

Instituto Politécnico de Lisboa

Escola Superior Tecnologias da Saúde de Lisboa

**"Detecção precoce da Doença de Parkinson com
recurso a imagens de Ressonância Magnética e
algoritmos de classificação Deep Learning"**

Autora: Mónica Alexandra Agostinho dos Santos Capinha

Orientadora: Prof^a Doutora Ana Luísa da Costa Ferreira Vieira

Jurí:

Prof^o Doutor Luís Manuel Carvalho Freire

Prof^o Doutor Nuno Miguel de Pinto Lobo e Matela

Mestrado em Radiações Aplicadas às Tecnologias da Saúde – Ressonância
Magnética

Lisboa, 2022

Instituto Politécnico de Lisboa

Escola Superior Tecnologias da Saúde de Lisboa

**"Deteção precoce da Doença de Parkinson com
recurso a imagens de Ressonância Magnética e
algoritmos de classificação Deep Learning"**

Autora: Mónica Alexandra Agostinho dos Santos Capinha

Orientadora: Prof^a Doutora Ana Luísa da Costa Ferreira Vieira

Jurí:

Prof^o Doutor Luís Manuel Carvalho Freire

Prof^o Doutor Nuno Miguel de Pinto Lobo e Matela

Mestrado em Radiações Aplicadas às Tecnologias da Saúde – Ressonância
Magnética

Lisboa, 2022

Agradecimentos

Ao longo destes dois anos que passaram no decurso do Mestrado em Radiações Aplicadas às Tecnologias da Saúde variante de Ressonância Magnética não poderia de deixar de agradecer a todos os professores que proporcionaram os ensinamentos que adquiri e que levaram ao desenvolvimento desta dissertação. A vossa compreensão de, ao longo das aulas, aceitarem mais uma aluna – a minha filha, que também gostava de aumentar os seus conhecimentos :) – assim como os meus colegas, tornaram este percurso cheio de altos e baixos mais ligeiros.

Gostaria também de agradecer especialmente à minha orientadora, a Professora Doutora Ana Luísa Vieira, que ao longo deste tempo se demonstrou sempre disponível em ajudar e a superar os momentos mais difíceis, por vezes desesperantes. Sem a sua orientação não teria sido possível chegar ao fim.

Aos meus colegas de trabalho e amigos que transformaram o meu desânimo em força nesta nova etapa da minha vida.

À minha família, a minha mana que me deu abrigo nos dias em que as aulas iam até tarde proporcionando um descanso mental, aos meus pais, mas em especial a minha mãe que, mesmo não estando presente, está sempre no meu coração acompanhando-me e dando-me força.

Aos meus filhos, Rodrigo e Margarida, por serem uns filhos excepcionais, e apesar das suas idades serem compreensíveis e tentarem dar sempre uma ajuda imprescindível – um abraço e um sorriso.

Ao meu marido César, um especial Obrigado, por me ter apoiado e encorajado a avançar para esta etapa, tornando-a possível, aguentando todo o meu desespero com todo o seu carinho e amizade – Obrigada!

A doença de Parkinson consiste num distúrbio neurológico que afeta cerca de 1% da população mundial. Esta doença tem origem na degeneração da substância nigra, uma porção do cérebro localizada no mesencéfalo. Sendo que os neurónios nesta região do cérebro estão revestidos por um pigmento rico em neuromelanina, com elevado teor de ferro, é possível assim a visualização da substância nigra em imagens de Ressonância Magnética.

A deteção de doenças com recurso a inteligência artificial tem sido uma área de estudo com elevada relevância científica e aplicada nos últimos anos. Nesse contexto, este trabalho tem como objetivo identificar a combinação de arquitetura de *Deep Learning* e ponderação de ressonância magnética utilizadas nos protocolos *standard* que permite alcançar melhor capacidade preditiva na deteção da doença de Parkinson. Para isso foi inicialmente efetuada uma revisão da literatura, a que se seguiu uma avaliação da capacidade preditiva de quatro arquiteturas *Deep Learning* distintas (LeNet; AlexNet; InceptionNetV3; ResNet50) considerando quatro ponderações de imagem (T1, T2, DP e FLAIR). A arquitetura ResNet50 com a ponderação T2 foi a que apresentou melhores *performance* preditiva, alcançando resultado considerados 'muito bons' (AUC=0,889). Da análise das relações entre valores de erro e idade e género dos utentes, verificou-se ainda a existência de uma relação significativa positiva entre idade e valores de erro para pessoas do género feminino, sugerindo uma maior dificuldade do modelo em efetuar o diagnóstico correto para mulheres de idade mais avançada.

Palavras-chave: *Deep Learning*, doença de Parkinson, processo neurodegenerativo, Ressonância Magnética

Abstract

Parkinson's disease is a neurological disorder that affects about 1% of the world's population. This disease originates in the degeneration of the substantia nigra, a portion of the brain located in the midbrain. Since neurons in this region of the brain are coated by a pigment rich in neuromelanin, with a high iron content, it is possible to visualize the substantia nigra on MRI images.

The detection of diseases using artificial intelligence has been an area of study with high scientific and applied relevance in recent years. In this context, this work aims to identify the combination of Deep Learning architecture and MRI weighting used in standard protocols that allows achieving better predictive ability in the detection of Parkinson's disease. To this end, a literature review was initially performed, followed by an evaluation of the predictive ability of four different Deep Learning architectures (LeNet; AlexNet; InceptionNetV3; ResNet50) considering four image weightings (T1, T2, DP and FLAIR). The ResNet50 architecture with T2 weighting was the one that presented the best predictive performance, achieving results considered as 'very good' (AUC=0.889). From the analysis of the relationships between error values and age and gender of patients, there was also a significant positive relationship between age and error values for females, suggesting greater difficulty of the model in making the correct diagnosis for older women.

Keywords: Deep Learning, MRI, neurodegeneration process, Parkinson Disease

Agradecimentos	i
Resumo	ii
Abstract	iii
Índice de Figuras	vi
Índice de Tabelas	viii
1. Introdução	1
1.1. Enquadramento	1
1.2. Estrutura da tese	3
2. Conceitos teóricos	5
2.1. Doença de Parkinson: história e conhecimento atual	5
2.1.1. Substância nigra	6
2.2. Princípios básicos da Ressonância Magnética	8
2.3. Neuroimagem em doença de Parkinson	19
2.3.1. Neuromelanina e ressonância magnética	20
2.3.2. Nigrossoma-1 e ressonância magnética	21
2.4. <i>Deep Learning</i>	22
2.4.1. Arquiteturas das CNN	25
2.4.1.1. LeNet	26
2.4.1.2. AlexNet	27
2.4.1.3. Inception	27
2.4.1.4. ResNet	28
3. Objetivos	31
4. Metodologia	33
4.1. Estado da Arte da aplicação de <i>Deep Learning</i> em Ressonância Magnética	33
4.2. Detecção doença de Parkinson com <i>Deep Learning</i>	34
4.2.1. Recolha dos dados	34

4.2.2. Treino e avaliação dos modelos	37
4.2.3. Avaliação do efeito do género e idade dos utentes na capacidade preditiva dos modelos	39
5. Estado da Arte da aplicação de <i>Deep Learning</i> em Ressonância Magnética	41
5.1. Resultados	41
5.2. Análise e Discussão	47
6. Detecção doença de Parkinson com <i>Deep Learning</i>	49
6.1. Resultados	49
6.1.1. Avaliação dos modelos	49
6.1.2. Efeito do género e idade dos utentes na capacidade preditiva dos modelos	52
6.2. Análise e Discussão	55
7. Conclusão	59
Bibliografia	61

Índice de Figuras

Figura 2.1: Corte transversal do mesencéfalo representando a localização da substância nigra.	6
Figura 2.2: Representação de um átomo de hidrogénio.	8
Figura 2.3: Representação dos spins quando expostos ao campo gravitacional da terra (A) e quando aplicado um campo magnético externo (B).	9
Figura 2.4: Demonstração da magnetização longitudinal.	11
Figura 2.5: Demonstração da magnetização transversal.	11
Figura 2.6: Relaxação T1.	12
Figura 2.7: Relaxação T2.	12
Figura 2.8: Magnetização longitudinal dos tecidos.	13
Figura 2.9: Magnetização transversal dos tecidos.	14
Figura 2.10: Sequência <i>Spin Echo</i> .	16
Figura 2.11: Sequência gradiente eco.	17
Figura 2.12: Variação Campo magnético no corpo humano.	18
Figura 2.13: Imagens RM da Substância Nigra	20
Figura 2.14: Imagem RM T2 SWI	21
Figura 2.15: Rede Neuronal Convolutacional, CNN	22
Figura 2.16: Desempenho do <i>Deep Learning</i> vrs redes tradicionais.	23
Figura 2.17: Arquitetura LeNet	26
Figura 2.18: Arquitetura AlexNet	27
Figura 2.19: Arquitetura InceptionV3	28
Figura 2.20: Arquitetura ResNet	29
Figura 4.1: Plataforma <i>Image and Data Archive</i> , recolha de dados	35
Figura 4.2: Biblioteca Virtual IDA	35
Figura 4.3: Seleção corte transversal para a base de dados.	36
Figura 5.1: Histograma do número de artigos publicados por ano.	41
Figura 5.2: Distribuição do número de trabalhos publicados por patologia.	42
Figura 5.3: Número de trabalhos por áreas anatómicas investigadas.	43
Figura 5.4: Desempenho do diagnóstico efetuado por redes neuronais profundas versus por humanos, por patologia considerada.	44
Figura 5.5: Número de estudos com aplicação de <i>Deep Learning</i> no estudo de patologias do foro neurológico.	45
Figura 6.1: Boxplot de valores de AUC agrupados por arquitetura Deep Learning	50

Figura 6.2: Boxplot de valores de AUC agrupados por ponderação de imagens.	51
Figura 6.3: totais de predições corretas e incorretas por género.	52
Figura 6.4: Distribuição das predições corretas e incorretas por género e idade.	53

Índice de Tabelas

Tabela 2.1: Tempos de relaxação T1 e T2 de diferentes tecidos - 1.5 Tesla.	15
Tabela 2.2: Diferentes ponderações na imagem em sequências <i>Spin Echo</i>	16
Tabela 5.1: Trabalhos científicos que procedem à aplicação de modelos <i>Deep Learning</i> no diagnóstico da doença de Parkinson com imagens de ressonância magnética.	46
Tabela 6.1: Valor de AUC para cada combinação de arquitetura, ponderação e número de épocas de treino testadas.	50
Tabela 6.2: Parâmetros do modelo linear de efeitos mistos na avaliação de diferenças de propensão para o erro entre gêneros.	53
Tabela 6.3: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade, considerando ambos os gêneros.	54
Tabela 6.4: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade para o gênero masculino.	54
Tabela 6.5: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade para o gênero feminino.	54

1.1 Enquadramento

A doença de Parkinson é uma doença neurodegenerativa comum, afetando cerca de 1% da população mundial. Segundo a Organização Mundial de Saúde (Tysnes et al., 2017) a relevância médica desta doença é crescente, sobretudo em consequência do envelhecimento da população, o que propicia o aumento da sua taxa de incidência ao longo do tempo. Simultaneamente outros fatores que contribuem para este acréscimo incluem contextos ambientais ou sociais crescentemente desfavoráveis, bem como a utilização de novas tecnologias no seu diagnóstico (Shrimanker et al., 2021).

Em Portugal, num estudo recente, verificou-se que a prevalência da doença de Parkinson era de pelo menos 180 casos por 100 mil habitantes, estimando-se desta forma a existência de cerca de 20 mil habitantes com esta doença (Ferreira et al., 2017).

A doença de Parkinson tem início com a morte de neurónios dopaminérgicos, localizado na substância nigra, levando a uma diminuição na produção de dopamina. A manifestação dos primeiros sintomas ocorre por vezes em idades jovens, sintomas estes que vão desde perdas na motricidade fina, a alterações na marcha e na expressão verbal (Doença de Parkinson— Distúrbios neurológicos, Manuais MSD).

De forma a facilitar o diagnóstico das doenças neurodegenerativas, em especial a doença de Parkinson, têm-se realizado diferentes estudos que visam uma melhor caracterização da doença, bem como a identificação de biomarcadores de relevância para o diagnóstico (Wang et al., 2019). A neuroimagem assume um papel de relevo neste âmbito, onde se destacam as técnicas de Tomografia por emissão de Positrões (PET), a Tomografia Computorizada por emissão de fotão único (SPECT), e com crescente relevo nas últimas décadas, a imagem de Ressonância Magnética. De entre as vantagens desta última, destacam-se o seu custo inferior ao das técnicas anteriores, assim como o seu baixo risco para a saúde (Carrete, 2017).

Recentemente assistiu-se também a um crescente interesse na utilização de imagens de Ressonância Magnética no diagnóstico automatizado de doenças neurológicas com recurso a 'Inteligência Artificial'. A vasta maioria destas aplicações assenta em modelos de 'redes neuronais artificiais', modelos matemáticos inspirados na forma como os sistemas nervosos biológicos processam a informação (Bishop, 1995). Nos anos recentes reconheceu-se uma elevada capacidade de algumas 'arquitecturas' de redes neuronais artificiais, nomeadamente as compostas por um

elevado número de camadas de 'neurónios' (ou redes neuronais artificiais profundas; do inglês '*Deep neural networks*'), no reconhecimento de padrões em imagens, incluindo imagens obtidas por dispositivos médicos (LeCun et al., 2015).

A aplicação destes algoritmos computacionais na classificação de imagem médica tem revolucionado a área de diagnóstico imagiológico nos anos recentes, levando a suscitar dúvidas acerca da possibilidade destas técnicas virem eventualmente a substituir o pessoal médico na realização desta tarefa (A Inteligência Artificial irá substituir os radiologistas, 2017). Independentemente desta possibilidade futura, é cada vez mais consensual que estas técnicas devem constar no leque de ferramentas de diagnóstico, atuando em complementaridade com outros meios de diagnóstico já utilizados pelo pessoal médico especializado (Li et al., 2014). A crescente utilização destas técnicas afigura-se como um novo paradigma na forma como a identificação patológica é efetuada na Radiologia, sendo por isso uma área de intensa investigação nos últimos anos (Wang et al., 2019).

Muito do foco das técnicas de diagnóstico automatizado tem incidido sobre imagens de ressonância magnética, incluindo na identificação de doenças neurodegenerativas. Da mesma forma a investigação acerca das próprias doenças também tem progredido a elevado ritmo. Devido a isso, o elevado volume de trabalhos publicados na área e a elevada diversidade de tópicos de aplicação abordados dificultam o reconhecimento acerca das tendências e estado atual desta área de investigação.

1.2 Estrutura da tese

Esta dissertação é composta por 6 capítulos, um inicial onde se abordam os conceitos teóricos da Ressonância Magnética, bem como os conceitos da doença de Parkinson, passando pela sua problemática e neuroanatomia para uma melhor compreensão da sua importância para o diagnóstico da doença. Neste capítulo também é efetuada uma abordagem à temática de *Deep Learning*, onde são descritas algumas arquiteturas mais conhecidas e amplamente utilizadas. Posteriormente inclui-se um segundo capítulo correspondente aos objetivos pretendido para este trabalho. Passando em seguida para um terceiro capítulo onde é abordada a metodologia utilizada, sendo este dividido em 2 subcapítulos. No primeiro referente a uma revisão da literatura, onde se identifica o estado atual de aplicação de redes neuronais no auxílio do diagnóstico da doença de Parkinson, tendo para isso sido realizada uma revisão bibliográfica exaustiva de forma a melhor fundamentar os objetivos deste trabalho. No segundo subcapítulo é descrita a metodologia utilizada, incluindo a recolha dos dados, o treino e avaliação dos modelos, assim como as análises efetuadas para avaliação do efeito do género e idade dos utentes na capacidade preditiva dos modelos. Num quinto capítulo é apresentado os resultados de ambos os trabalhos bem como a sua análise e discussão.

Por fim é incluído um quarto capítulo correspondente à conclusão geral.

2. Conceitos teóricos

2.1. Doença de Parkinson: história e conhecimento atual

A Doença de Parkinson foi descrita pela primeira vez em 1817 por James Parkinson, cirurgião inglês, num estudo intitulado “An Essay on the Shaking Palsy”, (em português: “*Um ensaio sobre a paralisia dos tremores*”), em que descreve a doença como uma *paralisia agitante*. Após 60 anos, Jean-Martin Charcot, neurologista francês, veio dar o nome atual à doença, em homenagem ao seu descobridor (Goedert et al., 2018) (Parent, 2018).

A doença de Parkinson atualmente é a segunda doença neurodegenerativa mais comum a nível mundial, sendo que a primeira é a doença de Alzheimer (Cabreira et al., 2019). Esta doença tem como principais sintomas os tremores, rigidez muscular e bradicinesia, resultando comumente em instabilidade postural e de movimentos, incluindo durante a marcha. No entanto, têm-se também verificado o aparecimento de sintomas não motores sobretudo em pessoas jovens, que cada vez mais se associam à doença de Parkinson. Estes sintomas não motores podem manifestar-se na forma de depressão, ansiedade, alterações no padrão do sono, alterações na função vesical ou fecal, assim como através de alterações de memória. A doença de Parkinson com início precoce está aparentemente associada a antecedentes familiares, resultando de uma alteração genética hereditária que aumenta o seu risco de desenvolvimento. Contudo esta hipótese está ainda em investigação (Barbosa et al., 2005).

À luz do conhecimento atual a doença de Parkinson que ainda não tem uma causa conhecida, sabendo-se, contudo que os sintomas motores da doença têm origem na morte de células na substância nigra, resultando na diminuição de produção de dopamina, um neurotransmissor.

2.1.1 Substância nigra

A substância nigra, parte constituinte dos gânglios da base, está localizada no mesencéfalo (Figura 2.1) (Cassidy et al., 2019). A substância nigra apresenta um papel muito importante na função cerebral, sendo responsável pela produção de dopamina, um neurotransmissor chave no funcionamento da coordenação motora, na movimentação ocular, bem como na regulação da sensação de recompensa, prazer e vício. Muitas destas funções são mediadas eletricamente pelo corpo estriado.

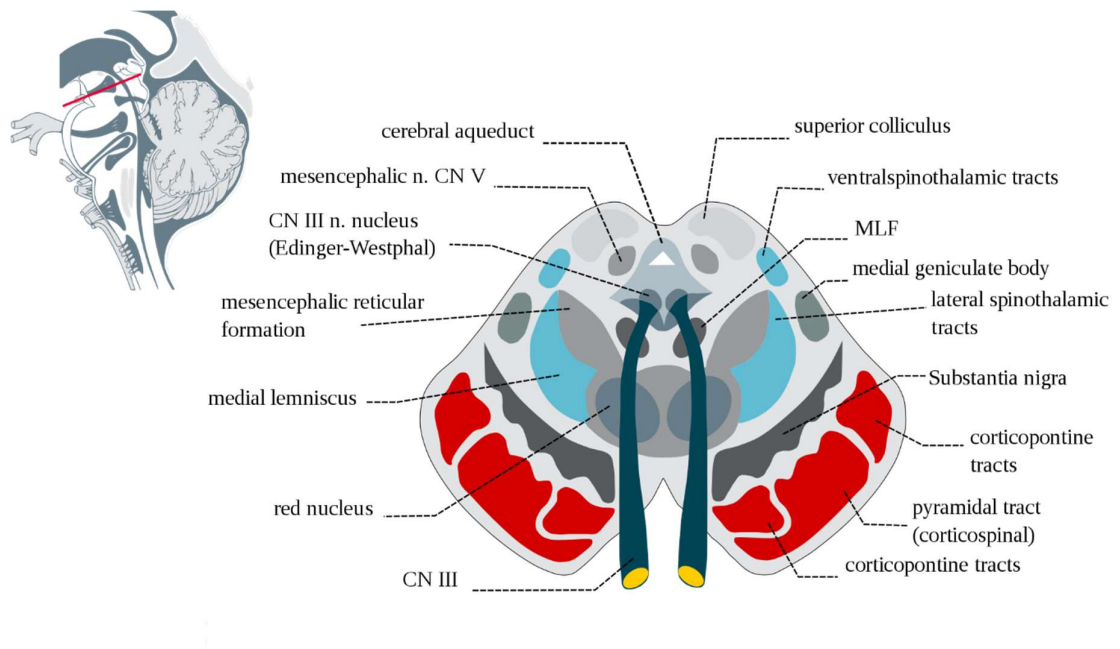


Figura 2.1: Corte transversal do mesencéfalo representando a localização da substância nigra
(*Substantia nigra*. (2021), *Wikipedia*.)

