



**TÉCNICO**  
LISBOA

## **Detecção de Melanoma usando Métodos de Aprendizagem Profunda**

**Joana Barbosa Miranda de Garcia Baleiras**

Tese para obter o Grau de Mestrado em

### **Engenharia Electrotécnica e de Computadores**

Orientador(es): Dr. Ana Catarina Fidalgo Barata  
Prof. Jorge dos Santos Salvador Marques

#### **Júri**

Presidente: Prof. João Fernando Cardoso Silva Sequeira

Orientador: Dr. Ana Catarina Fidalgo Barata

Vogal: Prof. Maria Margarida Campos da Silveira

**Junho 2019**



# Agradecimentos

Esta dissertação de mestrado representa uma fase muito importante e gratificante na minha formação académica, que só foi possível concretizá-la graças à paciência, ao apoio e ao conhecimento de um conjunto de pessoas, às quais quero exprimir um especial agradecimento.

Aos meus orientadores, expresso o meu mais sincero agradecimento por todo o apoio, disponibilidade, ajuda, conhecimento e (muita) paciência ao longo deste percurso, uma vez que foram decisivos e preciosos para a realização desta dissertação. Esta tese foi parcialmente apoiada pelo projecto FCT PTDC/EEIPRO/0426/2014.

Aos meus grandes amigos que este curso me deu a oportunidade de conhecer e que me acompanharam e apoiaram ao longo desta jornada, o meu gigante obrigado. Por não me fazerem desistir, por me fazerem rir e manter a motivação, pelos conselhos sábios na hora certa, pelos brindes, pelas noitadas, pelos jantares, pelas sessões de "terapia" aos Sábados, pelas tardes de "estudo", pelos serões no Beckett's e por me levantarem do chão (literalmente, nalgumas situações): Luís Rodrigues, Ana Rita Córias, Catarina Ferreira, Sara Barroso, Miguel Vasconcelos, Gonçalo Pais, Eduardo Creso, Ana Rita Machado, Mónica Paiágua e João Rodrigues.

E, por fim, um profundo e enorme obrigado à minha irmã, Mafalda Baleiras, ao meu avô Vivi e, em especial, aos meus pais, Rui Nuno Baleiras e Ana Cristina Miranda, pelos valores de vida que me transmitiram, por acreditarem sempre em mim, por toda a paciência nos momentos mais difíceis, por todo o carinho e apoio. A eles, as palavras nunca serão suficientes para agradecer. Porém, espero que o culminar desta etapa possa retribuir, de alguma forma, todo o amor e toda a dedicação que constantemente me proporcionaram e por tudo o que me ensinaram até hoje. Muito obrigada.



# Declaração

Declaro que o presente documento é um trabalho original da minha autoria e que cumpre todos os requisitos do Código de Conduta e Boas Práticas da Universidade de Lisboa.



## **Abstract**

Melanoma is the deadliest form of skin cancer due to its propensity to metastasize. Therefore, early detection is crucial to guarantee a successful treatment. Computer-Aided Diagnosis (CAD) systems have been developed to provide an automated melanoma detection. This thesis proposes a CAD system based on deep neural networks to distinguish between melanoma and benign skin lesions. The novelty lies in the use and comparison of three different Convolutional Neural Networks (CNN), namely AlexNet, VGG and ResNet. These CNN were trained using transfer learning and data augmentation. Additionally, a thorough assessment of the influence of dropout and the choice of learning rate is performed. The image dataset was extracted from the ISIC 2017 Challenge. The original dataset contains 2000 training images (1626 benign lesions and 374 melanomas), 150 validation images (120 benign lesions and 30 melanomas) and 600 test images (483 benign lesions and 117 melanomas). The training set was then augmented and contained 1626 benign examples and 1870 malign examples. For a test set of 600 images, ResNet reached SE of 79%, SP of 60%, BACC of 69% and ROC-AUC of 77%. Hence, the proposed system demonstrates the potential of CNN in CAD systems for automated melanoma detection.

**Keywords:** Melanoma, CAD system, automated melanoma detection, CNN, AlexNet, VGG, ResNet, transfer learning, data augmentation, ISIC Challenge





## Resumo

O melanoma é o cancro da pele mais mortífero. A detecção precoce deste tipo de cancro é essencial para a cura do paciente. Tem havido um esforço para o desenvolvimento de sistemas automatizados para a detecção e a análise de lesões da pele, denominados por sistemas CAD. Este trabalho apresenta um sistema CAD capaz de distinguir melanomas de não-melanomas, utilizando redes neuronais profundas. A principal contribuição é a comparação de três redes neuronais convolucionais (CNN) em separado, AlexNet, VGG e ResNet, nunca antes utilizadas nem comparadas no mesmo estudo de detecção automática de melanomas. Estas CNN foram treinadas recorrendo a transferência de aprendizagem e ao aumento de dados. Adicionalmente, foi avaliado o impacto de dois hiper-parâmetros relevantes para o treino das redes: o dropout e o ritmo de aprendizagem. A base de dados utilizada foi a do concurso ISIC da edição de 2017. O conjunto original de imagens dermatoscópicas consistia em 2000 imagens de treino (1626 lesões benignas e 374 malignas), 150 de validação (120 lesões benignas e 30 malignas) e 600 de teste (483 lesões benignas e 117 malignas). O conjunto de treino foi submetido *a posteriori* à técnica de aumento de dados, contendo 1626 lesões benignas e 1870 malignas. Para um conjunto de teste de 600 imagens, a ResNet atingiu 79% de SE, 60% de SP e 77% de ROC-AUC. O sistema desenvolvido realça o potencial das CNN usadas em sistemas CAD para a detecção automática de melanoma.

**Keywords:** Melanoma, sistemas CAD, detecção automática de melanomas, CNN, AlexNet, VGG, ResNet, transferência de aprendizagem, aumento de dados, concurso ISIC



# Conteúdos

<b>Lista de Tabelas</b>	<b>xi</b>
<b>Lista de Figuras</b>	<b>xiii</b>
<b>Acrónimos</b>	<b>xvii</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Motivação . . . . .	1
1.2 Diagnóstico Clínico . . . . .	2
1.3 Aprendizagem Automática . . . . .	3
1.4 Sistemas CAD para detecção de melanomas . . . . .	3
1.5 Objectivos e Estrutura da Tese . . . . .	5
<b>2 Detecção de Melanomas</b>	<b>7</b>
2.1 Introdução . . . . .	7
2.2 Lesões Cutâneas . . . . .	8
2.3 Estruturas Dermatoscópicas . . . . .	10
2.4 Métodos Clínicos . . . . .	11
2.4.1 Regra ABCD . . . . .	11
2.4.2 Lista dos Sete-Pontos . . . . .	14
2.5 Métodos Automáticos - <i>CAD systems</i> . . . . .	15
2.5.1 Pré-processamento . . . . .	16
2.5.2 Segmentação . . . . .	16
2.5.3 Extracção e selecção de características . . . . .	17
2.5.4 Classificação . . . . .	17
2.5.5 Revisão da Literatura . . . . .	17
<b>3 Redes Neurais Convolucionais</b>	<b>21</b>
3.1 Introdução às redes neuronais artificiais . . . . .	21
3.1.1 Inspiração Biológica . . . . .	21
3.1.2 Conceitos-chave . . . . .	23
3.1.2.1 Passo propagação directa . . . . .	24

3.1.2.2	Passo de retropropagação . . . . .	25
3.2	Arquitectura de uma CNN . . . . .	27
3.2.1	Camada de convolução . . . . .	28
3.2.2	Camada de <i>pooling</i> . . . . .	30
3.2.3	Camada totalmente interligada . . . . .	30
3.3	Concurso ILSVRC . . . . .	31
3.3.1	AlexNet . . . . .	31
3.3.2	VGG . . . . .	32
3.3.3	ResNet . . . . .	33
<b>4</b>	<b>Formulação do problema e sistema proposto</b>	<b>37</b>
4.1	Formulação do Problema . . . . .	37
4.2	Arquitectura do Sistema Proposto . . . . .	37
4.2.1	Pré-processamento . . . . .	38
4.2.2	Aumento de dados . . . . .	39
4.2.3	Classificação . . . . .	40
4.2.4	Treino . . . . .	41
<b>5</b>	<b>Resultados Experimentais</b>	<b>43</b>
5.1	Base de Dados . . . . .	43
5.2	Métricas de avaliação de desempenho . . . . .	44
5.3	Treino das CNN . . . . .	45
5.4	Escolha de hiper-parâmetros . . . . .	46
5.4.1	AlexNet . . . . .	47
5.4.1.1	Técnica de Aumento de Dados . . . . .	49
5.4.1.2	Escolha dos outros hiper-parâmetros . . . . .	49
5.4.2	VGG . . . . .	49
5.4.3	ResNet . . . . .	51
5.4.3.1	Escolha de hiper-parâmetros . . . . .	52
5.5	Avaliação do Desempenho: Comparação das três CNN . . . . .	52
<b>6</b>	<b>Conclusão e Trabalho Futuro</b>	<b>57</b>
	<b>Bibliografia</b>	<b>59</b>

# Lista de Tabelas

2.1	Conversão entre as diferentes categorias de profundidade do critério de Breslow e os níveis do indicador T. Adaptado de [1]. . . . .	9
2.2	Descrição de cada estágio do melanoma. Adaptado de [2], [3], [1]. . . . .	10
2.3	Principais estruturas dermatoscópicas para lesões melanocíticas. Adaptado com base em [4, 5].	11
2.4	Descrição de cada um dos quatro critérios usados pela regra ABCD e o intervalo de valores que cada um destes critérios pode assumir na fórmula TDS. Adaptado de [6]. . . . .	14
2.5	Conversão entre os possíveis valores de TDS de uma lesão e o diagnóstico provável. Adaptado de [6]. . . . .	14
2.6	Identificação e classificação individual de cada critério da lista dos sete-pontos. Adaptado de [7].	15
2.7	Apresentação de três trabalhos de investigação baseados na arquitetura clássica de reconhecimento de padrões. . . . .	18
2.8	Apresentação de três trabalhos de investigação baseados na arquitetura <i>end-to-end</i> . . . . .	19
5.1	Base de dados de imagens dermatoscópicas ISIC 2017. . . . .	43
5.2	Número de épocas definido para cada CNN utilizada. . . . .	46
5.3	Avaliação da AlexNet no conjunto de validação, sem o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem. . . . .	47
5.4	Avaliação da AlexNet no conjunto de validação, sem o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem. . . . .	47
5.5	Avaliação da AlexNet no conjunto de validação, sob o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem. . . . .	47
5.6	Avaliação da AlexNet no conjunto de validação, sob o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem. . . . .	48
5.7	Avaliação da VGG no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem. . . . .	50
5.8	Avaliação da VGG no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem. . . . .	50
5.9	Avaliação da ResNet no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem. . . . .	51

5.10 Avaliação da ResNet no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem. . . . .	51
5.11 Avaliação da AlexNet no conjunto de teste com a melhor configuração da rede obtida na fase de treino. . . . .	53
5.12 Avaliação da VGG no conjunto de teste com a melhor configuração da rede obtida na fase de treino. . . . .	53
5.13 Avaliação da ResNet no conjunto de teste com a melhor configuração da rede obtida na fase de treino. . . . .	53

# Lista de Figuras

1.1	Lesão benigna (queratose seborreica), à esquerda em a), e melanoma, à direita em b); lesão benigna (nevo), à esquerda em c), e melanoma, à direita em d). . . . .	2
1.2	Arquitectura de um modelo CAD convencional. . . . .	4
1.3	Modelo <i>end-to-end</i> de um sistema CAD. . . . .	5
2.1	Diagrama que ilustra as duas categorias de lesões do cancro da pele [8]. . . . .	8
2.2	Rede pigmentada: típica à esquerda em a) e atípica à direita em b). Extraído de [9]. . . . .	12
2.3	Pontos: regulares à esquerda em a) e irregulares à direita em b). Extraído de [9]. . . . .	12
2.4	Glóbulos regulares à esquerda em a) e cicatrizes regulares à direita em b). Extraído de [9]. . . . .	12
2.5	Mancha azulada (benigna) à esquerda em a) e véu azul-esbranquiçado à direita em b). Extraído de [9]. . . . .	12
2.6	Estrutura de regressão. Extraído de [9]. . . . .	13
2.7	Diagnóstico da lesão: melanoma, baseado na regra ABCD. Classificação final, TDS (7,6 pontos): A - 2 eixos de simetria (2 pontos $\times$ 1,3); B - 5 segmentos (5 pontos $\times$ 0,1); C - 5 cores [castanho claro, castanho escuro, cinzento azulado, preto e branco] (5 pontos $\times$ 0,5); D - 4 estruturas [áreas homogéneas, cicatrizes, pontos, glóbulos] (4 pontos $\times$ 0,5). Extraído de [6]. . . . .	13
2.8	Diagnóstico da lesão: melanoma, baseado na lista dos sete-pontos. Classificação final (7 pontos): rede pigmentada atípica (2 pontos), véu azul-esbranquiçado (2 pontos), pontos/glóbulos irregulares (1 ponto), pigmentação irregular (1 ponto) e cicatrizes irregulares (1 ponto). Adaptado de [7]. . . . .	15
3.1	Comparação entre o processo biológico e o artificial, adaptado de [10]. $n$ neurónios ( $x_1, x_2, \dots, x_n$ ) a transmitir um sinal ao neurónio $y$ , através das suas sinapses ( $w_1, w_2, \dots, w_n, b$ ). . . . .	22
3.2	Passo de propagação directa de uma arquitectura MLP (adaptado de [11]). . . . .	24
3.3	Passo de retropropagação de uma arquitectura MLP, extraído de [11]. . . . .	26
3.4	Arquitectura de uma CNN simples <sup>1</sup> . . . . .	28
3.5	Representação de um filtro de tamanho $3 \times 3 \times 3$ a convolver uma imagem de entrada $5 \times 5 \times 3$ , resultando um mapa de características de $3 \times 3 \times 1$ . Esta convolução utiliza um passo de dois píxeis ( $S = 2$ ) e um zero ( $P = 1$ ). Adaptado de [12]. . . . .	29

3.6	Operação de <i>max-pooling</i> , em que o mapa de características, à esquerda, é reduzido através de um filtro de dimensão $2 \times 2$ e passo de dois pixéis, produzindo o mapa de característica à direita de dimensão $2 \times 2$ . Adaptado de [12]. . . . .	30
3.7	Arquitetura da AlexNet, adaptado de [13]. . . . .	32
3.8	Representação da arquitetura VGG-16 ("ti" representa uma camada totalmente interligada). Adaptado de [14]. . . . .	33
3.9	Estrutura de um bloco residual [15]. . . . .	34
3.10	Arquitetura do modelo ResNet-50, adaptada de [15]. Os números indicados ao lado de cada bloco (e.g., $x3$ ) indicam o número de vezes em que esse bloco se repete. . . . .	34
4.1	Diagrama do sistema desenvolvido. . . . .	38
4.2	Imagem dermatoscópica original, à esquerda em a), e a mesma imagem recortada com base na sua caixa delimitador rectangular, à direita em b). . . . .	39
4.3	Aumento de dados de treino. A primeira imagem (mais à esquerda, a)) representa a imagem original de uma lesão maligna. As restantes imagens representam as inversões horizontal, b), e vertical, c), e as rotações de $90^\circ$ , d), e $180^\circ$ , e). . . . .	40
5.1	Exemplos extraídos da base de dados ISIC 2017. Em cima, imagem dermatoscópica maligna do conjunto de treino (à esquerda) e respectiva segmentação binária (à direita); no centro, imagem dermatoscópica benigna do conjunto de validação (à esquerda) e respectiva segmentação binária (à direita) e imagem dermatoscópica benigna do conjunto de teste (à esquerda) e respectiva segmentação binária (à direita). . . . .	44
5.2	Convergência do erro balanceado para a rede AlexNet no conjunto de treino aumentado. . . .	46
5.3	A melhor configuração obtida da AlexNet no conjunto de validação com aumento de dados, <i>dropout</i> e variação do ritmo de aprendizagem, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	48
5.4	A melhor configuração obtida da AlexNet no conjunto de validação, mas sem o efeito de DA, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	48
5.5	A melhor configuração obtida da AlexNet no conjunto de validação, mas sem <i>dropout</i> , dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	48
5.6	A melhor configuração obtida da AlexNet no conjunto de validação, mas sem a variação do ritmo de aprendizagem, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	49
5.7	A melhor configuração obtida da VGG no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	50
5.8	A segunda melhor configuração obtida da VGG no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	50
5.9	A melhor configuração obtida da ResNet no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	51



5.10	A segunda melhor configuração obtida da ResNet no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b). . . . .	52
5.11	Curva ROC da melhor configuração obtida da AlexNet no conjunto de teste, $ROC - AUC = 68\%$ .	53
5.12	Curva ROC da melhor configuração obtida da VGG no conjunto de teste, $ROC - AUC = 75\%$ .	54
5.13	Curva ROC da melhor configuração obtida da ResNet no conjunto de teste, $ROC - AUC = 77\%$ .	54
5.14	Exemplos de melanomas: correctamente classificados (à esquerda), incorrectamente classificados (à direita). . . . .	55
5.15	Exemplos de não-melanomas: correctamente classificados (à esquerda), incorrectamente classificados (à direita). . . . .	56



# Acrónimos

- ACS** *American Cancer Society*, Sociedade Americana do Cancro. 1
- AJCC** *American Joint Committee on Cancer*. 9
- BACC** *Balanced accuracy*, precisão balanceada. 45
- BCC** *Basal Cell Carcinoma*, cancro basocelular. 7
- CAD** *Computer-Aided Diagnosis*. 2, 10
- CNN** *convolutional neural network(s)*, rede(s) neuronal(ais) convolucional(ais). vii, 3, 57
- DA** *Data Augmentation*, técnica de aumento de dados. 16, 46
- DL** *Deep Learning*, Aprendizagem Automática Profunda. 3
- FN** *False Negative*, falso negativo. 44
- FP** *False Positive*, falso positivo. 44
- GPU** *Graphics Processing Unit(s)*. 3, 57
- IARC** *International Agency for Research on Cancer*. 1
- ILSVRC** *ImageNet Large-Scale Visual Recognition Challenge*. 2, 40
- ISIC** *International Skin Imaging Collaboration*. 4
- MLP** *Multi-Layer Perceptron*, modelo do perceptrão multicamada. 3, 22
- ROC** *Receiver Operating Characteristic*, característica de operação do receptor. 45
- ROC-AUC** *Area Under the ROC Curve*, área por baixo da curva ROC. 45
- SCC** *Squamous Cell Carcinoma*, cancro das células escamosas. 7
- SE** *Sensitivity*, sensibilidade. 45

**SP** *Specificity*, especificidade. 45

**TDS** *Total Dermatoscopy Score*. 12, 13

**TN** *True Negative*, verdadeiro negativo. 44

**TP** *True Positive*, verdadeiro positivo. 44

# 1

## Introdução

### 1.1 Motivação

O cancro da pele é um problema importante de saúde pública [16]. A *International Agency for Research on Cancer* (IARC)<sup>1</sup> previu que em 2018 surgiriam em todo o mundo cerca de 1,3 milhões de casos novos de cancro da pele e que ocorreriam 125 mil óbitos em consequência directa desta patologia [17].

De entre os vários tipos de cancro da pele, o melanoma merece destaque por ser o mais mortífero de todos. A nível mundial, este tumor é responsável por 75% das mortes provocadas por cancro da pele [18, 19], apesar de representar apenas 1% do total de cancros da pele [20]. Na Austrália, um dos países em que este tumor tem maior incidência, as previsões para 2018 da IARC afirmavam que 14260 pessoas seriam diagnosticadas com melanoma (576 indivíduos por milhão de habitantes), das quais 1602 iriam falecer [?] (65 por milhão de habitantes). Na Europa, o melanoma é o responsável por mais de 20000 mortes por ano [21]. No caso concreto de Portugal, previam-se 1320 novos casos de melanoma (128 por milhão de habitantes) e 356 mortes em 2018 [22] (35 por milhão de habitantes). Nos Estados Unidos, a IARC previa que em 2018 seriam diagnosticados 71434 novos casos de melanoma (218 por milhão de habitantes) e ocorreriam 9491 vítimas mortais (29 em 1000000 habitantes) [23].

O melanoma é o neoplasma que mais se metastiza, começando apenas como um simples sinal de pele. Porém, se for detectado precocemente (melanoma de estágio I ou melanoma *in situ*), o melanoma é tratado através de uma simples excisão da zona de pele lesada [20, 24, 18]. No estágio mais avançado da doença, com evidência de metástases disseminadas pelo corpo, a *American Cancer Society*, Sociedade Americana do Cancro (ACS), estima que somente 20% desses casos sobrevive aos primeiros cinco anos após o diagnóstico [25].

Deste modo, é essencial identificar o melanoma tão precocemente quanto possível. Só assim será possível travar a aceleração na incidência da doença a que se tem assistido nos últimos 30 anos em todo o mundo [26]. Para tal, são precisos dermatologistas experientes neste domínio, bem como equipamentos e investigação

---

<sup>1</sup>Trata-se da agência da Organização Mundial de Saúde especializada em assuntos cancerígenos.

clínica e tecnológica que auxiliem aqueles na produção de diagnósticos.

## 1.2 Diagnóstico Clínico

Clinicamente, o melanoma é identificado por observação visual da pele ou através de equipamentos, nomeadamente os dermatoscópios. Estes aparelhos permitem a visualização detalhada das estruturas internas de cada lesão cutânea, com um grau de detalhe não perceptível a olho nu [27]. Quando persistem dúvidas relativas à malignidade de uma lesão, o médico, por vezes, aconselha uma análise histológica.

Todavia, a análise clínica é um procedimento subjectivo, uma vez que depende da experiência do dermatologista e da sua acuidade visual. Para além destes dois factores, a semelhança morfológica entre lesões cutâneas benignas e o melanoma também dificulta a precisão do diagnóstico clínico (ver Figura 1.1).

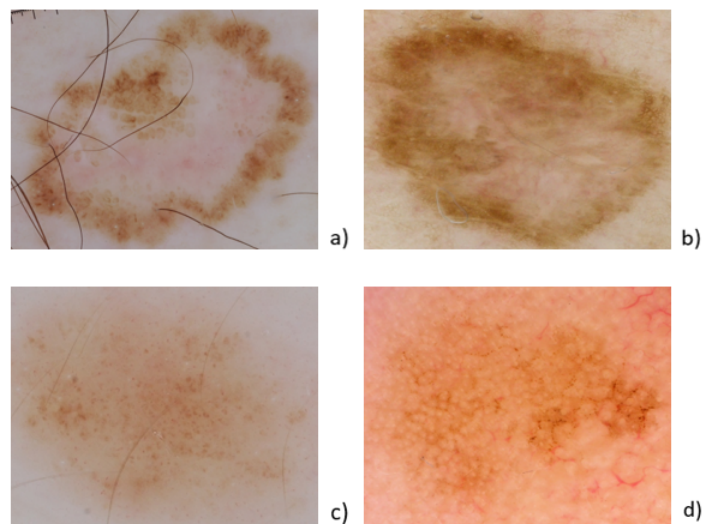


Figura 1.1: Lesão benigna (queratose seborreica), à esquerda em a), e melanoma, à direita em b); lesão benigna (nevo), à esquerda em c), e melanoma, à direita em d).

Por isso, tem havido um esforço para o desenvolvimento de sistemas automatizados para a detecção e a análise de lesões da pele (em inglês, denominados por *Computer-Aided Diagnosis (CAD) systems*) [28, 29, 30, 31, 32, 33]. Muitos destes sistemas baseiam-se em métodos de Aprendizagem Automática, como, por exemplo, as redes neurais, máquinas de vectores de suporte, florestas aleatórias [34, 35, 36]. No caso das redes neurais, estas têm conhecido um sucesso sem precedentes em problemas de análise de imagem e de vídeo, onde têm ultrapassado o estado-da-arte [37, 38, 39], e, até mesmo no caso do diagnóstico do cancro da pele, há um interesse recente por estes métodos [16, 32, 33]. Os sistemas CAD desenvolvidos com redes neurais, identificados nestas referências bibliográficas, têm apresentado resultados promissores na distinção entre lesões benignas e melanoma, equiparando-se aos do diagnóstico clínico de dermatologistas competentes [32]. Por esta razão, a presente dissertação interessa-se por analisar e comparar no mesmo estudo o desempenho de três modelos neuronais na tarefa de detecção de melanoma. Estes três modelos neuronais foram casos de sucesso em tarefas de detecção e classificação de objectos no concurso *ImageNet Large-Scale Visual Recognition Challenge (ILSVRC)* (mencionado na Secção seguinte) [13, 14, 15].

## 1.3 Aprendizagem Automática

Desde o início da computação que se tem procurado desenvolver máquinas que aprendam a partir da experiência, inspiradas no cérebro humano, e tomem decisões "inteligentes" sozinhas. Este conceito de aprendizagem automática [40] surgiu por volta dos anos 40 do século XX. McCulloch e Pitts (1943) foram os pioneiros no desenvolvimento de um modelo matemático simplificado de um neurónio [41]. As décadas seguintes são pautadas pelo esforço em interligar centenas de neurónios em redes neuronais (do qual nasceu o modelo do perceptrão multicamada, *Multi-Layer Perceptron*, modelo do perceptrão multicamada (MLP), e desenvolver algoritmos que permitam aprender os seus parâmetros (denominados por pesos) a partir de dados [42].

