

# 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



# Agenda

- Introdução
- Problema de pesquisa
- Objetivos
- Metodologia
- Resultados
- Considerações finais

# Introdução

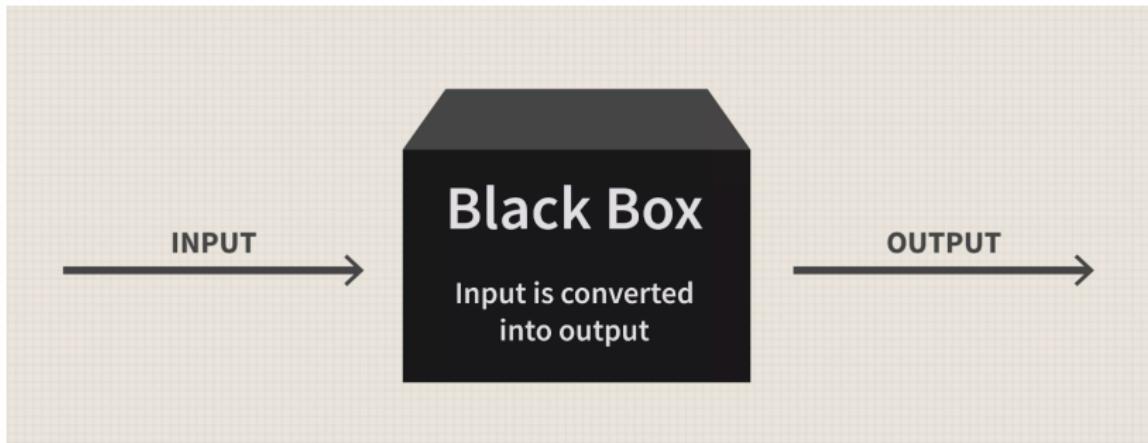


Figure: *Modelo "black box"*

(Julie Bang, Investopedia 2019)

## Problema de pesquisa

Como a utilização de diferentes técnicas de XAI fornecem insumos para a interpretabilidade de uma rede neural convolucional aplicada na classificação de lesões de pele.

# Objetivos

- **Objetivo geral:**  
Implementar, comparar e discutir diferentes técnicas de XAI em uma rede neural convolucional aplicada na classificação de lesões de pele.
- **Objetivos específicos:**
  1. Modelar uma rede neural convolucional para classificar lesões de pele.
  2. Aplicar duas técnicas de XAI no modelo desenvolvido.
  3. Discutir e comparar as técnicas e resultados das técnicas aplicadas.

# Metodologia

- Dados
- Modelo
- Treino
- Métricas
- Interpretabilidade

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Dados

# Dados

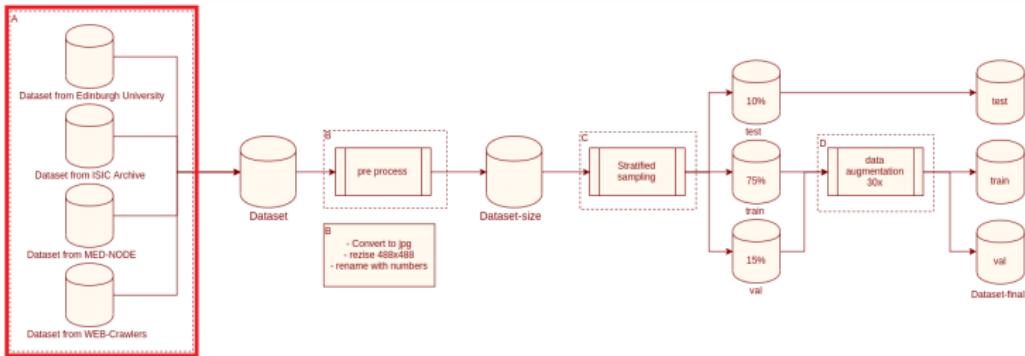


Figure: Pipeline de tratamento dos dados destacando a fase de aquisição das 4 bases usadas.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Dados

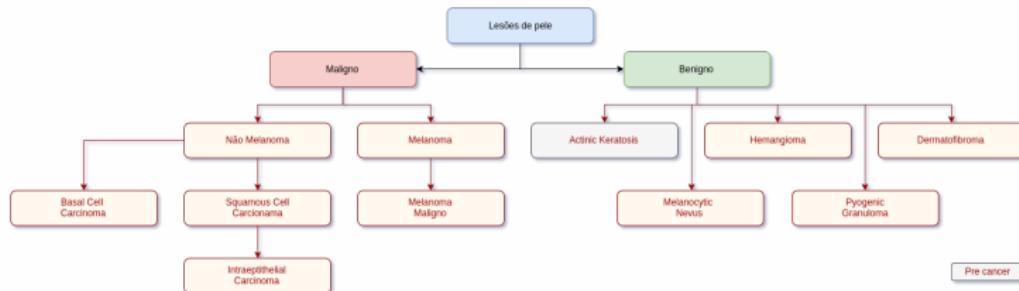


Figure: *Lesões de interesse adquiridas nas bases usadas*

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Dados

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

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

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Dados

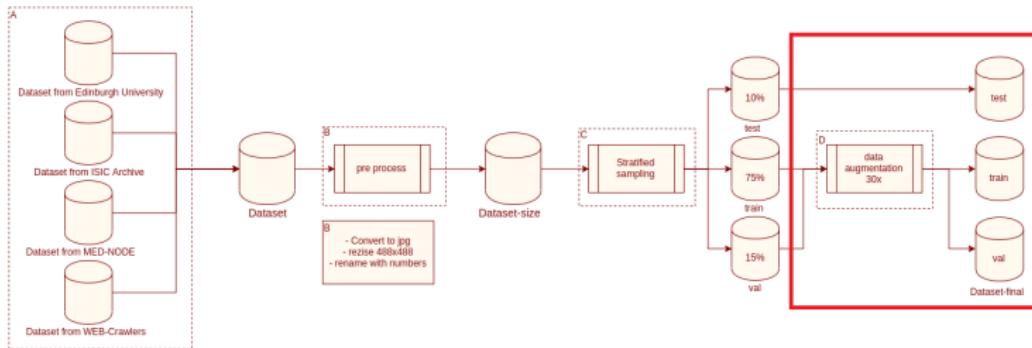


Figure: *Pipeline de tratamento dos dados destacando a fase Data Augmentation*

Autoria própria

## *Data Augmentation (PEREZ; WANG, 2017; CUBUK et al., 2018)*

Essa técnica é usada quando não possui uma quantidade grande de amostras para treinar o modelo. Desse modo, é feita criando dados sintéticos através da aplicação de transformações que visão, principalmente, não alterar as características da amostra base.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Dados