A substância nigra está dividida em *substantia nigra pars compacta* e *pars reticulata*. A substância *nigra pars compacta* apresenta-se como uma estrutura muito compacta do cérebro, sendo constituída por neurónios dopaminérgicos, e que apresentam uma coloração negra devido à presença do pigmento neuromelanina. A substância *nigra pars compacta* é responsável, de modo indireto, pelo controlo motor através da estimulação nervosa do corpo estriado. Em pacientes da doença de Parkinson estas estimulações nervosas encontram-se diminuídas, apresentando um decréscimo da neuromelanina. A porção *pars compacta* também é responsável pela sensação de recompensa, prazer e vício através das vias dopaminérgicas mesocortical (mesencéfalo para o córtex pré-frontal) e da via mesolímbica (núcleos *accumbens*).

Lesões nesta porção da *substantia nigra* estão frequentemente relacionadas com a ocorrência de memória episódica (Sonne et al., 2020).

Por outro lado, a substância *nigra pars reticulata* trata-se de uma estrutura mais difusa, composta por neurónios mais finos que os dopaminérgicos (presentes na porção *pars compacta*). Os neurónios da substância *nigra pars reticulata* são responsáveis pela produção do neuro-inibidor GABA (Kalia & Lang, 2015) e pela transmissão dos sinais dos gânglios da base para o tálamo e para o tectum óptico. A substância *nigra pars reticulata* tem também como função a inibição da função *pars compacta*, por intermédio da ação do neuro-inibidor GABA. No entanto, os mecanismos que controlam esta ação não estão ainda totalmente conhecidos (Sonne et al., 2020).

2.2. Princípios básicos da Ressonância Magnética

A Ressonância magnética está intrinsecamente relacionada com as propriedades magnéticas dos prótons de átomos de hidrogénio que, quando submetidos a um campo magnético externo forte e estimulados por ondas rádio próximas da frequência de Larmor, emitem um sinal que poderá ser captado por uma antena e transformado em imagem. Este princípio básico foi descrito por Purcell e Block, professores de Harvard e Stanford respetivamente, em 1946, através de experiências em sólidos e líquidos, valendo-lhes o prémio Nobel da Física em 1952.

Em 1947 surgiram as primeiras imagens de ressonância magnética do corpo humano. Estas foram realizadas por Paul Lauterbur, dando início a um conjunto significativo de desenvolvimentos, como a codificação em fase, codificação em frequência, a transformada de Fourier, a técnica eco-planar ou a utilização de um campo fechado, permitindo também adquirir imagens em vídeo de 30ms/imagem (Urm, 2017).

O corpo humano é constituído por vários elementos químicos, como hidrogénio, oxigénio, carbono, fósforo, cálcio, fluor, sódio, potássio e o nitrogénio. Todos os átomos destes elementos têm no seu núcleo prótons e neutrões. No entanto, é para o hidrogénio que se dirigem as atenções das técnicas de ressonância magnética, quer pela sua abundância na massa corporal quer pela simplicidade da sua estrutura atómica. Ao apresentar no seu núcleo apenas um próton – partícula com carga positiva – este elemento proporciona um maior momento angular ou *spin*, que corresponde a movimento em torno de si mesmo, dando origem a um momento magnético que o torna sensível à ressonância magnética (Mazzola, A. A., 2015) (Figura 2.2). O núcleo do átomo hidrogénio possui assim propriedades magnéticas, funcionando como um ‘pequeno íman’.

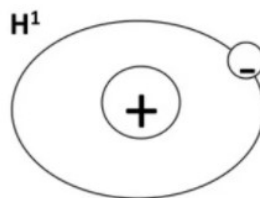


Figura 2.2. Representação de um átomo de hidrogénio.
(Mazzola, A.A., 2015)

Todas as partículas apresentam um momento angular em relação ao campo gravitacional da terra, sendo que cada vetor destes spins acaba por se anular, variando unicamente com as trocas térmicas que ocorrem com o meio, como é possível verificar na figura 2.3.A. No entanto, quando estas partículas ficam sujeitas a um campo magnético externo (Figura 2.3.B) ocorrem simultaneamente dois eventos distintos: o alinhamento dos spins em direção ao novo campo magnético, bem como a precessão destes em torno do campo externo.

Os spins quando expostos a um campo magnético externo (B_0), assume dois estados de energia distintos, verificando-se assim um alinhamento de um maior número de spins no sentido do vetor do campo magnético externo, designados paralelo ou *spin-up*, sendo estes os de menor energia e um número menor em sentido contrário, antiparalelo ou *spin-down*, sendo estes de maior energia, como é possível visualizar figura 2.3.B, dando origem a uma magnetização global de equilíbrio do tecido caracterizada por M_0 .

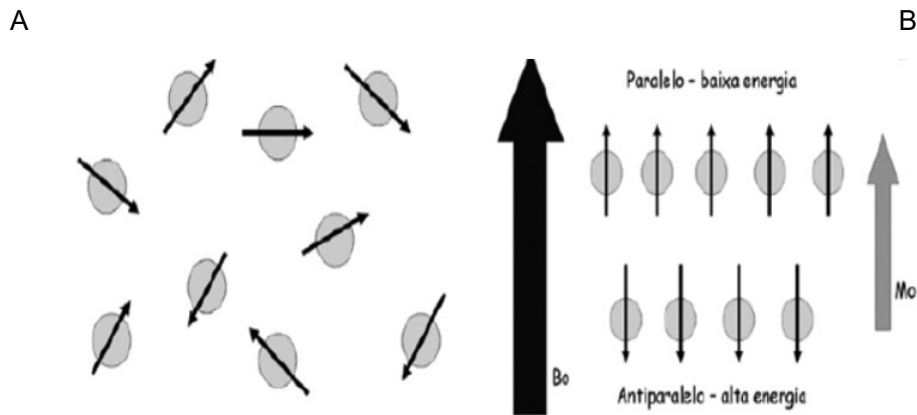


Figura 2.3. Representação dos spins quando expostos ao campo gravitacional da terra (A) e quando aplicado um campo magnético externo (B). (Watson, 1920)

Simultaneamente ao momento angular, ocorre o momento de precessão dos spins, isto é, para além do alinhamento destes com o campo magnético ocorre também a rotação dos núcleos sobre o seu próprio eixo. Pode-se fazer uma analogia simples com o pião, em rotação sobre o seu eixo.

Quanto mais intenso for o campo magnético externo B_0 , maior será a frequência de precessão e também maior diferença entre os dois estados de energia.

O movimento de precessão é determinado segundo a equação de Larmor:

$$\omega = \gamma B_0$$

Onde a frequência de Larmor ω é dada pelo produto da constante giromagnética da partícula γ e o campo magnético externo aplicado B_0 .

Os spins de um núcleo rodam em torno do campo magnético B_0 com uma frequência igual à frequência de Larmor, se aumentarmos o campo magnético maior será a frequência que o núcleo gira. Sabe-se que a frequência de precessão do hidrogénio num campo magnético de 1 Tesla é de 42MHz, o que significa que os seus spins são muito rápidos.

A distribuição dos spins pelos dois estados de energia é dada pela equação de Boltzmann:

$$\frac{N^-}{N^+} = e^{-\frac{\Delta E}{kT}}$$

Segundo esta equação, a razão entre o número de spins no estado de energia mais baixo (N^-) e o número de spins no estado de energia mais elevado (N^+) vai depender da temperatura (T) como também do campo magnético externo ($\Delta E = \gamma \frac{h}{2\pi} B_0$). À temperatura ambiente a diferença entre estes dois estados de energia é muito reduzida, por exemplo, vamos ter cerca de 6 spins em 10^6 num campo magnético de 1 Tesla. Se diminuirmos a temperatura ambiente o número de spins que estão no estado de energia mais baixo vai aumentar, permitindo deste modo, termos mais spins paralelos que antiparalelos.

Como podemos ver na figura 2.4, a magnetização dos spins que se encontram em antiparalelo vão se anular com a do mesmo número de spins que se encontram no estado paralelo. A componente transversal da magnetização dos restantes spins que se encontram no estado paralelo também se anula, permitindo assim ter uma magnetização ao longo do campo magnético B_0 , no eixo do z, denominando-se por magnetização longitudinal, deixando de existir magnetização transversal segundo o eixo x e y.

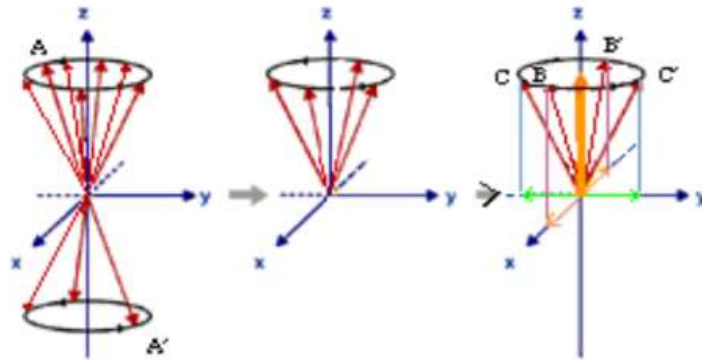


Figura 2.4. Demonstração da magnetização longitudinal (Rios, E, 1998).

O efeito combinado de todos estes spins segundo o campo magnético externo vai permitir a obtenção de imagens em ressonância magnética, no entanto este efeito por si só não nos permite captar as imagens, sendo para isso necessário aplicar também um impulso de radiofrequência. O impulso de radiofrequência aplicado terá de ter uma frequência próxima da frequência de Larmor, de forma a existirem trocas de energia com os núcleos. Estes núcleos absorvem a energia dos impulsos de radiofrequência, passando assim a existir magnetização transversal, entrando os spins em coerência de fase (Figura 2.5). Quando termina a emissão do impulso de radiofrequência, a energia absorvida pelos spins vai ser libertada.

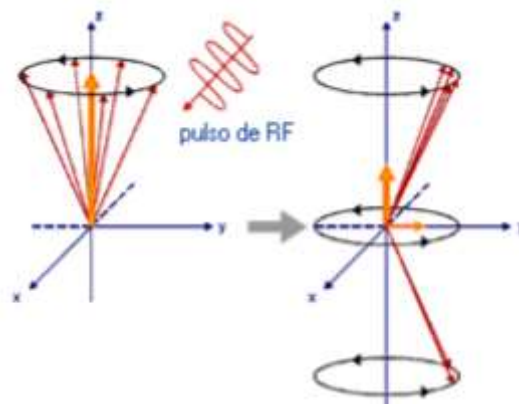


Figura 2.5. Demonstração da magnetização transversal (Rios, E, 1998).

A libertação da energia para o meio é efetuada através de dois processos de relaxação, designados por T1 e T2. O processo de relaxação T1, também conhecido por relaxação longitudinal, térmica ou spin-rede, caracteriza-se pela recuperação da

magnetização paralela (longitudinal) ao campo magnético, após estimulação eletromagnética. A constante de tempo deste processo é o tempo que decorre até a magnetização longitudinal (M_z) recuperar 63% da sua magnetização inicial. Esta relaxação é demonstrada graficamente por um crescimento exponencial da magnetização longitudinal, como é possível observar na figura 2.6.

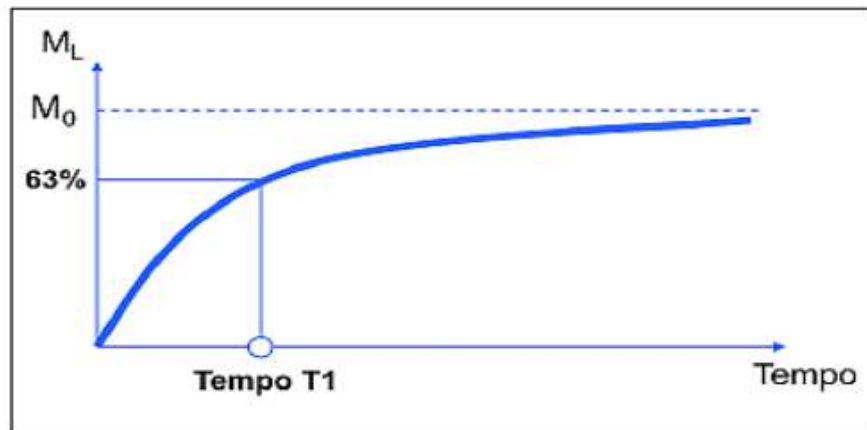


Figura 2.6. Relaxação T1. (Mazzola, 2009)

O processo de relaxação T2, relaxação transversal ou spin-spin, é caracterizado pela perda da magnetização perpendicular (transversal) ao campo magnético, após estimulação eletromagnética. A constante de tempo deste processo corresponde ao tempo que decorre até a magnetização transversal decair para 37% do seu valor inicial. Esta representa-se graficamente por um decréscimo da magnetização transversal (Figura 2.7).

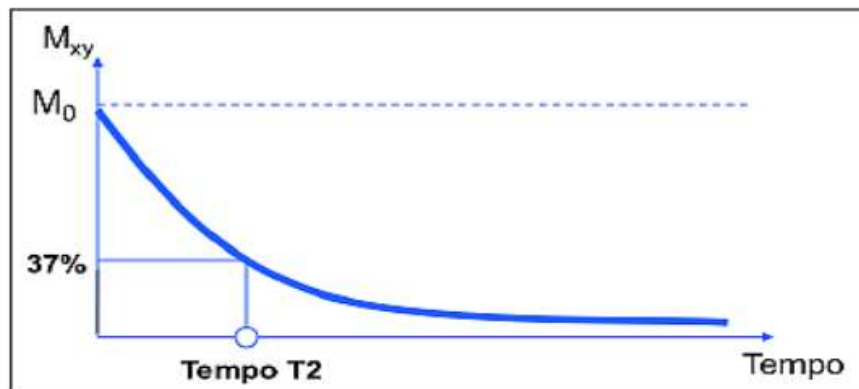


Figura 2.7. Relaxação T2. (Mazzola, 2009)

O decaimento da magnetização transversal, conhecido como T2 está relacionado com o tempo natural ou efetivo que os núcleos demoram a perder a coerência, no entanto esta decai muito mais rapidamente. Caracterizando-se este decaimento por T2*, designando-se por decaimento por indução livre. Resultante das heterogeneidades do campo magnético principal. Podendo estas estar relacionadas com propriedades intrínsecas do próprio magneto, bem como, com alterações do campo provocadas por suscetibilidade magnética dos tecidos ou materiais colocados no campo. Nas sequências *spin-eco*, que será abordada mais adiante, esta heterogeneidade é eliminada, com a aplicação do impulso de 180°, permitindo observar-se o T2 natural dos tecidos.

Existem diferentes parâmetros que vão influenciar os tempos de relaxação T1 e T2. Na relaxação longitudinal T1, estes tempos dependem da energia térmica dos prótons excitados e da transferência dessa energia com o meio envolvente. Se o meio envolvente for constituído por um líquido puro, ou até mesmo água esta recuperação demora mais tempo a acontecer, pois as moléculas da água movimentam-se a uma velocidade muito elevada dificultando a transferência de energia (ou seja, o retorno dos prótons ao seu estado de equilíbrio). Em tecidos compostos por diferentes tamanhos moleculares, a energia apresenta, à partida, tempos de relaxação longitudinal mais curtos, dada a troca de energia mais rápida que é proporcionada pelo movimento mais lento das moléculas.

Como podemos verificar na figura 2.8 abaixo, o líquido cefalorraquidiano (CFS) apresenta um T1 mais longo, necessitando de 2500 ms para recuperar 63% da sua magnetização longitudinal. Por outro lado, a gordura atinge a mesma percentagem de recuperação da magnetização num período de tempo muito mais curto, de 240 ms.

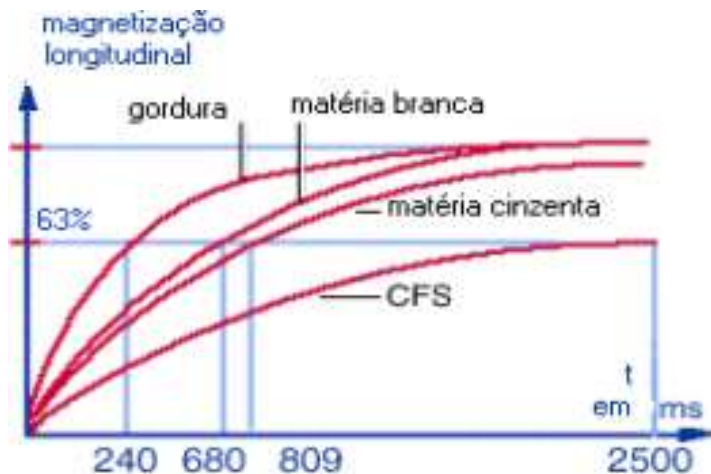


Figura 2.8. Magnetização Longitudinal dos tecidos (Rios, E, 1998).

A relaxação transversal T2 é resultante da heterogeneidade presente nos campos magnéticos externo e interno, bem como as interações moleculares, representando a taxa a que decai a magnetização transversal. Como já foi mencionado anteriormente as moléculas de água, que se movimentam rapidamente e têm tempos de precessão maiores, apresentam um tempo de relaxação transversal T2 mais longo que líquidos não puros ou tecidos compostos por moléculas de maiores dimensões. Estas diferenças estão representadas graficamente na figura 2.9.

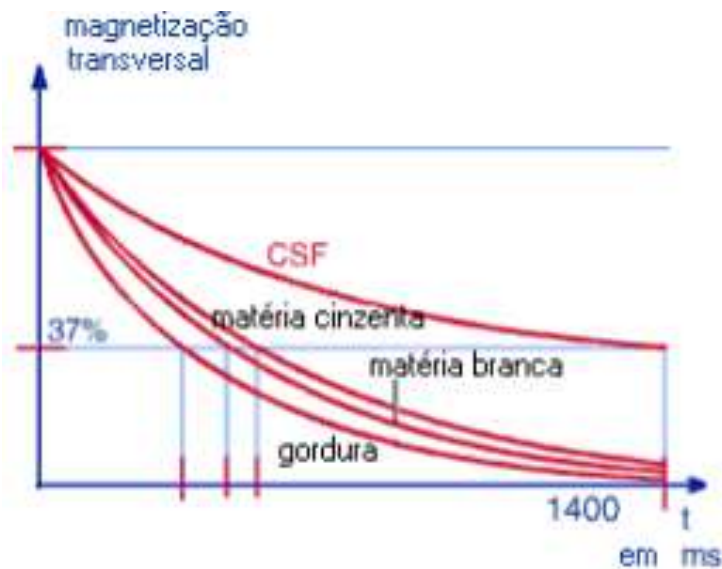


Figura 2.9 - Magnetização transversal dos tecidos. (Rios, E, 1998)

Deste modo, todos estes processos influenciam o contraste entre os tecidos em imagens de ressonância magnética. Para se obter o contraste desejado, é necessário ajustar parâmetros como a duração, amplitude e frequência de repetição dos impulsos de radiofrequência, bem como os gradientes do campo magnético. Na Tabela 2.1 apresentam-se os diferentes tempos de relaxação longitudinal e transversal dos tecidos do corpo humano quando exposto a um campo magnético de 1.5T, podendo através destes valores quantitativos ter como referência para os tempos de relaxação T1 e T2 dos tecidos quando expostos campo magnético.