Paralelamente, também foram desenvolvidos métodos estatísticos com o mesmo objectivo, mas sem inspiração biológica, e.g., discriminante de Fisher (Fisher, 1936), árvores de decisão (Hunt *et al.*, 1966; Breiman *et al.*, 1984; Quinlan, 1986), método de k-vizinhos mais próximos (Cover e Hart, 1967), Adaboost (Freund e Schapire, 1994) e máquinas de vectores de suporte (Vapnik e Cortes, 1995). Estes métodos, em particular as máquinas de vectores de suporte, tornaram-se os métodos favoritos para tarefas de classificação de dados durante as décadas de 90 e de 2000 [43].

Na última década, tem-se assistido a uma revolução nos métodos neuronais sob o nome de Aprendizagem Profunda (*Deep Learning*, Aprendizagem Automática Profunda (DL)). Esta revolução foi impulsionada pelo desafio ILSVRC<sup>2</sup> na edição de 2012, em que a equipa vencedora desenvolveu um método neuronal, em que as primeiras camadas se baseavam em operações de convolução com filtros, denominado por rede neuronal convolucional (CNN) [13]. Esta rede, designada por AlexNet, na altura era muito flexível, com várias camadas escondidas e cerca de 60 milhões de parâmetros (pesos). Depois deste sucesso, seguiram-se outros usando também CNNs. Nos anos seguintes o concurso ILSVRC foi ganho sucessivamente por redes convolucionais com um número cada vez maior de camadas escondidas. A revolução das redes neuronais assenta, então, na combinação de três factores: grandes quantidades de dados (milhões de imagens, *Big Data*), maior capacidade de cálculo computacional (GPU<sup>3</sup>) e avanços na arquitectura das redes e nos algoritmos de treino.

Diferentes tipos de redes profundas têm sido aplicadas na resolução de problemas complexos em áreas variadas. Alguns exemplos são: detecção de fraude em comunicações móveis [45], tradução automática de textos como o Google Tradutor [46], reconhecimento de fala (conversão de fala em texto) [47], condução automática de veículos autónomos [48], reconhecimento de objectos em imagens [49] e até em redes sociais, como o Facebook [50]. E, como já referido, esta tecnologia também já está a ser activamente utilizada na detecção do cancro da pele [16, 32, 33].

## 1.4 Sistemas CAD para detecção de melanomas

Os sistemas CAD aplicadas à detecção de melanoma, mencionados na Secção 1.1, baseiam-se em técnicas de processamento de imagem e de aprendizagem automática para avaliar de lesões cutâneas. O funciona-

---

<sup>2</sup>O concurso ILSVRC nasceu em 2010 e, desde então, é uma referência na classificação e detecção de objectos, através da criação de uma base de dados com milhões de imagens e respectivas categorias de objectos [44]

<sup>3</sup>*Graphics Processing Unit(s)* (GPU) é uma unidade de processamento altamente especializada no processamento de imagens, aumentando o desempenho computacional e a capacidade de armazenamento de dados, face aos CPUs, reduzindo o tempo de cálculo em uma a duas ordens de grandeza.

mento destes sistemas está descrito na Figura 1.2. Começa com o carregamento de um conjunto de imagens dermatoscópicas digitais, a base de dados de casos passados, que são usadas para treinar o sistema. Antes de serem processadas, as imagens tipicamente sofrem algumas alterações (*e.g.*, remoção de pelos e correcção da luminosidade) para melhorar a qualidade da imagem. A esta transformação dá-se o nome de *pré-processamento de imagem*, à qual se seguem então as três principais fases da arquitectura CAD: segmentação de imagem, extracção e selecção de características e classificação [28, 51, 8, 31].



Figura 1.2: Arquitectura de um modelo CAD convencional.

A segmentação consiste em dividir a imagem em duas regiões (lesão e pele saudável), explorando vários tipos de critérios (cor, textura, contornos). Esta operação pode ser realizada através de vários tipos de algoritmos, por exemplo: limiarização adaptativa, modelos deformáveis, classificação de pixels, sistemas difusos [52, 53].

A fase de extracção de características baseia-se na identificação das propriedades cromáticas e morfológicas que melhor descrevem a lesão cutânea [54]. Para esta fase, a maioria dos sistemas CAD é inspirada na regra ABCD (*ABCD rule*) [55] utilizada pelos dermatologistas na avaliação de lesões cutâneas, que consiste na análise de quatro aspectos: assimetria, fronteira, cor e estruturas diferenciais, relacionadas com textura [56].

O classificador recebe o vector de características e calcula uma decisão binária (melanoma/não-melanoma).

De notar, que as imagens usadas como base de dados para o treino do classificador (conjunto de treino) são previamente segmentadas e anotadas por dermatologistas experientes. A anotação das imagens consiste na descrição do tipo de lesão. Por este motivo, estes modelos enfrentam duas grandes dificuldades. Necessitam de um conjunto grande (centenas ou milhares) de imagens dermatológicas e também requerem que o conjunto inicial de imagens seja previamente analisado e anotado pelos médicos [57].

Até 2016, cada grupo de investigação desenvolvia a sua própria base de dados, havendo algumas tentativas de criação de bases de dados públicas disponíveis para todos (PH2 [58]). Desde 2016, a associação *International Skin Imaging Collaboration* (ISIC) [59] tem procurado superar estas dificuldades ao disponibilizar bases de dados públicas com muitas centenas de imagens dermatológicas, previamente anotadas, destinadas à detecção automática de lesões da pele (melanoma e outros tipos), bem como de estruturas dermatoscópicas. Isso permitiu a aplicação de sistemas CAD com redes neuronais profundas na detecção de melanomas. O método neuronal favorito é o das redes neuronais convolucionais (em inglês, *convolutional neural networks*, CNN) por permitirem uma fase de treino mais rápida e uma melhor generalização [11]. As CNN são desenhadas para processar imagens. Uma imagem a cores consiste em três matrizes 2D, em que cada vector representa um canal de cor e cada entrada da matriz representa a intensidade de um pixel da imagem. Este tipo de redes aplica iterativamente um conjunto de transformações lineares e invariantes no espaço (convoluções), seguido de não linearidades, com o intuito de extrair características relevantes, de forma automática. Estas redes são capazes de classificar novas imagens, às quais não teve acesso durante a fase de treino.

As redes neuronais podem ser usadas como classificadores na arquitectura da Figura 1.2, recebendo o



vector de características extraídas da imagem.

Contudo, tendem a ser usadas numa estratégia mais ambiciosa (ver Figura 1.3) que é designada por "end-to-end", em que compete à rede realizar todas as operações necessárias para processar uma imagem de entrada e produzir uma decisão [32].

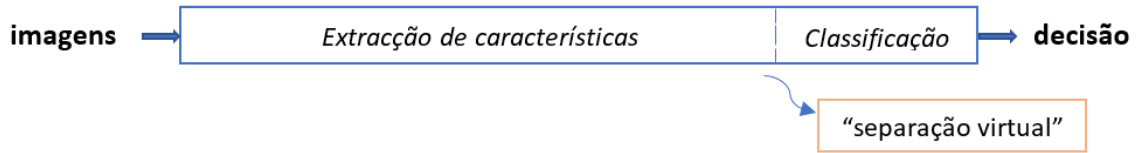


Figura 1.3: Modelo *end-to-end* de um sistema CAD.

Em ambas as arquiteturas a rede tem que ser treinada a partir de um conjunto de treino de imagens anotadas por especialistas. No entanto, na primeira arquitetura a entrada é um vector de características de reduzida dimensão quando comparada com a arquitetura *end-to-end*, cuja entrada da rede é uma imagem com dezenas de milhares de pixels. A rede tem que ser muito mais flexível, neste caso, e a tarefa de treino é muito mais exigente de forma a que a tarefa de classificação no conjunto de teste seja fiável. É aqui que as redes CNN podem dar um contributo importante.

Actualmente, existem já vários modelos de CNN aplicados ao cancro da pele [60, 16, 32, 33]. Porém, dada a enorme diversidade de CNN existente, coloca-se a questão da escolha: qual a arquitetura que melhor se adapta à detecção de melanomas? A presente dissertação visa, assim, estudar a aplicação de três arquiteturas CNN vencedoras em edições diferentes do concurso ILSVRC (a saber: AlexNet, VGG e ResNet) na detecção de melanoma. Para tal, recorre à base de dados de imagens do concurso ISIC da edição de 2017 [61].

## 1.5 Objectivos e Estrutura da Tese

A presente dissertação tem dois objectivos. O primeiro objectivo é o desenvolvimento de um sistema de detecção de melanomas baseado em redes neuronais convolucionais, que têm tido sucesso em vários problemas de análise de imagem [62]. Para tal, este sistema recorre a três modelos neuronais diferentes: AlexNet, VGG e ResNet. O segundo objectivo é a comparação do desempenho destes modelos neuronais e perceber qual deles melhor se ajusta ao problema em questão.

A tese está organizada da seguinte forma.

O Capítulo 1 explica a importância do diagnóstico precoce de melanoma e a relevância da aplicação da Aprendizagem Automática na sua detecção.

O Capítulo 2 discute as características visuais do melanoma e a importância do diagnóstico precoce. Apresenta ainda dois métodos clínicos que os dermatologistas utilizam para diagnosticar lesões da pele e uma breve descrição do estado da arte em sistemas automáticos aplicados no diagnóstico deste tipo de cancro da pele.

O Capítulo 3 apresenta uma breve introdução às redes neuronais e descreve três arquiteturas neuronais usadas na dissertação (AlexNet, VGG e ResNet).

O Capítulo 4 formula o problema tratado nesta dissertação e descreve o sistema proposto para o resolver.

O Capítulo 5 foca-se na avaliação experimental do sistema e na comparação entre as diferentes arquitecturas utilizadas, com base nos valores experimentais.

O Capítulo 6 apresenta algumas conclusões e sugestões de trabalho futuro.

# 2

## Detecção de Melanomas

Este Capítulo apresenta os conceitos-chave sobre lesões cutâneas e melanomas. Primeiro, explica a origem e o desenvolvimento das lesões cutâneas e do melanoma em particular, bem como o tratamento possível e a importância do diagnóstico precoce (Secção 2.2). Segundo, menciona as principais estruturas dermatoscópicas das lesões cutâneas (Secção 2.3). De seguida, descreve dois principais métodos clínicos para avaliar as lesões da pele (Secção 2.4). Por fim, apresenta métodos automáticos desenvolvidos com o intuito de potenciar uma detecção mais rápida do melanoma (Secção 2.5).

### 2.1 Introdução

O cancro está associado a uma proliferação anormal ou a uma malformação das células humanas (devido a mutações genéticas) [63]. No caso do cancro da pele, a sua origem deve-se ao desenvolvimento anómalo das células da epiderme ou da derme, as duas principais camadas da pele. Assim, a designação do tumor advém do tipo de células a partir do qual surgiu. As três formas mais comuns são [64]: cancro basocelular (*basal cell carcinoma*, BCC) proveniente das células basais da epiderme, cancro das células escamosas da epiderme (*squamous cell carcinoma*, SCC) e melanoma, que provém do crescimento maligno de células melanocíticas (ou melanócitos) da epiderme [65].

Os melanócitos são as células responsáveis pela produção de melanina, o pigmento que dá cor à pele e ao cabelo [18]. *A priori*, todas as lesões melanocíticas são sinais benignos, a menos que se desenvolvam de uma forma descontrolada, tornando-se, pois, numa lesão maligna, que é conhecida por melanoma. Uma das técnicas mais utilizadas para analisar lesões melanocíticas é a dermoscopia [66]. Esta técnica não-invasiva utiliza um aparelho óptico (chamado dermoscópico ou dermatoscópico) com lentes de ampliação, no qual é utilizada luz polarizada ou introduzido um líquido entre as lentes e a pele do paciente para reduzir o efeito de reflexão da luz [5, 53, 27]. A grande vantagem deste método clínico é a visualização mais detalhada das estruturas dermatoscópicas das lesões observadas, que a olho nu não são detectáveis (*e.g.*: pontos, glóbulos e rede pigmentada, estruturas explicadas na Secção 2.3) [67].

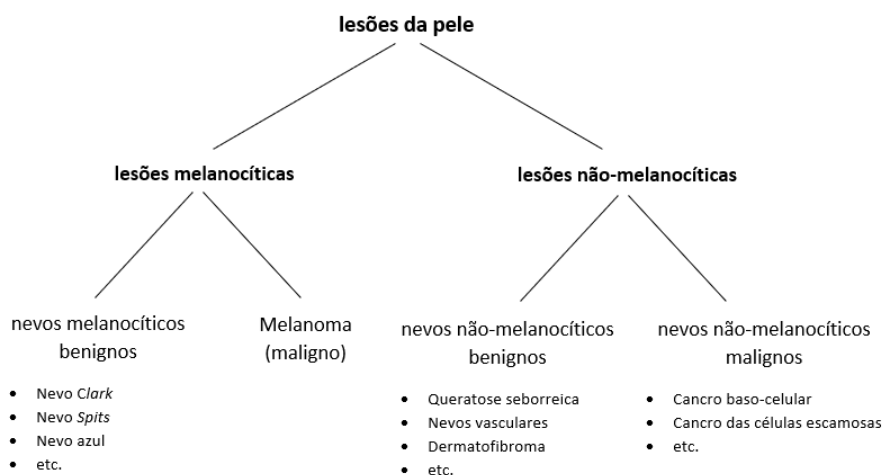


Figura 2.1: Diagrama que ilustra as duas categorias de lesões do cancro da pele [8].

Os dermatologistas avaliam as lesões da pele apoiando-se em métodos de diagnóstico clínico, que incluem um conjunto de critérios de avaliação. Dois dos mais conhecidos são a regra ABCD [55] e a lista dos sete-pontos (*Seven-point checklist*) [68], explicados em detalhe na Secção 2.4.

O diagnóstico de melanoma é subjectivo pois depende da experiência do médico e da sua acuidade visual [8, 52, 69]. Por isso, têm sido desenvolvidos sistemas automáticos de detecção de sinais da pele (*CAD systems*) para potenciar o diagnóstico precoce de melanoma. Estes sistemas tipicamente recebem uma imagem dermatoscópica e produzem uma tentativa de diagnóstico. Actualmente, não são sistemas infalíveis uma vez que o seu desempenho depende muito da qualidade e quantidade de imagens dermatoscópicas usadas no treino. Mas pretendem auxiliar os dermatologistas a fazer o diagnóstico. Na Secção 2.5 deste Capítulo encontram-se descritos alguns destes sistemas automáticos descritos na literatura.

## 2.2 Lesões Cutâneas

As lesões da pele dividem-se em duas categorias: melanocíticas e não-melanocíticas. Cada um destes grupos engloba diversas lesões cutâneas, que podem estar na origem das várias formas do cancro da pele (e.g., cancro basocelular que provém dos não-melanócitos basais; cancro das células escamosas, que advém dos não-melanócitos escamosos; melanoma, originado pelos melanócitos). A Figura 2.1 mostra a distinção entre as duas categorias dos sinais da pele.

Um sinal de pele, ou nevo (*nevus*, no singular, e nevus ou *nevi*, no plural), aparentemente inocente pode, rápida ou gradualmente, crescer de forma anormal, mudar de cor ou textura, formando assim um tumor [24, 18]. Por vezes, os tumores de pele também podem surgir *de novo*, i.e., através de novas lesões cutâneas que não existiam. Estas novas lesões são o resultado de mutações genéticas que levam a uma proliferação acelerada das células da pele [20]. Estas alterações genéticas podem ser potencialmente despoletadas pela exposição solar excessiva e acumulada [70], que é a maior causa dos cancros de pele [71, 25].

O melanoma é o tumor de pele que merece destaque por ser o que causa mais mortes [19, 1]. A sua origem deve-se à proliferação descontrolada e excessiva de melanócitos. Alguns dos sintomas são manchas claras na pele, que provocam comichão ou sangramento. Inicialmente, consiste numa pequena lesão castanha,

avermelhada ou preta e, à medida que o tempo passa, sofre modificações de forma, cor ou textura [24]. Os principais factores de risco são: inúmeros sinais de pele; uma cor clara de tez, cabelo e olhos; caso clínico individual ou de família de melanoma; várias ocorrências passadas de escaldões; exposição solar contínua; e alterações morfológicas de *nevi* existentes [71, 72, 73, 3].

Este neoplasma pode ser descrito por quatro fases distintas, desde o melanoma *in situ* até a um estágio em que se metastizou e espalhou-se pelo corpo humano. O melanoma *in situ* consiste na fase mais precoce deste tumor, que se localiza na zona mais superficial da epiderme e está apenas confinado a essa área, sem sinais de metastização nem de invasão dos gânglios linfáticos. Cada uma destas fases encontra-se descrita na Tabela 2.2 e é, então, caracterizada por um conjunto de três indicadores, cujo nome em inglês é representado pela sigla TNM. Desta sigla, T representa a espessura do tumor (em inglês, *Tumour*), N os gânglios linfáticos (*Nodes*, que avalia se o tumor atingiu estes gânglios ou não) e M as metástases (*Metastases*, que avalia se o melanoma já se alastrou para outras zonas do corpo, *i.e.*, se já se metastizou). A definição da sigla destes indicadores deve-se ao comité americano *American Joint Committee on Cancer* (AJCC) e está sumariamente descrita em [2].

Cada um destes indicadores divide-se em quatro patamares. O primeiro nível corresponde ao melhor cenário possível: menor espessura do tumor (avaliado pelo indicador T, que se baseia no critério de profundidade de Breslow [74] descrito na Tabela 2.1), sem sinal de invasão nos gânglios linfáticos (avaliado pelo indicador N) e sem metástases (avaliado pelo indicador M). Já no último patamar o tumor pode ser bastante profundo, ter atingido quatro ou mais gânglios linfáticos e ter-se alastrado para qualquer órgão do corpo.

Tabela 2.1: Conversão entre as diferentes categorias de profundidade do critério de Breslow e os níveis do indicador T. Adaptado de [1].

Espessura de Breslow (mm)	Níveis do (T)umor
≤ 1	T1
1,01 - 2	T2
2,01 - 4	T3
< 4,01	T4

Pela Tabela 2.2 e pelo significado da sigla TNM, é possível compreender que quanto menor for o estágio de desenvolvimento do melanoma, mais hipóteses de sobrevivência o paciente tem [18]. Se este tumor for detectado num estágio inicial (se estiver *in situ* ou no estágio I) e se for removido cirurgicamente, o paciente tem uma elevada probabilidade de cura [72, 3, 26]. Mais concretamente, 90% a 97% dos pacientes sobrevivem aos primeiros dez anos após o diagnóstico [30]. O tipo de tratamento cirúrgico disponível consiste numa simples excisão, *i.e.*, na remoção da zona cancerígena e de uma parte da pele saudável em redor. Por vezes, o paciente é submetido a uma segunda intervenção cirúrgica para remover mais tecido saudável, que estava à volta do melanoma previamente removido. No caso de o paciente apresentar sinais de metastização, o melanoma pode ser incurável e o tratamento apenas paliativo [18].

Clinicamente, os dermatologistas analisam as lesões da pele visualmente (a olho nu) ou através de um dermatoscópio. Este equipamento permite a visualização mais pormenorizada e aumentada das estruturas dermatoscópicas contidas em cada lesão observada. Por outras palavras, analisam as imagens clínicas ou dermatoscópicas de cada nevo. De forma a produzir um prognóstico, os médicos podem recorrer a um dos

Tabela 2.2: Descrição de cada estágio do melanoma. Adaptado de [2], [3], [1].

Estádio de melanoma	Descrição
I	Tem no máximo 1mm de profundidade e não apresenta sinais de metastização nem de invasão dos gânglios linfáticos. Por vezes, nesta fase o tumor pode estar ulcerado <sup>1</sup> e sem mitoses <sup>2</sup> ou não estar ulcerado e ter mitoses.
II	A espessura do melanoma está compreendida entre os níveis T2 e T4 e continua sem apresentar níveis de metastização nem de invasão dos gânglios linfáticos. Pode estar ou não ulcerado.
III	O melanoma invade os gânglios linfáticos e a sua espessura está compreendida entre os níveis T1 e T4. Não há sinais de metástases no corpo e o tumor pode estar ou não ulcerado.
IV	Estádio muito grave do melanoma. Apresenta invasão dos gânglios linfáticos, metástases numa ou em mais zonas do corpo distantes da origem do tumor e com espessura compreendida entre os níveis T1 e T4.

muitos métodos clínicos de diagnóstico existentes. A regra ABCD [55] e a lista dos sete-pontos [68] são dos métodos mais utilizados para este fim e encontram-se explicados na Secção 2.4.

No entanto, este procedimento de diagnóstico clínico é subjectivo, visto que se baseia numa observação humana visual (mesmo quando se recorre à dermoscopia pela utilização do dermatoscópio), na acuidade visual do dermatologista e também na sua experiência profissional [69]. De referir ainda que há lesões benignas melanocíticas e não-melanocíticas que apresentam propriedades dermatoscópicas e morfológicas muito idênticas às do melanoma [19], como o caso do nevo Clark [28] e das queratoses seborreicas [75], o que dificulta muito a tarefa do dermatologista. Deste modo, com o intuito de colmatar estes erros de diagnóstico e potenciar o diagnóstico precoce do melanoma, modelos de detecção automática têm sido desenvolvidos [36]. Estes sistemas são denominados por *Computer-Aided Diagnosis (CAD) systems* e encontram-se descritos na Secção 2.5.

## 2.3 Estruturas Dermatoscópicas

Cada lesão cutânea é constituída por uma ou várias estruturas dermatoscópicas. Na Tabela 2.3 estão listadas as principais estruturas dermatoscópicas para lesões melanocíticas benignas e malignas, baseadas em [4, 5, 65, 76]:

- lesões melanocíticas benignas: rede pigmentada típica, pontos/glóbulos regulares, cicatrizes regulares, manchas regulares;
- lesões melanocíticas malignas: rede pigmentada atípica, pontos/glóbulos irregulares, cicatrizes irregulares, manchas irregulares, estruturas de regressão, véu azul-esbranquiçado.

De seguida, as Figuras 2.2 a 2.6 mostram exemplos de estruturas dermatoscópicas.

Tabela 2.3: Principais estruturas dermatoscópicas para lesões melanocíticas. Adaptado com base em [4, 5].

<b>Lesões melanocíticas benignas</b>	<b>Descrição</b>
Rede pigmentada típica	Rede de linhas castanha (clara ou escura) e uniformemente distribuída por toda a lesão, embora por vezes mais fina nas extremidades.
Pontos/glóbulos regulares	Estruturas redondas/ovais acastanhadas, uniformes em tamanho e forma e com distribuição homogénea, concentrando-se, muitas das vezes, no centro da lesão.
Cicatrizes regulares	Estruturas lineares que apresentam um padrão <i>starburst</i> com simetria radial e cor castanha.
Manchas regulares	Manchas cinzento-azuladas ou branco-acastanhadas.
<b>Lesões melanocíticas malignas</b>	<b>Descrição</b>
Rede pigmentada atípica	Rede irregular de linhas, distribuídas de forma não uniforme por toda a lesão. Normalmente as linhas são mais grossas do que as da rede pigmentada típica. Apresentam cor preta, castanha ou cinzenta.
Pontos/glóbulos irregulares	Estruturas redondas/ovais irregulares, variam de tamanho e forma e não estão uniformemente distribuídos pela lesão. Apresentam cor preta ou castanha.
Cicatrizes irregulares	Estruturas lineares distribuídas irregularmente nas extremidades da lesão.
Manchas irregulares	Manchas pretas, castanhas ou cinzentas, assimétricas e sem forma regular.
Estruturas de regressão	Estruturas planas que ilustram a regressão do melanoma. Costumam apresentar duas cores: branco e azul. A parte branca da lesão é bem definida e assemelha-se a uma cicatriz superficial. Por vezes, as zonas azuladas são esfumadas e pontilhadas.
Véu azul-esbranquiçado ( <i>Blue-whitish veil</i> )	Estrutura com pigmentação azul localizada apenas numa parte da lesão e sobre a qual existe uma mancha branca ténue (tornando-se difícil de distinguir, muitas vezes, com as estruturas de regressão).

## 2.4 Métodos Clínicos

Existem vários métodos clínicos de diagnóstico cujo o intuito de guiar os dermatologistas quando estão a realizar o prognóstico dos sinais da pele do paciente. Na presente dissertação são apenas enunciados dois desses métodos: a regra ABCD e a lista dos sete-pontos.

### 2.4.1 Regra ABCD

A regra ABCD foi criada por Stolz *et al.* em 1991 e modificada em 1994 [55], sendo até aos dias de hoje fortemente utilizada pelos médicos para o diagnóstico de lesões cutâneas. Este procedimento simples, apenas

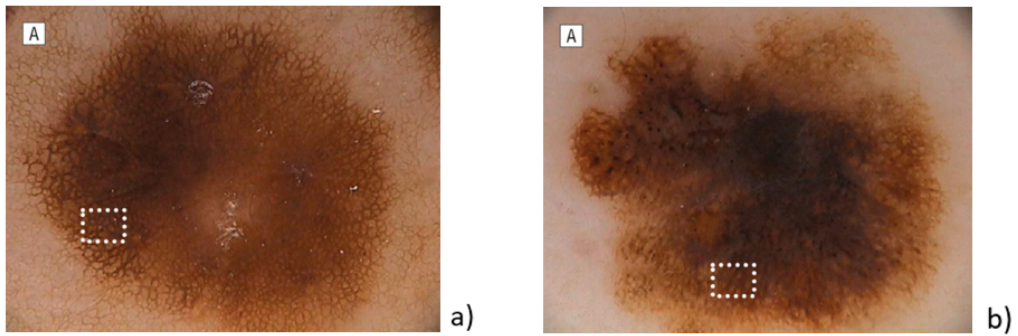


Figura 2.2: Rede pigmentada: típica à esquerda em a) e atípica à direita em b). Extraído de [9].

