

SBCAS2023

XXIII Simpósio Brasileiro de Computação Aplicada à Saúde



DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO
ASSOCIAÇÃO UFMA - UFPI



Explicando as decisões com IAs: Demonstrando sua aplicação em imagens médicas

**Elineide Santos, Justino Duarte,
Luis Henrique Vogado, Leonardo Sousa, Hércio
Soares e Rodrigo Veras**

- Elineide Silva dos Santos
- Graduada em Ciência da Computação - UFPI
- Mestra em Ciência da Computação na área Computação Visual - UFPI
- Doutoranda em Ciência da Computação na área de Computação Aplicada - DCCMAPI Associação UFMA/UFPI

- Introdução;
- Objetivo;
- Aprendizado Profundo;
- Métodos de Explicabilidade;
- Explicabilidade na Prática;
- Conclusões.

- Processar uma imagem envolve um conjunto de etapas, não é uma tarefa trivial;
- Atualmente a IA busca resolver tarefas difíceis de ser descritas formalmente, como reconhecimento de fala, padrões e imagens;
- É um grande desafio para um computador a habilidade de entender conhecimentos passados e prever/resolver problemas futuros.

- Em Aprendizado de máquina, a aprendizagem é feita por meio de treinamentos em banco de dados;
- Classificação de imagens é uma das aplicações da aprendizagem de máquina;
- Deep Learning é uma técnica desenvolvida a partir de redes neurais que são capazes de reconhecer e classificar padrões.

- Modelos baseados em redes neurais artificiais não têm explicação natural para seus resultados;
- Esses modelos são compostos por uma enorme quantidade de parâmetros que se relacionam de várias maneiras;
- As redes convolucionais são chamadas de modelos caixas pretas no contexto de interpretabilidade.

- O objetivo é abordar aspectos teóricos sobre explicabilidade, sua importância no contexto médico, e conheceremos vários métodos de explicação de modelos DNN.

Representação Computacional de Imagens



(a)



(b)



(c)



(d)

Figura 1: Imagens coloridas no padrão RGB.

(a) RGB, (b) Vermelho, (c) Verde, (d) Azul.

- Uma imagem digital pode ser descrita por uma matriz $m \times n$;
 - **Imagens binárias:** pixels tem valores 0 e 1;
 - **Imagens em tons de cinza:** valores podem variar entre 0 e 255;
 - **Imagem colorida:** esse pixel é representado por três valores que variam de 0 a 255.
- Imagens coloridas possuem múltiplas bandas de frequências.

Técnicas de Aumento de Dados (Data Augmentation)

- Dados limitados são um grande obstáculo na aplicação de modelos de aprendizagem profunda;
- Classes desequilibradas podem ser um obstáculo adicional;
- Técnicas capazes de modificar a forma, tamanho e posição dos componentes presentes dentro da imagem. Todos esses métodos são fundamentados em transformações geométricas e têm como característica principal a fácil implementação.

- A inversão na imagem original, tal inversão pode ser na horizontal ou na vertical.



(a) original



(b) Flip horizontal



(c) Flip Vertical

Figura 2: Exemplos da utilização da técnica Flipping.

- A técnica que realiza os aumentos girando a imagem para a direita ou esquerda em um eixo entre 1° e 359° . O parâmetro do grau de rotação é responsável por garantir a segurança dos aumentos realizados pela rotação.



(a)



(b)



(c)



(d)

Figura 3: Exemplos da utilização da técnica Rotação de 90° .

- Essa técnica pode mudar as imagens para a esquerda, direita, para cima ou para baixo. Essa característica pode ser vantajosa em conjuntos com distorções posicionais.



(a)



(b)



(c)

Figura 4: Exemplos da utilização da técnica Translação.

- Um zoom aleatório é obtido pelo argumento `zoom_range`. Um zoom menor que 1.0 amplia a imagem, enquanto um zoom maior que 1.0 diminui o zoom da imagem.



(a)



(b)



(c)

Figura 5: Exemplos da utilização da técnica Zoom.

- Essa transformação inclina a forma da imagem. O contraste entre rotação e cisalhamento refere-se à fixação de um eixo, durante a aplicação da transformação de cisalhamento.



(a)



(b)



(c)

Figura 6: Exemplos da utilização da técnica de cisalhamento.

Exemplo Colab



```
# Caminho da imagem de entrada e saída
IMAGE_PATH = "/content/drive/MyDrive/Minicurso SBCAS/gato.png"
OUTPUT_PATH = "/content/drive/MyDrive/Minicurso SBCAS/dataAumentation/"

# carregar a imagem original e converter em array
image = tf.keras.preprocessing.image.load_img(IMAGE_PATH)
image = tf.keras.preprocessing.image.img_to_array(image)

# adicionar uma dimensão extra no array
image = np.expand_dims(image, axis=0)

# criar um gerador (generator) com as imagens do data augmentation
imgAug = ImageDataGenerator(
    rotation_range=90)#rotação

imgGen = imgAug.flow(image, save_to_dir=OUTPUT_PATH,
                    save_format='jpg', save_prefix='cissa2')
```

Aprendizado Profundo

- Deep Learning ou Aprendizado Profundo é uma técnica de aprendizado de máquina desenvolvida a partir das Redes Neurais Artificiais (RNA);
- RNAs são capazes de reconhecer e classificar padrões e posteriormente generalizar o conhecimento adquirido.