Tabela 2.1 - Tempos de relaxação T1 e T2 de diferentes tecidos - 1.5 Tesla.

(Mazzola, 2009)

Tecido	T1 (ms)	T2 (ms)
Substância branca	790	90
Substância cinzenta	920	100
Líquido céfalo-raquidiano (líquor)	4000	2000
Sangue (arterial)	1200	50
Parênquima hepático	490	40
Miocárdio	870	60
Músculo	870	50
Lipídios (gordura)	260	80

Além dos tempos de relaxação, para gerarmos imagens é necessário ainda considerar a formação de ecos, surgindo relativamente a estas duas grandes famílias nas sequências de impulsos, a *Spin Echo* e a Eco de gradiente.

A sequência *Spin Echo* traduz-se essencialmente na emissão de impulso radiofrequência de 90° seguido de um 180°. Este procedimento foi descrito por Hahn em 1950 (Hahn, 1950), referindo que os prótons excitados com um impulso de radiofrequência inicial e um segundo impulso emitido após um tempo predefinido de pausa, permite gerar um eco resultante de um processo natural de desfasamento dos momentos magnéticos induzidos pelo segundo impulso. Controlando o tempo que decorre entre o primeiro e o segundo impulso dá-se à medição do tempo de eco (TE) que é resultante do intervalo que vai da aplicação do primeiro impulso ao pico do eco, determinado a relaxação longitudinal presente no eco, o TR será o tempo de repetição para dar início a um novo ciclo da sequência, podendo deste modo determinar a recuperação da magnetização longitudinal entre os dois pulsos de radiofrequência de 90° (Figura 2.10).

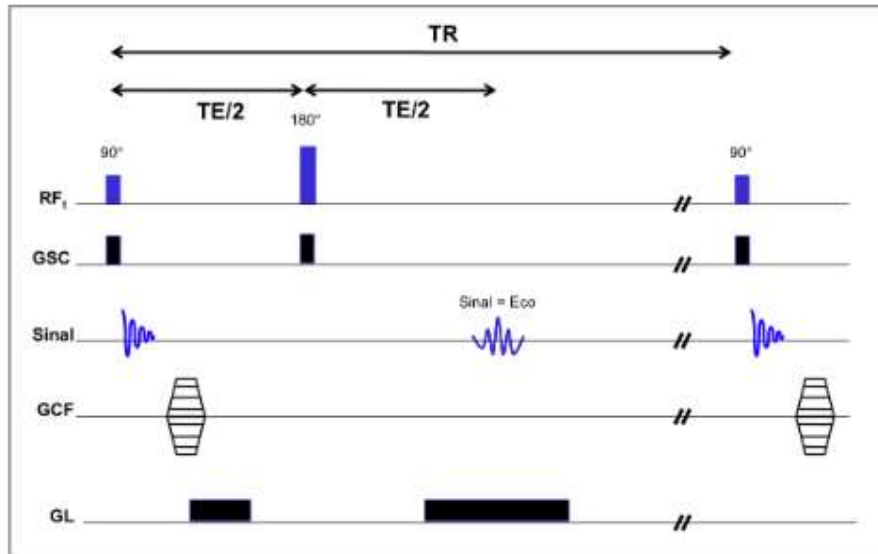


Figura 2.10 - Sequência *Spin Echo*. (Mazzola, 2009)

Em cada tempo de repetição (TR) é preenchido uma linha do espaço K, de acordo com o preenchimento convencional do espaço k (abordado na seção seguinte). Ao controlar os TR e TE das sequências, permite obter imagens com diferentes ponderações (tabela 2.2).

Tabela 2.2 -Diferentes ponderações na imagem em sequências *Spin Echo*

(Junge t al., 2013)

Tempo de repetição (TR)	Tempo de eco (TE)	Ponderação
TR Curto (< 500 ms)	TE Curto (5 a 25 ms)	T1
TR Longo (> 1500 ms)	TE Longo (> 90 ms)	T2
TR Longo (> 1500 ms)	TE Curto (5 a 25 ms)	DP

Nas sequências *Spin Echo* a presença de um impulso de 180° vai eliminar as heterogeneidades do campo magnético, permitindo deste modo obtermos sequências sempre ponderadas em T2 e não em T2*.

A sequência Eco de gradiente é caracterizada pela ausência do impulso de 180°, não eliminando assim a heterogeneidade magnética do campo B₀. Nesta sequência como são utilizados os gradientes do campo magnético para formar o eco, o tempo de TR diminui bastante, assim como a possibilidade de controlar a magnetização

longitudinal através da rotação do *flip angle*, α , sendo este último o parâmetro mais determinante para o contraste da imagem, nesta família de sequências (Figura 2.11).

Nesta sequência, com a utilização dos gradientes de campo magnético é possível aplicar inicialmente um gradiente com polaridade negativa. Isto irá permitir um desfasamento da precessão dos spins. Quando este gradiente é revertido positivamente, os spins retomam à sua precessão em fase gerando um eco.

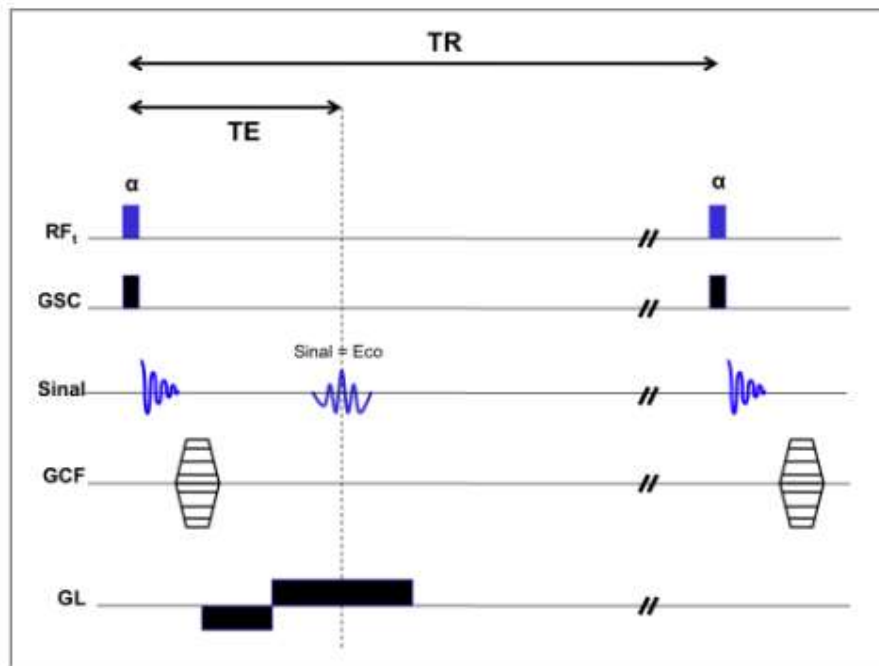


Figura 2.11 - Sequência gradiente eco. (Mazzola, 2009)

Como já foi referido anteriormente, o corpo humano quando exposto a um campo magnético vai ter todos os seus prótons a girar em torno do campo magnético externo. Para podermos obter as imagens em ressonância magnética temos que estimular estes prótons para ter informação de diagnóstico, sendo também necessário fazer a localização anatômica do sinal e a reconstrução do mesmo em imagem.

Para a localização do sinal no espaço é necessário fazer a localização do corte e do sinal dentro deste. Assim sendo, com o auxílio das bobines de gradiente é possível fazer a localização do sinal no espaço, aplicando gradientes lineares nas direções x, y e z. No que diz respeito à seleção do corte, aplica-se um gradiente de campo magnético ao longo do corpo do utente (Figura 2.12).

Com o auxílio das bobines de gradiente é possível variar numa determinada coordenada a frequência, tendo frequências diferentes com localizações diferentes.

Permitindo assim a determinação da seleção de corte e com o auxílio da largura de banda a espessura de corte.

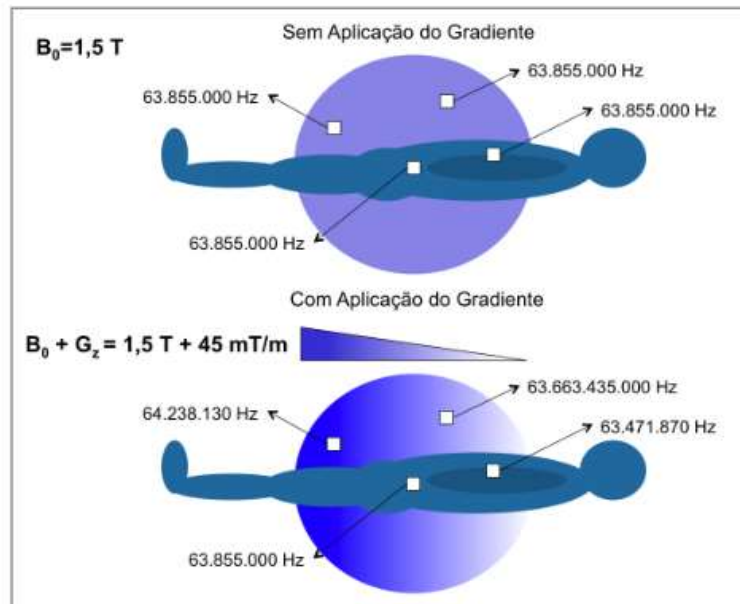


Figura 2.12 - Variação Campo magnético no corpo Humano. (Mazzola, 2009)

Essencialmente, sempre que é aplicado um impulso de radiofrequência tem que se ativar em simultâneo o gradiente de seleção de corte. Posteriormente para a localização dentro do corte é aplicado o gradiente de codificação de fase. Quando se desliga este gradiente ir-se-á ter os spins dos prótons do corte a processar numa fase específica. No momento de aquisição do sinal é ligado o gradiente de codificação em frequência, e nessa altura cada *pixel* do corte irá ter prótons com fases e frequências de processão distintas.

O registo das frequências e fases espaciais de cada *pixel* da imagem será guardado em cada ponto no espaço K, o qual através da transformada de Fourier é convertido na imagem final de ressonância magnética.

2.3. Neuroimagem em doença de Parkinson

Os gânglios da base têm sido alvo de estudo para a compreensão da doença de Parkinson. Anteriormente à década de 90 do século passado, a maioria do trabalho incidia na associação de défices reconhecidos em vida, com lesões cerebrais avaliadas visualmente em indivíduos *post mortem*. Os avanços tecnológicos que ocorreram desde então, nomeadamente a introdução de técnicas de imagem médica, permitiram a avaliação da morfologia cerebral *in vivo*, revolucionando o estudo e a possibilidade de diagnóstico da doença (Provost et al., 2015).

Inicialmente as técnicas de imagem mais utilizadas para o estudo da doença de Parkinson foram a Tomografia por emissão de positrões (PET) e a Tomografia Computorizada por emissão de fotão único (SPECT). Recentemente, a Ressonância Magnética tem ganho um papel de destaque. Esta envolve um menor custo económico que as técnicas anteriores, não requer a utilização de radiações ionizantes, e possibilita avaliar a conectividade anatómica e funcional com maior resolução espacial e temporal (Carrete, 2017).

Dadas estas vantagens, não é de surpreender o investimento forte na utilização de Ressonância Magnética para o estudo de doenças neurodegenerativas nos últimos anos. No entanto, é de referir que a maioria destas situações está direcionada apenas à doença de Alzheimer, sendo comparativamente reduzido o número de casos de aplicação no estudo da doença de Parkinson. Relativamente ao diagnóstico desta última doença, é de referir ainda a elevada prevalência de utilização de teste verbais.

2.3.1 Neuromelanina e ressonância magnética

A neuromelanina presente no nosso cérebro atua na proteção dos neurónios contra processos oxidativos do ferro. Este composto encontra-se em maiores quantidades na substância *nigra pars compacta*, resultando numa coloração negra nesta região do cérebro (Cassidy et al., 2019).

Como descrito anteriormente, a produção de dopamina no cérebro vai diminuindo com o envelhecimento em resultado da morte celular na substância nigra. Consequentemente assiste-se também a uma diminuição da neuromelanina nesta região do cérebro, o que relativamente à doença de Parkinson representa um aumento de danos oxidativos nas células.

A neuromelanina apresenta propriedade paramagnética, devido à sua capacidade de armazenamento de ferro. O ferro que se encontra no sistema nervoso encontra-se ligado às células sob a forma de ferritina (Fe^{3+}), protegendo deste modo as células do processo oxidativo do ferro livre (Fe^{2+}) (Zecca et al., 2004).

Devido às propriedades paramagnéticas da neuromelanina, a sua visualização é possível na Ressonância Magnética. Por exemplo, com a utilização de ponderações T1 consegue-se observar alterações no seu sinal, sendo que com elevadas concentrações desta temos hipersinal na substância nigra, ao passo que com a progressão da doença de Parkinson ocorre uma diminuição deste sinal, como representado na figura 2.13. (Schwarz et al., 2011, Trujillo et al., 2017).

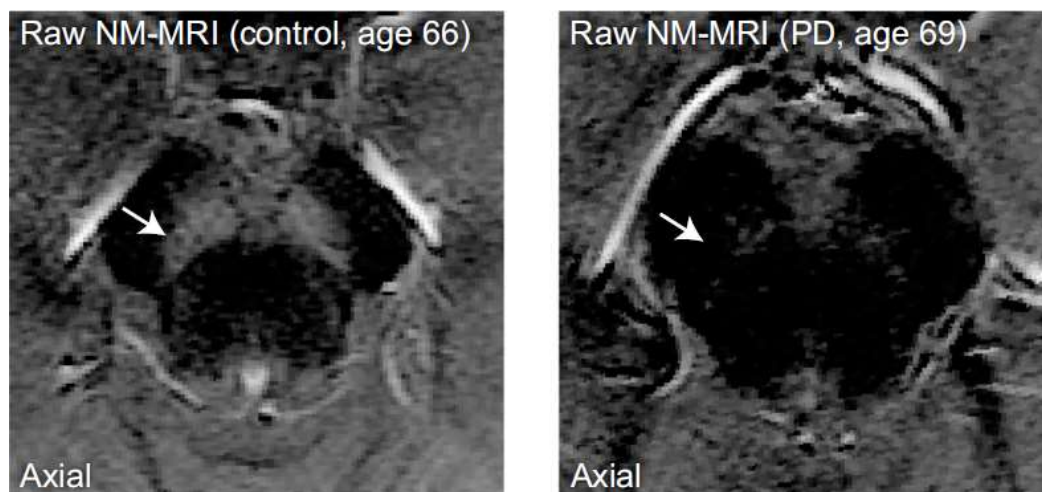


Figura 2.13: Imagens RM da Substância Nigra, setas brancas indicam as alterações de sinal da neuromelanina, imagem paciente controlo (Esquerda) e paciente com DP (Direita). (Schwarz et al., 2011, Trujillo et al., 2017)

No entanto, verificou-se que para efetuar este tipo de estudo, é importante ter em consideração o campo magnético utilizado, sendo que em campos magnéticos abaixo de 1.5 T ocorre aparentemente diminuição da capacidade de visualização e diagnóstico (Meijer et al., 2012).

2.3.2 Nigrossoma-1 e ressonância magnética

Os nigrossomas são aglomerados de pequenas células dopaminérgicas localizados na substância nigra. Existem cerca de cinco nigrossomas, sendo o maior o nigrossoma-1. Este está posicionado na substância nigra lateral e médio lateral, dando à sua configuração visual a aparência de uma cauda de andorinha (Meijer et al., 2016), sendo por isso conhecido simplesmente por “cauda de Andorinha”).

Este nigrossoma é um biomarcador importante da doença de Parkinson, uma vez que apresenta o maior número de neurónios que são afetados pela mesma (Oliveira et al., 2017). Estudos recentes verificaram que, apesar da sua reduzida dimensão de apenas alguns milímetros, este apresentou diferenças significativas na sua estrutura em doentes com doença de Parkinson versus doentes sem a patologia (Schwarz et al., 2014).

A estrutura em forma de cauda de andorinha do nigrossoma-1 consegue-se diferenciar em ressonância magnética, através de sequências eco de gradiente T2* (T2 SWI), sendo que a ausência deste biomarcador indica com elevada precisão o diagnóstico da doença de Parkinson, como é possível visualizar na figura 2.14 (Schwarz et al., 2014).

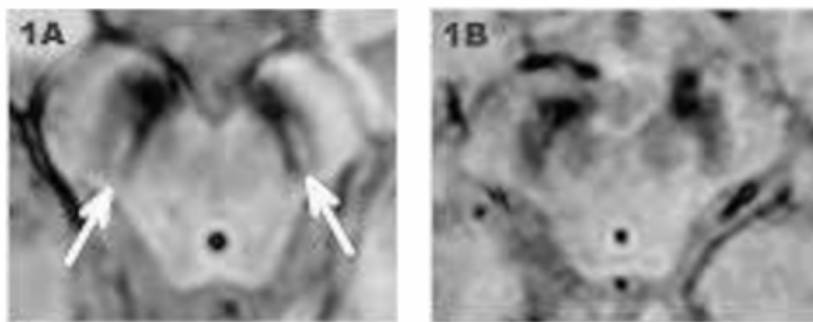


Figura 2.14: Imagem RM T2 SWI. (1A) Paciente controlo, visualização da cauda de andorinha; (1B) Paciente com DP, visualização da perda de diferenciação. (Schwarz et al., 2014)

2.4. Deep Learning

O *Deep Learning* é um subcampo da aprendizagem automática (do inglês 'machine learning'), ou mais genericamente da "inteligência artificial", que se dedica ao desenvolvimento de modelos de redes neuronais artificiais (Bishop, 1995).

A arquitetura das redes neuronais artificiais pode variar consideravelmente, mas de um modo geral caracteriza-se por possuir várias 'camadas', cada uma destas composta por 'neurónios', que são unidades de processamento de informação capazes de transformar os dados utilizando funções matemáticas distintas (por exemplo, lógicas ou sigmoidais) (LeCun et al., 2015). Em termos simples, uma rede neuronal artificial 'aprende' a distinguir as características relevantes nos dados ao fazer circular a informação ao longo das suas camadas. Este fluxo de informação serve para otimizar, de forma iterativa, os parâmetros das funções a serem calculados pelos neurónios, assim como os 'pesos' a dar aos diferentes atributos presentes nos dados, apresentando uma camada de entrada (*inputs*), algumas camadas intermédias também designadas por ocultas e uma camada de saída (*output*) (Nielsen, M. A., 2015).

No *Deep Learning* a arquitetura das suas redes neuronais tem sempre uma camada de entrada e uma camada de saída, tal como nas redes neuronais simples, o que o faz diferenciar são as diversas camadas intermédias ('ocultas'). As redes *Deep Learning* são redes neuronais artificiais compostas por múltiplas camadas de neurónios, capazes de identificar automaticamente características distintas em dados com estruturas complexas, como imagens, sequências genéticas, linguagem humana, ou séries temporais (Bengio et al., 2020; LeCun et al., 2015). As principais arquiteturas no *Deep Learning* incluem as redes *multilayer perceptrons*, as redes neuronais recorrentes, *long short-term memory* (LSTM), as máquinas de Boltzmann, as redes de Hopfield, as redes neuronais convolucionais entre outras.

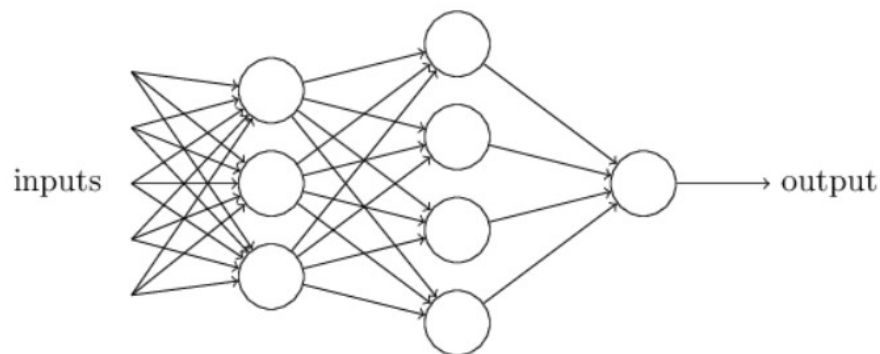


Figura 2.15: Rede Neuronal Convolucional, CNN (Nielsen, M. A., 2015).