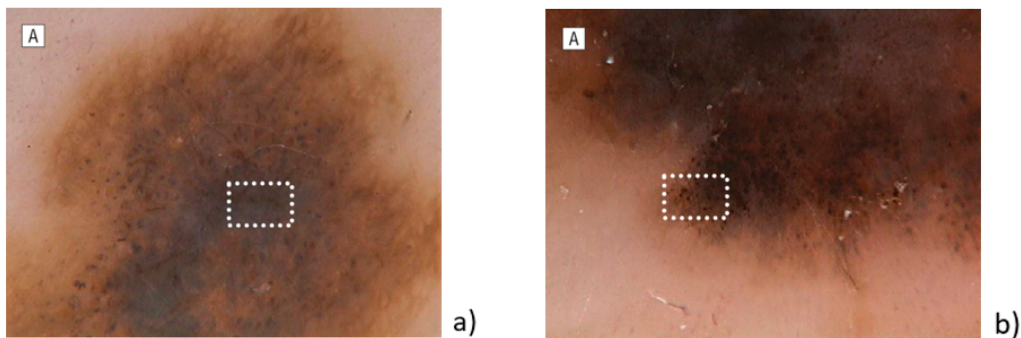


Figura 2.3: Pontos: regulares à esquerda em a) e irregulares à direita em b). Extraído de [9].

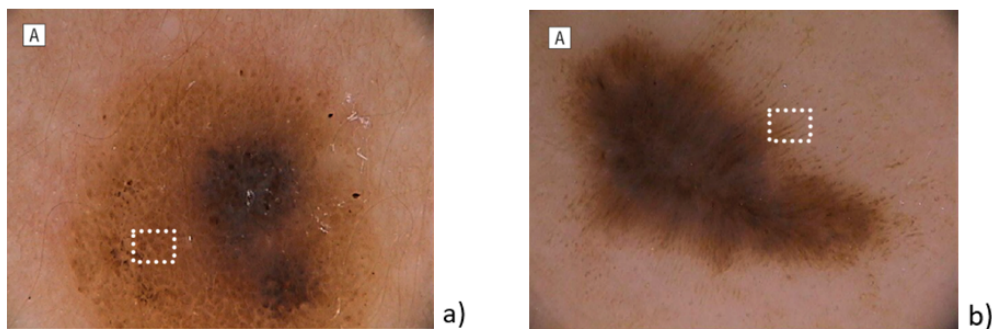


Figura 2.4: Glóbulos regulares à esquerda em a) e cicatrizes regulares à direita em b). Extraído de [9].

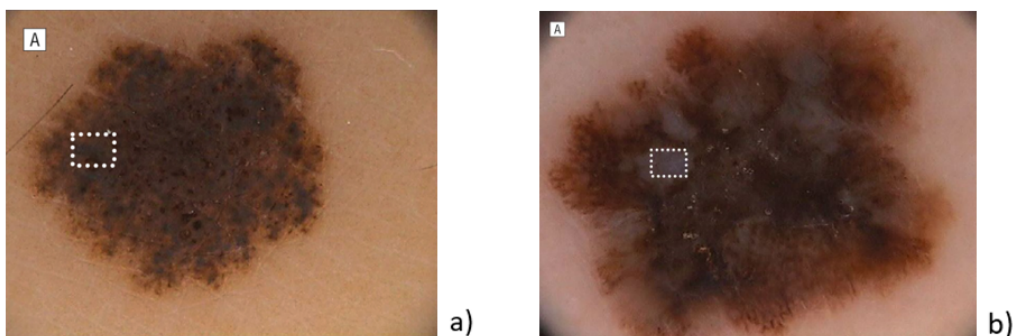


Figura 2.5: Mancha azulada (benigna) à esquerda em a) e véu azul-esbranquiçado à direita em b). Extraído de [9].

aplicável a lesões melanocíticas, permite identificar se uma lesão é benigna, maligna ou duvidosa (*i.e.*, não é possível determinar com exactidão a sua natureza) através do cálculo de uma fórmula, denominada por *Total*





Figura 2.6: Estrutura de regressão. Extraído de [9].

*Dermatoscopy Score (TDS) 2.1.*

$$TDS = A \times 1.3 + B \times 0.1 + C \times 0.5 + D \times 0.5 \quad (2.1)$$

Os factores A, B, C, D representam os quatro critérios pelos quais a regra ABCD se rege e são eles, respectivamente: assimetria (A), contorno (B), cor (C) e estruturas dermatoscópicas presentes (D). A Tabela 2.4 descreve sumariamente como são calculados cada um destes critérios e indica o intervalo de valores possível para cada um. Cada critério é multiplicado por um factor diferente, indicado numericamente em 2.1. Na Tabela 2.5 encontra-se a tradução entre os valores possíveis de classificação final da lesão e o consequente diagnóstico. Se uma lesão tiver uma classificação inferior a 4,75, trata-se de uma lesão benigna; entre 4,8 e 5,45 é uma lesão duvidosa; acima de 5,45 são potencialmente lesões malignas. De notar que os valores TDS não são totalmente fiáveis, na medida em que há falsos positivos ou falsos negativos. Por outras palavras, o diagnóstico baseado na regra ABCD pode classificar uma lesão como benigna e, afinal, ela ser maligna (falso positivo) e vice-versa (falso negativo). Quando o diagnóstico indica que uma lesão é duvidosa, normalmente os médicos aconselham o paciente a vigiar a evolução do sinal ou a fazer uma excisão. A Figura 2.7 mostra um exemplo de diagnóstico através da regra ABCD.

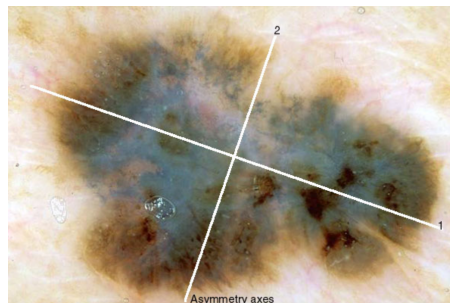


Figura 2.7: Diagnóstico da lesão: melanoma, baseado na regra ABCD. Classificação final, TDS (7,6 pontos): A - 2 eixos de simetria (2 pontos×1,3); B - 5 segmentos (5 pontos×0,1); C - 5 cores [castanho claro, castanho escuro, cinzento azulado, preto e branco] (5 pontos×0,5); D - 4 estruturas [áreas homogêneas, cicatrizes, pontos, glóbulos] (4 pontos×0,5). Extraído de [6].

Tabela 2.4: Descrição de cada um dos quatro critérios usados pela regra ABCD e o intervalo de valores que cada um destes critérios pode assumir na fórmula TDS. Adaptado de [6].

<b>Critério</b>	<b>Descrição</b>	<b>Intervalo de Classificação Individual</b>
Assimetria, A	Na lesão são desenhados e posicionados dois eixos de 90° de forma a produzir a menor pontuação possível de assimetria. Este critério vale dois pontos se se verificar uma assimetria dermoscópica nos dois eixos a nível de cor, forma e estruturas dermoscópicas. Se essa assimetria só se verificar num dos eixos, vale a um ponto. Caso contrário, vale zero pontos.	0-2
Contorno, B	A lesão é segmentada em oito partes. Se em cada uma delas for encontrada um corte abrupto do padrão de pigmentação na periferia, soma-se um ponto por segmento a este critério. Caso contrário, se o corte for gradual somam-se zero pontos por cada segmento.	0-8
Cor, C	Cada lesão pode conter uma a seis cores possíveis: branco, vermelho, castanho claro, castanho escuro, cinzento azulado e preto. Cada cor encontrada soma um ponto à classificação individual deste critério.	1-6
Estruturas dermoscópicas, D	Este critério avalia a presença de cinco estruturas dermoscópicas: rede pigmentada, áreas homogêneas, cicatrizes, pontos e glóbulos. Se 10% ou mais da lesão apresentar uma área homogênea, esta estrutura é contabilizada. As cicatrizes e os pontos são contabilizados apenas se a lesão tiver pelo menos duas amostras de cada uma destas estruturas. Os glóbulos são contabilizados se existir, pelo menos, um. Cada estrutura contabilizada soma um ponto à classificação individual deste critério, independentemente de aparecer uma ou mais vezes na lesão.	1-5

Tabela 2.5: Conversão entre os possíveis valores de TDS de uma lesão e o diagnóstico provável. Adaptado de [6].

<b>TDS</b>	<b>Diagnóstico</b>
< 4,75	Lesão benigna
4,8-5,45	Lesão duvidosa
> 5,45	Lesão maligna

## 2.4.2 Lista dos Sete-Pontos

Este método foi proposto por Argenziano em 1998 *et al.* [68] com o intuito de ajudar dermatologistas menos experientes.

Como o próprio nome indica, este procedimento de diagnóstico avalia a presença de sete estruturas dermoscópicas, designadas por critérios. Porém, três das sete estruturas são mais relevantes que as restantes, pertencendo, assim, ao conjunto denominado por principal. As restantes fazem parte do conjunto secundário. O conjunto principal é formado por: rede pigmentada atípica, véu azul-esbranquiçado e padrão vascular atípico. Cada um destes critérios contribui com dois pontos para a classificação final da lesão a ser diagnosticada. O conjunto secundário é representado pelas seguintes estruturas: cicatrizes irregulares, pigmentação irregular, pontos/glóbulos irregulares e estruturas de regressão. Neste caso, cada critério soma à classificação final um

ponto. A Tabela 2.6 pode ser consultada para uma compreensão mais clara dos dois conjuntos de avaliação da lista dos sete-pontos e respectivas pontuações. A classificação final deste método é obtida pela soma individual de cada uma das estruturas encontradas na lesão. Se o valor for igual ou superior a 3 pontos, a lesão é diagnosticada como sendo melanoma. Neste caso, a lesão tem, pelo menos, um critério principal e dois secundários ou três critérios secundários. Apenas com o objectivo de ilustrar este método de diagnóstico, a Figura 2.8 apresenta uma lesão e as estruturas nela presentes.

Tabela 2.6: Identificação e classificação individual de cada critério da lista dos sete-pontos. Adaptado de [7].

Conjunto "major"	Classificação individual
Rede pigmentada atípica	2
Véu azul-esbranquiçado	2
Padrão vascular atípico	2
Conjunto "minor"	Classificação individual
Cicatrizes irregulares,	1
Pigmentação irregular	1
Pontos/glóbulos irregulares	1
Estruturas de regressão	1

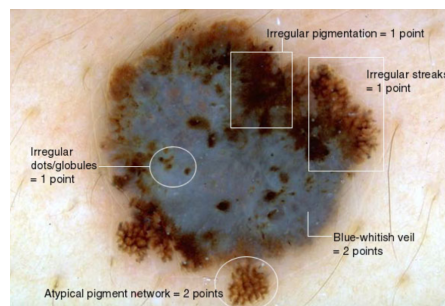


Figura 2.8: Diagnóstico da lesão: melanoma, baseado na lista dos sete-pontos. Classificação final (7 pontos): rede pigmentada atípica (2 pontos), véu azul-esbranquiçado (2 pontos), pontos/glóbulos irregulares (1 ponto), pigmentação irregular (1 ponto) e cicatrizes irregulares (1 ponto). Adaptado de [7].

## 2.5 Métodos Automáticos - CAD systems

Os sistemas CAD são sistemas de diagnóstico automático e têm como objectivo colmatar a subjectividade inerente aos métodos de diagnóstico clínico, através da detecção automática de melanomas. Estes sistemas ajudam os dermatologistas na avaliação clínica das lesões dermatoscópicas [58], recorrendo a técnicas de aprendizagem automática e processamento de imagem. O procedimento dos CAD (ver Figura 1.2) consiste maioritariamente em quatro fases (cada uma delas descrita nas próximas Secções): pré-processamento, segmentação, extracção e selecção de características e, por fim, a classificação das lesões [28, 51, 8, 31].

De realçar que é imprescindível e de extrema importância a utilização de uma base de dados de imagens dermatoscópicas de casos clínicos reais, para que os sistemas CAD sejam capazes de produzir uma detecção automática fiável de lesões dermatoscópicas. Para tal, é necessário que se disponibilizem (grandes) bases de dados de imagens dermatoscópicas e que cada uma dessas imagens esteja previamente segmentada e anotada

por dermatologistas experientes [58]. A anotação de uma imagem corresponde ao diagnóstico clínico da lesão e, nalguns casos mais raros, à identificação das estruturas dermatoscópicas mais relevantes.

Quando estas bases de dados contêm um número insuficiente de imagens dermatoscópicas, recorre-se, muitas vezes, à técnica de aumento de dados, *Data Augmentation*, técnica de aumento de dados (DA), para aumentar artificialmente o conjunto total de imagens (e.g. através de rotações, inversões e distorções não-lineares do conjunto original) e melhorar, assim, o desempenho dos sistemas CAD. Um outro problema das bases de dados públicas existentes é a falta de normalização das imagens dermatoscópicas. Por outras palavras, as imagens da base de dados são adquiridas por equipamentos diferentes e diferentes condições de iluminação, o que prejudica o desempenho dos sistemas automáticos de diagnóstico.

Nas Secções 2.5.1, 2.5.2, 2.5.3 e 2.5.4 são apresentadas, respectivamente, cada fase dos sistemas CAD. De seguida, são enunciados seis estudos de sistemas CAD que se destacam na literatura da detecção de melanoma. Estes estudos são divididos em dois grupos de sistemas CAD. No primeiro, as características da imagem são previamente escolhidas por um especialista e os CAD extraem e classificam automaticamente essas características recorrendo a técnicas específicas de extracção e de classificação. Portanto, neste caso, as características a extrair são previamente seleccionadas manualmente. No segundo grupo, os CAD extraem e classificam automaticamente as características, recorrendo a redes neuronais profundas.

### 2.5.1 Pré-processamento

O pré-processamento tenta corrigir imperfeições da imagem e melhorar a qualidade da mesma. As imperfeições correspondem a objectos que dificultam a visualização da lesão, como, por exemplo, pêlos, bolhas de ar, gel aplicado na pele aquando da aquisição da imagem dermoscópica, marcas de régua que indicam a dimensão em milímetros da lesão, entre outros [36]. O melhoramento da qualidade inclui a calibração dos níveis de cor e iluminação, bem como realça o contraste entre a lesão e o respectivo contorno [30]. Estes melhoramentos de cor e contraste são necessários, pois a dermoscopia torna difícil a distinção cromática entre a lesão e a pele saudável, em volta da lesão, e a distinção do contraste entre a lesão e o contorno desta. De notar ainda que as imagens dermatoscópicas podem ter grandes dimensões, o que não implica que a lesão seja grande. Na maior parte das vezes a lesão ocupa uma pequena área da imagem total, pelo que se recorre ao recorte da imagem original na tentativa de se seleccionar apenas a lesão e parte da pele saudável em seu redor. A este recorte à volta da lesão dá-se o nome de *bounding box*.

De maneira a corrigir as imperfeições e a melhorar a imagem são utilizadas diversas técnicas de filtragem (e.g. filtros de mediana [53] e filtros gaussianos [77]), compensação da cor [78], transformação do espaço de cores e *thresholding* ([79, 80]), entre outras [30, 36].

### 2.5.2 Segmentação

Esta fase divide a imagem em duas regiões: a região associada à lesão e a região que corresponde à pele saudável. A fronteira entre as duas regiões forma o contorno da lesão, uma das características a ser extraída e analisada posteriormente. As regiões obtidas por essa divisão, embora distintas, são homogéneas entre si explorando vários tipos de critérios: cor, textura ou contornos [36, 30].

Foram propostas várias técnicas de segmentação: segmentação baseada na região (e.g., *region-growing* [28]); segmentação baseada na cor ([81]); limiarização adaptativa, modelos deformáveis, classificação de pixéis, sistemas difusos ([52, 53]), *clustering* ([82]), sistemas difusos ([52]), entre outras [30, 36].

### 2.5.3 Extracção e selecção de características

O termo característica no contexto dos sistemas CAD pode ter dois significados: característica/estrutura clínica/dermatoscópica (e.g., assimetria, glóbulos regulares e rede pigmentada) ou característica de imagem (cor, textura e forma). Portanto, para além de poderem representar propriedades ao nível dos pixéis, as características da imagem podem igualmente ser interpretadas como estruturas dermatoscópicas, se assim forem construídas [30]. As textura, cor e geométricas são das características extraídas mais comuns [36, 83]. As características geométricas dizem respeito a propriedades do contorno da lesão: forma, tamanho, área e perímetro, rácios de espessura/circularidade, simetria, entre outros.

Depois da extracção de características poderá ocorrer a redução de dimensionalidade. Este passo pode ser feito de duas formas: redução de características e selecção de características. Na primeira transformam-se as características num vector de características modificadas, com menor dimensão. Na segunda, não se modificam as características mas escolhem-se as melhores por eliminação de características pouco informativas ou redundantes. Existem vários algoritmos para a redução de dimensionalidade, tais como *principal component analysis* (PCA), MIFS ([84]), ReliefF ([85]), CFS ([86]). Este passo de redução ou selecção de características contribui bastante para o desempenho da classificação, visto que não só reduz o tempo de cálculo como pode melhorar o desempenho, permitindo uma melhor generalização do diagnóstico do sistema CAD.

### 2.5.4 Classificação

A classificação tem como objectivo produzir um diagnóstico, a partir do vector de características. Dependendo do tipo do sistema CAD, esta fase pode produzir dois tipos de respostas: diagnóstico binário (a lesão é benigna ou maligna) e diagnóstico em  $n$  tipos de lesões. Para este último tipo de resposta, o número  $n$  de nevos corresponde ao número total de classes para o qual o CAD foi treinado para reconhecer. A classificação é tanto mais correcta quanto mais informativo for o conjunto de características a diferenciar as lesões da pele [30]. Este conjunto é aprendido pelo CAD durante uma fase de treino que utiliza todas as imagens segmentadas da base de dados inicial. Por isso, esse conjunto de características é tanto mais preciso quanto mais precisa for a segmentação, cujo sucesso depende da base de dados de imagens dermatoscópicas inicial.

Existem diversos algoritmos de classificação, sendo os seguintes os mais utilizados: máquinas de vectores de suporte (SVM), regressão logística, redes neuronais, k-vizinhos mais próximos (k-nn), árvores de decisão, floresta aleatória.

### 2.5.5 Revisão da Literatura

As redes neuronais convolucionais (CNN) são dos métodos de aprendizagem profunda que mais têm sido estudados [62], devido às suas capacidades para extrair e classificar automaticamente características relevantes a partir de imagens digitais. No caso da detecção de melanoma, as CNN conseguem extrair automaticamente

características que permitem diferenciar uma lesão benigna de uma maligna. Embora o diagnóstico automático produzido não seja ainda totalmente fiável, pretende-se que os sistemas CAD que utilizem CNN se tornem num método de diagnóstico de melanoma capaz de ajudar os médicos em ambiente clínico [32], visto que aceleram o processo da detecção de melanoma.

Nesta Secção são apresentados dois grupos de sistemas CAD. O primeiro grupo baseia-se na arquitectura clássica de reconhecimento de padrões (ver Figura 1.2). O segundo grupo baseia-se numa arquitectura *end-to-end* de extracção e classificação automáticas de características, recorrendo a redes neuronais convolucionais (CNN) (ver Figura 1.3).

Selecionaram-se três sistemas de cada um destes grupos que são resumidos nas Tabelas 2.7 e 2.8.

Grupo I de sistemas CAD					
Autores, [referência]	Pré-processamento + Segmentação	Características	Classificador	Base de Dados	Desempenho
Celebi <i>et al.</i> , [29]	Filtro mediano, redução de cor, <i>Otsu thresholding</i>	Forma, cor e textura. Algoritmos: CFS, MIFS e modificação do ReliefF	SVM	564 imagens dermatoscópicas das bases de dados dos Hospitais das Universidades de Graz (Áustria), de Nápoles e de Florença (Itália)	SP = 92,34% SE = 93,33%
Barata <i>et al.</i> , [69]	<i>Adaptive thresholding</i>	Cor e textura. Algoritmos: regra ABCD (sistema I) e BoF (sistema II)	k-nn, SVM e Adaboost (sistema I); BoF, histograma + k-nn (sistema II)	176 imagens dermatoscópicas da base de dados do Hospital Pedro Hispano, Matosinhos	Sistema I: SP = 80% SE = 96% Sistema II: SP = 75% SE = 100%
Jaworek-Korjakowska e Kleczek, [32]	<i>Black frame removal</i> , detecção de pêlos, <i>top-hat transform</i> , filtro mediano	Geométricas, cor, textura e assimetria. Algoritmos: <i>region-growing and merging</i> , CFS	k-nn, <i>logistic regression</i> , árvores de decisão, SVM	300 imagens dermatoscópicas das bases de dados dos Hospitais das Universidades de Nápoles (Itália) e de Graz (Áustria)	Exactidão = 92% (SVM)

*Legenda:* CFS - *Correlation based feature selection*; MIFS - *Mutual information based feature selection*; SVM - Máquinas de vectores de suporte; k-nn - k-vizinhos mais próximos; BoF - *Bag of Features*

Tabela 2.7: Apresentação de três trabalhos de investigação baseados na arquitectura clássica de reconhecimento de padrões.

A avaliação destes sistemas é feita em bases de dados diferentes, adquiridas em condições distintas e nível de dificuldade distinto. Por isso, não é possível uma comparação entre sistemas. A proposta da base de dados ISIC com elevado número de imagens dermoscópicas abre uma janela de esperança em relação ao futuro.

Como se pode constatar através dos estudos mencionados no Grupo 2, as redes neuronais convolucionais começam a apresentar um bom desempenho no diagnóstico de melanoma. O objectivo principal da aplicação deste método de aprendizagem profunda não é substituir o papel dos dermatologistas. A intenção é de ajudar, em ambiente clínico, os médicos a diagnosticar o melanoma mais cedo e evitar a excisão desnecessária de

Grupo II de sistemas CAD					
Autores, [referência]	Pré-processamento + Segmentação	Características	Classificador	Base de Dados	Desempenho
Codella <i>et al.</i> , [16]	CNN	Histograma de cor, FC6, DRN, U-net, entre outros	SVM	1279 imagens dermatoscópicas: 900 de treino e 379 de teste	SP = 92,34% SE = 93,33%
Esteva <i>et al.</i> , [33]	Arquitetura <i>end-to-end</i> : uso de uma CNN pré-treinada. Realização de três tarefas de classificação, cujo desempenho foi comparado ao diagnóstico clínico de 21 dermatologistas.			129.450 imagens dermatoscópicas e captadas por fotografia. Conjuntos de teste: I - 135 imagens; II - 130 imagens melanocíticas clínicas; III - 111 imagens dermatoscópicas de lesões melanocíticas.	Tarefa I: ROC-AUC = 96%  Tarefa II: ROC-AUC = 94%  Tarefa III: ROC-AUC = 91%  Em cada tarefa, a CNN teve melhor desempenho do que a média de diagnóstico feito pelos médicos.
Haenssle <i>et al.</i> , [34]	Arquitetura <i>end-to-end</i> : uso de uma CNN pré-treinada, cujo desempenho num dos conjuntos de teste foi comparado ao diagnóstico clínico de 58 dermatologistas.			+100.000 imagens de treino e validação. Conjuntos de teste: I - 300 imagens dermatoscópicas; II - 100 imagens dermatoscópicas; III - 100 imagens do concurso ISIC.	Tarefa II: CNN: SE = 95%  58 dermatologistas (média): SE = 88,90% SP = 75,70%

Legenda: FC6 – camada n.º 6 totalmente interligada de uma CNN; DRN – redes profundas residuais; U-net – um tipo de CNN; SVM - Máquinas de vetores de suporte

Tabela 2.8: Apresentação de três trabalhos de investigação baseados na arquitetura *end-to-end*.

não-melanomas, que clinicamente é uma tarefa, por vezes, extremamente difícil [32, 33, 87].





# 3

## Redes Neurais Convolucionais

Este Capítulo introduz vários conceitos relacionados com as redes neuronais convolucionais (CNN). Primeiro, explica a inspiração biológica e depois os conceitos básicos inerentes a todos os métodos neuronais (Secção 3.1). De seguida, descreve a arquitectura das CNN, composta principalmente por três tipos de camadas: convolução, *pooling* e totalmente interligadas (Secção 3.2). Por fim, apresenta o concurso ILSVRC [44], que despoletou a revolução das CNN na década actual, e descreve sumariamente a arquitectura de três modelos de sucesso desse concurso: AlexNet, VGG e ResNet [13, 14, 15].

### 3.1 Introdução às redes neuronais artificiais

#### 3.1.1 Inspiração Biológica

As células nervosas, neurónios, são a unidade de operação básica do sistema nervoso humano, que transmitem sinais entre si através de sinapses (zona de contacto entre dois neurónios, na qual se dá a transmissão do sinal através de neurotransmissores). Cada neurónio recebe um sinal de entrada através das suas dendrites (ramificações dos neurónios que recebem os sinais, *i.e.*, os estímulos nervosos) e, através do seu axónio (estrutura que conduz os impulsos nervosos desde as dendrites até à zona de transmissão com o neurónio vizinho), produz um sinal de saída, que é depois transmitido a outros neurónios através de sinapses [10]. Porém, o sinal de saída só é efectivamente transmitido tendo em conta a força da sinapse e se ocorreu a activação do neurónio, *i.e.*, se o valor do sinal processado no corpo do neurónio é superior a um determinado valor pré-definido (considerado limiar). As ligações entre neurónios baseiam-se na força da sinapse, *i.e.*, pela importância que um neurónio atribui a um sinal proveniente de um neurónio anterior.

As redes neuronais artificiais são inspiradas no processo biológico acima descrito. O conceito do neurónio artificial surgiu em 1943 através do modelo proposto por McCulloch e Pitts [41]. Este primeiro modelo matemático de um neurónio artificial somava vários sinais de entrada (que representavam as dendrites), previamente multiplicados por pesos, e produzia um único valor de saída. O sinal de saída era comparado com

um limiar. Se o valor do sinal estivesse acima do valor limiar, então o sinal de saída era activado e tomava o valor de um, caso contrário era zero. O neurónio artificial era, assim, representado pela soma ponderada dos vários sinais que recebia e pelo processamento dessa soma, à semelhança do neurónio humano.

