Faculdade de Engenharia da Universidade do Porto



Aprendizagem e Avaliação Computacional da Doença de Parkinson a partir de Parâmetros da Marcha

Marta Isabel Azevedo Sol Neves Ferreira

VERSÃO FINAL

Dissertação realizada no âmbito do
Mestrado em Engenharia Biomédica

Orientador: Prof. Doutor João Manuel R. S. Tavares
Faculdade de Engenharia da Universidade do Porto

Co-orientador: Prof. Doutor Fábio Augusto Barbieri
Universidade Estadual Paulista, Brasil

junho de 2021

© Marta Ferreira, 2021

Resumo

A doença de Parkinson (DP) é uma doença neurodegenerativa crónica e progressiva caraterizada por uma diminuição do número de neurónios dopaminérgicos na parte compacta da substância nigra. Nesta doença são manifestados, primeiramente, sintomas não-motores e à medida que progride começam a aparecer os sintomas motores. Esta doença não tem cura e o diagnóstico precoce constitui um grande desafio. Na verdade, apenas é possível fazer um diagnóstico, quando surgem os sintomas motores, numa fase equivalente a uma perda de cerca de 60% dos neurónios dopaminérgicos. A avaliação destes sintomas baseia-se, essencialmente, em testes clínicos subjetivos, o que pode contribuir para a existência de diagnósticos inconsistentes.

Deste modo, nesta Dissertação foi proposta uma metodologia que compreende o desenvolvimento e implementação de dois modelos de *Machine Learning*, um para a classificação automática da DP e outro para a classificação automática da severidade desta doença. A base de dados utilizada inclui 126 participantes recrutados na região de Bauru (São Paulo, Brasil), dos quais 63 eram doentes de Parkinson e encontravam-se na fase ligeira ou intermédia da doença. O valor de exatidão mais alto que o primeiro modelo obteve foi 84,6%, através do algoritmo *naive Bayes*. Os parâmetros da marcha identificados como fundamentais para distinguir indivíduos saudáveis dos doentes de Parkinson foram: comprimento do passo, velocidade do passo, largura do passo e variabilidade da largura do passo. Em relação ao segundo modelo, o valor AUC mais elevado foi 0,786, alcançado pelo algoritmo *Random Forest*. Neste caso, as características que melhor distinguem a DP ligeira da intermédia foram: a variabilidade da largura da passada e a variabilidade da duração do apoio duplo do passo.

Por fim, conclui-se que os modelos de *Machine Learning* propostos podem ajudar os profissionais de saúde a diagnosticar a DP, bem como a estimar a severidade da doença.

Palavras-chave: Diagnóstico, Doença de Parkinson, *Machine Learning*, Parâmetros da Marcha

Abstract

Parkinson's disease (PD) is a chronic progressive neurogenerative disorder, which is characterized by a decrease in the number of dopaminergic neurons in substantia nigra pars compacta. In this disease, non-motor symptoms are first manifested, and as it progresses, motor symptoms begin to appear. PD has no cure and early diagnosis is a great challenge. Indeed, only is possible to make a diagnosis, when motor symptoms appear, at equivalent phase of a loss of about 60% of dopaminergic neurons. Essentially, the evaluation of these symptoms is based on subjective clinical measures, which can contribute to the existence of inconsistent diagnoses.

Thus, in this dissertation, the proposed methodology comprehends the development and implementation of two Machine Learning models, one for the automatic classification of PD and the other for the automatic classification of severity of this disease. The database used includes 126 participants recruited from the Bauru region (São Paulo, Brazil), of which 63 were Parkinson's patients and were in the mild or moderate stage of the disease. The higher accuracy of the first model was 84.6%, with the naïve Bayes algorithm. The gait parameters identified as fundamental to distinguish healthy individuals from Parkinson's patients were step length, step velocity, step width and step width variability. In second model, the higher AUC value was 0.786, achieved by Random Forest algorithm. In this case, the features that best distinguish mild from moderate PD were stride width variability and step double support time variability.

Lastly, it is concluded that the proposed Machine Learning models can help health professionals to diagnosis PD, as well as to estimate the severity of disease.

Keywords: Diagnosis, Gait Parameters, Machine Learning, Parkinson's Disease

Agradecimentos

Ao orientador Professor João Manuel R. S. Tavares pela disponibilidade e orientação prestada durante o desenvolvimento da Preparação da Dissertação.

Ao Professor Fábio Augusto Barbieri pelas sugestões dadas, por ter disponibilizado a base de dados necessária para que o trabalho fosse avante e pela colaboração no mesmo. Também quero agradecer ao Tiago Penedo e ao Vinícius Christianini Moreno por terem ajudado na disponibilização da base de dados e no esclarecimento de todas as dúvidas.

Aos meus colegas que sempre se mostraram disponíveis para me ajudar e motivar e desempenharam um papel importante ao longo do mestrado. Também agradeço a todos os docentes da FEUP com os quais interagi, pois sem eles não teria alcançado as competências técnicas necessárias para chegar a esta fase.

Por último, mas não menos importante, quero agradecer a todos os amigos e família, em particular os meus pais, por todo o apoio que me deram, principalmente, nos momentos mais difíceis, incentivando-me a não desistir e por toda a confiança depositada em mim.

Marta Ferreira

"*All our dreams can come true, if we have the courage to pursue them.*"

Walt Disney

Índice

Capítulo 1 ..... **1**

Introdução **1**

1.1 - Motivação e enquadramento 1

1.2 - Principais objetivos e metodologia seguida 2

1.3 – Contribuições 2

1.4 - Estrutura da Dissertação 3

Capítulo 2 ..... **6**

Fundamentos Teóricos **6**

2.1 – Doença de Parkinson 6
- 2.1.1 – Fisiopatologia 7
- 2.1.2 – Sinais e sintomas 9
- 2.1.3 – Diagnóstico 11
- 2.1.4 – Tratamento 14

2.2 – Marcha Humana 15
- 2.2.1 - Ciclo de marcha "normal" 15
- 2.2.2 - Métodos de análise da marcha humana 16
- 2.2.3 - Marcha na doença de Parkinson 21

2.3 - *Machine Learning* 23
- 2.3.1 - Perspetiva histórica 23
- 2.3.2 - Aplicações do Machine Learning 25
- 2.3.3 - Tipos de Machine Learning 26
- 2.3.4 – Ciclo de vida de um modelo de *Machine Learning* 29

2.4 – Resumo 41

Capítulo 3 ..... **44**

***Machine Learning* na Avaliação da Doença de Parkinson: Estado da Arte** **44**

3.1 - Revisão sistemática da literatura 45
- 3.1.1 – Análise baseada na voz 46
- 3.1.2 – Análise baseada em sistemas não-*wearable e wearable* 47
- 3.1.3 – Análise baseada na escrita 52
- 3.1.4 – Análise baseada em dados de EEG e outros dados clínicos 53

3.2 - Comparação dos métodos revistos 56

3.3 – Resumo 57

Capítulo 4 ..... **60**

Caracterização da base de dados e Metodologia **60**

4.1 – Base de dados 60
- 4.1.1 - Caracterização da amostra 60
- 4.1.2 – Análise clínica 61
- 4.1.3 – Tarefa de locomoção 62

4.1.4 – Instrumentos e procedimento do estudo biomecânico 62

4.2 – Metodologia 64
4.2.1 – Análise do *dataset* 67
4.2.2 - Pré-processamento dos dados 68
4.2.3 - Seleção das caraterísticas 68
4.2.4 - Classificação 72
4.2.5 – Avaliação da classificação 74

4.3 - Resumo 74

Capítulo 5 77

Resultados e Discussão 77

5.1 – Classificação automática da DP 77
5.1.1 – Resultados do pré-processamento 77
5.1.2 - Resultados da seleção de caraterísticas 80
5.1.3 - Resultados da classificação 83

5.2 – Classificação automática da severidade da DP 89
5.2.1 – Resultados do pré-processamento 89
5.2.2 – Resultados da seleção de caraterísticas e classificação 92

5.3 - Resumo 99

Capítulo 6 101

Conclusões e Trabalhos Futuros 101

6.1 – Conclusões 101

6.2 - Trabalhos Futuros 102

Referências 104

Lista de figuras

Figura 2.1 - Núcleos da base: núcleo caudado, putamen, globo pálido, núcleo subtalâmico e substância nigra (adaptado de [13]). 7
Figura 2.2 - Diagrama ilustrativo do modelo clássico para o circuito organizacional dos núcleos da base na condição saudável e na DP, respetivamente (adaptado de [16]). 8
Figura 2.3 - Ciclo da marcha humana normal (adaptado de [15]). 16
Figura 2.4 - Exemplos de parâmetros espaciais da marcha (adaptado de [60]). 17
Figura 2.5 - Exemplo de uma plataforma de força, mostrando as três componentes da FRS (Fx, Fy, Fz) e os três momentos de força (Mx, My, Mz) medidos quando o indivíduo está em contacto com a mesma [71]. 19
Figura 2.6 - Exemplos de sensores *wearables* para análise da marcha: a) Sensor inercial posicionado no pé [74]; b) Protótipo equipado por um sistema com cinco sensores de força triaxiais para medição da FRS; c) Sensores de EMG *wireless* comercializados (incluem elétrodos e unidade de comunicação *wireless*) [73]. 20
Figura 2.7 - Exemplo de laboratório que combina sistemas não-*wearable* e *wearable*, equipado com a tecnologia BTS GaitLab, da empresa BTS Bioingenieering [70]. 21
Figura 2.8 - Postura típica de um doente de Parkinson (adaptado de [76]). 22
Figura 2.9 - Exemplo de aprendizagem supervisionada (adaptado de [95]). 27
Figura 2.10 - Exemplo de aprendizagem não supervisionada (adaptado de [95]). 27
Figura 2.11 - Aprendizagem por reforço (adaptado de [98]). 28
Figura 2.12 - Aprendizagem semi-supervisionada [101]. 28
Figura 2.13 - Exemplo de um ciclo de vida de um projeto de *Machine Learning* (adaptado de [102]). 29
Figura 2.14 - Método *Filter* (adaptado de [108]). 30
Figura 2.15 - Método *Wrapper* (adaptado de [108]). 30
Figura 2.16 - Método *Embedded* (adaptado de [108]). 31
Figura 2.17 - Hiperplano de classificação do algoritmo SVM num espaço *input* bidimensional baseado na maximização das margens (adaptado de [116]). 32
Figura 2.18 - Esquema de um *perceptron* (adaptado de [119]). 33
Figura 2.19 - *Artificial Neural Networks* com duas camadas ocultas (adaptado de [123]). ... 34
Figura 2.20 - Exemplo de classificação do algoritmo KNN. O círculo a verde é classificado consoante a classe do *k* vizinhos mais próximos (adaptado de [128]). 35
Figura 2.21 - Exemplo de uma árvore de decisão (adaptado de [136]). 36
Figura 2.22 - Exemplo de *underfitting* e *overfitting* (adaptado de [141]). 38
Figura 2.23 -Técnica de validação cruzada *5-fold* (adaptado de [143]). 39
Figura 2.24 - Exemplo de uma curva ROC (a vermelho) e o AUC correspondente (a amarelo) (adaptado de [146]). 41
Figura 3.1 - Número de publicações científicas, por ano, na Scopus, a partir dos termos pesquisados. 45
Figura 3.2 - Principais passos num sistema de classificação da DP. 57
Figura 4.1 - Fotografia tirada a um participante com os marcadores refletores posicionados segundo o protocolo Plug-in-gait *full body* (VICON). 62

Figura 4.2 - *Workflow* do modelo de ML para a deteção da DP proposto para esta Dissertação. 65
Figura 4.3 - *Workflow* do modelo de ML para a deteção da severidade da DP proposto para esta Dissertação. 66
Figura 4.4 - Distribuição demográfica dos indivíduos pertencentes ao GC e ao GDP. 67
Figura 4.5 - Distribuição demográfica dos indivíduos pertencentes ao GDPL e ao GDPI.....68
Figura 5.1 -Valores de curtose para cada parâmetro da marcha visualizados na *IPython Console*. 80
Figura 5.2 - Matriz de correlação do conjunto de treino.....80
Figura 5.3 - Matriz de correlação do conjunto de treino após remoção das caraterísticas altamente correlacionadas.....81
Figura 5.4 - *Performance* de cada algoritmo de aprendizagem em função do número de caraterísticas selecionado pelo *SelectKBest*.....83
Figura 5.5 - Análise da evolução da ACC no conjunto de treino e no conjunto de validação, para o algoritmo NB com diferentes valores do hiperparâmetro *var_smoothing*.....85
Figura 5.6 - Importância das caraterísticas para o algoritmo RF no conjunto de teste.....86
Figura 5.7 - *Heatmap* dos valores ACC obtidos para o algoritmo SVM (RBF), variando os valores dos hiperparâmetros *C* e *gamma*. 86
Figura 5.8 - *Heatmap* dos valores de ACC obtidos para o algoritmo LR, variando os valores dos hiperparâmetros *C* e *solver*. 87
Figura 5.9 - Impacto do método de seleção de caraterísticas no desempenho dos diferentes algoritmos de aprendizagem na VC 10-*fold*.....88
Figura 5.10 - Comparação entre os diferentes algoritmos de aprendizagem através da curva ROC na VC 10-*fold*.....89
Figura 5.11 - Valores de curtose para cada parâmetro da marcha visualizados na *IPython Console*. 92
Figura 5.12 - Matriz de correlação do conjunto de treino. 92
Figura 5.13 - Matriz de correlação do conjunto de treino após remoção das caraterísticas altamente correlacionadas.....93
Figura 5.14 - Distribuição das classes no conjunto de treino original.....93
Figura 5.15 - Importância das caraterísticas para o algoritmo RF no conjunto de treino do *dataset* original, utilizando os melhores hiperparâmetros. 94
Figura 5.16 - Distribuição das duas classes no conjunto de treino após a aplicação da técnica de sobreamostragem SMOTE. 95
Figura 5.17 - Importância das caraterísticas para o algoritmo RF **a)** utilizando a técnica SMOTE; **b)** utilizando a técnica ROS. 97
Figura 5.18 - *Performance* dos diferentes classificadores segundo as métricas AUC, PREC (macro avg), *Recall* (macro avg), FM (macro avg) e ACC. 98

Lista de tabelas

Tabela 2.1 – Nova versão da escala UPDRS proposta pela MDS [37]. 12
Tabela 2.2 - Escala HY modificada para a DP [42]. 13
Tabela 2.3 - Definições de alguns termos utilizados na análise da marcha. 17
Tabela 2.4 - Parâmetros da marcha em indivíduos saudáveis e nas diferentes faixas etárias ([66]–[68]). 18
Tabela 2.5 – Comparação entre as diferentes técnicas utilizadas na análise do ciclo de marcha [69]. 20
Tabela 2.6 - Exemplos de problemas de regressão, classificação e regressão e classificação simultaneamente (adaptado de [94]).....26
Tabela 2.7 - Algoritmos: tipos de problemas e principais vantagens e desvantagens.....37
Tabela 2.8 - Exemplo de matriz de confusão para um problema de classificação binária (adaptado de [125]). 40
Tabela 3.1 - Termos pesquisados nas bases de dados de referência para a pesquisa de artigos relacionados com o tema de Dissertação. 45
Tabela 3.2 - Resumo dos estudos encontrados na literatura científica relacionados com análise baseada na voz. 47
Tabela 3.3 - Resumo dos estudos encontrados na literatura científica relacionados com análise da marcha utilizando sistemas *não-wearable*. 49
Tabela 3.4 - Resumo dos estudos encontrados na literatura científica relacionados com a análise da marcha utilizando sistemas *wearable e não-wearable*.....51
Tabela 3.5 - Resumo das caraterísticas da marcha selecionadas pelos autores nos estudos encontrados na revisão literária. 52
Tabela 3.6 - Resumo dos estudos encontrados na literatura científica relacionados com a análise baseada na escrita. 53
Tabela 3.7 – Resultados da classificação do algoritmo SVM de acordo com o tipo de emoção [160].....54
Tabela 3.8 - Resumo dos estudos encontrados na literatura científica relacionados com análise baseada em outros dados clínicos. 55
Tabela 3.9 – Resultados obtidos na classificação em [164], de acordo com a intensidade do campo magnético.....56
Tabela 4.1 – Informações clínicas dos participantes. 61
Tabela 4.2 - Resumo dos parâmetros espaciotemporais calculados durante a marcha do participante. 63
Tabela 4.3 – Valores pré-definidos de cada algoritmo de aprendizagem na implementação do SciKit-Learn para a deteção da DP. 70
Tabela 4.4 – Valores pré-definidos de cada algoritmo de aprendizagem na implementação do SciKit-Learn para a classificação automática da severidade da DP. 71
Tabela 4.5 – Lista de valores atribuídos aos diferentes hiperparâmetros testados na classificação automática da DP.....73
Tabela 4.6 - Lista de valores atribuídos aos diferentes hiperparâmetros na classificação automática da severidade da DP. 73

Tabela 5.1 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função do passo dos participantes, no GC e no GDP. 78
Tabela 5.2 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função da passada dos participantes, no GC e no GDP. 79
Tabela 5.3 - Número de características ideal para cada algoritmo e respetiva descrição. 82
Tabela 5.4 - Resultados do *GridSearchCV* (VC 10-*fold*) com valores dos hiperparâmetros ideais e a ACC obtida na conjunto de treino e do conjunto de teste. 84
Tabela 5.5 - Resultados do *GridSearchCV* (LOOCV) com valores dos hiperparâmetros ideais e a ACC obtida na conjunto de treino e do conjunto de teste. 84
Tabela 5.6 - Resultados da *performance* dos diferentes algoritmos de aprendizagem após a utilização do método *Grid Search* (VC 10-*fold*). 87
Tabela 5.7 - Resultado da matriz de confusão após a utilização do método *Grid Search* (VC 10-*fold*). 88
Tabela 5.8 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função da passo dos participantes, no GDPL e no GDPI. 90
Tabela 5.9 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função do passada dos participantes, no GDPL e no GDPI. 91
Tabela 5.10 - Resultados da seleção de caraterísticas e do GridSearchCV com valores dos hiperparâmetros e os valores AUC obtidos no conjunto do treino e do teste (dados originais). 94
Tabela 5.11 - Resultado da matriz de confusão utilizando os dados originais. 95
Tabela 5.12 - Resultados do *GridSearchCV* com valores dos hiperparâmetros e com valores do AUC do treino e do teste (dados com sobre-amostragem). 96
Tabela 5.13 - Resultados da matriz de confusão utilizando as técnicas SMOTE e ROS. 97

Abreviaturas e Símbolos

Lista de abreviaturas

ACC	Exatidão (*Accuracy*)
A_i	Índice de Assimetria (*Asymmetry Index*)
ANN	Rede Neuronal Artificial (*Artificial Neural Networks*)
ANOVA	Análise de variância (*Analysis of Variance*)
AUC	Área sob a curva ROC (*Area Under the ROC Curve*)
BC	Classificador de Bayes (*Bayes classifier*)
BLR	*Boosted Logistic Regression*
BOLD	*Blood Oxygen level-dependent*
CT	Árvore de Classificação (*Classification Tree*)
DaT	Transportador Dopamínico (*Dopamine transporter*)
DBS	Estimulação Cerebral Profunda (*Deep Brain Stimulation*)
DL	*Deep Learning*
DP	Doença de Parkinson
DT	Árvores de decisão (*Decision Trees*)
EC	*Ensemble classifier*
EEG	Eletroencefalograma
EMG	Eletromiografia
FDA	*Food and Drug Administration*
FM	*F-Measure*
FN	Falsos Negativos
FoG	*Freezing* da marcha (*Freezing of Gait*)
FP	Falsos Positivos
FPR	Fração de Falsos Positivos (*False Positive Rate)*
FRS	Força de Reação do Solo
GC	Grupo de Controlo
GDP	Grupo de pacientes com Doença de Parkinson
GDPL	Grupo de pacientes com Doença de Parkinson Ligeira
GDPI	Grupo de pacientes com Doença de Parkinson Intermédia
GLCM	Matriz de co-ocorrência dos níveis de cinzento (*Grey Level CoOccurrence Matrix*)
GPi	Segmento interno do globo pálido
HY	Hoehn & Yahr
IA	Inteligência Artificial

ICC	Coeficiente de Correlação Intraclasse (*Intraclass Correlation Coefficient*)
IMC	Índice de Massa Corporal
KFD	*kernel Fisher discriminant*
KNN	*K-Nearest Neighbors*
LB	Corpo de Lewy
LCR	Líquido Cefalorraquidiano
LDA	*Linear Component Analysis*
LOOCV	*Leave-one-out cross-validation*
LOSO	*Leave-One-Subject-Out*
LR	*Logistic Regression*
MAO-B	Monoamina-oxidase B
MDS	Sociedade de distúrbios de movimento (*Movement Disorders Society*)
MDS-UPDRS	*Movement Disorders Society - Unified Parkinson's Disease Rating Scale*
ML	Aprendizagem Automática (*Machine Learning*)
MLP	*Multilayer Perceptron*
MMSE	Mini-Exame do Estado Mental (*Mini-mental state examination*)
MoCA	Avaliação Cognitiva de Montreal (*Montreal Cognitive Assessment*)
MUAPs	Potenciais de ação das unidades motoras
NB	naive Bayes
OPF	*Optimum-Path Forest*
PCA	*Principal Component Analysis*
PET	Tomografia por Emissão de Positrões (*Positron Emission Tomography*)
PPMI	*Parkinson's Progression Markers Initiative*
PREC	Precisão
PSD	função de densidade espectral (*Power Spectral Density*)
RBF	*Radial Basis Function*
REM	Movimento rápido dos olhos (*Rapid Eye Movement*)
RF	*Random Forest*
RM	Ressonância Magnética
ROC	*Receiver Operating Characteristic*
ROS	*Random Oversampling*
SMOTE	*Synthetic Minority Oversampling Technique*
SNC	Sistema Nervoso Central
SNP	Sistema Nervoso Periférico
SNpc	parte compacta da substância nigra (*substantia nigra pars compacta*)
SNpr	parte reticulada da substância nigra (*substantia nigra pars reticulata*)
SPEC	Especificidade (*Specificity*)
SPECT	Tomografia computorizada por emissão de fotão (*Single Photon Emission Computed Tomography)*

STFT	*Short Time Fourier Transform*
SVM	*Support Vector Machine*
TPR	Fração de Verdadeiros Positivos (*True Positive Rate)*
TUG	*Timed Up and Go*
UPDRS	*Unified Parkinson's Disease Rating Scale*
VC	Validação Cruzada
VN	Verdadeiros Positivos
VP	Verdadeiros Negativos

Capítulo 1

Introdução

O presente capítulo, denominado Introdução, apresenta a motivação e enquadramento ao tema, bem como os objetivos propostos, metodologia, contribuições e estrutura deste documento.

1.1 - Motivação e enquadramento

A doença de Parkinson (DP) [1] é a segunda doença neurodegenerativa mais comum, a seguir à doença de Alzheimer. Nos últimos anos apresentou um crescimento mais rápido em termos de prevalência, incapacidade e morte (6,1 milhões de pacientes em 2016 em contraste com 2,5 milhões de pacientes em 1990), com uma prevalência de 0,3% na população geral, 1% para pessoas acima dos 60 anos e 3% para pessoas com idade igual ou superior a 80 anos. A taxa de incidência da DP é de 8 em 18 pessoas por 100 000 pessoas por ano. A idade média de início da doença é 60 anos e a duração média da doença (desde o diagnóstico até à morte) é 15 anos.

Numa meta-análise realizada [2] em quatro populações norte-americanas, em 2018, a prevalência aumentou de pouco menos do que 1% em homens e mulheres com idades compreendidas entre 45 e 54 anos para 4% nos homens e 2% nas mulheres com idade igual ou superior a 85 anos. Na verdade, a mortalidade na primeira década após a DP ter sido diagnosticada não está a diminuir, em comparação com os indivíduos não afetados, mas sim a aumentar.

Também nos países nórdicos [3] se tem verificado um aumento da prevalência e da incidência da DP. Uma das razões apontadas para esse aumento é o envelhecimento generalizado da população. No entanto, estudos publicados mais recentemente apontam para um aumento do risco da DP, particularmente em homens. Algumas das fundamentações encontradas nos estudos são a mudança dramática no comportamento de fumar, em finais do século XX, aumento da poluição do ar e fatores de risco genéticos (p. ex. variante da proteína alfa-sinucleína) influenciados por fatores de risco ambientais (p. ex. fumo do cigarro, álcool, exposição da vitamina D e níveis de urato) e outros fatores genéticos desconhecidos. Contudo, outros estudos consideram como menos significativa a influência dos fatores ambientais.

Um estudo transversal realizado em Portugal [4], em 2017, com uma amostra populacional acima dos 50 anos de idade, calculou a prevalência da DP em 180 por 100 000 habitantes (IC 95% 30-327/100 000), embora com uma provável subestimativa.

O tratamento com Levodopa e estimulação cerebral profunda (DBS) têm ajudado a aumentar a sobrevivência média dos doentes, situando-se atualmente nos 14,6 (± 7,7) anos, comparativamente aos 23,3 anos estimados para um adulto de 60 anos da população geral. Porém, dada a complexidade e heterogeneidade da DP, é difícil estimar o prognóstico de um doente [4].

Uma vez que nesta doença são manifestados sintomas motores e não-motores, o diagnóstico é baseado em características clínicas. Um exemplo de sintoma motor comum que surge no início da doença e evolui com o tempo são as alterações da marcha. Contudo, os médicos recorrem a testes clínicos subjetivos para determinar a presença e a severidade da DP. Esta subjetividade na avaliação pode contribuir para a existência de diagnósticos inconsistentes [5], [6].

Na verdade, alguns dos parâmetros da marcha são específicos e pioram à medida que a doença progride. Um sistema de análise da marcha quantitativo e objetivo pode melhorar a atual forma de avaliar a marcha (semiquantitativa), auxiliando no diagnóstico, na monitorização dos sintomas, na gestão do tratamento, na reabilitação e até na prevenção e avaliação do risco de quedas nos pacientes com DP [7].

A identificação da combinação ideal de caraterísticas da marcha que melhor definem a DP é essencial, no sentido de auxiliar no diagnóstico e na gestão da doença. O *Machine Learning* ou Aprendizagem Automática (ML) fornece um método que permite identificar a melhor combinação de parâmetros espaciotemporais clinicamente relevantes [5]. Além disso, as técnicas de ML podem ajudar profissionais de saúde a diagnosticar a DP e a quantificar a progressão da doença. Porém, a análise da marcha não é, por si só, uma solução para a DP, devendo ser complementada por outras fontes de informação [8].

1.2 - Principais objetivos e metodologia seguida

Esta dissertação tem como objetivo principal o desenvolvimento e implementação de dois modelos de ML, o primeiro, para o diagnóstico da DP e, o segundo, para a previsão da severidade da DP com base nos parâmetros espaciotemporais provenientes de uma base de dados constituída por um grupo de controlo e um grupo de pacientes com DP de uma região de Bauru (São Paulo, Brasil). Pretende-se ainda testar e validar o modelo implementado com recurso a métricas de avaliação.

Para garantir uma boa fundamentação teórica, na primeira parte da dissertação é apresentada a revisão da literatura sobre DP, análise da marcha e métodos de ML. Para um melhor aprofundamento dos métodos de ML aplicados ao diagnóstico e previsão da severidade da DP foi realizado um levantamento do estado da arte, com principal foco, nos métodos aplicados no contexto da marcha humana.

1.3 – Contribuições

No âmbito desta Dissertação as principais contribuições são:

- Análise detalhada da literatura, com a inclusão de conceitos relacionados com DP, marcha humana e ML, assim como é realizada uma revisão e comparação dos modelos de ML que permitem quer a identificação da DP quer a previsão da severidade da doença nas mais diversas tecnologias de diagnóstico assistido por computador, com foco particular na análise da marcha.
- Proposta de dois modelos de ML, um para a classificação automática da DP e, outro, para a classificação automática da severidade da doença.
- Desenvolvimento, implementação, teste e validação dos modelos de ML baseados em algoritmos de aprendizagem supervisionada apresentados no estada da arte.
- Identificação dos parâmetros espaciotemporais de marcha que melhor distinguem os indivíduos neurologicamente saudáveis dos doentes de Parkinson e, dentro do grupo de doentes de Parkinson, identificação dos parâmetros que diferenciam a DP ligeira da DP intermédia. Além disso, os parâmetros selecionados são comparados com os estudos mencionados no estado da arte e interpretados de um ponto de vista clínico.
- Comparação de *performances*, no primeiro modelo, entre a utilização da validação cruzada (VC) 10-*fold* e a LOOCV (*Leave-one-out cross-validation*) na fase de aprendizagem e, no segundo modelo, entre a utilização e a ausência de técnicas sobreamostragem no conjunto de treino. Desta forma, é possível perceber que abordagens permitem melhorar os resultados da classificação.

1.4 - Estrutura da Dissertação

Para além deste capítulo introdutório, este documento contém mais cinco capítulos.

O Capítulo 2 apresenta os fundamentos teóricos relacionados com a DP (Secção 2.1), marcha humana (Secção 2.2) e ML (Secção 2.3). Relativamente à DP, é descrita a fisiopatologia, seguida dos sinais e sintomas mais comuns, os meios de diagnóstico e o tratamento. Na secção seguinte, é estudada marcha humana, nomeadamente, o ciclo de marcha "normal", os métodos de análise de marcha humana existentes e a marcha na DP. Na última secção, dedicada ao ML, é apresentada a definição, a perspetiva histórica, as aplicações e os tipos de ML, assim como o ciclo de vida de um modelo de ML. Dentro desta última subsecção são explicados os passos necessários para a implementação de um modelo: pré-processamento dos dados, seleção de caraterísticas, treino e avaliação da classificação. Ainda são abordadas formas de otimizar o modelo e a validação. Também são enunciadas algumas métricas de avaliação. No final é apresentado um resumo do capítulo (Secção 2.4).

No Capítulo 3 é apresentado o estado da arte relacionado com o ML aplicado ao diagnóstico e previsão do nível de severidade da DP. Na Secção 3.1 são abordados modelos de ML desenvolvidos e implementados por autores, com base na análise da voz, na análise em sistemas não-*wearable* e *wearable*, análise da escrita e na análise em dados de EEG e outros dados clínicos. Na Secção 3.2 , os diferentes modelos são comparados e discutidos. No final é apresentado um resumo do capítulo (Secção 3.3).

O Capítulo 4 inicia-se com a caraterização da base de dados (Secção 4.1), seguindo-se a descrição da análise clínica realizada aos participantes, bem como a tarefa de locomoção

executada, instrumentos e procedimento do estudo biomecânico. Na Secção 4.2 é exposta a metodologia proposta para esta Dissertação. A Secção 4.3 apresenta um resumo do capítulo.

No Capítulo 5 são apresentados os resultados da implementação dos dois modelos de ML no *dataset* utilizado, seguindo a metodologia proposta. Na Secção 5.1 encontram-se os resultados relativos à classificação automática da DP, enquanto na Secção 5.2 são apresentados os resultados relativos à classificação automática da severidade da DP. No final é apresentado um resumo do capítulo (Secção 5.3).

No Capítulo 6 são apresentadas as conclusões finais desta Dissertação, bem como propostas de trabalhos futuros.

Capítulo 2

Fundamentos Teóricos

A doença de Parkinson (DP) é a segunda doença neurodegenerativa mais comum a nível mundial. Em 1817, James Parkinson (1755-1824) descreveu-a, pela primeira vez, como uma doença neurológica, tendo sido realizadas importantes descobertas durante os mais de 200 anos de história da DP. Ao longo destes anos foram revelados vários aspetos sobre a DP, que incluem características clínicas e patológicas, anatomia e neuroquímica, fatores ambientais e genéticos, hipótese de disseminação periférica para central da patologia através do nervo vago, avanços na tecnologia de diagnóstico e nas abordagens terapêuticas. Todas estas descobertas contribuíram para uma melhor compreensão da fisiopatologia da doença e, por consequência, uma gestão mais abrangente da mesma [9].

Este capítulo está dividido em quatro secções. A Secção 2.1 "Doença de Parkinson" apresenta uma caraterização da DP, incidindo na sua fisiopatologia e nos sinais e sintomas que definem esta doença. Seguidamente, é abordado o diagnóstico clínico. Por fim, esta secção termina com uma breve abordagem sobre as diferentes formas de tratamento. A Secção 2.2 "Marcha Humana" apresenta conceitos relacionados com o ciclo de marcha em indivíduos saudáveis, ou seja, "normal", assim como aborda os vários parâmetros que são obtidos a partir da análise da marcha e os respetivos sistemas de análise. Por último, será explorada a marcha típica em indivíduos com DP. Na Secção 2.3 "*Machine Learning*" são abordados alguns conceitos relacionados com o ML, começando pela sua evolução de um ponto de vista histórico, as principais aplicações, os tipos de ML e, por último, são exploradas as diferentes etapas do ciclo de vida de um modelo de ML: pré-processamento dos dados, seleção de características, escolha do algoritmo de aprendizagem e treino e avaliação da classificação. Na Secção 2.4 é apresentado um resumo do capítulo.

2.1 – Doença de Parkinson

Em 1817, Parkinson descreveu no trabalho, intitulado "An Essay on the Skaking Palsy", várias caraterísticas da condição que, atualmente, tem o seu nome. Parkinson enfatizou o tremor em repouso, postura em flexão, marcha festinante, disartria, disfagia e constipação. Charcot e outros, mais tarde, consideraram o termo "paralisia agitante" inapropriado, porque, na DP, a

força era, normalmente, bem preservada e muitos pacientes não tremiam. Relativamente aos sintomas não-motores, no trabalho original de Parkinson, apenas são referidas a ansiedade e a depressão [10].

Nos primeiros anos da década de 60, a DP começou a ser tratada com um medicamento "milagroso" designado por Levodopa. Em 1967, foi aprovado pela FDA (*Food and Drug Administration*). Este avanço foi um dos marcos mais importantes na farmacoterapia das doenças neurológicas, na primeira metade do século XX. Este progresso deveu-se a uma série de grandes avanços na compreensão dos mecanismos neuroquímicos subjacentes à doença. Atualmente, apesar da evolução significativa da farmacologia da DP, a Levodopa continua a ser essencial no tratamento [11].

2.1.1 – Fisiopatologia

O SNC (sistema nervoso central) consiste no encéfalo e na medula espinhal. Por outro lado, o sistema nervoso periférico (SNP) é exterior ao SNC e inclui recetores sensoriais, nervos, gânglios e plexos. O encéfalo humano é constituído pelo cérebro, diencéfalo, cerebelo e tronco cerebral. A substância cinzenta na superfície exterior do cérebro é o córtex e, na profundidade do cérebro, constitui os núcleos da base. A substância branca, entre o córtex e os núcleos centrais, é o centro oval. O centro oval consiste em feixes nervosos que ligam diferentes partes do córtex cerebral e este a outras regiões do SNC [12].

Os núcleos da base (ou gânglios da base) são um grupo de cinco estruturas que se designam por núcleo caudado, putamen, globo pálido, núcleo subtalâmico e substância nigra (Figura 2.1). Os três primeiros núcleos localizam-se no córtex cerebral. O globo pálido é dividido em segmento interno (GPi) e segmento externo (GPe). A substância nigra é dividida em parte compacta (SNpc, *substantia nigra pars compacta)* e parte reticulada (SNpr, *substantia nigra pars reticulata)*. O núcleo caudado e o putamen designam-se em conjunto por estriado; o putamen e o globo pálido em conjunto são chamados núcleo lenticular [13]. Estes núcleos integram comandos motores, assegurando que determinados grupos musculares são ativados ou inibidos.

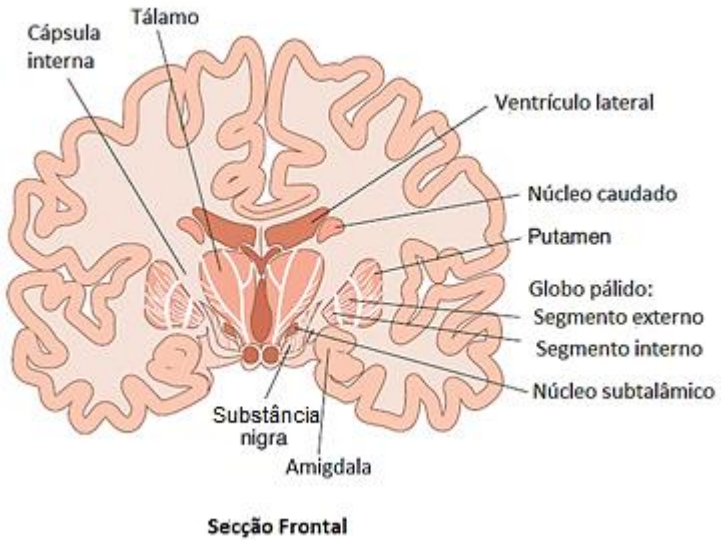


Figura 2.1 - Núcleos da base: núcleo caudado, putamen, globo pálido, núcleo subtalâmico e substância nigra (adaptado de [13]).

O tronco cerebral contém o bulbo raquidiano, a ponte e o mesencéfalo. O bulbo raquidiano encontra-se entre o cérebro e o cerebelo. Inclui centros reflexos para respostas visuais,

auditivas e táteis. A ponte contém feixes de axónios localizados entre o cerebelo e o resto do SNC. Além disso, a ponte, em conjunto com o mesencéfalo, regula a taxa de respiração e tem centros reflexos relacionados com movimentos da cabeça em resposta a estímulos visuais e auditivos. O mesencéfalo contém um número de centros reflexos para regular a frequência cardíaca e a vasoconstrição. Uma vez que se encontra superiormente à medula espinhal, contém vias que ascendem e descendem entre a medula espinhal e os centros cerebrais superiores [14].

Os núcleos da base estão ligados a alguns núcleos localizados no tronco cerebral que, juntos, atuam como um grupo funcional que forma uma via motora. O estriado é o elemento que recebe toda a informação do córtex. A transmissão de informação ocorre por duas vias: direta e indireta. A via direta liga o estriado ao GPi e à SNpr. A via indireta conecta o estriado ao GPe, ao núcleo subtalâmico e ao conjunto GPi/SNpr. A via direta origina desinibição do tálamo, enquanto que a via indireta causa, ou reforça, a normal inibição do tálamo. O tálamo tanto pode excitar o córtex (como resultado da via direta) como não excitar (como resultado da via indireta). O "interruptor" entre as duas vias é a SNpc, que se projeta para o estriado e liberta o neurotransmissor dopamina. Os recetores da dopamina podem ser excitatórios (recetores tipo D1) ou inibitórios (recetores tipo D2). A via direta é ativada pela dopamina e a via indireta é inibida pela dopamina. Quando a SNpc dispara, os núcleos da base recebem informação de que o corpo está no estado ativo e é mais provável o movimento. Quando a substância nigra está silenciosa, o corpo encontra-se no estado passivo e o movimento é inibido [15]. A Figura 2.2 apresenta o modelo clássico para o circuito organizacional dos núcleos da base. Na imagem da esquerda, a dopamina (azul) proveniente do SNpc desloca-se para o estriado para ativar a via direta (verde) e inibir a via indireta (vermelha). Este processo causa inibição do GPi, assim como inibição do tálamo e do córtex motor, promovendo o movimento. Na DP (imagem da direita), a diminuição de dopamina na SNpc causa hipoatividade da via direta e hiperatividade da via indireta, levando a uma excessiva atividade do GPi. Como resultado, há uma inibição excessiva do tálamo e do córtex motor, levando a uma supressão do movimento.

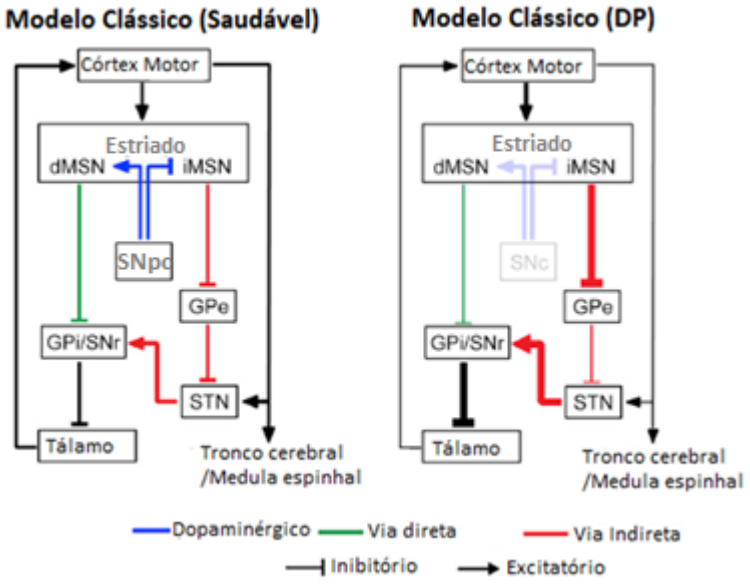


Figura 2.2 – Diagrama ilustrativo do modelo clássico para o circuito organizacional dos núcleos da base na condição saudável e na DP, respetivamente (adaptado de [16]).

As principais caraterísticas patológicas da DP são a perda dos neurónios dopaminérgicos com subsequente despigmentação da SNpc e a presença da proteína alfa-sinucleína mal dobrada nas inclusões intra-citoplasmáticas designadas por corpos de Lewy (LBs). Quando os pacientes são diagnosticados, pela primeira vez, com esta doença, uma proporção substancial de neurónios dopaminérgicos no SNpc, cerca de 60%, já desapareceu e a neurodegeneração continua a alastrar-se para outras regiões do SNC.

Os LBs são inclusões intraneuronais, arredondadas e eosinofílicas e têm dois componentes: alfa-sinucleína e ubiquitina. A alfa-sinucleína tem propensão para dobrar mal, tornando-se insolúvel e origina agregados da proteína beta-amiloide que se acumulam e formam inclusões intracelulares. A mutação do gene SNCA, responsável pela codificação da alfa-sinucleína, pode contribuir para um aumento da probabilidade de dobrar mal e agregar [17].

A deficiência da dopamina no estriado desempenha um papel importante no desenvolvimento dos sintomas motores, embora não seja claro se todos os sintomas sejam resultado apenas da deficiência de dopamina ou se então envolvidos outros neurotransmissores. Contudo, sabe-se que a perda neuronal na DP não se limita apenas aos neurónios dopaminérgicos da substância nigra, mas também afeta outras estruturas produtoras de noradrenalina, serotonina e acetilcolina [11].

Braak e os seus colegas propuseram um sistema de estadiamento de progressão da patologia da alfa-sinucleína na DP com seis etapas. Inicialmente, as alterações ocorrem no tronco cerebral inferior e no bulbo olfatório, com "disseminação" progressiva para níveis mais elevados do SNC, em etapas posteriores. Segundo esta teoria, as manifestações pré-clínicas da DP são não-motoras e, nas etapas mais tardias, deve estar associada com a demência cortical dos LBs [10].

Embora o mecanismo subjacente à patogénese da DP não seja ainda compreendido, a evidência sugere que a agregação da alfa-sinucleína solúvel tem um papel significativo no processo neurodegenerativo da DP ao prejudicar muitas funções subcelulares [18].

2.1.2 – Sinais e sintomas

Habitualmente, a DP é entendida como um distúrbio de movimento caraterizada, inicialmente, por lentidão de movimentos, tremor e rigidez muscular afetando uma parte do corpo.

Ao longo dos anos de progressão da doença, os sintomas motores tornam-se mais generalizados, o tronco tende a inclinar para a frente, o equilíbrio é afetado e a marcha torna-se mais difícil.

Nos dias de hoje, a compreensão da DP é mais clara e releva ser um distúrbio que envolve numerosos sintomas não-motores, que se podem manifestar no início da doença, mas, tipicamente, tornam-se mais proeminentes e incapacitantes após muitos anos de progressão da doença [19].

Sintomas motores

Nas últimas três décadas, um número significativo de critérios e diretrizes têm sido introduzidos para melhorar o diagnóstico da DP. Como resultado, a exatidão do diagnóstico subiu de 75% para 82%, porém, vários pacientes ainda precisam da opinião de um médico especialista em distúrbios de movimento [20].

