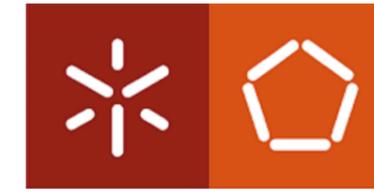




Técnicas de *machine learning* aplicadas à marcha para monitorização personalizada da toma da medicação levodopa em doentes com Parkinson

Paulo Jorge Martins da Silva

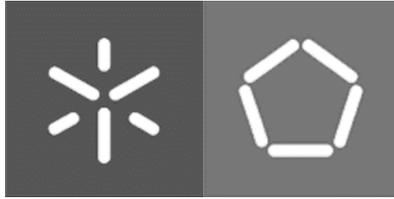


**Universidade do Minho**  
Escola de Engenharia

Paulo Jorge Martins da Silva

**Técnicas de *machine learning* aplicadas à  
marcha para monitorização personalizada  
da toma da medicação levodopa em  
doentes com Parkinson**





**Universidade do Minho**  
Escola de Engenharia

Paulo Jorge Martins da Silva

**Técnicas de *machine learning* aplicadas à  
marcha para monitorização personalizada  
da toma da medicação levodopa em doentes  
com Parkinson**

Dissertação de Mestrado Integrado em Engenharia Eletrónica  
Industrial e Computadores

Trabalho efetuado sob a orientação da  
**Professora Doutora Estela Bicho**  
**Doutora Flora Ferreira**

Novembro de 2021

## **DIREITOS DE AUTOR E CONDIÇÕES DE UTILIZAÇÃO DO TRABALHO POR TERCEIROS**

Este é um trabalho académico que pode ser utilizado por terceiros desde que respeitadas as regras e boas práticas internacionalmente aceites, no que concerne aos direitos de autor e direitos conexos.

Assim, o presente trabalho pode ser utilizado nos termos previstos na licença [abaixo](#) indicada.

Caso o utilizador necessite de permissão para poder fazer um uso do trabalho em condições não previstas no licenciamento indicado, deverá contactar o autor, através do RepositóriUM da Universidade do Minho.

### **Licença concedida aos utilizadores deste trabalho**



**Atribuição-NãoComercial-SemDerivações**  
**CC BY-NC-ND**

<https://creativecommons.org/licenses/by-nc-nd/4.0/>

# Agradecimentos

Em primeiro lugar, gostaria de agradecer do fundo do coração às duas pessoas que tornaram possível tudo o que sou hoje, os meus pais, Francisco Silva e Cristina Silva, que sempre foram modelos exemplares em todo o meu percurso académico em todas as suas vertentes psicológicas, financeiras e paternais.

Ao meu irmão, Marco Silva, por ter me mostrado que devemos seguir os nossos sonhos, independente do quanto assustador isso possa parecer.

À minha orientadora, Professora Doutora Estela Bicho, por ter depositado a sua confiança em mim para a realização deste projeto, por me ter incentivado a ser mais e melhor mesmo quando eu próprio duvidava das minhas capacidades.

À minha “co-orientadora”, Doutora Flora Ferreira, que foi o maior alicerce na construção desta dissertação, o meu muito obrigado. Pelo incentivo na procura da melhor solução possível, pelo rigor e perfeccionismo imposto na escrita e estruturação da dissertação, pela tranquilidade transmitida nos momentos mais desafiantes e por toda a disponibilidade mostrada ao longo deste percurso.

Aos meus amigos de curso, que foram um grande apoio e ajuda em todo o meu percurso académico. Em especial ao grupo “Boticas” que proporcionou estes cinco anos inesquecíveis. Um agradecimento particular ao meu parceiro Diogo Costa que sempre teve a palavra certa, no momento certo.

Ao Hospital da Nossa Senhora da Oliveira, em especial ao doutor Miguel Gago, pelos recursos imprescindíveis para a realização deste trabalho de investigação.

Ao engenheiro Carlos Fernandes pela orientação inicial e por toda a disponibilidade durante o desenvolvimento do estudo.

Por fim, à minha melhor amiga, minha companheira, Margarida Costa, por todo o afeto, paciência e apoio, por nunca me deixar desistir, nem desanimar. O meu muito obrigado. Um agradecimento muito especial também aos meus amigos de longa data – Catarina Cacho, João Costa, João Abreu, Helena Castro, Sara Marinho e Soraia Carvalho – por todos os momentos de pura felicidade e companheirismo.

À restante família e todas as outras pessoas que me apoiaram durante todos estes anos: um sincero obrigado!



## **DECLARAÇÃO DE INTEGRIDADE**

Declaro ter atuado com integridade na elaboração do presente trabalho académico e confirmo que não recorri à prática de plágio nem a qualquer forma de utilização indevida ou falsificação de informações ou resultados em nenhuma das etapas conducente à sua elaboração. Mais declaro que conheço e que respeitei o Código de Conduta Ética da Universidade do Minho.