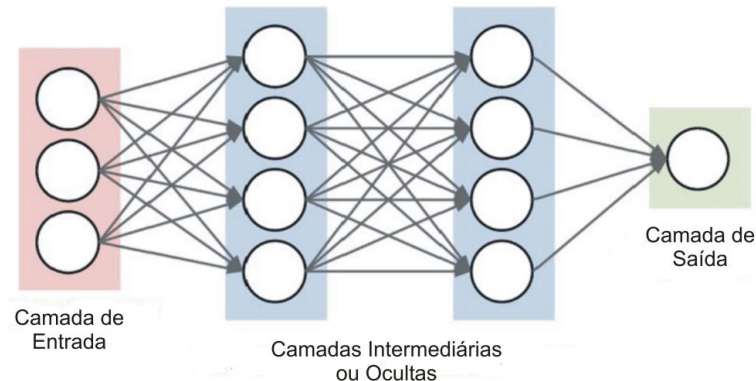


Figura 7: Exemplos de rede neural com duas camadas densas.

- A propriedade mais importante das redes neurais é a habilidade de aprender de seu ambiente e, com isso, melhorar seu desempenho;
- O Aprendizado Profundo passou a ser aplicado em diversas áreas como Visão Computacional, Processamento de Imagens e Computação Gráfica;
- Uma das vantagens do Aprendizado Profundo é sua capacidade de aprender utilizando um alto volume de dados de forma não-supervisionada.

- Deep Learning pode ser usada para:
 - **Pré-processamento:** que é uma maneira de realizar um ajuste ou melhoramento nos dados;
 - **Extração de características:** que é um processo para reconhecer algum padrão entre os dados e tem como intuito tornar o processo de decisão mais fácil durante a classificação;
 - **Classificação:** que é uma tarefa de predição de uma ou n classes.

- As Redes Neurais Convolucionais (CNNs) fazem parte do conjunto de técnicas de Aprendizado Profundo;
- Modelam abstrações em alto nível através de imagens e camadas convolucionais dispostas de forma sequencial ou não;
- Tem como principais desafios o alto custo computacional necessário e a exigência de grandes bases de dados para o treinamento.

- **Camadas Convolucionais:**

- São compostas por uma quantidade c de filtros com dimensões $d \times d$ que irão extrair mapas de características ao serem convoluídos com as imagens de entrada ou com saídas de outras camadas.

- **Camadas de Pooling:**

- reduz a dimensão espacial dos mapas gerados por camadas anteriores, conseqüentemente reduzindo o custo computacional e preservando as características mais relevantes.

- **Camadas Totalmente Conectadas ou Densas:**

- São constituídas de neurônios que representam pesos e guardam o aprendizado da rede neural. Localizadas ao final da arquiteturas e posteriores a camada de *Flatten* que transforma vetor de características em vetor de neurônios.

- A competição *ImageNet Large Scale Visual Recognition Challenge* (ILSVRC) aumentou a popularidade das CNNs a partir de 2012 com a AlexNet;
- A base de dados popularmente conhecida como ImageNet possui mais de 1 milhão de imagens e cerca de mil classes;
- Durante a competição surgiram CNNs do estado da arte como a InceptionV3, GoogLeNet, VGGNet e ResNet.



Figura 8: Exemplos de imagens apresentadas na base de dados ImageNet.

- Proposta por Simonyan and Zisserman em 2014, possui variações como 16 e 19 camadas convolucionais dispostas de forma sequencial, denominadas VGG-16 e VGG-19;
- A estratégia utilizada no desenvolvimento dessa CNN é a uniformidade no tamanho dos filtros convolucionais, sendo todos 3×3 ;
- No entanto, a quantidade de filtros dobra em quase todos os blocos convolucionais, onde apenas o último mantém a mesma quantidade;
- A arquitetura dispõe de duas camadas totalmente conectadas com 4096 elementos.