Primeiramente, para se estabelecer o diagnóstico da DP é necessário identificar se o paciente tem parkinsonismo. Trata-se de um diagnóstico clínico que se baseia em três elementos-

chave: bradicinesia, tremor e rigidez. Em qualquer um dos casos, a bradicinesia deve estar presente e, pelo menos, um dos outros dois sintomas referidos [21].

A bradicinesia é definida como uma lentidão de movimento e uma progressiva redução da frequência ou amplitude dos movimentos repetitivos. Na prática clínica, erradamente é comum aparecer este termo como sinónimo de hipocinesia (movimentos de pequena amplitude) ou acinesia (ausência ou muito limitado movimento voluntário espontâneo). A bradicinesia pode-se manifestar nos membros, na face, na voz e na marcha.

Por outro lado, o tremor de repouso é uma caraterística parkinsoniana que se manifesta como um movimento rítmico de baixa frequência (4-6 Hz) em todo o membro em repouso e desaparece no início do movimento [22]. É comum aparecer nas extremidades (envolvendo o polegar e o indicador na forma de *pill-rolling*), no queixo, nos lábios e na língua.

A rigidez dos membros é caraterizada por um aumento da resistência ao movimento passivo e é, normalmente, manifestado por um movimento em "roda dentada" [20].

Após o parkinsonismo ter sido identificado, é importante excluir outras causas desta síndrome. Para isso, devem ser investigadas todas as caraterísticas atípicas na história e no exame clínicos. Além da presença do parkinsonismo, existem outros indicadores que podem contribuir para um diagnóstico clínico da DP. A presença de sintomas não-motores é importante, uma vez que podem ser proeminentes no início da doença [21].

Além dos três elementos-chave, a instabilidade postural é outra caraterística comum na DP mais avançada. Neste sintoma motor há uma incapacidade para fazer ajustes na posição do corpo, no sentido de manter o equilíbrio e o impulso apropriado durante o movimento. A instabilidade postural surge na DP como resultado de vários fatores: função motora progressivamente degenerativa que inibe uma resposta adequada; perturbação dos reflexos posturais; distúrbios dos mecanismos de propriocepção (perceção da localização espacial) do corpo; e mudanças posturais, tais como inclinar-se, reduzir a oscilação do braço e diminuir a passada da perna [23].

A bradicinesia, o tremor, a rigidez e a instabilidade postural são considerados como sintomas cardinais da DP. No entanto, existem outros sintomas motores, tais como o *freezing* da marcha (FoG) e a postura em flexão, que são mais pronunciadas nas fases mais tardias da doença [24]. O FoG é caraterizado por um bloqueio motor na marcha, devido à falta de movimento dos pés, parecendo que estão colados ao chão. Muitas vezes, os pacientes tentam mover, mas não conseguem. Este fenómeno surge devido à incapacidade de gerar deslocamento da postura suficiente para iniciar o movimento para a frente. Muitas vezes, o FoG impossibilita a tentativa de iniciar a marcha. Por outro lado, a postura em flexão começa a surgir com o avançar da doença, sendo evidente a flexão no pescoço, tórax, cotovelos, ancas e joelhos [25].

Na DP existem outros sintomas motores secundários, nomeadamente, a micrografia e ausência de expressão facial. No primeiro caso, o doente escreve com um tamanho de letra cada vez mais pequeno. Já a expressão facial assemelha-se a uma máscara, uma vez que ocorre uma diminuição dos movimentos faciais involuntários [26].

Além disso, 90% dos pacientes com DP desenvolvem alterações na voz e na fala durante o percurso da doença. Estas alterações são caraterizadas, principalmente, por uma diminuição do volume da voz (hipofonia) e uma qualidade de voz ofegante e rouca (disfonia) [10].

Sintomas não-motores

Alguns dos sintomas não-motores já tinham sido identificados por James Parkinson: distúrbios de sono, disfunção gastrointestinal, disfunção da bexiga e fadiga (extrema exaustão).

Porém, nas últimas décadas, assistiu-se a um aumento do reconhecimento da prevalência e da importância da disfunção não-motora, tanto no diagnóstico como na gestão da DP. Atualmente é aceite que a doença, não só é caraterizada por determinados aspetos motores, como também tem numerosos sintomas não-motores [27].

Alguns dos sintomas não-motores comuns são: demência, alucinações, depressão, apatia e distúrbio do sono. A demência é caraterizada por uma diminuição na atenção, na memória e nas funções visuoespaciais que comprometem as atividades do dia-a-dia. Pode ser acompanhada por mudanças afetivas, alucinações e apatia. As alucinações são um fenómeno sensorial (p. ex. visual, auditivo, somático, tátil, olfativo) que ocorre na ausência de estímulos e pode ser distinguido de ilusões, que envolve a má interpretação na presença de um estímulo [28]. A depressão, comum em pacientes com DP, é considerada como clinicamente significante em 35% da população de pacientes. Em comparação com a depressão observada em pessoas sem DP, é mais moderada, embora envolva mais frequentemente apatia e anedonia. Muitas vezes surge antes de começarem os sintomas motores. A apatia ocorre em 90% dos pacientes com DP e coexiste com a depressão e a demência, apesar de poder ocorrer em separado. Tem sido reportada nos primeiros estágios da doença [29]. Distúrbios de sono é outro sintoma não-motor muito comum, afetando até 98% dos pacientes. Inclui fragmentação do sono, insónia, distúrbio de comportamento do sono durante o movimento rápido dos olhos (REM, do inglês *Rapid Eye Movement*), excessiva sonolência diurna, movimentos periódicos dos membros durante o sono e desregulação do ritmo circadiano.

Outro sintoma não-motor comum é a disfunção autonómica, reportada com uma prevalência entre 14% e 80%. Inclui a desregulação cardiovascular, gastrointestinal, urinária, pupilar e dos sistemas termorreguladores. Pode surgir em qualquer etapa da doença, embora seja mais evidente no início e até numa etapa pré-motora, com um aumento, ao longo do tempo, dos sintomas gastrointestinais, urinários e ortostáticos [30].

Por outro lado, a perda de olfato é outro sintoma não-motor que surge com frequência. Investigações mais recentes sugerem tratar-se de um dos primeiros sinais da presença de Parkinson [23].

2.1.3 – Diagnóstico

O diagnóstico clínico baseia-se no exame físico e em exames laboratoriais, nomeadamente testes sanguíneos, e em estudos por imagem do cérebro e da medula espinhal.

Os avanços observados na imagiologia têm permitido compreender vários distúrbios neurológicos como, por exemplo, a DP. Dentro das várias modalidade de imagem existentes destacam-se os estudos de tomografia computorizada por emissão de fotão, os exames de ressonância magnética (RM) e de tomografia por emissão de positrões [31].

Os estudos de tomografia computorizada por emissão de fotão, do inglês *Single Photon Emission Computed Tomography* (SPECT) com a utilização de um transportador dopamínico (DaT) permitem identificar a presença da disfunção neuronal da dopamina pré-sináptica, distinguindo a DP de outros tipos de parkinsonismo neurodegenerativos. No caso da DP, há uma redução da captação do DaT nos gânglios da base. Além disso, o exame de DaT SPECT tem uma elevada exatidão (98%-100% na sensibilidade e na especificidade) na deteção de perda de células nigroestriadas em pessoas com DP. Em 2011, a FDA aprovou o exame de DaT SPECT para a distinção entre DP e tremor essencial, embora não seja executado com frequência.

Por outro lado, as imagens por RM, tipicamente, não são muito úteis no diagnóstico da DP. Contudo, pode ajudar a distinguir a DP de outros tipos de parkinsonismo (p. ex. no parkinsonismo vascular são visíveis alterações vasculares nas imagens) [32].

Os estudos do metabolismo da glucose no cérebro por tomografia por emissão de positrões (PET, do inglês *Positron Emission Tomography*) são um importante contributo no diagnóstico e monitorização da DP. Este metabolismo pode ser medido com exatidão e fiabilidade, através do radiofármaco ^{18}F-FDG (fluorodesoxiglucose). A captação deste radiofármaco aumenta com a atividade sináptica e diminui na presença de disfunção neuronal. Estes estudos são relevantes na DP, porque a disfunção neuronal precede alterações estruturais não detetadas nas imagens cerebrais anatómicas [33]. Contudo, este tipo de exame tem como principal desvantagem uma limitada resolução temporal e espacial, além de expor o ser humano à radioatividade. Em alternativa, a ressonância magnética funcional deve ser considerada. Uma das técnicas mais usadas é a técnica BOLD (*Blood Oxygen level-dependent*) que consiste na deteção de alterações dos níveis de saturação de oxigénio no sangue [11]. Esta técnica, em combinação com medições eletrofisiológicas, permite medir o *input* e o processamento de informação neuronal dentro da região cerebral. Através da técnica BOLD têm sido realizados estudos para averiguar as consequências funcionais da perda de dopamina da via nigroestriatal na DP. Além disso, é possível avaliar as funções não-motoras, tais como a atenção e a memória, e os efeitos agudos de uma intervenção terapêutica [34].

Outro tipo de avaliação da DP consiste na medição subjetiva da DP, por meio da quantificação dos sinais e sintomas dos pacientes e da incapacidade causada pela doença, no sentido de monitorizar o percurso da doença e o efeito das intervenções terapêuticas [35]. Seguidamente são apresentadas algumas escalas e questionários que permitem fazer essa medição.

Escala UPDRS

A *Unified Parkinson's Disease Rating Scale* (UPDRS) tem sido amplamente utilizada e extensivamente testada devido às suas propriedades clinimétricas [36]. A escala UPDRS original é constituída por quatro partes ou subescalas: avaliação da atividade mental, cognição, comportamento e humor; avaliação das atividades de vida diária; avaliação da função motora; e complicações do tratamento. Quanto maior for o resultado obtido nesta escala, maior será a incapacidade associada à doença.

Em 2001, a Sociedade de Distúrbios de Movimento (MDS) recomendou uma nova versão da UPDRS, mantendo as quatro partes, mas com soluções para algumas ambiguidades identificadas, fornecendo instruções claras e incorporando problemas da DP clinicamente pertinentes. Em junho de 2008, foi apresentada uma nova versão da escala UPDRS proposta pela MDS (MDS-UPDRS) [37], [38]. Tal como a escala UPDRS, a nova versão é constituída por quatro partes (Tabela 2.1).

Tabela 2.1 – Nova versão da escala UPDRS proposta pela MDS [37].

Parte	Descrição
I	Avaliação dos aspetos não motores da vida diária
II	Avaliação dos aspetos motores da vida diária
III	Avaliação da função motora
IV	Complicações motoras

As partes I e II contêm 13 itens cada, a parte III consiste em 18 itens e a parte IV tem 6 itens [35]. No total são 50 itens ou questões e cada questão avalia uma determinada incapacidade com cinco opções de resposta (0 = normal, 1 = discreto, 2 = ligeiro, 3 = moderado, 4 = grave) [39]. A MDS-UPDRS-III é uma medida importante, tanto de ponto de vista de investigação como na prática clínica diária, uma vez que permite determinar objetivamente o desempenho motor e a sua progressão ao longo do tempo. A pontuação máxima possível é 132 pontos [40]. Martínez-Martín *et al*. [41] propuseram pontos de corte para subclassificar os pacientes da DP, a partir das pontuações da MDS-UPDRS. No caso da MDS-UPDRS-III, a DP encontra-se no estágio ligeiro/moderado para os pontos de corte entre 32 e 33 e no estágio moderado/grave para os pontos de corte entre 58 e 59.

Escala de Hoehn & Yahr (HY)

A escala HY é o sistema mais amplamente utilizado e aceite universalmente, com o intuito de avaliar a incapacidade funcional na DP [42]. Esta escala permite dividir a severidade dos sintomas em graus, utilizando uma escala de 1 a 5 para avaliar a extensão da incapacidade clínica dos pacientes. Pacientes cujos sintomas se situem nos estádios entre 1 e 3 são considerados minimamente desabilitados e conseguem ter uma vida independente. Se estiverem entre os estádios 4 e 5 são considerados severamente desabilitados [43]. Na versão modificada desta escala, existem 7 opções de resposta, ou seja, foram acrescentados os estádios 1,5 e 2,5 [36], tal como apresentado na Tabela 2.2.

Tabela 2.2 - Escala HY modificada para a DP [42].

Estádio	Descrição
1	Envolvimento unilateral
1,5	Envolvimento unilateral e axial
2	Envolvimento bilateral sem instabilidade postural
2,5	Doença bilateral ligeira com recuperação no teste de retropulsão
3	Doença bilateral ligeira a moderada com instabilidade postural e independência física
4	Incapacidade grave, embora consiga andar ou permanecer de pé sem ajuda
5	Confinado à cadeira de rodas ou à cama, a não ser que receba ajuda

A escala HY modificada descreve a severidade da doença de forma mais ampla. Os estádios 1 a 2 correspondem à fase ligeira da DP, os estádios 2,5 a 3 equivalem à fase moderada e os estádios 4 e 5 definem a fase avançada (grave) da doença [44].

Mini-Exame do Estado Mental (MMSE)

O MMSE (do inglês *Mini-Mental State Examination*) é um questionário que foi desenvolvido na década de 70, com o objetivo de identificar indivíduos com deterioração cognitiva em estudos de campo. Atualmente, é o teste cognitivo breve mais conhecido e com mais aplicações no âmbito da avaliação do estado mental, nomeadamente em estudos epidemiológicos, da investigação clínica e na prática clínica [45].

A pontuação máxima deste exame é 30. A demência é detetada se ponto de corte se situar entre 24 e 26. As pontuações mais elevadas são interpretadas como um funcionamento cognitivo normal. Este exame inclui tarefas, tais como avaliar a orientação no tempo e espaço, construção visual, memória, linguagem e aspetos práticos [46].

Avaliação Cognitiva de Montreal (MoCA)

Tal como o MMSE, o MoCA é um teste para avaliar a função cognitiva global e permite detetar a demência. Além disso, é sensível na identificação de défice cognitivo ligeiro, situado entre a demência e o normal (de acordo com a idade). Permite avaliar a performance em diferentes domínios cognitivos, incluindo memória, função executiva e capacidades visuoespaciais. Na DP tem sido utilizado para avaliar a demência e défice cognitivo ligeiro [47].

2.1.4 – Tratamento

Na verdade, existem três tipos de tratamento para a DP: farmacológico, neurocirúrgico e terapia de suporte (por exemplo, exercício físico, fisioterapia, terapia da fala e medidas dietéticas). Até 2016, os tratamentos disponíveis eram de natureza sintomática. Atualmente não existe uma terapia capaz de abrandar a progressão da DP ou até prevenir a sua manifestação [48].

O tratamento farmacológico adotado em caso de sintomas motores é baseado em medicamentos dopaminérgicos, ou seja, baseados em dopamina. A Levodopa é o medicamento standard, sendo o mais utilizado no início da terapia. Contudo, com a progressão da doença e uma menor capacidade do sistema para armazenar dopamina, a maioria dos pacientes experiencia uma resposta com duração mais curta a doses individuais, com fases alternadas de boas e más respostas, movimentos involuntários da cabeça, tronco e membros (discinesias) e outras complicações motoras [49]. A maioria dos pacientes com DP começa a ter uma resposta oscilante à Levodopa até 5 anos após o início da toma. As oscilações mais comuns consistem em alterações na função motora presumivelmente associadas com mudanças na concentração e eficácia da dopamina no corpo estriado. No estado *on*, os sintomas estão controlados e há uma elevada concentração e eficácia da dopamina no corpo estriado. No estado *off*, há uma baixa concentração ou reduzida eficácia da dopamina no corpo estriado, estando associada a uma reemergência dos sintomas motores parkinsonianos [11], [23], [50].

Quando os sintomas de DP são extremamente graves e a medicação não é capaz de revertê-los, deve ser considerada a neurocirurgia, nomeadamente, a estimulação cerebral profunda (DBS) como opção final para o tratamento. Esta terapia consiste no envio de impulsos elétricos para certas partes do cérebro (geralmente na substância nigra ou globo pálido, que comunica com a substância nigra), através de um dispositivo neuroestimulador que é um implante cerebral. A área alvo desta terapia situa-se no núcleo subtalâmico. A estimulação dos neurónios pode levar à neurogénese e neuroplasticidade e uma melhoria dos problemas motores a longo prazo, tais como a discinesia e tremor, e todos os sintomas em resposta à Levodopa [51].

No tratamento farmacológico para sintomas não-motores são utilizados medicamentos que tratam estes sintomas por intermédio de outros neurotransmissores. São tratamentos similares aos adotados à população geral (sem DP). No caso de demência é recomendado, pela MDS, a utilização da rivastigmina. Para o tratamento da depressão são sugeridos inibidores seletivos da recaptação da serotonina e da norepinefrina e antidepressivos tricíclicos. Além disso, são sugeridas abordagens não-farmacológicas, tais como terapia cognitivo-comportamental e DBS.

Relativamente à psicose, deve ser considerada a interrupção da administração de certos medicamentos, tais como anticolinérgicos, amantadina, agonistas da dopamina, inibidores da MAO-B (Monoamina-oxidase B) e, por vezes, a Levodopa. Se persistir, pode ser necessário um tratamento com três opções: pimavanserina, clozapina e quetiapina. No caso de distúrbios de sono e fadiga não existem tratamentos farmacológicos que se revelem eficazes na atenuação destes sintomas. Relativamente aos distúrbios de sono, pode ser recomendada a toma de melatonina e, se necessário, clonazepam, embora não exista evidência da sua eficácia [32]. Em relação às disfunções autonómicas, existem alguns tratamentos farmacológicos bem sucedidos, contudo devem ser testados, primeiro, os tratamentos não-farmacológicos [52].

Além do tratamento farmacológico e neurocirúrgico, deve ser incentivada a prática do exercício físico a todos os pacientes, uma vez que traz benefícios a nível motor e funcional. De facto, a sua prática é fundamental no tratamento de sintomas motores, nomeadamente nas alterações da marcha e do equilíbrio, que não melhoram com a toma da medicação. Porém, caso a prática de exercício físico seja interrompida, os benefícios serão perdidos após alguns meses [53].

2.2 – Marcha Humana

Marcha refere-se ao deslocamento do centro de gravidade durante a locomoção. Nos humanos é conseguido à custa do movimento sincronizado dos membros inferiores e do tronco, resultando no movimento de uma posição para outra. Trata-se de uma característica do comportamento do ser humano, influenciada por fatores mutuamente independentes, tais como o peso, o género e a idade [54].

A programação locomotora ocorre nos centros supraespinhais e envolve a conversão de uma ideia num padrão de atividade muscular que é necessária para a marcha. Esta programação supraespinhal pode ser considerada como um comando central locomotor que é transmitido pelo tronco cerebral e pela medula espinhal. A execução deste comando envolve a ativação dos centros neurais inferiores que, subsequentemente, estabelecem a sequência de padrões da atividade muscular e o feedback sensorial a partir dos músculos, articulações e outros recetores que modificam os movimentos [55].

A análise da marcha permite um estudo aprofundado de como um indivíduo caminha ou executa determinadas funções sempre que se desloca de um ponto a outro. Além disso, permite compreender melhor os movimentos do corpo enquanto caminha, para que o médico forneça o melhor tratamento possível ao paciente, caso este tenha um problema na marcha [56].

2.2.1 - Ciclo de marcha “normal”

O ciclo de marcha normal tem duas principais fases: a fase de apoio e a fase de balanço, tal como se pode observar na Figura 2.3. A fase de apoio subdivide-se em cinco pontos de referência e a fase de balanço subdivide-se em três pontos.

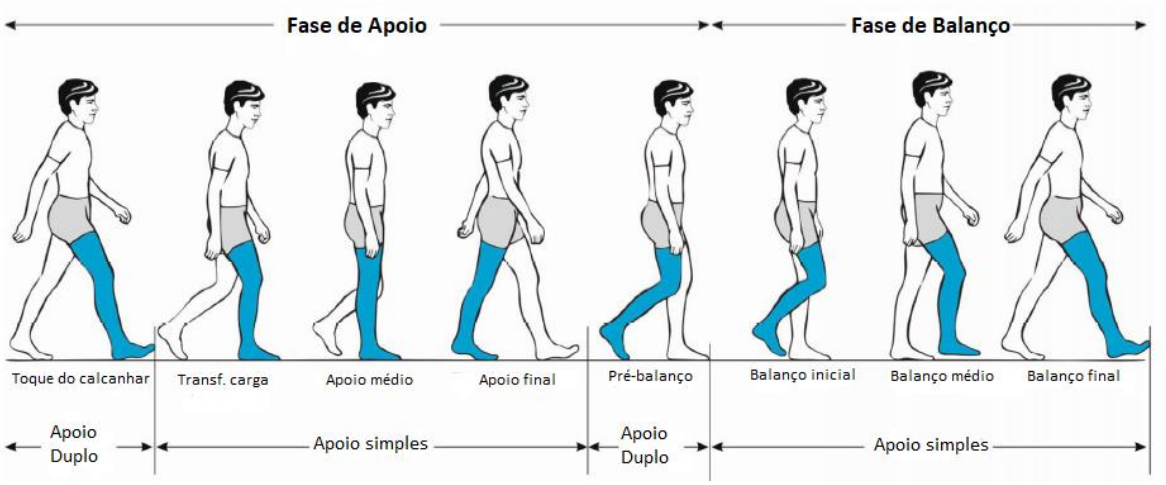


Figura 2.3 - Ciclo da marcha humana normal (adaptado de [57]).

Durante a fase de apoio, apenas um pé está em contacto com o chão. Esta fase tem os seguintes pontos de referência: 1) Toque do calcanhar é o momento em que o calcanhar toca o chão. Nesta fase, o peso do corpo gradualmente é transferido para o membro de apoio; 2) Transferência de carga é o momento em que todo o pé contacta o chão. Tal como anteriormente, o peso do corpo continua a ser transferido para o membro de apoio; 3) Apoio médio é o ponto médio da fase de apoio; 4) Apoio final é o momento em que o calcanhar deixa o chão; 5) Pré-balanço é o momento em que o hálux levanta e deixa completamente o chão.

A fase de balanço começa imediatamente após a fase de apoio. Corresponde ao momento do ciclo de marcha em que o pé não está em contacto com o chão. A perna permanece no ar e balança para a frente. Esta fase tem os seguintes pontos de referência: 1) Balanço inicial surge quando o indivíduo tenta acelerar as pernas; 2) Balanço médio é o momento em que a perna permanece no centro da fase de balanço; 3) Balanço final surge quando o indivíduo tenta desacelerar até parar a perna, preparando-a para o próximo toque do calcanhar [58].

Analisando a marcha, embora pareça que a fase de apoio e a fase de balanço sejam responsáveis por 50% do ciclo de marcha, tal suposição é falsa. A fase de apoio é responsável por 60% do ciclo de marcha, enquanto que a fase de balanço é responsável pelos restantes 40%. Isso deve-se há existência de um período de dupla sustentação dos membros, em que ambos os pés estão em contacto com o chão [59].

2.2.2 - Métodos de análise da marcha humana

A análise da marcha inclui a medição de parâmetros temporais, cinemáticos, cinéticos e eletromiografia (EMG) dinâmica, a partir das quais podem ser tiradas conclusões sobre o indivíduo (saúde, idade, tamanho, peso, velocidade, etc.) [60].

Os diferentes tipos de parâmetros são apresentados de seguida [60]:

- Antropométricos - consideram as dimensões corporais humanas e incluem a idade, o género, o peso, a altura, o comprimento do membro e o índice de massa corporal (IMC).
- Espaciotemporais - também conhecidos como parâmetros gerais da marcha. São responsáveis por fornecer uma forma simples de avaliar objetivamente a marcha (especificações da marcha). Os parâmetros da marcha consideram parâmetros temporais e espaciais, tais como comprimento do passo e da passada, largura do passo, cadência, velocidade, fases de apoio e de balanço e eventos espaciotemporais do apoio do pé e

da saída do hálux. A Figura 2.4 representa alguns parâmetros espaciais avaliados no ciclo da marcha. Na Tabela 2.3 são apresentadas definições de alguns termos utilizados na análise da marcha.

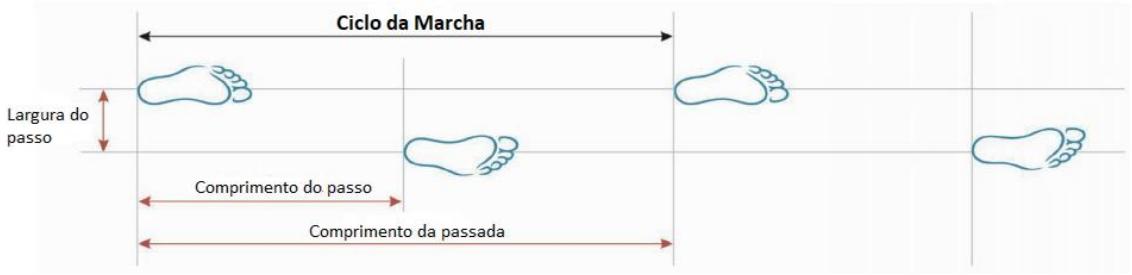


Figura 2.4 - Exemplos de parâmetros espaciais da marcha (adaptado de [57]).

- Cinemáticos – são medidas ou descrições geométricas de movimento dos segmentos corporais. É o movimento do corpo no espaço sem considerar forças. Exemplos de parâmetros cinemáticos são os ângulos das articulações ou dos segmentos, o movimento angular, a aceleração e a trajetória do segmento.
- Cinéticos - é um grupo de forças envolvido na produção da força de reação do solo (FRS), torque, padrões de pressão e forças articulares.
- Parâmetros de EMG - são requisitados no estudo da atividade muscular durante a marcha. Para a aquisição de potenciais de ação das unidades motoras (MUAPs) utilizam-se elétrodos de superfície ou de agulha.
- Combinados - para uma melhor análise e visualização, os investigadores combinam parâmetros como ângulos das articulações e FRS com parâmetros antropométricos.

Tabela 2.3 - Definições de alguns termos utilizados na análise da marcha.

Termo	**Definição**
Comprimento do passo [59]	Distância entre dois apoios de calcanhar sucessivos
Comprimento da passada [55]	Distância percorrida por um indivíduo durante um ciclo, entre um apoio do calcanhar e o próximo do mesmo pé
Largura da passada [55]	Distância mediolateral entre os dois pés
Cadência [61]	Número de passos por minuto
Duração do apoio simples [61]	Intervalo de tempo em que apenas um pé está em contacto com o chão (medida em s)
Duração do apoio duplo [61]	Duração entre a parte inicial e final da fase de apoio quando os dois pés estão em contacto com o chão (medida em s)
Duração da passada [62]	Duração do ciclo da marcha completo (medida em s)
Duração do passo [63]	Intervalo de tempo entre sucessivos contactos de cada pé com o chão (medida em s)
Velocidade da passada [62]	Comprimento da passada dividido pela duração da passada
Velocidade do passo [64]	Comprimento do passo dividido pela duração do passo para cada membro inferior separadamente

Através do estudo da marcha, em indivíduos saudáveis, é possível determinar os intervalos de valores normais relativamente aos parâmetros da marcha [65]. Na Tabela 2.4 são apresentados exemplos desses intervalos observados de forma independente, em vários estudos realizados, em diferentes faixas etárias.

Tabela 2.4 - Parâmetros da marcha em indivíduos saudáveis e nas diferentes faixas etárias ([66]–[68]).

Parâmetro espaciotemporal	Jovem (5 anos) média ± dp	Adulto média ± dp	Idoso média ± dp
Comprimento do passo (m)	0,44 ± 0,04	0,69 ± 0,002	-
Comprimento da passada (m)	0,87 ± 0,08	1,38 ± 0,05	1,34 ± 0,18
Largura do passo (m)	-	0,14 ± 0,01	-
Largura da passada (m)	-	-	0,1 ± 0,03
Cadência (passos/min)	141,1 ± 14,7	122 ± 4,27	-
Velocidade (m/s)	1,03 ± 0,18	1,40 ± 0,09	1,22 ± 0,23
Apoio simples (s)	0,35 ± 0,03	0,40 ± 0,02	0,41 ± 0,04
Apoio duplo (s)	0,16 ± 0,05	0,19 ± 0,02	0,29 ± 0,07
Duração do passo (s)	0,43 ± 0,06	0,50 ± 0,02	-
Duração da passada (s)	-	0,99 ± 0,03	1,12 ± 0,12

dp – desvio-padrão

A quantificação dos desvios dos parâmetros da marcha em relação aos valores normais forma a base do diagnóstico clínico e do potencial tratamento para pessoas com incapacidades motoras.

Por outro lado, existem dois tipos de sistemas que permitem analisar a marcha de forma objetiva: não-*wearable* e *wearable*. Os sistemas não-*wearable* são estacionários e requerem a aplicação de sensores que adquirem dados da marcha, enquanto o sujeito caminha. Por outro lado, os sistemas *wearable* permitem a captura e análise dos dados de marcha, fora do laboratório, durante as atividades de rotina diárias. Também há o sistema híbrido que combina os dois métodos [56].

Existem três tipos de sistemas *não-wearable*: sistemas óticos de captura de movimento, plataformas de força e tapetes instrumentados. Os sistemas óticos de movimento podem ser divididos em sistemas baseados em marcadores e sistemas sem marcadores. Os sistemas baseados em marcadores utilizam marcadores refletores colocados em pontos anatómicos (p. ex. articulações), permitindo a captura do movimento de cada parte do corpo. São sistemas altamente precisos (estimativa média de ruído = 0,03-0,05%) e reproduzíveis [coeficiente de correlação intraclasse (ICC) > 0,95]. Relativamente aos sistemas sem marcadores, não são tão precisos nem fidedignos, contudo, têm como vantagens, um reduzido tempo de preparação e a ausência de obstáculos ao movimento (provocados pelos marcadores). Existem dois tipos de sistemas sem marcadores: sistemas de visão ativos e passivos. Os sistemas ativos emitem luz infravermelha e visível, através de um laser ou pulsos de luz padronizados ou modulados. Os sistemas passivos utilizam técnicas de análise de imagem em tempo real [69].

Relativamente aos últimos dois sistemas *não-wearable,* tanto as plataformas de força (Figura 2.5) como os tapetes instrumentados são equipados com sensores localizados no chão, que medem parâmetros relacionados com a marcha. Existem dois tipos de sensores de chão:

plataformas de força e sistemas de medição da pressão. As plataformas de força diferem dos sistemas de medição de pressão, porque, embora também quantifiquem o centro de pressão, não medem diretamente o vetor de força aplicado. Os sistemas de medição de pressão são úteis para quantificar o padrão de pressão exercida pelo pé no chão, ao longo do tempo. Os sensores de chão, ao contrário dos sistemas óticos, analisam a FRS, ou seja, a força exercida pelos pés de um indivíduo no chão, enquanto este caminha [56], [70].

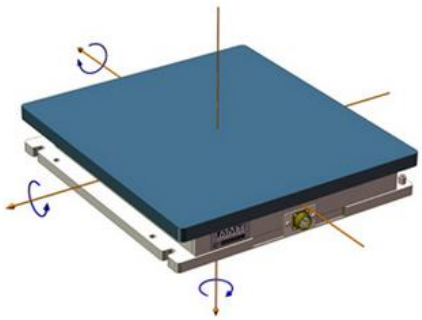


Figura 2.5 - Exemplo de uma plataforma de força, mostrando as três componentes da FRS (Fx, Fy, Fz) e os três momentos de força (Mx, My, Mz) medidos quando o indivíduo está em contacto com a mesma [71].

Os sistemas *wearable* utilizam sensores localizados em diversas partes do corpo, tais como pés, tornozelos, joelhos, coxas, ancas ou cintura. Estes sistemas incluem acelerómetros, giroscópios, magnetómetros, sensores de força, extensómetros, goniómetros, marcadores ativos, EMG, etc. Alguns destes sensores são abordados de seguida.

Os sensores inerciais são os mais utilizados na análise da marcha e permitem medir a velocidade, aceleração, orientação e forças gravitacionais. Combinam acelerómetros com giroscópios. Podem conter magnetómetros. A Figura 2.6 a) ilustra um exemplo de sensor inercial posicionado no pé.

Os goniómetros medem ângulos das articulações, tais como tornozelos, joelhos, ancas e metatarsos. Podem ser baseados em extensómetros, sensores indutivos ou mecânicos.

Os sensores ultrassónicos medem o tempo que demora para enviar e receber a onda produzida quando é refletida num objeto. Estes sensores permitem medir o comprimento do passo ou da passada, assim como a distância de separação entre os dois pés.

Os sensores de força medem a FRS produzida debaixo do pé e convertem-na, de modo proporcional, em tensão. Na Figura 2.6 b) é apresentado um protótipo de um sapato instrumentado com um sistema *wearable* constituído por cinco sensores de força triaxiais. Os sensores de pressão são, normalmente, usados para a medição da distribuição da pressão plantar, deteção da fase da marcha e deteção do passo. A maioria dos sensores utilizados nos modelos são do tipo resistivo, piezoelétrico, capacitivo e piezorresistivo [72].

A EMG mede, de forma indireta, a atividade muscular a partir de elétrodos de superfície ou em agulha, analisando a ação dos músculos na extremidade inferior durante a marcha humana. Os elétrodos funcionam como sensores de EMG que detetam potenciais de tensão e fornecem informação sobre o tempo e a intensidade da contração muscular, podendo ser comercializados em combinação com a tecnologia *wireless*, tal como se observa na Figura 2.6 c). Enquanto os elétrodos de superfície são utilizados quando se pretende obter informações gerais sobre a atividade muscular, os elétrodos de agulha limitam-se ao estudo de um músculo em particular [73].

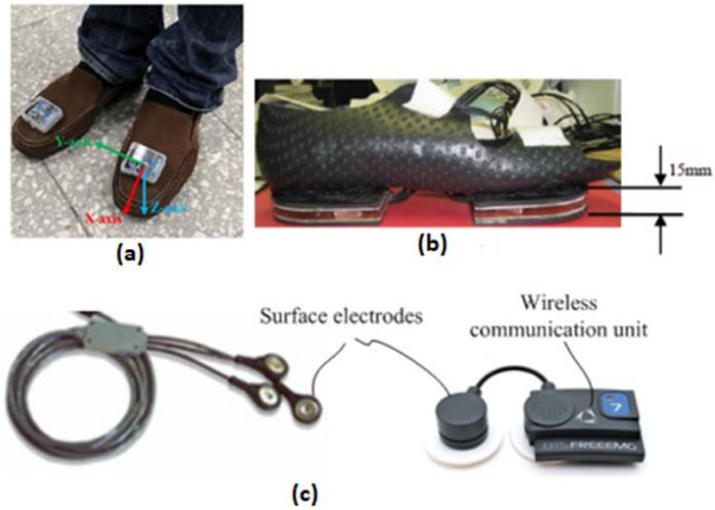


Figura 2.6 - Exemplos de sensores *wearables* para análise da marcha: a) Sensor inercial posicionado no pé [74]; b) Protótipo equipado por um sistema com cinco sensores de força triaxiais para medição da FRS; c) Sensores de EMG *wireless* comercializados (incluem elétrodos e unidade de comunicação *wireless*) [73].

Embora os sistemas não-*wearable* sejam altamente precisos, fidedignos e forneçam resultados detalhados, requerem equipamentos pesados e dispendiosos. Por outro lado, os sistemas *wearable* providenciam menos detalhe, mas podem ser usados em ambiente comunitário, garantindo feedback em tempo-real a médicos e pacientes [69]. A Tabela 2.5 apresenta, para cada técnica de análise de marcha, os parâmetros que podem ser extraídos e o respetivo ICC.

Tabela 2.5 - Comparação entre as diferentes técnicas utilizadas na análise do ciclo de marcha [69].

Técnica de Análise	Parâmetros obtidos	ICC
Captura de movimento baseada em marcadores	Espaciais e temporais; cinemáticos	0,95-1,00
Captura de movimento sem marcadores	Espaciais e temporais; cinemáticos	0,91-0,98
Plataformas de força	FRS; cinemáticos	0,22-0,97
Tapetes instrumentados	Espaciais e temporais	0,69-0,99
Sensores de pressão	Espaciais e temporais	0,90-0,99
Sensores inerciais	Espaciais e temporais; cinemáticos	0,90-0,99

Atualmente, muitos laboratórios e clínicas de análise de marcha combinam os dois tipos de sistemas. A Figura 2.7 ilustra um exemplo de um laboratório equipado com a tecnologia BTS GaitLab. Este laboratório contém câmaras de vídeo sensíveis aos infravermelhos, sensores inerciais, tapete instrumentado para medição da FRS, EMG *wireless*, *workstation*, sistema de gravação de vídeo, ecrã de TV e estação de controlo [70].



Figura 2.7 - Exemplo de laboratório que combina sistemas não-*wearable* e *wearable*, equipado com a tecnologia BTS GaitLab, da empresa BTS Bioingenieering [70].

A utilização dos dois sistemas em laboratório oferece uma maior sensibilidade e fiabilidade, contudo são dispendiosos, exigem muito tempo e espaço, assim como necessitam de certas capacidades técnicas para a operação. Em alternativa, existem sensores portáteis (p. ex. Kinect®, Nintendo Wii Balance e GAITRite®) e *wearable* que são menos dispendiosos e, também, medem de forma fidedigna parâmetros da marcha dentro e fora da clínica. Um exemplo da aplicação clínica são as intervenções de treino de equilíbrio na prevenção de quedas [69].

2.2.3 - Marcha na doença de Parkinson

Na DP, o comprometimento da marcha é uma condição em constante evolução e podem ser detetados diferentes padrões de marcha, à medida que a doença progride [7].

A marcha parkinsoniana [23], [31], [75] é caraterizada por um reduzido movimento dos músculos do tronco e dos membros superiores e inferiores, o que resulta numa marcha com uma base de apoio reduzida, lenta e arrastada com passos curtos e rápidos (Figura 2.8). Há uma tendência involuntária para aumentar a velocidade dos passos cada vez mais curtos, devido a uma alteração descontrolada do centro de gravidade para a frente, com uma incapacidade para terminar o movimento (festinação). O movimento das articulações é reduzido, em virtude da rigidez, ocasionando, geralmente, uma reduzida ou ausência de oscilação dos braços. A marcha na DP reflete uma combinação de acinesia, diminuição dos reflexos posturais, distonia, rigidez e tremor. Os pacientes da DP apresentam, muitas vezes, uma postura em flexão, com o tronco projetado para a frente, a cabeça inclinada para baixo, os braços moderadamente fletidos nos cotovelos e as pernas ligeiramente fletidas. Os dedos e punhos apresentam tremor de repouso, embora os braços se mantenham imóveis. À medida que o paciente caminha para a frente, as pernas permanecem fletidas nos joelhos, pernas e tornozelos. Além disso, costuma ter dificuldade em começar a andar, após ter estado parado ou sentado por um longo período de tempo. Os movimentos de marcha podem parar involuntariamente e o paciente "congela" a marcha, fenómeno designado por FoG, ao tentar passar por uma porta ou entrar no elevador. O FoG [25] é muito comum na DP e tanto surge nos estados *on* como *off*. Há, ainda, uma dificuldade para

alterar o sentido do movimento (virar "em bloco"). A velocidade da marcha é lenta, o equilíbrio é reduzido e há uma maior propensão para os pacientes cair.



Figura 2.8 - Postura típica de um doente de Parkinson (adaptado de [76]).

Os autores Zanardi *et al*. [77] realizaram uma revisão sistemática e meta-análise de 72 estudos, envolvendo 3027 participantes, de modo encontrar diferenças nos parâmetros da marcha entre 1510 pacientes de DP no estado *on* e um grupo de controlo com 1517 indivíduos saudáveis. Nos doentes de Parkinson foi observada uma velocidade da marcha mais reduzida em comparação com o grupo de controlo, assim como, uma diminuição do comprimento da passada e da duração da fase de balanço. Os autores consideraram a possibilidade de a bradicinesia e rigidez associados à inatividade física contribuírem para a diminuição da velocidade e, por esta estar estritamente relacionada com o comprimento da passada e a cadência, estes parâmetros serem igualmente afetados. Além disso, estes indivíduos apresentaram uma cadência e duração do apoio duplo superiores ao grupo de controlo. Segundos os autores, quer o aumento da cadência quer da duração do apoio duplo podem estar associados com uma reduzida estabilidade dinâmica da marcha na DP. Por outro lado, não foram encontradas diferenças entre os grupos relativamente à largura do passo nem à duração do apoio simples.

A marcha na DP pode também ser analisada consoante a severidade da sua progressão. Na fase inicial da doença, prevalece uma diminuição da velocidade da marcha, do comprimento do passo e da duração da fase de balanço, associados a um aumento parcial da cadência, em compensação. Este facto pode estar relacionado com a deficiência de dopamina, uma vez que esta responde muito bem à terapia de substituição de dopamina [10]. Além disso, os sintomas motores são, muitas vezes, assimétricos, devido ao facto de a neuropatologia dos núcleos da base ser assimétrica. A variabilidade da marcha é maior nos pacientes com DP do que em indivíduos saudáveis. Entre o estágio ligeiro e moderado da doença, grande parte dos parâmetros espaciotemporais afetados na fase inicial progridem bilateralmente, de modo que a assimetria diminui. É comum a fase de apoio duplo e a cadência aumentarem [78]. Os distúrbios da marcha são mais complexos (bloqueios motores, festinação) que, em conjunto com o desequilíbrio, tornam a DP mais incapacitante. Muitas vezes, tanto os distúrbios de marcha como os problemas de equilíbrio podem não ser suficientemente controlados pelo tratamento farmacológico ou pela DBS [79]. Em fases mais avançadas da doença, os bloqueios motores são mais frequentes

e acompanhados de um reduzido equilíbrio e controlo postural, com um grande risco de queda. Nesta fase é recomendado a utilização de auxiliares de marcha ou cadeira de rodas [78].

A avaliação da marcha de uma paciente com DP deve ser realizada em todas as fases do processo de acompanhamento. Dois exemplos de testes de marcha são o teste de marcha dos 10 m e o teste *timed up and go* (TUG). Ambos os testes apresentam validade e fiabilidade em casa e em ambiente clínico. No teste de macha dos 10 m, o paciente é solicitado para caminhar em linha reta, de modo a ser medida a velocidade de marcha, o comprimento do passo e a frequência do passo. No teste TUG, é pedido ao paciente para se levantar e caminhar 3 m para a frente, virar e voltar para a sua cadeira. Desta forma, é avaliada a rotação e o início e fim da marcha. Através dos dois testes, é ainda possível observar outras caraterísticas como a postura corporal, a velocidade e a fluência do passo, o comprimento da passada e a distância entre os pés (base) [11].

O tratamento com dopaminérgicos ou DBS ajudam a melhorar a marcha, embora não tanto como noutros sintomas motores. É difícil de tratar, mas no estado *off* pode recuperar com a medicação dopaminérgica. Além disso, a marcha pode melhorar através de pistas visuais ou auditivas rítmicas. O canto mental também pode ser uma mais-valia na melhoria da *performance* da marcha. A terapia física pode ajudar, embora se traduza num benefício de curto prazo [25].

2.3 - *Machine Learning*

Recentemente, o *Machine Learning* ou Aprendizagem Automática (ML) tem recebido uma considerável atenção devido à sua capacidade de prever com exatidão uma grande variedade de fenómenos complexos [80].

O ML é uma área da ciência da computação que estuda algoritmos e técnicas para solucionar automaticamente problemas difíceis de programar usando métodos de programação convencionais. Pesquisar na Internet ou reconhecimento de voz são alguns dos exemplos mais bem-sucedidos da aplicação do ML [81].

Por outro lado, a inteligência artificial (IA) é um conceito mais vasto e refere-se à área da ciência da computação dedicada ao desenvolvimento de sistemas que executam tarefas que, normalmente, necessitam de inteligência humana. Pode ser interpretada como a incorporação da inteligência humana em máquinas e baseia-se em regras e algoritmos estipulados.