# Resumo

Há cada vez mais evidências que sugerem que a avaliação da marcha pode ser uma ferramenta útil para apoiar a de doentes com Parkinson. A levodopa é o fármaco mais prescrito aos doentes com Parkinson e uma utilização adequada deste medicamento pode melhorar os sintomas e a qualidade de vida do paciente. Hoje em dia, com o desenvolvimento de sensores e dispositivos vestíveis, é possível utilizar um *smartwatch* para avaliar algumas características de marcha, que podem ser utilizadas na monitorização diária do estado de saúde. O objetivo deste estudo é investigar a possibilidade de monitorizar a toma do medicamento levodopa e ao mesmo tempo a presença do seu efeito na mobilidade paciente com Parkinson com base nas características de marcha possíveis de se obter com um *smartwatch*. Para tal, foi utilizado um conjunto de variáveis de marcha de 29 pacientes com Parkinson avaliados em dois estados, sem (estado OFF) e com (estado ON) o efeito da levodopa. Os dados foram normalizados tendo em conta as propriedades físicas do sujeito. Dois métodos estatísticos (ANOVA e Kendall) foram utilizados na seleção de variáveis. Os algoritmos *k-means* e *hierarchical clustering* foram aplicados para identificar subgrupos de pacientes com valores de marcha espaço-temporais semelhantes. De seguida, foram aplicadas cinco técnicas para gerar dados sintéticos a fim de aumentar o conjunto de dados utilizados na análise comparativa dos classificadores. A eficácia de diferentes classificadores de *machine learning*— *Support Vector Machine* (SVM), *Logistic Regression* (LR) e *Random Forest* (RF) baseados em medidas de marcha (tais como valor médio) e *Convolution Neural Network* (CNN) e *Long Short-Term Memory* (LSTM) baseados em séries temporais – foram avaliadas. Os melhores resultados foram obtidos com o classificador LR baseado na variabilidade da duração de um ciclo, com resultados de de aproximadamente 92%. Os resultados indicam que a utilização de técnicas de *machine learning* baseadas em variáveis de marcha, tais como a variabilidade da duração de um ciclo, que é possível aceder com um *smartwatch*, poderá ser um bom suporte para a monitorização do estado do paciente (ON/OFF). Note-se que um paciente classificado como "estado OFF" pode significar que ainda não tomou o medicamento ou que o medicamento já não está a fazer efeito.

**Palavras-Chave:** Avaliação da Marcha, Doença de Parkinson, *Machine Learning*, Medicamento Levodopa, *Smartwatch*

# Abstract

There is growing evidence suggesting that gait assessment can be a valuable tool to support the management of patients with Parkinson. Levodopa is the most prescribed medicine for Parkinson's Disease, and adequate use of this medication can improve the symptoms and the quality of patients' lives. Nowadays, with the development of wearable sensors and devices, it is possible to use a smartwatch to assess some gait characteristics that can be used to assist in daily health monitoring. This study aims to investigate the possibility of monitoring the levodopa medication-taking, and at the same time, the presence of levodopa effect in Parkinson patients' mobility based on gait features possible to obtain with a smartwatch.

For this purpose, a gait dataset from 29 patients with Parkinson Disease assessed in two states, without (OFF state) and with (ON state) the effect of levodopa, was used. Data were normalized considering the subject physical properties. Two statistical-based methods (ANOVA and Kendall) were used for feature selection. K-means and hierarchical clustering algorithms were applied to identify subgroups of patients with similar spatiotemporal gait values in the two states. Then, five techniques were applied to generate synthetic data to augment the dataset used in the comparative classification analysis. The effectiveness of different machine learning classifiers – Support Vector Machine (SVM), Logistic Regression (LR) and Random Forest (RF) based on gait measures (such as mean value) and Convolution Neural Network (CNN) and Long Short-Term Memory (LSTM) based on stride times series – were evaluated.

The best results were obtained with the LR classifier based on stride time variability, with accuracy results of 92% approximately. Results indicate that using machine learning techniques based on gait measures, such as stride time variability, that is possible to access with a smartwatch could be a good support for monitoring the patients state (ON/OFF). Note that a patient classified as “OFF state” can mean that he has not yet taken the medication or is no longer taking effect.