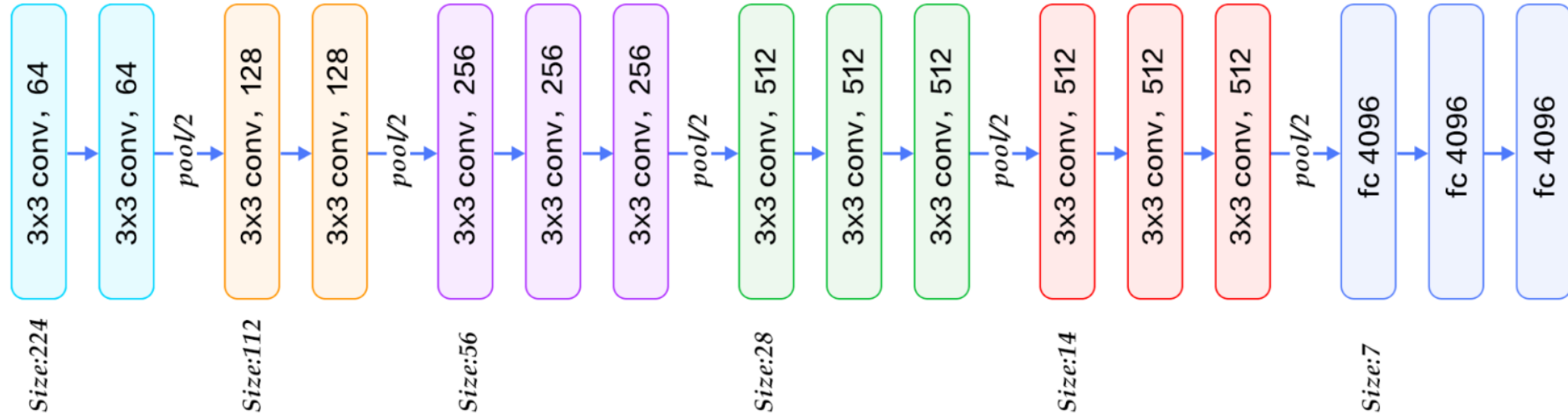


Figura 9: Arquitetura da VGG-16, um dos modelos do tipo VGGNet.

- Proposta em 2015, a Rede Neural Residual ou ResNet foi desenvolvida pelo time da *Microsoft* para solucionar o problema do desaparecimento do gradiente ou *vanishing gradient*;
- Esse problema ocorre quando são adicionadas muitas camadas em um modelo sequencial;
- Como consequência, o aprendizado se torna mais lento e ocasiona uma saturação no desempenho da CNN.

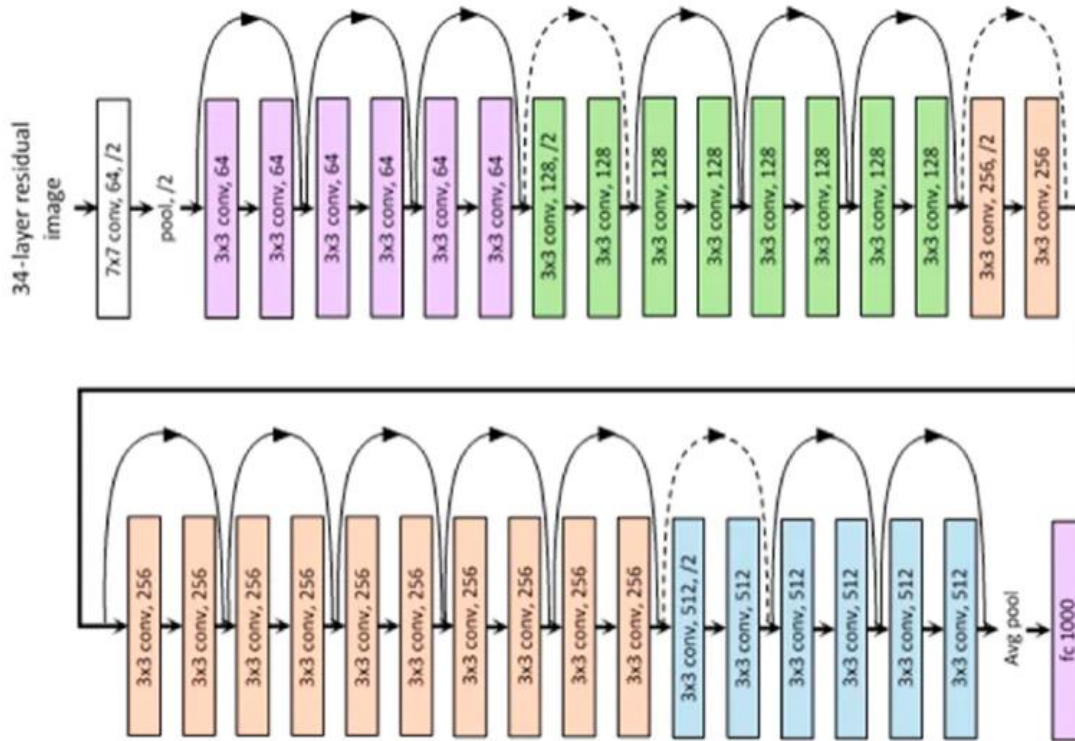


Figura 10: Arquitetura da ResNet34, um dos modelos do tipo ResNet.