Em 1957, Rosenblatt desenvolveu o modelo Perceptrão [88], em que a estrutura do neurónio artificial é idêntica à do modelo de McCulloch-Pits. O Perceptrão é um modelo de classificação binária para problemas linearmente separáveis, que consiste na combinação linear de sinais de entrada, com a adição de um termo limiar (*bias*). De seguida, à semelhança dos neurónios humanos, este modelo aplica uma função não-linear (chamada de função de activação) ao resultado anterior, obtendo, assim, um sinal de saída. O sinal de entrada é composto por vários neurónios, em que cada um tem uma ligação com o neurónio de saída. Cada ligação (que representa as dendrites e o axónio de um neurónio humano) tem um peso associado, que representa a importância que a ligação tem [10]. Os valores de pesos e de limiar podem ser reajustados através de um processo de aprendizagem (este conceito é explicado em 3.1.2). A Figura 3.1 ilustra a diferença entre o processo biológico e o modelo Perceptrão.

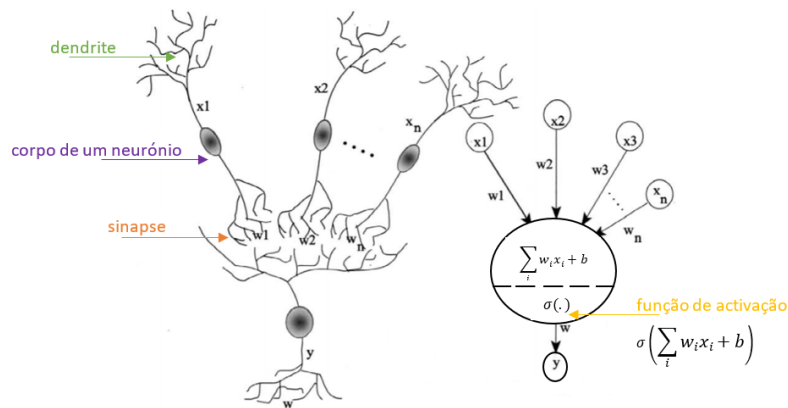


Figura 3.1: Comparação entre o processo biológico e o artificial, adaptado de [10].  $n$  neurónios ( $x_1, x_2, \dots, x_n$ ) a transmitir um sinal ao neurónio  $y$ , através das suas sinapses ( $w_1, w_2, \dots, w_n, b$ ).

Matematicamente, o Perceptrão pode ser formulado da seguinte forma

$$f(x) = \sum_i w_i x_i + b, \quad (3.1)$$

em que  $f(x)$  representa a equação do modelo Perceptrão,  $w_i, i = 1, 2, \dots, n$  os pesos,  $b$  o valor de limiar e  $\sigma$  a função de activação ( $\sigma(\sum_i w_i x_i + b)$ ).

Mais tarde, foi proposto o modelo do perceptrão multicamada (MLP) [89, 90, 91], que considera mais do que uma camada de neurónios. O MLP consiste então numa sequência de neurónios ligados entre si ao longo de camadas, em que a estrutura de cada neurónio baseia-se no modelo de McCulloch-Pits. Actualmente, muitas redes neuronais são inspiradas no modelo MLP. De uma forma geral, a arquitectura das redes é composta por uma camada de entrada, uma ou várias intermédias (chamadas camadas escondidas) e uma de saída. Quanto maior o número de camadas escondidas, mais profunda a rede se torna. Cada camada é então descrita por um ou múltiplos nós, ou neurónios. Cada neurónio está ligado a todos os neurónios da camada seguinte.

### 3.1.2 Conceitos-chave

Há duas classes de redes neuronais: redes de propagação e redes recorrentes [92]. As redes neuronais de propagação são as de interesse nesta dissertação e, por isso, doravante aprofundadas.

As redes neuronais de propagação são inspiradas no modelo MLP. A camada de entrada recebe os dados e propaga-os para as camadas seguintes [92]. As camadas escondidas, se existirem, estão contidas entre a de entrada e a de saída. A principal função das camadas escondidas é extrair implicitamente características (e respectivas relações) provenientes da camada anterior através de não-linearidades, aumentando o nível de abstracção de camada em camada [11]. A introdução de não-linearidades ao longo de toda a rede permite que esta seja capaz de representar dados que não são linearmente separáveis [93]. A camada de saída transforma o sinal recebido pela última camada escondida num vector de classes/categorias, em que cada uma contém um valor [11] (a que tiver maior valor é a categoria classificada pela rede). Se as redes neuronais tiverem mais do que duas camadas escondidas, tornam-se redes neuronais profundas e são tanto mais profundas quantas mais camadas escondidas tiverem [94].

O **treino** deste tipo de redes neuronais é feito de forma supervisionada e baseia-se no cálculo do erro entre o resultado final desejado e aquele predito pela rede, para um dado conjunto de treino. O erro pode ser avaliado por diversas formas, sendo a função de entropia cruzada (3.2) uma das mais utilizadas para tarefas de classificação [95, 96].

$$E(t, y) = - \sum_{p \in P} \sum_i t_i \log(y_i), \quad (3.2)$$

em que  $P$  representa o conjunto de treino (logo,  $p$  cada exemplo de treino),  $i$  o número de classes do classificador ( $i = 1, 2$  no caso de um classificador binário);  $t$  um indicador binário que é igual a um, se a classe  $i$  for a correcta para cada exemplo e  $y$  o vector de probabilidades obtidas para cada exemplo da classe  $i$ . O objectivo do treino da rede é então minimizar o erro de classificação, *i.e.*, encontrar um mínimo local da função de custo [97], através do ajuste iterativo dos seus parâmetros. Idealmente, o conjunto de treino deve ser grande e conter um número de exemplos representativo das diversas categorias, de forma a evitar que ocorra sobreaprendizagem [13, 98]. Caso contrário, se o conjunto de treino for pequeno, a rede consegue memorizar os exemplos e erradamente pensa-se que a rede "aprendeu" a classificar correctamente. A minimização da função de custo é conseguida através do algoritmo do gradiente, que executa a actualização dos pesos. O cálculo do gradiente é realizado em cada iteração do treino, que se divide em duas fases, **propagação directa** e **retropropagação**. Na primeira fase, o sinal passa da camada de entrada até à de saída, através das camadas escondidas, e é calculado o erro na última camada. A segunda fase dedica-se ao cálculo do gradiente. Nesta segunda fase, recorre-se muitas vezes ao algoritmo de retropropagação, dado que este algoritmo permite calcular o gradiente de forma eficiente para as várias camadas da rede [92].

Estas duas fases do treino são aprofundadas nas próximas sub-secções. Após o treino da rede, é possível aplicar o modelo a novos dados (**conjunto de teste**), e avaliar a sua capacidade de generalização [92]. Nesta fase, existe apenas uma passagem directa dos dados pela rede. Se tiver ocorrido sobreaprendizagem, a generalização da rede é mínima, constatando-se, então, que não aprendeu correctamente durante a fase de treino. Deste modo, só é possível afirmar que a rede aprendeu quando é capaz de classificar correctamente

novos exemplos de dados [99].

### 3.1.2.1 Passo propagação directa

Este passo no modelo MLP (Figura 3.2) consiste, então, na transmissão do sinal desde a camada de entrada até à de saída, passando sequencialmente pelas camadas escondidas.

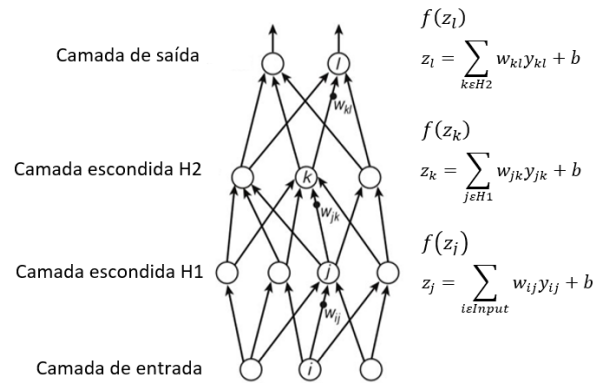


Figura 3.2: Passo de propagação directa de uma arquitectura MLP (adaptado de [11]).

Durante a fase de treino, os pesos são inicializados aleatoriamente na primeira vez em que é executado este passo. Existem várias técnicas de inicialização dos pesos [100], sendo a inicialização aleatória e a inicialização por transferência de aprendizagem duas possibilidades. A inicialização aleatória consiste em atribuir inicialmente valores aleatórios aos pesos de acordo com uma distribuição uniforme com um certo intervalo de valores, que depende do conjunto de dados inicial [100]. A inicialização por transferência de aprendizagem consiste na reutilização de pesos de uma rede já treinada para um problema diferente [101].

O sinal recebido por um neurónio de uma das camadas intermédias consiste na combinação linear de todas as suas ligações com os neurónios da camada anterior e adição de um termo contaste

$$z_j = \sum_i w_{ij} x_{ij} + b_j, \quad (3.3)$$

em que  $x_{ij}$  é o sinal recebido pela ligação do neurónio  $j$  ao neurónio  $i$  da camada anterior,  $w_{ij}$  o peso da ligação entre os neurónios  $i$  e  $j$  e  $b_j$  o termo limiar. Este último parâmetro também pode ser treinado à semelhança dos outros pesos.

A saída do neurónio  $j$  obtém-se aplicando uma função de activação ao sinal recebido pelo neurónio (que está representada na Figura 3.2 pela função  $f(\cdot)$ ). Os três tipos de activação mais conhecidos são [102]:

- Função da tangente hiperbólica:

$$f(z) = \tanh(z) = \frac{2}{1 + e^{-2z}} - 1 \quad (3.4)$$

- Função logística sigmoide:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3.5)$$

- Função ReLU (do inglês, *Rectified Linear Unit*):

$$f(z) = \max(z, 0) \quad (3.6)$$

### 3.1.2.2 Passo de retropropagação

O passo de retropropagação no modelo MLP permite o cálculo do gradiente e conseqüente ajuste dos pesos.

#### Custo a otimizar

O passo de retropropagação procura minimizar uma função de custo, dependente dos dados de treino e da saída da rede. Para problemas de classificação, recorre-se normalmente à função de entropia cruzada (3.2) [103].

O objectivo é obter o mínimo global de  $E$  (durante a fase de treino), dado que se pretende que o resultado actual da rede seja o mais próximo possível do resultado desejado (mas evitando o fenómeno de sobreaprendizagem).

#### Algoritmo de Gradiente - Optimização

De forma a otimizar o custo recorre-se ao algoritmo do gradiente [104]. Através deste algoritmo, os pesos são inicializados e repetidamente actualizados de acordo com a direcção do valor mínimo estimado [92]

$$w_{ij}(t+1) = w_{ij}(t) - \eta \frac{\partial E}{\partial w_{ij}}, \quad (3.7)$$

em que  $w_{ij}(t+1)$  é o valor do peso após o ajuste,  $w_{ij}(t)$  o peso actual (antes do ajuste),  $\eta$  o **ritmo de aprendizagem** (valor numérico que garante uma convergência mais rápida dos pesos na direcção do valor mínimo a estimar) e  $\frac{\partial E}{\partial w_{ij}}$  o gradiente. De notar que o ritmo de aprendizagem é um parâmetro muito importante no treino das redes neuronais, afectando a convergência do algoritmo do gradiente. Se o valor deste parâmetro for muito pequeno, a convergência pode ser extremamente lenta. Por outro lado, se for muito elevado, o gradiente poderá nunca estabilizar num mínimo (por vezes dá-se o caso de divergir, nunca permitindo a convergência do algoritmo) [105]. O gradiente pode ser calculado recorrendo à regra em cadeia

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial E}{\partial y_i} \frac{\partial y_i}{\partial w_{ij}}, \quad (3.8)$$

onde

$$\frac{\partial y_i}{\partial w_{ij}} = \frac{\partial y_i}{\partial z_i} \frac{\partial z_i}{\partial w_{ij}} \quad (3.9)$$

#### Cálculo do Gradiente - retropropagação

O algoritmo mais usado para efectuar o cálculo do gradiente é o de retropropagação [92], que foi desenvolvido em 1986 por Rumelhart [90, 99]. O algoritmo encontra-se esquematizado na Figura 3.3.

O algoritmo de retropropagação permite calcular sucessivamente o gradiente, propagando o erro desde a camada de saída até à de entrada. Porém, o cálculo do gradiente difere entre a camada de saída e as camadas escondidas relativamente ao termo  $\frac{\partial E}{\partial y_i}$ . O cálculo deste termo para a última camada, assumindo o uso da

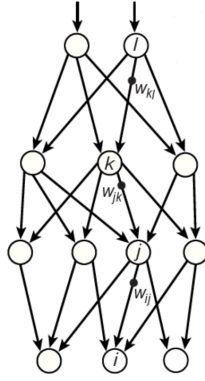


Figura 3.3: Passo de retropropagação de uma arquitectura MLP, extraído de [11].

função de entropia cruzada e a função de activação *softmax*, é determinado por (3.10) e para as camadas escondidas é dado por (3.11).

$$\frac{\partial E}{\partial y_i} = y_i - t_i, \quad (3.10)$$

em que  $i$  representa o neurónio da camada de saída

$$\frac{\partial E}{\partial y_i} = \sum_k \frac{\partial E}{\partial y_k} \frac{\partial y_k}{\partial y_i}, \quad (3.11)$$

onde

$$\frac{\partial y_k}{\partial y_i} = \frac{\partial y_k}{\partial z_k} \frac{\partial z_k}{\partial y_i} = f'(z_k) w_{ki}, \quad (3.12)$$

em que  $k$  corresponde a um neurónio da camada escondida mais próxima da camada de saída, com o qual o neurónio  $i$  da camada escondida imediatamente anterior tem uma ligação,  $w_{ki}$ .

Desta forma, é possível propagar o cálculo do gradiente até à camada de entrada e recorrer a 3.7 para actualizar os pesos das diferentes camadas. De referir ainda que há duas variantes do algoritmo de retropropagação a respeito do cálculo do gradiente e conseqüente ajuste dos pesos: gradiente estocástico e gradiente em mini-conjuntos.

A **variante do gradiente estocástico** selecciona aleatoriamente um único exemplo de treino, em cada iteração, para realizar o ajuste dos pesos, em vez de utilizar todos os exemplos de treino [106]. A vantagem deste método é a diminuição do tempo de computação por iteração: passa de  $n$  (exemplos) para um. Apesar desta simplificação, pretende-se na mesma que o comportamento da convergência para o erro mínimo global seja semelhante ao do gradiente, embora esta variante salte de mínimo local para um novo mínimo local (pois só tem em conta um elemento de cada vez) [104].

A **variante de gradiente em mini-conjuntos** procura juntar as vantagens da variante anterior e a do gradiente aplicado a todo o conjunto de treino [104]. Para tal, executa o ajuste dos pesos para cada subconjunto (ou mini-conjunto) de  $m$  elementos de treino. Esta variante deve ser usada para redes neuronais cujo o conjunto de treino é bastante abundante e redundante [105].

### Aceleração

Uma das técnicas para acelerar o treino das redes neuronais é a utilização de algoritmos de optimização do gradiente mais eficientes. Um dos algoritmos de optimização mais comumente utilizados é o *Adam* [107]. Este algoritmo recorre tanto à variante do gradiente estocástico como à variante de gradiente em mini-conjuntos e ajusta o valor do ritmo de aprendizagem para diferentes parâmetros a partir das estimativas dos dois primeiros momentos do gradiente (3.13), em cada iteração.

$$\theta_t = \theta_{t-1} - \eta \frac{\frac{m_t}{1-\beta_1^t}}{\sqrt{\frac{v_t}{1-\beta_2^t} + \epsilon}}, \quad (3.13)$$

onde  $\theta_t$  representa o vector de parâmetros a actualizar,  $\theta_{t-1}$  o vector de parâmetros da iteração anterior,  $\eta$  o ritmo de aprendizagem,  $m_t$  a estimativa do primeiro momento do gradiente,  $v_t$  a estimativa do segundo momento do gradiente,  $\beta_1^t$  e  $\beta_2^t$  os decaimentos exponenciais para as estimativas dos primeiro e segundo momentos.

### Regularização

Existem ainda outros métodos, técnicas de regularização, que podem ser aplicados durante a fase de treino, quando o conjunto de dados é reduzido. Estas técnicas de treino permitem que o treino (e, conseqüentemente, a generalização) das redes seja mais eficiente, pois previnem o fenómeno de sobreaprendizagem [108, 103]. Dois exemplos destas técnicas são decaimento dos pesos da rede e *dropout* [108].

A primeira técnica consiste em diminuir a gama de valor dos pesos por um factor proporcional à sua magnitude em cada iteração da fase de treino [109, 103]. Esta regularização é feita através da inserção de um termo à função de custo [109]. Este termo de regularização pode ser definido, por exemplo, segundo a norma  $L_1$  (3.14) ou a norma  $L_2$  (3.15) [103].

$$E(w_{ij}) = E_o(w_{ij}) + \frac{1}{2} \lambda \sum_{ij} |w_{ij}|, \quad (3.14)$$

$$E(w_{ij}) = E_o(w_{ij}) + \frac{1}{2} \lambda \sum_{ij} w_{ij}^2, \quad (3.15)$$

onde  $E_o(w_{ij})$  refere-se à função de custo original e  $\lambda$  um parâmetro de penalização dos pesos.

A técnica de *dropout* consiste em retirar temporária e aleatoriamente alguns neurónios (e respectivas ligações) da rede neuronal consoante uma certa probabilidade fixa  $p$  [98], somente durante o treino da rede.

## 3.2 Arquitectura de uma CNN

Nos MLP cada uma das camadas está totalmente ligada quer à camada anterior, quer à seguinte. Por este motivo, o número de parâmetros a optimizar pode ser muito elevado, o que requer um conjunto de treino com muitos exemplos, assim como um elevado poder computacional [110].

Por exemplo, uma imagem contém muitos pixéis, pelo que são necessários muitos parâmetros para a modelar. Uma imagem é representada por uma matriz de dimensões *altura*  $\times$  *largura*  $\times$  *profundidade* (em

que a profundidade corresponde ao número de canais de cores da imagem. Se a imagem for a cores, RGB, então a profundidade é igual a três), em que cada entrada representa um valor de intensidade de um pixel, que varia entre 0 e 255.

As redes neuronais convolucionais, CNN ou ConvNet, são o tipo de rede neuronal especializado em processar imagens, permitindo superar as limitações acima descritas. Recorrendo a um conjunto de imagens, as CNN têm a capacidade de aprender quais são as características mais significativas de cada imagem, começando por identificar características de baixo nível (como curvas, arestas e cantos) e, posteriormente, transformá-las em representações de nível mais abstracto (como objectos ou partes de objectos) [93, 11].

Um exemplo de uma CNN pode ser visto na Figura 3.4. Este tipo de redes substitui camadas totalmente interligadas por camadas definidas por operações de convolução entre a saída da camada anterior e um conjunto de filtros [93]. À semelhança do MLP, é aplicada uma função de activação de não-linearidade à saída de cada camada de convolução. A função de activação mais popular actualmente é a função ReLU por ter demonstrado mais benefícios quando comparado com os das funções tangente hiperbólica [13] e logística [108]. Normalmente, a seguir a uma camada de convolução (ou a um pequeno conjunto destas camadas), o sinal passa por uma camada de *pooling*, que permite a redução espacial desse sinal. As últimas camadas numa CNN são denominadas por totalmente interligadas, sendo que na última camada a função ReLU é normalmente substituída por uma *softmax*, que permite associar uma probabilidade a cada uma das classes de saída. As camadas de convolução, *pooling* e totalmente interligadas são explicadas nas Secções 3.2.1, 3.2.2 e 3.2.3.

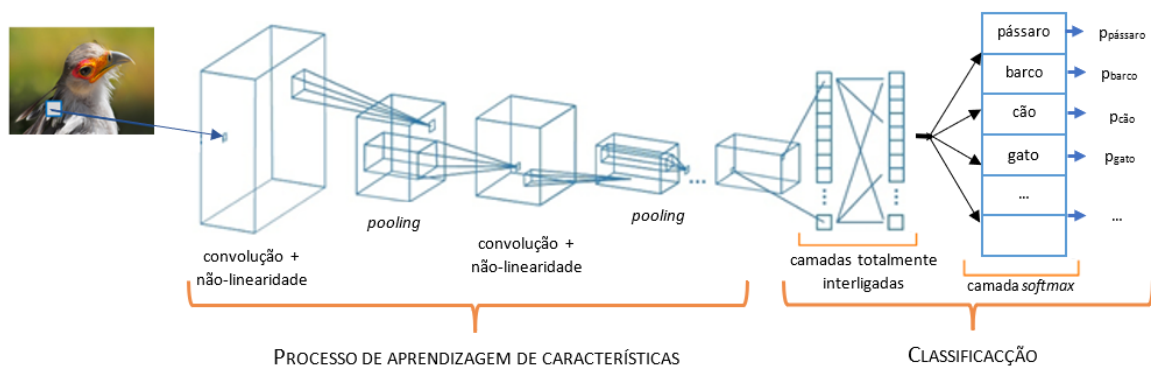


Figura 3.4: Arquitectura de uma CNN simples <sup>1</sup>.

### 3.2.1 Camada de convolução

As camadas de convolução correspondem à aplicação de um conjunto de filtros à imagem inicial ou à saída da camada anterior. A imagem inicial e a saída da camada anterior são referidas como matriz doravante. Esta aplicação é conseguida através da convolução entre cada filtro e a matriz, pelo que a profundidade do filtro tem de coincidir com a profundidade da matriz. O resultado da convolução é denominado por mapa de características. A Figura 3.5 mostra a convolução entre um filtro e a imagem de entrada e o respectivo mapa de características.

<sup>1</sup>Esta figura foi inspirada em duas figuras retiradas nos seguintes blogues: [www.adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks](http://www.adeshpande3.github.io/adeshpande3.github.io/A-Beginner's-Guide-To-Understanding-Convolutional-Neural-Networks) e [www.medium.com/RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148](http://www.medium.com/RaghavPrabhu/understanding-of-convolutional-neural-network-cnn-deep-learning-99760835f148)



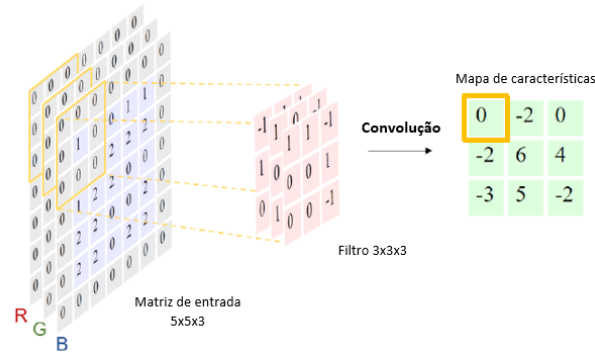


Figura 3.5: Representação de um filtro de tamanho  $3 \times 3 \times 3$  a convolver uma imagem de entrada  $5 \times 5 \times 3$ , resultando um mapa de características de  $3 \times 3 \times 1$ . Esta convolução utiliza um passo de dois pixéis ( $S = 2$ ) e um zero ( $P = 1$ ). Adaptado de [12].

Antes de continuar com a explicação das camadas de convolução, é importante realçar que estas são caracterizadas por um conjunto de hiper-parâmetros, *i.e.*, parâmetros que podem ser definidos pelo utilizador com o objectivo de melhorar o desempenho de classificação da rede [111]. Este conjunto de hiper-parâmetros consiste na seguinte lista:

- Número de filtros;
- Dimensão da máscara dos filtros;
- Profundidade do filtro ( $F$ ): à semelhança da profundidade da imagem, corresponde ao número de canais do filtro;
- Passo de filtro ( $S$ ): corresponde ao número de pixéis deslocados sobre a matriz de entrada;
- Número de zeros ( $P$ ): por vezes, há filtros que não encaixam totalmente na imagem inicial ou na saída da camada anterior. Por isso, a operação de *zero-padding* permite lidar com os pixéis na fronteira dos mapas de activação. Quando é utilizado, permite processar esses pixéis; caso contrário, os pixéis serão descartados. Deste modo, esta operação permite controlar as dimensões do mapa de saída [93].

Cada filtro é aplicado a toda a matriz, multiplicando os pesos do filtro com as entradas da matriz. Cada multiplicação contribui com um valor para o mapa de características. As dimensões de altura ( $H_{mc}$ ) e largura ( $W_{mc}$ ) de cada mapa de características depende da altura ( $H_{matriz}$ ), da largura ( $W_{matriz}$ ), de  $F$ ,  $S$  e  $P$ . A altura do mapa de características é dada por (3.16) e a largura do mesmo é dada por (3.17).

$$H_{mc} = \frac{H_{matriz} - F + 2P}{S} + 1 \quad (3.16)$$

$$W_{mc} = \frac{W_{matriz} - F + 2P}{S} + 1 \quad (3.17)$$

Um mapa de características contém, assim, as características extraídas por um mesmo filtro. O número total de mapas de características é igual ao número de filtros utilizados. Em cada camada de convolução, cada unidade está ligada a regiões locais dos mapas de características das camadas anteriores através de pesos, aos quais se dá o nome de "banco de filtro" [11].

O nível de abstracção das características extraídas vai aumentando sempre que é adicionada uma camada

de convolução, visto que os filtros vão sendo aplicados consecutivamente sobre características anteriormente extraídas [110].

Cada mapa de características numa camada de convolução é submetido a uma função de activação. No caso das CNN, a camada não-linear de activação mais utilizada é a ReLU.