De entre estas, destaca-se a arquitetura das redes neuronais convolucionais, CNN (do inglês Convolutional Neural Network) (Figura 2.15), desenvolvido por LeCun et al. (1998), e que teve origem numa aplicação para reconhecimento de dígitos manuscritos em imagens – designado por LeNet. Com este avanço foi possível desenvolver arquiteturas adicionais para a classificação de imagens, como abordarei mais à frente. As redes neuronais convolucionais são ainda das mais utilizadas atualmente no reconhecimento de imagens, o que após a sua parametrização, o seu treino, os respetivos pesos obtidos, pode determinar a probabilidade de uma imagem pertencer ao diagnóstico de doença de Parkinson positivo ou negativo.

O elevado potencial dos modelos *Deep Learning* foi reconhecido apenas recentemente, em virtude do surgimento de técnicas que permitiram o seu treino com recurso a grandes volumes de informação. O ano de 2012 marca de forma particular este ponto de viragem, com a apresentação de uma arquitetura *Deep Learning* que conquistou, com larga vantagem, uma competição de referência na área do reconhecimento computadorizado de imagens. Esta rede, que adquiriu a designação ‘AlexNet’ (Krizhevsky et al., 2012), serviu de referência à criação de um vasto número de novas arquiteturas, várias delas responsáveis por incrementos significativos na precisão do reconhecimento de imagens em anos seguintes (p. ex., ‘VGG-16’; ‘ResNet’, etc).

Esta evolução está linearmente relacionada com a evolução da tecnologia, pois o aumento no volume de dados disponíveis em conjunto com um aumento da capacidade de processamento computacional possibilitou aumentar o desempenho dos modelos de *Deep Learning*, representado na figura 2.16 (Ng. A., 2018).

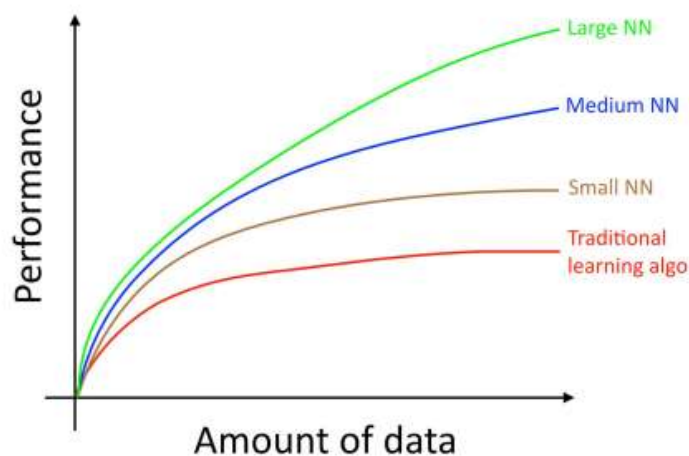


Figura 2.16: Desempenho do *Deep Learning* vrs redes tradicionais. (Ng. A., 2018)

Alguns estudos referem que o diagnóstico da doença de Parkinson efetuado por pessoal médico apresenta uma taxa de erro de aproximadamente 25% (Masoom Shah et al., 2018; Tolosa et al., 2006). Deste modo é aparentemente ainda difícil garantir que um doente seja corretamente diagnosticado e receba o tratamento necessário. Dada esta taxa de erro relativamente elevada, torna-se aparente a necessidade de avaliar a capacidade preditiva dos modelos de *Deep Learning* na identificação da doença de Parkinson.

2.4.1. Arquiteturas das CNN

Como já foi mencionado anteriormente, as redes neuronais convolucionais tiveram a sua origem na criação de redes em que entre as camadas de entrada e de saída apresentava múltiplas camadas intermédias, ou ocultas. A característica definidora das redes convolucionais é a presença de camadas que atuam como ‘filtros’ e que através de operações de convolução identificam automaticamente características relevantes nos dados de entrada. Estas características são identificadas pelas diferentes camadas e é através da combinação destas características, ou seja, dos ‘pesos’ aprendidos pelas diferentes camadas, que o modelo calcula a probabilidade de cada uma das classes representadas pelos dados.

Uma propriedade vantajosa das redes neuronais artificiais, é que a aprendizagem pode ser guardada e utilizada para tarefas de classificação futura. Ou seja, existe a possibilidade de ajustar modelos pré-treinados e que já foram ‘ensinados’ a reconhecer características com alguma relevância para a tarefa de classificação que se pretende executar. Por exemplo, várias das arquiteturas mais conhecidas como a ResNet, estão disponíveis *online* pré-treinadas com bases de dados de imagens de enorme dimensão, como a IMAGENET (Fei-Fei et al., 2009) que é atualmente composta por cerca de 14 milhões de imagens em representação de mais de 100 000 classes distintas, como ‘carro’, ‘peixe-balão’, ‘semáforo’, etc. Deste modo, estes modelos podem ser utilizados diretamente na classificação automática destas classes ou podem ser treinadas adicionalmente para reconhecer novas classes. A vantagem da segunda abordagem sobre um treino a partir do zero, é que o modelo já possui filtros capazes para reconhecer um conjunto alargado de características distintas em imagens, e portanto, este passo da aprendizagem já não necessita de ser realizado.

O treino (ou ‘aprendizagem’) de uma rede neuronal consiste na passagem dos dados por toda a arquitetura, desde a camada de entrada até à camada de saída. Este processo é iterativo, sendo cada ciclo completo (ou seja, a passagem da informação na camada de entrada até à camada de saída) designado de ‘época’. A aprendizagem das redes neuronais consiste no ajuste iterativo dos pesos das ligações entre os neurónios das diferentes camadas e que resultam das funções de ativação em cada neurónio. O objetivo contínuo do processo de aprendizagem é identificar os pesos que fazem ‘sobressair’ as características das imagens que melhor permitem proceder à sua correta classificação. A avaliação da classificação produzida pelo modelo a cada época é feita através de validação cruzada, ou seja, parte das imagens são excluídas do processo de treino, sendo utilizadas apenas para comparação com as classificações produzidas pelo

mesmo. A variação na taxa de sucesso desta classificação guia o processo de treino na escolha das características mais e menos relevantes na discriminação das classes. Durante este procedimento, um critério relevante a ter em conta diz respeito ao número de épocas que o modelo é treinado. Um número reduzido de épocas (eventualmente <10) pode ser insuficiente para identificar as características relevantes na distinção das classes de interesse. Por outro lado, um número muito elevado (eventualmente > 100) pode levar à extração de características demasiado específicas das imagens de treino e que não têm relevância além destas, reduzindo a capacidade de generalização, e como tal de capacidade preditiva, do modelo. Quando isto acontece diz-se que o modelo está sobre-ajustado (ou em inglês 'overfitted').

2.4.1.1. LeNet

A arquitetura LeNet (LeCun et al., 1998) é arquitetura inicial das redes convolucionais. No entanto, ainda é a partir desta que assentam a maioria dos conceitos base dos algoritmos '*Deep Learning*'. A LeNet apresenta na sua totalidade 7 camadas, sem contar com a camada de entrada (Figura 2.17). As camadas estão dispostas alternadamente, com 2 camadas convolucionais com filtros de 2x2 pixels, 2 camadas de *pooling* médias, servindo estas camadas para 'resumir' a informação extraída pelas camadas de filtros anteriores. De seguida são aplicadas mais 2 camadas totalmente conectadas entre si, e por fim a camada de saída (LeCun et al., 1998).

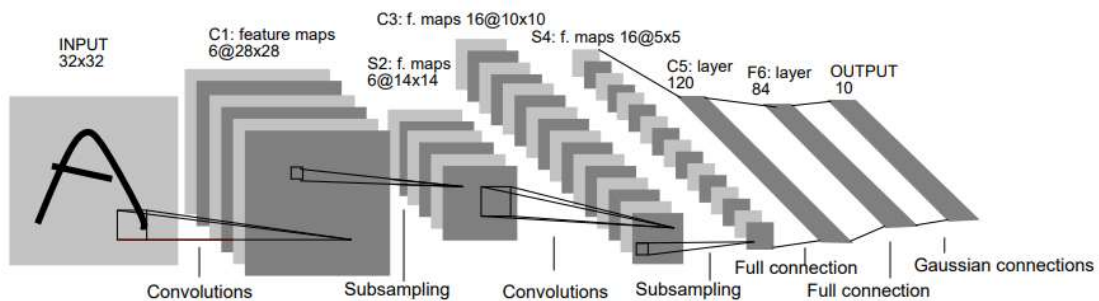


Figura 2.17: Arquitetura LeNet (LeCun et al., 1998)

2.4.1.2. AlexNet

A arquitetura AlexNet (Krizhevsky et al., 2012) alcançou notoriedade ao vencer um importante concurso de reconhecimento de imagens, o *ImageNet Large Scale Recognition Challenge* (ILSVRC), em 2012. Esta arquitetura apresenta uma estrutura muito semelhante à LeNet, com a introdução de mais algumas camadas de filtros. É de particular relevância a utilização pela primeira vez da função de ativação 'ReLU', uma função que transforma os valores de entrada em valores não negativos escalonados linearmente, e que ainda é utilizado na generalidade dos modelos *Deep Learning*.

A arquitetura AlexNet é composta por 8 camadas (Figura 2.18), não contando com a camada entrada, em que as 5 primeiras são camadas convolucionais, algumas delas seguidas de camadas de *pooling* e as 3 últimas são camadas totalmente interconectadas (Krizhevsky et al., 2012).

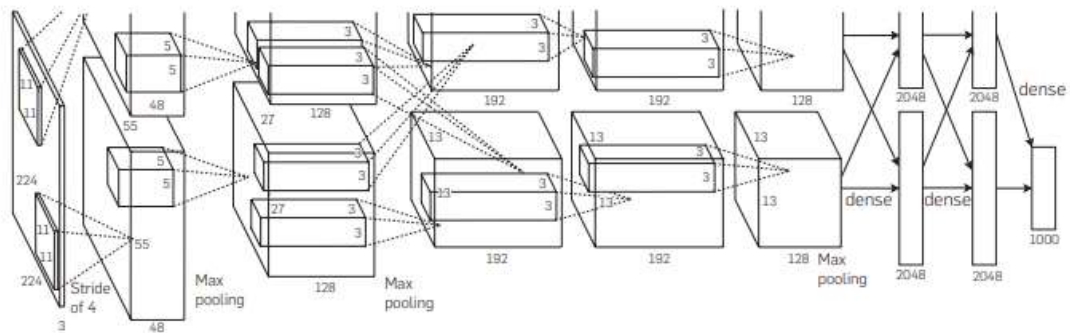


Figura 2.18: Arquitetura AlexNet (Krizhevsky et al., 2012)

2.4.1.3. Inception

A arquitetura Inception (InceptionV1), também conhecida por GoogLeNet, ganhou o concurso da ILSVRC em 2014. Apresentando o menor erro na classificação do conjunto de dados da ImageNet. Esta arquitetura teve como fundamento um melhor aproveitamento dos recursos computacionais através de um aumento da profundidade e largura da rede (Szegedy et al, 2014).

Esta arquitetura é composta por 22 camadas, várias destas convolucionais com tamanhos de filtro reduzidos e com fatores de ativação ReLu associados, de forma a melhorar o seu desempenho.

Foram desenvolvidas várias versões desta arquitetura após a sua apresentação em 2014, tendo em 2015 sido publicado um artigo com uma alteração significativa na arquitetura. Esta foi apresentada como a InceptionV3 (Szegedy et al, 2015), tendo um aumento de 2,5 vezes do desempenho preditivo, requerendo um menor número de parâmetros e sendo por isso 6 vezes menos exigente em termos computacionais (Figura 2.19).

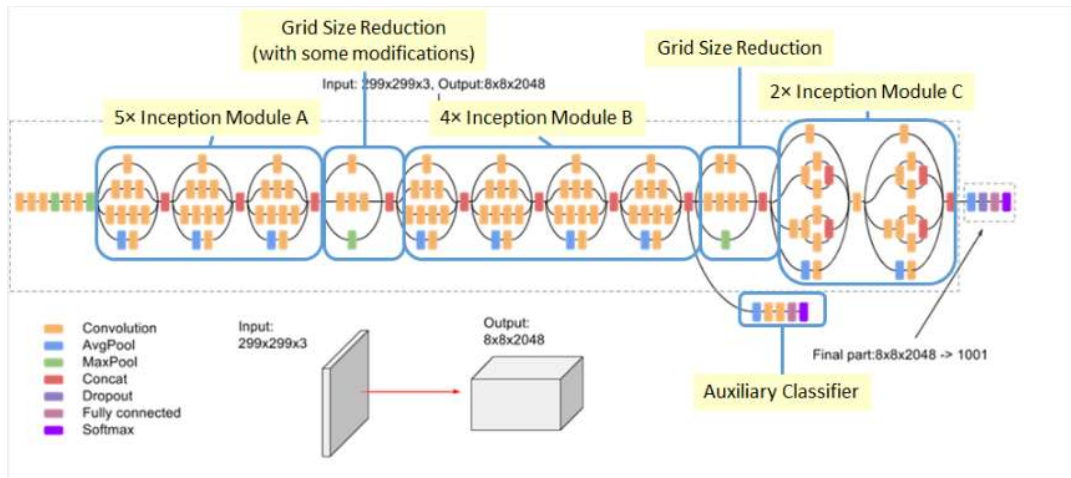


Figura 2.19: Arquitetura InceptionV3 (Tsang, S.-H, 2019)

2.4.1.4. ResNet

A ResNet, também conhecida por *Residual Neural Network* foi a primeira a apresentar um erro de classificação inferior ao ser humano, ao vencer o primeiro prémio em 2015 no concurso ILSVRC da ImageNet (He et al., 2015). A maior diferenciação desta arquitetura face às descritas anteriormente diz respeito à existência de um número muito mais elevado de camadas ocultas. Dependendo da sub-variante considerada, a arquitetura resnet é tipicamente composta por 34 camadas (variante 'ResNet34'), 50 camadas (variante 'ResNet50'), 101 camadas ('ResNet101'), ou mais (Figura 2.20).

3. Objetivos

O presente trabalho pretende efetuar uma revisão das tendências e padrões de publicação científica em redes neuronais profundas aplicadas à classificação de patologias em imagem de ressonância magnética. Pretende-se obter uma descrição dos padrões gerais de publicação, principais áreas de aplicação investigadas e resultados obtidos. O objetivo é ter uma visão geral do estado da arte na área, assim como na área da doença de Parkinson.

Após se efetuar uma revisão bibliográfica do estado de arte, dos estudos relacionados com a utilização de imagens de ressonância magnética no diagnóstico da doença de Parkinson com o auxílio de redes neuronais profundas, pretendeu-se testar de forma robusta a aplicação de redes neuronais profundas no diagnóstico da doença de Parkinson, bem como avaliar a relevância de vários fatores extrínsecos às arquiteturas, nomeadamente a ponderação de ressonância magnética utilizada e características dos utentes, na fiabilidade das classificações obtidas.

Para atingir estes objetivos, procedeu-se ao desenvolvimento de um conjunto alargado de combinações de arquiteturas de *Deep Learning* e utilização de diferentes ponderações em ressonância magnética e avaliou-se o desempenho de cada uma na deteção da doença. Adicionalmente, estudou-se também se existem fatores relacionados com características do utente que implicam maior ou menor capacidade preditiva dos modelos. Mais especificamente, após se identificar a combinação de arquitetura e ponderação de imagem que apresentam maior capacidade preditiva, testou-se se existiriam diferenças significativas na taxa de erro do mesmo em função do género e da idade dos utentes. A hipótese de partida relativamente ao fator idade é que imagens de utentes com idade mais avançada terão tendencialmente maiores taxas de erro pois os modelos deparam-se com uma maior diversidade nos padrões encefálicos (Cole et al., 2017). Quanto ao fator género, é do conhecimento científico que existe uma maior prevalência da doença no género masculino, uma relação de 3:2 (Twelves et al., 2003), desta forma é plausível considerar que possam existir mecanismos, ainda que desconhecidos, que façam diferir a capacidade de diagnóstico automático da doença em função do género. De igual modo, considerando um efeito plausível para este último fator, é também hipoteticamente razoável esperar que a relação entre taxas de erro e idade do utente seja distinta entre géneros. Ou seja que, por exemplo, pessoas do género masculino e de idade avançada possam apresentar taxas de erro distintas das de pessoas do género feminino com a mesma idade.

Deste modo, este trabalho vai um passo além da maioria dos trabalhos anteriores que aplicam *Deep Learning* no diagnóstico da doença de Parkinson. Ou seja, além de uma avaliação da *performance* preditiva de diferentes arquiteturas, intenta-se também identificar quais as ponderações de ressonância magnética que melhor se adequam a cada arquitetura assim como identificar fatores dos utentes com implicações nas *performances* preditivas.

4. Metodologia

4.1. Estado da Arte da aplicação do *Deep Learning* em Ressonância Magnética

No dia 10 de Fevereiro de 2021 foi realizada uma pesquisa bibliográfica no motor de busca ‘*Web of Science*’ (WoS) (<https://webofknowledge.com>) utilizando a seguinte expressão: “*Deep Learning*” AND “*mri*” AND “*diagnosis*”. Esta conjugação de termos pretende ir de encontro ao objetivo do trabalho, que se prende com a revisão bibliográfica de trabalhos que versam sobre a utilização de redes neuronais profundas para diagnóstico com recurso a imagens de ressonância. As opções de pesquisa foram parametrizadas para devolver apenas resultados referentes a artigos científicos (isto é, excluindo resumos de conferências, livros, capítulos de livros ou outros) e publicados entre 2015 e a data da pesquisa. Estes parâmetros (artigos científicos publicados de 2015 em diante) pretendem restringir os resultados da pesquisa a análises mais robustas (revistas com revisão por pares), sendo que o ano 2015 também corresponde ao início da investigação na temática das redes neuronais profundas aplicadas a imagens de ressonância (Figura 5.1), pelo que no plano temporal, os resultados são também representativos da totalidade de investigação produzida.

Após se obterem os resultados da pesquisa, efetuou-se o ‘*download*’ desta para uma tabela tipo ‘folha de cálculo’. De seguida, e com base na leitura do resumo (‘*abstract*’) de cada trabalho identificado, procedeu-se à sua classificação relativamente ao enquadramento no contexto do trabalho (aplicação de algoritmos *Deep Learning* para auxílio a diagnóstico com imagens de ressonância magnética). Verificou-se, por exemplo que um número elevado de trabalhos identificados era referente a segmentação cerebral e vertebral, e como tal não estava enquadrado. Para os trabalhos classificados como enquadrados, procedeu-se à sua diferenciação entre trabalhos de revisão e trabalhos de investigação original. Para os primeiros procedeu-se à sua classificação em revisão sistemática, não-sistemática e meta-análise. Os trabalhos de investigação original foram classificados quanto aos seguintes fatores: 1) patologia ou patologias investigadas; 2) parte (s) do corpo consideradas; 3) utilização de imagem de ressonância magnética ou desta em combinação com outros suportes de diagnóstico (p. ex. imagens TAC); 4) quais as tipologias de redes neuronais profundas consideradas (convolucionais 2D ou convolucionais 3D); 5) se a comparação de desempenho preditivo entre várias arquiteturas de redes é efetuada; 6) se a exatidão do diagnóstico

com o diagnóstico realizado por humanos é comparada (médicos especialistas), e por fim 7) se o desempenho do modelos é considerado pelos autores como “bem-sucedido”, “moderado/a melhorar” ou “insuficiente/ineficaz”.