A IA engloba o ML e o ML, por sua vez, engloba o *Deep Learning* (DL). Esta última vertente inclui modelos computacionais e algoritmos que imitem a arquitetura das redes neuronais biológicas no cérebro [*Artificial Neural Networks* (ANN) ou redes neuronais artificiais]. Sempre que o cérebro recebe novas informações, este tenta compará-las com informações já conhecidas para tentar entendê-las. O cérebro decifra as informações através de rótulos (*labels*) e atribui os itens a várias categorias. O DL emprega o mesmo conceito [82].

2.3.1 - Perspetiva histórica

A história do ML está intimamente relacionada com a história da IA. Os investigadores que estabeleceram e desenvolveram o conceito de ANN e ML, adquiriram este conhecimento através dos seus envolvimentos na evolução da tecnologia de aviação no início do século XX. Assim como a tecnologia aeronáutica não precisa de imitar um pássaro para atingir o seu propósito,

na IA não é necessário produzir máquinas que entendam e pensem como os seres humanos para atingir o desempenho necessário.

Por outro lado, em 1943, McCulloch e Pitts publicaram um artigo no qual sugerem redes neuronais como forma de imitarem cérebros humanos [83].

A origem do conceito IA também remonta à década de 40, mais especificamente em 1942, quando o escritor Isaac Asimov publicou um texto intitulado "Runaround". Este texto aborda a história de um robot desenvolvido pelos engenheiros Gregory Powell e Mike Donavan, além de enunciar as três leis da robótica:

- Lei 1: um robot não pode causar danos a um ser humano ou deixar que algum ser humano se magoe;
- Lei 2: Um robot tem de obedecer a ordens por um ser humano, exceto quando essas ordens entram em conflito com a primeira lei;
- Lei 3: um robot tem de proteger a sua própria existência desde que não entre em conflito com a primeira ou segunda lei.

O trabalho de Asimov inspirou gerações de cientistas na área da robótica, IA e ciência da computação [84].

Posteriormente, o conceito de usar computadores para simular o comportamento inteligente e pensamento crítico foi descrito, pela primeira vez, por Alan Turing, em 1950. No livro "Computers and Intelligence", Turing criou um simples teste, mais tarde designado por "Turing test", para averiguar se os computadores eram capazes de ter inteligência humana. Seis anos mais tarde, John McCarthy descreveu, pela primeira vez, AI como "a ciência e a engenharia de fazer máquinas inteligentes" [85].

Um dos primeiros exemplos de IA foi uma parte de software desenvolvida pela IBM, nomeadamente, por Arthur Samuel, através da qual era possível jogar damas de forma autónoma após a recolha de dados e análise de todas as opções disponíveis antes de tomar uma decisão [86]. Foi nesta altura, em 1959, que foi introduzido, pela primeira vez, o termo "*Machine Learning*" [87].

Entre os anos de 1950 e 1960, os computadores eletrónicos começaram a ser utilizados e os algoritmos foram desenvolvidos para permitir modelar e analisar grandes conjuntos de dados. No início emergiram três grandes ramos no ML: aprendizagem simbólica descrita por Huntt *et al.*, métodos estatísticos por Nilson e redes neuronais por Rosenblatt. Estes três ramos deram origem a métodos mais avançados: métodos de reconhecimento de padrões ou estatísticos, tais como KNN (do inglês *K-Nearest Neighbors*), análise discriminativa e classificadores bayesianos; aprendizagem indutiva de regras simbólicas, tais como algoritmos de indução das árvores de decisão (DT, do inglês *Decision Trees*); e as ANN [88].

Além disso, Rosenblatt foi pioneiro no desenvolvimento de *perceptrons*. Um *perceptron* era um dispositivo eletrónico, construído em conformidade com os princípios biológicos, com capacidade para aprender. Por outro lado, os engenheiros elétricos Widrow e Hoff criaram um sistema designado por ADALINE (ADAptive LInear Element), em 1960. O método utilizado para aprender era diferente do *perceptron*, uma vez que empregava a regra dos mínimos quadrados. Em 1962, Widrow e Hoff desenvolveram um procedimento de aprendizagem que examinava o valor antes que o peso o ajustasse [89].

Nos anos 70, o trabalho em IA levou à criação dos *Expert Systems*, embora a linguagem PROLOG tenha sido desenvolvida previamente, por Colmereuer. Tratava-se de uma linguagem mais adequada para a criação destes sistemas. Em 1971, Edward Shortliffe inventou o sistema

MYCIN cujo objetivo consistia no diagnóstico de infeções bacterianas no sangue, assim como apresentava uma proposta de terapia farmacológica.

Os anos 80 trouxeram uma grande mudança, visto que as corporações reconheceram a possibilidade de utilizar IA. Trabalhos realizados na IA resultaram em resultados significativos, tendo sido encontradas aplicações práticas, nomeadamente, em jogos lógicos. Este progresso levou a que, em 1994, um programa de computador derrotasse o campeão mundial de xadrez [90].

Em 2015, as redes neuronais artificiais regressaram na forma de DL, quando o AlphaGo, um programa desenvolvido pela Google, conseguiu derrotar o campeão mundial no jogo de tabuleiro Go, feito nunca alcançado por computadores. Atualmente, as ANN e a DL formam a base da maioria das aplicações que conhecemos em IA. São a base dos algoritmos de reconhecimento facial utilizados no Facebook, dos algoritmos de reconhecimento de voz e dos carros autónomos [84].

2.3.2 - Aplicações do Machine Learning

O ML fornece métodos automatizados que podem detetar padrões nos dados e usá-los para executar determinadas tarefas [91]. Na Subsecção 2.3.4 são abordados os tipos de tarefas ou modelos no ML.

Seguidamente são apresentados os critérios que podem ser seguidos de modo a identificar as tarefas para as quais os métodos de ML podem ser úteis [92]:

- A tarefa envolve uma função que mapeia *inputs* bem-definidos para *outputs* bem-definidos;
- Existência de grandes conjuntos de dados ou possibilidade de serem criados contendo pares *input-output*;
- A tarefa fornece *feedback* percetível com objetivos bem-definidos e métricas;
- A tarefa não envolve longas cadeias de lógica ou raciocínio que dependam de diversos conhecimentos prévios;
- A tarefa não requer explicações detalhadas de como a decisão foi feita;
- A tarefa tem uma tolerância ao erro e não necessita de soluções comprovadamente corretas;
- O fenómeno ou função sendo aprendida não deve mudar rapidamente com o tempo;
- Não é necessário nenhuma destreza especializada, capacidades físicas ou algum tipo de mobilidade.

Por sua vez, os problemas em ML dividem-se em duas categorias: classificação ou regressão [93]. No caso da classificação, se, por exemplo, o objetivo de um sistema for classificar carateres escritos à mão, então este recebe uma imagem e atribui uma de duas classes possíveis: Classe 1 (C1) ou Classe 2 (C2). Para representar o resultado da classificação em termos de variável *y*, podemos considerar que, se a imagem for classificada como C1, então *y* recebe o valor 1 e, se a imagem for classificada como C2, então *y* recebe o valor 0. Desta forma, é possível constatar que os valores atribuídos aos *inputs* são discretos ou categóricos. No caso dos problemas de regressão, os valores atribuídos são contínuos. Tanto a regressão como a classificação podem ser vistas como casos particulares de aproximação de uma função. No caso dos problemas de regressão, é a função de regressão que o sistema se pretende aproximar, enquanto nos problemas de classificação, as funções que o sistema procura aproximar é a probabilidade de estar associada a diferentes classes expressas em função das variáveis *input*. Na

Tabela 2.6 são apresentados mais exemplos de problemas de classificação, regressão ou classificação e regressão simultaneamente.

Tabela 2.6 - Exemplos de problemas de regressão, classificação e regressão e classificação simultaneamente (adaptado de [94]).

Regressão	Classificação	Regressão e Classificação
• Identificar proactivamente as peças do carro que podem falhar.	• Encontrar nomes de locais ou pessoas num texto.	• Identificar clientes mais rentáveis para a empresa.
• Prever a quantidade de dinheiro que uma pessoa gastará no produto X.	• Identificar pessoas com base em imagens ou gravações de voz.	• Encontrar campos de petróleo, minas de ouro ou sítios arqueológicos com base em locais existentes.
• Prever o número de erupções de um vulcão num período de tempo.	• Reconhecer pássaros com base no seu chilrear.	
• Prever a receita anual de uma empresa.	• Identificar tumores e doenças.	
	• Prever que equipa ganhará a Liga dos Campeões em futebol.	

Para resolver estes problemas de ML, existem funções de aproximação de diferentes tipos tais como modelos lineares, SVMs (do inglês *Support Vector Machines*), DT, processos Gaussianos, entre outros [91]. Tanto os problemas de classificação como os de regressão fazem parte dos métodos de aprendizagem supervisionada [81] que será explorado na Subsecção 2.3.4 .

2.3.3 - Tipos de Machine Learning

Os métodos de ML podem dividir-se em três categorias principais (aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço)[95], dependendo se são treinados ou não com supervisão. Alguns autores consideram a existência de uma quarta categoria que consiste na combinação das técnicas de aprendizagem supervisionada e não supervisionada (aprendizagem semi-supervisionada) [96].

Aprendizagem supervisionada

Na aprendizagem supervisionada, o conjunto de treino consiste em pares *input* e *output* desejado (*label*). O objetivo é aprender uma função de mapeamento de *inputs* para *outputs*. Na Figura 2.9, os *inputs* são os pontos no plano bidimensional e os *outputs* são as *labels* atribuídas a cada *input* (círculos ou cruzes). Neste exemplo, o objetivo é aprender a realizar classificação binária, isto é, a prever duas classes. Exemplos deste tipo de técnica de ML incluem a classificação das mensagens do email em spam ou não-spam [95].

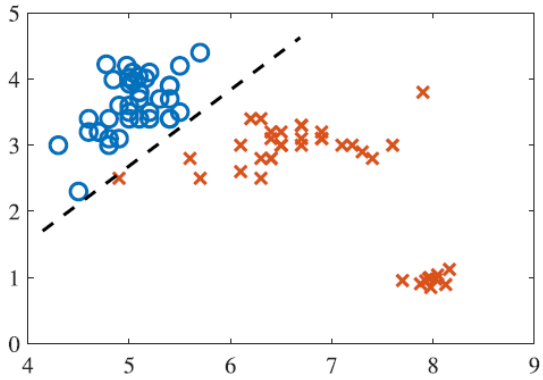


Figura 2.9 - Exemplo de aprendizagem supervisionada (adaptado de [95]).

Portanto, este tipo de modelo de ML apenas se aplica em dados com *labels* associadas, o que implica que haja interação humana para etiquetar os dados [94].

Os algoritmos mais populares de aprendizagem supervisionada são: SVM, *random forest* (RF) e *neural network* [97]. Estes algoritmos são abordados na Subsecção 2.3.4 .

Aprendizagem não supervisionada

Na verdade, a maioria dos grandes conjuntos de dados não tem *labels* associadas e, neste caso, a abordagem aprendizagem supervisionada não pode ser aplicada. Porém, é possível trabalhar nestes dados, através do estudo da distribuição dos dados e inferir verdades sobre os mesmos em diferentes partes da distribuição. É ainda possível estudar a estrutura e valores nos dados e inferir novos dados mais significativos e estruturas a partir deles [94].

No exemplo da Figura 2.10, os *inputs* são, novamente, pontos no plano bidimensional, mas não é fornecida informação sobre a *label* correspondente. Neste caso, o objetivo é agrupar conjuntos de pontos *input* que estão mais próximos uns dos outros, de modo a atribuir uma *label* a cada ponto. Desta forma, são criados *clusters* (delimitados pelas linhas a tracejado) [95].

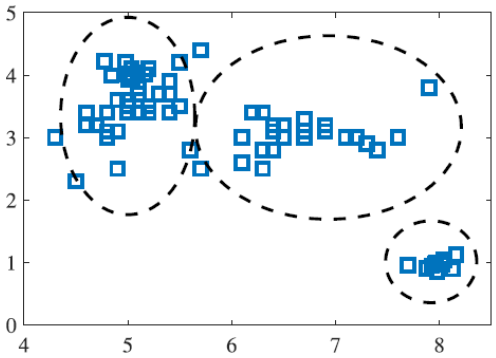


Figura 2.10 - Exemplo de aprendizagem não supervisionada (adaptado de [95]).

Os algoritmos mais populares de aprendizagem não supervisionada são: *K-means clustering* e *self-organizing neural networks*. Na primeira técnica, o algoritmo agrupa os dados em vários grupos que partilham características similares ou próximas. No segundo caso, o algoritmo organiza neurónios no sentido de minimizar os erros da função designada para o problema [97].

Aprendizagem por reforço

Aprendizagem por reforço é um tipo especial de método de aprendizagem que necessita de ser tratado em separado dos métodos de aprendizagem supervisionada e não supervisionada.

Este método envolve o feedback do ambiente, ou seja, não é exatamente sem supervisão, contudo, não tem um conjunto de amostras com *labels* associadas para o treino, logo não pode ser tratado como supervisionada. Neste tipo de ML, o sistema está continuamente a interagir com o ambiente em busca de produzir o comportamento desejado (atingir um objetivo) e a receber o feedback do ambiente [87]. Tal como se observa na Figura 2.11, esta técnica de aprendizagem envolve duas entidades: o Agente e o Ambiente. Durante o treino do Agente, este, primeiramente, toma uma ação que, por sua vez, afeta o Ambiente, causando mudança de estado. Este processo gera uma recompensa imediata e o Agente toma uma nova ação, dando continuidade ao ciclo [98].

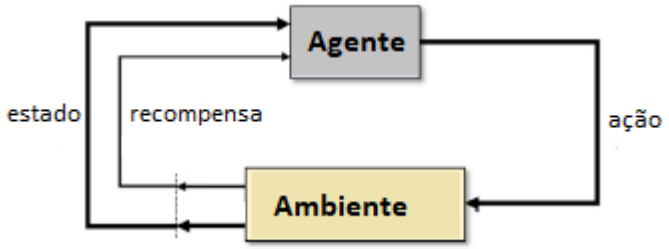


Figura 2.11 – Aprendizagem por reforço (adaptado de [98]).

O objetivo do Agente é maximizar a recompensa cumulativa esperada através da interação com o Ambiente desconhecido ao longo do tempo. Desta forma, aprende quais as ações que deve executar para aumentar a recompensa e atingir o objetivo [99].

Aprendizagem semi-supervisionada

Este tipo de ML combina simultaneamente a aprendizagem supervisionada e não-supervisionada. *A* aprendizagem semi-supervisionada permite obter uma performance superior ao abranger o conhecimento de supervisão, fornecido por um conjunto de treino pequeno com *labels* associadas e a informação sobre a distribuição das amostras provida pelo conjunto de amostras não supervisionado [100].

A Figura 2.12 mostra um exemplo de aprendizagem semi-supervisionada que inclui dados de treino parcialmente com *labels*, ou seja, uma pequena porção dos dados tem *labels* associadas e uma grande porção não tem *labels* [101]. O objetivo deste modelo é treinar com o menor número de dados etiquetados possível, de modo a tornar o processo menos dispendioso [94].

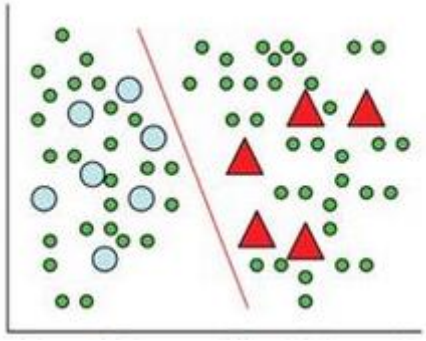


Figura 2.12 – Aprendizagem semi-supervisionada [101].

2.3.4 – Ciclo de vida de um modelo de *Machine Learning*

Para que seja implementado um modelo de ML é necessário seguir as seguintes etapas: escolha do *dataset* (conjunto de dados) e seleção de características, escolha do algoritmo de aprendizagem e um módulo de avaliação [102], tal como mostra a Figura 2.13.

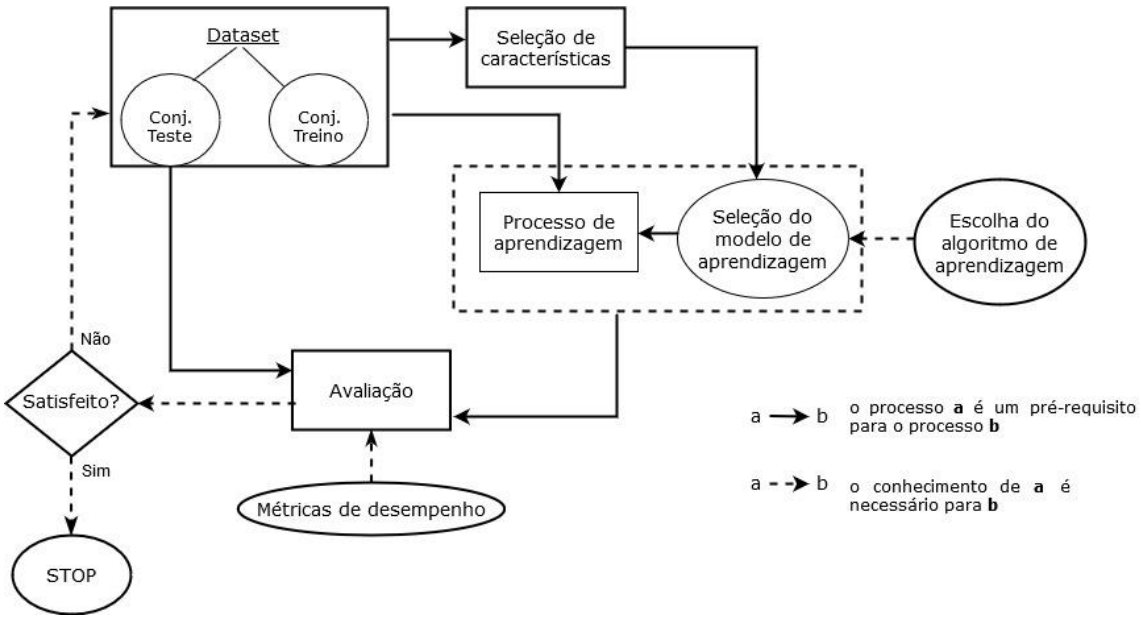


Figura 2.13 - Exemplo de um ciclo de vida de um projeto de *Machine Learning* (adaptado de [102]).

O *dataset* consiste no *input* de um modelo de ML. Posteriormente, pode ser dividido em conjunto de treino (do inglês *training set*) e conjunto de teste (do inglês *test set*). Após a escolha e aquisição do *dataset*, os dados passam por uma etapa de pré-processamento que consiste na preparação dos dados, extração de características, normalização do *dataset* e aplicação de técnicas de amostragem. Seguidamente é explicada mais detalhadamente cada etapa do pré-processamento dos dados.

A preparação do *dataset* consiste na escolha do formato mais adequado, ou seja, em matriz. Na extração de caraterísticas, os dados originais são transformados em caraterísticas potencialmente úteis na aprendizagem. Após a extração, os *datasets* são normalizados, para que contenham valores que variem dentro de um intervalo específico. Por exemplo, na normalização Min-Max, o *dataset* é normalizado para [0,1]. Por último, quando o *dataset* não está balanceado, ou seja, existe um desequilíbrio significativo se compararmos o número de amostras em cada classe, pode levar a performances insatisfatórias [97], [103]. Para contrariar isso, é necessário aplicar técnicas que aumentem ou diminuem o número de amostras de uma classe de modo a obter uma distribuição balanceada, nomeadamente, ao aplicar a técnica de sobre-amostragem (do inglês *oversampling*) ou a técnica de sub-amostragem (do inglês *undersampling*), respetivamente. No primeiro caso, as amostras pertencentes à classe minoritária são selecionadas e replicadas aleatoriamente. No segundo caso, as amostras pertencentes à classe maioritária são removidas aleatoriamente.

Um exemplo de técnica de sobre-amostragem é o SMOTE (*Synthetic Minority Oversampling Technique*) que gera amostras sintéticas minoritárias para aumentar o número de amostras da classe minoritária. Neste caso, não é necessário adicionar informação extra, pois os dados são

reutilizados. Contudo, o processo de aprendizagem consome mais tempo, uma vez que o *dataset* original contém um número muito pequeno de amostras minoritárias [104].

Seleção de características

Por vezes, o tamanho do *dataset* pode ser demasiado grande, sendo crucial investigá-lo em termos de informação irrelevante e redundante, antes de aplicar um algoritmo de aprendizagem. O processo de seleção de caraterísticas é utilizado no sentido de descobrir e remover esse tipo de informação (tanto quanto possível), com o objetivo de construir um pequeno subconjunto de caraterísticas altamente preditivo e evitar o *overfitting* dos dados de treino [102].

Existem dois tipos de métodos de seleção de características, também designados por métodos de redução da dimensionalidade do espaço de características: método *Filter* e método *Wrapper* [105]. O método *Filter* (Figura 2.14) efetua um processo de seleção de caraterísticas independente do algoritmo de aprendizagem. Neste caso, apenas são selecionados os atributos que ocupam os primeiros lugares no ranking, relativamente ao cumprimento de um determinado critério de seleção. Utiliza um critério de relevância e é, geralmente, uma etapa de pré-processamento. Trata-se de um método conhecido pela sua rapidez, mas com tendência para selecionar grandes subconjuntos de características [106], [107].

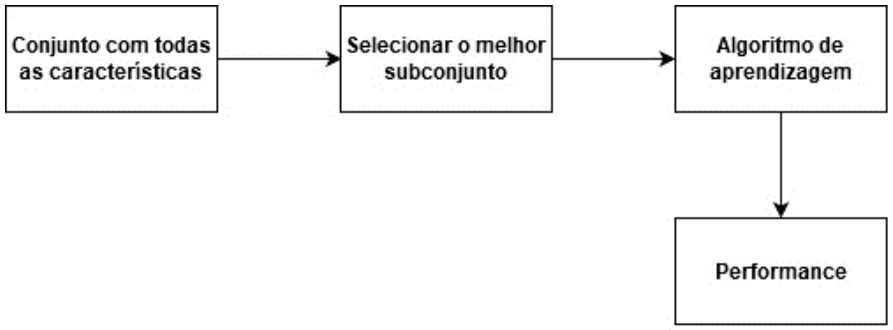


Figura 2.14 - Método *Filter* (adaptado de [108]).

Por outro lado, o método *Wrapper* (Figura 2.15) utiliza uma abordagem que seleciona iterativamente, através de um ciclo de feedback, os atributos que melhoraram a performance do algoritmo de aprendizagem utilizado para treinar o modelo. Ao contrário do método *Filter*, cada conjunto é avaliado no contexto específico do algoritmo de aprendizagem. Tem como vantagem a sua exatidão, mas com uma execução lenta e cada problema classificação requer uma nova seleção de um subconjunto [106], [107]. Ambos os métodos são independentes do tipo de técnica de ML [109].

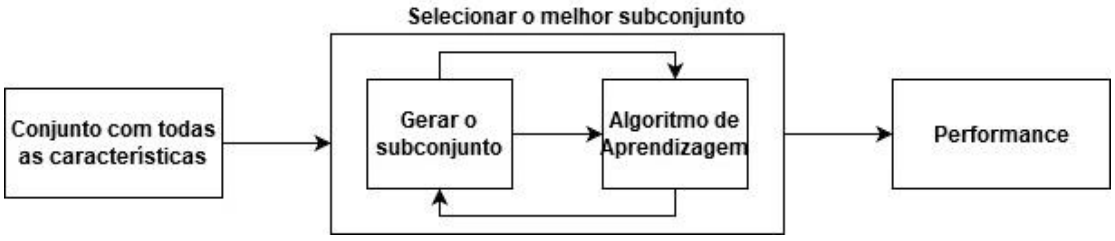


Figura 2.15 - Método *Wrapper* (adaptado de [108]).

Além dos métodos referidos, também existe o método *Embedded* (Figura 2.16) que combina as vantagens do método *Filter* com o método *Wrapper* [110]. Neste método, a seleção de características é uma parte integral do modelo de classificação, sendo baseada na pontuação calculada pelo modelo. O modelo de classificação é construído utilizando um conjunto de treino e pela avaliação da exatidão obtida na previsão no conjunto de teste. O modelo de classificação que foi construído fornece o valor da *performance* para cada característica. Características com pontuação mais baixa são removidas e continuam as mais elevadas, mas, ao contrário do método *Filter*, a medida de *performance* das características é calculada pelo modelo de classificação. Pelo facto de o modelo ser construído apenas uma vez para determinar a pontuação, apresenta um custo computacional mais baixo em comparação com o método *Wrapper*, que requer a construção de múltiplos modelos devido ao processo iterativo [111].

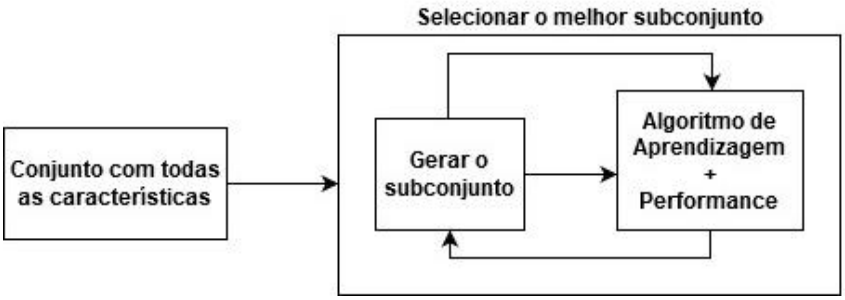


Figura 2.16 - Método *Embedded* (adaptado de [108]).

Portanto, a seleção de características permite a remoção de variáveis independentes que podem estar fortemente correlacionadas entre si e garante a manutenção das variáveis independentes que podem estar fortemente correlacionadas com a variável prevista ou dependente [107].

Algoritmos de aprendizagem

Após as características terem sido selecionadas, estas serão utilizadas no treino do algoritmo de aprendizagem, fixando diferentes valores de certos parâmetros. Para cada valor, o algoritmo irá selecionar uma hipótese diferente dentro de um conjunto de hipóteses. A hipótese escolhida será aquela que conseguir obter um melhor desempenho no conjunto de teste. Este desempenho pode ser avaliado através da utilização da função de custo associada a uma determinada tarefa. Esta função mede a diferença entre a *label* prevista pelo modelo e a *label* verdadeira. Por último, utilizando esta hipótese, será possível prever as *labels* das amostras no conjunto de teste [112].

As técnicas de aprendizagem supervisionada são muito utilizados em problemas que requerem processamento de dados "inteligente". Existe um conjunto muito alargado de métodos. As diferenças entre eles incluem complexidade computacional, capacidade de generalização e desempenho na classificação [113]. Seguidamente são apresentados alguns exemplos de algoritmos aprendizagem supervisionada.

• ***Support Vector Machine***

O SVM é um algoritmo de aprendizagem supervisionada que procura um hiperplano (uma superfície que tem menos uma dimensão do que as variáveis preditoras) capaz de separar classes. A Figura 2.17 apresenta um exemplo de um espaço de características bidimensional, cujo hiperplano é uma reta. A posição e a direção do hiperplano depende dos vetores de suporte: pontos que estão mais próximos da fronteira entre as duas classes [114]. A margem do hiperplano pode ser definida como a distância mínima entre um ponto do conjunto de treino e o hiperplano. Se o hiperplano tiver uma margem grande, então irá separar o conjunto de treino [115].

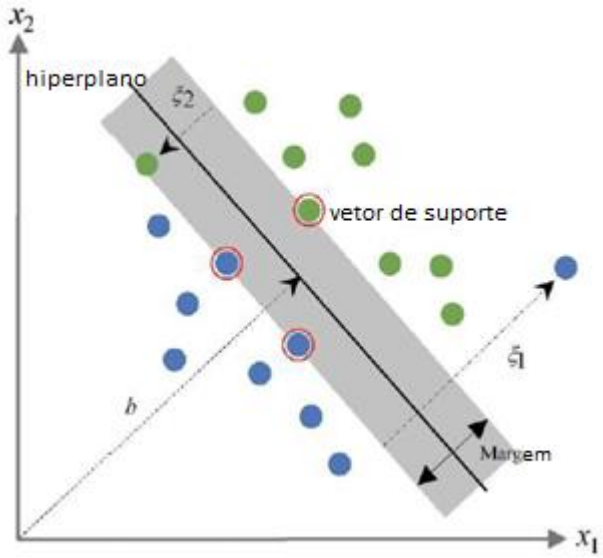


Figura 2.17 - Hiperplano de classificação do algoritmo SVM num espaço *input* bidimensional baseado na maximização das margens (adaptado de [116]**)**.

O objetivo deste algoritmo é encontrar o hiperplano que forneça a melhor capacidade de generalização. Para isso, a margem entre duas classes de dados deve ser a maior possível. Segundo a teoria do ML, a maximização da margem fornece ao SVM uma boa capacidade de generalização, porque minimiza o risco estrutural relacionado com a complexidade do SVM [117].

O algoritmo apresentado anteriormente apenas é aplicável em *datasets* que podem ser linearmente separáveis, ou seja, assume que existe um hiperplano capaz de separar corretamente o *dataset* em duas classes [112]. Quando o *dataset* não é completamente separável, é possível utilizar uma *soft-margin* que, além de encontrar o melhor hiperplano que separa as classes, também aceita pontos dentro da margem. Contudo, é penalizado quando existirem estes pontos. A severidade da penalidade é controlada por um hiperparâmetro que averigua se é *hard-margin* ou *soft-margin*. O *hard-margin* permite poucos pontos dentro da margem e o hiperplano dependerá de um reduzido número de vetores de suporte. O *soft-margin* permite mais pontos dentro da margem e o hiperplano dependerá de um maior número de vetores de suporte [114].

Em problemas lineares, a fronteira de decisão é um hiperplano, mas, em muitos problemas da vida real, a fronteira não é linear. Neste caso, os métodos empregues em problemas lineares são igualmente aplicáveis, contudo, é necessário que os dados sejam transformados de um espaço para outro de maior dimensão, de modo que a fronteira de decisão linear consiga separar

duas classes [118]. A transformação do espaço é conseguida à custa da função *kernel* que pode ser linear, radial (RBF, do inglês *Radial Basis Function*), sigmoide ou polinomial [116].

• ***Artificial Neural Networks***

As ANN compreendem um único ou um número de neurónios artificiais interconectados similar ao encontrado nos sistemas nervosos biológicos. A unidade fundamental de uma ANN é o *percetron* (Figura 2.18). Este elemento pode ter múltiplas entradas que são multiplicadas pelos respetivos pesos de conexão (pesos sinápticos). Estes pesos valorizam os *inputs* mais significativos, de modo a aumentar os seus impactos na reposta neuronal, para que, tal como acontece num neurónio biológico, sejam diferenciadas as forças sinápticas. Isto significa que cada *input* tem um peso associado. Posteriormente, as entradas ponderadas pelos pesos passam, juntamente com um *input* adicional *bias*, por uma função de combinação (somatório). No final, a função de transferência ou de ativação produz um resultado (*output*) [119], [120].

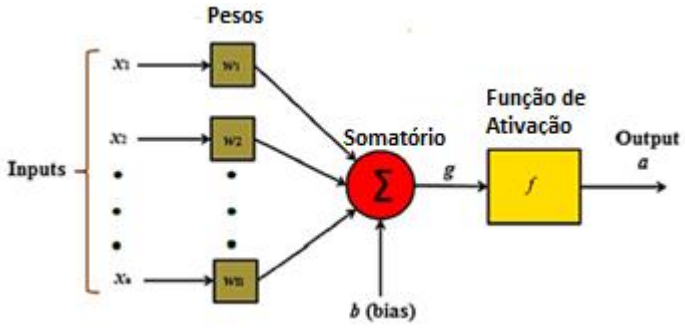


Figura 2.18 - Esquema de um *perceptron* (adaptado de [119]).

A função de ativação tem como objetivo normalizar os valores de *output* dentro de um intervalo. Este processo resulta numa melhor precisão em operações matemáticas executadas pela rede neuronal. As funções mais comuns são a função sigmoide logística, tangente hiperbólica, gaussiana e a função linear [121], [122].

É de salientar que os pesos sinápticos são coeficientes ajustáveis dentro da rede e regulam a força do sinal de entrada. Além disso, é possível ajustar as forças em resposta a diferentes conjuntos de treino, relacionando-as com a arquitetura de uma rede ou através das suas regras de aprendizagem [119], [120].

As ANN podem ser divididas em três partes, designadas por camadas: camada de entrada, camada oculta e camada de saída. A camada de entrada é responsável por receber dados, sinais, características ou medidas recolhidas do ambiente externo. As camadas ocultas, também designadas por intermediárias ou invisíveis, são constituídas por neurónios responsáveis por extrair padrões associados com o processo ou sistema a ser analisado. Estas camadas realizam grande parte do processamento interno da rede. A camada de saída é, também, constituída por um ou mais neurónios e é responsável por produzir e apresentar as saídas no final da rede, resultantes do processamento executado pelos neurónios das camadas anteriores [122]. Na Figura 2.19 está representada uma ANN com uma camada de entrada, duas camadas ocultas e uma camada de saída.

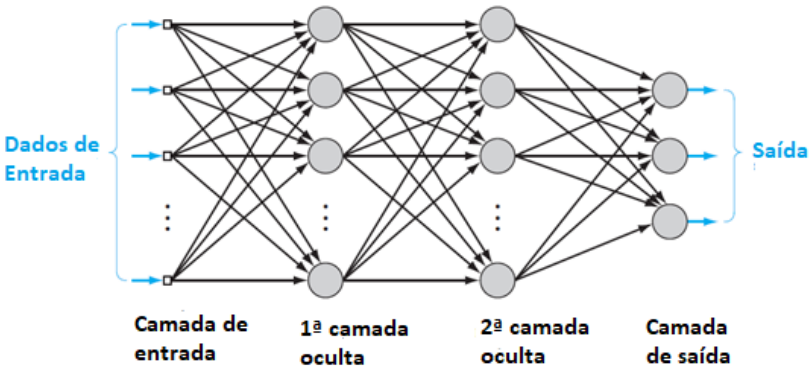


Figura 2.19 - *Artificial Neural Networks* com duas camadas ocultas (adaptado de [123]).

A forma como os neurónios artificiais estão interligados pode resultar em diferentes topologias possíveis. Essas topologias são agrupadas em dois tipos: *feedforward* e recorrente. Na topologia *feedforward*, a informação flui apenas numa única direção dos *inputs* para os *outputs*. Na segunda topologia, parte da informação não só flui dos inputs para os outputs como também flui na direção inversa [120]. Em termos práticos, a topologia *feedforward* é mais amplamente utilizada do que a recorrente em problemas do mundo real. Quando a rede neuronal com topologia *feedforward* tem múltiplas camadas ocultas pode ser designada por *multilayer feedforward network*, também conhecida por *Multilayer Perceptron* (MLP), topologia padrão das ANN [124].

O sistema de aprendizagem das ANN pode ser treinado utilizando variados métodos de adaptação como o *backpropagation*. Esta é a técnica mais comum para treinar redes neuronais com supervisão. O objetivo é ajustar iterativamente os pesos na rede neuronal para produzir um *output* desejado, por minimização do erro obtido na saída da rede [121].

• ***K-Nearest Neighbors***

O KNN é um algoritmo de aprendizagem supervisionada que se destaca pela sua simplicidade e eficácia [113]. Pode ser aplicado em problemas de classificação binários ou multi-classe [125] ou em problemas de regressão [81]. É um algoritmo não-paramétrico, ou seja, não tem parâmetros ou tem um número fixo de parâmetros independente do tamanho dos dados. Em vez disso, os parâmetros são determinados pelo tamanho do *dataset* de treino. Pode ser a melhor opção em estudos de classificação cujo conhecimento prévio é escasso ou inexistente acerca da distribuição dos dados.

Este algoritmo também é conhecido por ser "lazy", uma vez que todos os dados de treino são guardados e espera até que os dados de teste sejam produzidos, sem ter que criar um modelo de aprendizagem [126].

O objetivo do KNN é utilizar todos os dados disponíveis e classificar novos dados, baseados numa medida de similaridade. A ideia subjacente a este método é classificar todos os pontos não-classificados e atribuí-los a uma classe igual à classe que a maioria dos k vizinhos mais próximos pertence. O número de vizinhos mais próximos do novo ponto é dado pelo valor de k [127]. Na Figura 2.20 está representado um conjunto de dados dividido em duas classes (quadrados azuis e triângulos vermelhos) e pretende-se atribuir uma classe ao círculo verde (dentro do círculo interno). Se k=3 (círculo externo) então é atribuída a classe triângulos vermelhos porque há dois triângulos vermelhos e um quadrado azul dentro do círculo externo. Se, por

exemplo *k*=5, então irá pertencer à classe quadrados azuis (três quadrados contra dois triângulos fora do círculo externo) [128].

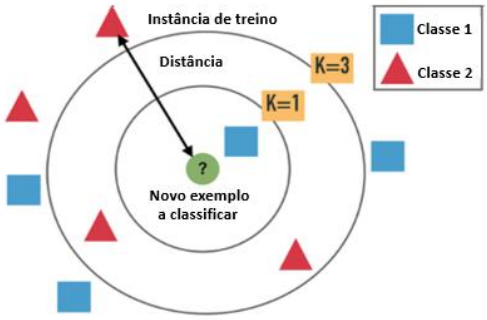


Figura 2.20 - Exemplo de classificação do algoritmo KNN. O círculo a verde é classificado consoante a classe do *k* vizinhos mais próximos (adaptado de [128]).

O método mais comum para medir a distância entre um ponto do conjunto de treino e o ponto não-classificado consiste no cálculo da distância euclidiana. Esta métrica é definida como a raiz das diferenças quadradas entre as coordenadas de um par de objetos [127].

• **Naive Bayes**

Os algoritmos *naive Bayes* (NB) pertencem a uma família de classificadores probabilísticos simples e são baseados na aplicação do teorema de Bayes que assume uma forte independência entre as características, daí o termo "naive". Este classificador calcula um conjunto de probabilidades através da contagem da frequência e combinação de valores dado um *dataset*. É uma técnica simples de construção de classificadores: modelos que atribuem *labels* a instâncias, representadas como vetores de características, onde as *labels* das classes são extraídas de um conjunto finito. Não se trata de um único algoritmo, mas uma família de algoritmos que têm como princípio assumir que o valor de uma característica é independente do valor de qualquer outra característica, dada a classe da variável [129], [130].

É uma técnica muito popular na filtragem de spam no email através do cálculo da probabilidade de um email ser uma mensagem spam ou não [94]. Algumas palavras têm uma probabilidade maior de ocorrer quando o email é spam ou não-spam. O algoritmo aprende as palavras mais comuns em cada um dos casos e calcula essa probabilidade [130].

• **Logistic Regression**

Logistic Regression (LR) é um algoritmo de aprendizagem supervisionada aplicável em problemas de classificação que prevê a probabilidade de uma determinada amostra do *dataset* pertencer a uma das classes do *target* [131]. Este algoritmo é útil quando se pretende prever respostas na forma booleana, 0/1 ou sim/não, ou seja, para estimar valores discretos. Neste caso é implementada a função sigmoide (ou logística) que opera dentro do intervalo entre 0 e 1 [81], [132].

• **Árvores de decisão**

DT são um modelo paramétrico, ou seja, é definido o modelo a partir de todo o espaço *input* e aprende os parâmetros ao considerar todos os dados de treino. Assim são utilizados o

mesmo modelo e o mesmo conjunto de parâmetros em qualquer *input* de teste. Trata-se de um modelo hierárquico aprendizagem supervisionada [133].

Um modelo de DT constrói uma árvore no sentido de classificar as instâncias, com base nos valores das características. Cada nó representa alguma decisão (condição de teste) sobre a característica da instância, enquanto cada ramo representa um valor possível para aquela característica. Com base no valor do nó, a árvore desce passando por outras derivações de nós até terminar na folha. Por fim, esta indica qual a categoria da classificação atribuída ou a decisão final. Utilizando este método, o foco está em como decidir qual a característica é o melhor classificador em nó [134], [135].

A Figura 2.21 mostra um exemplo de DT cujo objetivo é prever como fazer a deslocação para o trabalho cujo resultado pode ser ir de autocarro ou ir a pé. As questões (cor azul) representam os nós, as respostas são determinadas pelos ramos e a previsão obtida (cor laranja) é dada pela *label* da folha [136].

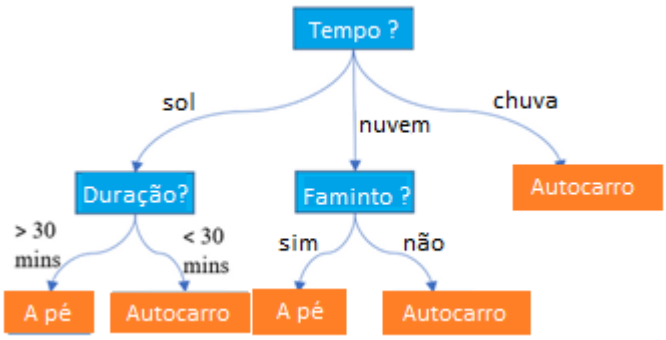


Figura 2.21 - Exemplo de uma árvore de decisão (adaptado de [136]).

Existem algoritmos que podem ser construídos a partir de mais do que uma árvore de decisão e designam-se por métodos *ensemble*. É o caso do algoritmo RF que é caracterizado por selecionar aleatoriamente as características. Este classificador prevê a *label* da classe de um dado de entrada pela maioria de votação em previsões realizadas por um conjunto de classificadores em árvore. Uma parte dos dados é guardada para o teste e o restante para o treino. Este método é designado por *leave-one-out cross-validation* (LOOCV). Múltiplas DT são construídas e, em cada uma, é escolhido aleatoriamente e criado um subconjunto do conjunto de treino com um certo número de amostras. Cada classificador em árvore divide o espaço de características num determinado número de partições e o resultado para um dado de entrada é determinado de acordo com a partição em que os dados estão [137].

Na Tabela 2.7 são apresentados, para cada algoritmo, os tipos de problemas em que podem ser aplicados, assim como as principais vantagens e desvantagens.

Tabela 2.7 - Algoritmos: tipos de problemas e principais vantagens e desvantagens.

Algoritmo	Problema	Vantagens	Desvantagens
SVM	Classificação e regressão [134]	Trabalha bem em *datasets* grandes e pequenos [138].	Não faz um escalamento correto do *dataset*; Em *datasets* grandes pode ter problemas em termos de tempo de execução e utilização da memória; Requer um pré-processamento dos dados atento e um afinamento dos parâmetros; Difícil de inspecionar e entender o porquê de uma previsão [138].
ANN	Classificação e regressão [138]	Capaz de construir modelos bastante complexos (em *datasets* grandes) [138].	Sensível ao escalamento do *dataset* e na escolha dos parâmetros; Modelos grandes precisam de mais tempo de treino [138].
KNN	Classificação e regressão [134]	Simples; Sem custo computacional durante a aprendizagem, mas sim durante a previsão; Não faz suposições sobre os dados e suas distribuições [114].	Não consegue trabalhar com variáveis categóricas; Pode ser computacionalmente caro se o conjunto de treino for grande; Não é possível interpretar o modelo em termos de relações nos dados do mundo real; A exatidão da previsão pode ser afetada pelo ruído e *outliers*; Fraco desempenho em *datasets* de grandes dimensões [114].
NB	Classificação [138]	Rápido no treino e na previsão; Fácil de entender o procedimento de treino; Trabalha muito bem com poucos dados e é relativamente robusto aos parâmetros [138].	Aplicável apenas em problemas de classificação; Muitas vezes, menos exato do que outros modelos lineares [138].
LR	Classificação [139]	Rápido no treino e na previsão; Trabalha bem em grandes *datasets* e em dados dispersos; Fácil de entender como a previsão é feita [138].	Não trabalha corretamente se não houver uma total separação entre as classes; Assume que as classes são linearmente separáveis [114].
DT	Classificação e regressão [81]	Fácil de interpretar; Consegue lidar com variáveis categóricas e contínuas; Não faz suposições sobre a distribuição das variáveis preditoras; Consegue lidar com valores omissos de forma sensata; Consegue lidar com variáveis contínuas em diferentes escalas [114].	Árvores individuais são suscetíveis ao *overfitting* [114].