### 3.2.2 Camada de *pooling*

As camadas *pooling* são geralmente intervaladas entre uma sequência de camadas de convolução. O principal objectivo das camadas *pooling* é reduzir o número total de parâmetros da rede, diminuindo, assim, a complexidade do modelo (e, conseqüentemente, diminui o tempo e a memória de computação) e também prevenindo o fenómeno de sobreaprendizagem pois reduz o número de características [112]. Esta redução espacial realizada em cada mapa de características retém a informação mais importante [113], pois recorre a um filtro que selecciona um determinado valor do mapa de características. Este valor é escolhido conforme a função de *pooling* usada. As duas funções de *pooling* mais utilizadas são: máximo, cujo o valor determinado é o máximo de todos os valores do mapa de características [114, 115]; média, cujo valor determinado corresponde à média aritmética de todos os valores [115]. A função de máximo é a mais utilizada e está demonstrada na Figura 3.6.

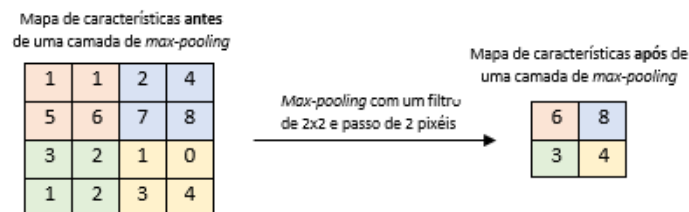


Figura 3.6: Operação de *max-pooling*, em que o mapa de características, à esquerda, é reduzido através de um filtro de dimensão  $2 \times 2$  e passo de dois pixéis, produzindo o mapa de característica à direita de dimensão  $2 \times 2$ . Adaptado de [12].

### 3.2.3 Camada totalmente interligada

Geralmente as camadas totalmente interligadas são as últimas camadas da rede, tendo uma estrutura semelhante à de um MLP, nas quais cada neurónio está ligado a todos os neurónios das camadas anteriores e posteriores. O principal objectivo deste tipo de camadas é a classificação, convertendo as características extraídas pelas camadas de convolução e *pooling* em classes [116] (e.g. pássaro, barco, cão ou gato, como no caso da Figura 3.4). Por outras palavras, procuram combinar todas as características aprendidas pela rede para classificar cada imagem atribuindo-lhe uma classe. É através da última camada totalmente interligada que se consegue atribuir classes, visto que é submetida a uma função de *softmax*. Esta função permite que a soma de todos os valores do vector seja igual a um e que cada valor emita uma certa probabilidade da imagem de entrada pertencer a uma certa classe.

### 3.3 Concurso ILSVRC

As CNN requerem uma base de dados muito grande de imagens anotadas para serem capazes de aprender a extrair automaticamente o conjunto de características relevantes. A revolução na utilização destas redes deu-se em 2012, graças ao concurso *ImageNet Large-Scale Visual Recognition Challenge* (ILSVRC) [11], no qual está associada a base de dados ImageNet. Esta base de dados contém milhões de imagens de cerca de 1000 categorias diferentes. Adicionalmente, o aparecimento de GPUs, que permitem acelerar o treino das redes, também impulsionou a sua utilização.

O projecto ImageNet surgiu em 2010 com dois objectivos [44]: criar uma base de dados para problemas de detecção de objectos em grande escala e incentivar o desenvolvimento de algoritmos dedicados à detecção de objectos e à classificação de imagens, promovido pelo concurso ILSVRC. Este desafio incentiva anualmente a competição entre equipas de investigação no desenvolvimento de algoritmos de classificação e detecção de objectos, disponibilizando uma versão reduzida da base de dados ImageNet. O desempenho de cada equipa é avaliado através do critério *top-5 error rate*, que indica a fracção de imagens de teste cuja categoria correcta não é nenhuma das cinco mais prováveis dadas pelo modelo.

Até 2012, os vencedores do concurso foram equipas que apresentaram técnicas de características artesanais. Desde 2012, os algoritmos vencedores são baseados em CNN. A primeira CNN apresentada foi a AlexNet que reduziu o *top-5 error rate* de 26.1% para 15.3% [13].

Os modelos apresentados nas edições de 2012 (AlexNet, [13]), 2014 (VGG, [14]) e 2015 (ResNet, [15]) são brevemente detalhados nas seguintes Secções 3.3.1, 3.3.2 e 3.3.3.

#### 3.3.1 AlexNet

A rede AlexNet [13] foi a vencedora da edição de 2012, tendo atingido um *top-5 error* de 15,3%. Os resultados obtidos pela AlexNet foram justificados pela função de activação empregue (ReLU), pela utilização da técnica de regularização de *dropout* e pelo aumento do número de imagens do conjunto de treino (técnica de aumento de dados) através de transformações geométricas e de cor. O uso de GPUs permitiu treinar o modelo mais rapidamente.

A AlexNet recebe imagens de  $224 \times 224 \times 3$  e é formada por oito camadas: cinco de convolução (entre algumas das quais se intercalam três camadas de *pooling*) e três totalmente interligadas. Todas as camadas de *pooling* usam a função de *max-pooling*, com filtros de tamanho  $3 \times 3$  e passo de dois píxeis. À excepção da primeira camada de convolução, que tem passo de quatro píxeis, todas as camadas de convolução têm passo de dois píxeis. De notar que algumas camadas de convolução (segunda, quarta e quinta) foram divididas ao meio, devido à inserção de dois GPUs em paralelo. Assim, os filtros destas camadas estão apenas ligados aos mapas de características da camada anterior que residem no mesmo GPU. As restantes camadas da AlexNet estão ligadas a todos os neurónios das camadas anteriores.

A Figura 3.7 mostra com maior detalhe a arquitectura desta CNN.

A primeira camada de convolução contém 96 filtros com dimensão de  $11 \times 11 \times 3$ . De seguida, têm-se três camadas, a de activação ReLU, uma de normalização local e, por fim, a primeira camada de *max-pooling*. A segunda camada de convolução é inserida com 256 filtros de tamanho  $5 \times 5 \times 48$  por GPU (96 no total).

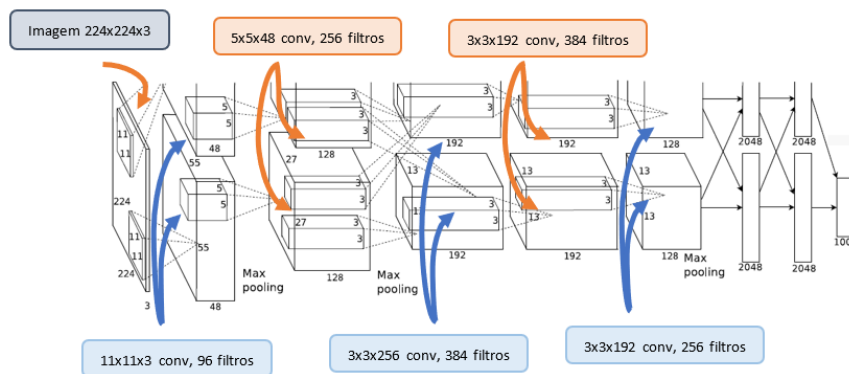


Figura 3.7: Arquitetura da AlexNet, adaptado de [13].

Posteriormente, é seguida por uma sequência de três camadas: ReLU, normalização local e, por fim, a segunda camada de *max-pooling*. A terceira camada de convolução é dividida em duas e, no total, tem 384 filtros de tamanho  $3 \times 3 \times 256$ . A quarta camada de convolução, também dividida em duas, percorre os mesmos número e tamanho de filtros que a camada anterior, mas com uma profundidade igual a 192 por GPU (384 no total). A quinta e última camada de convolução, igualmente dividida em duas, contém no total 256 filtros de tamanho  $3 \times 3 \times 192$  por GPU. Posteriormente, o sinal de saída desta camada é processado por uma camada de *max-pooling*, sendo depois convertido a um vector de 4096 unidades (*i.e.*, a uma camada totalmente interligada de dimensão  $1 \times 1 \times 4096$ ). A segunda camada totalmente interligada tem a mesma dimensão que a camada anterior e sofre igualmente *dropout* com a mesma probabilidade. A última camada contém 1000 neurónios de forma a classificar entre 1000 objectos/classes diferentes (recorrendo, para tal, à função de *softmax*).

### 3.3.2 VGG

VGG foi uma das CNN apresentadas na edição de 2014 do concurso ILSVRC. Embora não tenha sido a vencedora, a VGG demonstrou a importância da profundidade e como esta propriedade influencia o desempenho das CNN [14].

O aumento de profundidade, introduzido por esta rede, conseguido através do uso de camadas com filtros de reduzidas dimensões ( $3 \times 3$ ), o que leva à diminuição do número de parâmetros a estimar quando se compara com a arquitetura da AlexNet. A escolha de tamanho do filtro para a VGG deve-se ao facto de este ser o menor tamanho possível capaz de captar os conceitos de esquerda/direita, cima/baixo e centro [14].

Existem várias configurações deste tipo de rede, que se diferenciam maioritariamente pelo número de camadas de convolução (no caso da configuração VGG com 16 camadas, existem dois modelos: um em que os filtros são de tamanho  $3 \times 3$  e outro que têm dimensão  $1 \times 1$ ). Para qualquer uma das configurações, todas as camadas de convolução apresentam um passo de um pixel e as de *pooling* de dois pixels. A função de *pooling* utilizada é a de máximo, *max-pooling*. Ao contrário da AlexNet, não existem camadas de normalização local após cada camada de convolução.

A Figura 3.8 ilustra com maior pormenor a arquitetura da VGG-16 com filtros de dimensão  $3 \times 3$ . A VGG-16 contém 16 camadas: treze camadas de convolução, que se diferenciam pelo número de filtros usados, e três totalmente interligadas, em que as primeiras duas contêm 4096 unidades e a última 1000.

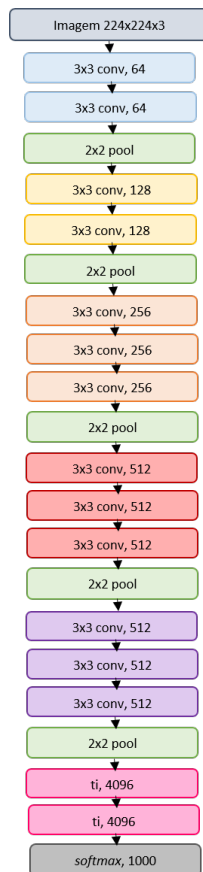


Figura 3.8: Representação da arquitectura VGG-16 ("ti" representa uma camada totalmente interligada). Adaptado de [14].

A não-linearidade ReLU é aplicada a cada sinal de saída de todas as camadas de convolução e totalmente interligadas. Existem cinco camadas de *max-pooling* com filtro de  $2 \times 2$ , no fim das seguintes camadas de convolução: segunda, quarta, sétima, décima e décima terceira. A última camada totalmente interligada passa pela função de *softmax* com o intuito de o modelo classificar entre 1000 classes diferentes.

### 3.3.3 ResNet

O modelo ResNet foi o vencedor na edição de 2015 do concurso ILSVRC por ter atingido o valor mais baixo de *top-5 error*, 3,57% [15]. À semelhança da VGG, a ResNet também pretende tirar partido da profundidade para melhorar o desempenho. Todavia, à medida que a profundidade aumenta, torna-se mais difícil treinar a CNN, devido ao gradiente tender para zero nas camadas iniciais, e maior é a sua complexidade. Deste modo, a ResNet procura solucionar o problema inerente à profundidade através da criação de blocos residuais (ver Figura 3.9). Esta inovação permite saltar algumas ligações na rede, à medida que a profundidade aumenta. Quando comparada com a VGG, a redução de complexidade não só se deve à introdução dos blocos residuais como também à eliminação das camadas totalmente interligadas, com excepção da última associada à classificação. Por estes dois motivos, a ResNet apresenta melhores resultados de classificação de imagens e detecção de objectos [15].

A ideia por detrás dos blocos residuais é ignorar algumas camadas da rede, através da minimização do resíduo entre a entrada e a saída de algumas camadas. De maneira a compreender este processo, é importante

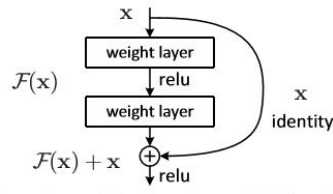


Figura 3.9: Estrutura de um bloco residual [15].

esclarecer primeiro os seguintes conceitos: resíduo e função de identidade. Supondo que a entrada de um bloco é  $x$  e que a saída verdadeira é dada por  $H(x)$ , a função de identidade mapeia  $H(x) = x$ , i.e., a entrada de um bloco é igual à sua saída. O resíduo corresponde à diferença entre  $x$  e  $H(x)$  e é dado por  $F(x)$  ( $F(x) = H(x) - x$ ). Como se pretende a saída do bloco,  $H(x) = F(x) + x$ . Desta forma, para saltar blocos, basta anular  $F(x)$  ( $F(x) = 0$ ) para que  $H(x) = x$ .

A principal diferença entre a ResNet e as CNN apresentadas anteriormente é que na primeira as camadas são treinadas para aprender o resíduo entre a entrada e saída de blocos, enquanto que nas restantes CNN as camadas aprendem directamente  $H(x)$ .

Existem várias configurações da ResNet, cujas diferenças relacionam-se pelo número de camadas de convolução. A configuração original contém 152 camadas: 151 camadas de convolução e uma totalmente interligada. A Figura 3.10 mostra a arquitectura da ResNet de 50 camadas: 49 camadas de convolução e uma totalmente interligada.

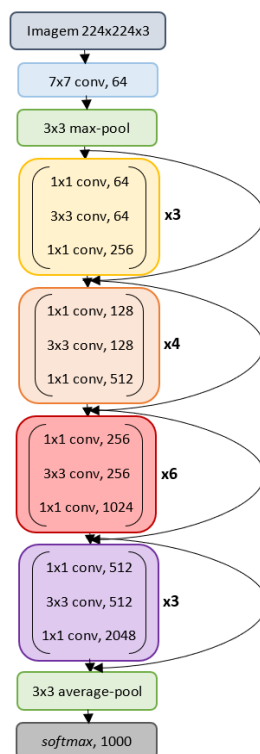


Figura 3.10: Arquitectura do modelo ResNet-50, adaptada de [15]. Os números indicados ao lado de cada bloco (e.g.,  $x3$ ) indicam o número de vezes em que esse bloco se repete.

A primeira camada de convolução tem 64 filtros de tamanho  $7 \times 7$ , passo de dois pixéis e é procedida por uma camada de *max-pooling* com filtro  $3 \times 3$  e passo de dois pixéis. Até à camada totalmente interligada

(última camada), as camadas de convolução estão organizadas por blocos residuais, em que cada bloco contém uma sequência de três camadas. O número de filtros em cada camada de convolução varia, mas há um padrão relativamente ao tamanho dos filtros em todos os blocos residuais:  $1 \times 1 - 3 \times 3 - 1 \times 1$ . A última camada da ResNet-50 é precedida por uma camada de *average-pooling*. Por fim, o resultado desta camada passa por uma camada de *softmax*, de forma a classificar imagens entre 1000 classes possíveis. De referir ainda que nesta CNN também se utilizam camadas de ReLU e normalização por conjunto (o tipo de normalização é diferente dos apresentados pelas CNN anteriores [117]), após cada camada de convolução.





# 4

## Formulação do problema e sistema proposto

Este Capítulo descreve o problema abordado nesta dissertação e descreve a solução proposta.

### 4.1 Formulação do Problema

Esta dissertação aborda o diagnóstico de melanomas em imagens dermatoscópicas. Em particular, pretende-se desenvolver um método para distinguir melanomas de outros tipos de lesões, nevos e queratoses seborreicas, que serão ambas consideradas como lesões benignas. Desta forma, formula-se o diagnóstico de melanomas como um problema de classificação binário. Pretende-se abordar este problema recorrendo a redes neuronais profundas, uma vez que estas têm obtido resultados promissores em diversas áreas.

É possível recorrer a métodos de aprendizagem profunda, uma vez que se tem acesso a uma base de dados recente, com 2750 imagens dermatoscópicas [61], contendo melanomas, nevos e queratoses seborreicas.

O problema apresenta várias dificuldades. As diferenças entre lesões benignas e malignas são, por vezes, muito subtis e difíceis de discriminar de forma automática. Mas há outras dificuldades relacionadas com o método de aquisição das imagens. As imagens contêm artefactos (por exemplo, pêlos e objectos estranhos) que podem influenciar a decisão. Outra agravante neste problema é o facto de a base de dados apresentar um conjunto de treino desequilibrado, visto que o número de lesões benignas é muito superior ao de lesões malignas.

### 4.2 Arquitectura do Sistema Proposto

A Figura 4.1 apresenta a arquitectura do sistema proposto para classificar lesões em benignas ou malignas.

No primeiro esquema (parte superior) está representada a fase de treino e no segundo a fase de classificação. Para este sistema foram consideradas, uma de cada vez, três redes neuronais diferentes.

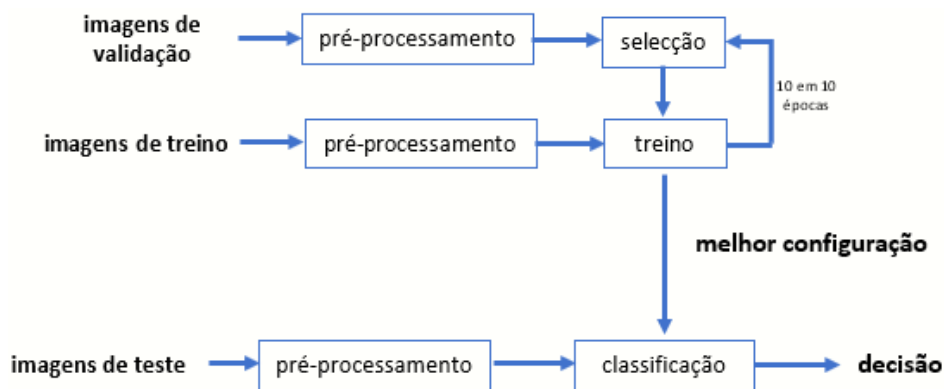


Figura 4.1: Diagrama do sistema desenvolvido.

Todas as imagens da base de dados (conjuntos de treino, validação e teste) são previamente submetidas a uma fase de pré-processamento para que contenham somente a região de lesão da pele. De seguida, segue-se a fase de treino. Começa-se por aumentar artificialmente o conjunto de treino pré-processado, de forma a equilibrar os exemplos das duas classes. Posteriormente, procede-se ao treino da rede neuronal, a partir do qual se pretende que a rede aprenda a distinguir lesões benignas de melanomas. De dez em dez épocas, o desempenho da rede que está a ser treinada é avaliado recorrendo ao conjunto de validação, previamente pré-processado, através de duas métricas: perda e erro balanceado. Esta fase termina com a selecção da melhor configuração de rede. Por fim, tem-se a fase de classificação, na qual se recorre ao conjunto de teste pré-processado e à melhor configuração da rede. Cada imagem de teste é classificada como benigna ou maligna usando a rede treinada.

#### 4.2.1 Pré-processamento

Dado que o módulo a seguir ao pré-processamento recorre a redes neuronais convolucionais, é necessário que o conjunto de dados a fornecer às CNN contenha certas características. Por este motivo, este bloco do diagrama tem a função de pré-processar e redimensionar todas as imagens. Deste modo, as imagens sofreram as seguintes transformações:

- Segmentação: O primeiro passo é a segmentação, no qual se faz o cálculo da máscara binária que distingue os pixels da lesão dos da pele saudável. Neste trabalho foram utilizadas máscaras de segmentação manuais, feitas por dermatologistas.
- Recorte: Para focar a rede na região de interesse (lesão), optou-se por recortar as imagens recorrendo a caixas delimitadoras rectangulares (ver Figura 4.2). Para tal, utilizam-se as máscaras de segmentação de cada imagem, a partir das quais se calculam as dimensões de cada caixa e, a partir dessas dimensões, recorta-se cada imagem.

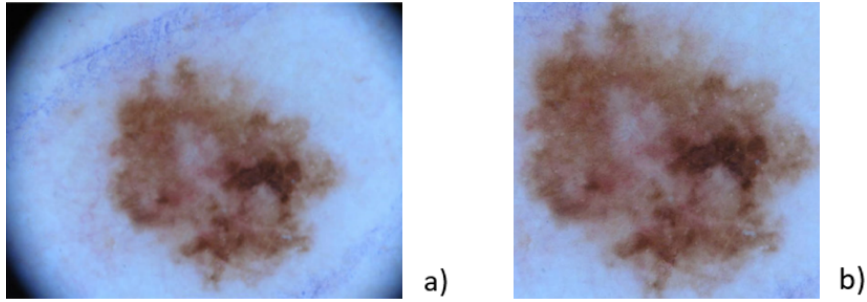


Figura 4.2: Imagem dermoscópica original, à esquerda em a), e a mesma imagem recortada com base na sua caixa delimitador rectangular, à direita em b).

- Redimensionamento: A imagem de entrada das CNN utilizadas tem dimensões fixas ( $224 \times 224$  pixéis). Por isso, cada imagem recortada é depois redimensionada.

- Normalização da imagem: Cada CNN usada nesta dissertação pressupõe uma normalização específica. No caso da AlexNet, é subtraído a cada pixel a intensidade média calculada a partir de todas as imagens do conjunto de treino e, de seguida, esse resultado é dividido pelo desvio-padrão (ver 4.1, onde  $i$  representa um dos três canais de cor,  $x$  o valor de intensidade de um pixel,  $x_{\text{médio}}$  o valor-médio para o pixel  $x$  e  $\sigma_i$  o desvio-padrão de cada canal). Quando é usada a VGG ou a ResNet, o método de normalização é o de subtracção da média (4.2, onde  $x_i$  é o valor de um pixel do canal  $i$  e  $x_{\text{médio}_i}$  o valor médio do pixel  $x$  do canal  $i$ ).

$$x_i < -\frac{x_i - x_{\text{médio}_i}}{\sigma_i} \quad (4.1)$$

$$x_i < -x_i - x_{\text{médio}_i} \quad (4.2)$$

## 4.2.2 Aumento de dados

O conjunto de imagens de treino utilizado é desequilibrado em termos do número de lesões malignas face ao das benignas. Para conjuntos de dados desequilibrados é frequente recorrer-se à técnica de aumento dados, pois tende a melhorar significativamente o desempenho do classificador e ainda a prevenir a sobreaprendizagem [118].

Deste modo, aumentou-se cinco vezes o conjunto original de lesões malignas de treino, através de rotações de  $90^\circ$  e  $180^\circ$  e de inversões verticais e horizontais (ver Figura 4.3).

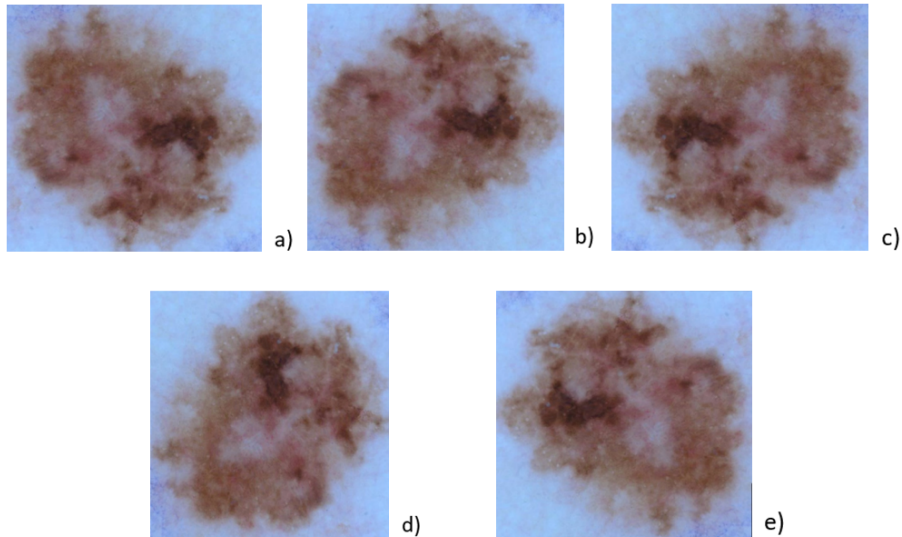


Figura 4.3: Aumento de dados de treino. A primeira imagem (mais à esquerda, a)) representa a imagem original de uma lesão maligna. As restantes imagens representam as inversões horizontal, b), e vertical, c), e as rotações de 90°, d), e 180°, e).

Para além das transformações de aumento de dados, é associado um peso às lesões benignas tendo em conta o número de imagens de lesões benignas no conjunto aumentado (4.3) e outro peso associado às lesões malignas tendo em conta o número de imagens de lesões malignas no conjunto aumentado (4.4).

$$\text{peso}_{\text{benigno}} = \frac{\text{número total do conjunto aumentado de treino}}{\text{número de imagens de lesões benignas no conjunto aumentado}} \quad (4.3)$$

$$\text{peso}_{\text{maligno}} = \frac{\text{número total do conjunto aumentado de treino}}{\text{número de imagens de lesões malignas no conjunto aumentado}} \quad (4.4)$$

### 4.2.3 Classificação

A classificação das lesões da pele é feita através do uso de uma rede CNN. Nesta dissertação foram testadas três arquitecturas: AlexNet [13], VGG [14] e ResNet [15]. Estas redes foram escolhidas pelo seu desempenho no concurso ILSVRC nos anos 2012 (ganhou), 2014 (neste caso não foi a vencedora, mas foi uma das CNN mais promissoras nesta edição) e 2015 (ganhou), respectivamente.

