

The logo for IADE (Instituto de Análise de Dados e Estatística) features the letters 'IADE' in a bold, white, sans-serif font. The letters are set against a black background that is shaped like a parallelogram, with the right side slanted upwards. A thick black horizontal line is positioned above the logo, and a thick black vertical line is positioned to the left of the logo, forming an L-shape that frames the top-left corner of the page.

**IADE**

**Inteligência Artificial e Machine Learning**

---

**Trabalho**

# **Deteção de Melanoma em Imagens de Pele Utilizando Redes Neurais Convolucionais**



## Objetivos

- Desenvolver um modelo de classificação multiclasse utilizando uma rede neural convolucional (CNN) no TensorFlow, capaz de detetar com precisão diferentes tipos de doenças de pele, incluindo melanoma.
- Compreender o processo completo de construção de um modelo de CNN, desde a preparação e visualização dos dados até ao treino e avaliação do modelo, incluindo estratégias para lidar com overfitting, underfitting e desequilíbrios de classe.

## Enunciado

- Nesta atividade, irá construir um modelo de classificação multiclasse utilizando uma rede neural convolucional (CNN) no TensorFlow. O objetivo é desenvolver um modelo personalizado capaz de detetar com precisão o melanoma, uma forma agressiva de cancro de pele. O melanoma é responsável por 75% das mortes relacionadas com cancro de pele, tornando crucial a implementação de soluções que possam auxiliar no diagnóstico precoce.
- O conjunto de dados fornecido (link a disponibilizar) é composto por 2.357 imagens de doenças oncológicas malignas e benignas, classificadas de acordo com a International Skin Imaging Collaboration (ISIC). As imagens estão distribuídas em várias categorias, sendo o melanoma e os nevos ligeiramente dominantes. As classes incluídas no conjunto de dados são:
  - Queratose actínica
  - Carcinoma de células basais
  - Dermatofibroma
  - Melanoma
  - Nevus
  - Queratose benigna pigmentada
  - Queratose seborreica
  - Carcinoma de células escamosas
  - Lesão vascular

- **Tarefas a Realizar:**

- 1. Leitura e Compreensão dos Dados:** Definir os caminhos para os dados de treino e teste. Realizar uma análise inicial para compreender a distribuição e o conteúdo das imagens.
- 2. Criação do Conjunto de Dados:** Criar os conjuntos de dados de treino e validação a partir da pasta de treino, com batch size 32, redimensionando as imagens para 180x180 pixels.
- 3. Visualização do Conjunto de Dados:** Implementar um código para visualizar instâncias de todas as nove classes presentes no conjunto de dados.
- 4. Criação e Treino do Modelo Inicial:**
  - Construir um modelo CNN capaz de classificar as nove classes.
  - Normalizar os valores dos pixels entre 0 e 1.
  - Escolher uma função de otimização e de perda adequada.
  - Treinar o modelo durante 20 épocas.
  - Analisar os resultados iniciais, verificando se há indícios de overfitting ou underfitting.
- 5. Aumento de Dados (Data Augmentation):** Aplicar uma estratégia de aumento de dados para mitigar possíveis problemas de overfitting ou underfitting.
- 6. Criação e Treino do Modelo após Aumento de Dados:**
  - Reconstruir e treinar a CNN com os dados aumentados.
  - Treinar o modelo durante 20 épocas.
  - Avaliar se o aumento de dados resolveu os problemas anteriores.
- 7. Distribuição e Tratamento de Desequilíbrios de Classe:**
  - Analisar a distribuição das classes no conjunto de treino.
  - Identificar as classes sub-representadas.
  - Utilizar a biblioteca Augmentor para equilibrar as classes do conjunto de dados.
- 8. Criação e Treino do Modelo após Correção de Desequilíbrios:**
  - Treinar novamente a CNN com as correções de desequilíbrio de classe.
  - Treinar o modelo durante 30 épocas.
  - Analisar se os problemas foram resolvidos.

9. **Observação:** Não utilize modelos pré-treinados. Todos os processos de construção devem ser baseados em um modelo personalizado. Devido ao tempo de treino ser elevado, recomenda-se a utilização de GPUs, como as disponibilizadas pelo Google Colab.

## Instruções de entrega

- Trabalho: **Individual.**
- Estrutura do Trabalho:
  - Notebook: **Os alunos devem entregar um Jupyter Notebook (.ipynb) contendo:**
    - **Introdução:** Descrição do problema e dos objetivos do trabalho, incluindo a importância da detecção precoce do melanoma.
    - **Análise Exploratória de Dados (EDA):** Passo a passo da análise exploratória, incluindo visualização de instâncias das diferentes classes, distribuição das classes e insights iniciais sobre o dataset.
    - **Pré-processamento:** Código e explicação das etapas de preparação dos dados, como redimensionamento das imagens, normalização e divisão em conjuntos de treino e teste.
    - **Modelagem:** Implementação do modelo de rede neural convolucional (CNN). Incluir o código detalhado e comentado para a construção da rede, treino inicial, aplicação de técnicas de aumento de dados (data augmentation) e tratamento de desequilíbrios de classe.
    - **Treino e Avaliação:** Cálculo e análise das métricas de desempenho do modelo, incluindo a discussão sobre problemas de overfitting ou underfitting, e melhorias após cada etapa de treino.
    - **Conclusão:** Reflexão sobre os resultados obtidos, discussões sobre o desempenho do modelo ao longo das diferentes etapas e sugestões para melhorias futuras.
  - **Formato de Entrega:**
    - **Notebook (.ipynb):** O código completo, incluindo gráficos, visualizações e análises, deve ser entregue em um único arquivo Jupyter Notebook.
    - **Plataforma de Submissão:** Canvas.



CREATING  
CREATORS

<https://www.iade.europeia.pt/>  
Geral: +351 213 939 600  
Av. Dom Carlos I, 4  
1200-649 Lisboa. Portugal