De modo geral, quando estamos a trabalhar num novo *dataset*, é preferível começar com um modelo simples como NB ou KNN, para ver até onde chegam. Após uma melhor compreensão dos dados, deve ser considerado um algoritmo que consiga construir modelos mais complexos como DT ou ANN [138].

Ainda relacionado com o processo de aprendizagem, um modelo de ML pode ser discriminativo ou generativo. Tanto SVM como DT podem aprender a partir da fronteira de decisão

dentro do *dataset*, logo é discriminativo. Modelos como NB, que aprendem a partir da distribuição da probabilidade dos dados, são generativos [140].

Treino e avaliação da classificação

Na aprendizagem supervisionada, o objetivo é construir um modelo através dos dados de treino, de modo a ser capaz de fazer previsões exatas em dados nunca anteriormente vistos pelo classificador com as mesmas características que as usadas no conjunto de treino. Se o modelo for capaz de fazer previsões exatas em dados não vistos anteriormente, significa que este é capaz de generalizar a partir do conjunto de treino para o conjunto de teste (Figura 2.22 b). O modelo construído deve ser capaz de generalizar com a maior exatidão possível. Se o conjunto de treino e o conjunto de teste forem semelhantes, espera-se que o modelo seja exato no conjunto de teste. Porém, há casos em que isso não acontece. O *overfitting* ocorre quando o modelo é excessivamente complexo para a quantidade de informação existente e, por isso, adapta-se em demasia às particularidades do conjunto de treino, resultando numa boa performance no conjunto de treino, mas, depois, não é capaz de generalizar para novos dados (Figura 2.22 c).

Por outro lado, o *underfitting* acontece quando o modelo é demasiado simples, não sendo capaz de capturar todos os aspetos e toda a variabilidade nos dados, resultando num mau desempenho do modelo, inclusive no conjunto de treino (Figura 2.22 a).

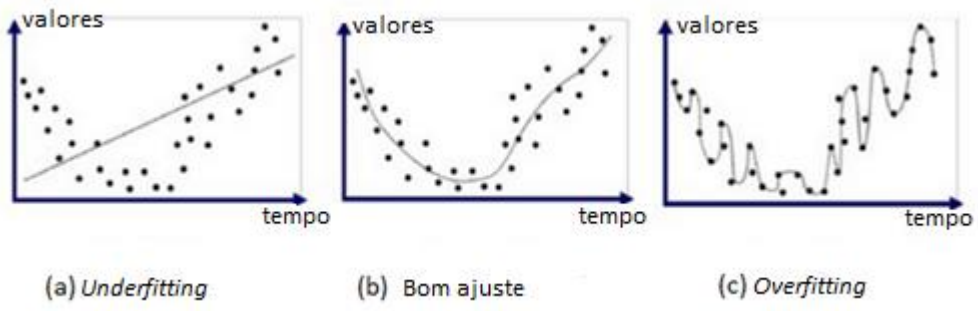


Figura 2.22 - Exemplo de *underfitting* e *overfitting* (adaptado de [141]).

Portanto, quanto mais complexo for o modelo, maior será a capacidade de prever nos dados de treino. Mas, se for excessivamente complexo, começará a focar-se exageradamente em cada ponto dos dados do conjunto de treino e o modelo não irá generalizar corretamente em novos dados [138].

Otimização do modelo

O treino de técnicas de ML permite que sejam encontrados os melhores parâmetros de aprendizagem assim como a estrutura que maximiza a classificação ou a exatidão da previsão. Este processo é designado por otimização [97].

Além disso, para que um modelo de ML possa ser aplicado em diferentes problemas, os hiperparâmetros devem ser afinados. Hiperparâmetros são parâmetros que são utilizados para configurar um modelo de ML ou para especificar o algoritmo utilizado que minimize a função de custo. Portanto, é selecionando um conjunto de hiperparâmetros utilizados pelo algoritmo no modelo para que a *performance* do modelo seja a melhor possível no que diz respeito à

exatidão da previsão. Para isso é, geralmente, utilizado o método *grid search* que busca exaustivamente a melhor configuração, através da avaliação de todas as combinações de hiperpâmetros possíveis [124], [142].

Validação

Uma das técnicas de validação do modelo utilizada para estimar como um modelo irá atuar quando generalizado é a validação cruzada (VC). Esta técnica é utilizada, muitas vezes, em processos iterativos cujo objetivo final é otimizar o modelo e garantir que este não se ajuste excessivamente aos dados, para que o modelo possa generalizar bem com novos dados e fazer boas previsões. Durante a VC, os dados são divididos em dados de treino e dados de teste e, seguidamente, os dados de treino são utilizados para treinar o modelo e, através dos dados de teste, é avaliado o seu desempenho. Este processo é repetido de forma iterativa e o modelo escolhido será o melhor do conjunto.

Existem dois tipos de VC: *holdout* e *K-fold*. O primeiro caso é uma técnica de VC simples, na qual o conjunto de teste é usado para calcular o erro absoluto médio do teste. Através do erro acumulado é efetuada a avaliação do modelo [124]. Na VC K-*fold*, o *dataset* é dividido em K subconjuntos separados (tipicamente K=10) e o método de validação é invocado K vezes. Em cada iteração, é utilizado um dos K subconjuntos como conjunto de teste e os subconjuntos K-1 restantes como conjunto de treino [102]. Posteriormente, é calculada a métrica, escolhida previamente, para avaliar a *performance* (p. ex., Exatidão), repete K vezes e, por fim, é obtida a média dos resultados [140].

Na Figura 2.23 é apresentado um exemplo de VC *5-fold*, ou seja, neste caso o processo é repetido cinco vezes, sendo que em cada iteração, são criados cinco subconjuntos de treino e de teste diferentes.

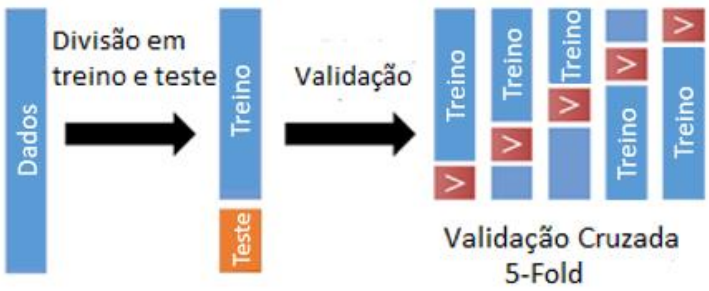


Figura 2.23 -Técnica de VC *5-fold* (adaptado de [143]).

Métricas de avaliação

Existem diferentes métricas para avaliar a performance do classificador [118]. Seguidamente serão apresentadas as métricas mais utilizadas em problemas de classificação.

Em problemas binários, a melhor forma de avaliar o sucesso ou insucesso de um classificador durante o seu treino é através de uma matriz de confusão. Neste caso, a matriz terá duas linhas e duas colunas, tal como exemplifica a Tabela 2.8. No exemplo apresentado, as linhas correspondem à classe prevista pelo algoritmo e as colunas correspondem à classe verdadeira. A partir da matriz é possível obter o número de verdadeiros positivos (VP), ou seja, o número de instâncias positivas que foram corretamente classificadas como positivas e o número de verdadeiros negativos (VN), isto é, o número de instâncias negativas que foram corretamente classificadas como negativas. Além disso, através da matriz de confusão, é possível obter o número de instâncias negativas incorretamente classificadas como positivas (FP) e número de instâncias positivas incorretamente classificadas como negativas (FN) [125], [144].

Tabela 2.8 - Exemplo de matriz de confusão para um problema de classificação binária (adaptado de [125]).

	Classe Verdadeira	
Classe Prevista	**1**	**0**
1	VP	FP
0	FN	VN

Por outro lado, a Exatidão (*ACC)* é a métrica mais utilizada quer em problemas de classificação binários quer multi-classe, permitindo avaliar a qualidade da solução produzida com base na percentagem de previsões corretas sobre o total de instâncias [88], [144]:

$$ACC = \frac{VP + VN}{VN + FP + FN + VP} \tag{2.1}$$

A Precisão (*PREC)* é a medida mais utilizada para averiguar a qualidade dos resultados. No contexto da classificação binária, a Precisão representa o número de verdadeiros positivos (p. ex. , o número de instâncias corretamente classificadas como pertencentes à classe 1) dividido pela soma de verdadeiros positivos e falsos positivos (p. ex. , o número de instâncias que foram incorretamente classificadas como pertencentes à classe 1). Então, uma Precisão de 1.0 (ou 100 %) é conseguida se cada instância prevista pelo classificador como pertencente à classe 1 for, na verdade, pertencente à classe 1 (ou seja, não há falsos positivos) [88], [135]:

$$PREC = \frac{VP}{VP + FP} \tag{2.2}$$

O *Recall* (ou sensibilidadade) define-se como o número de verdadeiros positivos dividido pela soma de verdadeiros positivos e falsos negativos (p. ex., o número de instâncias que pertenciam à classe 1, mas que foram classificadas como classe 0 pelo modelo). Se o *Recall* for 1.0 (ou 100%) significa que o modelo não falhou em nenhuma instância que estava na classe 1 (isto é, não há falsos negativos) [88], [135]:

$$Recall = \frac{VP}{VP + FN} \tag{2.3}$$

Geralmente, a precisão e o *Recall* estão inversamente relacionados, uma vez que, uma elevada Precisão está associada a um baixo *Recall* e vice-versa [135].

A Especificidade (SPEC) dá-nos a proporção total de verdadeiros negativos corretamente classificados pelo número total de instâncias que são realmente negativas [124]:

$$SPEC = \frac{VN}{FP + VN} \tag{2.4}$$

A *F-Measure* (*FM*), também designada por *F1 score* representa a média harmónica entre o *Recall* e a Precisão [88], [144]:

$$FM = \frac{2.Recall.PREC}{Recall + PREC} = \frac{2VP}{2VP + FP + FN} \tag{2.5}$$

A curva ROC (*Receiver Operating Characteristic*) é um tipo de curva bidimensional que permite visualizar a performance de um classificador. O eixo dos XX representa a fração de falsos positivos (*False Positive Rate*, FPR) e o eixo dos YY representa a fração de verdadeiros positivos (*True Positive Rate*, TPR). O TPR também é conhecido por sensibilidade e o FPR equivale a 1 - especificidade [145].

Por outro lado, a partir da curva ROC é possível obter uma métrica independente da aplicação, considerada como uma das propriedades mais importantes de um classificador. Esta métrica designa-se por área sob a curva ROC (*Area Under Curve*, AUC). Quanto maior for a área, melhor é o classificador [87]. Na Figura 2.24 está representado um exemplo de curva ROC e o AUC correspondente.

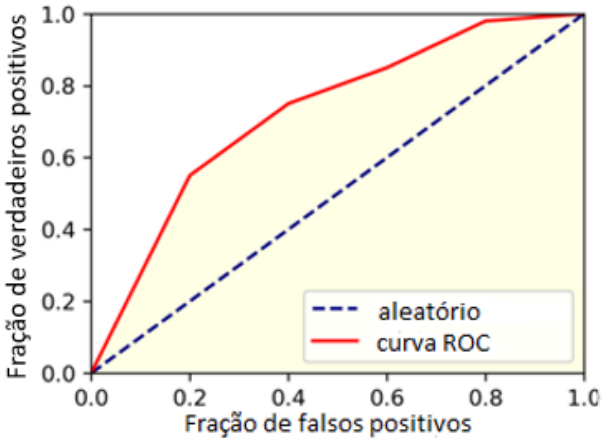


Figura 2.24 - Exemplo de uma curva ROC (a vermelho) e o AUC correspondente (a amarelo) (adaptado de [146]).

Se o classificador for perfeito, o AUC será 1,0. Mas se o classificador for aleatório nas suas suposições, então o AUC será 0,5. No mundo real, não se espera que um classificador obtenha um AUC de 1,0, contudo se variar entre 0,6 e 0,9, então é considerado como um bom classificador [145].

2.4 – Resumo

Ao longo deste capítulo foram abordados alguns conceitos que permitem uma melhor compreensão da DP, da macha humana e de técnicas de ML.

Embora a DP tenha sido descrita, pela primeira vez, em 1817, por James Parkinson, atualmente, ainda não existe uma terapia capaz de abrandar a sua progressão ou até prevenir a sua manifestação. Além disso, os mecanismos subjacentes à patogénese da DP ainda não são totalmente compreendidos. Sabe-se que esta doença apresenta numerosos sintomas não-motores, que se podem manifestar em primeiro lugar, mas só quando surgem os sintomas motores é possível estabelecer o diagnóstico clínico.

À medida que a doença progride, o doente de Parkinson apresenta uma postura em flexão cada vez mais acentuada e uma redução do movimento dos músculos do tronco e dos membros superiores e inferiores que, por consequência, comprometem a marcha. Para determinar a presença e a severidade da DP, os médicos fazem uma avaliação subjetiva da marcha com base em escalas, tais como MDS-UPDRS e HY.

Nos últimos anos foram desenvolvidos métodos de análise da marcha baseados na aquisição de dados a partir de sistemas *wearable* e não-*wearable*. Estes sistemas são capazes de medir parâmetros da marcha de forma objetiva e fidedigna.

Outro avanço significativo que se assistiu nas últimas décadas foi o desenvolvimento da IA e do ML. Atualmente, o ML forma a base da maioria das aplicações que conhecemos em IA, desde algoritmos de reconhecimento facial e de voz até a identificação de tumores e doenças. Nestas aplicações de software são executados algoritmos que fazem previsões automáticas de resultados sem serem explicitamente programados. Através de um processo de aprendizagem, o algoritmo é treinado e testado num conjunto de dados até ser encontrado um modelo que maximize o seu desempenho. Desta forma, o modelo tem potencial para efetuar previsões de qualidade em dados futuros.

Capítulo 3

Machine Learning na Avaliação da Doença de Parkinson: Estado da Arte

Em virtude dos grandes avanços tecnológicos que assistimos na presente época, a identificação de condição anormal de saúde é da maior importância. No caso da DP, o diagnóstico clínico assenta em vários exames (p. ex., ressonância magnética, EEG, entre outros), de forma a avaliar os sintomas motores do paciente e, depois, atribuir um nível de severidade, com base numa escala (p. ex., UPDRS, HY) [6]. Contudo, os sintomas são analisados com base no julgamento individual, pelo que a escala UPDRS apresenta um problema de inconsistência entre avaliadores. Além disso, a pontuação da UPDRS é calculada quando os pacientes visitam os centros de saúde, mas os sintomas variam de dia para dia. Uma consulta de curta duração não é suficiente para monitorizar os sintomas e os efeitos da medicação de forma precisa. Portanto, há uma necessidade de monitorizar o paciente remotamente, a fim de verificar os sintomas de uma forma mais consistente e eficiente [147].

A utilização de medidas clínicas subjetivas na avaliação da DP pode comprometer a validade e a fiabilidade do diagnóstico. Por esse motivo, é fundamental recorrer a métodos quantitativos, tais como a análise dos parâmetros da marcha, a fim de diagnosticar a DP na fase inicial.

Por outro lado, o principal objetivo do ML é fornecer interpretação direta e extrair padrões úteis nos dados processados, através da conceção de algoritmos que aprendem por experiência nos dados etiquetados [6]. Baseado nas alterações do padrão de marcha, o ML pode ser introduzido para identificar padrões de marcha de pessoas em diferentes estágios da DP e distinguilos de pessoas saudáveis [148].

No presente capítulo é apresentada a revisão da literatura realizada nas bases de dados científicas mais relevantes (Secção 3.1). Uma vez que a DP apresenta uma grande variedade de sinais e sintomas, são exploradas e comparadas as diferentes tecnologias de diagnóstico por computador (Secção 3.2). No final é apresentado um resumo do capítulo (Secção 3.3).

3.1 - Revisão sistemática da literatura

A pesquisa das publicações científicas foi realizada entre outubro de 2020 e janeiro de 2021, nas seguintes bases de dados de referência: Pubmed[1], Enginnering Village[2], IEEE Xplore[3] e Scopus[4]. A Tabela 3.1 apresenta os termos relacionados com o tema da dissertação que foram pesquisados nas referidas bases de dados.

Tabela 3.1 - Termos pesquisados nas bases de dados de referência para a pesquisa de artigos relacionados com o tema de Dissertação.

Base de dados	Termos pesquisados
Pubmed Engineering Village IEEE Xplore Scopus	“parkinson's disease” AND “machine learning” AND gait AND (prediction OR detection OR diagnosis)

Na Scopus foram encontrados 120 artigos relacionados com os termos: “parkinson's disease” AND “machine learning” AND gait AND (prediction OR detection OR diagnosis). Os resultados obtidos podem ser analisados na Figura 3.1. É possível constatar uma tendência crescente no número de publicações científicas publicadas, a partir de 2013, embora, entre 2016 e 2017, tenha havido uma estabilização no número. Nos últimos anos é evidente um maior interesse por parte dos investigadores nesta área.

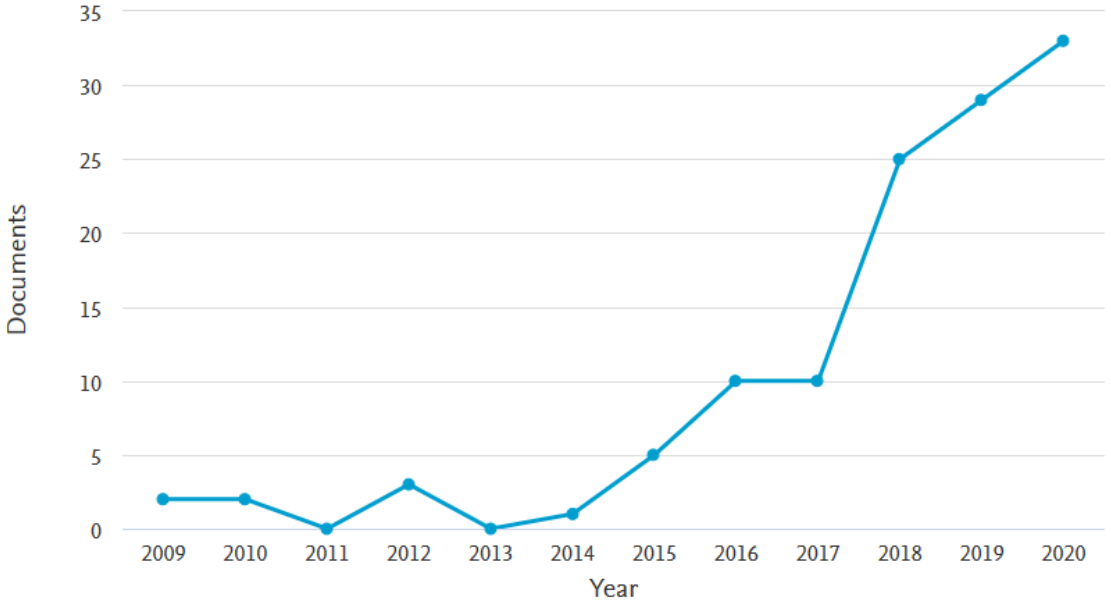


Figura 3.1 - Número de publicações científicas, por ano, na Scopus, a partir dos termos pesquisados.

Tal como apresentado na Subsecção 2.1.2 os sintomas na DP dividem-se em motores e não-motores. Tendo em consideração os sintomas que diagnosticam a DP, o tipo e a origem dos dados, é possível subdividir as tecnologias de diagnóstico assistido por computador, para a

[1] https://pubmed.ncbi.nlm.nih.gov
[2] https://www.engineeringvillage.com
[3] https://ieeexplore.ieee.org
[4] https://www.scopus.com

classificação, previsão e monitorização dos pacientes, em seis subcategorias: análise baseada na voz, análise baseada em sistemas *wearable* e não-*wearable*, análise baseada na escrita e análise baseada em dados de EEG e outros dados clínicos, que incluem a avaliação do olfato e do sono, imagens de ressonância magnética, entre outros. A análise dos dados provenientes dos sistemas *wearable* e não-*wearable* incluem sintomas motores, tais como as alterações da marcha [147].

Para encontrar estudos baseados em cada sintoma, foram alterados os termos pesquisados para a seguinte forma: "parkinson's disease" AND "machine learning" AND (voice OR gait OR handwritting OR EEG OR imaging OR sleep OR olfact) AND (prediction OR detection OR diagnosis). Alguns dos estudos encontrados são descritos nas Subsecções 3.1.1 a 3.1.4 .

3.1.1 – Análise baseada na voz

Um dos sinais mais comuns da DP é a dificuldade na voz, atingindo 90% das vítimas de DP na fase inicial desta doença. Os sistemas de diagnóstico especializados em distúrbios vocais são uma das áreas mais importantes na deteção da DP [149].

Os autores Hasan e Hasan [149] desenvolveram uma técnica de seleção de caraterísticas, designada por análise de variância (ANOVA, do inglês *Analysis of Variance*), para selecionar as 50 caraterísticas mais relevantes do *dataset* proveniente do UCI *Machine Learning repository*, de acordo com o *F-score* da ANOVA. O *dataset* contém 188 pacientes com DP, dos quais 107 são homens e 81 são mulheres, com idades compreendidas entre 33 e 87 anos. Também tem um grupo de controlo constituído por 64 indivíduos saudáveis (23 homens e 41 mulheres), com idades que variam entre 41 e 82 anos. Para a aquisição de dados, foi utilizado um microfone com uma frequência de resposta de 44,1 kHz. Cada participante emitiu a vogal/a/letra em 3 repetições. O *dataset* foi dividido em 20% para o conjunto de treino e 80% para o conjunto de teste. O algoritmo RF obteve melhor ACC do que os algoritmos KNN, DT, LR, NB, SVM (*kernel* linear e RBF), com 91%, uma PREC de 93%, um *Recall* de 96%, um *F1-score* de 94% e um AUC de 83%. Apenas o algoritmo DT obteve melhor AUC com 86%.

Por outro lado, Avuçlu e Elen [150] propuseram um sistema de apoio à decisão clínica que auxiliasse médicos a diagnosticar a DP. Para isso, utilizaram uma base de dados que continha um conjunto de medições biomédicas da voz de 31 pessoas, das quais 23 com DP. No total foram recolhidas 195 amostras de som de cada participante, numa média de 6 gravações. O *dataset* foi dividido num conjunto de treino e num conjunto de teste e 11 combinações diferentes foram testadas (45 × 55, 50 × 50, 55 × 45, 60 × 40, 65 × 35, 70 × 30, 75 × 25, 80 × 20, 85 × 15, 90 × 10, 95 × 5). Os resultados obtidos foram comparados com a VC 10-*fold*. Os classificadores utilizados foram: KNN, RF, NB e SVM com *kernel* polinomial. Os melhores resultados foram obtidos quando o *dataset* foi dividido em 75% para o conjunto de treino e 25% para o conjunto de teste. O conjunto de treino classificado pelo algoritmo KNN obteve 100% de sucesso. Já o algoritmo RF foi o mais bem-sucedido no conjunto de teste com 85,81%.

Na Tabela 3.2 é apresentado um resumo dos pontos mais importantes encontrados nos estudos mencionados anteriormente, nomeadamente, os resultados obtidos na classificação para cada algoritmo.

Tabela 3.2 - Resumo dos estudos encontrados na literatura científica relacionados com análise baseada na voz.

Referência	No. atributos	Algoritmo	*ACC*	*Recall*	FM	Validação
(Hasan e Hasan, 2020)[149]	50	RF	91%	96%	94%	VC *holdout* (20-80)
		KNN	85%	90%	89%	
		DT	86%	87%	91%	
		LR	88%	97%	93%	
		NB	84%	88%	89%	
		SVM (linear)	84%	92%	91%	
		SVM (RBF)	86%	96%	91%	
(Avuçlu e Elen, 2020)[150]	23	KNN	85,64%	76,45%	78,76%	VC 10-*fold*
		NB	70,26%	76,76%	68,07%	
		RF	85,13%	83,82%	81,32%	
		SVM	88,72%	84,10%	84,58%	

3.1.2 – Análise baseada em sistemas não-*wearable e wearable*

Várias tecnologias podem ser aplicadas na análise quantitativa da marcha, subdividindo-se em dois tipos: não-*wearable* e *wearable* [7], tal como apresentado na subsecção 2.2.2 . Estas tecnologias podem ajudar clínicos a medir os sintomas motores com maior precisão em comparação com a avaliação clínica subjetiva (inclui escalas UPDRS e HY) [147].

Os autores Buongiorno *et al*. [151] desenvolveram um sistema de baixo custo com o intuito de analisar os parâmetros espaciotemporais e angulares, através do sensor Microsoft Kinect v2, para identificação e classificação da severidade da DP. No estudo participaram 16 pacientes com DP, cuja avaliação da severidade da DP foi efetuada segundo a escala MDS-UPDRS, e 14 indivíduos saudáveis. Foram colocados 25 marcadores refletores para localizar e registar a posição do corpo.

Umas das tarefas analisadas foi a marcha. Enquanto o participante andava, uma câmara, (sensor Kinect) posicionada a 3,5 m de distância do ponto inicial da marcha, captava o seu movimento. Através do Microsoft SDK foi construída automaticamente a posição 3D dos 25 marcadores e extraídas 16 caraterísticas (temporais, espaciais e angulares). Para a classificação das caraterísticas foram treinados 2 conjuntos de dados, um conjunto com 6 caraterísticas e outro com 9 caraterísticas, tendo sido selecionadas, em ambos os casos, a partir do software *Weka*, com base no classificador *CfsSubsetEval* e no método de pesquisa *BestFirst*. A fim de avaliar o modelo treinado, foi usada a VC 5-*fold*. O algoritmo ANN obteve um melhor desempenho, com uma ACC de 89,4% (para 9 caraterísticas) e 95% (para 6 caraterísticas).

Por outro lado, os autores Wahid *et al*. [152] realizaram um estudo cujo objetivo consistia, em primeiro lugar, na normalização das caraterísticas espaciotemporais a partir de equações standard adimensionais e da regressão linear múltipla, no sentido de encontrar diferenças entre os pacientes com DP e o grupo de controlo. Posteriormente, foi testada a eficácia da aplicação de ferramentas de ML na classificação da marcha parkinsoniana. A normalização teve em consideração as caraterísticas do paciente: idade, altura, massa corporal, género e velocidade de marcha preferencial. Para o estudo participaram 23 pacientes com DP e 26 indivíduos saudáveis (grupo de controlo) com idades equivalentes. Foram utilizados 15 marcadores refletores,

segundo o modelo Helen Hayes modificado. Cada indivíduo realizou uma marcha de 8 m com uma velocidade da sua preferência, repetindo-a 10 vezes. Seguidamente foi utilizado um sistema com 8 câmaras de vídeo de análise de movimento (Vicon, Oxford Metrics, Ltd., Oxford), que adquiria os dados em simultâneo com duas plataformas de força, para a obtenção da FRS. Os resultados obtidos, utilizando os dados normalizados pela regressão linear múltipla, foram superiores, em comparação com o outro método. A ACC obtida no algoritmo RF foi 92,6%, em comparação com 80,4% (SVM), 86,2% (KFD, *kernel Fisher discriminant*), 84,4% (KNN) e 82,0% (NB). Os parâmetros internos dos classificadores foram otimizados através do método *Grid Search* e da VC 10-*fold*.

Outro estudo, publicado por Rehman *et al*. [5], explora quais as caraterísticas espaciotemporais da marcha clinicamente relevantes que devem ser consideradas para a classificação da DP. Os objetivos deste estudo consistiam no desenvolvimento de um modelo de ML para aplicar nas caraterísticas, de modo a distinguir indivíduos com DP dos saudáveis e na identificação da melhor combinação de caraterísticas para classificar a DP na fase inicial. Neste estudo participaram 303 indivíduos, 119 com DP no estado *on* e 184 pertencentes ao grupo de controlo. Além disso, foram registados parâmetros demográficos, tais como idade, altura, peso e IMC. Os sintomas motores da DP foram avaliados segundo a MDS-UPDRS-III e a escala HY, enquanto a cognição foi avaliada segundo o questionário MMSE. Cada participante foi instruído para caminhar com uma velocidade da sua preferência, de modo contínuo durante 2 minutos, num circuito oval de 25 m, constituído por um tapete instrumentado (GAITRite®). No total, foram extraídas 16 caraterísticas, a partir do Microsoft Access. Para a classificação foram testados 12 algoritmos, com ajuda da VC 10-*fold*. No final foram escolhidos 3 modelos de diferentes domínios (RF, LR, SVM com *kernel* RBF) e identificadas cinco caraterísticas mais relevantes: velocidade média do passo, comprimento médio do passo, variabilidade do comprimento do passo, largura média do passo e variabilidade da largura do passo. A classificação da DP na fase inicial obteve uma ACC que variou entre 73 e 97%, um *Recall* entre 63% e 100% e uma SPEC entre 79% e 94%.

Os autores Urcuqui *et al*. [153] desenvolveram um modelo de ML para classificar a marcha em pessoas diagnosticadas com DP. Foram extraídos parâmetros espaciotemporais a partir de câmaras de RGBD (MS Kinect) para captura de movimento e a marcha foi realizada num corredor equipado com Kinect. Participaram 60 voluntários (24 mulheres e 36 homens), com idades compreendidas entre 44 e 88 anos, agrupados pelo género e idade. O grupo com DP continha 18 homens e 12 mulheres. Cada participante repetiu 3 vezes a marcha e os dados foram adquiridos a partir de três pontos localizados em cada tornozelo. Os algoritmos escolhidos para realizar a classificação foram: LR, NB, RF e DT. O algoritmo RF obteve melhor ACC (82%), com uma baixa FPR (12%) e baixa TPR (23%). O comprimento da passada e a idade foram as melhores variáveis para treinar o modelo DT, com uma ACC de 94%.

Na Tabela 3.3 é apresentado um resumo dos estudos que utilizaram sistemas *não-wearable*, incluindo uma descrição detalhada do tipo de análise de marcha, os parâmetros da marcha extraídos e os resultados da classificação obtidos, de acordo com o algoritmo e o número de atributos.

Tabela 3.3 - Resumo dos estudos encontrados na literatura científica relacionados com análise da marcha utilizando sistemas *não-wearable*.

Referência	Técnica de análise	Parâmetros da Marcha	Algoritmo	No. atributos	*ACC*	*Recall*	*SPEC*
(Buongiorno *et al.*, 2019) [151]	Captura de movimento baseada em marcadores	Espaciotemporais; Angulares	SVM	9	78,5%	81,7%	74,8%
				6	88,7%	78,9%	96,3%
			ANN	9	89,4%	87,0%	91,8%
				6	95%	90,0%	99,0%
(Wahid *et al.*, 2015) [152]	Captura de movimento baseada em marcadores; Plataformas de força	Espaciotemporais	RF	9	92,6%	96%	90%
			SVM		80,4%	73%	100%
			KFD		86,2%	86%	91%
			NB		82,0%	85%	81%
			KNN		84,4%	69%	100%
(Rehman *et al.*, 2019)[5]	Tapete instrumentado (GAITRite®)	Espaciotemporais	RF	16	94,28%	100%	89%
				5	97,14%	100%	94%
			LR	16	82,85%	71%	89%
				5	84,99%	76%	94%
			SVM-RBF	16	81,92%	71%	89%
				5	85,71%	79%	94%
(Urcuqui *et al.*, 2018)[153]	Captura de movimento sem marcadores	Espaciotemporais; cinemáticos	RF	17	82%	88,2%	76,5%
			DT	2	94%	76,5%	35,3%

De outro modo, as técnicas de ML podem ser combinadas com sistemas *wearable* para detetar e avaliar os sintomas motores da DP. É o exemplo dos autores Aich *et al.* [154], que desenvolveram e validaram um algoritmo baseado em dados adquiridos por um acelerómetro, durante a marcha. Este estudo inclui caraterísticas de sinal e parâmetros espaciotemporais. Através destas caraterísticas pretenderam classificar o estado da medicação, ou seja, estado *on* ou *off,* em 20 pacientes com DP. Para estimar a severidade da doença foram utilizadas as escalas UPDRS-III e HY. Os dados foram extraídos em duas "condições" de marcha: no estado *on*, 1 a 1,5 horas depois da toma da medicação dopaminérgica e no estado *off*, no mínimo, 12 horas depois da toma da medicação dopaminérgica. Para cada tarefa, foi colocado um acelerómetro triaxial em cada joelho do participante. A partir do acelerómetro foram adquiridos dois tipos de caraterísticas: caraterísticas estatísticas (média, desvio-padrão, desvio absoluto mediano, mínimo, máximo, energia, intervalo interquartil, área de magnitude do sinal, *skewness* e curtose) e parâmetros da marcha (comprimento do passo, comprimento da passada, duração do passo, duração da passada, velocidade do passo e velocidade da passada). Para a validação do modelo foi utilizada a VC LOSO (*leave-one-subject-out*). Os algoritmos de ML escolhidos foram RF, SVM, KNN e NB. O classificador RF obteve o melhor resultado com uma ACC de 96,72%, um *Recall* de 97,35% e uma PREC média de 96,92%, na deteção dos dois estados.

Por outro lado, Mittra e Rustagi [155] publicaram um estudo no qual foi analisada a marcha com a finalidade de distinguir os pacientes com DP dos indivíduos saudáveis. Para a sua concretização foi utilizada a base de dados da *Physionet* designada por *Gait Analysis Database*. Esta base de dados consiste em 93 pacientes com DP idiopática e 73 indivíduos saudáveis. As características foram extraídas a partir de 8 sensores colocados em cada planta do pé do participante, para assim calcular a componente vertical das FRS. Cada participante andou durante 2 minutos. De modo a reduzir o volume de dados, foram aplicadas funções estatísticas, seguido do escalamento dos dados em -1 e 1. O algoritmo KNN obteve a melhor ACC (93,08%), em comparação com os algoritmos DT, RF, SVM (com *kernel* linear, RBF e polinomial) e LR. A PREC, o *Recall* e o F1-score obtidos pelo KNN foram 89,58%, 84,31% e 86,86%, respetivamente.

Noutro estudo, os autores Balaji *et al*. [156] utilizaram a mesma base de dados que [155], no sentido de diagnosticar automaticamente a DP e apresentar o nível de severidade. Esta base de dados contém 10 pacientes com o estádio 3 na escala HY, 28 pacientes com o estádio 2,5 na escala HY e 55 pacientes com o estádio 2 na escala HY. Ou seja, a maioria dos pacientes encontrava-se na fase inicial ou intermédia da doença. A partir dos 16 sensores colocados nos pés, foram extraídos medidas estatísticas e parâmetros espaciotemporais. De forma a encontrar o melhor conjunto de caraterísticas foi aplicada uma técnica de seleção de caraterísticas baseada na correlação (através do coeficiente de Pearson). Inicialmente, foi executada uma análise estatística dos dados dos sensores, para diferenciar o grupo de controlo dos pacientes de DP. Desta análise, foram extraídos a média, a mediana e o desvio-padrão de ambos os grupos. Seguidamente, para avaliar a eficiência dos classificadores, foi utilizada a VC 10-*fold*. Além disso, foi aplicado o teste de Shapiro-Wilk com um nível de significância de 5%, de modo a garantir que a distribuição normal dos atributos da marcha fosse avaliada. Relativamente aos parâmetros espaciotemporais, foram analisados: a duração do passo, a duração da fase de apoio, a duração da fase de balanço, a razão entre a fase de apoio e a fase de balanço, a duração da passada, a cadência, a velocidade e o comprimento do passo e da passada. Neste estudo verificou-se que a cadência, o comprimento da passada, a duração da passada e a velocidade foram os parâmetros nos quais se observou um maior desfasamento entre os dois grupos. No caso da cadência, em pacientes com severidade mais elevada houve uma redução significativa, enquanto nos estágios ligeiro e moderado não se encontraram diferenças significativas em relação ao grupo de controlo. Os classificadores testados foram DT, SVM, EC (*Ensemble classifier*) e classificador de Bayes (BC, *Bayes classifier*). O classificador DT foi construído de modo a dividir a severidade da DP em quatro níveis, baseados na escala HY: saudável (valor previsto < 1), ligeiro (valor previsto >=1 e <2,25), médio (valor previsto >=2,25 e <2,75) e elevado (valor previsto >= 2,75). Além disso, o classificador DT faz uma distinção entre o nível ligeiro e o médio com base no valor do sensor (ligeiro: valor>=122,418; médio: valor<122,418). Em relação aos resultados, os classificadores DT e SVM obtiveram, ambos, a melhor ACC com 99,4%, assim como o melhor *Recall* com 99,6% e SPEC com 99,8%.

Há ainda estudos que combinam sistemas *wearable* com não-*wearable*. Os autores Aich *et al*. [157] propuseram um sistema que recebe dados adquiridos por acelerómetros *wearable* para estimar as caraterísticas da marcha, juntamente com um sistema de captura de análise de movimento 3D (VICON®, Oxford, UK). O estudo foi validado pelo teste de mobilidade Tinetti. Este sistema inclui um algoritmo de ML para distinguir 48 pacientes com DP (estado *on*) de um grupo de indivíduos idosos saudáveis (22 homens e 18 mulheres). A fim de averiguar a

severidade da doença, foram utilizadas as escalas UPDRS e HY. Os acelerómetros foram colocados em ambos os joelhos dos participantes, a uma distância de 34 cm do solo. Cada participante realizou uma marcha de 6 m, tendo sido adquiridas cinco importantes caraterísticas da marcha: tempo de passo, tempo de passada, comprimento do passo, comprimento de passada e velocidade de marcha. Para a classificação foram testados cinco algoritmos, utilizando a VC 5-*fold*. O algoritmo DT obteve melhor ACC com 88,46%. O *Recall* e a SPEC foram 92,86% e 90,91%, respetivamente.

A Tabela 3.4 apresenta um resumo dos pontos principais encontrados nos estudos explorados anteriormente para análise da marcha utilizando sistemas *wearable* e não-*wearable*.

Tabela 3.4 - Resumo dos estudos encontrados na literatura científica relacionados com a análise da marcha utilizando sistemas *wearable e não-wearable*.

Referência	Tipo de sensor	Parâmetros da Marcha	Algoritmo	ACC	*Recall*	SPEC	Validação
(Aich *et al.*, 2020)[154]	Acelerómetros triaxiais	Espaciotemporais	RF	96%	97%	-	VC LOSO
			SVM	93%	92%	-	
			KNN	86%	84%	-	
			NB	88%	86%	-	
(Mittra e Rustagi, 2018)[155]	Força	Cinéticos	KNN	93,08%	84,31%	-	-
			DT	87,78%	92,00%	-	
			RF	90,39%	90,91%	-	
			SVM-linear	89,95%	87,86%	-	
			SVM-RBF	90,39%	90,38%	-	
			SVM-Poly*	88,64%	94,44%	-	
			LR	90,06%	87,76%	-	
(Balaji *et al.*, 2020)[156]	Força	Espaciotemporais	DT	99,4%	99,6%	99,8%	VC 10-*fold*
			SVM	97,6%	95,2%	99,2%	
			EC	95,1%	91,5%	98,2%	
			BC	69,7%	46,0%	89,2%	
(Aich *et al.*, 2020)[157]	Acelerómetro e captura de movimento 3D	Espaciotemporais	KNN	85,57%	91,25%	88,36%	VC 5-*fold*
			SVM	87,23%	91,89%	89,97%	
			NB	84,41%	89,56%	87,35%	
			DT	88,46%	92,86%	90,91%	

*Poly - polinomial

Na Tabela 3.5 é apresentado um resumo dos parâmetros espaciotemporais, antropométricos e clínicos utilizados como caraterísticas no desenvolvimento e implementação dos modelos de ML referidos anteriormente e que os autores consideraram como fundamentais para identificar a DP e, em alguns casos, classificar o nível de severidade.

Tabela 3.5 - Resumo das caraterísticas da marcha selecionadas pelos autores nos estudos encontrados na revisão literária.

Referência	Parâmetros espaciotemporais		Antropométricos e clínicos
(Buongiorno *et al.*, 2019) [151]	Duração do apoio duplo Duração da passada Comprimento da passada		-
(Wahid *et al.*, 2015) [152]	Comprimento da passada Cadência Duração da fase de apoio Duração da fase de balanço Duração do apoio duplo	Comprimento do passo Duração do passo Duração da passada Largura do passo	-
(Rehman *et al.*, 2019)[5]	Velocidade média do passo Comprimento médio do passo Variabilidade do comprimento do passo Largura média do passo Variabilidade da largura do passo		-
(Urcuqui *et al.*, 2018)[153]	Velocidade da marcha Duração do passo Duração da fase de balanço Duração do apoio Percentagem da fase de balanço	Percentagem da fase de apoio Cadência do passo Comprimento do passo Comprimento da passada	Idade
(Aich *et al.*, 2020)[154]	Comprimento do passo Comprimento da passada Duração do passo	Duração da passada Velocidade do passo Velocidade da passada	-
(Balaji *et al.*, 2020)[156]	Duração do passo Duração da fase de apoio Duração da fase de balanço Razão fase de balanço/apoio Duração da passada	Cadência Velocidade da marcha Comprimento do passo Comprimento da passada	-
(Aich *et al.*, 2020)[157]	Duração do passo Duração da passada Comprimento do passo Comprimento da passada Velocidade da marcha		-

3.1.3 – Análise baseada na escrita

A micrografia é outro meio de diagnóstico amplamente utilizado no diagnóstico da DP e consiste, basicamente, num exame escrito. Esta técnica é considerada como uma medida objetiva na caraterização da doença, uma vez que a diminuição do tamanho da letra e o tremor nas mãos são os principais sintomas da DP.

Os autores Pereira *et al.* [158] apresentaram resultados relativamente ao estudo desenvolvido na área do diagnóstico automático da DP, através de técnicas de visão de computador. Para isso utilizaram um *dataset* constituído por imagens extraídas de exames de escrita de 55 indivíduos, dos quais 37 sofriam de DP e os 18 restantes faziam parte do grupo de controlo. O exame consistia no preenchimento de algumas tarefas, tais como desenhar círculos, espirais e curvas. Seguiu-se uma etapa de processamento de imagem e extração de caraterísticas, para separar o esqueleto da espiral desenhada pelo participante. Foram utilizados classificadores de

aprendizagem supervisionada NB, OPF (*Optimum-Path Forest*) e SVM com kernel RBF. Além disso, foi executada a VC 10-*fold*. O classificador NB obteve a melhor classificação com uma ACC de 78,9%, seguida do OPF, embora os resultados tenham sido maus no grupo de controlo em comparação com o grupo de pacientes. A única exceção relativamente ao NB foi um melhor *Recall* no grupo de controlo.

Goel *et al*. [159] propuseram uma nova forma de implementar métodos de aprendizagem supervisionada e não supervisionada na deteção da DP. Além disso, utilizaram técnicas de PCA (*Principal Component Analysis*) e LDA (*Linear Component Analysis*) a fim de reduzir a dimensionalidade do espaço de caraterísticas. Para a realização do estudo, foram utilizados dois *datasets*, um com 4 curvas e outro com 4 espirais escritas pelos pacientes de Parkinson. Foram incluídas 13 caraterísticas (idade, género, lateralidade da mão, distância máxima e mínima entre o *template* do exame e o traço da escrita, altura mínima e máxima, etc.) e 633 instâncias. O modelo foi treinado com 70% do *dataset* e foram testados os algoritmos SVM, KNN, DT e RF. Em ambos os *datasets*, após ter sido utilizado o PCA, as caraterísticas foram reduzidas de 13 para 6. Em relação ao *dataset* com desenhos de curvas, o classificador RF obteve a ACC mais elevada com 87,36% e uma SPEC de 94,53%. Relativamente ao *dataset* com desenho de espirais, o classificador SVM obteve o melhor resultado, com uma ACC de 84,73% e SPEC de 96,09%.

Na Tabela 3.6 é apresentado um resumo dos resultados obtidos para cada algoritmo nos dois estudos relacionados coma análise baseada na escrita.

Tabela 3.6 - Resumo dos estudos encontrados na literatura científica relacionados com a análise baseada na escrita.