Estas categorias permitem ter uma visão abrangente das áreas médicas investigadas, das opções técnicas e de avaliação das redes neuronais utilizadas, assim como das percepções acerca da capacidade e utilidade deste tipo de modelos.

4.2. Detecção da doença de Parkinson com *Deep Learning*

4.2.1 Recolha dos dados

De forma a alcançar os objetivos propostos, o trabalho foi dividido em duas etapas. Na primeira fase iniciou-se pela recolha de imagens de Ressonância Magnética a partir de uma base de dados de acesso e utilização gratuita, a *Image and Data Archive* (IDA; Crawford et al., 2016). Esta base de dados é administrada pelo Laboratório de Neuro imagem (LONI) do *USC Mark and Mary Stevens Neuroimaging and Informatics Institute* e está em funcionamento desde 1990. No IDA é possível ter acesso a imagens de diversas modalidades na área da neuroimagem, entre as quais a ressonância magnética. De forma a ter acesso a esta base de dados, foi apenas necessário efetuar o registo na plataforma, podendo assim ter livre acesso aos dados disponibilizados pela mesma.

Para a identificação e recolha das imagens integrantes deste estudo tiveram-se em consideração os seguintes critérios: utentes com a doença de Parkinson (PD), utentes sem a doença, idade acima dos 20 anos e as Ponderações T1, T2, DP e FLAIR (Figura 4.1).

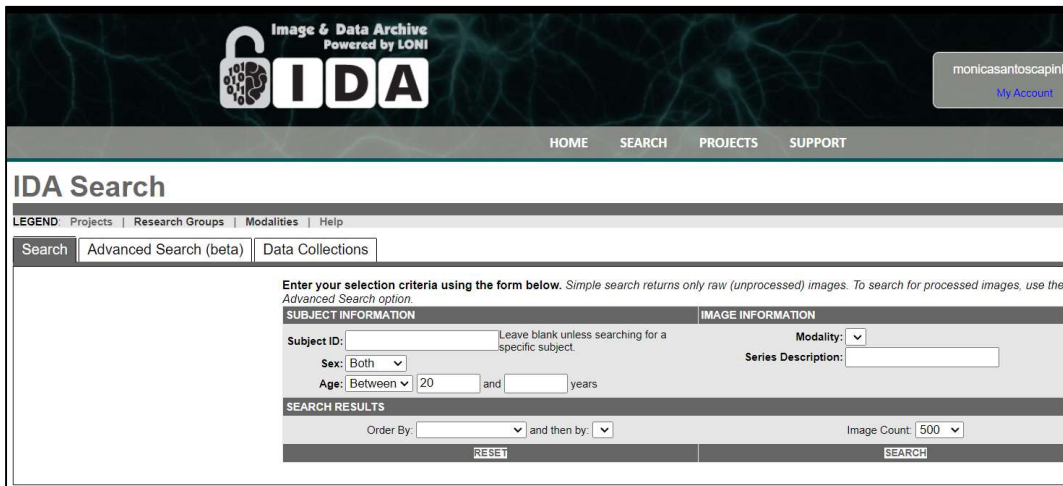


Figura 4.1 – Plataforma *Image and Data Archive*, recolha de dados.

Após fazer a pesquisa e identificação das imagens efetuou-se o *download* das mesmas. Estas imagens ficam armazenadas numa biblioteca virtual da plataforma, *Data Collections* (Figura 4.2), sendo separadas entre casos positivos (utentes com Parkinson) e casos controlo (utentes sem Parkinson).

Subject	Group	Sex	Age	Visit	Modality	Description A	Type	Acq Date	Format	Downloaded	All
3277	Control	M	66	1	MRI	AX FLAIR T2	Original	11/09/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3260	Control	F	63	1	MRI	AX FLAIR T2	Original	6/27/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3270	Control	M	55	1	MRI	AX FLAIR T2	Original	9/16/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3267	Control	F	53	1	MRI	AX FLAIR T2	Original	3/17/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3274	Control	M	81	1	MRI	AX FLAIR T2	Original	11/18/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3600	Control	M	55	1	MRI	AX PD ± T2	Original	4/18/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3611	Control	F	42	1	MRI	AX PD ± T2	Original	6/25/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3620	Control	M	69	1	MRI	AX PD ± T2	Original	11/30/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3613	Control	F	56	1	MRI	AX PD ± T2	Original	6/16/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3636	Control	M	57	1	MRI	AX PD ± T2	Original	2/07/2013	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3616	Control	M	66	1	MRI	AX PD ± T2	Original	7/23/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3611	Control	F	42	1	MRI	AX PD ± T2	Original	6/25/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3627	Control	M	59	1	MRI	AX PD ± T2	Original	6/22/2012	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3613	Control	F	56	1	MRI	AX PD ± T2	Original	6/16/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3600	Control	M	55	1	MRI	AX PD ± T2	Original	4/18/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3637	Control	M	57	1	MRI	AX PD ± T2	Original	2/27/2013	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3614	Control	M	62	1	MRI	AX PD ± T2	Original	6/16/2011	DCM	<input type="checkbox"/>	<input type="checkbox"/>
3624	Control	M	61	1	MRI	AX PD ± T2	Original	2/07/2012	DCM	<input type="checkbox"/>	<input type="checkbox"/>

Figura 4.2 – Biblioteca Virtual IDA

Como já foi mencionado anteriormente no capítulo 2.1 a doença de Parkinson manifesta-se consideravelmente na substância nigra. Nesse sentido, procedeu-se à seleção de cortes transversais na região do mesencéfalo, de forma a garantir sempre a mesma área de estudo, e tendo em consideração a necessidade de incluir toda a área anatómica envolvente, sendo, por isso, o tamanho das imagens variável (Figura 4.3).

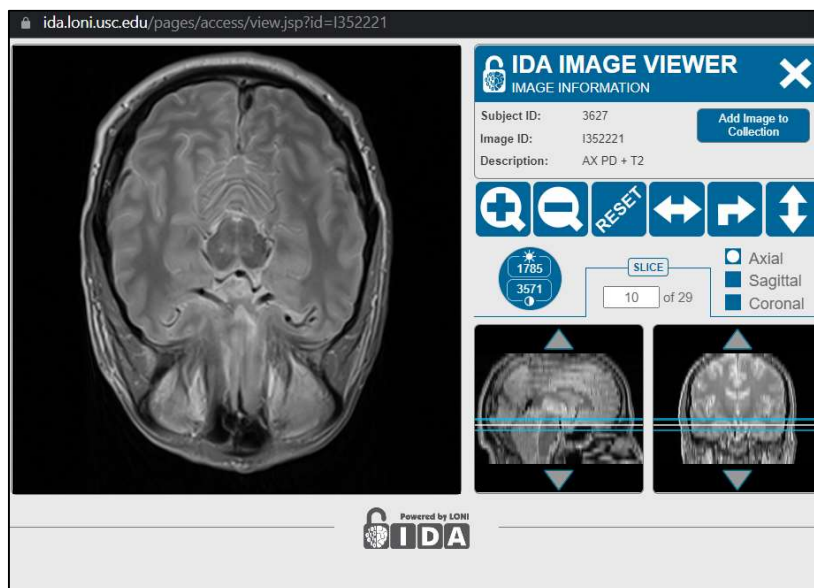


Figura 4.3 – Seleção corte transversal para a base de dados.

Estas imagens foram guardadas em formato PNG, em pastas separadas por ponderações (Ponderação T1, Ponderação T2, Ponderação FLAIR, Ponderação DP) e dentro destas em subpastas correspondentes a Positivos ou Negativos. Durante este processo, tentou-se obter um número próximo de imagens referentes a casos positivos e a casos negativos, de forma a não enviesar as redes neuronais no sentido de uma classe mais prevalente.

Simultaneamente, foi criada uma base de dados num ficheiro Excel onde foram registadas todas as informações relativas a cada imagem, nomeadamente o identificador de utente ('ID'), género, idade e ponderação (DP, FLAIR, T1 e T2).

No total foi criada uma base de dados composta por 1895 imagens, sendo estas subdivididas em:

- Ponderação DP:
 - Negativos: 218 imagens
 - Positivos: 218 imagens
 - Número de utentes: 104 (61 Masculinos; 43 Femininos)
- Ponderação FLAIR:
 - Negativos: 223 imagens
 - Positivos: 230 imagens
 - Número de utentes: 87 (52 Masculinos; 35 Femininos)

- Ponderação T1:
 - Negativos: 238 imagens
 - Positivos: 240 imagens
 - Número de utentes: 89 (56 Masculinos; 33 Femininos)
- Ponderação T2:
 - Negativos: 238 imagens
 - Positivos: 251 imagens
 - Número de utentes: 113 (72 Masculinos; 41 Femininos)

4.2.2. Treino e avaliação dos modelos

Para o treino dos modelos utilizou-se o *software* Weka (Hall et al., 2009). Este *software* é de acesso livre, tendo o seu desenvolvimento sido iniciado em 1993 na Universidade de Waikato, Nova Zelândia. Atualmente é uma das aplicações informáticas mais utilizadas no emprego de modelos de aprendizagem automática. A versão de base do *software* não possui implementações de modelos *Deep Learning*, sendo que essa oferta é possibilitada pela biblioteca WekaDeeplearning4j (Lang et al., 2019). Esta biblioteca, dedicada às redes neuronais profundas, disponibiliza um leque alargado de arquiteturas, nomeadamente a LeNet, AlexNet, ResNet e a InceptionNet.

Neste contexto, utilizou-se o *software* WEKA com a biblioteca WekaDeeplearning4j para avaliar a combinação da arquitetura *Deep Learning* e ponderação de imagem que permite obter melhores resultados na identificação de doença de Parkinson. Para isso procederam-se aos seguintes passos:

- Escolha das arquiteturas a testar: arquiteturas LeNet, AlexNet, InceptionNet e ResNet. Considerando as implementações disponibilizadas pelo *software*, utilizou-se a arquitetura LeNet pré-treinada com a base de dados MNIST, as arquiteturas InceptionNetV3 e ResNet variante 50 (ResNet50) pré-treinadas com a base de dados IMAGENET e a arquitetura AlexNet, sem qualquer tipo de pré-treino.
- A escolha das ponderações: foram testadas as imagens com ponderações T2, T1, DP e FLAIR, cada uma destas individualizadas numa única pasta, estando cada uma destas subdivididas em duas pastas, correspondentes aos casos positivos e negativos.
- Parametrização das arquiteturas: cada arquitetura foi treinada, de forma independente, duas vezes com o mesmo conjunto de dados. Um treino foi

efetuado para um total de 25 épocas e outro a um total de 50 épocas. A utilização destes dois conjuntos de épocas distintas prende-se com a tentativa de identificar qual dos mesmos fornece melhor capacidade preditiva, considerando possíveis sub-ajustamentos (no caso de um modelo ter poucas épocas de treino) ou sobre-ajustamentos (quando o número de épocas é excessivo) aos dados de treino.

- Parametrização dos modelos: foi realizada com um valor de *batch size* de 16. Este valor corresponde ao número total de imagens consideradas pelo modelo em cada época de treino. O valor de 16 permite que cada imagem seja utilizada pelo menos uma vez no processo de treino a 25 épocas ($16 \times 25 = 400$) e duas vezes para 50 épocas (800 imagens consideradas). O valor por defeito de *batch size* do *software* é de 100, mas valores elevados dependem também de uma capacidade computacional elevada, pelo que se considerou a sua redução para um valor menor. É de referir que o valor do parâmetro *batch size* não interfere diretamente na capacidade preditiva do modelo, mas apenas no impacto que cada época de treino tem sobre o ajuste dos pesos na rede. Um *batch size* menor com elevado número de épocas de treino é tecnicamente equivalente a um *batch size* elevado combinado com um número reduzido de épocas de treino. É de referir também que cada arquitetura necessita de imagens com dimensão fixa pré-definida. Este redimensionamento é feito de forma automática pelo *software* Weka, em função dos requisitos de cada arquitetura e correspondeu a uma matriz de 28x28 pixels para a LeNet, 224x224 para as arquiteturas ResNet50 e AlexNet, e 299x299 para a arquitetura InceptionNetV3.

Os restantes parâmetros utilizaram os valores por defeito do *software*, *early stopping* = 0, *attribute normalization* = *standardize training data*, *number of GPUs* = 1, *Seed* = 1, *debug* = *False*, *do not check Capabilities* = *False*.

- Avaliação da performance preditiva: para avaliar a capacidade preditiva de cada arquitetura e número de épocas de treino (25 ou 50) utilizou-se uma validação cruzada. Neste procedimento, 80% das imagens são escolhidas aleatoriamente para treinar o modelo e as restantes 20% são deixadas inteiramente de parte e são comparadas com classificações obtidas do modelo após o treino.

- Medição da capacidade preditiva de cada modelo: para medir a capacidade preditiva de cada modelo (ou seja, a concordância entre as predições obtidas e as efetivamente constantes nos 20% de imagens deixados de fora do processo de treino) utilizou-se a AUC. Esta medida resulta da avaliação conjunta da 'especificidade' e 'sensibilidade'. A especificidade corresponde à proporção de verdadeiros negativos (ou seja, casos em que o modelo classifica a imagem como diagnóstico negativo e o utente é efetivamente negativo), a sensibilidade à proporção de verdadeiros positivos. A AUC é uma métrica que integra estas duas componentes, e para a qual se considera que valores acima de 0.8 são correspondentes a uma 'muito boa' capacidade preditiva e valores acima de 0.9 a uma 'excelente' capacidade preditiva (Šimundić, 2009). Valores iguais ou inferiores a 0.5 correspondem a um modelo sem utilidade preditiva.

No total, foram testadas 32 combinações de arquiteturas, ponderações de imagens e número de épocas de treino. Todos os modelos foram treinados e avaliados num computador com processador AMD Ryzen 7 3700X 8-Core Processador 3.59 GHz com 32GB RAM e uma GPU: NVIDIA GeForce GTX 1660 Super 6GB.

Com base nos valores de AUC, identificou-se a combinação de arquitetura, ponderação de imagem e número de épocas de treino que permitiu alcançar melhor performance preditiva.

4.2.3. Avaliação do efeito do género e idade dos utentes na capacidade preditiva dos modelos

Após se proceder à identificação da combinação de arquitetura e ponderação que alcançam melhor performance preditiva, utilizou-se a mesma na classificação do conjunto total de imagens da respetiva ponderação. Após isso, procedeu-se à criação de uma tabela onde se agruparam, para cada imagem, a existência ou inexistência de erro preditivo (codificado como 1 ou 0, respetivamente), a idade do utente, o seu género ('M' ou 'F') e o seu identificador ('ID').

De seguida, procedeu-se à avaliação da existência de relações significativas entre a obtenção de predições corretas ou incorretas do modelo e as variáveis idade e género. Para avaliar estas relações, utilizaram-se modelos lineares de efeitos mistos. Estes são

modelos que testam a existência de relações lineares significativas entre uma ou mais variáveis explicativas e uma variável resposta. Uma vantagem desta tipologia de modelos relativamente aos modelos lineares generalizados ‘simples’ é que permite considerar a existência de um número desigual de observações para cada unidade de análise, tal como no caso dos utentes que são representados na amostra. Por outras palavras, a base de dados inclui um número variável de imagens por utente e se esta variação não for tida em conta, os utentes com maior número de imagens têm maior ‘peso’ nos resultados do modelo estatístico. Os modelos lineares de efeitos mistos permitem remover este efeito ao realizar as estimativas de significância com base na variação observada entre diferentes grupos de unidades de análise (utentes, neste caso), e não cada observação individual.

No total foram implementados quatro modelos. O primeiro destes avaliou a existência de diferenças significativas no número de classificações incorretas (‘erros’) entre géneros, tendo-se para isso definido como variável dependente a existência ou inexistência de erro e como variável explicativa o género. O segundo modelo avaliou a existência de uma relação entre a idade e a ocorrência ou não de erro, tendo esta última sido definida como variável dependente e a anterior como variável explicativa. Neste modelo não foi efetuada diferenciação entre géneros. Portanto, para considerar um efeito desta diferenciação, testaram-se ainda associações entre idade e erro, considerando isoladamente os resultados obtidos para o género masculino (terceiro modelo) e para o género feminino (quarto modelo). Estes modelos foram implementados em R (v 4.0.5) com recurso à biblioteca glmmTMB (Brooks et al., 2017).

5.Estado da Arte da aplicação do *Deep Learning* em Ressonância Magnética

5.1. Resultados

A pesquisa bibliográfica devolveu um total de 418 referências, das quais 298 (71%) corresponderam a trabalhos enquadrado na temática do trabalho. De entre os trabalhos não enquadrados destacam-se os que não se aplicavam em diagnóstico, mas sim em segmentação/morfótipo, bem como estudos preliminares. O número de trabalhos enquadrados, cresceu de uma forma quase exponencial desde 2015, ou seja, duplicou aproximadamente a cada ano (Figura 5.1). O ano 2015, marca o ano 'zero', em que apenas um único artigo foi publicado, sendo que em 2020 foram já publicados um total de 151. Os dados referentes aos anos 2021 são ainda incompletos, mas há data da pesquisa bibliográfica (10 de fevereiro) já se encontravam publicados 17.

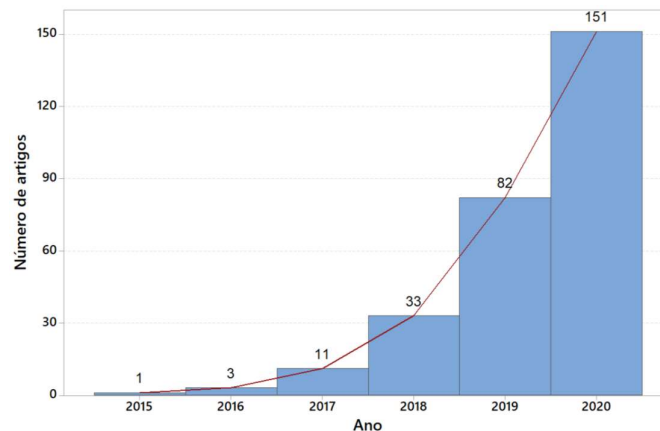


Figura 5.1 – Histograma do número de artigos publicados por ano.

Destaca-se a enorme prevalência de trabalhos de investigação original, sendo que apenas 3 trabalhos se enquadravam na tipologia de 'revisão', dois deles correspondendo a revisões sistemáticas e um a revisão sistemática com meta-análise. Nos trabalhos de investigação original destaca-se a elevada diversidade de patologias consideradas (n=38), abrangendo desde problemas de ordem neurológica, a problemas tumorais e lesões físicas (p. ex. fraturas ósseas) (Figura 5.2). Pese embora esta diversidade, algumas patologias concentram grande parte dos esforços de investigação,

nomeadamente Alzheimer, com um total de 63 trabalhos, seguido dos tumores cerebrais com 59 trabalhos, cancro da próstata com 25, doenças cardíacas com 22 e a deteção de alterações cerebrais com 19.