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

## Modelo

### RUSSAKOVSKY et al., 2014

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

### *Transfer Learning*

Essa estratégia utiliza o modelo pre treinado, porém é alterado a última camada da rede, geralmente a camada *fully-connected*, para o número de classes da nova base de dados e então atualiza os parâmetros via *backpropagation*. Dessa forma o treinamento aproveita os parâmetros já ajustados para extração de características.

# Treino

## Processo de treinamento

Foram conduzido 8 experimentos com diferentes combinações de parâmetros livres com a finalidade alcançar o modelo as melhores métricas. Cada experimento teve um tempo de duração de aproximadamente 12 horas ininterruptas o que resultou em aproximadamente 10 - 14 *epochs*.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Métricas

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

### Metodologia: Métricas

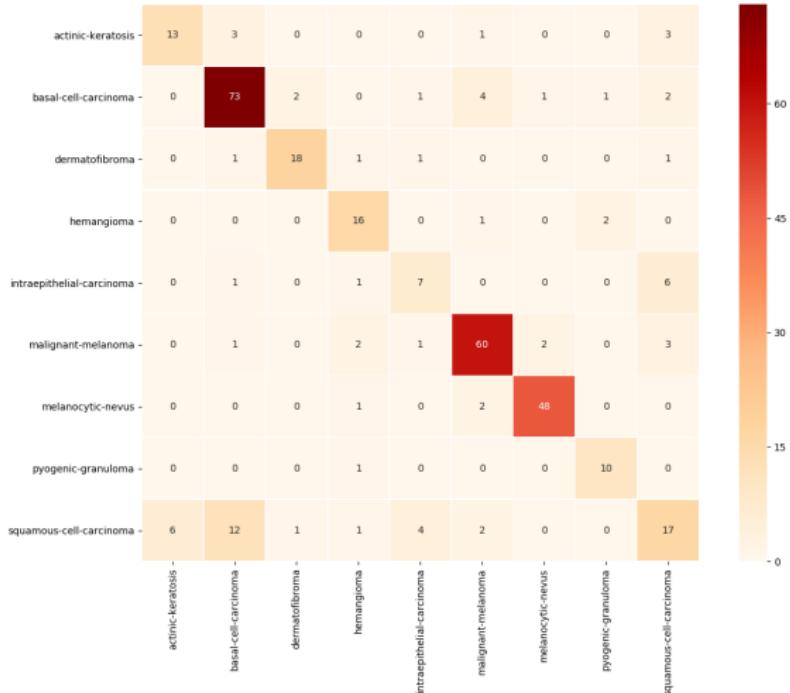


Figure: Matriz de confusão do modelo para as 9 lesões alvo

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Métricas

Table: *Comparativo da AUC entre os trabalhos referências e o presente trabalho.*

<b>Tipo de lesão</b>	(ESTEVA et al., 2017)	(HAN et al., 2018)	(MENDES; SILVA, 2018)	<b>Atual</b>
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

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Interpretabilidade

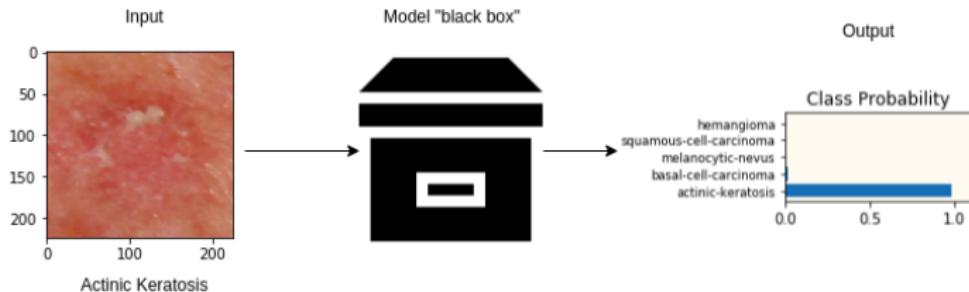


Figure: *Resultado subjetivo do modelo*

Autoria própria

## Interpretabilidade

**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).

**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.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Interpretabilidade

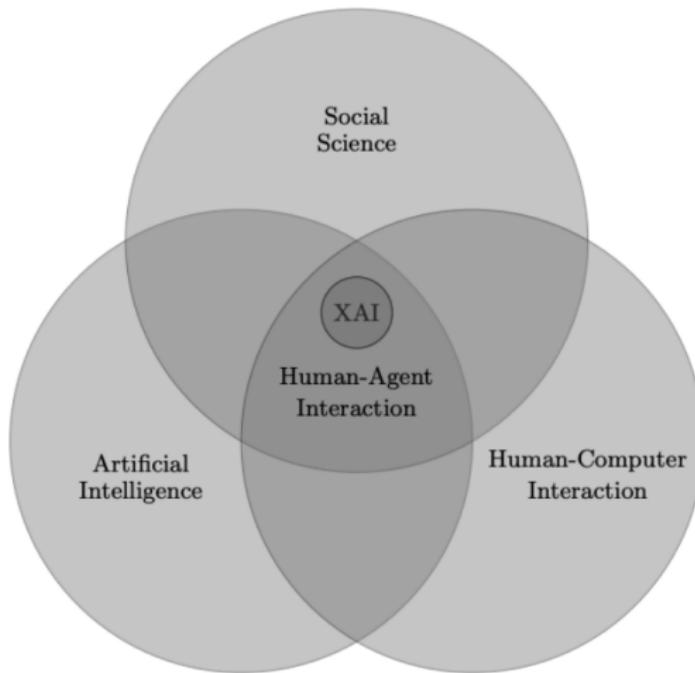


Figure: *Escopo da área XAI. (MILLER, 2017)*

## Escopo

### 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).

## Métodos de *perturbation-based*

Usa entradas com pequenas perturbações para testar se a decisão final do modelo muda em relação a entrada sem perturação, exemplos desses métodos; SHAP, Shapley Values e o LIME.