Referência	*Dataset*	Algoritmo	*ACC*	*Recall*	SPEC	Validação
(Pereira *et al.*, 2015)[158]	Desenho de espirais + curvas + círculos	NB OPF SVM	78,9% 77,1% 75,8%	GC:91%; GDP:34% GC:32%; GDP:80% GC:24%; GDP:83%	-	VC 10-*fold*
(Goel *et al.*, 2020)[159]	Desenho de curvas	SVM KNN DT RF	85,26% 85,26% 81,05% 87,36%	-	94,53% 93,75% 88,28% 94,53%	VC *holdout* (70-30)
	Desenho de espirais	SVM KNN DT RF	84,73% 81,57% 77,89% 80,52%		96,09% 96,09% 83,59% 88,28%	

GC – grupo de controlo; GDP – grupo de pacientes com DP

3.1.4 – Análise baseada em dados de EEG e outros dados clínicos

De modo geral, os sintomas não-motores na DP envolvem cognição e emoções, tendo merecido uma especial atenção nos últimos anos. Os sinais do eletroencefalograma (EEG) podem refletir o estado emocional de uma pessoa. Yuvaraj *et al*. [160] desenvolveram uma *framework* computacional a fim de classificar 20 pacientes com DP e compará-los com 20 indivíduos

saudáveis, através da informação emocional proveniente da atividade elétrica cerebral. O objetivo do estudo consistia em encontrar uma relação entre os estados emocionais e as bandas de frequência de EEG nos dois grupos de indivíduos. A cada participante foi pedido o mínimo movimento possível e máxima concentração no estímulo emocional. Foi utilizado o Emotiv EPOC para adquirir os sinais de EEG *wireless* e, durante o teste, o participante assistiu a pequenos vídeos num ecrã. No final de cada vídeo, o participante tinha de preencher um questionário acerca do estado emocional sentido, durante a experiência, numa escala de 1 a 5, sendo 1 muito baixo e 5 muito alto. Seguidamente, os sinais foram pré-processados e extraíram-se caraterísticas para avaliar seis estados emocionais (alegria, tristeza, receio, raiva, surpresa e repugnância). Os algoritmos utilizados para testar os estados emocionais foram: KNN e SVM. O SVM teve um desempenho superior ao KNN em ambos os grupos. Os resultados obtidos com o algoritmo SVM podem ser consultados na Tabela 3.7. Neste modelo foi utilizada a VC 10-*fold*.

Tabela 3.7 – Resultados da classificação do algoritmo SVM de acordo com o tipo de emoção [160].

Referência	Tipo de dados	Algoritmo	Tipo de emoção	*ACC*	Validação
(Yuvaraj *et al.*, 2014)[160]	EEG	SVM	Alegria	GDP: 94,33% GC: 92,50%	VC 10-*fold*
			Tristeza	GDP: 75,00% GC: 84,67%	
			Receio	GDP: 80,33% GC: 77,50%	
			Raiva	GDP: 78,00% GC: 82,67%	
			Surpresa	GDP: 88,00% GC: 96,67%	
			Repugnância	GDP: 72,67% GC: 66,50%	

GDP – grupo de pacientes com DP; GC – grupo de controlo.

Os autores Neehal *et al.* [161] desenvolveram um modelo de ML que analisa imagens de imagens de ressonância magnética funcional para detetar a DP numa fase inicial e auxiliar o médico a tomar medidas que beneficiem o paciente. Nesse sentido utilizaram a base de dados *Parkinson's Progression Markers Initiative* (PPMI) e escolheram 8 pacientes (7 homens e 1 mulher) com DP, no estado inicial, para proceder à análise. Neste estudo, a partir de cada voxel, foi extraída a caraterística função de densidade espectral (PSD, *Power Spectral Density*), no domínio tempo-frequência, através da *Short Time Fourier Transform* (STFT). Para classificar as imagens, foi escolhido o algoritmo SVM com o *kernel* RBF, atribuindo o estágio da DP correspondente, em conjunto com a VC 5-*fold*. Os resultados obtidos mostraram que o algoritmo desenvolvido obteve uma ACC, um *Recall* e uma SPEC de 100% em 7 indivíduos e, num indivíduo, obteve uma ACC de 99,76%, um *Recall* de 99,53% e uma SPEC de 100%.

Na verdade, a maioria das pessoas estão familiarizadas com os sintomas motores da DP. Contudo, os sintomas não-motores precedem os motores, facto este que levou a um maior interesse por parte dos investigadores em prever a DP numa fase inicial. É o caso dos autores

Challa *et al.* [162] que desenvolveram um modelo de ML que permite prever a DP nessa fase, com base nos sintomas não-motores, tais como distúrbios de sono e perda olfativa. Para o estudo foram considerados 184 pacientes saudáveis e 402 pacientes com DP no estágio inicial, provenientes da base de dados PPMI. As caraterísticas incluídas no estudo foram a pontuação obtida no teste de identificação do olfato da universidade da Pensilvânia, a pontuação obtida no questionário de triagem do distúrbio comportamental do sono REM, os biomarcadores do líquido cefalorraquidiano (LCR) e as imagens de DaTscan SPECT. Os algoritmos utilizados neste estudo foram: MLP, Bayesian Network (BayesNet), RF e *Boosted Logistic Regression* (BLR). O *dataset* foi dividido em 70% para o conjunto de treino e 30% para o conjunto de teste, mantendo a mesma proporção de população com e sem DP em cada conjunto. O algoritmo BLR obteve a melhor performance com 97,16% de ACC e AUC de 98,9%.

Outro modelo de ML foi proposto por Prashanth *et al.* [163]. Este modelo utiliza apenas as caraterísticas relacionadas com a pontuação obtida no teste de identificação do olfato da universidade da Pensilvânia e a pontuação obtida no questionário de triagem do distúrbio comportamental do sono REM da base de dados PPMI. Para o estudo foram consideradas 195 indivíduos saudáveis para o grupo de controlo e 423 indivíduos com DP no estágio inicial. 70% dos dados foram utilizados no treino e 30% foram utilizados para testar o modelo. A fim de distinguir a DP no estágio inicial dos indivíduos saudáveis foram utilizados dois algoritmos: SVM e CT (*Classification Tree*). O SVM com o *kernel* RBF apresentou uma melhor performance com ACC de 85,48%, um *Recall* de 90,55%, uma SPEC de 74,58% e um AUC de 88,22%. Na Tabela 3.8 é apresentado um resumo dos principais pontos encontrados nos estudos mencionados.

Tabela 3.8 - Resumo dos estudos encontrados na literatura científica relacionados com análise baseada em outros dados clínicos.

Referência	Tipo de dados	Algoritmo	No. indivíduos	ACC	*Recall*	SPEC	Validação
(Neehal *et al.*, 2020)[161]	Imagens ressonância magnética funcional	SVM (RBF)	7 1	100% 99,76%	100% 99,53%	100% 100%	VC 5-*fold*
(Challa *et al.*, 2016)[162]	Questionários; Biomarcadores de LCR; Imagens de DatScan SPECT;	MLP BayesNet RF BLR	586	95,45% 96,02% 96,59% 97,16%	95,5% 96,0% 96,6% 97,2%	- - - -	VC *holdout* (70-30)
(Prashanth *et al.*, 2014)[163]	Questionários para identificação do olfato e triagem do distúrbio de sono REM	SVM (linear) SVM (RBF) CT	618	84,95% 85,48% 81,72%	88,19% 90,55% 83,46%	77,97% 74,58% 77,97%	VC *holdout* (70-30)

Por fim, os autores Solana-Lavalle e Rosas-Romero [164] apresentaram um novo método de classificar imagens 3D de RM, ao funcionar como uma ferramenta de apoio ao diagnóstico da DP. Para a realização do estudo utilizaram a maior base de dados com imagens de RM (PPMI), tendo sido selecionados um grupo de 330 pacientes com DP (226 homens e 104 mulheres) e um

grupo de controlo com 150 indivíduos (86 homens e 64 mulheres). Para este estudo foram escolhidas imagens ponderadas em T1. Primeiramente, para cada imagem foi detetada a região de interesse, através da morfologia baseada no voxel, no interior da substância cinzenta do cérebro. No final foi criado um mapa de vóxeis resultante das diferenças de concentração de substância cinzenta entre os *templates* que representavam os dois grupos de indivíduos. Seguiu-se a extração das regiões de interesse, utilizando o teste *t* para comparação entre os 2 *templates*. Após a definição 3D da região de interesse para análise, as caraterísticas foram extraídas com base na matriz de co-ocorrência dos níveis de cinzento (GLCM, do inglês *Grey Level CoOccurrence Matrix*). Para reduzir o número de caraterísticas foi escolhido o PCA, seguido do método *Wrapper*. Seguidamente, cada classificador foi testado através da VC 10-*fold*. Nos homens, o algoritmo NB obteve a ACC mais alta com 99,01%, o algoritmo SVM conseguiu o *Recall* mais elevado com 99,35% e o algoritmo NB obteve a melhor SPEC com 100%. Nas mulheres, o algoritmo LR obteve a melhor ACC com 96,77%, o BayesNet conseguiu o melhor *Recall* com 100% e o MLP obteve a melhor SPEC com 96,15%. Na Tabela 3.9 são apresentados os resultados obtidos neste estudo, em função da intensidade do campo magnético 1,5T e 3T.

Tabela 3.9 - Resultados obtidos na classificação em [164], de acordo com a intensidade do campo magnético.

Referência	Intensidade do campo magnético	Algoritmo	Género	ACC	*Recall*	SPEC	Validação
(Solana-Lavalle e Rosas-Romero, 2021)[164]	1,5 T	LR	F	96,77%	97,22%	96,15%	VC 10-*fold*
		NB	M	99,01%	98,72%	100%	
		BayesNet	F	88,71%	100%	73,08%	
		MLP	M	98,02%	98,72%	95,65%	
			F	93,55%	91,67%	96,15%	
	3T	LR	F	93,28%	95,06%	89,47%	
			M	91,56%	92,81%	88,89%	
		MLP	F	85,71%	98,06%	65,79%	
		SVM	M	95,56%	99,35%	87,50%	

F-feminino; M-masculino

3.2 - Comparação dos métodos revistos

Após uma análise da literatura científica verifica-se a existência de um número assinalável de modelos de ML capazes de identificar a DP e de prever o nível de severidade numa fase inicial. Uma vez que a DP tem inúmeros sintomas motores e não-motores, existem, atualmente, diferentes métodos de ML, adaptados aos tipos de dados. No caso das alterações da marcha, há investigadores que optaram por empregar, nos seus estudos, sistemas *wearable* e outros optaram pelos sistemas não-*wearable* ou a combinação dos dois. Os autores Balaji *et al.* [156] utilizaram sensores de força *wearable* e obtiveram os valores de ACC, *Recall* e SPEC mais elevados, com 99,4%, 99,6% e 99,8%, respetivamente. Estes resultados foram obtidos quando aplicado o classificador DT.

Em relação aos algoritmos de aprendizagem, SVM e RF foram os mais escolhidos para treinar e testar um modelo de ML, cuja exatidão variou entre 72,67% e 100%. Em relação ao tipo de VC, a 10-*fold* foi a mais referida pelos autores.

Se analisarmos os parâmetros da marcha, o comprimento do passo, o comprimento da passada, a duração da passada e a duração do passo foram os parâmetros espaciotemporais mais mencionados pelos autores. Por outro lado, quanto aos parâmetros antropométricos e clínicos, apenas um autor escolheu a idade como caraterística.

De modo geral, os diferentes autores construíram um sistema de classificação da DP constituído por cinco fases, tal como mostra a Figura 3.2.

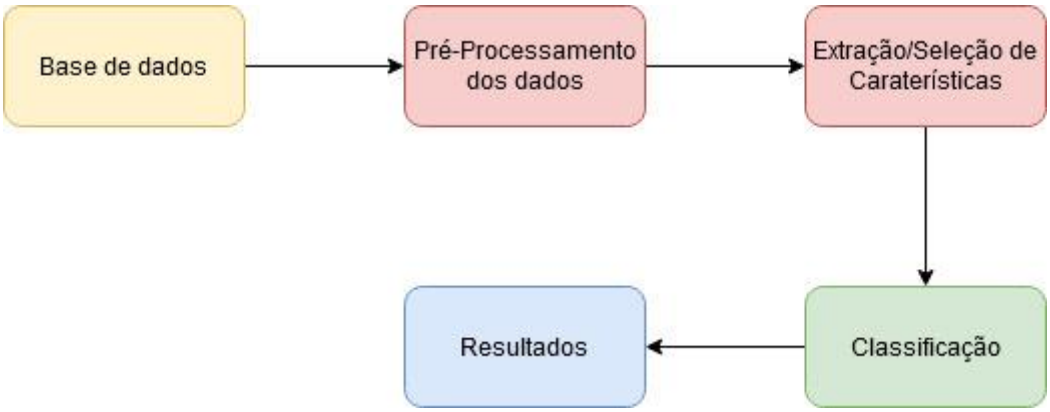


Figura 3.2 – Principais passos num sistema de classificação da DP.

Primeiramente, os dados pertencentes aos pacientes com DP e indivíduos saudáveis são adquiridos, seguindo-se o pré-processamento. Posteriormente, as caraterísticas mais relevantes são selecionadas. Estas caraterísticas fornecerão informação que permitirá distinguir a marcha parkinsoniana da marcha "normal". Finalmente, a classificação é executada de modo a identificar os indivíduos como pertencentes ao grupo de pacientes com DP ou ao grupo saudável. O desempenho da classificação é avaliado de acordo com as métricas de avaliação.

Contudo, os métodos de ML revistos apresentam algumas limitações. Grande parte dos estudos utilizaram *datasets* pequenos, o que pode comprometer a fiabilidade dos sistemas desenvolvidos. Além disso, o aumento do número de amostras tem como vantagem a possibilidade de implementar técnicas de DL. Outra limitação encontrada em alguns estudos diz respeito à utilização de *datasets* desequilibrados, sobretudo, em relação ao género.

Por outro lado, seria interessante o desenvolvimento de mais métodos de ML que fizessem a previsão da DP considerando quer os sintomas motores quer os sintomas não-motores. Relativamente aos estudos que analisaram parâmetros da marcha com base em sistemas *wearable*, uma vez que maioria adquiriu esses dados em contexto clínico, numa fase posterior, o modelo de ML deveria ser testado em casa. Em relação aos sistemas não-*wearable*, é necessário desenvolver mais sistemas de baixo custo para a análise da marcha.

3.3 – Resumo

Devido à sua capacidade de extrair e de classificar automaticamente padrões, o ML permite apoiar o diagnóstico da DP e identificar o nível de severidade. Nos últimos anos foram desenvolvidas várias tecnologias assistidas por computador para o diagnóstico e tratamento da DP, incluindo análise da voz, análise da escrita, análise da marcha a partir de sistemas *wearable* e não-*wearable*, análise de imagens médicas, análise de sinais de EEG e de outros dados clínicos.

Os modelos desenvolvidos obtiveram melhores resultados quando treinados por algoritmos de aprendizagem supervisionada, utilizando apenas parâmetros espaciotemporais.

Por outro lado, um sistema de classificação da DP é constituído pelas seguintes fases: aquisição de dados; pré-processamento de dados; extração e/ou seleção de caraterísticas; classificação; e resultados.

Por fim, embora o interesse, por parte dos investigadores, nesta área, tenha crescido nos últimos anos e os resultados obtidos sejam promissores, há aspetos que devem ser melhorados.

Capítulo 4

Caracterização da base de dados e Metodologia

O presente capítulo encontra-se dividido em duas secções. Na Secção 4.1 é apresentada a base de dados utilizada no trabalho, com ênfase na caraterização da amostra, seguida da análise clínica realizada a todos os participantes. Além disso, é descrita a tarefa de locomoção executada, assim como são mencionados todos os instrumentos e procedimentos que permitiram a aquisição dos parâmetros da marcha.

Na Secção 4.2 é apresentada a metodologia implementada a fim de detetar a DP e identificar o nível de severidade. Na verdade, foram desenvolvidos dois modelos de ML diferentes, mas com vários pontos em comum. Esta secção está subdividida em cinco partes que coincidem com os principais passos da metodologia proposta para esta Dissertação.

4.1 – Base de dados

4.1.1 - Caracterização da amostra

A amostra para este estudo é constituída por indivíduos que foram recrutados da região de Bauru (São Paulo, Brasil). Todos os participantes deram o seu consentimento escrito, de acordo com a Declaração de Helsínquia.

Participaram no estudo 126 indivíduos distribuídos em dois grupos: grupo de pacientes com DP (GDP) que inclui 63 participantes, sendo 34 homens e 29 mulheres; e grupo de pessoas neurologicamente saudáveis (grupo de controlo, GC) constituído por 63 participantes sem DP, dos quais 18 são homens e 45 são mulheres. Para o estudo, os participantes foram selecionados de acordo com os seguintes critérios de exclusão e de inclusão: idade abaixo dos 50 anos; visão ou audição limitada e doença musculoesquelética ou ortopédica que pudesse afetar a capacidade para executar o protocolo experimental; declínio cognitivo (abaixo dos 24 pontos no questionário MMSE, de acordo com o nível de escolaridade [165]); e/ou outras doenças neurodegenerativas (p. ex., doença de Alzheimer).

Relativamente aos pacientes com DP, os participantes foram diagnosticados com a DP idiopática pelo seu neurologista particular, de acordo com os critérios do Banco de Cérebros de Londres [166]. Além disso, foram incluídos aqueles que obtiveram uma pontuação igual ou

inferior a 3 na escala HY. Todos os pacientes tomavam medicação anti-parkinsoniana, no mínimo, 3 meses antes da aquisição dos dados.

Para caracterização da amostra foi realizada uma anamnese, tendo sido registado o género, idade, massa corporal, altura, IMC e MMSE de todos os participantes. No caso dos pacientes com DP foi ainda fornecida informação relativa à duração da doença e a pontuação da escala HY e da terceira parte da escala MDS-UPDRS. Na Tabela 4.1 são apresentadas as informações clínicas dos participantes de ambos os grupos, referentes à média e desvio-padrão. Além disso é apresentado o teste t-*Student* para amostras independentes, sendo que os valores p mostram a diferença estatística entre os dois grupos ($p<0,05$).

Tabela 4.1 – Informações clínicas dos participantes.

Parâmetro	GC (n=63) Média ± dp	GDP (n=63) Média ± dp	t(gl)	p
Género	18 H / 45 F	34 H / 29 F	-	-
Idade (anos)	68,32 ± 6,28	68,79 ± 7,51	t(124)=-0,386	0,700
Massa corporal (Kg)	72,05 ± 13,21	69,41 ± 13,75	t(124)=1,10	0,273
Altura (m)	1,60 ± 0,08	1,62 ± 0,09	t(124)=-1,52	0,131
IMC (Kg/m^2)	28,20 ± 4,67	26,38 ± 4,66	t(124)=2,20	0,030*
MMSE (pts)	27,95 ± 2,02	27,49 ± 2,00	t(124)=1,29	0,201
Duração da doença (anos)	-	5,33 ± 3,67	-	-
HY (pts)	-	2,12 ± 0,49	-	-
HY I (n)	-	42	-	-
HY II (n)	-	21	-	-
MDS-UPDRS-III (pts)	-	27,46 ± 10,60	-	-

dp, desvio-padrão; t, valor t de *Student*; gl, graus de liberdade; p, significância

*estatisticamente significativo para $p<0,05$

4.1.2 – Análise clínica

O GDP foi avaliado por um experimentador experiente, através da terceira parte da escala MDS-UPDRS e da escala HY. Por intermédio da primeira escala foi possível aferir o grau de acometimento motor da DP. Por outro lado, a escala HY foi utilizada para identificar o estágio da doença, avaliar a existência da unilateralidade/bilateralidade da doença e obter o nível de resposta dos reflexos posturais.

A atribuição de *labels* no GDP teve como referência a escala HY. Neste grupo existem 4 pacientes com HY=1, 5 pacientes com HY=1,5, 33 pacientes com HY=2, 14 pacientes com HY=2,5 e 7 pacientes com HY=3. Portanto, no GDP há 42 pacientes que se encontram no estágio ligeiro da doença (HY I) e 21 pacientes no estágio intermédio (HY II).

A função cognitiva foi avaliada em todos os participantes segundo o questionário MMSE [167]. As pontuações obtidas no MMSE foram ajustadas de acordo com a escolaridade de cada participante [168].

4.1.3 – Tarefa de locomoção

Para a realização do estudo, foi pedido a cada participante que executasse a marcha, com uma velocidade da sua preferência, num percurso retilíneo de, aproximadamente, 8 metros, sem restrições ou obstáculos, num total de 3 tentativas. Os participantes executaram a marcha num tapete normal com 8,5 m de comprimento e 1,5 m de largura. O GDP foi avaliado sob efeito da medicação dopaminérgica (uma hora após a ingestão da medicação).

4.1.4 – Instrumentos e procedimento do estudo biomecânico

Para a recolha dos parâmetros espaciotemporais da marcha foi utilizado um sistema de captura de movimento tridimensional da Vicon® (Vicon Motion Systems Ltd; Oxford, UK), constituído por 10 câmaras de infravermelhos, com uma frequência de aquisição de 100 Hz. Para a captura do movimento, através de marcadores refletores colocados sobre o corpo dos participantes, foi utilizado o software Nexus da Vicon®.

Os marcadores refletores foram posicionados de acordo com o modelo standard *full-body* (Plug-in Gait) da Vicon [169], tal como apresentado na Figura 4.1. Contudo, para a análise e cálculo dos parâmetros espaciotemporais apenas foram utilizados os marcadores refletores colocados em ambos os pés (segundo metatarso e calcâneo).



Figura 4.1 - Fotografia tirada a um participante com os marcadores refletores posicionados segundo o protocolo Plug-in-gait *full body* (VICON).

A fim de calcular os parâmetros espaciotemporais da marcha foi utilizado o software *MATLAB R2018a*. Além disso, foi aplicado um filtro *Butterworth* passa-baixo de 5ª ordem (*zero-lag*), com uma frequência de corte de 6 Hz, nos dados adquiridos. A Tabela 4.2 apresenta a descrição dos parâmetros espaciotemporais calculados para o passo e passada dos participantes.

Tabela 4.2 - Resumo dos parâmetros espaciotemporais calculados durante a marcha do participante.

Parâmetros espaciotemporais	
Passo	**Passada**
Comprimento (cm)	Comprimento (cm)
Largura (cm)	Largura (cm)
Duração (s)	Duração (s)
Velocidade (cm/s)	Velocidade (cm/s)
Duração do apoio simples (s)	Duração do apoio simples (s)
Duração do apoio duplo (s)	Duração do apoio duplo (s)
Apoio simples (% do ciclo)	Apoio simples (% do ciclo)
Apoio duplo (% do ciclo)	Apoio duplo (% do ciclo)
Cadência (passos/s)	Cadência (passada/s)

Além dos parâmetros referidos, foi calculada a variabilidade do passo e da passada e a assimetria do passo. Na Subsecção 3.1.2 do Capítulo 3 foi apresentado um estudo publicado por Rehman *et al*. [5] cujo objetivo consistia na seleção dos parâmetros da marcha clinicamente relevantes na DP numa fase inicial, entre os quais figuravam a variabilidade e assimetria do passo. Enquanto o cálculo da variabilidade pode ajudar a prever quedas em pacientes idosos e com DP, a assimetria pode ser útil na identificação da DP ligeira, visto que a degeneração das células dopaminérgicas ocorre com uma distribuição assimétrica. Por esse motivo, foi decidido incluir os referidos parâmetros nesta Dissertação.

Segundo os autores Simieli *et al*.[170] a variabilidade pode ser calculada a partir do coeficiente de variação. Este coeficiente é determinado com base na média e no desvio-padrão de cada parâmetro do passo e da passada durante cada ensaio. De acordo com Orcioli-Silva *et al*. [171], a assimetria do passo pode ser determinada através do índice de assimetria (A_i, do inglês *Asymmetry Index*), que quantifica a assimetria entre o lado mais afetado (MAS) e o lado oposto ao afetado (LAS) no GDP e a assimetria entre o lado dominante (DL) e o lado não-dominante (NDL) no GC:

$$A_i = \left| \frac{valor\ do\ LAS\ ou\ DL - valor\ do\ MAS\ ou\ NDL}{valor\ do\ LAS\ ou\ DL - valor\ do\ MAS\ ou\ NDL} \right| \times 100 \quad (4.1)$$

A fim de comparar os valores médios de cada variável entre os dois grupos foi aplicado o teste t-*Student* para amostras independentes, recorrendo ao Software *SPSS® Statistics* (versão 26, IBM®). Uma vez que a dimensão da amostra é superior a 30, este teste pode ser aplicado sem ser necessário verificar a normalidade nos dois grupos [172]. Juntamente com o teste t-*Student* foi empregue o teste de *Levene* para verificar a homogeneidade das variâncias [173].

Numa fase posterior, o GDP foi dividido em dois grupos: GDP ligeira (GDPL) e GDP intermédia (GPDI). No caso do GDPI, a sua dimensão é inferior a 30, tendo sido necessário verificar a normalidade neste grupo através do teste de Shapiro-Wilk. Foi aplicado o teste t-*Student* nos parâmetros que seguiam a normalidade e o teste de Mann-Whitney nos restantes casos.

4.2 – Metodologia

No Capítulo 3 foram apresentados alguns estudos que propuseram modelos de ML capazes de apoiar a avaliação clínica durante a deteção da DP, assim como identificar o nível de severidade. Na Secção 3.2 desse capítulo foram destacados alguns pontos essenciais na fase de desenvolvimento e implementação dos modelos de ML revistos.

Com base na informação encontrada na revisão da literatura, é apresentado o *workflow* do modelo de ML para a deteção da DP proposto para esta Dissertação (Figura 4.2).

A partir dos passos seguidos no *workflow* apresentado para a deteção da DP, foi construído outro para a deteção da severidade da DP, tal como é observável na Figura 4.3. As diferenças entre as figuras são relativas à proporção entre o conjunto de treino e de teste e, no caso do segundo *workflow*, a utilização das técnicas de sobre-amostragem SMOTE e ROS. Posteriormente, entre as Subsecções 4.2.2 - 4.2.5 é detalhado cada passo do *workflow* referente à classificação automática da DP e da sua severidade.

Para o desenvolvimento e implementação dos modelos de ML foi utilizada a versão 3.3.6 do software *Spyder* (*Scientific Python Development Environment*). Trata-se de um editor integrado de desenvolvimento (IDE) *open-source* em programação em Python para a visualização, processamento e análise de dados num ambiente interativo [174]. Algumas das bibliotecas fundamentais para o desenvolvimento de modelos de ML encontram-se listadas de seguida.

- SciKit-Learn [175] – inclui algoritmos de ML de aprendizagem supervisionada e não-supervisionada, assim como ferramentas para afinação de modelos, pré-processamento de dados, seleção e avaliação de modelos, entre outras.
- Pandas [176] – indicada para trabalhar em dados armazenados em tabelas, tais como bases de dados. Esta biblioteca auxilia na pesquisa, limpeza e processamento dos dados.
- Numpy [177] – utilizada, principalmente, para facilitar a manipulação de *arrays* multidimensionais e a realização de cálculos.
- Matplotlib [178] e Seaborn [179] – para a visualização de dados na forma de gráficos, tais como gráficos de barras, linhas, circulares, etc…

Além disso, o computador utilizado tem um processador Intel® Core™ i3-2350M CPU @ 2.30 GHz com 6 GB de RAM e NVIDIA GeForce 610M com 2GB de RAM.

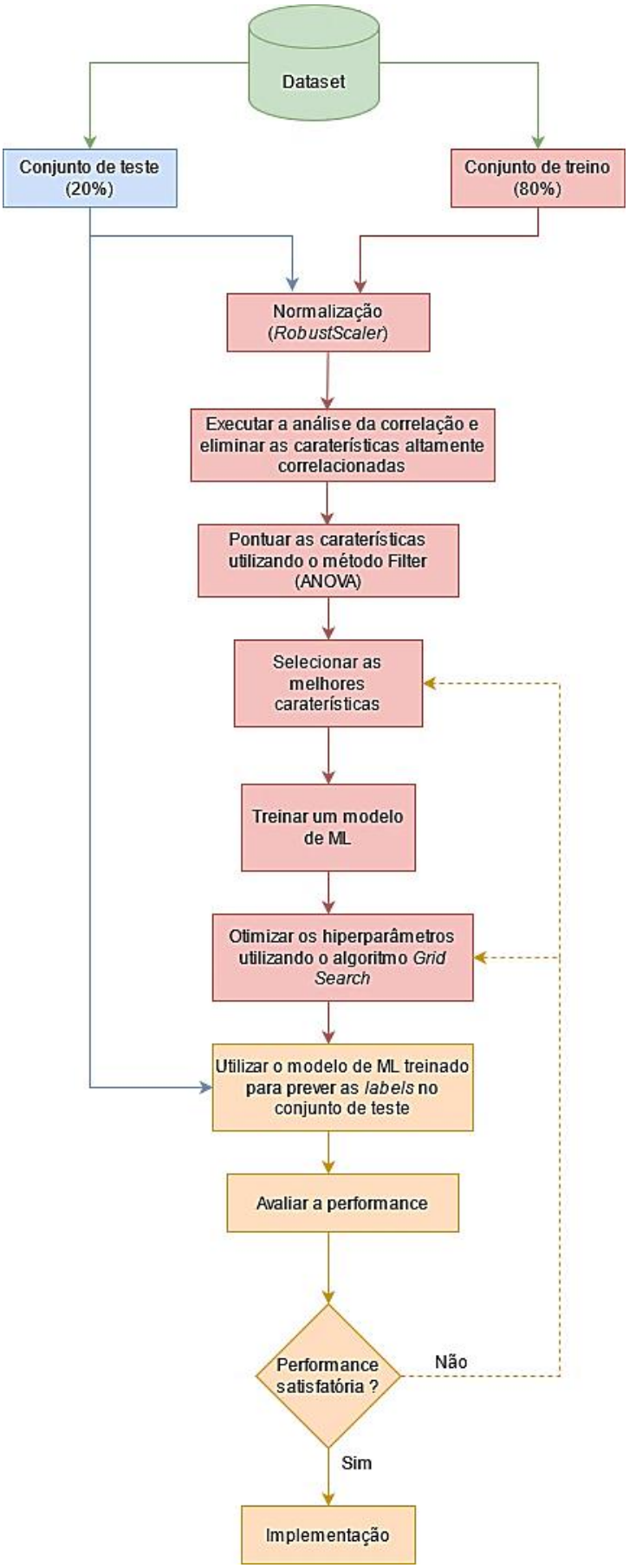


Figura 4.2 - *Workflow* do modelo de ML para a deteção da DP proposto para esta Dissertação.

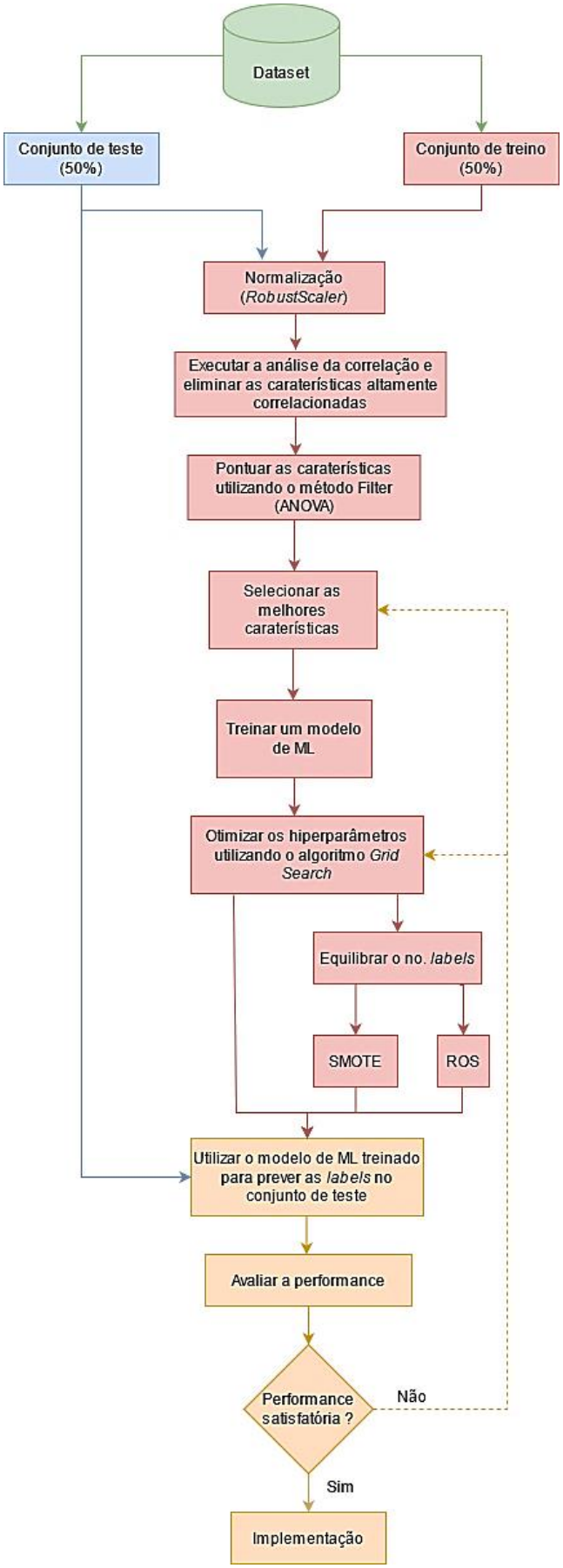


Figura 4.3 - *Workflow* do modelo de ML para a deteção da severidade da DP proposto para esta Dissertação.

No caso da deteção da severidade da DP, uma vez que a dimensão do *dataset* considerado é muito pequena e as classes não têm a mesma distribuição (proporção 1:2), foram testadas duas técnicas de sobre-amostragem de dados: SMOTE e ROS (*Random Oversampling*). Enquanto a técnica SMOTE gera novas amostras sintéticas por interpolação, a técnica ROS clona aleatoriamente amostras da classe minoritária até esta ter o mesmo tamanho da classe maioritária [139]. No *workflow* proposto, estas técnicas são aplicadas depois da seleção de caraterísticas. No caso da técnica SMOTE, Blagus e Lusa [180] referem que a sua utilização antes da seleção de caraterísticas requer algum cuidado, porque a maioria dos métodos de seleção de caraterísticas assumem a independência das amostras.

4.2.1 – Análise do *dataset*

O *dataset* original encontrava-se separado em dois ficheiros em Excel, sendo que um continha informações relativas ao GDP e o outro era referente ao GC. Cada ficheiro tinha quatro folhas de cálculo: a primeira apresentava as informações demográficas e clínicas dos participantes, a segunda e terceira folhas incluíam os parâmetros da marcha dos participantes em função do passo e da passada, respetivamente, e a quarta folha continha os valores da assimetria do passo.

O passo seguinte consistiu na leitura dos dois ficheiros e no armazenamento de todos os parâmetros da marcha numa estrutura de dados bidimensional do tipo *DataFrame* da biblioteca *Pandas*. Um *Dataframe* assemelha-se a uma matriz, no qual cada coluna corresponde a um parâmetro e cada linha corresponde aos registos de um participante. Nesta fase, também se verificou se existiam valores omissos, duplicados ou *outliers*. Uma forma de considerar uma amostra como *outlier* é medir a sua curtose (valor aceitável entre -2 e 2) [181]. A curtose pode ser calculada a partir da classe *kurtosis* da biblioteca *scipy.stats*.

Além disso, foi necessário atribuir uma *label* a cada linha do *DataFrame*, dependendo se o participante tinha ou não DP (*label* 1 ou *label* 0, respetivamente). Na Figura 4.4 é apresentada a distribuição demográfica dos indivíduos pertencentes ao GC e ao GDP. Esta base de dados é constituída por um número de indivíduos do género feminino superior ao masculino, embora o género masculino seja representado em maior número no GDP.

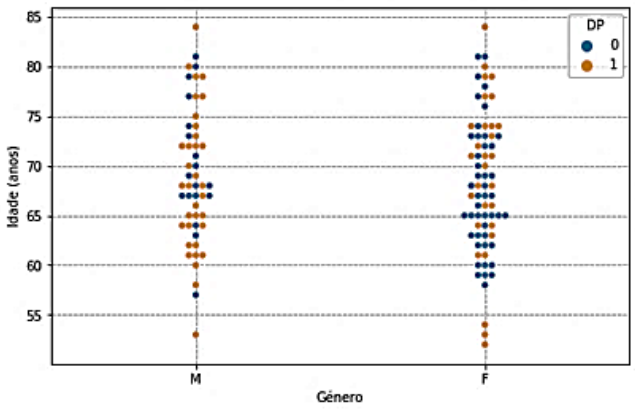


Figura 4.4 - Distribuição demográfica dos indivíduos pertencentes ao GC e ao GDP.

Numa fase posterior, para a identificação do nível de severidade da DP, considerou-se que a *label* 0 representa o GDPL e a *label* 1 representa o GDPI. A atribuição da *label* teve como base a pontuação obtida na escala HY. Na Figura 4.5 está representada a distribuição

demográfica dos indivíduos pertencentes aos grupos referidos. É de notar que o género masculino é o mais representado no GDPI, enquanto no GDPL há um maior equilíbrio entre géneros.

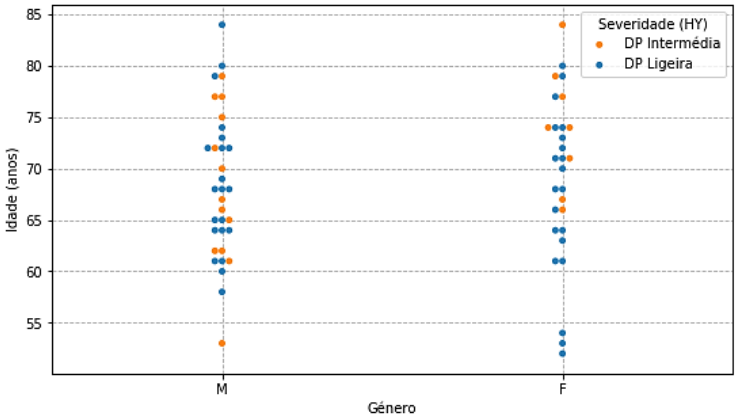


Figura 4.5 - Distribuição demográfica dos indivíduos pertencentes ao GDPL e ao GDPI.

4.2.2 - Pré-processamento dos dados

Primeiramente, o *dataset* foi dividido em conjunto de treino (80% do *dataset*) e conjunto de teste (20% do *dataset*), através da classe *train_test_split* da biblioteca *sklearn.model_selection*. Segundo Raschka e Mirjalili [182], na prática, as proporções mais comumente utilizadas são 60:40, 70:30 ou 80:20, dependendo do tamanho inicial do *dataset*.

Uma vez que o *dataset* utilizado para a deteção da DP tinha um tamanho pequeno, foram escolhidos algoritmos de aprendizagem mais simples para construção do modelo. Para a fase de identificação da severidade da DP, foi utilizado um *dataset* com uma dimensão ainda mais pequena, relativo ao GDP, tendo sido necessário equilibrar as classes. A proporção do *dataset* escolhida foi 50:50. Além disso, foi definido o parâmetro *stratify* na classe *train_test_split*, para garantir que os conjuntos de treino e de teste tivessem uma proporção de *labels* igual à do *dataset* [183].

De modo a que as diferentes caraterísticas tivessem importâncias proporcionais foi necessário proceder à normalização dos valores, utilizando a classe *RobustScaler* da biblioteca *sklearn.preprocessing*. Esta classe funciona de forma semelhante à classe *StandardScaler*, ou seja, assegura que diferentes caraterísticas estejam todas na mesma escala, contudo a classe *RobustScaler* utiliza a mediana e os quartis, ao invés da média e da variância. Desta forma, esta classe não é tão afetada pela presença de *outliers*, ao contrário da classe *StandardScaler* [138].

4.2.3 - Seleção das caraterísticas

Tal como discutido na Secção 2.3.4 do Capítulo 2, a seleção de caraterísticas permite que sejam incluídas apenas as caraterísticas com uma forte correlação com a variável prevista ou dependente. No caso do método *Filter*, as caraterísticas são ordenadas com base num critério e são selecionadas de acordo com a posição que ocuparem no ranking. Como este método, em comparação com o método *Wrapper*, apresenta um custo computacional mais baixo, foi escolhido para ser implementado neste trabalho.

Um tipo de método *Filter* é baseado em testes estatísticos univariados. Um dos exemplos deste tipo de método é a ANOVA que analisa individualmente cada caraterística para encontrar uma relação estatisticamente significativa entre cada caraterística e o *target*. Embora seja um

método rápido, não tem a capacidade de capturar a redundância entre caraterísticas [138], [181]. Por esse motivo, neste trabalho, as caraterísticas foram selecionadas, primeiramente, através do método *Filter* baseado na correlação, seguido da ANOVA.

O método *Filter* baseado na correlação permite identificar as caraterísticas correlacionadas. Através da correlação é possível medir quão fortemente uma variável depende de outra, sendo demonstrada na forma de matriz (*heatmap*). Um dos critérios mais utilizados em ML é o coeficiente de correlação de Pearson [-1,1]. Quando o coeficiente se aproxima de -1 significa que existe uma correlação negativa, quando se aproxima de 1 existe correlação positiva e quando se aproxima de 0 não há correlação linear entre as duas variáveis. Portanto, quando duas variáveis são altamente correlacionadas entre si significa que é possível prever uma a partir da outra, tal como deve acontecer entre as caraterísticas e o *target*. Mas, se as caraterísticas forem altamente correlacionadas entre si, surge um problema de redundância no *dataset*, sendo necessário eliminar uma delas [184].

Através da função *corr()* com definição do método "Pearson" é possível obter uma matriz de correlação do *dataset*. Posteriormente, a cada duas caraterísticas que apresentasse um coeficiente de correlação de *Pearson* superior a 0,8 entre elas, uma continuou e a outra foi eliminada.

Na fase seguinte, as melhores caraterísticas foram selecionadas a partir da classe *SelectKBest* da biblioteca *sklearn.feature_selection*. O tipo de teste escolhido foi *f_classif*. Neste caso as caraterísticas são escolhidas com base nos valores p obtidos no teste. Todas as caraterísticas que apresentarem um elevado valor p (significa que é improvável estarem relacionadas com o *target*) são eliminadas, sendo definido previamente um número fixo de caraterísticas (*k*) [138]. Uma vez que o número de caraterísticas ideal varia de algoritmo para algoritmo, antes da fase de classificação, cada algoritmo foi treinado com diferentes números de caraterísticas até se obter o número ideal de acordo com a métrica ACC durante a VC 10-*fold*. Para isso foi utilizada a classe *cross_val_score* da biblioteca *sklearn.model_selection*. Esta classe gera os resultados (valor ACC) de cada um dos 10 *folds*. Através de um ciclo FOR, para cada número de caraterísticas selecionado, foi estimada a média ACC dos 10 *folds* e guardada num *numpy array*. O número ideal de caraterísticas equivale ao valor ACC mais elevado presente no *array*.

Antes da afinação dos hiperparâmetros, foram gerados modelos com base nos valores pré-definidos dos parâmetros dos algoritmos. No caso do parâmetro *hidden_layer_sizes* do algoritmo MLP, equivalente ao número de camadas ocultas, o seu valor padrão é 100, sendo demasiado grande, tendo em conta o tamanho do *dataset*. Por esse motivo que foi alterado para 10 camadas ocultas. Os parâmetros e os respetivos valores para cada algoritmo encontram-se especificados na Tabela 4.3.

Tabela 4.3 - Valores pré-definidos de cada algoritmo de aprendizagem na implementação do SciKit-Learn para a deteção da DP.