Métodos de Explicabilidade

- Explicabilidade basicamente ajuda na interpretação da saída de um modelo por um usuário;
- Uma característica importante das explicações é sua semelhança com o tipo de dado utilizado no treinamento.

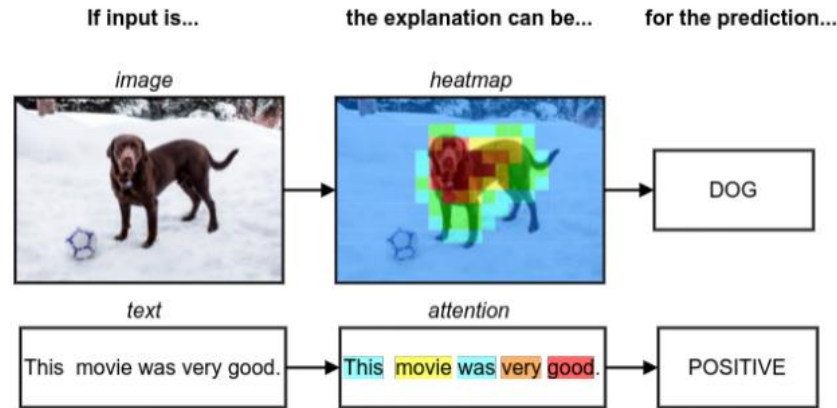


Figura 11: Exemplos de como as explicações podem parecer na prática.

- Diante da enorme quantidade de parâmetros que caracterizam as CNN's, é necessário inspecionar cuidadosamente antes de ser implantado em qualquer ambiente médico;
- Segundo Xie et al. [Xie et al. 2020] os métodos de explicabilidade podem ser categorizados em métodos de visualização, destilação de modelo e métodos intrínsecos.

- A explicabilidade é demonstrada destacando as regiões que influenciam fortemente na saída da rede, os chamados saliency maps ou heatmaps;
- Subdivididos em métodos baseados em retropropagação e visualização baseada em perturbação;
- Ex: Activation Maximization, Deconvolution, CAM and Grad-CAM, Occlusion Sensitivity.

- Modelo inerentemente explicável, popularmente conhecido como “white-bo”;
- Busca-se distinguir as regras de decisão ou características de entrada que induzem as saídas da rede;
- Subdivididos em métodos de aproximação local e tradução de modelos;
- Ex: LIME, SHAP, FSA Based e Rule Based.

- Essas estruturas são facilmente explicáveis, pois eles têm a capacidade de aprender as saídas precisas por entrada e as saídas que expressam uma explicação da ação da rede;
- O usuário precisa ter um conhecimento amplo sobre o campo de aplicação adotado, por essa razão os métodos intrínsecos são mais difíceis de serem implementados e provavelmente exigem um maior custo computacional;
- O *Single-Modal Weighting*, *Multi-Modal Interaction*, *Text Explanation*, *Explanation Association*, *Model Prototype*.

- Base de Imagens;
- Redes neurais;
- Métricas de avaliação;
- Métodos de explicabilidade.

Explicabilidade na Prática - Base de Imagens

- Base ISIC 2016, uma das base de imagens dermatoscópicas encontrada em diversas literaturas;
- Essa base de dados possui 900 imagens, divididas em imagens de lesões benignas e malignas. Vale ressaltar que utilizamos o conjunto de treino do Desafio (Part 3: Lesion Classification), disponível em: ***<https://challenge.isic-archive.com/landing/2016/39/>***
- A resolução das imagens de entrada variam entre 767×576 e 3024×2016. Na Figura 3.12, apresentamos alguns exemplos da base utilizada no experimentos

