

# **Inteligência Computacional**

## **Meta I | Skin Diseases**

2025/2026

Daniel Silva

2023144551

Guilherme Martins

2023144573

## Conteúdo

Capítulo 1 – Descrição do Caso de Estudo e Objetivos do Problema .....	3
1.1 Descrição do Caso de Estudo .....	3
1.2 Objetivos do Problema .....	4
Capítulo 2 – Descrição do Modelo e Implementação dos Algoritmos .....	4
2.1 Preparação dos Dados.....	4
2.2 Modelos Implementados .....	5
2.2.1 CNN Simples.....	5
2.2.2 CNN Melhorado (Com Flatten) .....	5
2.2.3 CNN Melhorado (Com Global Average Pooling) .....	6
2.3 Processo de Treino e Validação .....	6
2.4 Avaliação dos Modelos .....	6
Capítulo 3 – Análise de Resultados.....	7
3.1 Metodologia de Avaliação .....	7
3.2 Resultados Quantitativos.....	7
3.3 Matrizes de Confusão .....	8
3.4 Análise Crítica dos Resultados.....	10
Conclusão.....	12

# Capítulo 1 – Descrição do Caso de Estudo e Objetivos do Problema

## 1.1 Descrição do Caso de Estudo

As doenças de pele representam um problema de saúde pública que afeta milhões de pessoas em todo o mundo. A deteção precoce e o diagnóstico correto destas patologias são fundamentais para evitar complicações graves e permitir tratamentos mais eficazes. Contudo, o acesso a médicos especialistas em dermatologia nem sempre é possível, sobretudo em regiões mais remotas ou com recursos limitados, o que pode atrasar o diagnóstico e comprometer o tratamento adequado.

Neste contexto, as técnicas de Inteligência Computacional e Aprendizagem Automática têm vindo a assumir um papel cada vez mais relevante na área da saúde. Estas abordagens permitem o desenvolvimento de sistemas capazes de analisar imagens médicas e auxiliar os profissionais no processo de diagnóstico, tornando-o mais rápido, objetivo e acessível.

O presente projeto tem como caso de estudo a classificação automática de doenças de pele a partir de imagens. Para o efeito, foi utilizado o conjunto de dados “Skin Diseases”, disponível na plataforma Kaggle, que contém cerca de 38.760 amostras distribuídas por seis classes distintas:

- Ekzama
- Akne
- Benign
- Malign
- Enfeksiyonel

Cada amostra corresponde a uma imagem representando uma lesão cutânea, sendo o problema formulado como uma tarefa de classificação multiclasse. O conjunto de dados é amplamente utilizado como benchmark em estudos e competições na área da visão computacional, o que permite comparar o desempenho de diferentes modelos. De acordo com o estado da arte, alguns modelos conseguem alcançar taxas de acerto superiores a 85%, o que demonstra o potencial desta abordagem no apoio ao diagnóstico médico.

## 1.2 Objetivos do Problema

O principal objetivo deste projeto consiste em desenvolver um modelo de rede neuronal capaz de classificar corretamente o tipo de doença de pele presente numa imagem. Pretende-se, desta forma, criar uma ferramenta que possa contribuir para o diagnóstico precoce e apoiar os profissionais de saúde na triagem de casos dermatológicos.

De forma mais detalhada, o trabalho tem como objetivos:

- Compreender o problema de classificação de imagens médicas, identificando os principais desafios associados à variabilidade e qualidade das imagens;
- Preparar e analisar o conjunto de dados, assegurando uma divisão equilibrada entre os subconjuntos de treino, validação e teste;
- Implementar e comparar diferentes modelos de redes neurais, nomeadamente CNN
- Avaliar o desempenho dos modelos com base em métricas como a acurácia, sensibilidade, especificidade, F1-score e AUC;
- Refletir sobre possíveis melhorias e aplicações futuras, como a integração do modelo em aplicações móveis de triagem ou em sistemas hospitalares de apoio ao diagnóstico dermatológico.

# Capítulo 2 – Descrição do Modelo e Implementação dos Algoritmos

## 2.1 Preparação dos Dados

Antes da implementação dos modelos, foi necessário preparar o conjunto de dados “Skin Diseases”. O dataset original contém cerca de 38.760 imagens, distribuídas por seis classes de doenças cutâneas. As imagens apresentam diferentes dimensões, níveis de iluminação e variações de cor, o que exigiu uma fase de pré-processamento para garantir uniformidade e qualidade dos dados.