Porém, os modelos de CNN utilizados são simplificações das arquitecturas originais, visto que, como o número de exemplos de treino é reduzido, é necessário reduzir o número de pesos a treinar. Assim, referente à AlexNet original (ver Figura 3.7), as três camadas totalmente interligadas de 4096, 4096 e 1000 neurónios são substituídas por uma camada totalmente interligada de duas unidades apenas. Relativamente à arquitectura original da VGG (ver Figura 3.8), as três últimas camadas totalmente interligadas de 4096, 4096 e 1000 unidades são substituídas também por uma camada totalmente interligada de dois neurónios. Quanto à arquitectura original da ResNet (ver Figura 3.10), a camada totalmente interligada de passa a ter apenas dois em vez 1000 neurónios pois o problema de classificação estudado nesta dissertação é um problema binário.

#### 4.2.4 Treino

De uma forma geral, a fase de treino das três CNN é executada da mesma forma. Todavia, o número de épocas de treino diferiu entre as três arquitecturas. Os pesos são inicializados através de transferência de aprendizagem e não de forma aleatória, visto que o conjunto de dados é reduzido. Desta forma, importaram-se os pesos de redes pré-treinadas na base de dados ImageNet. No passo de retropropagação, a actualização dos pesos é definida pelo optimizador Adam (3.13). O cálculo do gradiente é executado usando o algoritmo de *Retropropagação*, cuja função de custo a minimizar é a de entropia cruzada (3.2) por ser a função de erro mais usada para problemas de classificação [103]. Como o uso de todos os padrões de treino em cada iteração (modo *batch*) é lento, recorre-se a subconjuntos mais pequenos para estimar o gradiente em cada iteração (modo *mini-batch*).

A melhor configuração de cada CNN é seleccionada com base numa das duas seguintes métricas, avaliadas no conjunto de validação a cada 10 épocas: erro balanceado e perda. A primeira métrica é definida por  $1 - BACC$ , onde  $BACC$  é a precisão balanceada definida na Secção 5.2. A segunda métrica corresponde à função de custo utilizada (entropia cruzada).

Para cada uma das redes foram treinadas várias configurações, tendo em conta os seguintes hiper-parâmetros: ritmo de aprendizagem (variável ou fixo) e utilização de *dropout*, com diferentes probabilidades.



# 5

## Resultados Experimentais

Este Capítulo é dedicado aos resultados experimentais. Primeiro, começa por detalhar a base de dados utilizada. Segundo, enuncia as quatro métricas de avaliação de desempenho usadas. De seguida, descreve os resultados experimentais obtidos para cada CNN. Por fim, avalia o desempenho da solução desenvolvida na presente dissertação.

### 5.1 Base de Dados

O conjunto de imagens utilizado neste trabalho é a base de dados disponibilizada em 2017 pelo ISIC (*International Skin Imaging Collaboration*) para suporte a um concurso internacional. Esta iniciativa surgiu em 2016, com o apoio da IBM, com o intuito de promover o desenvolvimento de algoritmos de diagnóstico automático de melanomas, disponibilizando, para tal, uma grande base de dados pública de imagens dermatoscópicas de lesões da pele [59, 61], anotadas por médicos especialistas.

A base de dados ISIC 2017 contém 2000 imagens de treino, 150 de validação e 600 de teste. Na Tabela 5.1 é possível observar a divisão de cada um destes conjuntos de imagens. De notar que as lesões benignas contêm exemplos de nevos e queratoses seborreicas. Nomeadamente, o conjunto de treino contém 254 queratoses, o de validação 42 e o de teste 90. Na Figura 5.1 encontram-se exemplos para cada tipo de lesão, acompanhado da sua máscara de segmentação.

	Lesões melanocíticas	
Conjunto de imagens	Melanomas	Não-melanomas
Treino	374	1626
Validação	30	120
Teste	117	483

Tabela 5.1: Base de dados de imagens dermatoscópicas ISIC 2017.

Cada lesão foi diagnosticada por dermatologistas experientes. Adicionalmente, para cada lesão é providenciada a respectiva máscara de segmentação binária (ver Figura 5.1). Estas foram obtidas recorrendo a

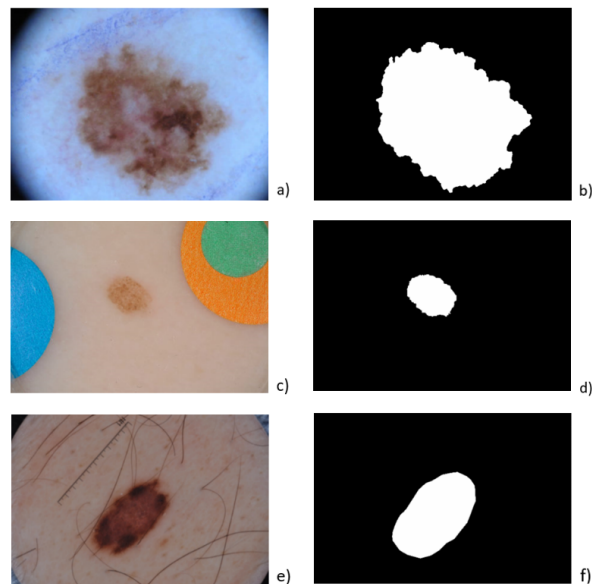


Figura 5.1: Exemplos extraídos da base de dados ISIC 2017. Em cima, imagem dermatoscópica maligna do conjunto de treino (à esquerda) e respectiva segmentação binária (à direita); no centro, imagem dermatoscópica benigna do conjunto de validação (à esquerda) e respectiva segmentação binária (à direita) e imagem dermatoscópica benigna do conjunto de teste (à esquerda) e respectiva segmentação binária (à direita).

segmentações manuais e métodos semi-automáticos, ambas coordenadas por dermatologistas. Estas imagens de segmentação permitem obter as dimensões das caixas delimitadoras rectangulares.

A base de dados contém as imagens originais de cada lesão cutânea e o diagnóstico clínico para cada imagem, *i.e.*, o tipo de classe de cada imagem de entrada (benigno ou maligno) a fornecer a cada CNN. O veredicto clínico acerca das lesões malignas resultou do consenso de especialistas e de informações patológicas. De mencionar ainda que, embora a base de dados contenha três tipos de classes diferentes (lesões melanocíticas benignas e malignas e queratoses seborreicas), a decisão realizada pelas CNN é binária, discriminando melanomas dos outros dois tipos de lesões benignas.

## 5.2 Métricas de avaliação de desempenho

Existem quatro tipos de resultados de um classificador binário: verdadeiro positivo (TP), verdadeiro negativo (TN), falso positivo (FP) e falso negativo (FN). Para este tipo de classificadores, existem algumas métricas que permitem avaliar o seu desempenho. Nesta dissertação recorreu-se às seguintes métricas: sensibilidade, especificidade, precisão balanceada e área por baixo da característica de operação.

- Verdadeiro positivo: exemplo positivo correctamente classificado como positivo.
- Verdadeiro negativo: exemplo negativo correctamente classificado como negativo.
- Falso positivo: exemplo negativo classificado como positivo.
- Falso negativo: exemplo positivo classificado como negativo.



A **sensibilidade** (SE), também apelidada de taxa de verdadeiros positivos, ilustra a proporção de classificações correctas de casos positivos (5.1).

$$\text{sensibilidade} = \frac{TP}{(TP + FN)} \quad (5.1)$$

A **especificidade** (SP), também conhecida por taxa negativa verdadeira, quantifica a proporção de classificações correctas de casos negativos tendo em conta toda a população negativa (5.2).

$$\text{especificidade} = \frac{TN}{(TN + FP)} \quad (5.2)$$

A **precisão balanceada** (BACC) exprime a precisão média obtida para cada uma das classes [119] e é utilizada quando o conjunto de dados é desequilibrado numa das classes (5.3).

$$BACC = \frac{1}{2} \left( \frac{TP}{TP + FP} + \frac{TN}{TN + FN} \right) \quad (5.3)$$

A **curva de característica de operação do receptor** (ROC) [120] é uma curva de probabilidade. Esta métrica relaciona graficamente a sensibilidade e a probabilidade de falso alarme (que corresponde a  $1 - \text{especificidade}$ ) do teste de classificação. A **área por baixo da curva ROC** (ROC-AUC) é uma medida de separabilidade. Por outras palavras, esta métrica indica a capacidade do classificador em distinguir entre classes. Quanto maior for o valor de ROC-AUC, melhor é o modelo a classificar exemplos positivos como positivos e exemplos negativos como negativos. Se ROC-AUC=1, o classificador distingue perfeitamente exemplos positivos de negativos. Se ROC-AUC=0.7, significa que 70% de hipótese de o classificador diferenciar entre classe positiva e a classe negativa. Se ROC-AUC=0.5, o classificador não é capaz de distinguir a classe positiva da negativa.

### 5.3 Treino das CNN

O treino das redes neuronais foi feito com o conjunto de treino da base de dados ISIC com 1626 exemplos positivos e 374 exemplos negativos (ver Tabela 5.1). Quando se usou aumento de dados o número de padrões negativos passou a ser de 1870.

A validação e o teste das redes foram feitos com os conjuntos de validação e teste descritos na Tabela 5.1.

Referente ao equipamento usado para a realização da dissertação, foram disponibilizados dois computadores cada um com um processador Intel Core i7 com 16GB de memória RAM, sem o recurso a uma placa adicional de GPU. A implementação das CNN é feita na linguagem Python com a biblioteca TensorFlow [121], que é uma ferramenta para modelos de aprendizagem automática, tais como as redes neuronais convolucionais.

O treino é realizado tal como foi descrito na Secção 4.2.4. O número de épocas e o número de subconjuntos do conjunto de validação, aquando da selecção da melhor configuração, são dois parâmetros que se alteram conforme a CNN utilizada.

A Tabela 5.2 indica o número de épocas durante o qual cada CNN foi treinada. O número de épocas é diferente para cada uma delas, pois cada uma tem um ritmo de convergência diferente devido à sua arquitectura e ao respectivo desempenho de classificação.

Tabela 5.2: Número de épocas definido para cada CNN utilizada.

CNN	Número de épocas
AlexNet	1000
VGG	600
ResNet	500

A AlexNet é a CNN que mais épocas tem, pois demora mais tempo a convergir, e a ResNet é a que menos épocas tem. Estima-se que a convergência de uma rede é alcançada sempre que o valor de perda e o de erro balanceado estabilizam, *i.e.*, que deixam de oscilar significativamente ao fim de algumas épocas. A título de exemplo, pode-se consultar a Figura 5.2, que ilustra a convergência da Alexnet para uma das experiências realizadas.

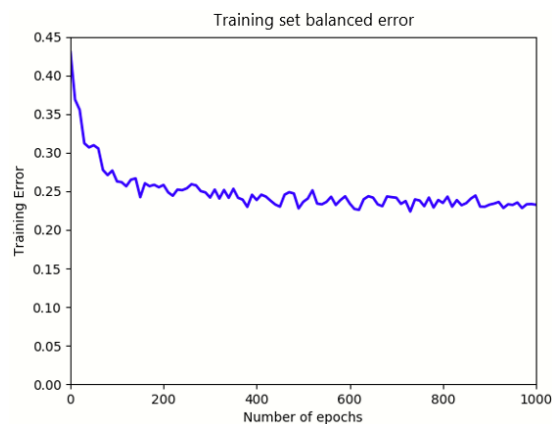


Figura 5.2: Convergência do erro balanceado para a rede AlexNet no conjunto de treino aumentado.

Os pesos são inicializados através de transferência de aprendizagem para todas as CNN utilizadas, como mencionado na Secção 4.2.4. Estes pesos advêm dos modelos de CNN previamente treinados com a base de dados ImageNet. A arquitectura da rede AlexNet é retirada de [122] e o ficheiro dos pesos é retirado de [123]. Os pesos da configuração original da VGG de 16 camadas e da ResNet de 50 camadas são extraídos de [124]. A implementação de cada CNN não respeita integralmente as arquitecturas originais de cada uma das três CNN, uma vez que foram modificadas (ver Secção 4.2.3).

As três redes foram treinadas em modo *mini-batch* com tamanho de *batches* de 46 imagens.

## 5.4 Escolha de hiper-parâmetros

Nos testes realizados foi avaliado o efeito de dois hiper-parâmetros (*dropout* e ritmo de aprendizagem) para cada CNN. Para cada uma das redes, são ilustrados os resultados de classificação através de tabelas e gráficos e discutido o efeito de cada hiperparâmetro medido no conjunto de validação. Adicionalmente, é também analisada a técnica de aumento de dados (DA) somente numa das CNN, a AlexNet.

São testados quatro valores diferentes de probabilidade de *dropout*: 0% (*i.e.*, sem efeito de *dropout*), 30%, 50% e 70%. Relativamente ao ritmo de aprendizagem, são testados dois valores: um valor fixo de  $1 \times 10^{-4}$  e um valor variável de 100 em 100 épocas.

A selecção do melhor conjunto de hiper-parâmetros é realizada com base no desempenho de cada CNN no conjunto de validação, calculado de dez em dez épocas pelas métricas de perda e erro balanceado (ver Secção 4). Assim, os melhores valores dos hiper-parâmetros para cada uma das medidas de desempenho são escolhidos com base no melhor conjunto de valores calculados pelas métricas de avaliação SE, SP e BACC (explicadas na Secção 5.2). De seguida, comparam-se estes três valores da medida de perda com os da medida de erro balanceado e escolhe-se aquela que apresentar melhores resultados. A partir da melhor medida, retira-se a melhor época de treino da rede, que traduz então a melhor configuração possível da rede (*i.e.*, o melhor conjunto de hiper-parâmetros). A capacidade de generalização da melhor configuração é posteriormente testada, recorrendo ao conjunto de teste, e discutida na Secção 5.5.

### 5.4.1 AlexNet

As Tabelas 5.3 e 5.4 mostram os resultados de classificação da AlexNet sem o efeito de DA no conjunto de treino para os diferentes valores de *dropout* e ritmo de aprendizagem. As Tabelas 5.5 e 5.6 mostram os resultados de classificação da AlexNet com o efeito de DA no conjunto de treino para os diferentes valores de *dropout* e ritmo de aprendizagem.

Tabela 5.3: Avaliação da AlexNet no conjunto de validação, sem o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem.

ALEXNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =20	época <sub>erro</sub> =230	época <sub>perda</sub> =50	época <sub>erro</sub> =440	época <sub>perda</sub> =20	época <sub>erro</sub> =80	época <sub>perda</sub> =30	época <sub>erro</sub> =70
Sensibilidade	0,08	0,19	0,13	<b>0,23</b>	0,17	0,20	0,13	0,20
Especificidade	0,98	0,95	0,97	<b>0,93</b>	0,95	0,97	0,97	0,97
BACC	0,53	0,57	0,55	<b>0,58</b>	0,56	0,58	0,55	0,58

Tabela 5.4: Avaliação da AlexNet no conjunto de validação, sem o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem.

ALEXNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =90	época <sub>erro</sub> =170	época <sub>perda</sub> =50	época <sub>erro</sub> =260	época <sub>perda</sub> =20	época <sub>erro</sub> =110	época <sub>perda</sub> =50	época <sub>erro</sub> =120
Sensibilidade	0,13	0,17	0,13	<b>0,20</b>	0,10	0,20	0,17	0,20
Especificidade	0,98	0,97	0,98	<b>0,98</b>	0,98	0,98	0,97	0,98
BACC	0,55	0,57	0,55	<b>0,59</b>	0,54	0,59	0,57	0,59

Tabela 5.5: Avaliação da AlexNet no conjunto de validação, sob o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem.

ALEXNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =140	época <sub>erro</sub> =220	época <sub>perda</sub> =100	época <sub>erro</sub> =60	época <sub>perda</sub> =120	época <sub>erro</sub> =30	época <sub>perda</sub> =110	época <sub>erro</sub> =530
Sensibilidade	0,53	<b>0,70</b>	0,43	0,63	0,40	0,60	0,27	0,47
Especificidade	0,75	<b>0,73</b>	0,83	0,76	0,83	0,72	0,86	0,83
BACC	0,64	<b>0,71</b>	0,63	0,70	0,62	0,66	0,56	0,65

O gráfico da Figura 5.3 mostra as curvas de convergência para a melhor configuração obtida para esta CNN: sob efeito de DA, com *dropout* a 30% e ritmo de aprendizagem variável. O gráfico da Figura 5.4 ilustra a melhor configuração de rede, mas sem a influência do aumento da técnica de dados. O gráfico da Figura

Tabela 5.6: Avaliação da AlexNet no conjunto de validação, sob o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem.

ALEXNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =870	época <sub>erro</sub> =520	época <sub>perda</sub> =950	época <sub>erro</sub> =210	época <sub>perda</sub> =990	época <sub>erro</sub> =300	época <sub>perda</sub> =990	época <sub>erro</sub> =130
Sensibilidade	0,57	0,60	0,43	0,67	0,53	<b>0,70</b>	0,53	0,70
Especificidade	0,82	0,81	0,79	0,73	0,78	<b>0,75</b>	0,78	0,69
BACC	0,69	0,70	0,61	0,70	0,66	<b>0,73</b>	0,65	0,70

5.5 mostra a melhor configuração da rede sem o efeito de *dropout*. Por último, o gráfico da Figura 5.6 ilustra a melhor configuração, mas sem a variação do ritmo de aprendizagem.

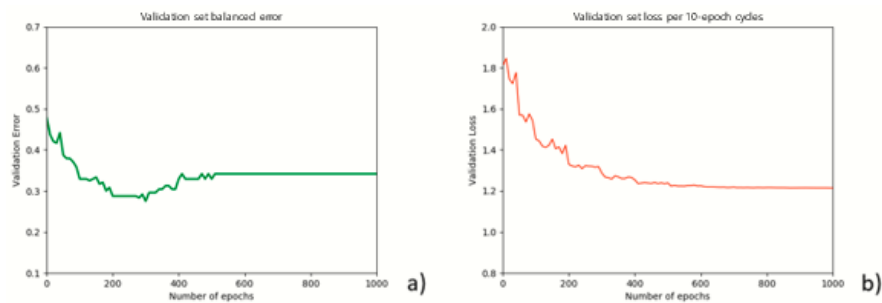


Figura 5.3: A melhor configuração obtida da AlexNet no conjunto de validação com aumento de dados, *dropout* e variação do ritmo de aprendizagem, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

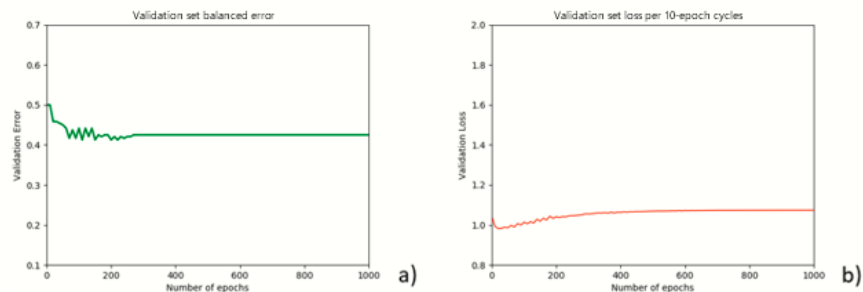


Figura 5.4: A melhor configuração obtida da AlexNet no conjunto de validação, mas sem o efeito de DA, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

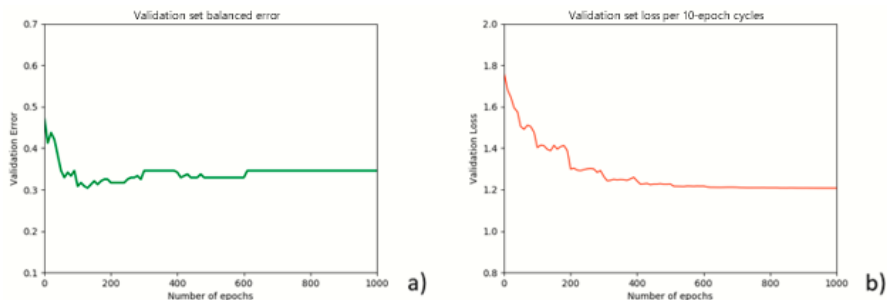


Figura 5.5: A melhor configuração obtida da AlexNet no conjunto de validação, mas sem *dropout*, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

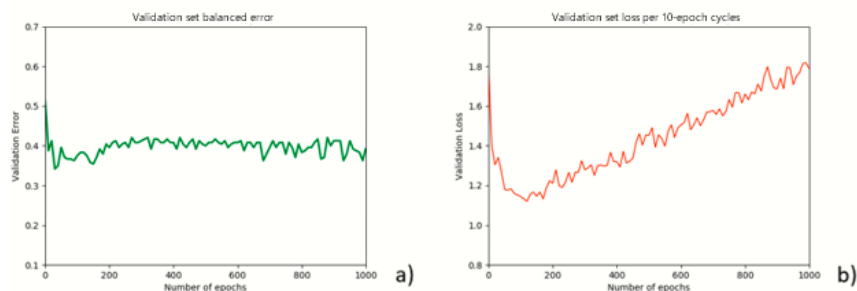


Figura 5.6: A melhor configuração obtida da AlexNet no conjunto de validação, mas sem a variação do ritmo de aprendizagem, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

#### 5.4.1.1 Técnica de Aumento de Dados

Ao analisar as Tabelas 5.3, 5.4, 5.5 e 5.6, é possível observar que o efeito do aumento de dados melhora substancialmente o desempenho da CNN, independentemente da métrica de selecção, pois consegue-se distinguir com muito maior precisão melanomas de não-melanomas. As experiências sem DA não permitem afirmar a capacidade de a rede diferenciar entre melanomas e não-melanomas, dado que a SE no máximo é 23%. Ao observarem-se os gráficos das Figuras 5.3 e 5.4, é possível constatar que a rede converge (a partir da época 600) com a técnica de DA. Através da curva do erro com DA também se pode inferir que é possível evitar o efeito de sobreaprendizagem.

Dados os bons resultados da AlexNet com DA, os testes realizados para as restantes CNN foram somente feitos sob o efeito de DA.

#### 5.4.1.2 Escolha dos outros hiper-parâmetros

##### Ritmo de Aprendizagem

Por observação das Tabelas 5.5 e 5.6 os resultados são melhores para um valor variável do ritmo de aprendizagem (excepto na experiência com *dropout* a 70%). Pela análise dos gráficos das Figuras 5.3 e 5.6, observa-se que com a variação do ritmo de aprendizagem a rede converge e o efeito de sobreaprendizagem e a frequência de oscilações entre as épocas diminuem significativamente, tornando mais precisa a classificação.

##### **Dropout**

Pela análise das Tabelas 5.5 e 5.6, observa-se uma melhoria significativa nos valores de SE à medida que o valor da probabilidade de *dropout* aumenta, quando o ritmo de aprendizagem é fixo. Sob a variação do ritmo de aprendizagem, a introdução do *dropout* não revela tanto impacto nos resultados. Por observação dos gráficos das Figuras 5.3 e 5.5, a presença de *dropout* acentua a redução do efeito de sobreaprendizagem e acelera a convergência da rede.

#### 5.4.2 VGG

As Tabelas 5.7 e 5.8 mostram os resultados de classificação da VGG no conjunto de validação, para os diferentes valores testados de *dropout* e ritmo de aprendizagem e sob o efeito de DA no conjunto de treino.

A melhor configuração da rede VGG ( $SE = 87\%$ ,  $SP = 73\%$ ,  $BACC = 80\%$ ) é apresentada no gráfico da

Tabela 5.7: Avaliação da VGG no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem.

VGG	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =170	época <sub>erro</sub> =160	época <sub>perda</sub> =380	época <sub>erro</sub> =420	época <sub>perda</sub> =290	época <sub>erro</sub> =150	época <sub>perda</sub> =110	época <sub>erro</sub> =250
Sensibilidade	0,53	0,83	0,53	0,73	0,63	0,70	0,53	0,80
Especificidade	0,82	0,69	0,84	0,87	0,85	0,84	0,90	0,72
BACC	0,68	0,76	0,69	0,80	0,74	0,77	0,72	0,76

Tabela 5.8: Avaliação da VGG no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem.

VGG	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =599	época <sub>erro</sub> =200	época <sub>perda</sub> =599	época <sub>erro</sub> =280	época <sub>perda</sub> =510	época <sub>erro</sub> =60	época <sub>perda</sub> =170	época <sub>erro</sub> =599
Sensibilidade	0,63	<b>0,87</b>	0,63	<b>0,77</b>	0,57	0,77	0,77	0,63
Especificidade	0,78	<b>0,73</b>	0,82	<b>0,79</b>	0,77	0,71	0,72	0,78
BACC	0,70	<b>0,80</b>	0,73	<b>0,78</b>	0,67	0,74	0,74	0,71

Figura 5.7: *dropout* a 70% e ritmo de aprendizagem variável. O gráfico da Figura 5.8 ilustra a segunda melhor configuração desta CNN ( $SE = 77\%$ ,  $SP = 79\%$ ,  $BACC = 78\%$ ): *dropout* a 50% e ritmo de aprendizagem variável.