Explicabilidade na Prática - Base de Imagens

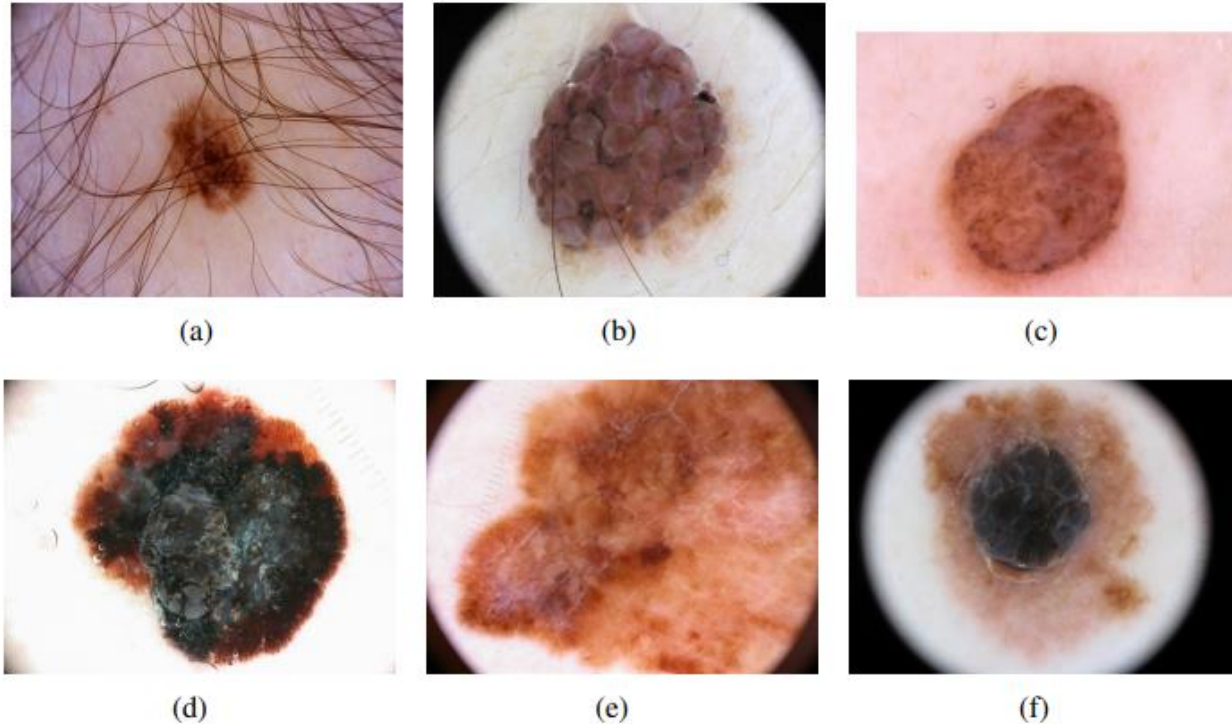


Figura 12: Exemplos da base ISIC 2016.

Explicabilidade na Prática - CNNs

- VGG-16;
- ResNet.

- Acurácia;
- Precisão;
- Recall;
- F1-score;
- Kappa.

- **Localização da pasta dos arquivos:**

- Drive
- Link: <https://drive.google.com/drive/folders/1jDvVrO9AywyTcKp3P9xHyCM33Abi0DFw?usp=sharing>

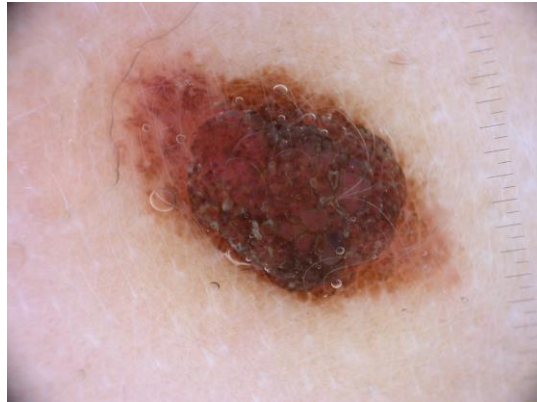
- **Ambiente:**

- Colab

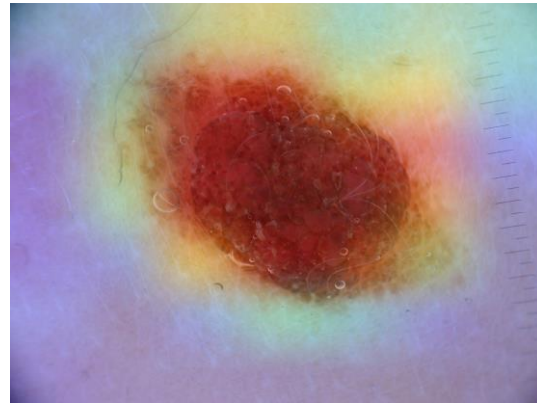
- **Linguagem de programação:**

- Python

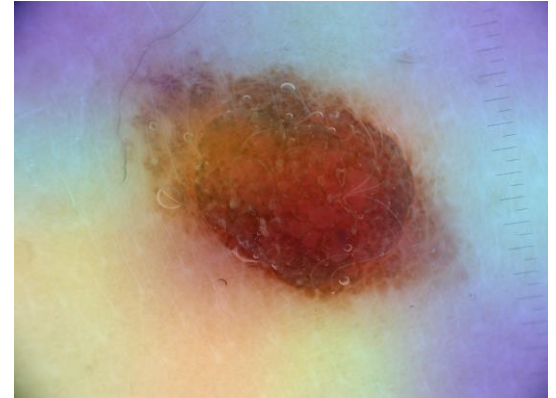
- Baseada no Class Activation Mapping (CAM), consiste em gerar mapas de calor sobre a imagem, baseados nos gradientes da rede;
- Assim pode-se visualizar regiões específicas da imagem que apresentaram maior contribuição para a classificação de uma determinada classe na camada de predição do modelo.