As principais etapas de preparação foram as seguintes:

- Redimensionamento das imagens para resoluções fixas de 64x64 e 128x128 pixels, permitindo a compatibilidade com as redes neurais;
- Normalização dos valores dos pixels para o intervalo [0, 1], assegurando uma escala adequada para o treino da rede;
- Divisão do dataset em três subconjuntos:
  - Pasta Train

- Pasta Test
- Pasta Val
- Aumento artificial dos dados (data augmentation), através de pequenas rotações, inversões horizontais, com o objetivo de aumentar a robustez do modelo e reduzir o risco de *overfitting*.

Esta fase garantiu que o modelo fosse treinado com um conjunto diversificado de imagens, refletindo melhor a variabilidade real das lesões cutâneas.

## 2.2 Modelos Implementados

Com o dataset preparado, foram implementadas e testadas três arquiteturas distintas de redes neurais convolucionais (CNNs), utilizando a biblioteca TensorFlow/Keras.

### 2.2.1 CNN Simples

O primeiro modelo desenvolvido foi uma CNN simples, composta por três blocos convolucionais seguidos de uma camada *Flatten* e duas camadas densas. A estrutura geral foi a seguinte:

```
model = Sequential([
    Conv2D(32, (3,3), activation='relu', input_shape=(img_size,
img_size, 3)),
    MaxPooling2D((2,2)),
    Conv2D(64, (3,3), activation='relu'),
    MaxPooling2D((2,2)),
    Conv2D(128, (3,3), activation='relu'),
    Flatten(),
    Dense(100, activation='relu'),
    Dense(num_classes, activation='softmax')
])
```

Este modelo serviu como base para avaliar o impacto de diferentes resoluções e número de épocas de treino. Apesar da sua simplicidade, apresentou resultados satisfatórios, alcançando acuráncias próximas de 70% no conjunto de teste.

### 2.2.2 CNN Melhorado (Com Flatten)

O segundo modelo corresponde a uma versão melhorada da CNN base, incluindo camadas adicionais de normalização em lote (Batch Normalization), camadas de Dropout para reduzir o overfitting, e um maior número de filtros convolucionais em cada bloco.

A arquitetura é composta por quatro blocos convolucionais, seguidos de camadas densas totalmente ligadas. O Dropout foi aplicado de forma progressiva (entre 0.2 e 0.5), aumentando a regularização e a generalização do modelo. Esta estrutura permitiu melhorar a estabilidade do treino e obter resultados mais consistentes.

### 2.2.3 CNN Melhorado (Com Global Average Pooling)

A terceira abordagem manteve a base da CNN melhorada, substituindo a camada *Flatten* por uma camada de Global Average Pooling 2D, uma técnica frequentemente utilizada em arquiteturas modernas, como a ResNet. Esta camada reduz o número de parâmetros e foca a atenção do modelo nas regiões mais relevantes das imagens.

## 2.3 Processo de Treino e Validação

Os modelos foram treinados com recurso à função de perda categorical crossentropy, utilizando o otimizador Adam com uma taxa de aprendizagem inicial de 0.001. O número de épocas variou entre 10 e 30, dependendo da configuração testada, e o tamanho do *batch* foi fixado em 16 imagens.

Durante o treino, foi monitorizada a evolução da acurácia e da função de perda nos conjuntos de treino e validação, de forma a identificar o ponto ótimo de convergência. O treino foi interrompido sempre que se observava uma estabilização ou degradação do desempenho de validação.

## 2.4 Avaliação dos Modelos

A avaliação foi realizada com base em várias métricas de desempenho:

- Acurácia (Accuracy): proporção de classificações corretas;
- Sensibilidade (Recall): capacidade do modelo para identificar corretamente os casos positivos;
- Especificidade: capacidade de reconhecer corretamente os casos negativos;
- F1-Score: média harmónica entre precisão e sensibilidade;
- AUC (Área Sob a Curva ROC): medida global de desempenho do classificador.

Estas métricas foram calculadas para cada classe e posteriormente agregadas, permitindo comparar objetivamente o desempenho entre os diferentes modelos e configurações.

# Capítulo 3 – Análise de Resultados

## 3.1 Metodologia de Avaliação

Os resultados apresentados nesta seção resultam do treino e avaliação dos diferentes modelos de redes neurais convolucionais (CNN), aplicados ao conjunto de dados *Skin Diseases*. Foram testadas várias configurações, variando a resolução das imagens ( $64 \times 64$  e  $128 \times 128$  píxeis), o número de épocas (10 e 30) e a arquitetura da rede. A avaliação foi efetuada através das métricas acurácia, sensibilidade, especificidade, F1-score e AUC, complementadas pelas respectivas matrizes de confusão.