## Métodos baseados em cálculos do gradiente

Essa metodologia é baseada na importância de ativação de cada pixel da imagem de entrada em relação a predição de saída. Esses métodos são bons para explicar amostras únicas porém não performa tão bom para obter entendimento geral da classe observada.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Interpretabilidade

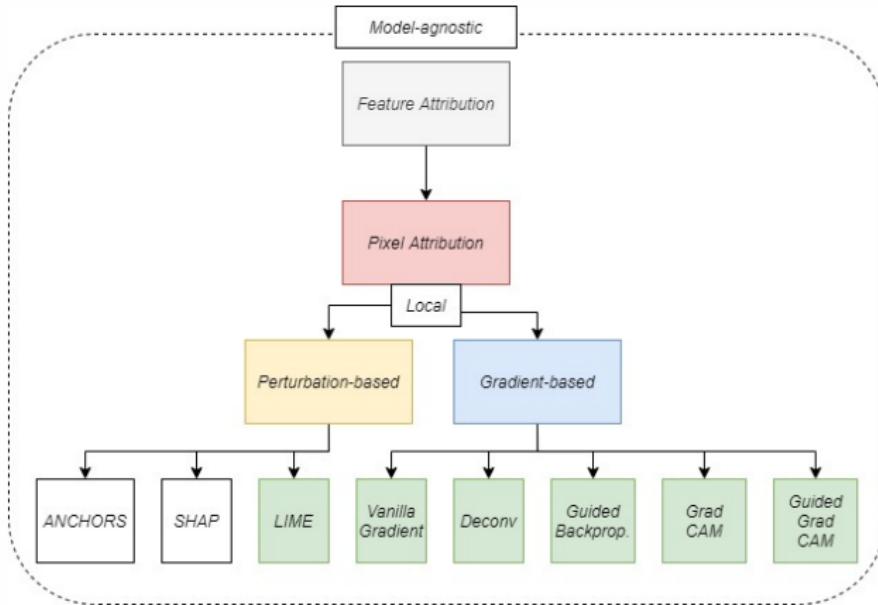


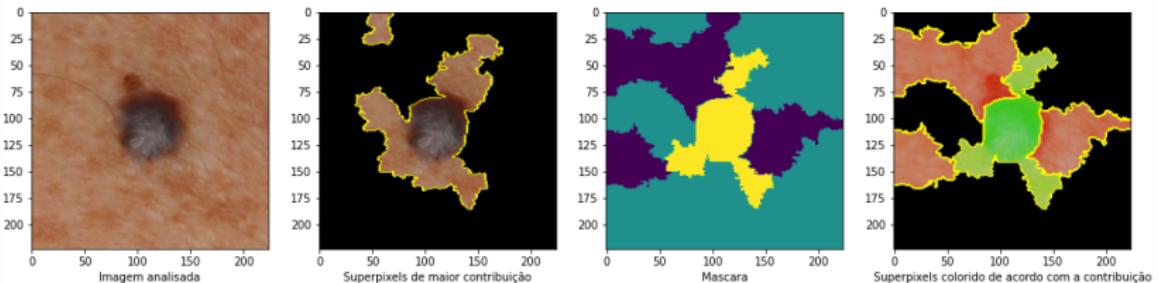
Figure: Estrutura das técnicas de interpretabilidade pertinentes no presente trabalho. Os blocos em verde são as técnicas implementadas no presente trabalho.

Autoria própria

## LIME

*Local Interpretable Model-agnostic Explanations* é uma metodologia apresentada por (RIBEIRO; SINGH; GUESTRIN, 2016) que implementa um modelo local que explica predições individuais.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Metodologia: Interpretabilidade



Resultado da análise usando LIME classificando a lesão como hemangioma com score de 0.798

Figure: *Explicação usando LIME em uma amostra de imagem clínica da lesão Hemangioma e seus respectivos superpixels que mais contribuem para a predição local da imagem e os scores de fidelidade local.*

Autoria própria

## *Gradient-based*

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).

- *Vanilla Gradient*
- *DeconvNet*
- *Guided Backpropagation*
- *Grad-CAM*
- *Guided Grad-CAM*

## *Vanilla Gradient*

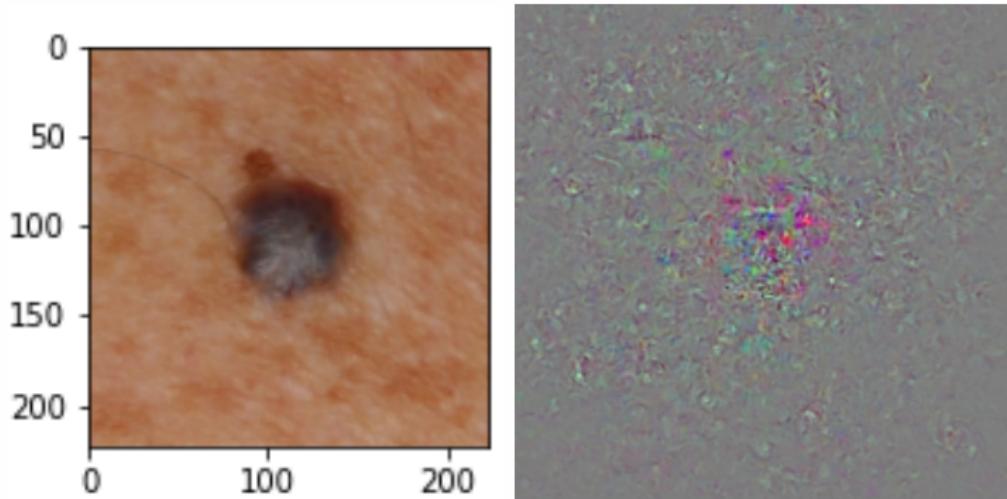


Figure: *Explicação usando Vanilla Gradient em uma amostra de imagem clínica da lesão Hemangioma.*