(a) Original

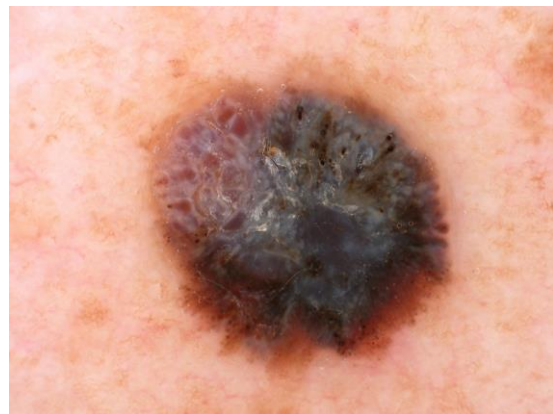


(b) VGG-16

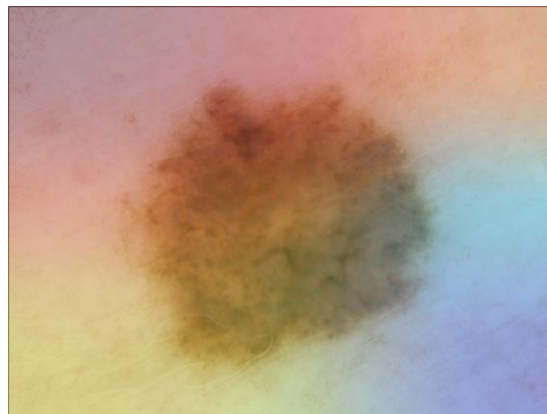


(c) ResNet

Figura 13: Exemplos da aplicação do Grad- CAM em lesão benigna.



(a) Original



(b) VGG-16

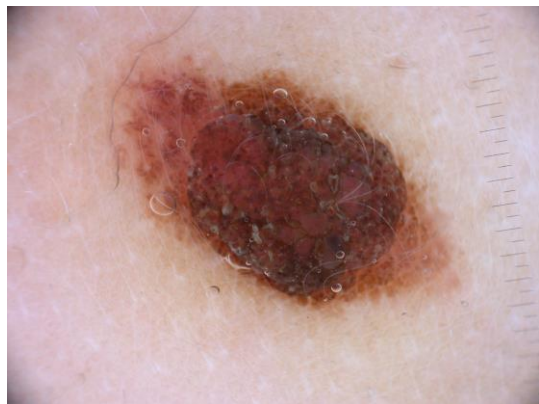


(c) ResNet

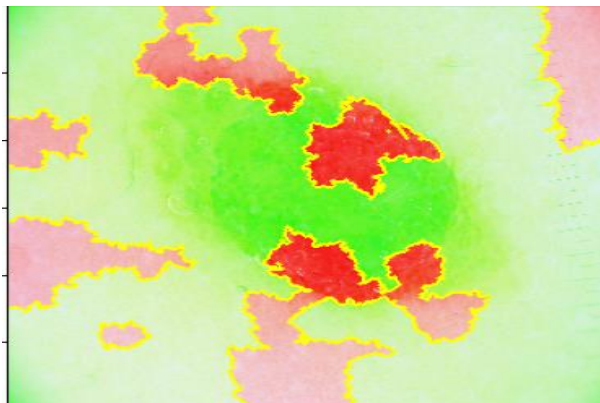
Figura 14: Exemplos da aplicação do Grad-CAM em lesão maligna.

Métodos de Explicabilidade - LIME

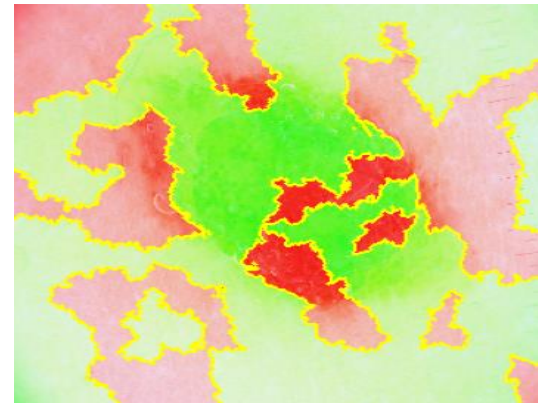
- Essa técnica é um “modelo-agnóstico”, o que significa que pode ser aplicado a qualquer tipo de modelo, desde que haja uma função de predição disponível;
- Ele usa uma abordagem de interpretabilidade local, que explora as decisões tomadas por um modelo em relação a instâncias específicas de dados;
- O LIME produz explicações em termos de recursos (features) relevantes para uma instância específica, o que ajuda os usuários a entender o motivo pelo qual o modelo tomou uma decisão.