## 3.2 Resultados Quantitativos

**Tabela 1 – Resultados obtidos nos diferentes modelos e configurações**

Modelo	Resolução	Épocas	Acurácia	AUC	Akne (F1)	Benign (F1)	Ekzama (F1)	Enfeksiyonel (F1)	Malign (F1)
CNN Simples	$128 \times 128$	10	0.602	0.868	0.498	0.707	0.534	0.519	0.610
CNN Simples	$128 \times 128$	30	0.647	0.893	0.634	0.733	0.564	0.562	0.668
CNN Simples (todas as imagens)	$64 \times 64$	30	0.693	0.920	0.637	0.781	0.612	0.602	0.716
CNN Simples (todas as imagens)	$128 \times 128$	30	0.686	0.915	0.639	0.781	0.583	0.579	0.724
CNN Melhorado (Flatten)	$128 \times 128$	30	0.682	0.911	0.579	0.787	0.585	0.596	0.700
CNN Melhorado (GlobalAvgPooling)	$128 \times 128$	30	0.667	0.907	0.578	0.780	0.561	0.577	0.668

Experimento	Akne Sens	Akne Spec	Benign Sens	Benign Spec	Ekzama Sens	Ekzama Spec	Enfeksiyonel Sens	Enfeksiyonel Spec	Malign Sens	Malign Spec
10 épocas — resolução 128 — modelo simples	0.469	0.961	0.656	0.888	0.486	0.948	0.645	0.793	0.596	0.897
30 épocas — resolução 128 — modelo simples	0.596	0.973	0.659	0.921	0.520	0.950	0.708	0.799	0.668	0.904
Todas imagens — 30 épocas — resolução 64 — modelo simples	0.568	0.980	0.764	0.893	0.576	0.952	0.720	0.834	0.671	0.941
Todas imagens — 30 épocas — resolução 128 — modelo simples	0.571	0.980	0.757	0.899	0.543	0.950	0.689	0.830	0.697	0.934
30 épocas — resolução 128 — modelo avançado (Flatten)	0.438	0.993	0.773	0.893	0.553	0.948	0.757	0.807	0.641	0.945
30 épocas — resolução 128 — modelo avançado (GlobalAverage2D)	0.432	0.994	0.799	0.859	0.525	0.946	0.721	0.808	0.582	0.953

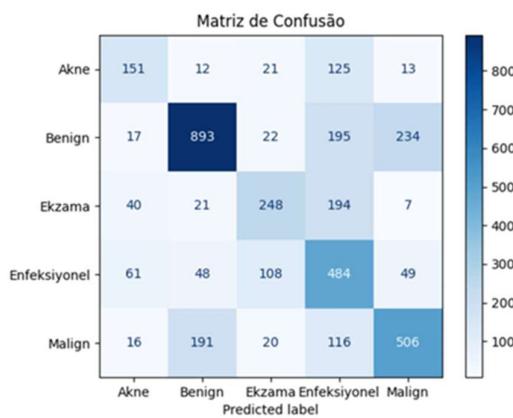
### 3.3 Matrizes de Confusão

A análise qualitativa foi complementada pela observação das matrizes de confusão obtidas para cada modelo.

A Tabela 2 apresenta as matrizes de confusão simplificadas (em percentagem de acertos por classe) para os modelos principais.

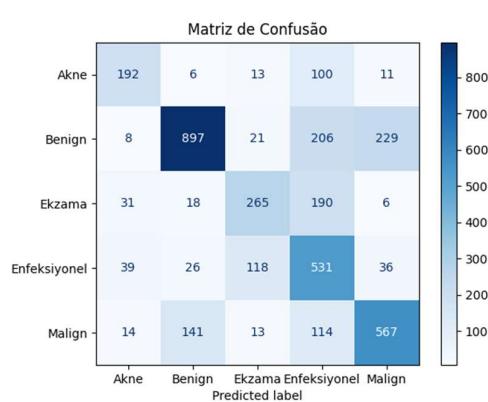
**Tendo em conta 2000 imagens por classe (Comparar Épocas):**

Para 10 épocas com 128 resolução no modelo simples:



A matriz de confusão do modelo simples treinado com 10 épocas demonstra um desempenho limitado, sendo o acerto na classe Akne o mais baixo (151) e a sua principal confusão a classe Enfeksiyonel (125). O erro clinicamente mais grave, classificar lesões Malignas como benignas (Falsos Negativos), ocorre em 191 casos.