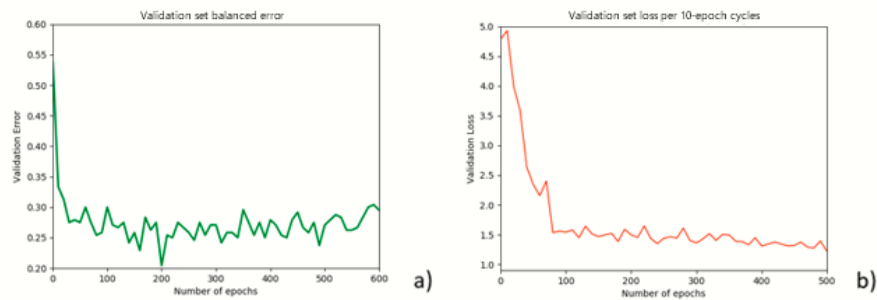


Figura 5.7: A melhor configuração obtida da VGG no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

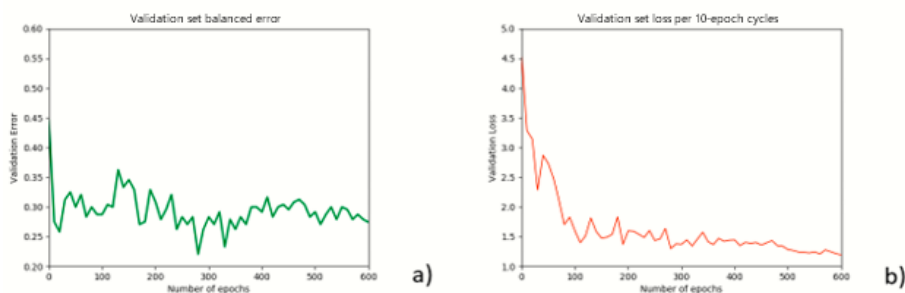


Figura 5.8: A segunda melhor configuração obtida da VGG no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

### Ritmo de aprendizagem

Pela análise das Tabelas 5.7 e 5.8, a variação do ritmo de aprendizagem apresenta maioritariamente melhores resultados de classificação do que um valor fixo do ritmo de aprendizagem. Observando os gráficos das Figuras 5.7 e 5.8, compreende-se que a introdução da variação do ritmo de aprendizagem reduz consideravelmente a amplitude das oscilações entre épocas, acentua a redução do efeito de sobreaprendizagem e apresenta

valores mais baixos para as curvas de erro balanceado e perda, quando se compara com as experiências feitas com um ritmo de aprendizagem fixo (que por falta de espaço não se ilustram).

### Dropout

Por observação das Tabelas 5.7 e 5.8, verifica-se que quanto maior o valor de *dropout*, melhor a classificação da rede no conjunto de validação, principalmente para as experiências em que se varia o ritmo de aprendizagem. Através dos gráficos das Figuras 5.8 e 5.7, um maior valor de *dropout* justifica a convergência mais rápida da rede e reduz o número de oscilações, principalmente analisando a curva de perda.

### 5.4.3 ResNet

As Tabelas 5.9 e 5.10 mostram os resultados de classificação da ResNet sob o efeito de DA para os diferentes valores testados de *dropout* e ritmo de aprendizagem.

Tabela 5.9: Avaliação da ResNet no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor fixo do ritmo de aprendizagem.

RESNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =100	época <sub>erro</sub> =410	época <sub>perda</sub> =150	época <sub>erro</sub> =320	época <sub>perda</sub> =320	época <sub>erro</sub> =70	época <sub>perda</sub> =140	época <sub>erro</sub> =40
Sensibilidade	0,80	0,80	0,73	0,80	0,70	0,80	0,70	0,80
Especificidade	0,72	0,73	0,74	0,75	0,76	0,73	0,78	0,75
BACC	0,76	0,77	0,74	0,78	0,73	0,76	0,74	0,78

Tabela 5.10: Avaliação da ResNet no conjunto de validação com o efeito de DA, aplicado no conjunto de treino, e para um valor variável do ritmo de aprendizagem.

RESNET	dropout=70%		dropout=50%		dropout=30%		dropout=0%	
	época <sub>perda</sub> =320	época <sub>erro</sub> =180	época <sub>perda</sub> =499	época <sub>erro</sub> =280	época <sub>perda</sub> =220	época <sub>erro</sub> =200	época <sub>perda</sub> =420	época <sub>erro</sub> =90
Sensibilidade	0,70	0,77	0,73	<b>0,80</b>	0,73	<b>0,80</b>	0,63	0,87
Especificidade	0,78	0,78	0,77	<b>0,78</b>	0,78	<b>0,77</b>	0,80	0,69
BACC	0,74	0,78	0,75	<b>0,79</b>	0,75	<b>0,78</b>	0,72	0,78

O gráfico da Figura 5.9 ilustra a melhor configuração da ResNet ( $SE = 80\%$ ,  $SP = 78\%$ ,  $BACC = 79\%$ ): *dropout* 50% e ritmo de aprendizagem variável. O gráfico da Figura 5.10 mostra a segunda melhor configuração: *dropout* a 30% e ritmo de aprendizagem variável.

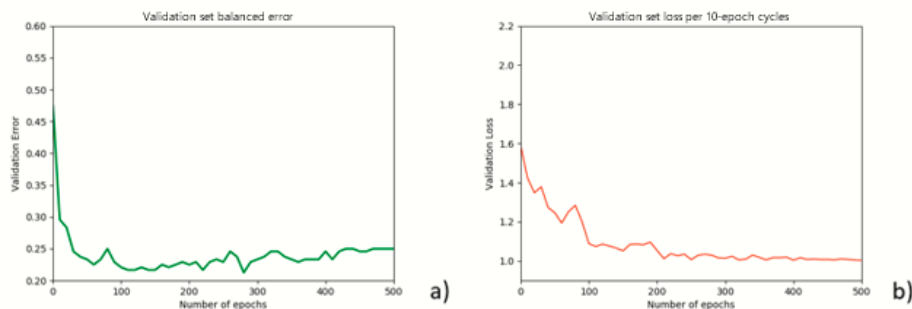


Figura 5.9: A melhor configuração obtida da ResNet no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

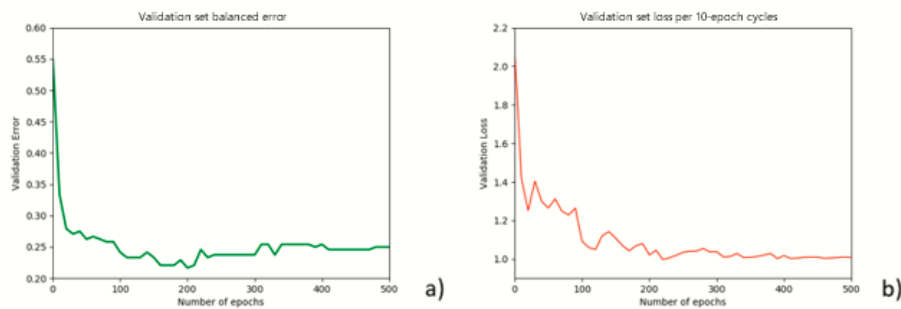


Figura 5.10: A segunda melhor configuração obtida da ResNet no conjunto de validação, dada pela métrica de erro balanceado, à esquerda em a), e pela métrica de perda, à direita em b).

### 5.4.3.1 Escolha de hiper-parâmetros

#### Ritmo de aprendizagem

As Tabelas 5.9 e 5.10 ilustram que a variação do ritmo de aprendizagem melhora os resultados de classificação face a um valor fixo deste hiperparâmetro, embora as diferenças não sejam muito significativas. Através dos gráficos das Figuras 5.9 e 5.10 não é possível discutir o impacto da variação deste hiperparâmetro face a um valor fixo, dado que ambas as configurações variam o ritmo de aprendizagem. Porém, dadas as experiências realizadas com a ResNet neste trabalho para um ritmo fixo e um ritmo variável de aprendizagem (que por falta de espaço não se ilustram as respectivas curvas de convergência), a introdução da variação do hiperparâmetro acentua a redução do efeito de sobreaprendizagem e acelera a convergência da rede.

#### **Dropout**

Através das Tabelas é possível verificar que, embora sem *dropout* a BACC seja superior face a alguns valores de BACC com *dropout*, a rede ainda não convergiu nas épocas 40 e 90 (que são as melhores épocas para as experiências sem *dropout*). Deste modo, a presença de *dropout* promove a convergência da rede. O melhor valor de probabilidade é 50%, visto que apresenta os melhores resultados (SE a 80%, SP entre 75% a 78%, BACC entre 78% e 79%). A partir dos gráficos das Figuras 5.9 e 5.10 as diferenças não são muito significativas. Porém, com *dropout* a 50% a frequência de oscilações é relativamente menor e a rede converge ligeiramente mais rápido.

## 5.5 Avaliação do Desempenho: Comparação das três CNN

Após a selecção da melhor configuração de cada rede, cada CNN é testada, recorrendo-se ao conjunto de teste. De seguida, as capacidades de generalização das três CNN são comparadas entre si, com o intuito de compreender qual das CNN é a mais indicada para a classificação de melanomas e de lesões benignas (mais concretamente, nevos benignos e queratoses seborreicas).

As Tabelas 5.11, 5.12 e 5.13 apresentam os resultados de classificação no conjunto de teste para a AlexNet, VGG e ResNet, respectivamente.

Os gráficos das Figuras 5.3, 5.7 e 5.9 ilustram a curva ROC no conjunto de teste de cada uma das CNN, respectivamente.

Através das Tabelas 5.11, 5.12 e 5.13, a ResNet é a CNN que apresenta o melhor resultado de classificação



ALEXNET	dropout=30%, ritmo de aprendizagem variável	
	época <sub>perda</sub> =990	época <sub>erro</sub> =300
Sensibilidade	0,54	0,56
Especificidade	0,73	0,70
BACC	0,64	0,63
ROC-AUC	0,69	0,68

Tabela 5.11: Avaliação da AlexNet no conjunto de teste com a melhor configuração da rede obtida na fase de treino.

VGG	dropout=70%, ritmo de aprendizagem variável	
	época <sub>perda</sub> =599	época <sub>erro</sub> =200
Sensibilidade	0,73	0,78
Especificidade	0,65	0,59
BACC	0,69	0,68
ROC-AUC	0,76	0,75

Tabela 5.12: Avaliação da VGG no conjunto de teste com a melhor configuração da rede obtida na fase de treino.

RESNET	dropout=50%, ritmo de aprendizagem variável	
	época <sub>perda</sub> =499	época <sub>erro</sub> =280
Sensibilidade	0,79	0,79
Especificidade	0,61	0,60
BACC	0,70	0,70
ROC-AUC	0,77	0,77

Tabela 5.13: Avaliação da ResNet no conjunto de teste com a melhor configuração da rede obtida na fase de treino.

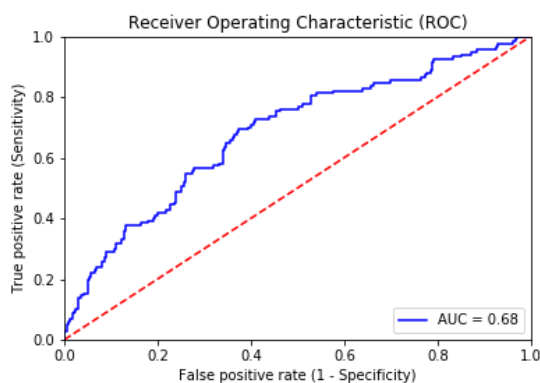


Figura 5.11: Curva ROC da melhor configuração obtida da AlexNet no conjunto de teste,  $ROC - AUC = 68\%$ .

com  $BACC = 70\%$  e  $ROC - AUC = 77\%$ . Esta CNN identificou correctamente 93 lesões como malignas e as restantes 24 classificou-as erradamente como benignas. A VGG é a segunda melhor na tarefa de classificação, com apenas menos 2% de BACC e ROC-AUC face à ResNet. Tal como esperado, a AlexNet é a pior das três CNN a classificar, sendo a sua capacidade de distinguir melanomas de não-melanomas inferior a 60%, alcançado 63% de BACC. Classificou correctamente 66 lesões malignas e as restantes 51 lesões malignas foram falsos negativos, *i.e.*, classificadas erradamente como benignas. Pela análise das curvas ROC, a ResNet é o melhor classificador binário, pois é a CNN que apresenta a melhor relação da sensibilidade ( $SE$ ) com a

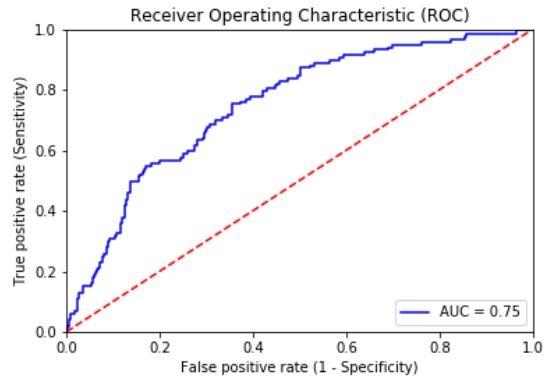


Figura 5.12: Curva ROC da melhor configuração obtida da VGG no conjunto de teste,  $ROC - AUC = 75\%$ .

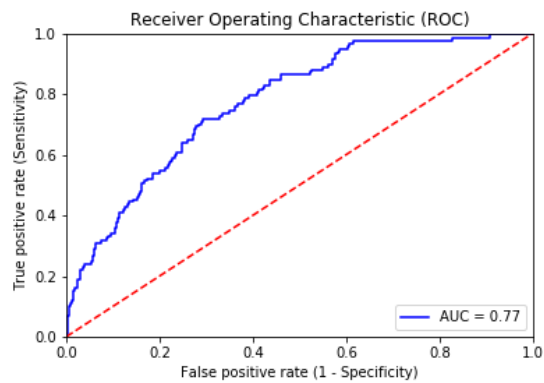


Figura 5.13: Curva ROC da melhor configuração obtida da ResNet no conjunto de teste,  $ROC - AUC = 77\%$ .

probabilidade de falso alarme ( $1 - SP$ ) e o valor mais alto da ROC-AUC.

As Figuras 5.14 e 5.15 ilustram seis imagens de lesões malignas e seis de lesões benignas, após aplicação da respectiva caixa delimitadora rectangular: à esquerda correctamente classificadas pela ResNet e, à direita, incorrectamente classificadas pela ResNet.

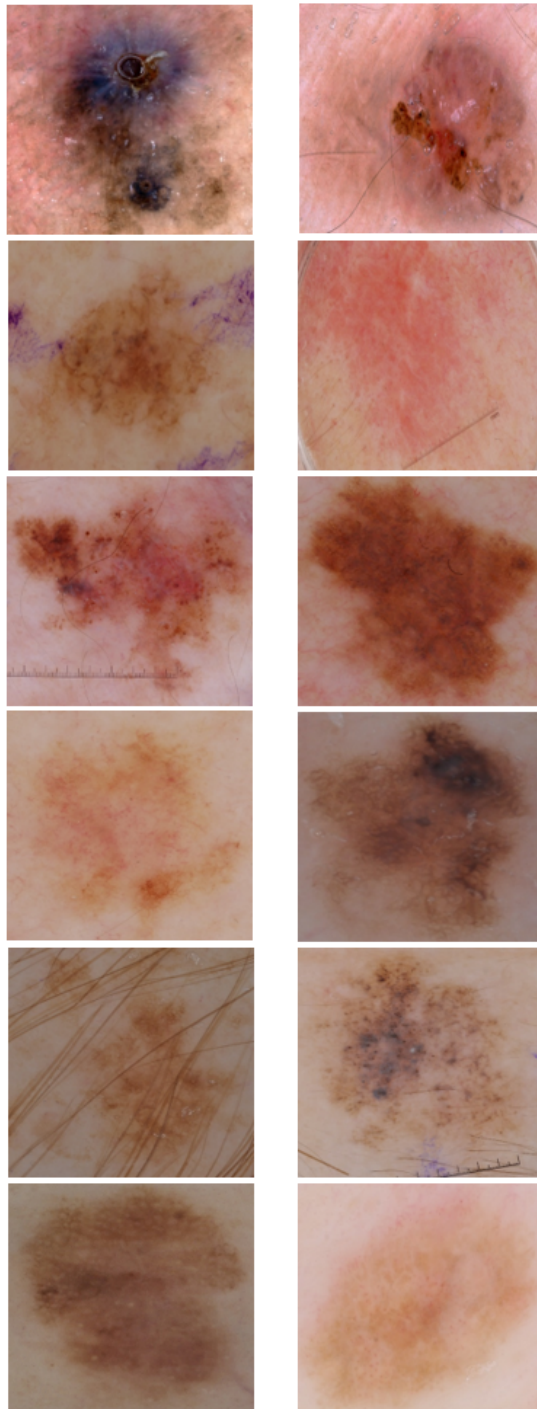


Figura 5.14: Exemplos de melanomas: correctamente classificados (à esquerda), incorrectamente classificados (à direita).

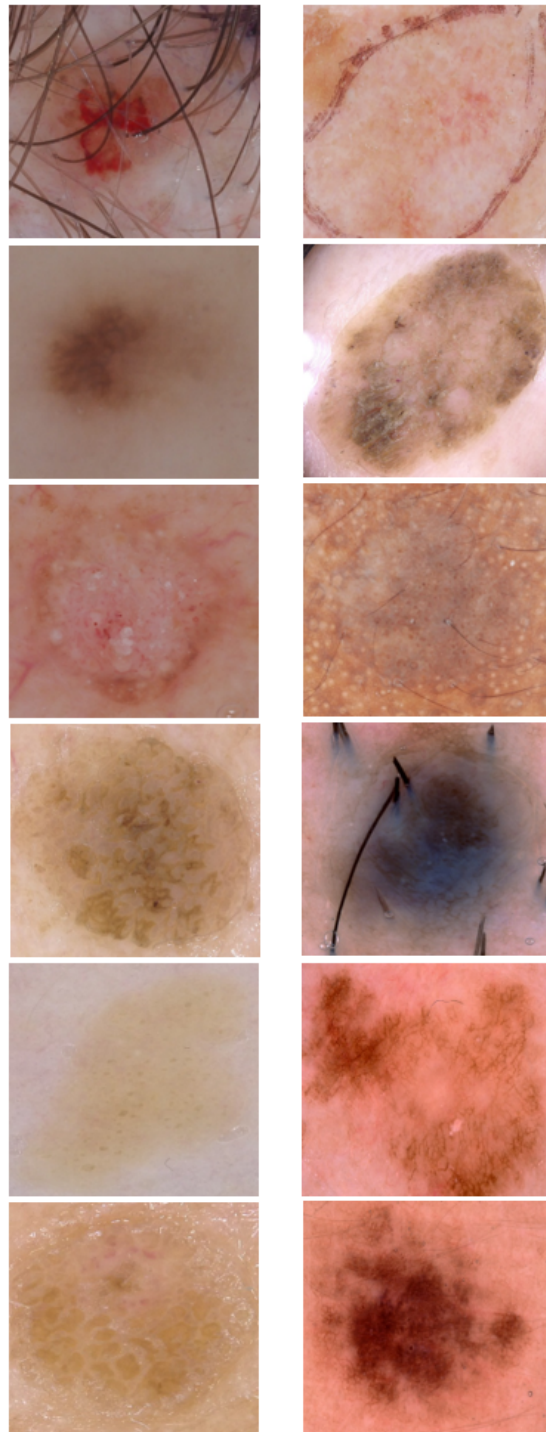


Figura 5.15: Exemplos de não-melanomas: correctamente classificados (à esquerda), incorrectamente classificados (à direita).

# 6

## Conclusão e Trabalho Futuro

Este Capítulo visa sintetizar os principais resultados alcançados nesta tese, bem como propor algumas sugestões para investigação futura.

O principal objectivo desta tese foi desenvolver um sistema de detecção automática de melanomas baseado em redes neuronais convolucionais. O sistema desenvolvido consistiu em duas fases, uma de treino e uma de classificação. Para cada uma destas fases foram avaliadas três redes neuronais convolucionais alternativas: AlexNet, VGG e ResNet. Estas CNNs foram escolhidas por serem relativamente pouco exigentes em matéria de custo computacional, já que o equipamento disponível no início para realizar esta investigação não dispunha de um GPU.

Este trabalho procurou comparar o desempenho das três redes na classificação de melanomas. Os resultados de classificação de cada rede foram obtidos através de quatro métricas de avaliação de desempenho: sensibilidade, especificidade, precisão balanceada e área por baixo da curva de característica de operação do receptor.

A CNN que teve melhor desempenho a classificar lesões da pele em benignas ou malignas foi a ResNet. No conjunto de teste, esta rede alcançou 70% de BACC e 77% de ROC-AUC, tendo classificado correctamente 79% melanomas no conjunto total de melanomas e 60% não-melanomas no conjunto total de não-melanomas. Esta CNN é a mais complexa das três redes estudadas, mas é a que converge num menor número de épocas. A CNN que exigiu mais memória computacional durante a fase de treino e, conseqüentemente, mais tempo a correr foi a VGG. Esta última obteve resultados de classificação muito próximos dos da ResNet, BACC a 68% e ROC-AUC a 75%. A CNN mais simples em termos de arquitectura e, por conseguinte, aquela que menor memória e tempo computacionais exigiu foi a AlexNet. Porém, foi a rede que exigiu um maior número de épocas durante a fase de treino.

Apesar dos resultados de classificação entre a ResNet e a VGG serem semelhantes e a VGG ter tido um melhor desempenho no conjunto de validação ( $SE = 87\%$  contra a  $SE = 80\%$  alcançada pela ResNet), a primeira CNN é a mais indicada na detecção automática de melanomas. Embora seja mais complexa, demora consideravelmente menos tempo a treinar do que a VGG devido aos blocos residuais. Adicionalmente, a

frequência e a amplitude das oscilações das curvas de convergência de erro balanceado e de perda são menores e mais suaves na ResNet do que na VGG.

As métricas de selecção aplicadas no conjunto de validação para escolher a melhor configuração de cada uma das CNNs foram o erro balanceado e a perda. Embora a comparação aprofundada destas duas métricas não tenha sido objecto de estudo neste trabalho, de uma forma geral, a métrica do erro balanceado apresentou resultados mais consistentes do que a métrica da perda. Os melhores resultados obtidos pela métrica de perda ocorreram na maioria das vezes nas últimas épocas de treino (ver Tabelas da secção 5.5), o que pode significar que os resultados estejam mascarados devido à sobreaprendizagem nas últimas épocas de treino.

Para além da limitação do equipamento computacional, teve-se um conjunto total de imagens dermatológicas limitado e desequilibrado na relação entre lesões benignas e malignas. Por estes motivos, recorreu-se à técnica de aumento de dados para aumentar artificialmente o número total de lesões malignas.

Os resultados apresentados nesta tese são satisfatórios e demonstrem, de facto, que é possível classificar automaticamente lesões da pele em benignas ou malignas. Porém, podem ainda ser melhorados e possivelmente mais promissores. Deste modo, apresentam-se de seguida algumas sugestões para trabalho futuro.

O primeiro ponto de melhoria seria implementar a solução proposta recorrendo a um ou mais GPUs<sup>1</sup>, de maneira a acelerar o tempo de execução da fase de treino (devido à computação dos parâmetros do modelo). Com o emprego de GPUs, seria igualmente interessante estudar outras redes neuronais convolucionais. Por exemplo, utilizar CNNs com arquitecturas mais complexas que tenham mostrado excelentes resultados em problemas de análise de imagem ou recorrer a modificações de CNNs utilizadas em estudos concretos de detecção automática de melanomas [32, 33].

Outro ponto de melhoria seria investigar outras formas de potenciar os resultados obtidos para os hiper-parâmetros usados neste trabalho ou estudar em detalhe a selecção de outros hiper-parâmetros das CNNs (por exemplo, a função de activação, o tamanho do *batch* e a inicialização dos pesos). A escolha de outros hiper-parâmetros pode potenciar a sensibilidade obtida pela CNN e, dessa forma, melhorar o desempenho da rede na classificação de melanomas.

A terceira sugestão seria utilizar três classes, em vez de duas, e avaliar o impacto nos resultados de classificação, dado que o conjunto de imagens continha ainda um número relevante de queratoses seborreicas.

Como quarta sugestão, a utilização de uma base de dados com um número muito maior de imagens dermatológicas e, de preferência, equilibrado.

Por fim, seria interessante analisar aprofundadamente o tema das métricas de selecção. Começar por investigar com detalhe qual das duas métricas, erro balanceado e perda, permite obter melhores resultados de classificação e discutir a possibilidade de haver outras métricas de selecção.

---

<sup>1</sup>No início da execução da tese, não havia nenhum GPU disponível. Quando surgiu disponibilidade, optou-se por não utilizar.