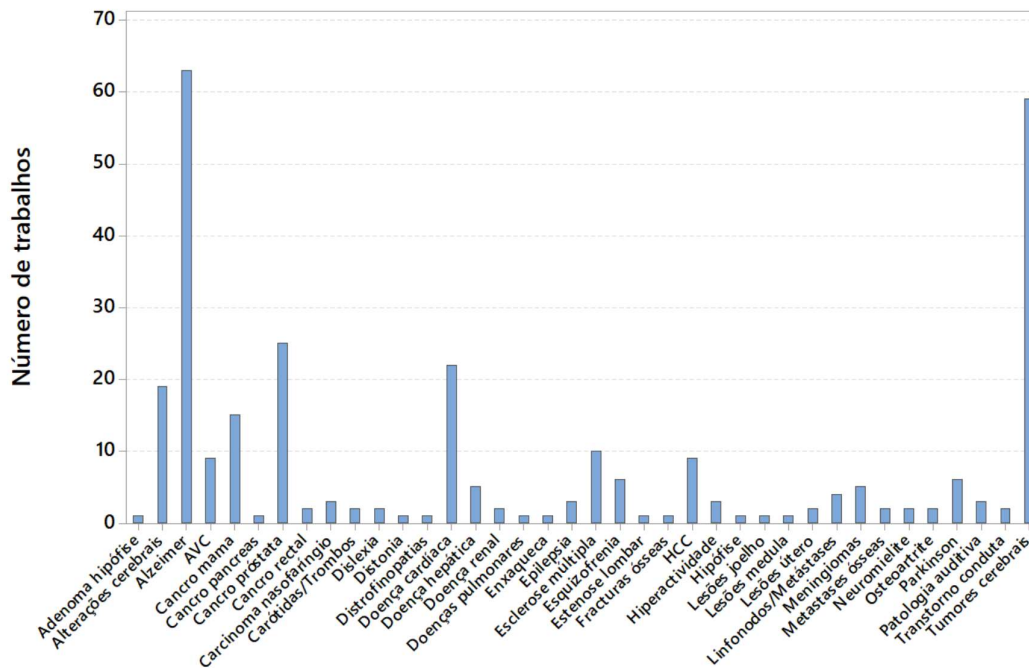


Figura 5.2 – Distribuição do número de trabalhos publicados por patologia.

Dada a grande concentração de esforços de investigação em doenças ou alteração do foro neurológico, não é surpreendente que esta seja a área anatómica mais frequentemente investigada, representando cerca de 65% da totalidade dos trabalhos (189 de um total de 293) (Figura 5.3). Qualquer uma das outras áreas anatómicas é apresentada por menos de 10% da totalidade dos estudos, sendo a mais frequente a pélvis, com 9,9%, e o coração com 7,5%.

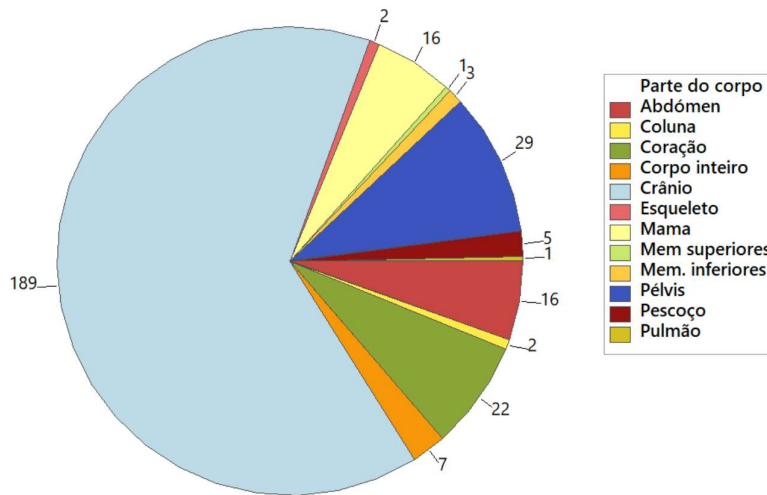


Figura 5.3 – Número de trabalhos por áreas anatómicas investigadas.

A larga maioria dos trabalhos circunscreveu-se apenas à análise de imagens de ressonância (n=280; 95%). No entanto, um número diminuto de trabalhos conjuga estas imagens com outros suportes de diagnóstico, sendo o mais comum a conjugação com PET (n=8), seguido de Ultrassom (n=4), TAC (n=3) e mamografia (n=1). De igual modo, a grande maioria dos trabalhos utiliza redes neuronais profundas ‘clássicas’ ou de tipologia 2D (que consideram apenas uma imagem) (n=239; 80%), enquanto uma porção mais diminuta de trabalhos testam redes capazes de analisar dados 3D (ou seja combinações volumétricas de imagens 2D) (n=61; 20%). Uma quantidade substancial dos trabalhos debruça-se sobre a comparação de desempenhos preditivos entre diferentes arquiteturas de redes neuronais profundas (n=55; 19%), ainda assim a maioria destes visa a aplicação clínica dos modelos sem efetuar este tipo de comparações (n=235; 81%).

Apenas 20 trabalhos efetuaram comparação do desempenho de diagnósticos realizados pelos modelos com diagnósticos realizados por humanos, sendo que 12 (60%) estudos indicam a superioridade da capacidade de diagnóstico dos modelos sobre os humanos, 5 (25%) indicam uma capacidade idêntica e apenas 1 (5%) indica uma capacidade inferior dos modelos (na deteção de fraturas ósseas) (Figura 5.4). É interessante verificar que este tipo de comparações abrange um leque relativamente diversificado de patologias (Figura 5.4), não estando enviesado no sentido das patologias do foro neurológico, como verificado para a totalidade dos estudos (Figura 5.3). Verifica-se ainda que as classes de menor desempenho dos modelos (i.e., igual ou

inferior ao diagnóstico humano), não se encontram concentradas em patologias específicas, mas sim relativamente dispersas por estas. Ainda relativamente à qualidade dos modelos, verifica-se que nos estudos que reportaram uma prestação qualitativa dos desempenhos dos modelos, a larga maioria indica que estes foram 'bem-sucedidos' (n=260; 88%), sendo que 11% (n=32) indicam que o seu desempenho foi moderado ou 'a melhorar' e apenas 1% (n=3) os referem como ineficazes.

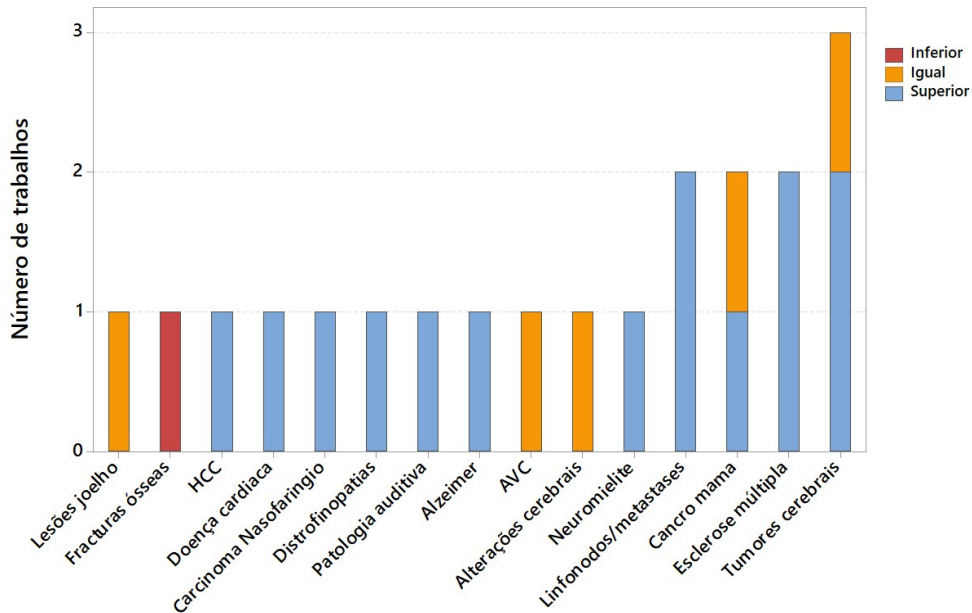


Figura 5.4 – Desempenho do diagnóstico efetuado por redes neuronais profundas versus por humanos, por patologia considerada.

Apesar na grande concentração dos esforços estarem no foro neurológico, verificou-se que apenas 37% está relacionado com as demências, prevalecendo os trabalhos dirigidos à doença de Alzheimer (n=63; 34%). A doença de Parkinson está entre as patologias com menor número de trabalhos (n=6; 3%), tal como as doenças relacionadas com a Hipófise, Meningismos entre outras (Figura 5.5).

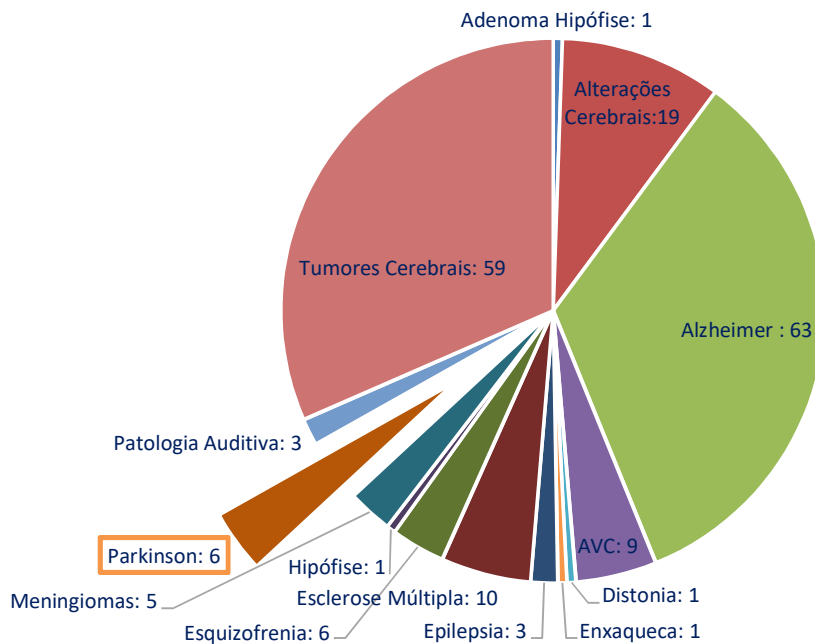


Figura 5.5 – Número de estudos com aplicação de *Deep Learning* no estudo de patologias do foro neurológico.

Nos seis trabalhos aplicados ao estudo da doença de Parkinson (Figura 5.5), o objetivo comum foi o de diagnosticar a doença, tendo sido para isso avaliada a capacidade dos modelos em diferenciar imagens relativas a indivíduos saudáveis de outras relativas a indivíduos portadores da doença. Para isso, os autores dos trabalhos utilizaram maioritariamente imagens com ponderação T2 de Ressonância Magnética (n=3; tabela 5.1) e na aplicação de arquiteturas desenvolvidas pelos próprios, variando nestas, por exemplo, o número das camadas convolucionais utilizadas e na conjugação destas com funções de *pooling* e de ativação. O desempenho preditivo dos modelos apresentou uma elevada variação entre estudos, sendo que a utilização de diferentes métricas de medição de erro não facilita a comparação. Por exemplo, Yin et al., (2020), verificaram que as alterações nas redes por eles propostas permitiram melhorar a capacidade preditiva dos modelos apresentando uma acurácia ('percentagem de predições corretas') de 71.2% em relação à arquitetura AlexNet (68.8%). Por outro lado, Tremblay et al., (2020), para um outro conjunto de imagens, conseguiu alcançar uma acurácia de 88.3%. Sivaranjini e Sujatha (2020), utilizando a mesma arquitetura AlexNet conseguiu um melhor desempenho, apresentando uma acurácia de 88.9%. Já o trabalho de Yasaka et al., (2021), focou-se na testagem de vários parâmetros da difusão tendo obtido para o conjunto de todos os parâmetros testados uma média de AUC (*Area Under*

the Curve) métrica que considera simultaneamente a percentagem de predições corretas referentes a casos de doença e casos sem doença; ver secção '4.2.2 treino e avaliação dos modelos' para mais detalhes) de 0,844. Já o trabalho de Kiryu et al. (2019), utiliza a ponderação em T1 num único corte em sagital da linha média obtendo uma acurácia de 96,8%, equivalendo a uma AUC de 0,995.

Tabela 5.1: Trabalhos científicos que procedem à aplicação de modelos *Deep Learning* no diagnóstico da doença de Parkinson com imagens ressonância magnética

Artigo	Rede Neuronal Profunda	Ponderação em RM / corte	Amostra	Estratégia de avaliação	Métrica e resultados de classificação
Yasaka et al., (2021)	CNN 4 Camadas (8,16,16,32)	Difusão / Axial	115 Saudáveis 115 Doença Parkinson (DP)	<i>Cross validation</i> 5-fold	AUC (média múltiplos parâmetros imagem): 0,844
Yin et al., (2020)	CNN Aprimorada versus AlexNet	T2 / Axial	2544 Imagens saudáveis 1183 Imagens DP	Treino- validação 80%-20%	Acurácia: CNN Aprimorada:71,2% AlexNet:68,8%
Sivaranjini e Sujatha., (2020)	AlexNet pré-treinado	T2 / Axial	82 (3526) Saudáveis 100 (4120) DP	80%-20%	Acurácia: 88,9% Sensibi.: 89,3% Especif.: 88,4% AUC: 0,9618
Nguyen et al., (2020)	CNN	T1 / Axial	4 Anos 42 Saudáveis 116 DP	80%-20% <i>Cross-validation</i> 5-fold	Não fornecido
Tremblay et al., (2020)	CNN 7 Camadas	T2 / Coronal	15 Saudáveis 15 DP 15 DOP-PV	Não fornecido	Acurácia: 88,3% Sensibi.: 88,4% Especif.: 88,2%
Kiryu et al., (2019)	CNN 8 Camadas	T1 / Sagital	142 Saudáveis 125 DP	85%-15%	Acurácia: 96,8% AUC: 0,995

5.2. Análise e Discussão

O presente trabalho apresenta uma breve revisão das tendências e padrões de publicação científica em redes neuronais profundas aplicadas à classificação de patologias em imagem de ressonância magnética. Os resultados obtidos indicam que esta é uma área de investigação muito recente, de crescimento repentino e explosivo, com enfoque numa elevada diversidade de patologias, mas sobretudo nas que têm na sua génese o cérebro. Os diagnósticos fornecidos pelas redes são na larga maioria das vezes considerados como úteis e tendencialmente mais fiáveis que os efetuados por humanos, ainda que com algumas exceções aparentes. Estes resultados estão, pelo menos parcialmente, enviesados pela publicação preferencial de 'resultados positivos', mas ainda assim a relevância destes modelos parece incontestável.

Ainda assim, é relevante a identificação de um número relativamente reduzido de trabalhos a empregar esta tipologia de modelos no diagnóstico da doença de Parkinson. Acresce a isto, o facto de que todos os trabalhos existentes se focarem apenas na avaliação da capacidade de diagnóstico da doença e no aperfeiçoamento das características das redes ou na comparação de arquiteturas diferentes. Por outro lado, nenhum trabalho procede à avaliação do impacto que características extrínsecas aos modelos, tais como a ponderação de imagens em Ressonância Magnética ou atributos dos indivíduos representados (por exemplo, idade ou género), têm sobre a capacidade dos modelos em classificar corretamente a presença ou não da doença.

É importante identificar também algumas limitações na interpretação dos resultados deste trabalho. Dado o carácter resumido que lhe é imposto e limitações de tempo inerentes ao seu contexto académico, diversos aspetos das publicações analisadas foram apenas considerados de forma resumida. Por exemplo, seria potencialmente interessante investigar em maior detalhe as tipologias das redes que são utilizadas e comparadas nos trabalhos. Uma avaliação que poderia ser feita a partir daí, seria se os trabalhos que reportam desempenhos mais reduzidos partilham de alguma característica em comum (p. ex. nas redes que utilizam). Outros aspetos que ficam por avaliar prendem-se com a geografia das publicações, p. ex. as regiões do mundo onde o conhecimento na área está concentrado, e que de alguma forma medeia o acesso futuro a estas tecnologias. Este tipo de análises é possibilitado pelo *Web of Science*, mas dada a inexistência de uma conta académica individual (e não institucional), não é possível efetuar as mesmas após se executar uma filtragem ou classificação dos resultados da pesquisa 'off-line', tal como realizado aqui.

Em conclusão, apesar de se ter verificado que a área mais explorada é a da neurociência, também é importante salientar que a doença de Parkinson é das patologias com menos trabalhos publicados, reforçando assim a necessidade de explorar esta área de aplicação em maior detalhe. A utilização de modelos de redes neuronais profunda na classificação de imagens de ressonância magnética apresenta um elevado potencial de crescimento futuro e de avanços promissores no diagnóstico por imagem. Este potencial aparenta passar não só pela melhoria da performance dos modelos, mas também por compreender quais os fatores que determinam uma maior ou menor capacidade preditiva dos mesmos.

6. Detecção da doença de Parkinson com *Deep Learning*

6.1. Resultados

6.1.1. Avaliação dos modelos

Inicialmente foram avaliadas as capacidades preditivas de cada combinação de arquitetura *Deep Learning*, ponderação de imagem e número de épocas de treino. Com base nos valores de AUC, verificou-se a existência de uma elevada diferença de desempenho preditivo entre diferentes arquiteturas e ponderações de imagem (Tabela 6.1). A arquitetura ResNet50 foi a que alcançou, de forma consistente, os melhores resultados (Figura 6.1), tendo sido também a que devolveu o modelo com a melhor capacidade preditiva (próxima de 'excelente'; AUC = 0,889), correspondendo a um treino de 50 épocas com imagens em ponderação T2. A arquitetura InceptionNetV3 também apresentou performances elevadas, particularmente para T1 e após um treino de 50 épocas, com um valor de AUC de 0,823. A arquitetura LeNet, apresentou também uma performance preditiva inferior à arquitetura ResNet50, mas ainda assim para o caso específico da ponderação T2 atingiu um valor de AUC considerado muito bom (AUC = 0,812). Para as restantes ponderações, a performance preditiva foi inferior, não ultrapassando um valor de AUC de 0,8. No caso da AlexNet, nenhum modelo produziu capacidade preditiva para poder ser considerado útil, tendo em todos os casos produzido classificações de imagens inteiramente de uma única classe (positivos), ou seja, sem capacidade discriminativa.

Tabela 6.1: Valor de AUC para cada combinação de arquitetura, ponderação e número de épocas de treino testadas.

	LeNet		ResNet50		AlexNet		InceptionNetV3	
	25 Épocas	50 Épocas	25 Épocas	50 Épocas	25 Épocas	50 Épocas	25 Épocas	50 Épocas
FLAIR	0.632	0.500	0.797	0.813	0.5	0.5	0.445	0.793
DP	0.665	0.488	0.665	0.678	0.5	0.5	0.42	0.654
T1	0.634	0.417	0.597	0.522	0.5	0.5	0.644	0.823
T2	0.812	0.548	0.886	0.889	0.5	0.5	0.671	0.789

- Sem utilidade preditiva
- Muito boa capacidade preditiva

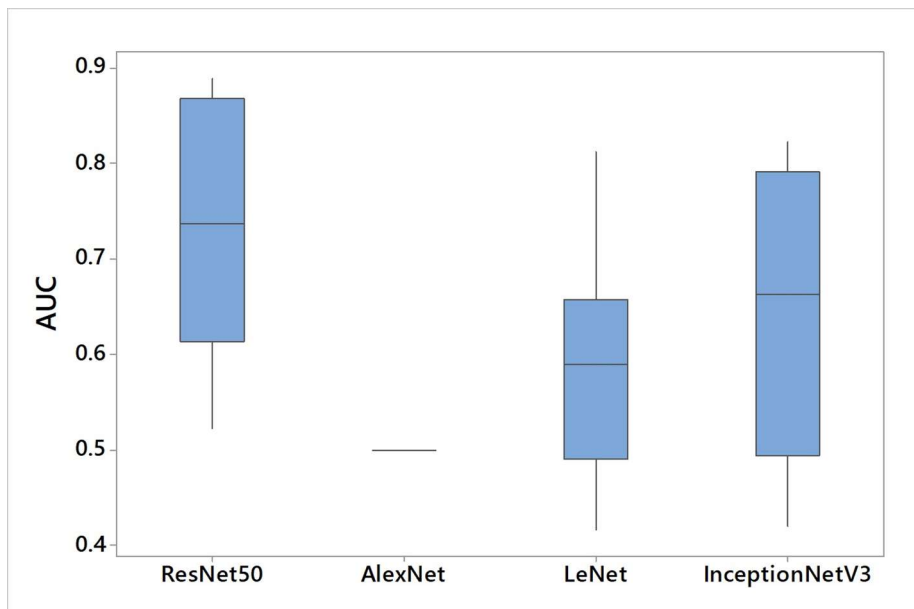


Figura 6.1 – Boxplot de valores de AUC agrupados por arquitetura Deep Learning.