Autoria própria

## DeconvNet

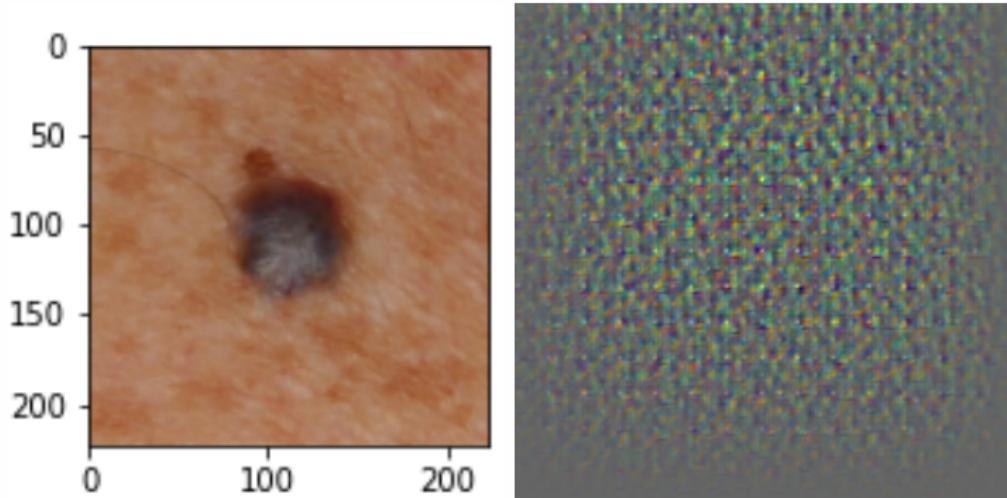


Figure: *Explicação usando deconvNet em uma amostra de imagem clínica da lesão Hemangioma.*

Autoria própria

## *Guided Backpropagation*

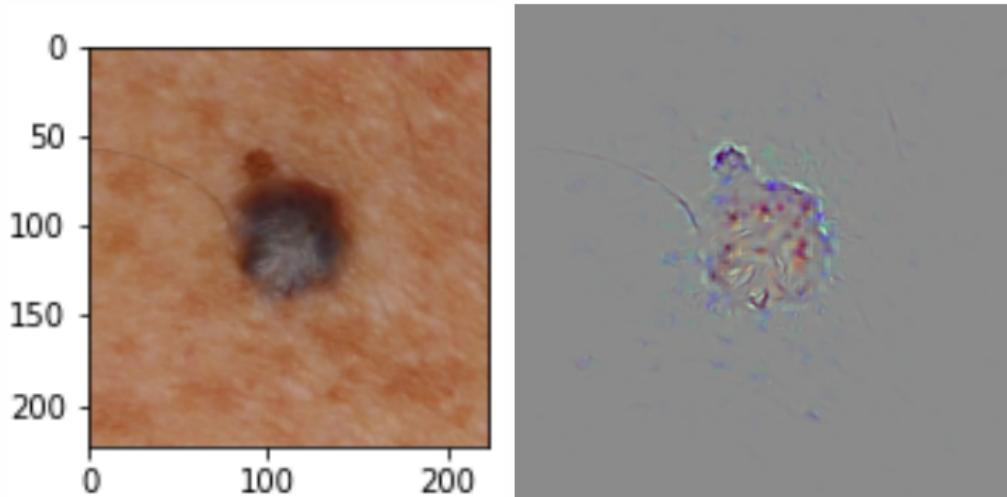


Figure: *Explicação usando Guided Backpropagation em uma amostra de imagem clínica da lesão Hemangioma.*

Autoria própria

## *GradCAM*

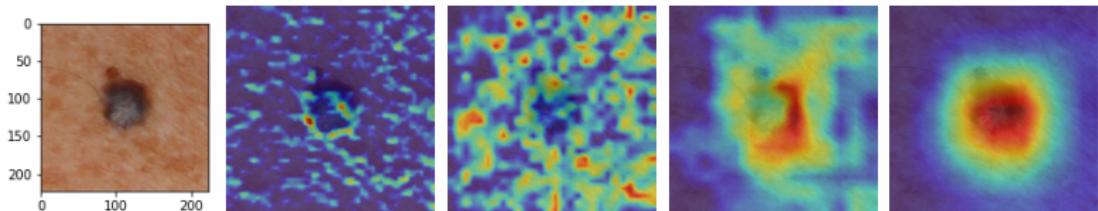


Figure: *Explicação usando GradCAM em uma amostra de imagem clínica da lesão Hemangioma.*

Autoria própria

## *Guided Grad-CAM*

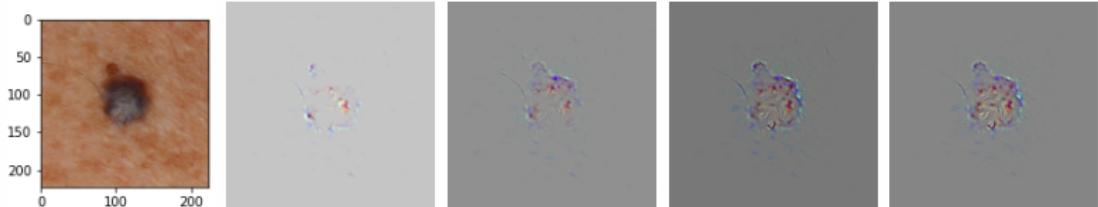


Figure: *Explicação usando Guided Grad-CAM em uma amostra de imagem clínica da lesão Hemangioma.*

Autoria própria

- **Resultados.**
  - Experimentos
  - Indicadores
  - Interpretabilidade de amostras locais
- **Conclusão.**
  - Considerações finais
  - Trabalhos futuros

## Experimento LIME

- Aplicada em toda a base de teste.
- Repositório<sup>1</sup> do autor (RIBEIRO; SINGH; GUESTRIN, 2016a).
- Parâmetros fixos e interação única.