(a) Original

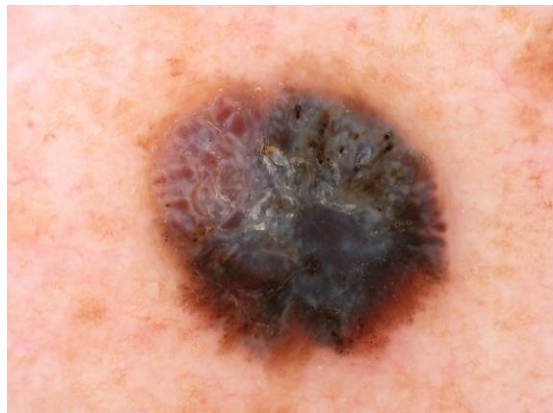


(b) VGG-16

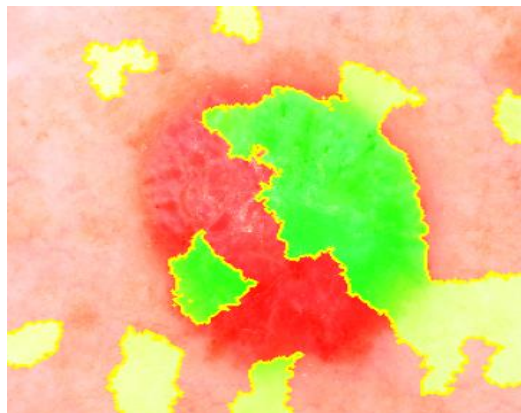


(c) ResNet

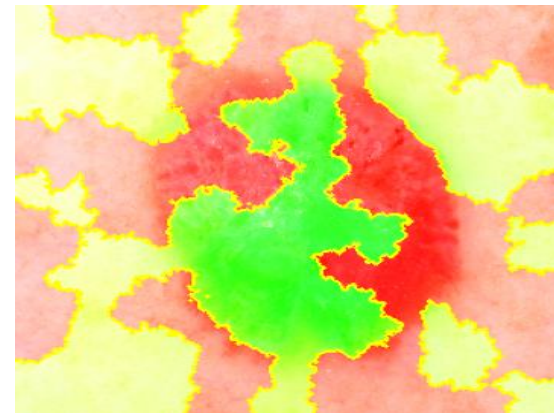
Figura 15: Exemplos da aplicação do LIME em lesão benigna.



(a) Original



(b) VGG-16



(c) ResNet

Figura 16: Exemplos da aplicação do LIME em lesão maligna.

- O t-SNE é um algoritmo que permite a redução e visualização de dados de altas dimensões;
- Por meio dessa técnica, é possível representar características extraídas por CNNs em alto nível;
- Permite criar novas representações que justifiquem como o aprendizado ocorreu e quais características são cruciais para a separação das classes utilizando um classificador linear.

Métodos de Explicabilidade - t-SNE

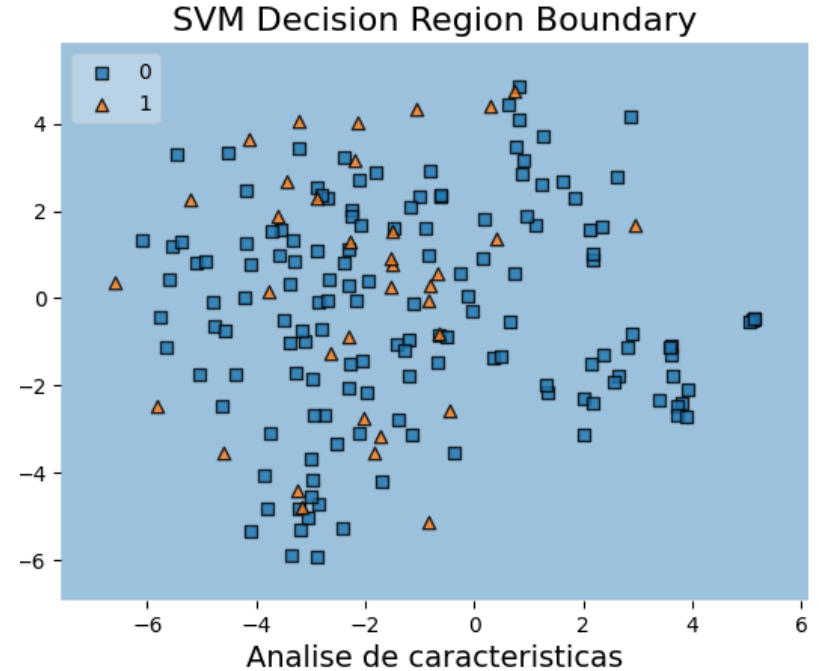
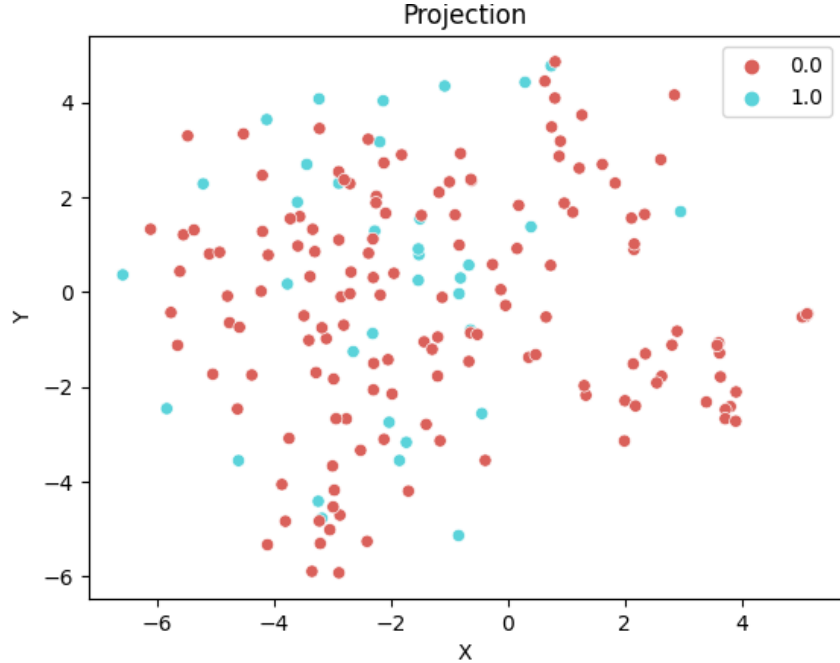