Algoritmo	Parâmetro	Valor
NB	*var_smoothing*	10^{-9}
LR	*C*	1,0
	solver	lbfgs
	random_state	42
SVM-RBF	*C*	1,0
	gamma	*scale*
	kernel	rbf
	random_state	42
	probability	*True*
RF	*n_estimators*	100
	criterion	gini
	max_depth	ilimitado
	min_samples_split	2
	min_samples_leaf	1
	max_leaf_nodes	ilimitado
	max_features	auto
	random_state	0
MLP	*hidden_layer_sizes*	10
	activation	relu
	solver	adam
	alfa	0
	learning_rate	*constant*
	learning_rate_init	0,001
	max_iter	200
	random_state	0
	tol	10^{-4}
	early_stopping	*False*

Na Tabela 4.4 encontram-se os parâmetros e os respetivos valores especificados para cada algoritmo de aprendizagem na fase de classificação automática da severidade da DP. Esta tabela inclui os parâmetros definidos para cada técnica de sobre-amostragem (SMOTE e ROS).

Tabela 4.4 - Valores pré-definidos de cada algoritmo de aprendizagem na implementação do SciKit-Learn para a classificação automática da severidade da DP.

Algoritmo	Parâmetro	Valor	Parâmetro (SMOTE/ROS)	Valor
NB	*var_smoothing*	10^{-9}	-	-
LR	*C*	1.0	*random_state*	2
	solver	lbfgs		
	random_state	2		
SVM	C	1.0	*random_state*	2
	gamma	scale		
	kernel	linear		
	random_state	2		
	probability	True		
RF	*n_estimators*	100	*random_state*	2
	criterion	gini	*sampling_strategy*	0,7
	max_depth	ilimitado		
	min_samples_split	2		
	min_samples_leaf	1		
	max_leaf_nodes	ilimitado		
	max_features	auto		
	random_state	0		
DT	*criterion*	gini	*random_state*	2
	splitter	best	*sampling_strategy*	0,7
	max_depth	ilimitado		
	min_samples_split	2		
	min_samples_leaf	1		
	max_features	ilimitado		
	random_state	0		
	max_leaf_nodes	ilimitado		

Por defeito, a classe *cross_val_score* utiliza a VC *K-fold*, contudo para *datasets* mais pequenos é recomendado um tipo especial de VC K-*fold*, designada por LOOCV. Neste caso, o número de *folds* é igualado ao número de amostras no conjunto de treino (*K*=*n*), de tal forma que apenas uma amostra do conjunto de treino é utilizada no teste durante cada iteração. Por ser um processo demorado, é desaconselhado em *datasets* grandes [185]. Para efeitos de comparação, o modelo foi treinado, utilizando cada abordagem. A LOOCV é implementada a partir da classe *LeaveOneOut* da biblioteca *sklearn.model_selection*.

4.2.4 - Classificação

Nesta subsecção são apresentados os principais passos seguidos na etapa de classificação, quer na classificação automática da DP quer na classificação automática da severidade da DP. Tal como explicado na Subsecção 2.3.4 do Capítulo 2, nesta etapa, vários algoritmos de aprendizagem são treinados e avaliados, a partir do conjunto de treino, e, posteriormente, é verificada a capacidade de generalização de cada algoritmo no conjunto de teste.

Os algoritmos de aprendizagem escolhidos no modelo proposto são os algoritmos mais mencionados pelos autores dos artigos apresentados na revisão sistemática da literatura (Secção 3.1 do Capítulo 3). O modelo começa a ser treinado com algoritmos mais simples, tais como NB e LR, e termina com algoritmos mais complexos (SVM, MLP e RF). Além disso, na Tabela 2.7 (Subsecção 2.3.4 do Capítulo 2) é possível consultar as principais vantagens e desvantagens dos respetivos algoritmos.

Para a classificação automática da severidade da DP, foram escolhidos os algoritmos NB, LR, SVM linear, RF e DT. Segundo Brownlee [183], a maioria dos algoritmos de ML assumem que todos os erros de classificação cometidos pelo modelo são iguais. Contudo, em *datasets* desequilibrados isso não acontece, uma vez que o erro de classificação de um caso pertencente à classe minoritária é mais grave do que a classificação incorreta de um exemplo da classe maioritária. Para solucionar esse problema, podem ser aplicadas técnicas de sobre-amostragem que transformam o conjunto de treino num novo conjunto com uma distribuição de classes mais equilibrada. Todavia, há algumas evidências [186] de que a aplicação das referidas técnicas em algoritmos, tais como NB ou DT, pode não ter grande impacto na sua *performance*.

Afinação dos hiperparâmetros

No sentido de otimizar o desempenho do modelo deve ser encontrado um conjunto de hiperparâmetros que minimizem a função de custo. O método *Grid Search* é uma técnica poderosa de otimização dos hiperparâmetros que pode ajudar a melhorar a *performance* do modelo ao procurar exaustivamente a melhor combinação de hiperparâmetros, sendo apenas necessário especificar uma lista de valores dos diferentes hiperparâmetros [182].

Este algoritmo pode ser implementado através da classe *GridSearchCV* da biblioteca *sklearn.model_selection*. Esta classe utiliza a estratégia de divisão dos dados em VC para avaliar a *performance* de cada combinação de hiperparâmetros [187]. No modelo implementado foi definido o parâmetro *cv* com o valor de 10, para que fosse executada a VC 10-*fold*. Posteriormente, o desempenho foi comparado com a utilização da LOOCV no *Grid Search*.

A métrica escolhida para avaliar a *performance* do modelo durante a VC, definida através do parâmetro *scoring*, foi a ACC. Na Tabela 4.5 é apresentada a lista de valores dos diferentes hiperparâmetros testados em cada algoritmo de aprendizagem.

Tabela 4.5 - Lista de valores atribuídos aos diferentes hiperparâmetros testados na classificação automática da DP.

Classificador	Hiperparâmetros	Valores
NB	*var_smoothing*	[10^{-9}, 1]
LR	*C*	[10^{-3}, 10^{3}]
	solver	{'newton-cg', 'lbfgs', 'liblinear', 'saga'}
SVM	*C*	[10^{-3},10^{4}]
	gamma	[10^{-7},1]
RF	*n_estimators*	{40, 80, 100}
	max_depth	{4,5,6}
	min_samples_split	{2,3,4}
	criterion	{gini', 'entropy'}
MLP	*activation*	{‘identity’,'logistic', 'relu', 'tanh'}
	solver	{"sgd','adam'}
	alpha	{10^{-6},10^{-5}}
	early_stopping	{True, False}
	learning_rate_init	{0,01; 0,1; 0,2}
	learning_rate	{'constant', 'adaptive'}

Para o treino do modelo que prevê a severidade da DP, foi escolhida a VC estratificada 10-*fold* repetida, através da implementação da classe *RepeatedStratifiedKFold* da biblioteca *sklearn.model_selection*. Segundo Brownlee [183], é importante utilizar a VC estratificada em *datasets* desequilibrados, de modo a garantir que cada *fold* tenha a mesma distribuição de classes do conjunto de treino. Na Tabela 4.6 é apresentada a lista de valores dos diferentes hiperparâmetros testados em cada classificador durante o *Grid Search*.

Tabela 4.6 - Lista de valores atribuídos aos diferentes hiperparâmetros na classificação automática da severidade da DP.

Classificador	Hiperparâmetros	Valores
NB	*var_smoothing*	[10^{-9}, 1]
LR	*C*	[10^{-3}, 10^{3}]
SVM	*C*	[10^{-3},10^{4}]
	kernel	{‘linear’}
RF	*n_estimators*	{40, 50,100}
	max_features	{'auto', 'sqrt', 'log2'}
	max_depth	{2,3,4,5}
	min_samples_split	{2,3,4}
	criterion	{gini', 'entropy'}
DT	*max_leaf_nodes*	[2,20]
	max_depth	{2,4,6,8}
	min_samples_split	{2,3}
	criterion	{gini', 'entropy'}

Na fase de treino, foram testadas três diferentes situações: conjunto de treino com distribuição original de classes, conjunto de treino com distribuição de classes após técnica de sobreamostragem SMOTE e conjunto de treino com distribuição de classes após técnica de sobreamostragem ROS. A técnica SMOTE foi implementada através da classe *SMOTE* da biblioteca *imblearn.over_sampling* e a técnica ROS através da classe *RandomOverSampler* da mesma biblioteca. Ambas as técnicas foram aplicadas durante a execução da classe *GridSearchCV*. A métrica escolhida para avaliar a *performance* do modelo durante a VC, definida através do parâmetro scoring, foi o *roc_auc*.

Em problemas de classificação não-balanceados, a ACC é inapropriada porque esta métrica equivale ao número total de previsões corretas dividido pelo número total de previsões realizadas para o *dataset*. Em alternativa devem ser consideradas as métricas PREC, *Recall*, FM ou AUC. No caso da métrica AUC, todas as classes são consideradas igualmente importantes [138], [183].

4.2.5 – Avaliação da classificação

Para que o modelo tenha capacidade de generalização, ou seja, não haja *overfitting*, deve ser utilizado um conjunto de teste para se obter uma estimativa da sua *performance*. Na Subsecção 2.3.4 do Capítulo 2 são descritas as métricas de avaliação mais comuns: ACC, PREC, Recall e FM. Através da função *classification_report()* da biblioteca *sklearn.metrics* podem ser determinadas todas as métricas referidas. A matriz de confusão também foi considerada, sendo obtida a partir da função *confusion_matrix* da biblioteca *sklearn.metrics*.

Além disso, é possível avaliar a *performance* do modelo através da curva ROC. Já o valor AUC pode ser calculado a partir da função *roc_auc_score()* da biblioteca *sklearn.metrics*.

Em problemas de ML com classes não-balanceadas, Müller e Guido [138] recomendam a utilização da "macro" *averaging* (macro avg) das métricas PREC, *Recall* e FM, uma vez que é realizada uma média não ponderada na qual é atribuído igual peso a cada classe, independentemente do seu tamanho.

4.3 - Resumo

Neste capítulo foi apresentada a base de dados utilizada neste trabalho, bem como a metodologia proposta para o desenvolvimento de dois modelos de ML: o primeiro, para a classificação automática da DP e, o segundo, para a classificação automática da severidade da DP. Os principais passos seguidos para a implementação do primeiro modelo são apresentados no *workflow* da Figura 4.2. O *dataset* é dividido em conjunto de treino e em conjunto de teste, com posterior normalização pelo *RobustScaler*. Utilizando apenas o conjunto de treino, as caraterísticas mais discriminativas do *dataset* são selecionadas, consoante o desempenho obtido na VC 10-*fold*. Como termo de comparação, o mesmo processo é repetido utilizando a LOOCV. Seguidamente, os hiperparâmetros dos algoritmos de aprendizagem são afinados e a *performance* dos mesmos é avaliada no conjunto de teste.

Para a implementação do segundo modelo de ML foram seguidos os passos apresentados no *workflow* da Figura 4.3. Como apenas foram utilizados os dados relativos ao GDP, a dimensão do *dataset* é mais pequena e o desempenho dos algoritmos pode ser afetado pelo desequilíbrio entre classes. Nesta metodologia é proposta a divisão dos dados estratificada, assim como da

VC estratificada 10-*fold* repetida. Além disso, a *performance* dos algoritmos é comparada com a utilização de técnicas de sobre-amostragem SMOTE e ROS.

Capítulo 5

Resultados e Discussão

Ao longo do presente capítulo são apresentados os resultados obtidos, seguindo a metodologia apresentada no Capítulo 4. Uma vez que foram implementados dois modelos diferentes, um para a classificação automática da DP e outro para a classificação automática da severidade da DP, este capítulo encontra-se igualmente dividido em duas partes: Secção 5.1 e Secção 5.2

5.1 – Classificação automática da DP

5.1.1 – Resultados do pré-processamento

Na Tabela 5.1 encontram-se os valores das médias e dos desvios-padrão dos parâmetros espaciotemporais obtidos em função do passo dos participantes para o GC e GDP, respetivamente. Nesta tabela os valores encontram-se agrupados em três domínios: média das 3 tentativas, características de variabilidade (através do coeficiente de variação) e assimetria do passo. Analisando esta tabela é possível concluir que o comprimento do passo e a velocidade do passo, assim como todos os parâmetros de assimetria do passo são significativamente diferentes entre os dois grupos analisados, visto que apresentam um nível de significância inferior a 0,05. Estes resultados confirmam o que foi referido na Subsecção 2.2.3 do Capítulo 2, ou seja, a marcha na DP é, efetivamente, caraterizada por uma diminuição da velocidade da marcha e do comprimento do passo. Além disso, os resultados dos parâmetros de assimetria mostram que o GDP apresenta um padrão de marcha mais assimétrico em relação ao GC.

Tabela 5.1 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função do passo dos participantes, no GC e no GDP.

Parâmetros da marcha	GC (n=63)		GDP (n=63)		p
	Média	dp	Média	dp	
Média das 3 tentativas					
Comprimento do passo (cm)	61,82	6,29	55,76	7,88	<0,001*
Largura do passo (cm)	13,64	4,32	12,53	4,51	0,173
Duração do passo (s)	0,52	0,05	0,53	0,05	0,191
Velocidade do passo (cm/s)	121,24	17,48	107,25	18,31	<0,001*
Duração do apoio simples (s)	0,33	0,06	0,33	0,06	0,577
Duração do apoio duplo (s)	0,19	0,05	0,20	0,07	0,436
Apoio simples (%)	62,46	12,43	63,22	10,91	0,714
Apoio duplo (%)	36,79	10,07	37,28	11,51	0,795
Cadência (passos/s)	1,96	0,20	1,92	0,18	0,191
Caraterísticas de variabilidade					
Comprimento do passo (%)	4,61	2,72	5,00	2,55	0,402
Largura do passo (%)	23,54	11,68	20,72	8,38	0,122
Duração do passo (%)	3,95	2,18	4,34	2,60	0,370
Velocidade do passo (%)	5,49	2,66	6,09	2,72	0,214
Duração do apoio simples (%)	8,15	4,26	8,24	4,08	0,926
Duração do apoio duplo (%)	11,21	7,55	11,33	6,82	0,907
Apoio simples (%)	6,46	4,32	6,28	4,27	0,813
Apoio duplo (%)	11,38	7,52	11,20	6,81	0,889
Cadência (%)	3,93	2,18	4,33	2,55	0,310
Assimetria					
Comprimento do passo (%)	-0,128	2,62	5,67	3,81	<0,001*
Largura do passo (%)	0,600	10,7	23,0	14,5	<0,001*
Duração do passo (%)	-0,260	1,83	5,27	4,09	<0,001*
Velocidade do passo (%)	0,102	2,61	7,10	4,85	<0,001*
Duração do apoio simples (%)	-0,250	5,18	10,6	7,47	<0,001*
Duração do apoio duplo (%)	-0,251	6,97	12,5	9,61	<0,001*
Apoio simples (%)	-0,020	4,13	7,34	5,99	<0,001*
Apoio duplo (%)	-0,005	7,39	12,9	9,11	<0,001*

dp, desvio-padrão; p, significância

*estatisticamente significativo para p<0,05

Na Tabela 5.2 são apresentados os valores das médias e dos desvios-padrão dos parâmetros espaciotemporais obtidos em função da passada dos participantes para o GC e GDP, respetivamente. Nesta tabela os valores encontram-se agrupados em dois domínios: média das 3 tentativas e características de variabilidade (através do coeficiente de variação) da passada.

Tabela 5.2 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função da passada dos participantes, no GC e no GDP.

Parâmetros da marcha	GC (n=63)		GDP (n=63)		p
	Média	dp	Média	dp	
Média das 3 tentativas					
Comprimento da passada (cm)	123,51	12,75	111,73	16,00	<0,001*
Largura da passada (cm)	13,67	4,43	12,62	4,65	0,177
Duração da passada (s)	1,03	0,10	1,05	0,10	0,196
Velocidade da passada (cm/s)	121,21	18,21	107,13	18,48	<0,001*
Duração do apoio simples (s)	0,65	0,12	0,66	0,12	0,494
Duração do apoio duplo (s)	0,38	0,11	0,39	0,13	0,687
Apoio simples (%)	63,11	9,95	63,39	10,90	0,880
Apoio duplo (%)	36,89	9,95	36,61	10,90	0,880
Cadência (passadas/s)	0,98	0,10	0,96	0,09	0,176
Caraterísticas de variabilidade					
Comprimento da passada (%)	2,88	2,22	3,17	2,06	0,445
Largura da passada (%)	16,09	8,71	15,19	8,25	0,552
Duração da passada (%)	2,71	2,10	2,53	1,56	0,591
Velocidade da passada (%)	4,01	2,56	4,23	2,33	0,611
Duração do apoio simples (%)	5,11	3,32	4,97	2,87	0.809
Duração do apoio duplo (%)	7,46	4,95	8,56	6,74	0,299
Apoio simples (%)	3,93	2,39	4,23	2,94	0,539
Apoio duplo (%)	7,00	4,85	7,88	6,36	0,383
Cadência (%)	2,71	2,16	2,54	1,60	0,620

dp, desvio-padrão; p, significância

*estatisticamente significativo para p<0,05

Relativamente à Tabela 5.2 é possível observar que, quer o comprimento da passada, quer a velocidade da passada são significativamente diferentes entre os dois grupos, uma vez que apresentam um nível de significância inferior a 0,05, tal como foi analisado para os parâmetros do passo.

Posteriormente, a fim de averiguar a existência de *outliers* no *dataset*, foi calculada a curtose de todos os parâmetros. O *output* resultante foi visualizado na consola do software (*IPython Console*), tal como é observável na Figura 5.1.

```
Curtose:
cp. passo                          -0.095796  dur. apoio duplo (passada)          -0.237784
lar. passo                          0.286214  apoio duplo (passada)               -1.163018
dur. passo                          0.320943  dur. apoio simples (passada)        -0.645679
vel. passo                          0.146770  apoio simples (passada)             -1.163018
dur. apoio duplo (passo)            0.593857  cad. passada                         0.574382
apoio duplo (passo)                -0.923099  var. cp. passada                     6.439298
dur. apoio simples (passo)         -0.707155  var. lar. passada                    0.570321
apoio simples (passo)               4.351799  var. dur. passada                   10.406114
cad. passo                          0.565890  var. vel. passada                    1.599353
var. cp. passo                      1.713167  var dur apoio duplo (passada)        6.182133
var. lar. passo                     4.387437  var. apoio duplo (passada)           5.948108
var. dur. passo                     5.745734  var. dur. apoio simples (passada)    5.037941
var. vel. passo                     1.308731  var. apoio simples (passada)         5.775768
var. dur. apoio duplo (passo)       0.590238  var. cad. passada                   12.328822
var. apoio duplo (passo)            1.322022  assim. cp. passo                     0.879518
var. dur. apoio simples (passo)     4.229678  assim. lar. passo                    0.322871
var. apoio simples (passo)         12.541343  assim. dur. passo                    2.407908
var. cad. passo                     5.459755  assim. vel. passo                    0.913857
cp. passada                        -0.063613  assim. dur. apoio duplo (passo)      2.818310
lar. passada                        0.128520  assim. apoio duplo (passo)           3.024099
dur. passada                        0.343296  assim. dur. apoio simples (passo)    5.945600
vel. passada                        0.207973  assim. apoio simples (passo)        11.120952
```

Figura 5.1 - Valores de curtose para cada parâmetro da marcha visualizados na *IPython Console*.

De acordo com esta figura, 18 dos 44 parâmetros da marcha analisados apresentam *outliers*, uma vez que a curtose encontra-se fora do intervalo [-2,2]. Este facto reforça a necessidade de utilizar o *RobustScaler* na fase de pré-processamento dos dados.

5.1.2 - Resultados da seleção de caraterísticas

Após o *dataset* ter sido dividido em conjunto de treino e conjunto de teste, foram selecionadas as caraterísticas mais relevantes no conjunto de treino. Primeiramente foi gerada uma matriz de correlação entre os 44 parâmetros da marcha (Figura 5.2). Maioritariamente, a correlação entre caraterísticas é baixa ou inexistente. Contudo, algumas caraterísticas apresentam uma elevada correlação entre si, o que significa que não são independentes, podendo gerar redundância no conjunto de treino.

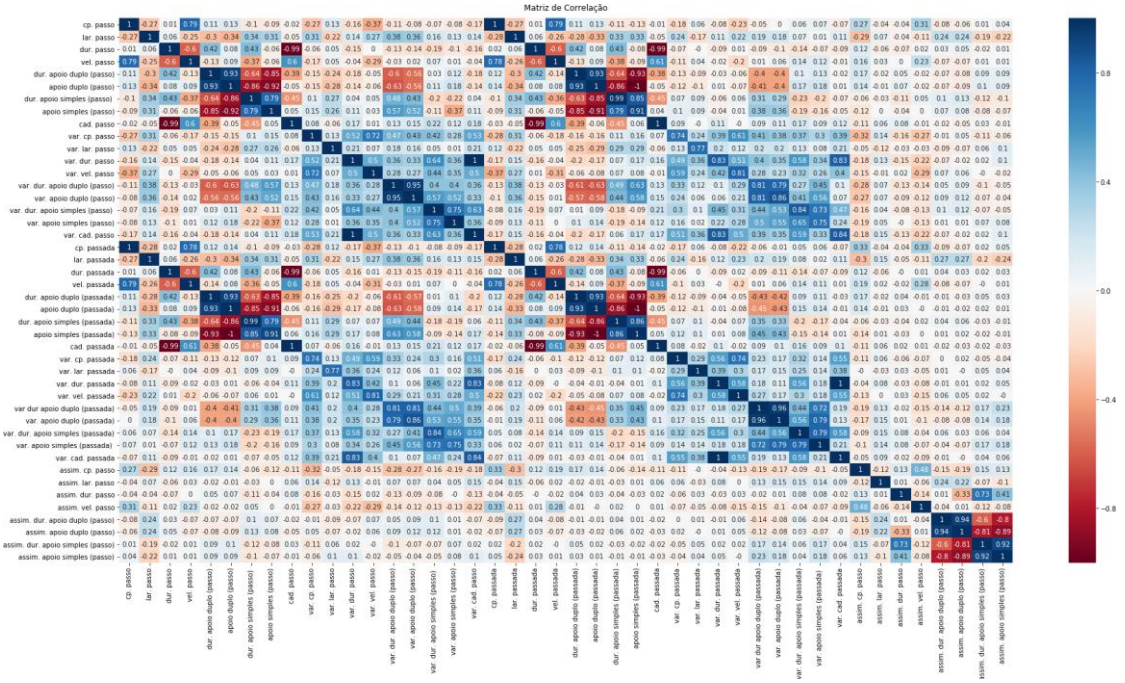


Figura 5.2 - Matriz de correlação do conjunto de treino.

Na Figura 5.3 são apresentados os 24 parâmetros da marcha selecionados. É importante realçar que, parâmetros como comprimento do passo e velocidade do passo, assim como 6 dos 8 parâmetros de assimetria do passo foram escolhidos, parâmetros estes que apresentam um nível de significância inferior a 0,05, tal como é apresentado na Tabela 5.1. Além disso, os parâmetros da marcha relativos à passada foram praticamente todos eliminados, à excepção de três: variabilidade do comprimento da passada, variabilidade da largura da passada e variabilidade do apoio simples da passada.

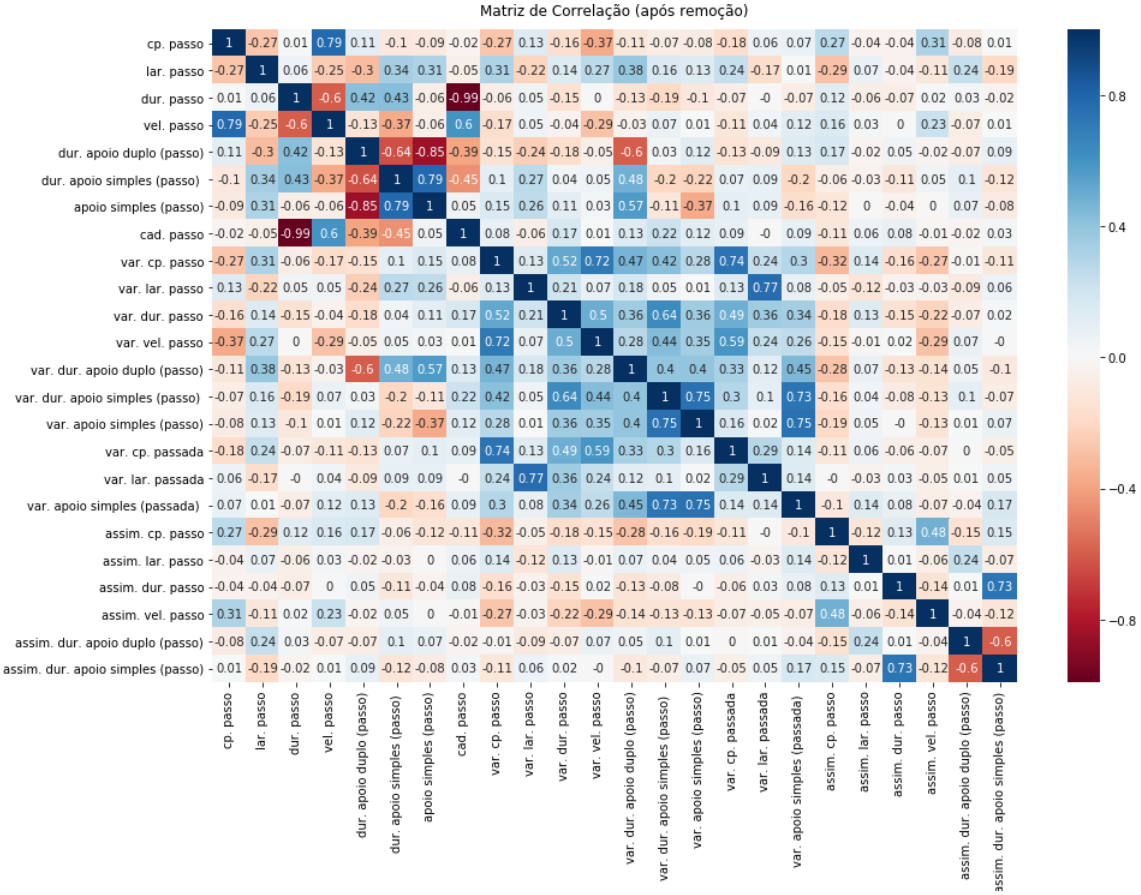


Figura 5.3 – Matriz de correlação do conjunto de treino após remoção das caraterísticas altamente correlacionadas.

Seguidamente foi aplicado o método *Filter* baseado na ANOVA, de modo a selecionar as melhores caraterísticas. Numa primeira fase, a avaliação do desempenho de cada algoritmo foi realizada através da VC 10-*fold*. A Tabela 5.3 apresenta o número de caraterísticas ideal de acordo com o algoritmo utilizado no processo de aprendizagem. Na descrição encontram-se os nomes das caraterísticas, por ordem de relevância, de acordo com os valores p obtidos no teste.

Tabela 5.3 - Número de características ideal para cada algoritmo e respetiva descrição.

Algoritmo	No. caraterísticas	Descrição
NB, SVM (RBF)	4	Comprimento do passo, velocidade do passo, largura do passo e variabilidade da largura do passo
LR, MLP	6	Comprimento do passo, velocidade do passo, largura do passo, variabilidade da largura do passo, variabilidade da duração do passo e assimetria da velocidade do passo
RF	12	Comprimento do passo, velocidade do passo, largura do passo, variabilidade da largura do passo, variabilidade da duração do passo, assimetria da velocidade do passo, assimetria da duração da apoio simples do passo, apoio simples do passo, variabilidade da largura da passada e variabilidade da velocidade do passo e duração do apoio simples do passo

Nesta tabela, o comprimento do passo, a velocidade do passo e a largura do passo destacam-se nas primeiras posições. De facto, estes parâmetros foram mencionados em vários estudos apresentados na Secção 3.1 do Capítulo 3. A variabilidade da largura do passo também foi considerada como um dos parâmetros da marcha mais clinicamente relevantes na identificação da DP no estágio inicial por Rehman *et al*. [5]. A assimetria da velocidade do passo e a assimetria da duração da apoio simples do passo foram selecionadas pelos algoritmos LR, MLP e RF, ocupando posições menos relevantes. Estes resultados estão em linha com o que foi concluido por Rehman *et al*. [5], visto que, nesse estudo, a assimetria também ocupa um lugar menos relevante.

Na Figura 5.4 estão representadas as perfomances dos diferentes classificadores em função do número de caraterísticas, durante a VC 10-*fold*. O algoritmo LR obteve a ACC mais elevada com 71,0%, ao selecionar 6 caraterísticas. Seguem-se os algoritmos MLP e SVM (RBF) ambos com ACC de 69,0% (com 6 e 4 caraterísticas, respetivamente), o algoritmo NB com ACC de 67,0% (4 caraterísticas) e, por último, o algoritmo RF obteve uma ACC mais baixa (61,0%), tendo selecionado 12 caraterísticas.

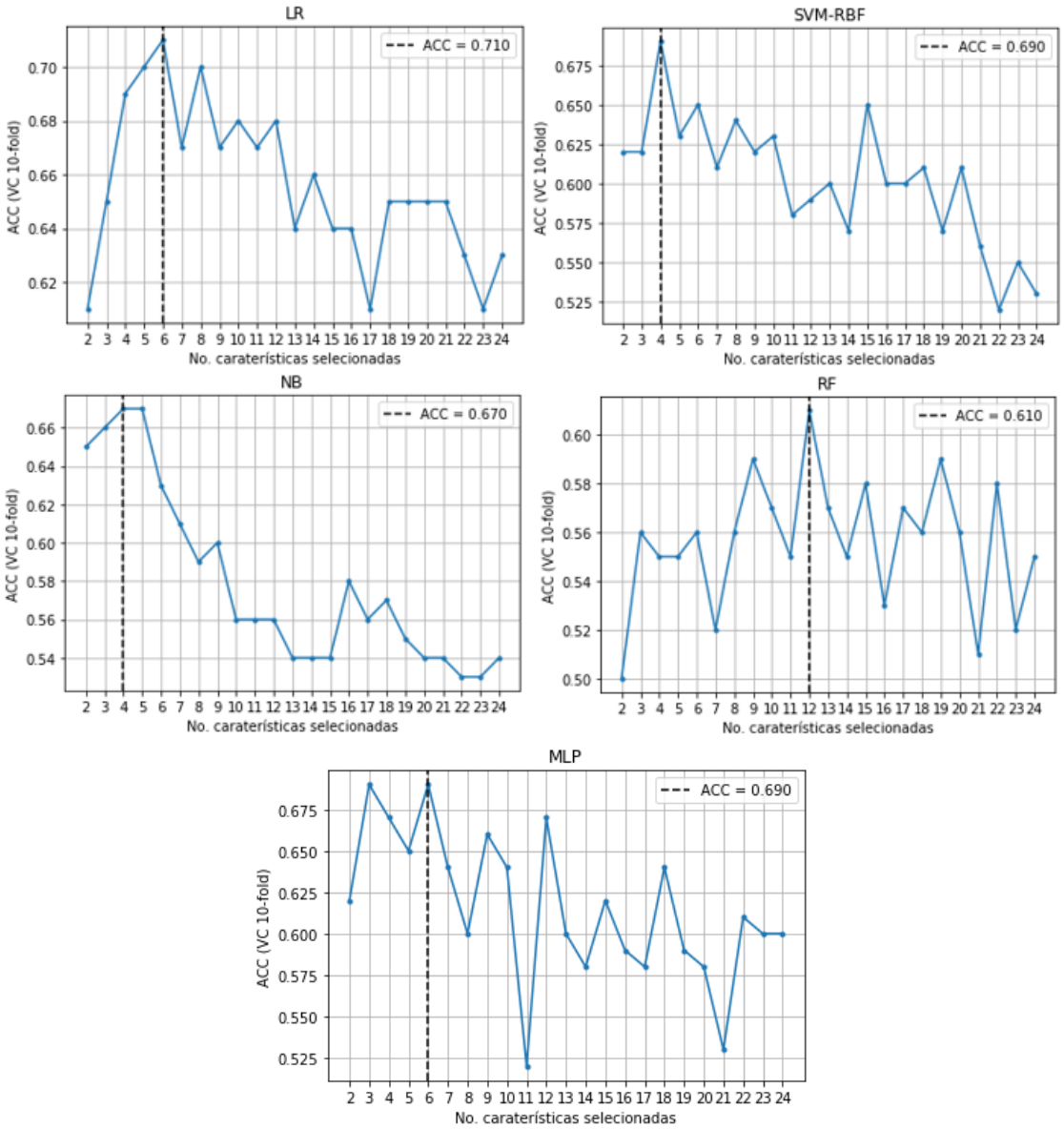


Figura 5.4 - *Performance* de cada algoritmo de aprendizagem em função do número de características selecionado pelo *SelectKBest*.

5.1.3 - Resultados da classificação

Após a seleção de características, iniciou-se o processo de aprendizagem de cada algoritmo, durante o qual foi escolhida a melhor combinação de hiperparâmetros no modelo treinado. Para auxiliar no processo de escolha desses hiperparâmetros utilizou-se o método *Grid Search*. Os melhores hiperparâmetros cujo modelo apresentou o melhor desempenho podem ser consultados na Tabela 5.4, bem como a ACC obtida no conjunto de treino (baseada na VC 10-*fold*) e no conjunto de teste. Analisando esta tabela é possível concluir que, com a exceção do algoritmo LR, todos os algoritmos beneficiaram da afinação dos hiperparâmetros, uma vez que a ACC subiu no conjunto de treino. Através destes hiperparâmetros, o desempenho do modelo foi avaliado no conjunto de teste, no qual se verifica um aumento da ACC em todos os algoritmos. Tanto o algoritmo NB como o algoritmo SVM (RFB) obtiveram uma ACC superior a 80%, com 84,6% e 80,8% respetivamente, seguidos dos algoritmos LR e MLP com a mesma ACC de 76,9%. O algoritmo RF alcançou uma ACC mais baixa com 69,2%.

Tabela 5.4 - Resultados do *GridSearchCV* (VC 10-*fold*) com valores dos hiperparâmetros ideais e a ACC obtida na conjunto de treino e do conjunto de teste.

Algoritmo	Hiperparâmetros	ACC	
		Treino (%)	Teste (%)
NB	*var_smoothing:* 0,15199110829529336	68,0	84,6
LR	*C:* 1,0 *solver:* newton-cg	71,0	76,9
SVM-RBF	*C:* 100 *gamma*: 0,001	71,0	80,8
RF	*n_estimators:*40 *max_depth:3* *min_samples_split:2* *criterion: entropy*	63,0	69,2
MLP	*activation: logistic* *solver: sgd* *alpha: 1e-06* *learning_rate_init:* 0,1 *learning_rate:* adaptative	72,0	76,9

Posteriormente, comparou-se o desempenho do mesmo modelo utilizando a LOOCV durante o treino. Os resultados obtidos e os hiperparâmetros escolhidos são apresentados na Tabela 5.5.

Tabela 5.5 - Resultados do *GridSearchCV* (LOOCV) com valores dos hiperparâmetros ideais e a ACC obtida na conjunto de treino e do conjunto de teste.

Algoritmo	No. caraterísticas	Hiperparâmetros	ACC	
			Treino (%)	Teste (%)
NB	4	*var_smoothing:* 0,01873817422860384	67,0	80,8
LR	5	*C:* 10 *solver:* newton-cg	70,0	76,9
SVM-RBF	16	*C*:10 Gamma: 0,01	67,0	80,8
RF	15	*n_estimators:*80 *max_depth:4* *min_samples_split:2* *criterion: gini*	63,0	65,4
MLP	3	*activation: identity* *solver: sgd* *alpha:* 1e-06 *early_stopping: True* *learning_rate_init: 0,2* *learning_rate:* constant	70,0	76,9

Ao comparar a Tabela 5.4 com a Tabela 5.5 constata-se que o desempenho da VC 10-fold, de modo geral, é superior à LOOCV. No caso da LOOCV, o algoritmo NB obteve uma pior *performance* nos dois conjuntos. Os algoritmos LR, SVM (RBF) e MLP tiveram um pior desempenho no conjunto de treino e o algoritmo RF obteve um pior resultado no conjunto de teste. Neste caso, a utilização da VC 10-*fold* é mais vantajosa do que a LOOCV.

Para uma análise mais aprofundada do processo de otimização dos hiperparâmetros, é possível analisar graficamente o impacto da alteração de um determinado hiperparâmetro na *performance* de um modelo. Na Figura 5.5 está representada a evolução da ACC no conjunto de treino e no conjunto de validação, para o algoritmo NB com diferentes valores do hiperparâmetro *var_smoothing*. À medida que o valor deste hiperparâmetro aumenta, a ACC é aproximadamente constante em ambos os conjuntos, cuja ACC no conjunto de validação é 67%, ligeiramente inferior ao conjunto de treino (aproximadamente 68%). No intervalo entre 0,01 e 1, há uma diminuição da ACC no conjunto de treino e uma subida da ACC no conjunto de validação (cerca de 68%). Quando o *var_smoothing* se aproxima de 1, a ACC desce abruptamente no conjunto de validação para 66%. No fim, volta a subir e aproxima-se do conjunto de treino, com uma ACC de 67%. Portanto, neste algoritmo não se verifica a existência de *overfitting*.

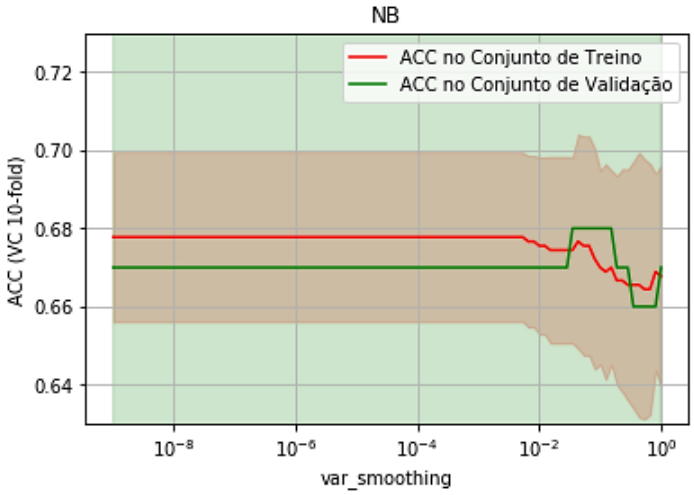


Figura 5.5 - Análise da evolução da ACC no conjunto de treino e no conjunto de validação, para o algoritmo NB com diferentes valores do hiperparâmetro *var_smoothing*.

Na Figura 5.6 está representada a importância das caraterísticas para o algoritmo RF. O comprimento do passo é a caraterística mais importante para este algoritmo, ou seja, é a caraterística mais discriminativa do *dataset*, uma vez que apresenta a importância relativa mais elevada. Seguem-se a variabilidade da velocidade do passo e a velocidade do passo, com a mesma importância relativa. A caraterística com a menor importância relativa é a assimetria da duração do apoio simples do passo.

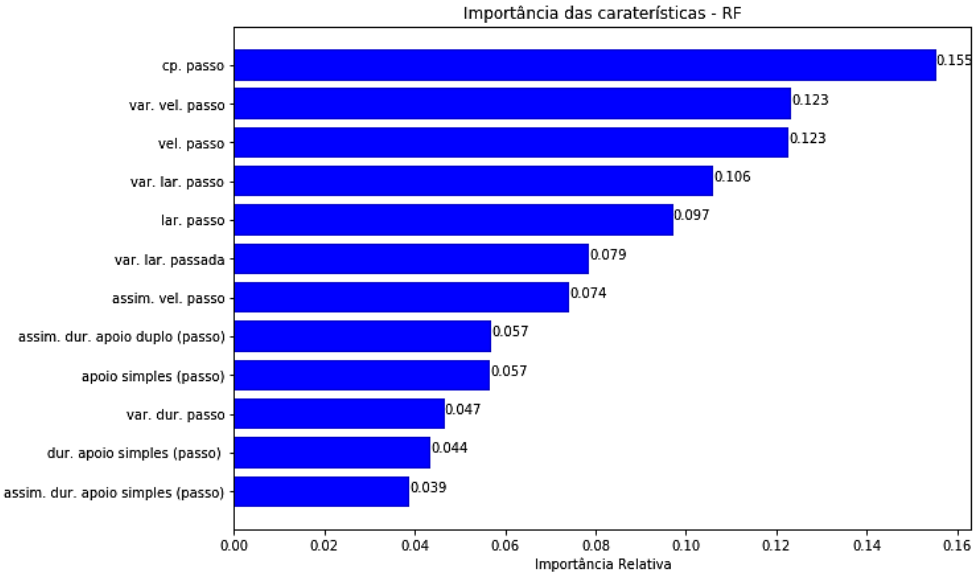


Figura 5.6 - Importância das caraterísticas para o algoritmo RF no conjunto de teste.

A Figura 5.7 mostra o *heatmap* com os valores de ACC obtidos em função dos valores *C* e *gamma* para o algoritmo SVM (RBF), durante afinação dos hiperparâmetros. *C* e *gamma* controlam a complexidade do modelo, de modo que valores mais elevados correspondem a um modelo mais complexo. No *heatmap*, as cores mais claras representam uma ACC mais elevada, enquanto as cores mais escuras representam uma ACC mais baixa. Para a maioria das combinações apresentadas, a ACC ronda os 52%. Segundo o *heatmap*, há três combinações através das quais é possível obter uma ACC mais elevada, cerca de 71%.

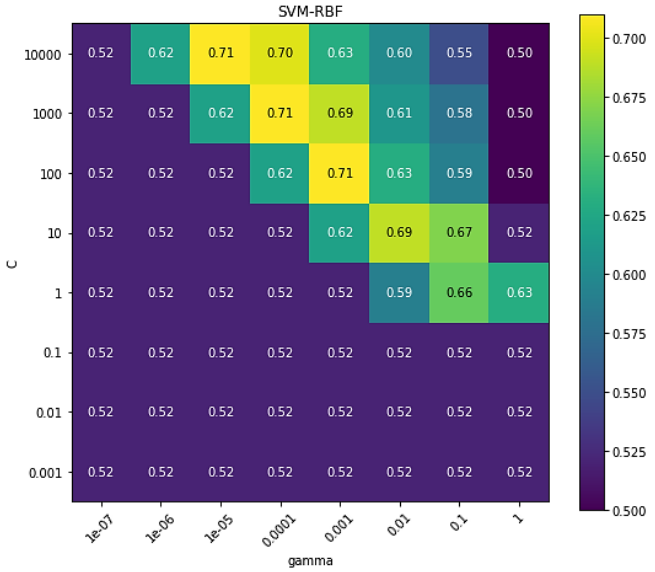


Figura 5.7 - *Heatmap* dos valores ACC obtidos para o algoritmo SVM (RBF), variando os valores dos hiperparâmetros *C* e *gamma*.

Na Figura 5.8 é apresentado outro *heatmap* dos valores de ACC obtidos utilizando o algoritmo LR, quando se variou os hiperparâmetros C e *solver*, durante a implementação do método *Grid Search*. De todas as combinações treinadas, quatro sobressaíram com uma ACC que ronda os 71%.

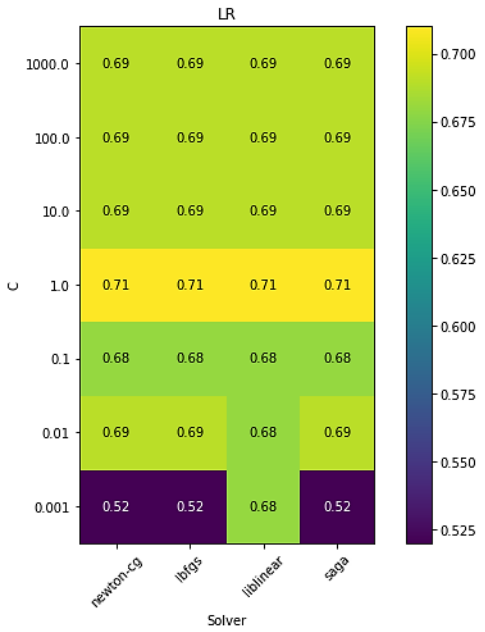


Figura 5.8 - *Heatmap* dos valores de ACC obtidos para o algoritmo LR, variando os valores dos hiperparâmetros *C* e *solver*.

A partir das figuras anteriores é possível verificar que o ajuste dos hiperparâmetros é fundamental para se obter um bom desempenho do modelo no conjunto de teste.