---

<sup>1</sup>Repositório: <https://github.com/marcotcr/lime/>

## Experimento *Gradient-Based*

### Parte 1

- *Vanilla Gradient, DeconvNet e Guided Backpropagation.*
- Resultado para as cinco classes de maior confiança usando Guided Grad-CAM e Grad-CAM.

### Parte 2 - Grad-CAM

- O objetivo é de forma visual observar ao longo de toda a rede neural como as características da imagem estão sendo avaliadas pela rede.
- Apenas para a classe de maior confiança.

Universidade de Brasília - UnB  
 Faculdade UnB Gama - FGA  
 Resultados: Experimento *Gradient-Based*

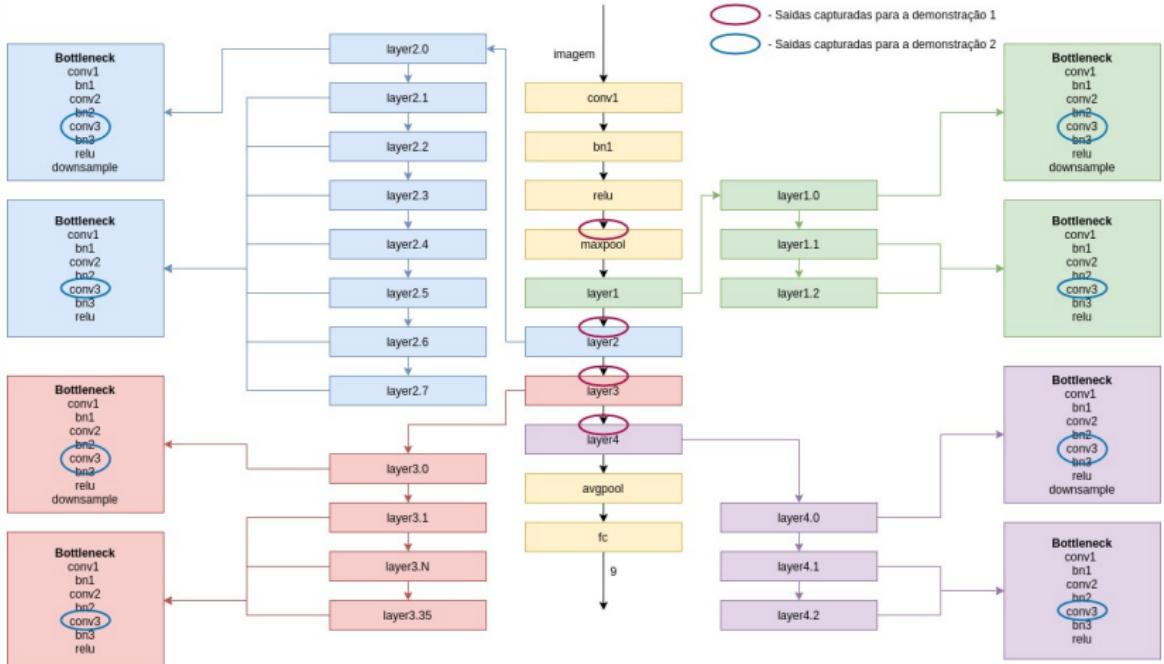


Figure: Diagrama de blocos da arquitetura da rede neural utilizada. Os círculos representam as camadas onde foram gerados os resultados dos experimentos.

## Indicadores

### Complexidade de implementação técnica

Dificuldade de implementação em código.

### Tempo na geração dos resultados

Resultados de tempo de execução, tempo de utilização de *CPU* e picos de memória para a mesma imagem

### Interpretabilidade visual local

Percepção da interpretabilidade visual das amostras utilizadas.

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Indicadores

Table: *Tabela de resumo de indicadores considerando as observações do autor.*

Resumo indicadores	
<b>Complexidade de implementação da técnica</b>	<b>Complexidade</b>
LIME	Facil
Gradient-based	Médio
<b>Tempo na geração dos resultados</b>	<b>CPU Times</b>
LIME	12.85 s
Gradient-based	436,5 ms
<b>Complexidade de implementação da técnica</b>	<b>Intepretabilidade visual</b>
LIME	Média
Vanilla Gradient	Baixa
DeconvNet	Baixa
Guided Backpropagation	Média
Grad-CAM	Alta
Guided Grad-CAM	Alta

## Interpretabilidade de amostras locais

- Técnicas *LIME*, *guided grad-CAM* e *grad-CAM*.
- Squamos Cell Carcinoma.
- Malignant Melanoma.
- Exemplo de *Overfitting*.