# Bibliografia

- [1] M. D. McCarter, "Melanoma," in *Abernathy's Surgical Secrets*. Elsevier, 2018, pp. 311–318.
- [2] A. J. C. on Cancer. (2009) Melanoma of the skin staging. [Online]. Available: <https://cancerstaging.org/references-tools/quickreferences/Documents/MelanomaSmall.pdf>
- [3] M. Trakatelli, E. de Vries, G. Katsanos, D. Tzachanis, and A. Eggermont, "Cutaneous malignant melanoma," in *Managing Skin Cancer*. Springer Berlin Heidelberg, 2010, pp. 79–110.
- [4] C. Massone, A. D. Stefani, and H. P. Soyer, "Dermoscopy for skin cancer detection," *Current Opinion in Oncology*, vol. 17, no. 2, pp. 147–153, março 2005.
- [5] R. P. Braun, H. S. Rabinovitz, M. Oliviero, A. W. Kopf, and J.-H. Saurat, "Dermoscopy of pigmented skin lesions," *Journal of the American Academy of Dermatology*, vol. 52, no. 1, pp. 109–121, janeiro 2005.
- [6] Dermoscopy tutorial: Abcd rule. DS ME S.r.l. [Online]. Available: <http://www.dermoscopy.org/consensus/2b.asp>
- [7] Dermoscopy tutorial: 7-point checklist. DS ME S.r.l. [Online]. Available: <http://www.dermoscopy.org/consensus/2d.asp>
- [8] H. Iyatomi, K. Norton, M. E. Celebi, G. Schaefer, M. Tanaka, and K. Ogawa, "Classification of melanocytic skin lesions from non-melanocytic lesions," in *2010 Annual International Conference of the IEEE Engineering in Medicine and Biology*. IEEE, agosto 2010, pp. 5407–5410.
- [9] G. Pellacani, C. Longo, J. Malvehy, S. Puig, C. Carrera, S. Segura, S. Bassoli, and S. Seidenari, "In vivo confocal microscopic and histopathologic correlations of dermoscopic features in 202 melanocytic lesions," *Archives of Dermatology*, vol. 144, no. 12, dezembro 2008.
- [10] I. A. Basheer and M. Hajmeer, "Artificial neural networks: fundamentals, computing, design, and application," *Journal of Microbiological Methods*, vol. 43, no. 1, pp. 3–31, dezembro 2000.
- [11] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, maio 2015.
- [12] A. Karpathy. (2018) Stanford cs class: Cs231n convolutional neural networks for visual recognition. [Online]. Available: <http://cs231n.github.io/convolutional-networks/>

- [13] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in Neural Information Processing Systems 25*, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2012, pp. 1097–1105, iLSVRC 2012 winner.
- [14] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, vol. abs/1409.1556, 2014, iLSVRC 2014.
- [15] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *arXiv preprint arXiv:1512.03385*, vol. abs/1512.03385, 2015, iLSVRC 2015 winner.
- [16] N. C. F. Codella, Q.-B. Nguyen, S. Pankanti, D. A. Gutman, B. Helba, A. C. Halpern, and J. R. Smith, "Deep learning ensembles for melanoma recognition in dermoscopy images," *IBM Journal of Research and Development*, vol. 61, no. 4/5, pp. 5:1–5:15, julho 2017.
- [17] I. A. for Research on Cancer. (2018) Incidence, mortality and prevalence by cancer site (world). [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/840-united-states-of-america-fact-sheets.pdf>
- [18] K. Korotkov and R. Garcia, "Computerized analysis of pigmented skin lesions: A review," *Artificial Intelligence in Medicine*, vol. 56, no. 2, pp. 69–90, outubro 2012.
- [19] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *arXiv preprint arXiv:1703.00577*, vol. abs/1703.00577, 2017.
- [20] S. Pathan, K. G. Prabhu, and P. Siddalingaswamy, "Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review," *Biomedical Signal Processing and Control*, vol. 39, pp. 237–262, janeiro 2018.
- [21] A. Forsea, V. del Marmol, E. de Vries, E. Bailey, and A. Geller, "Melanoma incidence and mortality in europe: new estimates, persistent disparities," *British Journal of Dermatology*, vol. 167, no. 5, pp. 1124–1130, setembro 2012.
- [22] T. G. C. O. International Agency for Research on Cancer. (2018) Incidence, mortality and prevalence by cancer site (portugal). [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/620-portugal-fact-sheets.pdf>
- [23] I. A. for Research on Cancer. (2018) Incidence, mortality and prevalence by cancer site (us). [Online]. Available: <https://gco.iarc.fr/today/data/factsheets/populations/840-united-states-of-america-fact-sheets.pdf>
- [24] A. Victor and M. Ghalib, "Automatic detection and classification of skin cancer," *International Journal of Intelligent Engineering and Systems*, vol. 10, no. 3, pp. 444–451, junho 2017.
- [25] A. C. Society, *Cancer Facts & Figures 2018*, 2018. [Online]. Available: <https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-2018.pdf>



- [26] R. B. Oliveira, M. E. Filho, Z. Ma, J. P. Papa, A. S. Pereira, and J. M. R. Tavares, "Computational methods for the image segmentation of pigmented skin lesions: A review," *Computer Methods and Programs in Biomedicine*, vol. 131, pp. 127–141, julho 2016.
- [27] M. E. Vestergaard, P. Macaskill, P. E. Holt, and S. W. Menzies, "Dermoscopy compared with naked eye examination for the diagnosis of primary melanoma: a meta-analysis of studies performed in a clinical setting," *British Journal of Dermatology*, pp. 669–676, junho 2008.
- [28] M. E. Celebi, H. A. Kingravi, B. Uddin, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, and R. H. Moss, "A methodological approach to the classification of dermoscopy images," *Computerized Medical Imaging and Graphics*, vol. 31, no. 6, pp. 362–373, setembro 2007.
- [29] J. F. Alcon, C. Ciuhu, W. ten Kate, A. Heinrich, N. Uzunbajakava, G. Krekels, D. Siem, and G. de Haan, "Automatic imaging system with decision support for inspection of pigmented skin lesions and melanoma diagnosis," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 14–25, fevereiro 2009.
- [30] M. E. Celebi, T. Mendonca, and J. S. Marques, *Dermoscopy Image Analysis*. Taylor & Francis, 2015. [Online]. Available: [https://www.ebook.de/de/product/25065260/dermoscopy\\_image\\_analysis.html](https://www.ebook.de/de/product/25065260/dermoscopy_image_analysis.html)
- [31] J. Jaworek-Korjakowska and P. Kłeczek, "Automatic classification of specific melanocytic lesions using artificial intelligence," *BioMed Research International*, pp. 1–17, 2016.
- [32] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau, and S. Thrun, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, janeiro 2017.
- [33] H. A. Haenssle, C. Fink, R. Schneiderbauer, F. Toberer, T. Buhl, A. Blum, A. Kalloo, A. B. H. Hassen, L. Thomas, A. Enk, and L. Uhlmann, "Man against machine: diagnostic performance of a deep learning convolutional neural network for dermoscopic melanoma recognition in comparison to 58 dermatologists," *Annals of Oncology*, vol. 29, pp. 1836–1842, maio 2018.
- [34] I. Maglogiannis and C. N. Doukas, "Overview of advanced computer vision systems for skin lesions characterization," *IEEE Transactions on Information Technology in Biomedicine*, vol. 13, no. 5, pp. 721–733, setembro 2009.
- [35] S. Dreiseitl, L. Ohno-Machado, H. Kittler, S. Vinterbo, H. Billhardt, and M. Binder, "A comparison of machine learning methods for the diagnosis of pigmented skin lesions," *Journal of Biomedical Informatics*, vol. 34, no. 1, pp. 28–36, fevereiro 2001.
- [36] A. Masood and A. A. Al-Jumaily, "Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms," *International Journal of Biomedical Imaging*, pp. 1–22, 2013.
- [37] H. C. Burger, C. J. Schuler, and S. Harmeling, "Image denoising: Can plain neural networks compete with BM3d?" in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, junho 2012, pp. 2392–2399.

- [38] D. Cireşan, U. Meier, and J. Schmidhuber, "Multi-column deep neural networks for image classification," in *2012 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, junho 2012, pp. 3642–3649.
- [39] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and L. Fei-Fei, "Large-scale video classification with convolutional neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, junho 2014, pp. 1725–1732.
- [40] I. E. Naqa, R. Li, and M. J. Murphy, Eds., *Machine Learning in Radiation Oncology: Theory and Applications*. Springer International Publishing, 2015.
- [41] W. S. McCulloch and W. Pitts, "A logical calculus of the ideas immanent in nervous activity," *The Bulletin of Mathematical Biophysics*, vol. 5, no. 4, pp. 115–133, dezembro 1943.
- [42] D. E. Rumelhart, B. Widrow, and M. A. Lehr, "The basic ideas in neural networks," *Communications of the ACM*, vol. 37, no. 3, pp. 87–92, março 1994.
- [43] V. Bettadapura, "Face expression recognition and analysis: The state of the art," pp. 1,16–21, 2012.
- [44] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "Imagenet large scale visual recognition challenge," *arXiv preprint arXiv:1409.0575*, vol. abs/1409.0575, 2014.
- [45] A. Chouiekh and E. H. I. E. Haj, "ConvNets for fraud detection analysis," *Procedia Computer Science*, vol. 127, pp. 133–138, 2018.
- [46] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, J. Klingner, A. Shah, M. Johnson, X. Liu, Łukasz Kaiser, S. Gouws, Y. Kato, T. Kudo, H. Kazawa, K. Stevens, G. Kurian, N. Patil, W. Wang, C. Young, J. Smith, J. Riesa, A. Rudnick, O. Vinyals, G. Corrado, M. Hughes, and J. Dean, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, vol. abs/1609.08144, 2016.
- [47] D. Yu and L. Deng, *Automatic Speech Recognition*. Springer London, 2015.
- [48] L. Fridman, D. E. Brown, M. Glazer, W. Angell, S. Dodd, B. Jenik, J. Terwilliger, J. Kindelsberger, L. Ding, S. Seaman, H. Abraham, A. Mehler, A. Sipperley, A. Pettinato, B. Seppelt, L. Angell, B. Mehler, and B. Reimer, "Mit autonomous vehicle technology study: Large-scale deep learning based analysis of driver behavior and interaction with automation," vol. abs/1711.06976, 2017.
- [49] D. Erhan, C. Szegedy, A. Toshev, and D. Anguelov, "Scalable object detection using deep neural networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, junho 2014, pp. 2155–2162.
- [50] K. Hazelwood, S. Bird, D. Brooks, S. Chintala, U. Diril, D. Dzhulgakov, M. Fawzy, B. Jia, Y. Jia, A. Kalro, J. Law, K. Lee, J. Lu, P. Noordhuis, M. Smelyanskiy, L. Xiong, and X. Wang, "Applied

- machine learning at facebook: A datacenter infrastructure perspective," in *2018 IEEE International Symposium on High Performance Computer Architecture (HPCA)*. IEEE, fevereiro 2018, pp. 620–629.
- [51] G. D. Leo, A. Paolillo, P. Sommella, and G. Fabbrocini, "Automatic diagnosis of melanoma: A software system based on the 7-point check-list," in *2010 43rd Hawaii International Conference on System Sciences*. IEEE, janeiro 2010, pp. 1–10.
- [52] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. S. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, pp. 35–45, fevereiro 2009.
- [53] M. E. Celebi, H. A. Kingravi, H. Iyatomi, J. Lee, Y. A. Aslandogan, W. V. Stoecker, R. Moss, J. M. Malters, and A. A. Marghoob, "Fast and accurate border detection in dermoscopy images using statistical region merging," in *Medical Imaging 2007: Image Processing*, J. P. W. Pluim and J. M. Reinhardt, Eds., vol. 6512. SPIE, março 2007.
- [54] R. Sumithra, M. Suhil, and D. Guru, "Segmentation and classification of skin lesions for disease diagnosis," *Procedia Computer Science*, vol. 45, pp. 76–85, 2015.
- [55] F. Nachbar, W. Stolz, T. Merkle, A. B. Cagnetta, T. Vogt, M. Landthaler, P. Bilek, O. Braun-Falco, and G. Plewig, "The ABCD rule of dermatoscopy," *Journal of the American Academy of Dermatology*, vol. 30, no. 4, pp. 551–559, abril 1994.
- [56] C. Barata, M. E. Celebi, and J. S. Marques, "Melanoma detection algorithm based on feature fusion," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, agosto 2015, pp. 2653–2656.
- [57] C. Barata, M. E. Celebi, and J. Marques, "Development of a clinically oriented system for melanoma diagnosis," *Pattern Recognition*, vol. 69, pp. 270–285, setembro 2017.
- [58] T. Mendonca, P. M. Ferreira, J. S. Marques, A. R. S. Marcal, and J. Rozeira, "PH<sup>2</sup> - a dermoscopic image database for research and benchmarking," in *2013 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, julho 2013, pp. 5437–5440.
- [59] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," *arXiv preprint arXiv:1605.01397*, vol. abs/1605.01397, 2016.
- [60] E. Nasr-Esfahani, S. Samavi, N. Karimi, S. M. R. Soroushmehr, M. H. Jafari, K. Ward, and K. Najarian, "Melanoma detection by analysis of clinical images using convolutional neural network," in *2016 38th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, agosto 2016, pp. 1373–1376.
- [61] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A

- challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," *arXiv preprint arXiv:1710.05006*, vol. abs/1710.05006, 2017.
- [62] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciampi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42, pp. 60–88, dezembro 2017.
- [63] D. Trichopoulos, F. P. Li, and D. J. Hunter, "What causes cancer?" *Scientific American*, vol. 275, no. 3, pp. 80–87, setembro 1996.
- [64] O. Abuzaghleh, B. D. Barkana, and M. Faezipour, "Noninvasive real-time automated skin lesion analysis system for melanoma early detection and prevention," *IEEE Journal of Translational Engineering in Health and Medicine*, vol. 3, pp. 1–12, 2015.
- [65] Dermoscopy tutorial. DS ME S.r.l. [Online]. Available: <http://www.dermoscopy.org>
- [66] G. Argenziano, S. Puig, I. Zalaudek, F. Sera, R. Corona, M. Alsina, F. Barbato, C. Carrera, G. Ferrara, A. Guilbert, D. Massi, J. A. Moreno-Romero, C. Muñoz-Santos, G. Petrillo, S. Segura, H. P. Soyer, R. Zanchini, and J. Malvehy, "Dermoscopy improves accuracy of primary care physicians to triage lesions suggestive of skin cancer," *Journal of Clinical Oncology*, vol. 24, no. 12, pp. 1877–1882, abril 2006.
- [67] M.-L. Bafounta, A. Beauchet, P. Aegerter, and P. Saiag, "Is dermoscopy (epiluminescence microscopy) useful for the diagnosis of melanoma?" *Archives of Dermatology*, vol. 137, no. 10, outubro 2001.
- [68] G. Argenziano, G. Fabbrocini, P. Carli, V. D. Giorgi, E. Sammarco, and M. Delfino, "Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions," *Archives of Dermatology*, vol. 134, no. 12, pp. 1563–1570, dezembro 1998.
- [69] C. Barata, M. Ruela, M. Francisco, T. Mendonca, and J. S. Marques, "Two systems for the detection of melanomas in dermoscopy images using texture and color features," *IEEE Systems Journal*, vol. 8, no. 3, pp. 965–979, setembro 2014.
- [70] K. H. Kraemer, "The role of sunlight and DNA repair in melanoma and nonmelanoma skin cancer," *Archives of Dermatology*, vol. 130, no. 8, p. 1018, agosto 1994.
- [71] C. D. J. Holman, B. K. Armstrong, P. J. Heenan, J. B. Blackwell, F. J. Cumming, D. R. English, S. Holland, G. R. H. Kelsall, L. R. Matz, I. L. Rouse, A. Singh, R. E. J. T. Seldam, J. D. Watt, and Z. Xu, *Epidemiology of Malignant Melanoma*, R. P. Gallagher, Ed. Springer Berlin Heidelberg, 1986.
- [72] S. N. Markovic, L. A. Erickson, R. D. Rao, R. R. McWilliams, L. A. Kottschade, E. T. Creagan, R. H. Weenig, J. L. Hand, M. R. Pittelkow, B. A. Pockaj, A. Bardia, C. M. Vachon, S. E. Schild, S. D. Laman, W. J. Maples, J. S. Pulido, and J. D. Cameron, "Malignant melanoma in the 21st century, part 1: Epidemiology, risk factors, screening, prevention, and diagnosis," *Mayo Clinic Proceedings*, vol. 82, no. 3, pp. 364–380, março 2007.
- [73] U. Leiter and C. Garbe, "Epidemiology of melanoma and nonmelanoma skin cancer—the role of sunlight," in *Sunlight, Vitamin D and Skin Cancer*. Springer New York, 2008, pp. 89–103.

- [74] N. B. J., *Rook's textbook of dermatology*, B. T., B. S., C. N., and G. C., Eds. Oxford: Wiley-Blackwell, 2010.
- [75] L. T. Thao and N. H. Quang, "Automatic skin lesion analysis towards melanoma detection," in *2017 21st Asia Pacific Symposium on Intelligent and Evolutionary Systems (IES)*, novembro 2017, pp. 106–111.
- [76] G. Argenziano and H. P. Soyer, "Dermoscopy of pigmented skin lesions – a valuable tool for early," *The Lancet Oncology*, vol. 2, no. 7, pp. 443–449, julho 2001.
- [77] I. Maglogiannis, E. Zafiroopoulos, and C. Kyranoudis, "Intelligent segmentation and classification of pigmented skin lesions in dermatological images," in *Advances in Artificial Intelligence*. Springer Berlin Heidelberg, 2006, pp. 214–223.
- [78] A. F. Barata, E. Celebi, and J. Marques, "Improving dermoscopy image classification using color constancy," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2014.
- [79] P. Schmid-Saugeona, J. Guillodb, and J.-P. Thirana, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Computerized Medical Imaging and Graphics*, vol. 27, no. 1, pp. 65–78, janeiro 2003.
- [80] K. Møllersen, H. M. Kirchesch, T. G. Schopf, and F. Godtlielsen, "Unsupervised segmentation for digital dermoscopic images," *Skin Research and Technology*, vol. 16, no. 4, pp. 401–407, setembro 2010.
- [81] D. Nie, "Classification of melanoma and clark nevus skin lesions based on medical image processing techniques," in *2011 3rd International Conference on Computer Research and Development*, vol. 3. IEEE, março 2011, pp. 31–34.
- [82] D. D. Gomez, C. Butakoff, B. K. Ersboll, and W. Stoecker, "Independent histogram pursuit for segmentation of skin lesions," *IEEE Transactions on Biomedical Engineering*, vol. 55, no. 1, pp. 157–161, janeiro 2008.
- [83] A. C. F. Barata, E. M. Celebi, and J. Marques, "A survey of feature extraction in dermoscopy image analysis of skin cancer," *IEEE Journal of Biomedical and Health Informatics*, pp. 1–1, 2018.
- [84] R. Battiti, "Using mutual information for selecting features in supervised neural net learning," *IEEE Transactions on Neural Networks*, vol. 5, no. 4, pp. 537–550, julho 1994.
- [85] I. Kononenko and E. Simec, "Induction of decision trees using relief," in *Proceedings of the ISSEK94 Workshop on Mathematical and Statistical Methods in Artificial Intelligence*. Springer Vienna, 1995, vol. 363, pp. 199–220.
- [86] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, 1999.
- [87] V. J. Mar and H. P. Soyer, "Artificial intelligence for melanoma diagnosis: how can we deliver on the promise?" *Annals of Oncology*, vol. 29, pp. 1625–1628, maio 2018.
- [88] F. Rosenblatt, "The perceptron: A probabilistic model for information storage and organization in the brain," *Psychological Review*, vol. 65, no. 6, pp. 386–408, 1958.

- [89] C. V. D. Malsburg, "Frank rosenblatt: Principles of neurodynamics: Perceptrons and the theory of brain mechanisms," in *Brain Theory*. Springer Berlin Heidelberg, 1986, pp. 245–248.
- [90] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," in *Parallel distributed processing: Explorations in the microstructure of cognition*, pdp research group ed. MIT Press, 1986, vol. 1.
- [91] R. Hecht-Nielsen, "Counterpropagation networks," *Applied Optics*, vol. 26, no. 23, p. 4979, dezembro 1987.
- [92] M. W. Gardner and S. Dorling, "Artificial neural networks (the multilayer perceptron)—a review of applications in the atmospheric sciences," *Atmospheric Environment*, vol. 32, no. 14-15, pp. 2627–2636, agosto 1998.
- [93] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*. MIT Press, 2016. [Online]. Available: <http://www.deeplearningbook.org>
- [94] G. Hinton, L. Deng, D. Yu, G. Dahl, A. rahman Mohamed, N. Jaitly, A. Senior, V. Vanhoucke, P. Nguyen, B. Kingsbury, and T. Sainath, "Deep neural networks for acoustic modeling in speech recognition," *IEEE Signal Processing Magazine*, vol. 29, pp. 82–97, agosto 2012.
- [95] Y. Tang, "Deep learning using linear support vector machines," *arXiv preprint arXiv:1306.0239*, vol. abs/1306.0239, 2013.
- [96] N. S. Keskar, D. Mudigere, J. Nocedal, M. Smelyanskiy, and P. T. P. Tang, "On large-batch training for deep learning: Generalization gap and sharp minima," *arXiv preprint arXiv:1609.04836*, vol. abs/1609.04836, 2016.
- [97] D. Svozil, V. Kvasnicka, and J. Pospichal, "Introduction to multi-layer feed-forward neural networks," *Chemometrics and Intelligent Laboratory Systems*, vol. 39, no. 1, pp. 43–62, novembro 1997.
- [98] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *Journal of Machine Learning Research*, vol. 15, pp. 1929–1958, 2014.
- [99] D. E. Rumelhart, R. Durbin, R. Golden, and Y. Chauvin, *Developments in connectionist theory. Back-propagation: Theory, architectures, and applications*, Y. Chauvin and D. E. Rumelhart, Eds. Lawrence Erlbaum Associates, 1995.
- [100] G. Thimm and E. Fiesler, "High-order and multilayer perceptron initialization," *IEEE Transactions on Neural Networks*, vol. 8, no. 2, pp. 349–359, março 1997.
- [101] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?" in *Advances in Neural Information Processing Systems 27*, Z. Ghahramani, M. Welling, C. Cortes, N. D. Lawrence, and K. Q. Weinberger, Eds. Curran Associates, Inc., 2014, pp. 3320–3328.

- [102] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, ser. Proceedings of Machine Learning Research, G. Gordon, D. Dunson, and M. Dudík, Eds., vol. 15. Fort Lauderdale, FL, USA: PMLR, abril 2011, pp. 315–323.
- [103] S. Scardapane, D. Comminiello, A. Hussain, and A. Uncini, "Group sparse regularization for deep neural networks," *Neurocomputing*, vol. 241, pp. 81–89, junho 2017.
- [104] S. Ruder, "An overview of gradient descent optimization algorithms," *arXiv preprint arXiv:1609.04747*, vol. abs/1609.04747, 2016.
- [105] G. Hinton, N. Srivastava, and K. Swersky. (2012) Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. [Online]. Available: [https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture\\_slides\\_lec6.pdf](https://www.cs.toronto.edu/~tijmen/csc321/slides/lecture_slides_lec6.pdf)
- [106] L. Bottou, "Large-scale machine learning with stochastic gradient descent," in *Proceedings of COMPSTAT/2010*. Physica-Verlag HD, 2010, pp. 177–186.
- [107] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *3rd International Conference for Learning Representations, San Diego*, vol. abs/1412.6980, 2015.
- [108] M. D. Zeiler and R. Fergus, "Stochastic pooling for regularization of deep convolutional neural networks," in *Proceedings of the International Conference on Learning Representation (ICLR)*, 2013.
- [109] A. Krogh and J. A. Hertz, "A simple weight decay can improve generalization," in *Advances in Neural Information Processing Systems*, vol. 4. Morgan Kaufmann, 1992, pp. 950–957.
- [110] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998.
- [111] W.-Y. Lee, S.-M. Park, and K.-B. Sim, "Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm," *Optik*, vol. 172, pp. 359–367, novembro 2018.
- [112] P. Pujari, M. Sewak, and M. R. Karim, *Practical Convolutional Neural Network Models*. PACKT PUB, 2018. [Online]. Available: [https://www.ebook.de/de/product/31928702/pradeep\\_pujari\\_mohit\\_sewak\\_md\\_rezaul\\_karim\\_practical\\_convolutional\\_neural\\_network\\_models.html](https://www.ebook.de/de/product/31928702/pradeep_pujari_mohit_sewak_md_rezaul_karim_practical_convolutional_neural_network_models.html)
- [113] Y.-L. Boureau, J. Ponce, and Y. Lecun, "A theoretical analysis of feature pooling in visual recognition," in *Proceedings of the 27th International Conference on International Conference on Machine Learning*. Omnipress, 2010, pp. 111–118.
- [114] D. Scherer, A. Müller, and S. Behnke, "Evaluation of pooling operations in convolutional architectures for object recognition," in *Artificial Neural Networks – ICANN 2010*. Springer Berlin Heidelberg, 2010, pp. 92–101.

- [115] D. Yu, H. Wang, P. Chen, and Z. Wei, "Mixed pooling for convolutional neural networks," in *Rough Sets and Knowledge Technology*. Springer International Publishing, 2014, vol. 8818, pp. 364–375.
- [116] W.-Y. Lee, S.-M. Park, and K.-B. Sim, "Optimal hyperparameter tuning of convolutional neural networks based on the parameter-setting-free harmony search algorithm," *Optik*, vol. 172, pp. 359–367, novembro 2018.
- [117] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, vol. abs/1502.03167, 2015.
- [118] S. C. Wong, A. Gatt, V. Stamatescu, and M. D. McDonnell, "Understanding data augmentation for classification: When to warp?" in *2016 International Conference on Digital Image Computing: Techniques and Applications (DICTA)*, novembro 2016, pp. 1–6.
- [119] K. H. Brodersen, C. S. Ong, K. E. Stephan, and J. M. Buhmann, "The balanced accuracy and its posterior distribution," in *2010 20th International Conference on Pattern Recognition*. IEEE, agosto 2010, pp. 3121–3124.
- [120] C. E. Metz, "Basic principles of ROC analysis," *Seminars in Nuclear Medicine*, vol. 8, no. 4, pp. 283–298, outubro 1978.
- [121] M. Abadi, P. Barham, J. Chen, Z. Chen, A. Davis, J. Dean, M. Devin, S. Ghemawat, G. Irving, M. Isard, M. Kudlur, J. Levenberg, R. Monga, S. Moore, D. G. Murray, B. Steiner, P. Tucker, V. Vasudevan, P. Warden, M. Wicke, Y. Yu, and X. Zheng, "Tensorflow: A system for large-scale machine learning," in *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 2016, pp. 265–283.
- [122] F. Kratzert. (2017) Finetuning alexnet with tensorflow. [Online]. Available: <https://kratzert.github.io/2017/02/24/finetuning-alexnet-with-tensorflow.html>
- [123] Alexnet implementation + weights in tensorflow. [Online]. Available: [http://www.cs.toronto.edu/~guerzhoy/tf\\_alexnet](http://www.cs.toronto.edu/~guerzhoy/tf_alexnet)
- [124] N. Silberman and S. Guadarrama. Tensorflow-slim image classification model library - pre-trained models. [Online]. Available: <https://github.com/tensorflow/models/tree/master/research/slim#Pretrained>