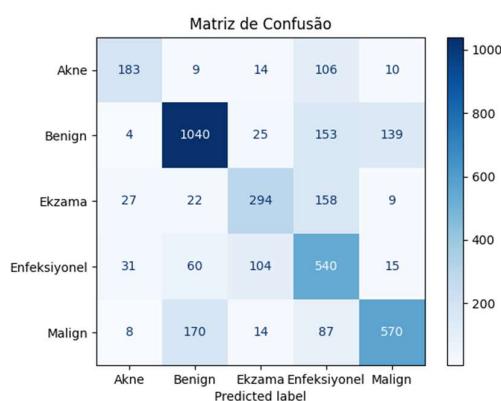
Para 30 épocas com 128 resolução no modelo simples:



Ao aumentar o treino para 30 épocas, verifica-se uma melhoria generalizada: o modelo acerta mais casos de Maligno (567 vs 506), e o erro de classificar Maligno como Benigno é reduzido para 141 casos, indicando que o treino prolongado aumentou a capacidade de discriminação da rede.

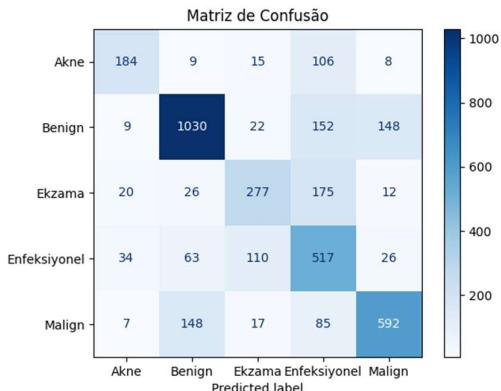
## Com todas as Imagens (Comparar resoluções e modelos):

Para 30 épocas com 64 resolução no modelo simples:



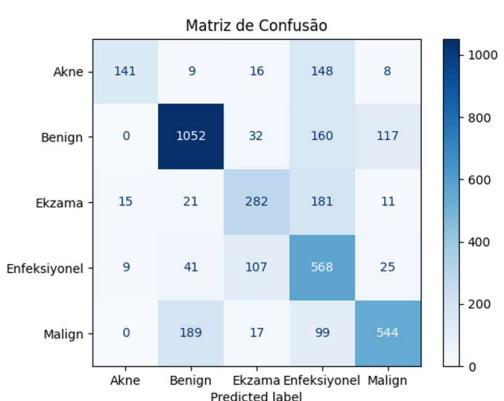
A comparação da matriz de confusão entre as duas resoluções do CNN Simples (30 épocas) revela um equilíbrio distinto. O modelo com resolução 64x64 alcançou o maior número de acertos na classe Benign (1040) e na classe Ekzama (294), mantendo o erro de classificar Maligno como Benigno em 170 casos.

resolução no modelo simples:



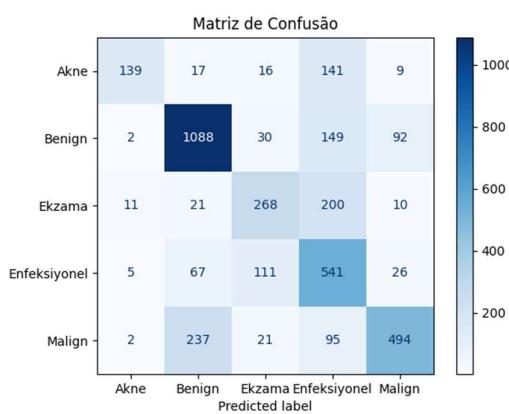
Por sua vez, o modelo com 128x128 demonstra uma capacidade ligeiramente superior na deteção de Malignos, com o maior número de acertos desta classe (592 vs 570) e uma menor taxa de Falsos Negativos críticos (148 vs 170 classificados como Benign) principal dificuldade em ambas a resolução continua a ser a distinção entre Akne/Ekzama e Enfeksiyonel.

Para 30 épocas com 128 resolução no modelo avançado (Flatten):



As matrizes dos modelos melhorados, que incluem *Batch Normalization* e *Dropout*, mostram padrões de confusão específicos. O modelo CNN Melhorado (Flatten) é o que apresenta o menor número de Falsos Positivos para a classe Malign (117), mas demonstra uma elevada confusão em Akne, classificando 148 casos como Enfeksiyonel.

Para 30 épocas com 128 resolução no modelo avançado (GlobalAverage2D):



Em contraste, o modelo CNN Melhorado (GlobalAvgPooling) é excelente na classificação de Benign (1088 acertos) e com o menor número de Falsos Positivos para Malign (92). No entanto, esta robustez no reconhecimento de negativos é compensada pelo seu ponto fraco mais perigoso: o maior erro de classificar lesões Malignas como Benignas (237 casos), o que o torna a opção mais arriscada.