## Squamous Cell Carcinoma

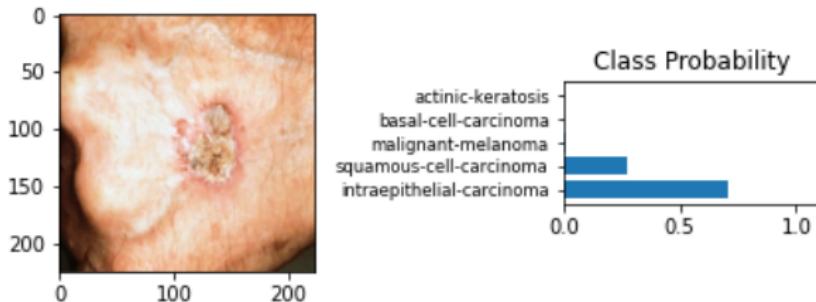


Figure: Probabilidade de classes da imagem de referência para Squamous Cell Carcinoma, falso-negativo.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

(a) (b) (c) (d) (e)

Figure: Guided Grad-CAM para a imagem da Figura 16. Imagens sequenciais das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Intraepithelial Carcinoma a classe predita pelo modelo, (b) Squamous cell Carcinoma a verdadeira classe da imagem, (c) Malignant Melanoma, (d) Basal Cell Carcinoma e (e) Actinic Keratosis.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

(a) (b) (c) (d) (e)

Figure: *Grad-CAM para a imagem da Figura 16. Imagens sequenciais das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Intraepithelial Carcinoma a classe predita pelo modelo, (b) Squamous cell Carcinoma a verdadeira classe da imagem, (c) Malignant Melanoma, (d) Basal Cell Carcinoma e (e) Actinic Keratosis.*

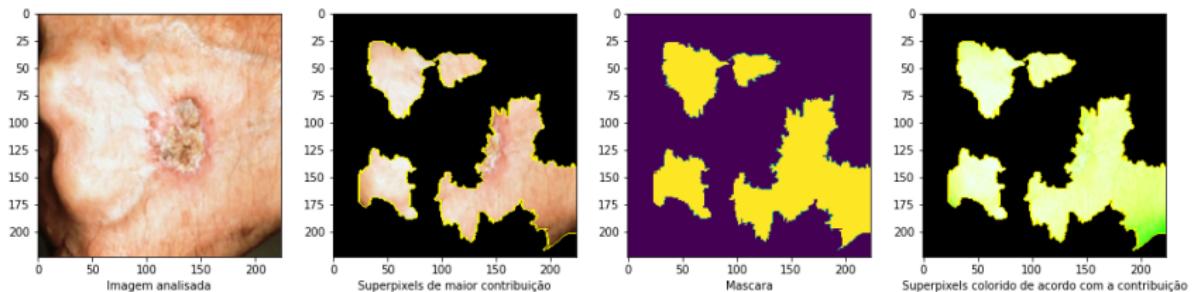
Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

Figure: *Grad-CAM de toda a rede neural para a imagem da Figura 16, com a classificação dada como Intraepithelial Carcinoma, contudo a verdadeira classe é Squamous cell Carcinoma. São 50 imagens sequenciais das saídas da terceira camada convolucional de cada bloco da rede neural.*

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma



Resultado da análise usando LIME classificando a lesão como intraepithelial-carcinoma com score de 0.553

Figure: Resultado do LIME para a imagem da Figura 16, falso-negativo.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

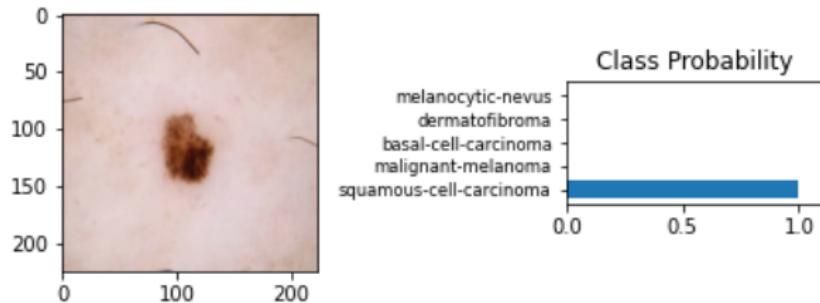


Figure: Probabilidade de classes da imagem de referência para Squamos Cell Carcinoma, verdadeiro-positivo

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

(a) (b) (c) (d) (e)

Figure: Guided Grad-CAM para a imagem da Figura 24. Imagens sequenciais das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Squamous cell Carcinoma a classe predita pelo modelo e classe real da imagem, (b) Malignant Melanoma, (c) Basal Cell Carcinoma, (d) Dermatofibroma e (e) Melanocytic nevus.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

(a) (b) (c) (d) (e)

Figure: *Grad-CAM para a imagem da Figura 16, Imagens sequenciais das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Squamous cell Carcinoma a classe predita pelo modelo, (b) Malignant Melanoma, (c) Basal Cell Carcinoma, (d) Dermatofibroma e (e) Melanocytic nevus.*

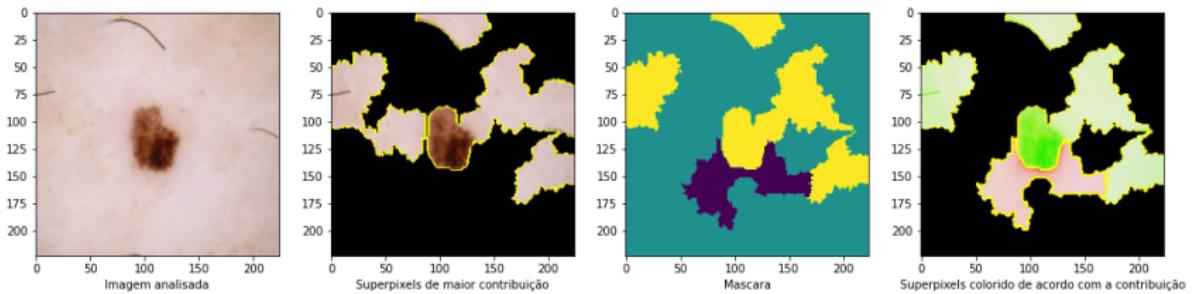
Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma

Figure: *Grad-CAM de toda a rede neural para a imagem da Figura 16, com a classificação sendo Squamous cell Carcinoma. São 50 imagens sequenciais das saídas da terceira camada convolucional de cada bloco da rede neural.*

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Squamos Cell Carcinoma



Resultado da analise usando LIME classificando a lesão como squamous-cell-carcinoma com score de 0.999