Figura 17: (a) Projeção do t-SNE no plano espacial de características. (b) Predição via SVM Linear para verificar a separação das classes.

Métodos de Explicabilidade - t-SNE (Leucemia)

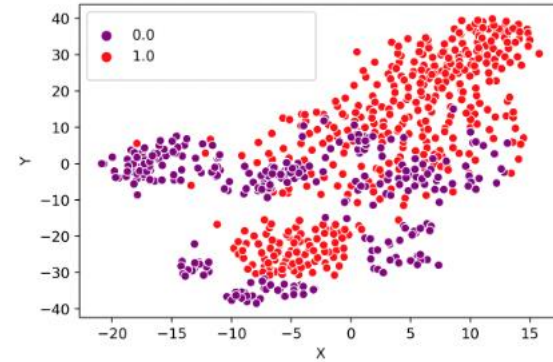
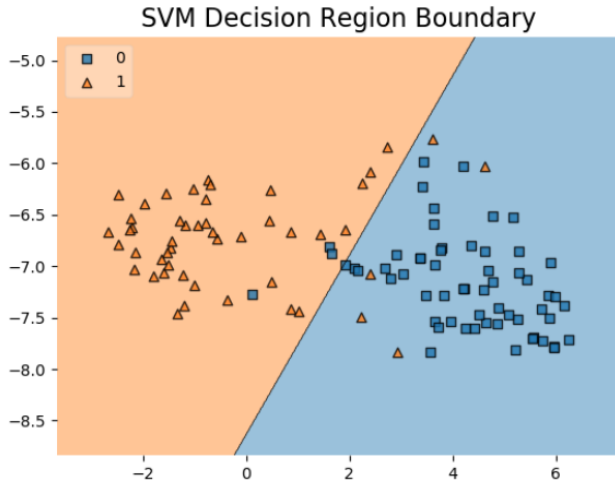
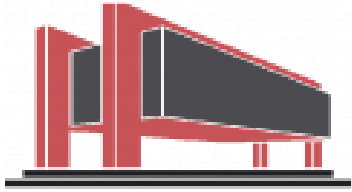


Figura 18: (a) Projeção do t-SNE para a base de Leucemia de acordo como Vogado et al. (2020).

- Nos experimentos foi definida uma base de imagens, e demonstrado seu pré-processamento;
- Criação de duas CNN a partir do conjunto ImageNet;
- As CNN's foram treinadas e em seguidas testadas com a finalidade de verificar a quantidade de predições realizadas corretamente;
- Aplica-se a explicabilidade nas imagens preditas corretamente, com a finalidade de justificar a tomada de decisão.

- Logo, a utilização da explicabilidade traz um grande benefício para o meio científico e conseqüentemente uma maior aceitação em trabalhos clínicos, pois aumenta a distância do padrão “caixa-preta”;
- Um direcionamento futuro para trabalhar com explicabilidade seria buscar otimizar o desempenho das CNNs ou seria a aplicação de métodos intrínsecos sobre as imagens para uma explicação mais transparente.



SBCAS2023

XXIII Simpósio Brasileiro de Computação Aplicada à Saúde



DOUTORADO EM CIÊNCIA DA COMPUTAÇÃO
ASSOCIAÇÃO UFMA - UFPI



Explicando as decisões com IAs: Demonstrando sua aplicação em imagens médicas

**Elineide Santos, Justino Duarte,
Luis Henrique Vogado, Leonardo Sousa, Hércio
Soares e Rodrigo Veras**