No que diz respeito às ponderações, verificou-se uma superioridade na performance de modelos treinados com imagens de ponderação de T2, seguido da ponderação FLAIR (Figura 6.2). A ponderação DP, foi a que apresentou consistentemente piores desempenhos preditivos, com todos os modelos a apresentarem valores de AUC abaixo de 0.8. A ponderação T1, apresentou para alguns modelos valores ligeiramente melhores, mas ainda assim inferiores aos obtidos com as ponderações FLAIR e T2.

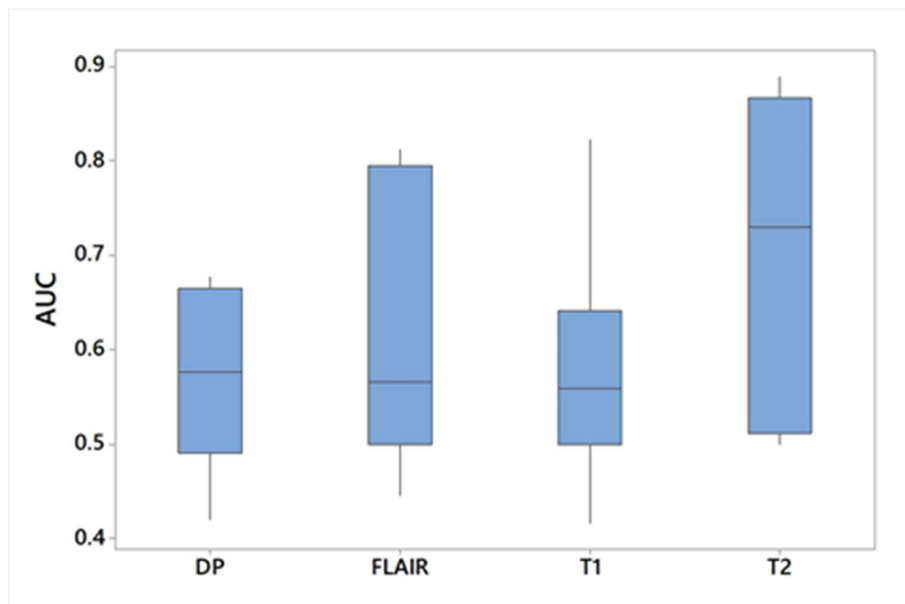


Figura 6.2: *Boxplot* de valores de AUC agrupados por ponderação de imagens.

6.1.2. Efeito do género e idade dos utentes na capacidade preditiva dos modelos

Como esperado, dado o elevado valor de AUC atingido, a larga maioria das predições efetuadas pelo modelo ResNet50 treinado em imagens com ponderação T2 durante 50 épocas são corretas (Figura 6.3). A análise visual da proporção das predições corretas e incorretas entre géneros não evidencia diferenças substanciais entre os mesmos (Figura 6.3).

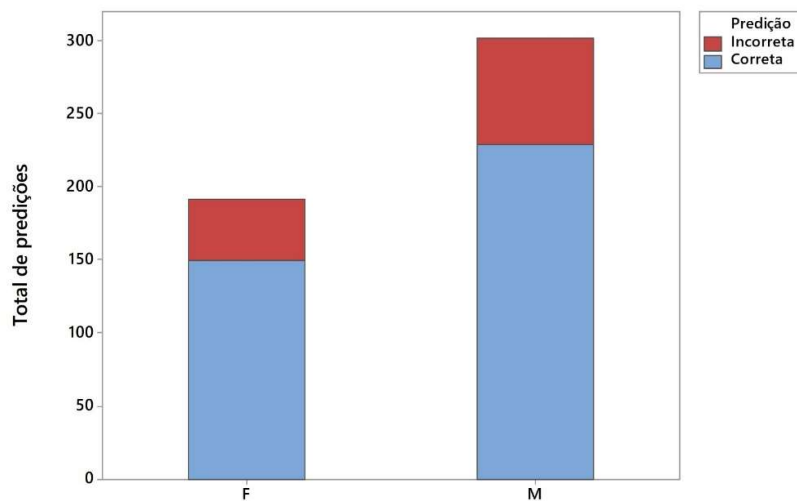


Figura 6.3: Totais de predições corretas e incorretas por género.

No entanto, é aparente a existência de diferenças na distribuição dos erros obtidos entre géneros. Mais especificamente, aparenta existir uma maior prevalência de predições incorretas para o género feminino em idade mais avançada (Figura 6.4, painéis superiores). Contrariamente, à distribuição de erros de predição no género masculino apresenta uma distribuição razoavelmente homogénea ao longo do gradiente de idade (Figura 6.4, painéis inferiores).

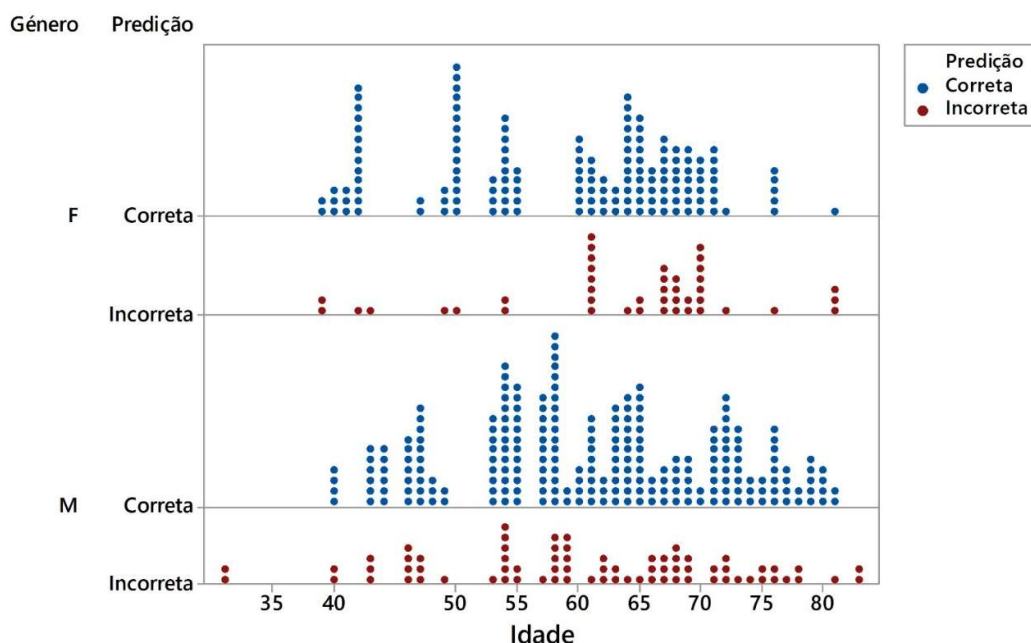


Figura 6.4: Distribuição das previsões corretas e incorretas por gênero e idade.

Os modelos lineares de efeitos mistos aplicados aos erros do ResNet50 com ponderação T2 e 50 épocas de treino permitiram identificar a ausência de diferenças significativas na preponderância de erro entre gêneros ($p=0,907$; tabela 6.2).

Tabela 6.2: Parâmetros do modelo linear de efeitos mistos na avaliação de diferenças de propensão para o erro entre gêneros.

	Coeficiente	Erro padrão	P
Constante	-1,423	0,278	<0,001
Gênero (M)	0,040	0,344	0,907

De igual modo, também não se verificou uma associação significativa entre a idade e uma maior ou menor propensão para erro quando os dois gêneros são considerados em conjunto ($p=0,339$; Tabela 6.3). O mesmo resultado verificou-se quando os utentes do gênero masculino foram considerados isoladamente ($p=0,635$; Tabela 6.4).

Tabela 6.3: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade, considerando ambos os gêneros.

	Coefficiente	Erro padrão	<u>P</u>
Constante	-2,290	0,959	0,01
Idade	0,015	0,956	0,339

Tabela 6.4: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade para o gênero masculino.

	Coefficiente	Erro padrão	<u>P</u>
Constante	-0,807	1,094	0,460
Idade	-0,008	0,018	0,635

No entanto, verificou-se uma associação significativa positiva entre propensão para o erro e idade para o gênero feminino ($p=0,040$; Tabela 6.5), tal como já sugerido anteriormente na figura 6.4. Por outras palavras, este resultado sugere que a taxa de erro tende a aumentar com o avançar da idade para o gênero feminino.

Tabela 6.5: Parâmetros do modelo linear de efeitos mistos na avaliação de uma associação entre propensão para o erro e idade para o gênero feminino.

	Coefficiente	Erro padrão	<u>P</u>
Constante	-5,199	1,872	0,005
Idade	0,061	0,030	0,040

6.2. Análise e Discussão

O presente trabalho teve como principal objetivo identificar as combinações de arquitetura *Deep Learning* e ponderação em ressonância magnética que proporcionam melhor capacidade preditiva na identificação da doença de Parkinson. Complementarmente, procedeu-se também à avaliação de possíveis relações entre esta mesma capacidade e os fatores género e idade dos utentes.

Verificou-se que todas as combinações testadas com a arquitetura AlexNet (AlexNet + FLAIR; AlexNet + DP; AlexNet + T1; AlexNet + T2), independentemente do número de épocas de treino (25 e 50), não originaram modelos com utilidade preditiva, apresentando invariavelmente valores de AUC = 0,5. Este resultado não era o esperado, visto esta arquitetura ter sido utilizada em trabalhos anteriores devolvendo valores de AUC muito elevados (p.ex. 0,9618) (Yin et al., 2020; Sivaranjini e Sujatha, 2020). Esta discrepância de resultados poderá estar relacionada com o número de épocas de treino dos modelos, bem como esta arquitetura no *software Weka* não estar pré treinada com bibliotecas de imagens genéricas (ao contrário, por exemplo, das arquiteturas ResNet50 e InceptionNetV3 que se encontram pré-treinadas com a biblioteca IMAGENET). Portanto, é possível que a arquitetura AlexNet necessite de um maior número de épocas de treino para eventualmente atingir os resultados em linha com os referidos em publicações anteriores. Ainda assim, é de referir que o número de épocas de treino reflete-se de forma linear no tempo de treino dos modelos, e que face ao número de combinações testadas (que perfizeram um total de 32 modelos treinados), aumentar o número de épocas de treino (por. ex. para 100) implicaria um dispêndio de tempo que não seria possível de adequar com trabalho desta natureza. A título de exemplo, uma arquitetura InceptionNetV3 levou em média cerca de 3 horas 30 minutos para completar o treino de 50 épocas. Isto significa que, somente para efetuar o treino deste modelo durante 100 épocas para o total das quatro ponderações distintas seriam necessárias pelo menos de 28 horas de processamento.

A combinação ResNet50 + T2 foi a que apresentou a melhor capacidade preditiva, com um valor de AUC (0,889), no limiar do que é considerado uma excelente capacidade discriminativa ($AUC \geq 0.9$). Este valor foi obtido para um modelo com 50 épocas de treino, no entanto a mesma combinação de arquitetura e ponderação com 25 épocas de treino também se aproximou deste limiar ($AUC=0,886$). Estes resultados reforçam como escolha preferencial a utilização da combinação desta arquitetura (ResNet50) e ponderação (T2) no diagnóstico automatizado da doença de Parkinson. Os resultados para esta combinação indicam uma capacidade de diagnóstico da doença

'média' comparativamente às obtidas em trabalhos anteriores (tabela 5.1), sendo por exemplo ligeiramente superior à obtida por Yasaka et al., (2021), com uma AUC de 0.84, mas inferior às obtidas por Kiryu et al., (2019) e Sivaranjini e Sujatha., (2020), ambos com uma AUC acima de 0,95. A comparação com outros trabalhos anteriores não é direta, uma vez que nestes foi utilizada a acurácia como métrica de performance preditiva, mas ainda assim o desempenho aqui obtido aparenta, pelo menos, ser superior ao obtido por Yin et al., (2020) onde 71,2% de imagens foram corretamente classificadas.

Detaca-se ainda que um conjunto de outras combinações de arquiteturas e ponderações atingiram capacidades preditivas de relevo. Verificou-se que a combinação InceptionNetV3 + T1 apresentou uma capacidade preditiva muito boa com um valor de AUC de 0,823. Este valor foi obtido para um modelo treinado a 50 épocas, no entanto a mesma combinação com um treino de 25 épocas não apresentou um desempenho muito bom, registando um valor de AUC de 0,644. Outra combinação de arquitetura com ponderação FLAIR que apresentou um valor considerável, foi a ResNet50, onde o modelo com 50 épocas de treino apresentou um valor de AUC de 0,813, enquanto com um treino de 25 épocas mostrou um valor de AUC de 0,797. A combinação LeNet + T2 também demonstrou uma boa capacidade preditiva da doença de Parkinson com 25 épocas de treino. Esta combinação apresentou um valor de AUC de 0,812, o que se pode considerar uma muito boa *performance* do modelo, no entanto a mesma combinação com 50 épocas de treino apresentou um desempenho no limiar de utilidade preditiva (AUC = 0,548).

Estes resultados sugerem fortemente que, entre as combinações testadas, a mais adequada na deteção da doença de Parkinson é a ResNet50 + T2. Coincidentemente, a ponderação T2 é a que apresenta um maior número de estudos para a avaliação da doença (Yin et al., 2020; Sivaranjini e Sujatha, 2020; Tremblay et al., 2020). O desempenho mais elevado registado para esta ponderação, tal como a sua utilização recorrente em trabalhos anteriores pode ser justificada por, na ponderação T2, o sinal de ressonância magnética diminui a sua intensidade em relação à distribuição normal do ferro cerebral (Drayer, B., 1986). Sendo esta diminuição de sinal resultado do encurtamento de relaxação do T2 das moléculas de água e conseqüentemente uma queda do sinal dos tecidos afetados (Kosta et al., 2006)

Foi também efetuada uma análise entre a existência ou não de erros na classificação de imagens e fatores relacionados com os utentes, nomeadamente o género e a idade. Os resultados desta análise permitiram verificar que não existem diferenças significativas na proporção de erros entre géneros. No entanto, foi possível

verificar uma relação significativa positiva entre erro e idade para o género feminino ($p=0,040$), ou seja, a taxa de erro aumenta com o avançar da idade para este género. Esta relação, no entanto, não se verificou para o género masculino ($p=0,635$).

A diferença nos resultados da relação entre erro e idade para os dois géneros, poderá estar relacionada com concentração de estrogénio no género feminino. O estrogénio é uma hormona que está associada à ovulação e às características femininas do género, no entanto funciona também como um neuroprotetor (Green et al., 1997). A diminuição acentuada de estrogénio leva à menopausa originando um declínio cognitivo a partir dos 45 anos de idade, evidenciando-se sobretudo a partir dos 65 anos (Davey, D. A., 2013). Estas alterações de estrogénio desencadeiam alterações no sistema nervoso central, entre as quais um aumento da incidência da doença de Alzheimer (Rahman et al., 2020), caracterizando-se esta pela formação de placas senis na região do hipocampo (Davey, D. A., 2013). Coincidentemente, verificou-se aqui que a incidência dos erros para este género é mais frequente em idades mais avançadas, enquanto em idades jovens esta frequência é significativamente mais reduzida. Este resultado, portanto, sugere que a neuroprotecção do estrogénio enquanto mediador de alterações do sistema nervoso central, influenciam a capacidade do modelo em identificar corretamente a doença. Por outras palavras, a quebra acentuada da concentração de estrogénio em pessoas em idade avançada do género feminino leva a alterações diversas no sistema nervoso central, que por sua vez induz uma acentuação do erro por parte do modelo. Por outro lado, no género masculino a ocorrência de erros do modelo é homogénea ao longo da idade, o que também reforça o papel plausível do mecanismo também identificado, uma vez que as concentrações de estrogénio neste género não sofrem variações substanciais em função da idade.

Pese embora o leque alargado de arquiteturas e ponderações testadas, algumas limitações impediram uma análise mais exaustiva. Nomeadamente, a falta de recursos computacionais com maior capacidade e o tempo disponível não permitiram explorar um maior número de combinações de arquiteturas e ponderações, assim como de épocas de treino. De igual forma, após esta identificação de uma combinação com desempenho preditivo elevado, seria potencialmente interessante criar uma base de dados de treino mais alargada, por exemplo com um maior número de imagens e recolhidas de múltiplas bases de dados, de forma a atingir um modelo operacional de utilização clínica.

Posto isto, este trabalho, permitiu levantar novas questões para trabalhos futuros. Nomeadamente, o efeito que o estrogénio, seja endógeno ou via medicamentosa, tem na incapacidade de diagnosticar a doença precocemente. Será interessante avaliar em

trabalhos futuros se este efeito também se verifica em métodos de diagnóstico não automatizado, ou seja, em diagnóstico clínico. Uma primeira abordagem possível para a avaliação explícita deste efeito em algoritmos *Deep Learning*, poderá ser com recurso à criação de bases de treino em que informação acerca da frequência e duração de tomas de medicação combinada de estrogénio esteja disponível.

7. Conclusão

A doença de Parkinson é um distúrbio neurológico que afeta milhões de pessoas pelo mundo. Para diminuir os seus efeitos é importante ter um diagnóstico precoce. Neste contexto, o presente trabalho forneceu uma avaliação exaustiva da combinação de arquiteturas *Deep Learning* e ponderação de imagem no reconhecimento automatizado da doença. A aplicação destes procedimentos automatizados permite auxiliar os especialistas médicos no diagnóstico e detecção precoce.

Para melhor fundamentar as opções metodológicas a testar, foi efetuado inicialmente uma revisão de literatura acerca da aplicação de redes neuronais profundas em diagnóstico com base em imagem médica, incluindo na identificação da doença de Parkinson com recurso a imagens de ressonância magnética. A pesquisa efetuada permitiu verificar que esta é uma área de foco crescente, tendo o número de trabalhos publicados crescido fortemente de ano para ano, desde 2015. No entanto verificou-se que, de forma algo surpreendente, a doença de Parkinson é das demências que menor número de estudos apresenta.

Acresce a isto, o fato de os estudos existentes se focarem maioritariamente em avaliar ou comparar o desempenho de arquiteturas *Deep Learning* pré-existentes ou de redes desenvolvidas pelos próprios autores. Por outro lado, verificou-se uma quase completa ausência de esforços na avaliação da combinação de arquitetura e ponderação que com melhor capacidade de identificação da doença de Parkinson, assim como na análise do efeito de fatores relacionados com os utentes nesta mesma capacidade.

Nesse sentido, este trabalho, testou várias combinações de arquiteturas e ponderações de ressonância magnética, onde se verificou que a combinação ResNet50+T2 foi aquela que apresentou o melhor valor de AUC. Esta combinação alcançou um desempenho preditivo no limiar do que é considerado uma excelente capacidade discriminativa. Verificou-se também que a ponderação T2 apresentou melhores resultados para a maioria das arquiteturas testadas, podendo este resultado estar relacionada com a perda de sinal desta ponderação quando existe fixação de ferro na substancia nigra causado pela doença de Parkinson.

Analisou-se ainda uma possível relação entre a ocorrência de erros nas predições do modelo e o género e a idade dos utentes. Estas análises permitiram verificar que não existe uma relação significativa da existência de erro entre os géneros. Ou seja, nenhum género aparenta possuir uma maior apetência à ocorrência de erros. No entanto, ao

efetuar uma análise isolada para cada género, verificou-se uma relação significativa entre erro e idade para o género feminino, relação essa que não se verificou para o género masculino. Por outras palavras, para pessoas do género feminino o erro está concentrado nas idades mais avançadas, ao passo que para o género masculino o erro se encontra distribuído de forma razoavelmente homogénea ao longo do gradiente de idade. Considerando esta diferença, surge a hipótese de um efeito causal para o estrogénio no género feminino. Esta hormona tem um efeito neuroprotetor que se perde com o início da menopausa, levando a alterações acentuadas no sistema nervoso central em idades mais avançadas no género feminino, e desta forma potenciando a ocorrência de erros de diagnóstico por parte do modelo. Trabalhos futuros poderão avaliar esta hipótese.