### 3.4 Análise Crítica dos Resultados

A análise comparativa do desempenho dos modelos, resoluções e número de épocas é fundamental para determinar a configuração mais eficaz para a tarefa de classificação automática de doenças de pele, com base nas métricas de desempenho (Acurácia e AUC) e no F1-score por classe, conforme apresentado na Tabela 1.

#### 1. Comparação do Impacto das Épocas (CNN Simples, 128x128, 2000 amostras cada classe)

Aumentar o número de épocas de treino no modelo CNN Simples com resolução 128x128 levou a uma melhoria significativa no desempenho global e por classe.

Modelo	Resolução	Épocas	Acurácia	Melhoria
CNN Simples	128x128	10	0,602	—
CNN Simples	128x128	30	0,647	+0,045

Conclusões:

- O aumento de 10 para 30 épocas resultou num ganho de 0,045 na Acurácia, indicando que o modelo necessitou de mais iterações para convergir e aprender as características relevantes do dataset.
- Esta melhoria foi acompanhada por ganhos em todas as classes, sendo o F1-score para Akne o que mais se beneficiou, passando de 0,498 para 0,634.

## 2. Comparação do Impacto da Resolução (CNN Simples, 30 Épocas, Imagens todas)

A mudança na resolução das imagens para o treino com 30 épocas no modelo CNN Simples demonstrou que a resolução mais baixa obteve o melhor desempenho global

Modelo	Resolução	Épocas	Acurácia	AUC
CNN Simples	128x128	30	0,686	0,915
CNN Simples	64x64	30	0,693	0,920

Conclusões:

- As duas resoluções apresentaram um desempenho global muito próximo. No entanto, a resolução de 64x64 alcançou a melhor Acurácia (0,693) e o melhor AUC (0,920).
- A diferença nas métricas globais é marginal (apenas 0,007 na Acurácia e 0,005 no AUC a favor de 64x64).
- Mas tendo em conta o problema a tratar que é doenças de pele seria mais eficaz usar imagens com mais resolução pois é necessário buscar características específicas para a deteção da classe respetiva (neste modelo por acaso a menor resolução deu melhor, mas com modelos melhores, mais resolução poderá influenciar bastante na decisão).

## 3. Comparação de Modelos/Arquiteturas (30 Épocas, 128x128)

A comparação entre o modelo Simples e as duas versões do Modelo Melhorado (avançado) na resolução 128x128 e 30 épocas avalia o impacto das camadas de *Batch Normalization* e *Dropout*.

Modelo	Resolução	Épocas	Acurácia	AUC
CNN Simples	128x128	30	0,686	0,915
CNN Melhorado ( <i>Flatten</i> )	128x128	30	0,682	0,911
CNN Melhorado ( <i>GlobalAvgPooling</i> )	128x128	30	0,667	0,907

Conclusões:

- O CNN Simples com o novo dado de Acurácia (0,686) e AUC (0,915) apresenta o melhor desempenho global nesta resolução de 128x128.

- Apesar disso, os Modelos Melhorados alcançaram *F1-scores* superiores para a classe Benign (0,787 e 0,780 vs 0,733 no Simples), e o *CNN Melhorado (Flatten)* obteve o melhor *F1-score* para a classe Malign (0,700).
- Isto sugere que a adição de técnicas de regularização (*Dropout* e *Batch Normalization*) nos modelos Melhorados, embora não tenha maximizado a Acurácia global, melhorou a robustez e a capacidade de classificação das classes individuais mais críticas.
- O modelo CNN Melhorado (Flatten) obteve o desempenho mais equilibrado entre as arquiteturas avançadas, superando ligeiramente o modelo que utilizou *Global Average Pooling*.

## Conclusão

Os resultados experimentais estabelecem uma base promissora para a classificação automática de doenças de pele. A *performance* global mais eficiente foi obtida com o modelo CNN Simples utilizando a resolução mais baixa, indicando que este nível de detalhe da imagem foi suficiente para a arquitetura base. Contudo, a análise das matrizes de confusão sublinha o risco clínico associado aos Falsos Negativos na deteção de lesões Malignas, uma vez que um dos modelos avançados demonstrou a maior propensão para classificar Maligno como Benigno. A próxima fase do projeto deverá focar-se na otimização do algoritmo através de estratégias como o balanceamento de classes e o aprofundamento das arquiteturas (*transfer learning*), visando maximizar a Sensibilidade e o *F1-score* para as classes críticas, nomeadamente a Maligna, garantindo a robustez necessária para o apoio ao diagnóstico médico.