Na Subsecção 2.3.4 do Capítulo 2 foram apresentadas diferentes métricas para avaliar a *performance* do classificador. Os resultados obtidos podem ser consultados na Tabela 5.6. Nesta tabela destaca-se o algoritmo NB com ACC de 84,6%, PREC de 0,923 e *Recall* de 0,800, seguido do algoritmo SVM (RBF) com ACC de 80,8%, PREC de 0,917 e *Recall* de 0,733. Os algoritmos LR e MLP obtiveram a mesma ACC com 76,9%, PREC de 0,909 e *Recall* de 0,667.

Todos algoritmos alcançaram uma PREC superior ao *Recall*, com valores próximos de 0,9. Ou seja, o modelo tem um baixo número de FP, mas um número de FN mais elevado. Portanto, sempre que houver erros de classificação, é mais provável que pacientes com DP sejam classificados como saudáveis do que o contrário. Em contexto médico, o número de FP deve ser o mínimo possível, devido aos elevados custos associados ao tratamento da DP.

Tabela 5.6 – Resultados da *performance* dos diferentes algoritmos de aprendizagem após a utilização do método *Grid Search* (VC 10-*fold*).

Algoritmo	ACC (%)	PREC	*Recall*	FM
NB	84,6	0,923	0,800	0,857
LR	76,9	0,909	0,667	0,769
SVM (RBF)	80,8	0,917	0,733	0,815
RF	69,2	0,889	0,533	0,667
MLP	76,9	0,909	0,667	0,769

Os resultados da matriz de confusão são apresentados na Tabela 5.7. Tanto o algoritmo NB como o algoritmo SVM (RBF) obtiveram um número de VP superior ao número de VN, ou seja, estes algoritmos detetaram corretamente mais casos de pacientes com DP. O algoritmo DT obteve igual número de VP e VN. Os algoritmos LR e RF obtiveram um número de VN superior ao número de VP, isto é, identificaram corretamente mais casos de indivíduos pertencentes ao

GC. Todos os algoritmos apresentaram um baixo número de FP, mas, em alguns casos, o número de FN é elevado. Esta dificuldade em identificar corretamente a DP prende-se com o facto de, numa fase inicial, o seu diagnóstico ser desafiante.

Tabela 5.7 - Resultado da matriz de confusão após a utilização do método *Grid Search* (VC 10-*fold*).

Algoritmo	VP	FP	FN	VN
NB	12	1	3	10
LR	8	0	7	11
SVM-RBF	11	1	4	10
RF	8	1	7	10
DT	10	1	5	10

A Figura 5.9 mostra o impacto da utilização de todas as caraterísticas, da sua seleção após a remoção das caraterísticas altamente correlacionadas e da sua seleção após o *SelectKBest* no desempenho dos diferentes algoritmos de aprendizagem na VC 10-*fold*. Todos os algoritmos de aprendizagem melhoraram a sua ACC após a seleção de caraterísticas. O algoritmo NB foi o mais beneficiado, visto que a sua ACC subiu significativamente, de 61,5% para 84,6%. Para o algoritmo LR, a utilização do *SelectKBest* não teve qualquer impacto, pois manteve a ACC em 76,9%.

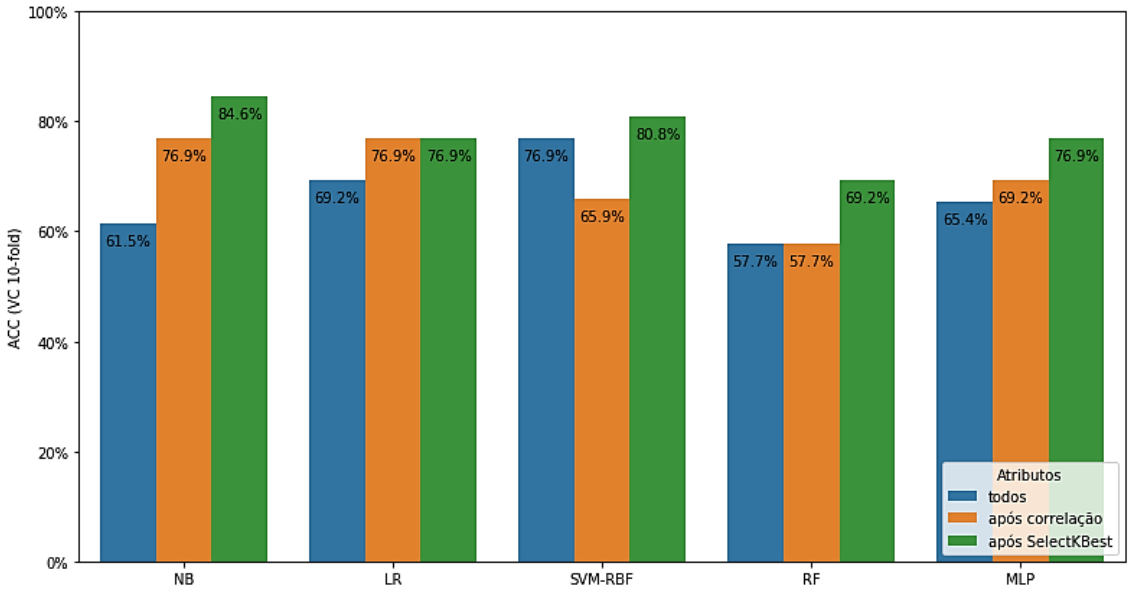


Figura 5.9 - Impacto do método de seleção de caraterísticas no desempenho dos diferentes algoritmos de aprendizagem na VC 10-*fold*.

Seguidamente é apresentada a curva ROC e respetivo valor AUC para os diferentes algoritmos (Figura 5.10).

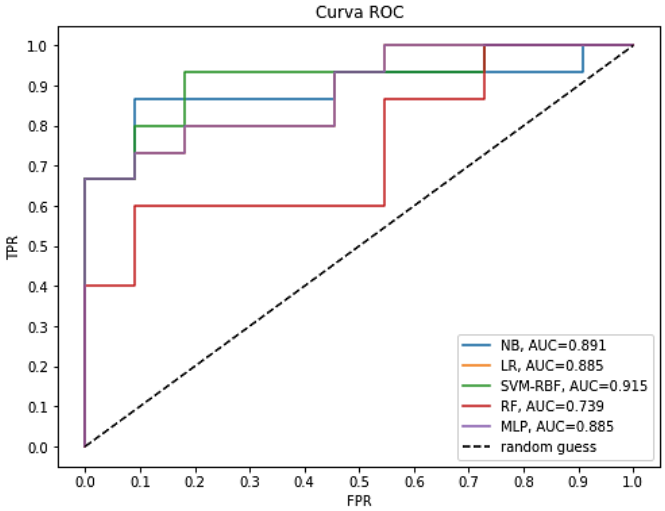


Figura 5.10 - Comparação entre os diferentes algoritmos de aprendizagem através da curva ROC na VC 10-*fold*.

O algoritmo SVM (RBF) obteve uma *performance* superior em relação aos restantes algoritmos com um valor AUC de 0,915. Este algoritmo conseguiu mais facilmente distinguir os dois grupos em análise. Do lado contrário encontra-se o algoritmo RF, que teve o pior desempenho com um valor AUC de 0,739.

Por fim, os resultados obtidos mostram que o modelo proposto apresenta uma boa exatidão podendo ser aplicado clinicamente no diagnóstico da DP. Contudo, o modelo apenas é válido para este *dataset*, sendo necessário validar de forma independente noutros *datasets*. Como o *dataset* utilizado na fase de aprendizagem é pequeno, provavelmente, haveria *overfitting*. Para evitar isso, o *dataset* teria que ser maior e mais representativo da população.

Além disso, os resultados da classificação podem ter sido afetados pelas discrepâncias encontradas, principalmente, relativamente ao género, IMC e velocidade escolhida pela participante que podem ter aumentado a dispersão dos parâmetros da marcha.

Um ponto interessante do modelo é a inclusão da largura do passo como parâmetro clinicamente relevante para o diagnóstico da DP, juntamente com o comprimento e a velocidade do passo (parâmetros convencionalmente discriminantes). Ao contrário de estudos anteriores relacionados com parâmetros discriminantes da marcha em pacientes com DP, a estabilidade mediolateral parece ser um fator importante para distinguir pacientes com DP de pessoas sem a doença. Pacientes com DP apresentam um agravamento da estabilidade mediolateral, necessitando de aumentar a largura do passo para desempenhar a marcha devido a uma diminuição na margem de estabilidade. Este é um ponto inovador deste trabalho que deve ser aprofundado em trabalhos futuros.

5.2 – Classificação automática da severidade da DP

5.2.1 – Resultados do pré-processamento

A fase de pré-processamento para classificação automática da severidade da DP seguiu uma metodologia similar à utilizada para a classificação automática da DP. Na Tabela 5.8 encontram-se os valores das médias e dos desvios-padrão dos parâmetros espaciotemporais obtidos

em função do passo dos participantes para o GDPL e GDPI, respetivamente. Nesta tabela, os valores encontram-se agrupados em três domínios: média das 3 tentativas, características de variabilidade (através do coeficiente de variação) e assimetria do passo.

Tabela 5.8 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função da passo dos participantes, no GDPL e no GDPI.

Parâmetros da marcha	GDPL (n=42)		GDPI (n=21)		p
	Média	dp	Média	dp	
Média das 3 tentativas					
Comprimento do passo (cm)	56,83	7,36	53,78	8,29	0,142
Largura do passo (cm)	12,29	4,40	13,20	3,88	0,422
Duração do passo (s)	0,53	0,05	0,52	0,05	0,997
Velocidade do passo (cm/s)	109,6	18,47	102,55	16,27	0,776
Duração do apoio simples (s)	0,34	0,06	0,33	0,06	0,554
Duração do apoio duplo (s)	0,20	0,07	0,19	0,06	0,347
Apoio simples (%)	62,55	10,64	64,55	10,93	0,343
Apoio duplo (%)	38,21	11,51	35,40	10,90	0,262
Cadência (passos/s)	1,92	0,19	1,92	0,18	0,975
Caraterísticas de variabilidade					
Comprimento do passo (cm)	4,56	2,00	5,88	3,27	0,129
Largura do passo (cm)	19,28	8,16	23,59	8,25	0,054
Duração do passo (s)	3,93	1,79	5,15	3,66	0,374
Velocidade do passo (cm/s)	5,57	2,04	7,12	3,57	0,088
Duração do apoio simples (s)	7,30	2,69	10,11	5,60	0,066
Duração do apoio duplo (s)	9,42	5,90	15,15	7,07	0,002*
Apoio simples (%)	5,23	2,27	8,37	6,25	0,012*
Apoio duplo (%)	9,44	5,96	14,73	7,17	0,003*
Cadência (passos/s)	3,94	1,84	5,11	3,51	0,343
Assimetria					
Comprimento do passo (%)	5,09	3,42	6,82	4,35	0,162
Largura do passo (%)	22,5	14,4	24,1	15,0	0,827
Duração do passo (%)	4,74	3,23	6,33	5,37	0,382
Velocidade do passo (%)	6,46	4,49	8,40	5,37	0,185
Duração do apoio simples (%)	9,75	5,24	12,3	10,6	0,705
Duração do apoio duplo (%)	10,6	8,67	16,1	10,5	0,044*
Apoio simples (%)	6,19	3,45	9,63	8,87	0,054
Apoio duplo (%)	11,4	9,02	16,0	8,70	0,023*

dp, desvio-padrão; p, significância

*estatisticamente significativo para p<0,05

Analisando esta tabela é possível constatar que a variabilidade da duração do apoio duplo, a variabilidade do apoio duplo, a variabilidade da duração do apoio simples e a variabilidade do apoio simples são significativamente diferentes entre os dois grupos analisados. Além disso, a assimetria da duração do apoio duplo e a assimetria do apoio duplo também são significativamente diferentes. Para qualquer um dos parâmetros referidos, o nível de significância é inferior a 0,05. De facto, a DP, numa fase intermédia, apresenta uma ligeira instabilidade postural, o que aumenta o risco de queda. Mais ainda, nesta fase, esta doença afeta, bilateralmente, os membros. Embora não sejam significativamente diferentes, a velocidade do passo e o comprimento do passo são menores no GDPI.

Na Tabela 5.9 são apresentados os valores das médias e dos desvios-padrão dos parâmetros espaciotemporais obtidos em função da passada dos participantes para o GDPL e GDPI, respetivamente. Nesta tabela os valores encontram-se agrupados em dois domínios: média das 3 tentativas e características de variabilidade (através do coeficiente de variação) da passada.

Tabela 5.9 - Médias e desvios-padrão dos parâmetros espaciotemporais, em função do passada dos participantes, no GDPL e no GDPI.

Parâmetros da marcha	GDPL (n=42)		GDPI (n=21)		p
	Média	dp	Média	dp	
Média das 3 tentativas					
Comprimento da passada (cm)	113,81	14,95	107,6	16,81	0,141
Largura da passada (cm)	12,31	4,46	13,24	3,92	0,423
Duração da passada (s)	1,05	0,10	1,05	0,10	0,985
Velocidade da passada (cm/s)	109,29	18,46	102,83	16,96	0,185
Duração do apoio simples (s)	0,66	0,12	0,68	0,12	0,460
Duração do apoio duplo (s)	0,40	0,11	0,38	0,14	0,351
Apoio simples (%)	62,57	10,68	65,03	10,61	0,336
Apoio duplo (%)	37,43	10,68	34,97	10,61	0,336
Cadência (passadas/s)	0,96	0,09	0,96	0,08	0,996
Caraterísticas de variabilidade					
Comprimento da passada (%)	2,91	1,86	3,71	2,39	0.126
Largura da passada (%)	13,42	7,49	18,72	8,72	0,012*
Duração da passada (%)	2,35	1,21	2,88	2,09	0,590
Velocidade da passada (%)	3,88	1,94	4,95	2,88	0,141
Duração do apoio simples (%)	4,37	2,32	6,17	3,50	0,038*
Duração do apoio duplo (%)	7,04	5,58	11,58	7,91	0,010*
Apoio simples (%)	3,61	2,22	5,46	3,78	0,024*
Apoio duplo (%)	6,46	5,40	10,71	7,28	0,002*
Cadência (%)	2,36	1,23	2,89	2,16	0,610

dp, desvio-padrão; p, significância

*estatisticamente significativo para p<0,05

Nesta tabela, a variabilidade da largura da passada, a variabilidade da duração do apoio simples, a variabilidade do apoio simples, a variabilidade da duração do apoio duplo e a variabilidade do apoio duplo são significativamente diferentes entre os grupos estudados, pois o

nível de significância é inferior a 0,05. No GDPI, os valores obtidos nos parâmetros mencionados são superiores ao GDPL.

Tal como para a classificação automática da DP, os valores de curtose foram analisados, tendo sido identificados 16 parâmetros com *outliers* (Figura 5.11). Novamente, este facto justificou a utilização do *RobustScaler* na fase de pré-processamento dos dados.

```
Curtose:
cp. passo                          -0.487605
lar. passo                          0.943518
dur. passo                          0.761222
vel. passo                          0.270362
dur. apoio duplo (passo)            0.613861
apoio duplo (passo)                -0.828522
dur. apoio simples (passo)         -0.269805
apoio simples (passo)              -1.181129
cad. passo                          0.600019
var. cp. passo                      2.374530
var. lar. passo                    -1.082628
var. dur. passo                     5.980346
var. vel. passo                     2.042400
var. dur. apoio duplo (passo)      -0.144336
var. apoio duplo (passo)            1.318316
var. dur. apoio simples (passo)     6.642330
var. apoio simples (passo)         15.093821
var. cad. passo                     4.945133
cp. passada                        -0.517302
lar. passada                        0.633387
dur. passada                        0.847052
vel. passada                        0.167222
dur. apoio duplo (passada)         -0.191289
apoio duplo (passada)              -1.130390
dur. apoio simples (passada)       -0.144673
apoio simples (passada)            -1.130390
cad. passada                        0.648429
var. cp. passada                    6.006776
var. lar. passada                  -0.413024
var. dur. passada                   1.842987
var. vel. passada                   3.315535
var dur apoio duplo (passada)       6.164945
var. apoio duplo (passada)          6.092232
var. dur. apoio simples (passada)   5.024582
var. apoio simples (passada)        6.870785
var. cad. passada                   2.732964
assim. cp. passo                    0.396754
assim. lar. passo                   0.090096
assim. dur. passo                   2.384085
assim. vel. passo                   0.749109
assim. dur. apoio duplo (passo)     0.723847
assim. apoio duplo (passo)          0.600212
assim. dur. apoio simples (passo)   3.446886
assim. apoio simples (passo)        7.199756
```

Figura 5.11 - Valores de curtose para cada parâmetro da marcha visualizados na *IPython Console*.

5.2.2 – Resultados da seleção de caraterísticas e classificação

De modo semelhante à classificação automática da DP, para a seleção de caraterísticas, primeiramente foram eliminadas as que fossem altamente correlacionadas entre si. A Figura 5.12 mostra a matriz de correlação resultante das caraterísticas presentes no conjunto de treino.

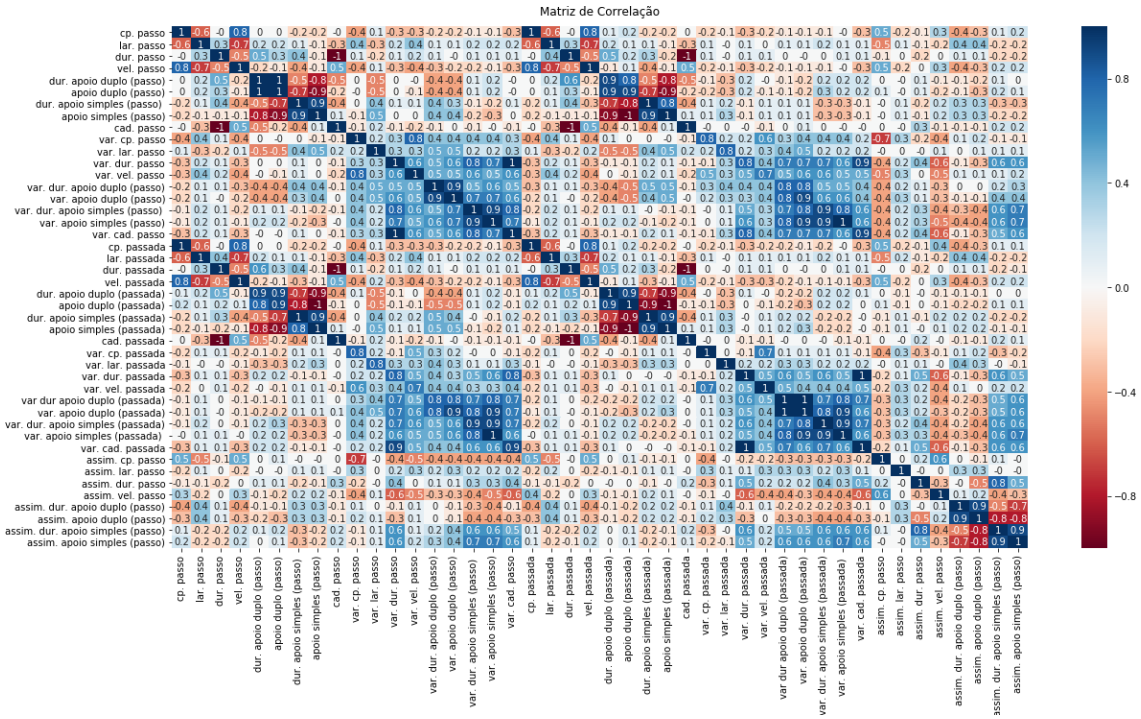


Figura 5.12 - Matriz de correlação do conjunto de treino.

Na Figura 5.13 é apresentada a matriz resultante com as 20 caraterísticas selecionadas. Se compararmos com a Tabela 5.8, a variabilidade da duração do apoio duplo e do apoio simples

do passo foram ambas selecionadas, assim como a assimetria da duração do apoio duplo. Em relação aos parâmetros da passada apenas foram escolhidos três: variabilidade do comprimento da passada, variabilidade da largura da passada e variabilidade da duração da passada. Dos três parâmetros, só a variabilidade da largura da passada é significativamente diferente entre os grupos, de acordo com a Tabela 5.9.

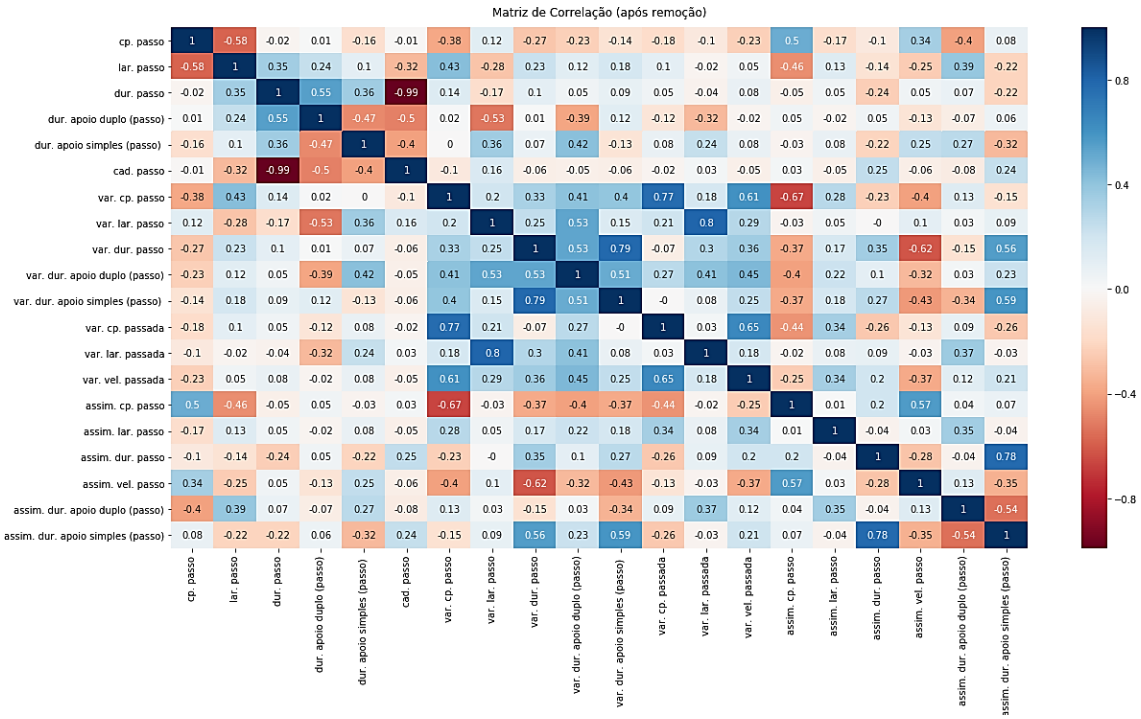


Figura 5.13 - Matriz de correlação do conjunto de treino após remoção das caraterísticas altamente correlacionadas.

Dados originais

Na fase de aprendizagem do modelo foram utilizados, numa primeira fase, conjuntos de treino e de teste com a distribuição de classes igual ao *dataset* original. Posteriormente, o conjunto de treino foi equilibrado com técnicas de sobre-amostragem de dados. A Figura 5.14 apresenta ocorrência das duas classes no conjunto de treino original, antes da sobre-amostragem. A *label* 0, correspondente ao GDPL, está em superioridade numérica em relação à label 1 (GDPI).

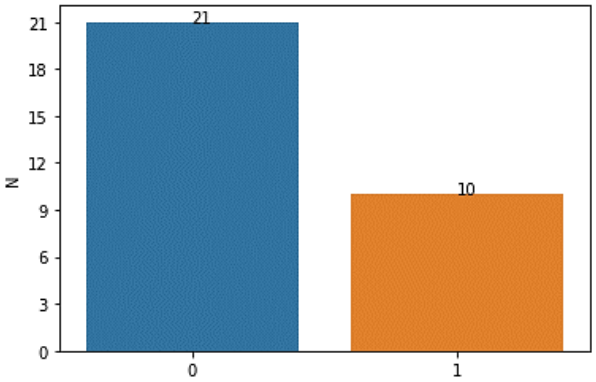


Figura 5.14 - Distribuição das classes no conjunto de treino original.

De forma similar à classificação automática da DP, as caraterísticas foram selecionadas através da classe *SelectKBest* do *SciKit-Learn*. Os nomes das caraterísticas selecionadas encontram-se na Tabela 5.10. Todos os algoritmos escolheram apenas duas caraterísticas: variabilidade da largura da passada e variabilidade da duração do apoio duplo do passo. Nesta tabela encontram-se os valores AUC obtidos no conjunto de treino (através da VC 10-*fold*) e no conjunto de teste. Os algoritmos NB, LR, SVM linear e RF alcançaram o mesmo resultado no treino com um AUC de 0,800. Relativamente ao conjunto de teste, houve pequenas diferenças nos resultados, entre os quais se destaca o algoritmo NB com um valor AUC de 0,766. O algoritmo DT obteve um pior desempenho em ambos os conjuntos, com um valor AUC de 0,725 e 0,695, respetivamente.

Tabela 5.10 - Resultados da seleção de caraterísticas e do GridSearchCV com valores dos hiperparâmetros e os valores AUC obtidos no conjunto do treino e do teste (dados originais).

Algoritmo	**Caraterísticas selecionadas pelo *SelectKBest***	**Hiperparâmetros**	**AUC**	
			Treino	**Teste**
NB	Variabilidade da largura da passada e variabilidade da duração do apoio duplo do passo	*var_smoothing*:1,0	0,800	0,766
LR		*C*: 0,001 *solver*: 'newton-cg'	0,800	0,762
SVM linear		*C*: 0,001	0,800	0,762
RF		*n_estimators*:40 *max_depth*:2 *min_samples_split*:2 *criterion: gini*	0,800	0,760
DT		*max_depth*:2 *min_samples_split*:2 *criterion: gini* *splitter: random* *max_leaf_nodes: 6*	0,725	0,695

Na Figura 5.15 está representada a importância das caraterísticas para o algoritmo RF. A variabilidade da duração do apoio duplo do passo é a caraterística mais importante para este algoritmo e, portanto, a mais discriminativa do *dataset*. Em segundo lugar ficou a variabilidade da largura da passada, embora a diferença entre as importâncias relativas seja pequena.

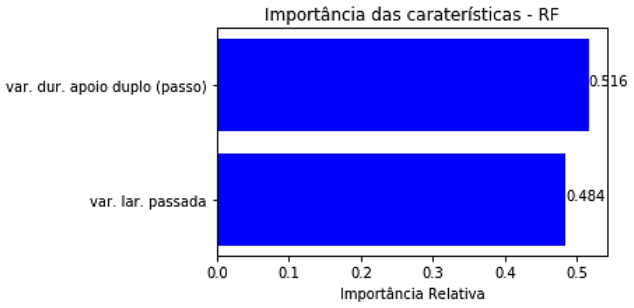


Figura 5.15 - Importância das caraterísticas para o algoritmo RF no conjunto de treino do *dataset* original, utilizando os melhores hiperparâmetros.

Os resultados da matriz de confusão obtidos utilizando os dados originais são apresentados na Tabela 5.11. Todos os algoritmos obtiveram um número de VN superior ao número de VP, ou seja, há mais casos de GDPL corretamente identificados do que casos de GDPI. Quer o algoritmo LR quer o algoritmo SVM linear não foram capazes de identificar o GDPI. Os resultados mostram que, neste caso, os dois algoritmos apresentaram uma sensibilidade maior perante a existência de um desequilíbrio entre classes e que é necessário utilizar técnicas de sobre-amostragem.

Tabela 5.11 - Resultado da matriz de confusão utilizando os dados originais.

Algoritmo	VP	FP	FN	VN
NB	5	2	6	19
LR	0	0	11	21
SVM linear	0	0	11	21
RF	6	3	5	18
DT	9	9	2	12

Dados com sobre-amostragem

Uma vez que o *dataset* tem uma dimensão pequena, a diferença no número de *labels* entre classes pode ter um grande impacto no desempenho do modelo. Por esse motivo foram aplicadas técnicas de sobre-amostragem nos dados. Primeiramente, aplicou-se a técnica SMOTE de modo a equilibrar as duas classes no conjunto de treino e comparou-se os resultados com os obtidos pela técnica ROS. A distribuição das duas classes após a aplicação da técnica SMOTE encontra-se representada na Figura 5.16.

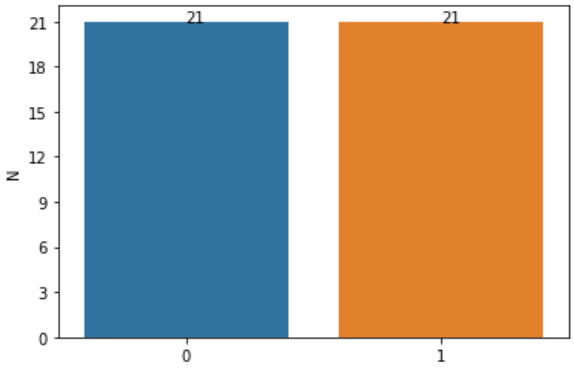


Figura 5.16 - Distribuição das duas classes no conjunto de treino após a aplicação da técnica de sobre-amostragem SMOTE.

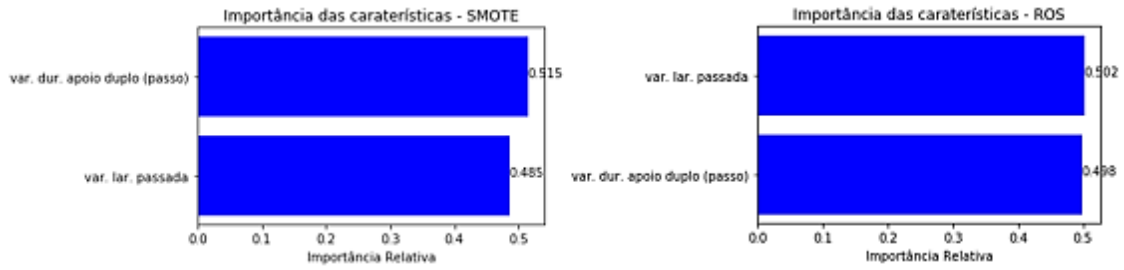
Os resultados relativos aos dados com sobre-amostragem são apresentados na Tabela 5.12. À semelhança dos dados originais, todos os algoritmos escolheram apenas duas caraterísticas: variabilidade da largura da passada e variabilidade da duração do apoio duplo do passo. Os algoritmos NB, LR, SVM linear obtiveram o mesmo resultado no treino com um AUC de 0,800, tal como tinha acontecido com os dados originais. O algoritmo RF teve uma *performance* mais baixa, diferindo, também, nos resultados entre a técnica SMOTE e a técnica ROS, com um AUC de 0,683 e 0,750, respetivamente. Contudo, este algoritmo alcançou, no conjunto de teste, o melhor resultado, em comparação com os restantes algoritmos. Além disso, não se encontraram diferenças entre as duas técnicas de sobre-amostragem. O algoritmo DT também obteve valores AUC diferentes, se forem comparadas as duas técnicas, no conjunto de treino. Estes valores

são superiores ao valor obtido pelo algoritmo DT no conjunto de treino com os dados originais. No conjunto de teste, há uma diminuição do desempenho nas duas técnicas. Os hiperparâmetros selecionados são diferentes entre técnicas, sendo que a técnica SMOTE obteve um valor AUC de 0,706, superior ao obtido pela técnica ROS com 0,600.

Tabela 5.12 - Resultados do *GridSearchCV* com valores dos hiperparâmetros e com valores do AUC do treino e do teste (dados com sobre-amostragem).

Algoritmo	No. atributos	Técnica de amostragem	Hiperparâmetros	AUC	
				Treino (%)	Teste (%)
NB	2	SMOTE	*var_smoothing*:1,0	0,800	0,771
		ROS		0,800	0,762
LR	2	SMOTE	*C:* 0,001	0,800	0,766
		ROS	*solver*: 'newton-cg'	0,800	0,775
SVM linear	2	SMOTE	*C:* 0.001	0,800	0,766
		ROS		0,800	0,775
RF	2	SMOTE	*n_estimators:4*0	0,683	0,786
		ROS	*max_depth:2* *min_samples_split:*2 *criterion: gini*	0,750	0,786
DT	2	SMOTE	*max_depth:4* *min_samples_split:*2 *criterion: gini* *splitter: random* *max_leaf_nodes: 7*	0,800	0,706
		ROS	*max_depth:6* *min_samples_split:3* *criterion: gini* *splitter: random* *max_leaf_nodes: 10*	0,775	0,600

Na Figura 5.17 é apresentada a importância das caraterísticas selecionadas pelo algoritmo RF, utilizando a técnica SMOTE (a) e a técnica ROS (b). A técnica SMOTE atribuiu maior importância à variabilidade da duração do apoio duplo do passo do que à variabilidade da largura da passada, com importâncias relativas muito próximas das obtidas com os dados originais (Figura 5.15). A técnica ROS considerou a variabilidade da largura da passada mais importante do que a variabilidade da duração do apoio duplo do passo, embora a diferença entre as importâncias relativas seja muito pequena. Pelo facto de estas técnicas terem dado importâncias diferentes às caraterísticas, isso explica as diferenças obtidas nos valores AUC no conjunto de treino.



(a) **(b)**

Figura 5.17 - Importância das caraterísticas para o algoritmo RF **a)** utilizando a técnica SMOTE; **b)** utilizando a técnica ROS.

Os resultados da matriz de confusão obtidos utilizando as técnicas SMOTE e ROS são apresentados na Tabela 5.13. Apenas o algoritmo DT obteve resultados diferentes, se forem comparadas as duas técnicas. De modo geral, os algoritmos identificaram corretamente mais casos de GDPL, visto que o número de VN é superior ao VP. Além disso, utilizando as técnicas de sobre-amostragem, os algoritmos LR e SVM linear foram capazes de identificar o GDPI.

Tabela 5.13 - Resultados da matriz de confusão utilizando as técnicas SMOTE e ROS.

Técnica de Amostragem	Algoritmo	VP	FP	FN	VN
SMOTE	NB	8	5	3	16
	LR	7	5	4	16
	SVM	7	5	4	16
	RF	7	4	4	17
	DT	7	4	4	17
ROS	NB	8	5	3	16
	LR	7	5	4	16
	SVM	7	5	4	16
	RF	7	4	4	17
	DT	6	3	5	18

Os resultados das métricas AUC, PREC (macro avg), *Recall* (macro avg), FM (macro avg) e ACC dos diferentes algoritmos encontram-se representados na Figura 5.18. Analisando a figura é possível concluir que as técnicas de sobre-amostragem favoreceram mais os algoritmos LR e SVM linear, devido às diferenças substanciais nas métricas PREC (macro avg), *Recall* (macro avg) e FM (macro avg) em comparação com os dados originais. Porém, no caso dos algoritmos NB e RF, a precisão baixou ligeiramente quando se aplicaram as referidas técnicas. Em relação à ACC, estes algoritmos também não foram particularmente beneficiados, uma vez que se manteve igual. Em todos os algoritmos, as métricas *Recall* (macro avg) e FM (macro avg) subiram, quando foram utilizadas as referidas técnicas. Relativamente ao AUC, no algoritmo DT, o seu valor foi inferior ao obtido quando utilizados os dados originais no treino do modelo.

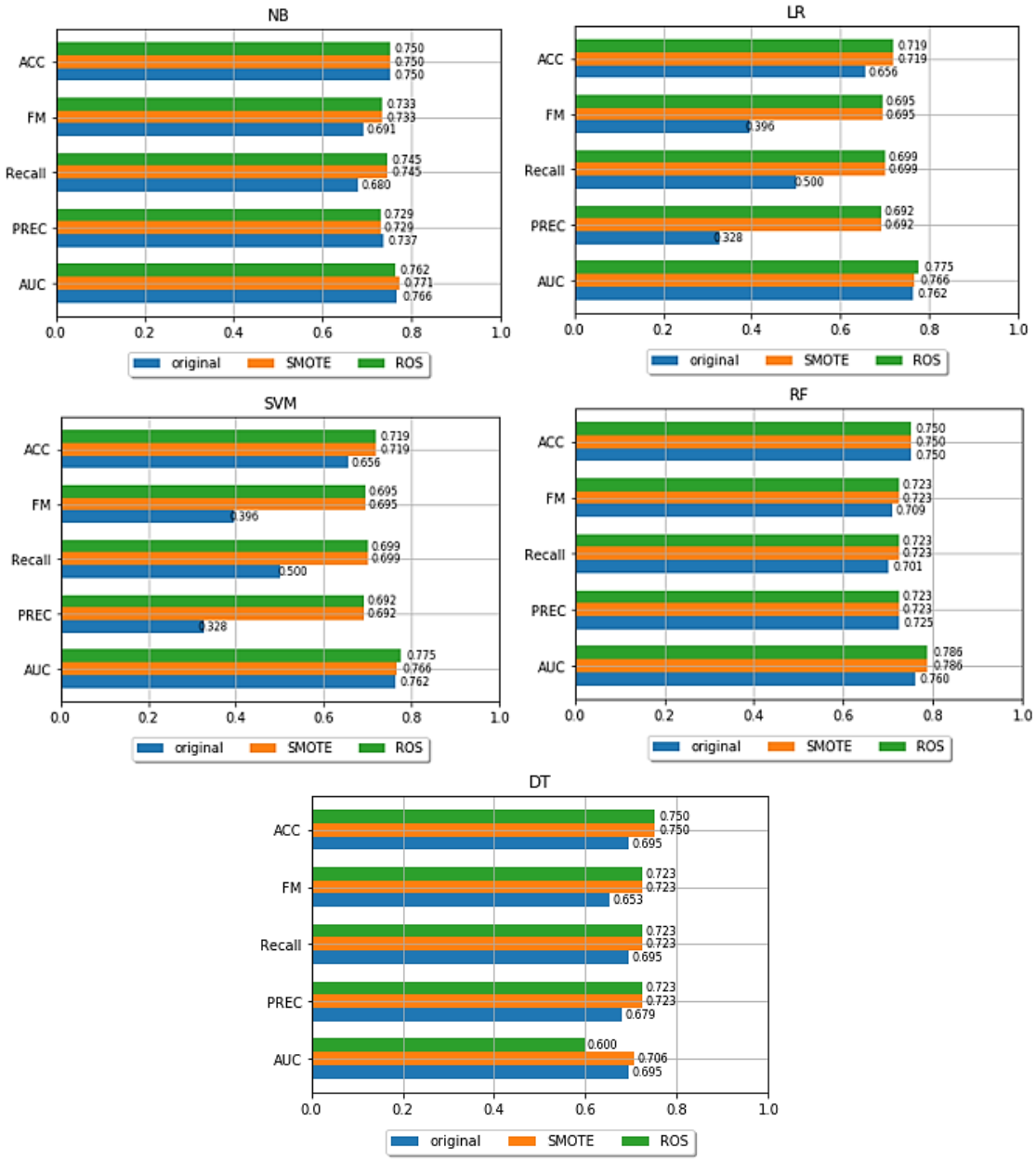


Figura 5.18 - *Performance* dos diferentes classificadores segundo as métricas AUC, PREC (macro avg), *Recall* (macro avg), FM (macro avg) e ACC.

Por fim, os resultados obtidos mostram que o modelo proposto pode ajudar os médicos a identificar o estágio ligeiro e intermédio da DP de forma objetiva, embora apenas seja válido no *dataset* analisado. Além disso, o modelo pode facilitar os médicos em processos de tomada de decisão sobre a terapêutica a adotar. O plano terapêutico pode, inclusive, ser personalizado e ajustado ao paciente.

Além das limitações apontadas na Subsecção 5.1.3 alguns erros obtidos na classificação podem ter como causa o facto de não haver um número equilibrado de pacientes entre os diferentes estágios da HY e haver um maior número de pacientes entre o estádio 2 e 2,5.

Outro fator que pode explicar a existência dos erros de classificação é a atribuição de *labels* baseada exclusivamente nos valores da escala HY. Neste trabalho foram encontradas algumas diferenças relativas à identificação da severidade para cada paciente, quando comparada a

escala HY com a escala UPDRS. Há casos de pacientes que, para uma escala, se encontram no estágio ligeiro da doença e, para a outra, encontram-se no estágio intermédio. Para resolver esta situação, além dos valores das escalas, deveria haver informação clínica mais detalhada de cada paciente, a fim de averiguar se há erros nas pontuações dadas pelos médicos.

Interessantemente o modelo para identificar a severidade da DP incluiu apenas variáveis relacionadas com a variabilidade da marcha. Estas variáveis estão relacionadas com os mecanismos que controlam o ritmo do movimento, o gerador de padrão central e o equilíbrio [188]. O aumento da variabilidade da marcha indica uma progressão da DP, ou seja, há uma diminuição mais acentuada dos núcleos da base que controlam os movimentos da marcha. Assim, outras regiões do cérebro aumentam a sua atividade para ser possível o controlo da marcha, o que diminui a automaticidade da mesma [170]. Desta forma, a variabilidade da marcha parece ser o parâmetro mais importante para identificar o estágio da DP.

5.3 - Resumo

No presente capítulo foram apresentados os resultados obtidos, após dois modelos diferentes terem sido aplicados a um *dataset* com os dados do GC e do GDP.

O primeiro modelo tinha como objetivo classificar automaticamente a DP. Os algoritmos NB e SVM (RBF) obtiveram os melhores resultados com ACC de 84,6% e 80,8%, respetivamente. Além disso, a grande maioria dos algoritmos testados beneficiaram da seleção de caraterísticas através do método *Filter* baseado na correlação seguido do método baseado na ANOVA. Neste caso, a utilização da VC 10-*fold* foi mais vantajosa do que a LOOCV. Também foi possível verificar que as quatro caraterísticas mais discriminativas do *dataset* são: comprimento do passo, velocidade do passo, largura do passo e variabilidade da largura do passo.

Para a classificação automática da severidade da DP foi desenvolvido um segundo modelo aplicado apenas ao *dataset* com os dados do GDP. Os resultados mostraram que os algoritmos respondem de maneira diferente ao desequilíbrio das classes no *dataset*. O desempenho dos algoritmos LR e SVM linear foi mais afetado por esse fator do que os restantes. Deste modo, estes algoritmos foram os que mais beneficiaram com as técnicas de sobre-amostragem. O algoritmo RF obteve o melhor desempenho com um AUC de 0,786. Entre a técnica SMOTE e a técnica ROS não se encontraram grandes diferenças em termos de *performance* dos algoritmos no conjunto de teste, com exceção do algoritmo DT. Todos os algoritmos consideraram a variabilidade da largura da passada e a variabilidade da duração do apoio duplo do passo como as caraterísticas que melhor distinguem a fase ligeira da fase intermédia da DP.

Embora os modelos propostos tenham conseguido distinguir pacientes de DP de indivíduos saudáveis, bem como identificar os vários estágios da doença, apenas são válidos na amostra analisada.

Capítulo 6

Conclusões e Trabalhos Futuros

O capítulo final do presente documento apresenta as principais conclusões do trabalho desenvolvido, bem como as perspetivas de trabalhos futuros.

6.1 – Conclusões

Embora muitos aspetos da DP tenham sido desvendados, ainda não foi encontrada uma terapia capaz de abrandar a progressão da DP ou prevenir a sua manifestação. À medida que a doença evolui, há um comprometimento da marcha que piora com o passar do tempo. Contudo, a avaliação clínica da DP baseia-se, essencialmente, em escalas subjetivas, o que dificulta o estabelecimento de um diagnóstico. A exatidão do mesmo varia entre 75% e 82%. Por outro lado, a análise quantitativa dos parâmetros da marcha pode fornecer informações úteis relacionadas com a doença. O ML, devido à sua capacidade de extrair padrões úteis nos dados processados, possibilita a deteção automática da alteração do padrão de marcha na DP, apoiando, assim, a avaliação clínica no início da doença.

A presente Dissertação tinha como objetivo desenvolver e implementar dois modelos de ML, um para o diagnóstico da DP e outro para a previsão da severidade da doença com base nos parâmetros espaciotemporais provenientes de uma base de dados constituída por um grupo de controlo que incluía 63 indivíduos saudáveis com idades comparáveis ao grupo de pacientes com DP. Todos os participantes foram recrutados na região de Bauru (São Paulo, Brasil) e encontravam-se na fase ligeira ou moderada da doença, entre o estádio 1 e 3 da escala HY.