Figure: *Resultado do LIME para a imagem da Figura 21, verdadeiro-positivo,*

Autoria própria

# Malignant Melanoma

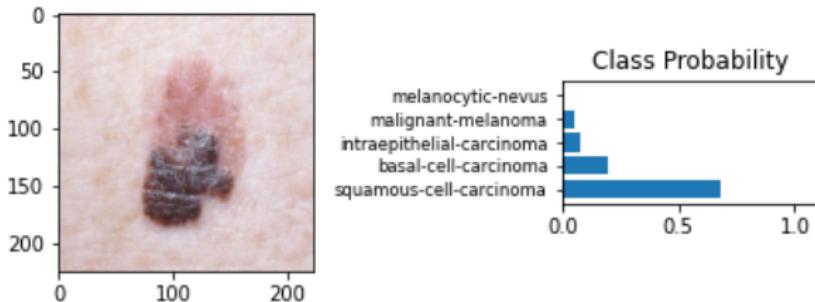


Figure: Probabilidade de classes da imagem de referência para Malignant Melanoma,  
Falso-negativo

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

(a)

(b)

(c)

(d)

(e)

Figure: Guided Grad-CAM para a Figura 26 Falso-negativo. As imagens das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Squamous cell Carcinoma a classe predita pelo modelo, (b) Basal Cell Carcinoma, (c) Intraepithelial Carcinoma, (d) Malignant Melanoma real classe da imagem e (e) Melanocytic Nevus.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

(a) (b) (c) (d) (e)

Figure: *Grad-CAM para A Figura 26 Falso-negativo mostrando imagens sequenciais das saídas da camada ReLu, layer1, Layer2, Layer3 e Layer4. (a) Squamous cell Carcinoma a classe predita pelo modelo, (b) Basal Cell Carcinoma, (c) Intraepithelial Carcinoma, (d) Malignant Melanoma real classe da figura e (e) Melanocytic Nevus.*

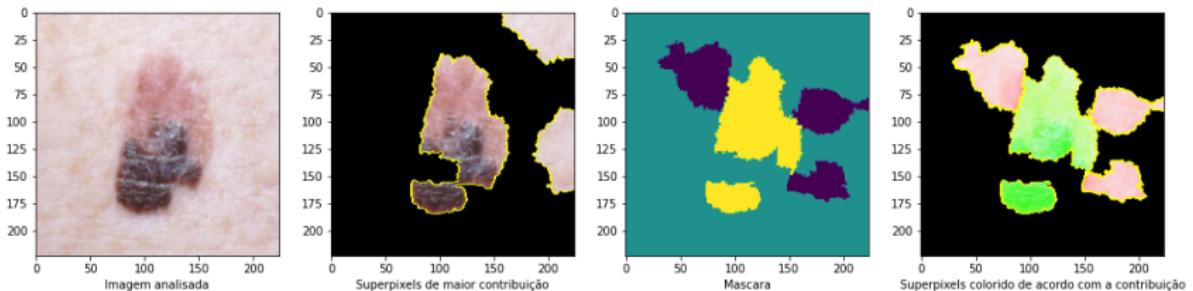
Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

Figure: *Grad-CAM de toda a rede neural para a imagem da Figura 26, com a classificação sendo Squamous cell Carcinoma a real classe é Malignant Melanoma. São 50 imagens sequenciais das saídas da terceira camada convolucional de cada bloco da rede neural.*

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma



Resultado da análise usando LIME classificando a lesão como squamous-cell-carcinoma com score de 0.765

Figure: Resultado do LIME para a imagem da Figura 26, Falso-negativo.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

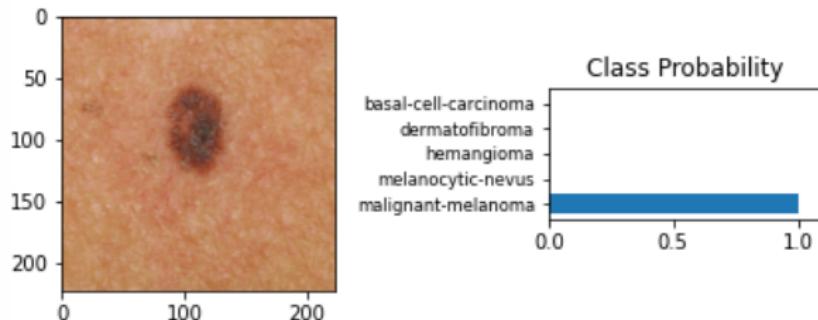


Figure: Probabilidade de classes da imagem de referência para Malignant Melanoma, verdadeiro-positivo.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

(a) (b) (c) (d) (e)

Figure: Guided Grad-CAM para a Figura 31 verdadeiro-positivo. As imagens das saídas das camadas ReLu, layer1, Layer2, Layer3 e Layer4 em sequência. (a) Malignant Melanoma a classe predita pelo modelo e a real classe da imagem, (b) Melanocytic Nevus, (c) Hemangioma, (d) Dermatofibroma e (e) Basal Cell Carcinoma.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

(a) (b) (c) (d) (e)

Figure: *Grad-CAM para A Figura 31 verdadeiro-positivo mostrando imagens sequenciais das saídas das camadas ReLu, layer1, Layer2, Layer3 e Layer4.* (a) *Malignant Melanoma* a classe predita pelo modelo e a real classe da imagem, (b) *Melanocytic Nevus*, (c) *Hemangioma*, (d) *Dermatofibroma* e (e) *Basal Cell Carcinoma*.

