syst. Appl.*, v. 42, p. 6578–6585, 2015.
- HAN, S. S. et al. Classification of the clinical images for benign and malignant cutaneous tumors using a deep learning algorithm. *The Journal of investigative dermatology*, v. 138 7, p. 1529–1538, 2018.
- CUA, A. B.; WILHELM, K.; MAIBACH, H. I. Elastic properties of human skin: relation to age, sex, and anatomical region. *Archives of Dermatological Research*, v. 282, p. 283–288, 1990.
- CUBUK, E. D. et al. Autoaugment: Learning augmentation policies from data. *ArXiv*, abs/1805.09501, 2018.

Referências

68 | 70

- CLANCEY, W. J.; SHORTLIFFE, E. H. Readings in medical artificial intelligence: the first decade. In: . [S.l.: s.n.], 1984.
- CLANCEY, W. J. The epistemology of a rule-based expert system – a framework for explanation. Artif. Intell., v. 20, p. 215–251, 1981.
- CHANDRASEKARAN, B.; TANNER, M. C.; JOSEPHSON, J. R. Explaining control strategies in problem solving. IEEE Expert, v. 4, p. 9–15, 1989.
- BIRAN, O.; COTTON, C. V. Explanation and justification in machine learning : A survey or. In: . [S.l.: s.n.], 2017.

Referências

69 | 70

- MILLER, T. Explanation in artificial intelligence: Insights from the social sciences. *Artif. Intell.*, v. 267, p. 1–38, 2017.
- MOLNAR, C. Interpretable Machine Learning: A guide for making black box models explainable. [S.l.: s.n.], 2019.
[jhttps://christophm.github.io/interpretable-ml-book/](https://christophm.github.io/interpretable-ml-book/)
- RIBEIRO, M. T.; SINGH, S.; GUESTRIN, C. Model-agnostic interpretability of machine learning. *ArXiv*, abs/1606.05386, 2016.
- ERHAN, D. et al. Visualizing higher-layer features of a deep network. In: . [S.l.: s.n.], 2009.

Referências

70 | 70

- SMILKOV, D. et al. Smoothgrad: removing noise by adding noise. ArXiv, abs/1706.03825, 2017.
- SUNDARARAJAN, M.; TALY, A.; YAN, Q. Axiomatic attribution for deep networks. In: ICML. [S.l.: s.n.], 2017.
- OZBULAK, U. PyTorch CNN Visualizations. [S.l.]: GitHub, 2019. [jhttps://github.com/utkuozbulak/pytorch-cnn-visualizations](https://github.com/utkuozbulak/pytorch-cnn-visualizations).
- SELVARAJU, R. R. et al. Grad-cam: Why did you say that? visual explanations from deep networks via gradient-based localization. ArXiv, abs/1610.02391, 2016.