Os parâmetros da marcha identificados como fundamentais para a deteção da DP são: comprimento do passo, velocidade do passo, largura do passo e variabilidade da largura do passo. Na Secção 3.1 do Capítulo 3 foram apresentados alguns estudos que utilizaram sistemas *wearable* e não-*wearable* para auxiliarem na deteção da DP. Dos sete autores mencionados na revisão literária, seis concluíram que o comprimento do passo é um dos parâmetros da marcha mais importante na classificação da DP. Rehman *et al.* [5] identificaram cinco parâmetros da marcha clinicamente relevantes na deteção da DP numa fase inicial, sendo que quatro coincidem com os referidos anteriormente. Além disso, estes autores concluíram que a assimetria é um parâmetro menos relevante na identificação da doença. Neste trabalho, esse facto também se verificou, uma vez que a assimetria da velocidade do passo ocupou o sexto lugar no ranking,

após a seleção de caraterísticas, nos algoritmos LR, MLP e RF. Os restantes algoritmos não selecionaram esses parâmetros.

Relativamente ao desempenho dos algoritmos de aprendizagem na classificação automática da DP, destacam-se NB e SVM (RBF), com a utilização da VC 10-*fold* em detrimento da LOOCV. Após a seleção de caraterísticas, o algoritmo NB melhorou consideravelmente a sua *performance*, ao subir a ACC de 61,5% para 84,6%.

Na segunda parte do trabalho, foram identificados dois parâmetros da marcha como fundamentais para distinguir a DP ligeira da DP intermédia: variabilidade da largura da passada e variabilidade da duração do apoio duplo do passo. Porém, de todos os estudos analisados, apenas Rehman *et al*. [5] estudaram a influência da variabilidade na DP, mas não incluíram os parâmetros da passada, nem incluíram pacientes com DP intermédia no estudo.

Em relação ao desempenho dos algoritmos de aprendizagem na classificação automática da severidade da DP, foram encontradas diferenças face à existência de um *dataset* desequilibrado. Enquanto os algoritmos LR e SVM linear "reagiram" negativamente a esse fator, os restantes não foram afetados. O algoritmo RF obteve o melhor desempenho com um AUC de 0,786. Na maioria dos algoritmos não foram encontradas diferenças significativas entre a utilização da técnica SMOTE e a técnica ROS, à exceção do algoritmo DT.

Por fim, o trabalho desenvolvido nesta Dissertação mostrou que os modelos de ML propostos podem ajudar os profissionais de saúde a diagnosticar a DP, bem como a estimar a severidade da doença com base na escala HY, apesar de apenas serem válidos na amostra analisada.

6.2 - Trabalhos Futuros

Em relação a possíveis trabalhos futuros propõe-se a criação de uma base de dados com uma dimensão maior e mais heterogénea, permitindo, assim, utilizar algoritmos de aprendizagem mais complexos, com a inclusão de abordagens de DL. Além disso, a utilização de mais dados possibilitaria a previsão de três classes (normal, DP ligeira e DP intermédia) a partir da implementação de um único modelo.

Para além disso, seria interessante combinar os parâmetros espaciotemporais com parâmetros cinemáticos e/ou cinéticos. Outros sistemas de aquisição de dados também podem ser explorados, tais como sistemas de captura de movimento com marcadores de baixo custo ou sistemas *wearable*. No caso dos sistemas *wearable*, os dados poderiam ser adquiridos em tempo real num ambiente doméstico, a fim de uma monitorização contínua dos pacientes.

Por último, para uma análise mais completa dos sintomas motores, poderiam ser incluídos exercícios que permitissem avaliar o tremor, com o intuito de melhorar a classificação da DP. Esta análise ainda poderia ser complementada, com a inclusão da análise dos sintomas não-motores.

Referências

[1] A. Draoui, O. El Hiba, A. Aimrane, A. El Khiat, and H. Gamrani, "Parkinson's disease: From bench to bedside," *Rev. Neurol. (Paris).*, vol. 176, no. 7, pp. 543–559, 2020.

[2] D. K. Simon, C. M. Tanner, and P. Brundin, "Parkinson disease epidemiology, pathology, genetics, and pathophysiology," *Clin. Geriatr. Med.*, vol. 36, no. 1, pp. 1–12, 2020.

[3] O.-B. Tysnes and A. Storstein, "Epidemiology of Parkinson's disease," *J. Neural Transm.*, vol. 124, no. 8, pp. 901–905, 2017.

[4] V. Cabreira and J. Massano, "Doença de Parkinson: Revisão Clínica e Atualização," *Acta Med. Port.*, vol. 32, no. 10, 2019.

[5] R. Z. U. Rehman, S. Del Din, Y. Guan, A. J. Yarnall, J. Q. Shi, and L. Rochester, "Selecting clinically relevant gait characteristics for classification of early parkinson's disease: A comprehensive machine learning approach," *Sci. Rep.*, vol. 9, no. 1, pp. 1–12, 2019.

[6] N. Kour, Sunanda, and S. Arora, "Computer-Vision Based Diagnosis of Parkinson's Disease via Gait: A Survey," *IEEE Access*, vol. 7, pp. 156620–156645, 2019.

[7] L. di Biase *et al.*, "Gait analysis in Parkinson's disease: an overview of the most accurate markers for diagnosis and symptoms monitoring," *Sensors*, vol. 20, no. 12, p. 3529, 2020.

[8] E. Rastegari, S. Azizian, and H. Ali, "Machine learning and similarity network approaches to support automatic classification of parkinson's diseases using accelerometer-based gait analysis," in *Proceedings of the 52nd Hawaii International Conference on System Sciences*, 2019, pp. 4231–4242.

[9] S. Li and W. Le, "Milestones of Parkinson's disease research: 200 years of history and beyond," *Neurosci. Bull.*, vol. 33, no. 5, pp. 598–602, 2017.

[10] R. Pahwa and K. E. L. P. D, *Handbook of Parkinson's Disease*, 5th ed. Boca Raton, FL: Taylor & Francis, 2013.

[11] S. A. Factor and William J. Weiner, *Parkinson's Disease: Diagnosis and Clinical Management*, 2nd ed. New York: Demos Medical Publishing, 2007.

[12] R. R. Seeley, T. D. Stephens, and P. Tate, *Anatomy and Physiology*, 6th ed. New York: McGraw-Hill Higher Education, 2003.

[13] K. E. Barrett, S. M. Barman, S. Boitano, and H. L. Brooks, *Ganong's Review of Medical Physiology (Enhanced EB)*, 23rd ed. New York: McGraw-Hill Education, 2009.

[14] S. S. Mader and P. Galliart, *Understanding Human Anatomy and Physiology*, 5th ed. New York: McGraw-Hill Higher Education, 2005.

[15] J. G. Betts *et al.*, "Anatomy & physiology," 2013. [Online]. Available: https://openstax.org/details/anatomy-and-physiology. [Accessed: 16-Nov-2020].

[16] M. M. McGregor and A. B. Nelson, "Circuit Mechanisms of Parkinson's Disease," *Neuron*, vol. 101, no. 6, pp. 1042–1056, 2019.

[17] R. Balestrino and A. H. V Schapira, "Parkinson disease," *Eur. J. Neurol.*, vol. 27, no. 1, pp. 27–42, 2020.

[18] E. M. Rocha, B. De Miranda, and L. H. Sanders, "Alpha-synuclein: Pathology, mitochondrial dysfunction and neuroinflammation in Parkinson's disease," *Neurobiol. Dis.*, vol. 109, no. Pt B, pp. 249–257, 2018.

[19] H. H. P. Nguyen and M. A. Cenci, *Behavioral Neurobiology of Huntington's Disease and Parkinson's Disease*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2015.

[20] S. Caproni and C. Colosimo, "Diagnosis and differential diagnosis of Parkinson disease," *Clin. Geriatr. Med.*, vol. 36, no. 1, pp. 13–24, 2020.

[21] J. C. Greenland and R. A. Barker, "The differential diagnosis of Parkinson's disease," in *Parkinson's Disease: Pathogenesis and Clinical Aspects*, 1st ed., Brisbane, Australia: Codon Publications, 2018, pp. 109–128.

[22] K. Bhatia, K. R. Chaudhuri, and M. Stamelou, *Parkinson's Disease*, 1st ed. Cambridge, San Diego, Oxford, London: Academic Press, 2017.
[23] A. D. Mosley, D. S. Romaine, and A. Samii, *The Encyclopedia of Parkinson's Disease*, 2nd ed. New York: Facts On File, 2009.
[24] A. Tarakad and J. Jankovic, "Diagnosis and Management of Parkinson's Disease," *Semin Neurol*, vol. 37, no. 2, pp. 118–126, 2017.
[25] J. Jankovic, M. Hallett, and S. Fahn, *Principles and Practice of Movement Disorders*, 2nd ed. Edinburgh: Elsevier Health Sciences, 2011.
[26] N. PADMAN, R. SWARNALATHA, V. VENKATESH, and N. KUMAR, "TELEDIAGNOSIS OF PARKINSON'S DISEASE SYMPTOM SEVERITY USING H&Y SCALE," *J. Eng. Sci. Technol.*, vol. 15, no. 3, pp. 1466–1480, 2020.
[27] R. F. Pfeiffer, "Non-motor symptoms in Parkinson's disease," *Parkinsonism Relat. Disord.*, vol. 22 Suppl 1, pp. S119–S122, 2016.
[28] J. Marinus, K. Zhu, C. Marras, D. Aarsland, and J. J. van Hilten, "Risk factors for non-motor symptoms in Parkinson's disease," *Lancet Neurol.*, vol. 17, no. 6, pp. 559–568, 2018.
[29] A. H. V Schapira, K. R. Chaudhuri, and P. Jenner, "Non-motor features of Parkinson disease," *Nat. Rev. Neurosci.*, vol. 18, no. 7, p. 435, 2017.
[30] A. W. Amara and A. A. Memon, "Effects of Exercise on Non-motor Symptoms in Parkinson's Disease," *Clin. Ther.*, vol. 40, no. 1, pp. 8–15, 2018.
[31] J. Jankovic and E. Tolosa, *Parkinson's Disease and Movement Disorders*, 6th ed. Philadelphia, PA: Wolters Kluwer, 2015.
[32] M. J. Armstrong and M. S. Okun, "Diagnosis and Treatment of Parkinson Disease: A Review," *JAMA*, vol. 323, no. 6, pp. 548–560, Feb. 2020.
[33] S. K. Meles, L. K. Teune, B. M. de Jong, R. A. Dierckx, and K. L. Leenders, "Metabolic imaging in Parkinson disease," *J. Nucl. Med.*, vol. 58, no. 1, pp. 23–28, 2017.
[34] I. Litvan, *Atypical Parkinsonian Disorders: Clinical and Research Aspects*, 1st ed. Totowa, New Jersey: Humana Press, 2007.
[35] R. Bhidayasiri and P. Martinez-Martin, "Clinical Assessments in Parkinson's Disease: Scales and Monitoring," *Int Rev Neurobiol*, vol. 132, pp. 129–182, 2017.
[36] P. Martinez-Martin, C. Rodriguez-Blazquez, M. J. Forjaz, and K. R. Chaudhuri, *Guide to Assessment Scales in Parkinson's Disease*. London: Springer Healthcare Limited, 2014.
[37] C. G. Goetz *et al.*, "MDS UPDRS: The MDS-sponsored Revision of the Unified Parkinson's Disease Rating Scale," *International Parkinson and Movement Disorder Society*, vol. 1, no. 414. Milwaukee, WI, pp. 1–33, 2008.
[38] S. Pires *et al.*, "Pistas auditivas musicais na fisioterapia em grupo de doentes com Parkinson," *Arq. Med.*, vol. 28, no. 6, pp. 162–166, 2014.
[39] C. G. Goetz *et al.*, "Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): process, format, and clinimetric testing plan," *Mov. Disord.*, vol. 22, no. 1, pp. 41–47, 2007.
[40] J. G. Hentz, S. H. Mehta, H. A. Shill, E. Driver-Dunckley, T. G. Beach, and C. H. Adler, "Simplified conversion method for unified Parkinson's disease rating scale motor examinations," *Mov. Disord.*, vol. 30, no. 14, pp. 1967–1970, 2015.
[41] P. Martínez-Martín *et al.*, "Parkinson's disease severity levels and MDS-Unified Parkinson's Disease Rating Scale," *Parkinsonism Relat. Disord.*, vol. 21, no. 1, pp. 50–54, 2015.
[42] C. Sampaio, C. G. Goetz, and A. Schrag, *Rating Scales in Parkinson's Disease: Clinical Practice and Research*, 1st ed. Oxford: OUP USA, 2012.
[43] E. J. Modestino, A. Reinhofer, K. Blum, C. Amenechi, and P. O'Toole, "Hoehn and Yahr staging of Parkinson's disease in relation to neuropsychological measures," *Front. Biosci. (Landmark Ed.*, vol. 23, pp. 1370–1379, 2018.
[44] M. Schenkman *et al.*, "Profile of Functional Limitations and Task Performance Among People With Early- and Middle-Stage Parkinson Disease," *Phys. Ther.*, vol. 91, no. 9, pp. 1339–1354, Sep. 2011.
[45] I. Santana *et al.*, "Mini-Mental State Examination: Screening and diagnosis of cognitive decline, using new normative data," *Acta Med. Port.*, vol. 29, no. 4, pp. 240–248, 2016.

[46] J. F. Scheffels, L. Fröhlich, E. Kalbe, and J. Kessler, “Concordance of Mini-Mental State Examination, Montreal Cognitive Assessment and Parkinson Neuropsychometric Dementia Assessment in the classification of cognitive performance in Parkinson’s disease,” *J. Neurol. Sci.*, vol. 412, p. 116735, 2020.
[47] C. R. Smith, J. Cavanagh, M. Sheridan, K. A. Grosset, B. Cullen, and D. G. Grosset, “Factor structure of the Montreal Cognitive Assessment in Parkinson disease,” *Int. J. Geriatr. Psychiatry*, vol. 35, no. 2, pp. 188–194, Feb. 2020.
[48] W. Oertel and J. B. Schulz, “Current and experimental treatments of Parkinson disease: a guide for neuroscientists,” *J. Neurochem.*, vol. 139 Suppl, pp. 325–337, 2016.
[49] S. Sveinbjornsdottir, “The clinical symptoms of Parkinson’s disease,” *J. Neurochem.*, vol. 139 Suppl, pp. 318–324, 2016.
[50] “Freezing,” *Parkinson’s UK*, 2018. [Online]. Available: https://www.parkinsons.org.uk/information-and-support/freezing. [Accessed: 24-Nov-2020].
[51] F. N. Emamzadeh and A. Surguchov, “Parkinson’s Disease: Biomarkers, Treatment, and Risk Factors,” *Front. Neurosci.*, vol. 12, p. 612, 2018.
[52] J. Palma and H. Kaufmann, “Treatment of autonomic dysfunction in Parkinson disease and other synucleinopathies,” *Mov. Disord.*, vol. 33, no. 3, pp. 372–390, 2018.
[53] H. Homayoun, “Parkinson’s Disease,” *In the Clinic*. Annals of Internal Medicine, pp. 1–19, 2018.
[54] A. S. Alharthi, S. U. Yunas, and K. B. Ozanyan, “Deep Learning for Monitoring of Human Gait: A Review,” *IEEE Sens. J.*, vol. 19, no. 21, pp. 9575–9591, 2019.
[55] C. L. Vaughan, B. L. Davis, and J. C. O’Connor, *Dynamics of human gait*, 2nd ed. Cape Town: Human Kinetics Publishers, 1992.
[56] R. S. Khandpur, *Compendium of Biomedical Instrumentation*, 1st ed. Chennai: Wiley, 2020.
[57] W. Pirker and R. Katzenschlager, “Gait disorders in adults and the elderly: A clinical guide,” *Wien. Klin. Wochenschr.*, vol. 129, no. 3–4, pp. 81–95, Oct. 2016.
[58] K. Sugandhi, F. F. Wahid, and G. Raju, “Detection of human gait cycle: An overlap based approach,” in *2017 International Conference on Infocom Technologies and Unmanned Systems (Trends and Future Directions)(ICTUS)*, 2017, pp. 1–3.
[59] J. E. Muscolino, *Kinesiology: The Skeletal System and Muscle Function*, 3rd ed. St. Louis, Missouri: Elsevier, 2017.
[60] C. Prakash, R. Kumar, and N. Mittal, “Recent developments in human gait research: parameters, approaches, applications, machine learning techniques, datasets and challenges,” *Artif. Intell. Rev.*, vol. 49, no. 1, pp. 1–40, 2018.
[61] B. Müller, S. Wolf, G. P. Brueggemann, Z. Deng, F. Miller, and W. S. Selbie, *Handbook of Human Motion*, 1st ed. Cham: Springer International Publishing, 2018.
[62] J. H. Hollman, M. K. Watkins, A. C. Imhoff, C. E. Braun, K. A. Akervik, and D. K. Ness, “A comparison of variability in spatiotemporal gait parameters between treadmill and overground walking conditions,” *Gait Posture*, vol. 43, pp. 204–209, 2016.
[63] A. J. J. Smith and E. D. Lemaire, “Temporal-spatial gait parameter models of very slow walking,” *Gait Posture*, vol. 61, pp. 125–129, 2018.
[64] H. Nagano, R. K. Begg, W. A. Sparrow, and S. Taylor, “A comparison of treadmill and overground walking effects on step cycle asymmetry in young and older individuals,” *J. Appl. Biomech.*, vol. 29, no. 2, pp. 188–193, 2013.
[65] D. T. H. Lai, R. K. Begg, and M. Palaniswami, “Computational Intelligence in Gait Research: A Perspective on Current Applications and Future Challenges,” *IEEE Trans. Inf. Technol. Biomed.*, vol. 13, no. 5, pp. 687–702, 2009.
[66] K. Guffey, M. Regier, C. Mancinelli, and P. Pergami, “Gait parameters associated with balance in healthy 2- to 4-year-old children,” *Gait Posture*, vol. 43, pp. 165–169, 2016.
[67] D. Meldrum, C. Shouldice, R. Conroy, K. Jones, and M. Forward, “Test–retest reliability of three dimensional gait analysis: Including a novel approach to visualising agreement of gait cycle waveforms with Bland and Altman plots,” *Gait Posture*, vol. 39, no. 1, pp. 265–271, 2014.
[68] O. Beauchet *et al.*, “Guidelines for assessment of gait and reference values for spatiotemporal gait parameters in older adults: the biomathics and Canadian gait

consortiums initiative," *Front. Hum. Neurosci.*, vol. 11, p. 353, 2017.
[69] C. J. Shanahan *et al.*, "Technologies for Advanced Gait and Balance Assessments in People with Multiple Sclerosis," *Front. Neurol.*, vol. 8, pp. 1–10, 2018.
[70] A. Muro-De-La-Herran, B. Garcia-Zapirain, and A. Mendez-Zorrilla, "Gait analysis methods: An overview of wearable and non-wearable systems, highlighting clinical applications," *Sensors*, vol. 14, no. 2, pp. 3362–3394, 2014.
[71] "Choosing a force plate," *Advanced Mechanical Technology, Inc.* [Online]. Available: https://www.amti.biz/fps-guide.aspx. [Accessed: 26-Nov-2020].
[72] D. TarniŢă, "Wearable sensors used for human gait analysis," *Rom J Morphol Embryol*, vol. 57, no. 2, pp. 373–382, 2016.
[73] W. Tao, T. Liu, R. Zheng, and H. Feng, "Gait Analysis Using Wearable Sensors," *Sensors (Basel).*, vol. 12, pp. 2255–2283, Dec. 2012.
[74] Y.-L. Hsu *et al.*, "Gait and Balance Analysis for Patients With Alzheimer's Disease Using an Inertial-Sensor-Based Wearable Instrument," *IEEE J. Biomed. Heal. informatics*, vol. 18, pp. 1822–1830, Nov. 2014.
[75] M. Gago *et al.*, *Manual para Pessoas com Parkinson*, 1st ed. Lisboa: EP Health Marketing, SL, 2014.
[76] P. Degwekar, "Parkinson's Disease," *PhysioRehab*, 2017. [Online]. Available: http://physiorehab.in/parkinsons-disease/. [Accessed: 19-Nov-2020].
[77] A. P. J. Zanardi *et al.*, "Gait parameters of Parkinson's disease compared with healthy controls: a systematic review and meta-analysis," *Sci. Rep.*, vol. 11, no. 1, p. 752, 2021.
[78] A. Mirelman *et al.*, "Gait impairments in Parkinson's disease," *Lancet Neurol.*, vol. 18, no. 7, pp. 697–708, 2019.
[79] G. Ebersbach, C. Moreau, F. Gandor, L. Defebvre, and D. Devos, "Clinical syndromes: Parkinsonian gait," *Mov. Disord.*, vol. 28, no. 11, pp. 1552–1559, 2013.
[80] W. J. Murdoch, C. Singh, K. Kumbier, R. Abbasi-Asl, and B. Yu, "Interpretable machine learning: definitions, methods, and applications," *arXiv Prepr. arXiv1901.04592*, pp. 1–11, 2019.
[81] G. Rebala, A. Ravi, and S. Churiwala, *An Introduction to Machine Learning*, 1st ed. Cham: Springer International Publishing, 2019.
[82] D. Jakhar and I. Kaur, "Artificial intelligence, machine learning and deep learning: definitions and differences," *Clin. Exp. Dermatol.*, vol. 45, no. 1, pp. 131–132, 2020.
[83] W. J. Park and J.-B. Park, "History and application of artificial neural networks in dentistry," *Eur. J. Dent.*, vol. 12, no. 4, p. 594, 2018.
[84] M. Haenlein and A. Kaplan, "A brief history of artificial intelligence: On the past, present, and future of artificial intelligence," *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 5–14, 2019.
[85] V. Kaul, S. Enslin, and S. A. Gross, "History of artificial intelligence in medicine," *Gastrointest. Endosc.*, vol. 92, no. 4, pp. 807–812, 2020.
[86] S. A. Bini, "Artificial Intelligence, Machine Learning, Deep Learning, and Cognitive Computing: What Do These Terms Mean and How Will They Impact Health Care?," *J. Arthroplasty*, vol. 33, no. 8, pp. 2358–2361, 2018.
[87] A. V Joshi, *Machine Learning and Artificial Intelligence*, 1st ed. Redmond, WA: Springer International Publishing, 2019.
[88] I. Kononenko, "Machine learning for medical diagnosis: history, state of the art and perspective," *Artif. Intell. Med.*, vol. 23, no. 1, pp. 89–109, 2001.
[89] B. Macukow, "Neural Networks – State of Art, Brief History, Basic Models and Architecture BT - Computer Information Systems and Industrial Management," 2016, pp. 3–14.
[90] M. Furmankiewicz, A. Sołtysik-Piorunkiewicz, and P. Ziuziański, "Artificial intelligence systems for knowledge management in e-health: the study of intelligent software agents," in *Latest Trends on Systems: The Proceedings of 18th International Conference on Systems, Santorini Island, Greece*, 2014, pp. 551–556.
[91] V. François-Lavet, P. Henderson, R. Islam, M. G. Bellemare, and J. Pineau, *An introduction to deep reinforcement learning*, vol. 11, no. 3–4. Boston, Delft: now, 2018.
[92] O. Simeone, "A brief introduction to machine learning for engineers," *Foundations and*

Trends R in Signal Processing, vol. 12, no. 3–4. now Publishers Inc., Boston, Delft, pp. 200–431, 2017.

[93] C. M. Bishop, “Neural networks for pattern recognition,” *J. Am. Stat. Assoc.*, vol. 92, p. 482, 1995.

[94] D. Ceilen, D. B. A. Meysman, and M. Ali, *Introducing Data Science*, 1st ed., vol. 1. SHELTER ISLAND, NY: Manning Publications, 2016.

[95] O. Simeone, “A very brief introduction to machine learning with applications to communication systems,” *IEEE Trans. Cogn. Commun. Netw.*, vol. 4, no. 4, pp. 648–664, 2018.

[96] R. Saravanan and P. Sujatha, “A state of art techniques on machine learning algorithms: a perspective of supervised learning approaches in data classification,” in *2018 Second International Conference on Intelligent Computing and Control Systems (ICICCS)*, 2018, pp. 945–949.

[97] S. Mirjalili, H. Faris, and I. Aljarah, *Evolutionary Machine Learning Techniques: Algorithms and Applications*, 1st ed. Singapore: Springer Singapore, 2019.

[98] J. Shin, T. A. Badgwell, K.-H. Liu, and J. H. Lee, “Reinforcement Learning – Overview of recent progress and implications for process control,” *Comput. Chem. Eng.*, vol. 127, pp. 282–294, 2019.

[99] C. Jin, Z. Yang, Z. Wang, and M. I. Jordan, “Provably efficient reinforcement learning with linear function approximation,” in *33rd Annual Conference on Learning Theory*, 2020, pp. 2137–2143.

[100] S. De Vito, G. Fattoruso, M. Pardo, F. Tortorella, and G. Di Francia, “Semi-supervised learning techniques in artificial olfaction: A novel approach to classification problems and drift counteraction,” *IEEE Sens. J.*, vol. 12, no. 11, pp. 3215–3224, 2012.

[101] A. Mehta, “An Ultimate Guide to Understanding Semi-Supervised Learning,” 2019. [Online]. Available: https://www.digitalvidya.com/blog/semi-supervised-learning/. [Accessed: 21-Oct-2020].

[102] J. E. Mason, I. Traoré, and I. Woungang, *Machine Learning Techniques for Gait Biometric Recognition: Using the Ground Reaction Force*, 1st ed. Cham: Springer International Publishing, 2016.

[103] R. Garreta and G. Moncecchi, *Learning scikit-learn: Machine Learning in Python*, 1st ed. Birmingham, UK.: Packt Publishing, 2013.

[104] R. Longadge and S. Dongre, “Class imbalance problem in data mining review,” *Int. J. Comput. Sci. Netw.*, vol. 2, no. 1, 2013.

[105] I. G. Maglogiannis, *Emerging Artificial Intelligence Applications in Computer Engineering: Real Word AI Systems with Applications in EHealth, HCI, Information Retrieval and Pervasive Technologies*. Amsterdam: IOS Press, 2007.

[106] A. Dumortier, E. Beckjord, S. Shiffman, and E. Sejdic, “Classifying smoking urges via machine learning,” *Comput. Methods Programs Biomed.*, vol. 137, pp. 203–213, Sep. 2016.

[107] V. Kotu and B. Deshpande, *Predictive Analytics and Data Mining: Concepts and Practice with RapidMiner*, 1st ed. Waltham, MA: Elsevier Science, 2014.

[108] S. Kaushik, “Introduction to Feature Selection methods with an example (or how to select the right variables?),” 2016. [Online]. Available: https://www.analyticsvidhya.com/blog/2016/12/introduction-to-feature-selection-methods-with-an-example-or-how-to-select-the-right-variables/. [Accessed: 26-Oct-2020].

[109] J. Suto, S. Oniga, and P. P. Sitar, “Comparison of wrapper and filter feature selection algorithms on human activity recognition,” in *2016 6th International Conference on Computers Communications and Control (ICCCC)*, 2016, pp. 124–129.

[110] S. Wang *et al.*, “PATHOLOGICAL BRAIN DETECTION BY ARTIFICIAL INTELLIGENCE IN MAGNETIC RESONANCE IMAGING SCANNING (INVITED REVIEW),” *Prog. Electromagn. Res.*, vol. 156, pp. 105–133, Jan. 2016.

[111] A. Suppers, A. van Gool, and H. Wessels, “Integrated Chemometrics and Statistics to Drive Successful Proteomics Biomarker Discovery,” *Proteomes*, vol. 6, pp. 1–17, Apr. 2018.

[112] M. Mohri, A. Rostamizadeh, and A. Talwalkar, *Foundations of Machine Learning*, 1st ed. Cambridge: MIT Press, 2012.

[113] Z. S. Hippe, J. L. Kulikowski, and T. Mroczek, *Human – Computer Systems Interaction: Backgrounds and Applications 2: Part 2*, 1st ed. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011.
[114] H. I. Rhys, *Machine Learning with R, the tidyverse, and mlr*, 1st ed. Shelter Island, NY: Manning Publications, 2020.
[115] S. Shalev-Shwartz and S. Ben-David, *Understanding Machine Learning: From Theory to Algorithms*, Draft ed. New York: Cambridge University Press, 2014.
[116] M. Achirul Nanda, K. Boro Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," *Information*, vol. 9, no. 1, pp. 1–14, 2018.
[117] M. M. Adankon and M. Cheriet, "Support Vector Machine," *Encyclopedia of biometrics*. Springer, New York, pp. 1303–1308, 2009.
[118] M. Gopal, *Applied Machine Learning*, 1st ed. Chennai: McGraw-Hill Education, 2018.
[119] R. F. Turkson, F. Yan, M. K. A. Ali, and J. Hu, "Artificial neural network applications in the calibration of spark-ignition engines: An overview," *Eng. Sci. Technol. an Int. J.*, vol. 19, no. 3, pp. 1346–1359, 2016.
[120] D. E. Babatunde, A. N. Anozie, and J. Omoleye, "Artificial Neural Network and its Applications in the Energy Sector–An Overview," *Int. J. Energy Econ. Policy*, vol. 10, no. 2, pp. 250–264, 2020.
[121] K. L. Priddy and P. E. Keller, *Artificial Neural Networks: An Introduction*, 1st ed. Bellingham: SPIE Press, 2005.
[122] I. N. da Silva, D. Hernane Spatti, R. Andrade Flauzino, L. H. B. Liboni, and S. F. dos Reis Alves, "Artificial Neural Networks," pp. 21–29, 2017.
[123] S. Haykin, *Neural networks-A comprehensive foundation*, 3rd ed. New Jersey: Prentice Hall, 1994.
[124] R. Bali, D. Sarkar, B. Lantz, and C. Lesmeister, *R: Unleash Machine Learning Techniques*, 1st ed. Birmingham: Packt Publishing, 2016.
[125] S. Rogers and M. Girolami, *A First Course in Machine Learning*, 1st ed. Boca Raton, FL: CRC Press, 2016.
[126] V. B. Prasath *et al.*, "Distance and Similarity Measures Effect on the Performance of K-Nearest Neighbor Classifier--A Review," *arXiv Prepr. arXiv1708.04321*, pp. 1–39, 2017.
[127] I. Saini, D. Singh, and A. Khosla, "QRS detection using K-Nearest Neighbor algorithm (KNN) and evaluation on standard ECG databases," *J. Adv. Res.*, vol. 4, no. 4, pp. 331–344, 2013.
[128] A. Bronshtein, "A Quick Introduction to K-Nearest Neighbors Algorithm," *Noteworthy - The Journal Blog*, 2017. [Online]. Available: https://blog.usejournal.com/a-quick-introduction-to-k-nearest-neighbors-algorithm-62214cea29c7. [Accessed: 27-Oct-2020].
[129] M. Granik and V. Mesyura, "Fake news detection using naive Bayes classifier," in *2017 IEEE First Ukraine Conference on Electrical and Computer Engineering (UKRCON)*, 2017, pp. 900–903.
[130] N. F. Rusland, N. Wahid, S. Kasim, and H. Hafit, "Analysis of Naïve Bayes algorithm for email spam filtering across multiple datasets," in *Proceedings of the IOP Conference Series: Materials Science and Engineering*, 2017, pp. 1–9.
[131] E. Bisong, *Building Machine Learning and Deep Learning Models on Google Cloud Platform: A Comprehensive Guide for Beginners*, 1st ed. Ottawa, ON: Apress, 2019.
[132] G. Bonaccorso, *Machine Learning Algorithms*, 1st ed. Birmingham: Packt Publishing, 2017.
[133] E. Alpaydin, *Introduction to Machine Learning*, 1st ed. Massachusetts: MIT Press, 2004.
[134] J. Alzubi, A. Nayyar, and A. Kumar, "Machine learning from theory to algorithms: an overview," in *Journal of physics: conference series*, 2018, vol. 1142, no. 1, pp. 1–15.
[135] N. Pentreath, *Machine Learning with Spark*, 1st ed. BIRMINGHAM, MUMBAI: Packt Publishing, 2015.
[136] J. Hoare, "What is a Decision Tree?" [Online]. Available: https://www.displayr.com/what-is-a-decision-tree/. [Accessed: 28-Oct-2020].
[137] F. Noroozi, T. Sapiński, D. Kamińska, and G. Anbarjafari, "Vocal-based emotion recognition using random forests and decision tree," *Int. J. Speech Technol.*, vol. 20, no. 2, pp. 239–246, 2017.

[138] A. C. Müller and S. Guido, *Introduction to Machine Learning with Python: A Guide for Data Scientists*, 1st ed. Sebastopol, CA: O'Reilly Media, 2016.
[139] T. Amr, *Hands-On Machine Learning with scikit-learn and Scientific Python Toolkits: A practical guide to implementing supervised and unsupervised machine learning algorithms in Python*, 1st ed. Birmingham, UK: Packt Publishing, 2020.
[140] W.-H. Weng, "Machine Learning for Clinical Predictive Analytics," in *Leveraging Data Science for Global Health*, 1st ed., L. A. Celi, M. S. Majumder, P. Ordóñez, J. S. Osorio, K. E. Paik, and S. Melek, Eds. Cham: Springer, 2020, pp. 199–217.
[141] A. Bhande, "What is underfitting and overfitting in machine learning and how to deal with it," 2018. [Online]. Available: https://medium.com/greyatom/what-is-underfitting-and-overfitting-in-machine-learning-and-how-to-deal-with-it-6803a989c76. [Accessed: 23-Oct-2020].
[142] L. Yang and A. Shami, "On hyperparameter optimization of machine learning algorithms: Theory and practice," *Neurocomputing*, vol. 415, pp. 295–316, 2020.
[143] A. Pant, "Workflow of a Machine Learning project," 2019. [Online]. Available: https://towardsdatascience.com/workflow-of-a-machine-learning-project-ec1dba419b94. [Accessed: 21-Oct-2020].
[144] M. Hossin and M. N. Sulaiman, "A review on evaluation metrics for data classification evaluations," *Int. J. Data Min. Knowl. Manag. Process*, vol. 5, no. 2, pp. 1–11, 2015.
[145] J. Thanaki, *Machine Learning Solutions: Expert techniques to tackle complex machine learning problems using Python*, 1st ed. Birmingham: Packt Publishing, 2018.
[146] H. Bui, "ROC Curve Transforms the Way We Look at a Classification Problem," 2020. [Online]. Available: https://towardsdatascience.com/a-simple-explanation-of-the-roc-curve-and-auc-64db32d75541. [Accessed: 23-Oct-2020].
[147] J. Goyal, P. Khandnor, and T. C. Aseri, "Classification, Prediction, and Monitoring of Parkinson's disease using Computer Assisted Technologies: A Comparative Analysis," *Eng. Appl. Artif. Intell.*, vol. 96, pp. 1–29, 2020.
[148] C. Caramia *et al.*, "IMU-Based Classification of Parkinson's Disease From Gait: A Sensitivity Analysis on Sensor Location and Feature Selection," *IEEE J. Biomed. Heal. Informatics*, vol. 22, no. 6, pp. 1765–1774, 2018.
[149] K. A. Hasan and M. A. M. Hasan, "Classification of Parkinson's Disease by Analyzing Multiple Vocal Features Sets," in *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 758–761.
[150] E. Avuçlu and A. Elen, "Evaluation of train and test performance of machine learning algorithms and Parkinson diagnosis with statistical measurements," *Med. Biol. Eng. Comput.*, vol. 58, no. 11, pp. 2775–2788, 2020.
[151] D. Buongiorno, I. Bortone, G. D. Cascarano, G. F. Trotta, A. Brunetti, and V. Bevilacqua, "A low-cost vision system based on the analysis of motor features for recognition and severity rating of Parkinson's Disease," *BMC Med. Inform. Decis. Mak.*, vol. 19, no. 9, pp. 1–13, 2019.
[152] F. Wahid, R. K. Begg, C. J. Hass, S. Halgamuge, and D. C. Ackland, "Classification of Parkinson's disease gait using spatial-temporal gait features," *IEEE J. Biomed. Heal. informatics*, vol. 19, no. 6, pp. 1794–1802, 2015.
[153] C. Urcuqui *et al.*, "Exploring Machine Learning to Analyze Parkinson's Disease Patients," in *2018 14th International Conference on Semantics, Knowledge and Grids (SKG)*, 2018, pp. 160–166.
[154] S. Aich *et al.*, "A Supervised Machine Learning Approach to Detect the On/Off State in Parkinson's Disease Using Wearable Based Gait Signals," *Diagnostics (Basel, Switzerland)*, vol. 10, no. 6, pp. 1–18, 2020.
[155] Y. Mittra and V. Rustagi, "Classification of Subjects with Parkinson's Disease Using Gait Data Analysis," in *2018 International Conference on Automation and Computational Engineering (ICACE)*, 2018, pp. 84–89.
[156] B. E., B. D., and B. R., "Supervised machine learning based gait classification system for early detection and stage classification of Parkinson's disease," *Appl. Soft Comput.*, vol. 94, p. 106494, 2020.
[157] S. Aich *et al.*, "Design of a Machine Learning-Assisted Wearable Accelerometer-Based Automated System for Studying the Effect of Dopaminergic Medicine on Gait Characteristics of Parkinson's Patients," *J. Healthc. Eng.*, vol. 2020, pp. 1–11, 2020.

[158] C. R. Pereira *et al.*, “A step towards the automated diagnosis of parkinson’s disease: Analyzing handwriting movements,” in *2015 IEEE 28th international symposium on computer-based medical systems*, 2015, pp. 171–176.
[159] N. Goel, A. Khanna, D. Gupta, and N. Gupta, “Detection of Parkinson’s Disease Using Machine Learning Techniques for Voice and Handwriting Features,” in *International Conference on Innovative Computing and Communications. Advances in Intelligent Systems and Computing*, 2020, pp. 631–643.
[160] R. Yuvaraj *et al.*, “Detection of emotions in Parkinson’s disease using higher order spectral features from brain’s electrical activity,” *Biomed. Signal Process. Control*, vol. 14, pp. 108–116, 2014.
[161] A. H. Neehal, M. N. Azam, M. S. Islam, M. I. Hossain, and M. Z. Parvez, “Prediction of Parkinson’s Disease by Analyzing fMRI Data and using Supervised Learning,” in *2020 IEEE Region 10 Symposium (TENSYMP)*, 2020, pp. 362–365.
[162] K. N. R. Challa, V. S. Pagolu, G. Panda, and B. Majhi, “An improved approach for prediction of Parkinson’s disease using machine learning techniques,” in *2016 International Conference on Signal Processing, Communication, Power and Embedded System (SCOPES)*, 2016, pp. 1446–1451.
[163] R. Prashanth, S. D. Roy, P. K. Mandal, and S. Ghosh, “Parkinson’s disease detection using olfactory loss and REM sleep disorder features,” in *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2014, pp. 5764–5767.
[164] G. Solana-Lavalle and R. Rosas-Romero, “Classification of PPMI MRI scans with voxel-based morphometry and machine learning to assist in the diagnosis of Parkinson’s disease,” *Comput. Methods Programs Biomed.*, vol. 198, p. 105793, 2021.
[165] “Mini Exame do Estado Mental (MEEM).” [Online]. Available: http://www.eerp.usp.br. [Accessed: 08-Feb-2021].
[166] A. J. Hughes, S. E. Daniel, L. Kilford, and A. J. Lees, “Accuracy of clinical diagnosis of idiopathic Parkinson’s disease: a clinico-pathological study of 100 cases.,” *J. Neurol. Neurosurg. & Psychiatry*, vol. 55, no. 3, pp. 181 LP – 184, Mar. 1992.
[167] O. P. Almeida, “Mini exame dos estado mental e o diagnóstico de demência no Brasil,” *Arq. Neuropsiquiatr.*, vol. 56, no. 3B, pp. 605–612, 1998.
[168] S. Brucki, R. Nitrini, P. Caramelli, P. H. F. Bertolucci, and I. H. Okamoto, “Sugestões para o uso do mini-exame do estado mental no Brasil,” *Arq. Neuropsiquiatr.*, vol. 61, no. 3B, pp. 777–781, 2003.
[169] “The standard vicon full-body model (plug-in gait) marker placement scheme.” [Online]. Available: http://www.idmil.org/mocap/Plug-in-Gait+Marker+Placement.pdf. [Accessed: 26-Nov-2020].
[170] L. Simieli *et al.*, “The variability of the steps preceding obstacle avoidance (approach phase) is dependent on the height of the obstacle in people with Parkinson’s disease,” *PLoS One*, vol. 12, no. 9, p. e0184134, Sep. 2017.
[171] D. Orcioli-Silva *et al.*, “Double obstacles increase gait asymmetry during obstacle crossing in people with parkinson’s disease and healthy older adults: A pilot study,” *Sci. Rep.*, vol. 10, no. 1, pp. 1–8, 2020.
[172] C. H. Brase and C. P. Brase, *Understanding Basic Statistics*, 8th ed. Boston, MA: Cengage Learning, 2016.
[173] D. George and P. Mallery, *IBM SPSS Statistics 26 Step by Step: A Simple Guide and Reference*, 16th ed. New York: Taylor & Francis, 2019.
[174] “Spyder - The Scientific Python Development Environment.” [Online]. Available: https://www.spyder-ide.org/. [Accessed: 11-Feb-2021].
[175] “scikit-learn 0.24.1,” *scikit-learn developers (BSD License)*. [Online]. Available: https://scikit-learn.org/stable/getting_started.html. [Accessed: 26-Apr-2021].
[176] “pandas documentation,” *the pandas development team*. [Online]. Available: https://pandas.pydata.org/docs/index.html. [Accessed: 26-Apr-2021].
[177] “What is NumPy?,” *The SciPy community*. [Online]. Available: https://numpy.org/doc/stable/user/whatisnumpy.html. [Accessed: 26-Apr-2021].
[178] “Pyplot tutorial,” *The Matplotlib development team*. [Online]. Available: https://matplotlib.org/stable/tutorials/introductory/pyplot.html. [Accessed: 27-Apr-2021].

[179] M. Waskom, “An introduction to seaborn.” [Online]. Available: https://seaborn.pydata.org/introduction.html. [Accessed: 27-Apr-2021].
[180] R. Blagus and L. Lusa, “SMOTE for high-dimensional class-imbalanced data,” *BMC Bioinformatics*, vol. 14, no. 1, p. 106, 2013.
[181] V. Jain and J. M. Chatterjee, *Machine Learning with Health Care Perspective: Machine Learning and Healthcare*, 1st ed. Cham: Springer International Publishing, 2020.
[182] S. Raschka and V. Mirjalili, *Python Machine Learning*, 1st ed. Birmingham, UK: Packt Publishing, 2017.
[183] J. Brownlee, *Imbalanced classification with Python: better metrics, balance skewed classes, cost-sensitive learning*, 1st ed. Calle de San Francisco: Machine Learning Mastery, 2020.
[184] Y. Charfaoui, “Hands-on with Feature Selection Techniques: Filter Methods,” 2020. [Online]. Available: https://heartbeat.fritz.ai/hands-on-with-feature-selection-techniques-filter-methods-f248e0436ce5. [Accessed: 15-Feb-2021].
[185] S. Raschka, *Python Machine Learning*, 1st ed. Birmingham, UK: Packt Publishing, 2015.
[186] F. Provost, “Machine learning from imbalanced data sets 101,” in *Proceedings of the AAAI’2000 workshop on imbalanced data sets*, 2000, vol. 68, no. 2000, pp. 1–3.
[187] “sklearn.model_selection.GridSearchCV.” [Online]. Available: https://scikit-learn.org/stable/modules/generated/sklearn.model_selection.GridSearchCV.html. [Accessed: 02-Mar-2021].
[188] A. Gabell and U. S. L. Nayak, “The effect of age on variability in gait,” *J. Gerontol.*, vol. 39, no. 6, pp. 662–666, 1984.