Como conclusão geral, refere-se que se reconheceu uma capacidade efetiva destes modelos na identificação da doença de Parkinson em imagens de ressonância magnética, mas também que esta capacidade é fortemente mediada pela conjugação de arquitetura e ponderação de imagem utilizadas. Adicionalmente, também se verificou que fatores externos, ou seja, relativamente aos utentes, podem condicionar de forma significativa a capacidade preditiva dos modelos. De forma a avançar nesta área de investigação, trabalhos futuros devem, portanto, considerar não apenas as questões relacionadas com a implementação e desenvolvimento de arquiteturas *Deep Learning* – onde a maior parte do foco tem estado – mas também na exploração dos efeitos que fatores como o género, idade ou historial clínico dos utentes têm sobre a sua capacidade preditiva.

Bibliografia

- A Inteligência Artificial irá substituir os radiologistas? (2017, Dezembro 19). *NeuralMind Inteligência Artificial*. <https://neuralmind.ai/blog/a-inteligencia-artificial-ira-substituir-os-radiologistas/>
- Al-Radaideh, A. M., & Rababah, E. M. (2016). The role of magnetic resonance imaging in the diagnosis of Parkinson's disease: A review. *Clinical Imaging*, 40(5), 987–996. <https://doi.org/10.1016/j.clinimag.2016.05.006>
- Barbosa, E. R., & Sallem, F. A. S. (2005). *Doença de Parkinson – Diagnóstico*. 8.
- Bengio, Y., LeCun, I., & Hinton, G. (2020). Deep Learning for AI. *Communications of the ACM*, 64(7), 58–65.
- Billones, C., Demetria, O., Hostallero, D. E., & Naval, P. (2016). *DemNet: A Convolutional Neural Network for the detection of Alzheimer's Disease and Mild Cognitive Impairment*. 3724–3727. <https://doi.org/10.1109/TENCON.2016.7848755>
- Bishop, C. M., & Bishop, P. of N. C. C. M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press.
- Brooks, M., E., Kristensen, K., Benthem, K., J., van, Magnusson, A., Berg, C., W., Nielsen, A., Skaug, H., J., Mächler, M., & Bolker, B., M. (2017). GlimmTMB Balances Speed and Flexibility Among Packages for Zero-inflated Generalized Linear Mixed Modeling. *The R Journal*, 9(2), 378. <https://doi.org/10.32614/RJ-2017-066>
- Cabreira, V., & Massano, J. (2019). Parkinson's Disease: Clinical Review and Update. *Acta Médica Portuguesa*, 32(10), 661–670. <https://doi.org/10.20344/amp.11978>
- Carrete, H. (2017). A doença de Parkinson e os parkinsonismos atípicos: A importância da ressonância magnética como potencial biomarcador. *Radiologia Brasileira*, 50, 5–6. <https://doi.org/10.1590/0100-3984.2017.50.4e1>
- Cassidy, C. M., Zucca, F. A., Girgis, R. R., Baker, S. C., Weinstein, J. J., Sharp, M. E., Bellei, C., Valmadre, A., Vanegas, N., Kegeles, L. S., Brucato, G., Kang, U. J., Sulzer, D., Zecca, L., Abi-Dargham, A., & Horga, G. (2019). Neuromelanin-sensitive MRI as a noninvasive proxy measure of dopamine function in the human brain. *Proceedings of the National Academy of Sciences*, 116(11), 5108–5117. <https://doi.org/10.1073/pnas.1807983116>

- Cole, J. H., & Franke, K. (2017). Predicting Age Using Neuroimaging: Innovative Brain Ageing Biomarkers. *Trends in Neurosciences*, 40(12), 681–690. <https://doi.org/10.1016/j.tins.2017.10.001>
- Crawford, K. L., Neu, S. C., & Toga, A. W. (2016). The image and data archive at the laboratory of neuro imaging. *Neuroimage*, 124, 1080-1083.
- Doença de Parkinson—Distúrbios neurológicos.* (sem data). Manuais MSD edição para profissionais. Obtido 25 de Junho de 2021, de <https://www.msmanuals.com/pt-pt/profissional/dist%C3%BArbios-neurol%C3%B3gicos/transtornos-de-movimento-e-cerebelares/doen%C3%A7a-de-parkinson>
- Davey, D. A. (2013). Alzheimer’s disease, dementia, mild cognitive impairment and the menopause: A «window of opportunity»? *Women’s Health* (London, England), 9(3), 279–290. <https://doi.org/10.2217/whe.13.22>
- Drayer, B., Burger, P., Darwin, R., Riederer, S., Herfkens, R., & Johnson, G. A. (1986). Magnetic Resonance Imaging of Brain Iron. *American Journal of Neuroradiology*, 7(3), 373–380.
- Fei-Fei, L., Deng, J., & Li, K. (2009). ImageNet: Constructing a large-scale image database. *Journal of vision*, 9(8), 1037-1037.
- Ferreira, J. J., Gonçalves, N., Valadas, A., Januário, C., Silva, M. R., Nogueira, L., Vieira, J. L. M., & Lima, A. B. (2017). Prevalence of Parkinson’s disease: A population-based study in Portugal. *European Journal of Neurology*, 24(5), 748–750. <https://doi.org/10.1111/ene.13273>
- Gao, W., & Zhou, Z.-H. (2016). Dropout Rademacher complexity of deep neural networks. *Science China Information Sciences*, 59(7), 072104. <https://doi.org/10.1007/s11432-015-5470-z>
- Goedert, M., & Compston, A. (2018). Parkinson’s disease—The story of an eponym. *Nature Reviews. Neurology*, 14(1), 57–62. <https://doi.org/10.1038/nrneurol.2017.165>
- Green, P. S., Bishop, J., & Simpkins, J. W. (1997). 17 β -Estradiol Exerts Neuroprotective Effects on SK-N-SH Cells. 5.
- Gur, R. C., Turetsky, B. I., Matsui, M., Yan, M., Bilker, W., Hughett, P., & Gur, R. E. (1999). Sex Differences in Brain Gray and White Matter in Healthy Young Adults: Correlations with Cognitive Performance. *Journal of Neuroscience*, 19(10), 4065–4072. <https://doi.org/10.1523/JNEUROSCI.19-10-04065.1999>
- Hahn, E. L. (1950). Spin Echoes. *Physical Review*, 80(4), 580–594. <https://doi.org/10.1103/PhysRev.80.580>

- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Deep Residual Learning for Image Recognition (arXiv:1512.03385). arXiv. <http://arxiv.org/abs/1512.03385>
- Jung, B. A., & Weigel, M. (2013). Spin echo magnetic resonance imaging. *Journal of Magnetic Resonance Imaging*, 37(4), 805–817. <https://doi.org/10.1002/jmri.24068>
- Kalia, L. V., & Lang, A. E. (2015). Parkinson's disease. *Lancet (London, England)*, 386(9996), 896–912. [https://doi.org/10.1016/S0140-6736\(14\)61393-3](https://doi.org/10.1016/S0140-6736(14)61393-3)
- Kosta, P., Argyropoulou, M. I., Markoula, S., & Konitsiotis, S. (2006). MRI evaluation of the basal ganglia size and iron content in patients with Parkinson's disease. *Journal of Neurology*, 253(1), 26–32. <https://doi.org/10.1007/s00415-005-0914-9>
- Kiryu, S., Yasaka, K., Akai, H., Nakata, Y., Sugomori, Y., Hara, S., Seo, M., Abe, O., & Ohtomo, K. (2019). Deep learning to differentiate parkinsonian disorders separately using single midsagittal MR imaging: A proof of concept study. *European Radiology*, 29(12), 6891–6899. <https://doi.org/10.1007/s00330-019-06327-0>
- Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25.
- Lang, S., Bravo-Marquez, F., Beckham, C., Hall, M., & Frank, E. (2019). WekaDeeplearning4j: A deep learning package for Weka based on Deeplearning4j. *Knowledge-Based Systems*, 178, 48–50. <https://doi.org/10.1016/j.knosys.2019.04.013>
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature*, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>
- LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11), 2278-2324.
- Li, R., Zhang, W., Suk, H.-I., Wang, L., Li, J., Shen, D., & Ji, S. (2014). Deep Learning Based Imaging Data Completion for Improved Brain Disease Diagnosis. *Medical image computing and computer-assisted intervention: MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention*, 17(0 3), 305–312.

- MAZZOLA, Alessandro A. Ressonância magnética: princípios de formação da imagem e aplicações em imagem funcional. *Revista Brasileira de Física Médica*. 2009; 3(1): 117-29
- MAZZOLA, A. A. (2015). *Princípios Físicos da Ressonância Magnética*. Pg, 2(3), 4.
- Martins, P. V. L. (sem data). *Gene prediction using Deep Learning*. 93.
- Masoom Shah, P., Zeb, A., Shafi, U., Zaidi, S. F., & Shah, M. (2018). *Detection of Parkinson Disease in Brain MRI using Convolutional Neural Network*. 1–6. <https://doi.org/10.23919/IConAC.2018.8749023>
- Meijer, F. J. A., Aerts, M. B., Abdo, W. F., Prokop, M., Borm, G. F., Esselink, R. A. J., Goraj, B., & Bloem, B. R. (2012). Contribution of routine brain MRI to the differential diagnosis of parkinsonism: A 3-year prospective follow-up study. *Journal of Neurology*, 259(5), 929–935. <https://doi.org/10.1007/s00415-011-6280-x>
- Meijer, F. J. A., Steens, S. C., van Rumund, A., van Cappellen van Walsum, A.-M., Küsters, B., Esselink, R. A. J., Verbeek, M. M., Bloem, B. R., & Goraj, B. (2016). Nigrosome-1 on Susceptibility Weighted Imaging to Differentiate Parkinson's Disease From Atypical Parkinsonism: An In Vivo and Ex Vivo Pilot Study. *Polish Journal of Radiology*, 81, 363–369. <https://doi.org/10.12659/PJR.897090>
- Mozhdehfarahbakhsh, A., Chitsazian, S., Chakrabarti, P., Chakrabarti, T., Kateb, B., & Nami, M. (2021). An MRI-based Deep Learning Model to Predict Parkinson's Disease Stages. *MedRxiv*, 2021.02.19.21252081. <https://doi.org/10.1101/2021.02.19.21252081>
- Ng, A. (2018). *Machine Learning Yearning*. <https://itbook.store/books/1001590486081>
- Nielsen, M. A. (2015). *Neural Networks and Deep Learning*. <http://neuralnetworksanddeeplearning.com>
- Nguyen, A. A., Maia, P. D., Gao, X., Damasceno, P. F., & Raj, A. (2020). Dynamical Role of Pivotal Brain Regions in Parkinson Symptomatology Uncovered with Deep Learning. *Brain Sciences*, 10(2), 73. <https://doi.org/10.3390/brainsci10020073>
- Oliveira, R. V. de, & Pereira, J. S. (2017). The role of diffusion magnetic resonance imaging in Parkinson's disease and in the differential diagnosis with atypical parkinsonism. *Radiologia Brasileira*, 50(4), 250–257. <https://doi.org/10.1590/0100-3984.2016-0073>
- Osório, F. (sem data). *Controle Inteligente de Veículos Autônomos: Automatização do Processo de Estacionamento de Carros*. 12.

- Parent, A. (2018). A Tribute to James Parkinson. *The Canadian Journal of Neurological Sciences. Le Journal Canadien Des Sciences Neurologiques*, 45(1), 83–89. <https://doi.org/10.1017/cjn.2017.270>
- Provost, J.-S., Hanganu, A., & Monchi, O. (2015). Neuroimaging studies of the striatum in cognition Part I: Healthy individuals. *Frontiers in Systems Neuroscience*, 9. <https://doi.org/10.3389/fnsys.2015.00140>
- Rahman, A., Schelbaum, E., Hoffman, K., Diaz, I., Hristov, H., Andrews, R., Jett, S., Jackson, H., Lee, A., Sarva, H., Pahlajani, S., Matthews, D., Dyke, J., Leon, M. J. de, Isaacson, R. S., Brinton, R. D., & Mosconi, L. (2020). Sex-driven modifiers of Alzheimer risk: A multimodality brain imaging study. *Neurology*, 95(2), e166–e178. <https://doi.org/10.1212/WNL.00000000000009781>
- Rios, E. (1998). Ressonância Magnética Nuclear. 25.
- Roberts, M., Driggs, D., Thorpe, M., Gilbey, J., Yeung, M., Ursprung, S., Aviles-Rivero, A. I., Etmann, C., McCague, C., Beer, L., Weir-McCall, J. R., Teng, Z., Gkrania-Klotsas, E., Rudd, J. H. F., Sala, E., & Schönlieb, C.-B. (2021). Common pitfalls and recommendations for using machine learning to detect and prognosticate for COVID-19 using chest radiographs and CT scans. *Nature Machine Intelligence*, 3(3), 199–217. <https://doi.org/10.1038/s42256-021-00307-0>
- Schwarz, S. T., Afzal, M., Morgan, P. S., Bajaj, N., Gowland, P. A., & Auer, D. P. (2014). The ‘Swallow Tail’ Appearance of the Healthy Nigrosome – A New Accurate Test of Parkinson’s Disease: A Case-Control and Retrospective Cross-Sectional MRI Study at 3T. *PLoS ONE*, 9(4), e93814. <https://doi.org/10.1371/journal.pone.0093814>
- Schwarz, S. T., Rittman, T., Gontu, V., Morgan, P. S., Bajaj, N., & Auer, D. P. (2011). T1-weighted MRI shows stage-dependent substantia nigra signal loss in Parkinson’s disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 26(9), 1633–1638. <https://doi.org/10.1002/mds.23722>
- Shrimanker, I., Tadi, P., & Sánchez-Manso, J. C. (2021). Parkinsonism. Em *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK542224/>
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., & Rabinovich, A. (2014). Going Deeper with Convolutions (arXiv:1409.4842; Versão 1). arXiv. <http://arxiv.org/abs/1409.4842>
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2015). Rethinking the Inception Architecture for Computer Vision (arXiv:1512.00567). arXiv. <https://doi.org/10.48550/arXiv.1512.00567>

- Simonyan, K., & Zisserman, A. (2015). Very Deep Convolutional Networks for Large-Scale Image Recognition (arXiv:1409.1556). arXiv. <http://arxiv.org/abs/1409.1556>
- Šimundić, A. M. (2009). Measures of diagnostic accuracy: basic definitions. *eJIFCC*, 19(4), 203.
- Sivaranjini, S., & Sujatha, C. M. (2020). Deep learning based diagnosis of Parkinson's disease using convolutional neural network. *Multimedia Tools and Applications*, 79(21), 15467–15479. <https://doi.org/10.1007/s11042-019-7469-8>
- Sonne, J., Reddy, V., & Beato, M. R. (2020). Neuroanatomy, Substantia Nigra. *Em StatPearls [Internet]*. StatPearls Publishing. <https://www.ncbi.nlm.nih.gov/books/NBK536995/>
- Thakur, R. (2020). Step by step VGG16 implementation in Keras for beginners. Medium. <https://towardsdatascience.com/step-by-step-vgg16-implementation-in-keras-for-beginners-a833c686ae6c>
- Tolosa, E., Wenning, G., & Poewe, W. (2006). The diagnosis of Parkinson's disease. *The Lancet Neurology*, 5(1), 75–86. [https://doi.org/10.1016/S1474-4422\(05\)70285-4](https://doi.org/10.1016/S1474-4422(05)70285-4)
- Tremblay, C., Mei, J., & Frasnelli, J. (2020). Olfactory bulb surroundings can help to distinguish Parkinson's disease from non-parkinsonian olfactory dysfunction. *NeuroImage : Clinical*, 28, 102457. <https://doi.org/10.1016/j.nicl.2020.102457>
- Trujillo, P., Summers, P. E., Ferrari, E., Zucca, F. A., Sturini, M., Mainardi, L. T., Cerutti, S., Smith, A. K., Smith, S. A., Zecca, L., & Costa, A. (2017). Contrast mechanisms associated with neuromelanin-MRI. *Magnetic Resonance in Medicine*, 78(5), 1790–1800. <https://doi.org/10.1002/mrm.26584>
- Tsang, S.-H. (2019, março 23). Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015. Medium. <https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c>
- Twelves, D., Perkins, K. S. M., & Counsell, C. (2003). Systematic review of incidence studies of Parkinson's disease. *Movement Disorders: Official Journal of the Movement Disorder Society*, 18(1), 19–31. <https://doi.org/10.1002/mds.10305>
- Tysnes, O.-B., & Storstein, A. (2017). Epidemiology of Parkinson's disease. *Journal of Neural Transmission*, 124(8), 901–905. <https://doi.org/10.1007/s00702-017-1686-y>
- Urm. (2017). História da Ressonância Magnética. *URM Ressonância Magnética*. Obtido 18 de Janeiro de 2022, de <https://www.urngv.com.br/historia-da-ressonancia-magnetica/>

- VGGNet-16 Architecture: A Complete Guide. (2020). Obtido 21 de julho de 2022, de <https://kaggle.com/code/blurredmachine/vggnet-16-architecture-a-complete-guide>
- Wang, J., Knol, M. J., Tiulpin, A., Dubost, F., Bruijne, M. de, Vernooij, M. W., Adams, H. H. H., Ikram, M. A., Niessen, W. J., & Roshchupkin, G. V. (2019). Gray Matter Age Prediction as a Biomarker for Risk of Dementia. *Proceedings of the National Academy of Sciences*, 116(42), 21213–21218. <https://doi.org/10.1073/pnas.1902376116>
- Watson, W. (1920). *General Physics*. Longmans, Green and Company. <https://books.google.pt/books?id=\ Zq3AAAAMAAJ>
- Yasaka, K., Kamagata, K., Ogawa, T., Hatano, T., Takeshige-Amano, H., Ogaki, K., Andica, C., Akai, H., Kunitatsu, A., Uchida, W., Hattori, N., Aoki, S., & Abe, O. (2021). Parkinson's disease: Deep learning with a parameter-weighted structural connectome matrix for diagnosis and neural circuit disorder investigation. *Neuroradiology*, 63(9), 1451–1462. <https://doi.org/10.1007/s00234-021-02648-4>
- Yin, D., Zhao, Y., Wang, Y., Zhao, W., & Hu, X. (2020). Auxiliary diagnosis of heterogeneous data of Parkinson's disease based on improved convolution neural network. *Multimedia Tools and Applications*, 79(33), 24199–24224. <https://doi.org/10.1007/s11042-020-08984-6>
- Young Onset Parkinson's. (sem data). Parkinson's Foundation. Obtido 25 de Junho de 2021, de <https://www.parkinson.org/Understanding-Parkinsons/What-is-Parkinsons/Young-Onset-Parkinsons>
- Zecca, L., Stroppolo, A., Gatti, A., Tampellini, D., Toscani, M., Gallorini, M., Giaveri, G., Arosio, P., Santambrogio, P., Fariello, R. G., Karatekin, E., Kleinman, M. H., Turro, N., Hornykiewicz, O., & Zucca, F. A. (2004). The role of iron and copper molecules in the neuronal vulnerability of locus coeruleus and substantia nigra during aging. *Proceedings of the National Academy of Sciences of the United States of America*, 101(26), 9843–9848. <https://doi.org/10.1073/pnas.0403495101>
- Shrimanker, I., Tadi, P., & Sánchez-Manso, J. C. (2021). Parkinsonism. Em *StatPearls*. StatPearls Publishing. <http://www.ncbi.nlm.nih.gov/books/NBK542224/>