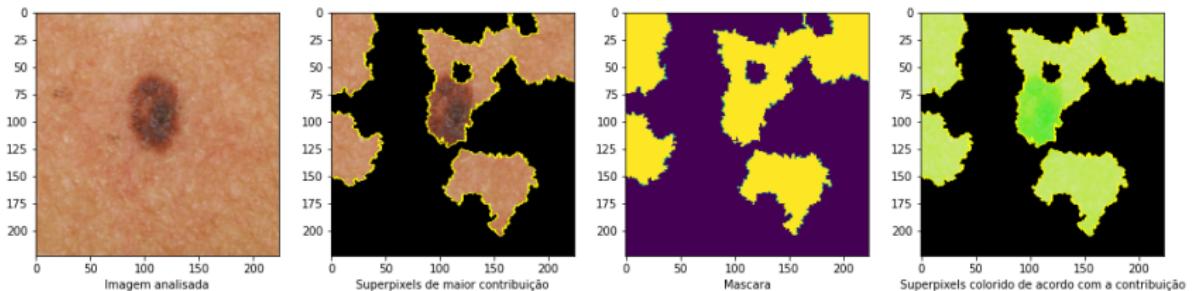
Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma

Figure: *Grad-CAM de toda a rede neural para a imagem da Figura 31, com a classificação correspondendo a real classe: Malignant Melanoma. São 50 imagens sequenciais das saídas da terceira camada convolucional de cada bloco da rede neural.*

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: Malignant Melanoma



Resultado da analise usando LIME classificando a lesão como malignant-melanoma com score de 0.573

Figure: *Resultado do LIME para a imagem da Figura 31, verdadeiro-positivo.*

Autoria própria

## Exemplo de *Overfitting*

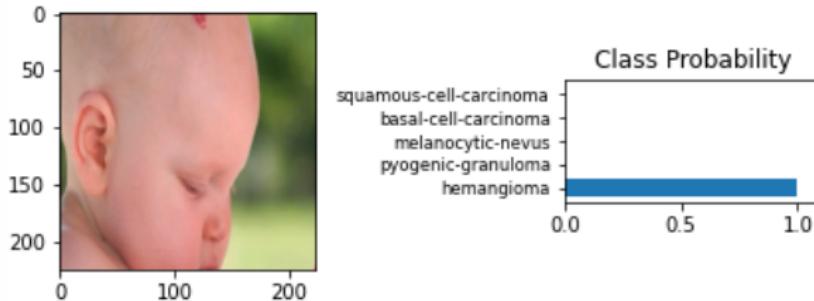


Figure: Probabilidade de classes da imagem de referência para Hemangioma, verdadeiro-positivo com possível Overfitting.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: *Overfitting*

(a) (b) (c) (d) (e)

Figure: Guided Grad-CAM para a Figura 36 verdadeiro-positivo com possível overfitting.  
As imagens das saídas das camadas ReLu, layer1, Layer2, Layer3 e Layer4 em sequência.  
(a) Hemangioma a classe predita pelo modelo e a real classe da imagem, (b) Pyogenic Granuloma, (c) Melanocytic Nevus, (d) Basal Cell Carcinoma e (e) Squamous Cell Carcinoma.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: *Overfitting*

(a) (b) (c) (d) (e)

Figure: *Grad-CAM para a Figura 36 verdadeiro-positivo com possível overfitting. As imagens das saídas das camadas ReLu, layer1, Layer2, Layer3 e Layer4 em sequência.* (a) *Hemangioma* a classe predita pelo modelo e a real classe da imagem, (b) *Pyogenic Granuloma*, (c) *Melanocytic Nevus*, (d) *Basal Cell Carcinoma* e (e) *Squamous Cell Carcinoma*.

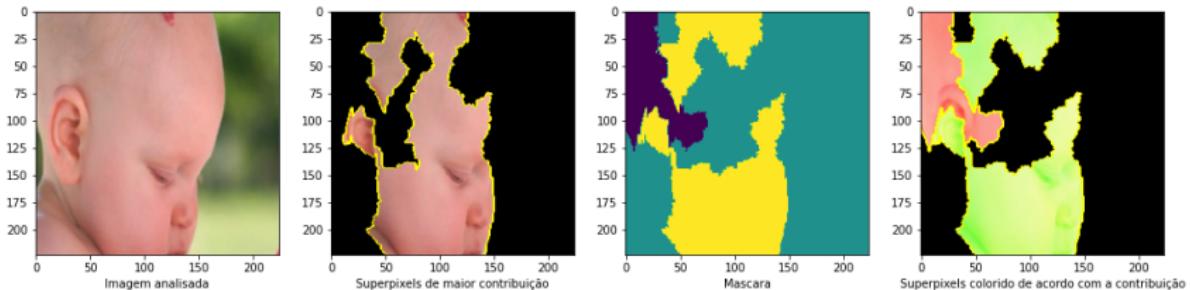
Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: *Overfitting*

Figure: *Grad-CAM* de toda a rede neural para a imagem da Figura 36, com a classificação correspondendo a real classe: *Hemangioma*. São 50 imagens sequenciais das saídas da terceira camada convolucional de cada bloco da rede neural.

Autoria própria

Universidade de Brasília - UnB  
Faculdade UnB Gama - FGA  
Interpretabilidade de amostras locais: *Overfitting*



Resultado da analise usando LIME classificando a lesão como hemangioma com score de 0.320

Figure: *Resultado do LIME para a imagem da Figura 36, verdadeiro-positivo com possível overfitting.*

Autoria própria

## Considerações finais

- Utilização de "caixas-pretas" em aplicações clínicas.
- Criação do modelo usando técnicas estado da arte.
- Aplicação de técnicas de visualização.
- Limitações, fragilidades e vulnerabilidades.
- Visualizações locais, auditar vieses, casos de *overfittings* e *underfittings*.

## Trabalhos futuros

- Coletar mais dados ou focar em outras aplicações críticas.
- Salientando detalhadamente tais fragilidades.
- Trabalhos conjuntos entre especialidades da área de interesse e os pesquisadores de XAI

"We believe that a true AI system should not only be intelligent, but also be able to reason about its beliefs and actions for humans to trust and use it." (Selvaraju et al., 2017)

## Referências

- 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

- 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

- 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

- 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.  
j<https://christophm.github.io/interpretable-ml-book/>j
- 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

- 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.