**Keywords:** Gait Assessment, Levodopa Medication, Machine Learning, Parkinson Disease, Smartwatch

## Índice

Agradecimentos .....	v
Resumo .....	viii
Abstract .....	ix
Lista de Figuras .....	xii
Lista de Tabelas .....	xiv
Acrónimos e Siglas .....	xv
Capítulo 1 Introdução.....	1
1.1 Motivação.....	1
1.2 Enquadramento .....	2
1.3 Objetivos e Resultados.....	3
1.4 Organização e Estrutura da Dissertação .....	3
Capítulo 2 Estado da Arte .....	4
2.1 Dispositivos vestíveis para avaliação da marcha.....	4
2.2 Avaliação da marcha nos doentes com Parkinson.....	6
2.3 Algoritmos de <i>machine learning</i> na análise da marcha.....	7
Capítulo 3 Base de Dados e Métodos.....	10
3.1 Base de Dados .....	10
3.2 Machine Learning .....	12
3.2.1 Normalização dos Dados.....	15
3.2.2 Escalonamento de Variáveis.....	15
3.2.3 Redução de dimensionalidade .....	17
3.2.4 Clustering .....	23
3.2.4.1 K-means .....	24
3.2.4.2 Hierarchical Clustering.....	25
3.2.5 Dados Sintéticos.....	26
3.2.6 Classificadores .....	29
3.2.6.1 Support Vector Machine.....	30
3.2.6.2 Random Forest Classifier .....	31
3.2.6.3 Logistic Regression .....	32
3.2.6.4 <i>Convolutional Neural Networks</i> (CNN) .....	33
3.2.6.5 Long Short-Term Memory (LSTM) .....	35
3.2.7 Métricas para avaliar o desempenho do Classificador (“ <i>Evaluate Classifier</i> ”) .....	36
Capítulo 4 Resultados .....	38
4.1 Seleção das variáveis de marcha mais relevantes.....	38
4.2 Análise de clusters.....	41
4.2.1 Com base na variabilidade da duração de um ciclo (GCT_SD) .....	41
4.2.2 Com base no comprimento de um ciclo (série temporal) .....	44
4.3 Dados sintéticos .....	45
4.4 Classificação .....	48
4.4.1 Com base na variabilidade da duração de um ciclo (GCT_SD) .....	48
4.4.2 Com base na variabilidade da duração de um ciclo (GCT_SD) e comprimento de um ciclo (SL) .....	51
4.4.3 Com base no comprimento de um ciclo (série temporal) .....	54
4.4.4 Com base no comprimento de um ciclo (série temporal) e duração de um ciclo (série temporal).....	56
4.5 Discussão .....	58
Capítulo 5 Conclusão, Limitações e Trabalho Futuro.....	60
5.1 Conclusão.....	60
5.2 Limitações.....	60
5.3 Trabalho Futuro.....	61
Referências.....	62

Anexos ..... 71

# Lista de Figuras

Figura 1 Esquema Geral do Projeto (ilustração <i>cartoon</i> à direita retirada de (B. K., 2019))	2
Figura 2 Variáveis espaço-temporais (baseado em (up, 2018))	6
Figura 3 Base de dados referente a um paciente em estado <i>OFF</i> . *Valores normalizados de acordo com as características físicas (idade, género, peso e altura) e velocidade da passada (Fernandes et al., 2021) (ver Secção 3.2.1)	11
Figura 4 Esquema Geral das etapas e técnicas de <i>Machine Learning</i> aplicadas	14
Figura 5 Exemplo de dados standardizados (baseado em (Sugiyama, 2015))	16
Figura 6 Exemplo de dados normalizados (baseado em (Sugiyama, 2015))	16
Figura 7 Avaliação do processo de treino (baseado em (Gollapudi, 2016))	17
Figura 8 <i>The Curse of Dimensionality</i> (baseado em (Gollapudi, 2016))	18
Figura 9 Fluxo dos Métodos de <i>Filter</i> (baseado em (Chandrashekar & Sahin, 2014))	19
Figura 10 Fluxo do Método de <i>Wrapper</i> (baseado em (Visalakshi & Radha, 2014))	20
Figura 11 Fluxo do Método <i>Embedded</i> (baseado em (J., 2016))	20
Figura 12 Tipos de Variáveis (baseado em (Kuhn & Johnson, 2013))	21
Figura 13 Seleção de uma técnica de <i>Filter</i> (baseado em (Kuhn & Johnson, 2013))	21
Figura 14 Correlação entre duas variáveis (baseado em (Akoglu, 2018))	23
Figura 15 Mapa de Calor (baseado em (Akoglu, 2018))	23
Figura 16 <i>Clustering</i> (baseado em (Swarndeeep Saket & Pandya, 2016))	24
Figura 17 Método do Cotovelo (baseado em (Na et al., 2010))	25
Figura 18 <i>K-means</i> (baseado em (Na et al., 2010))	25
Figura 19 Exemplo de um dendrograma (baseado em (Nielsen, 2016))	26
Figura 20 <i>Jittering</i>	27
Figura 21 <i>Scaling</i>	27
Figura 22 <i>Downsampling</i>	28
Figura 23 <i>Cutting</i>	28
Figura 24 <i>Permutation</i>	29
Figura 25 <i>Maximal Margin Classifier</i> (baseado em (Bhargava et al., 2020, Kumar et al., 2020))	30
Figura 26 <i>Support Vector Machine Illustration</i> (baseado em (Bhargava et al., 2020, Kumar et al., 2020))	30
Figura 27 Exemplo a 1 dimensão (baseado em (Bhargava et al., 2020, Kumar et al., 2020))	31
Figura 28 Exemplo a 2 dimensões (baseado em (Bhargava et al., 2020, Kumar et al., 2020))	31
Figura 29 Funcionamento Geral do <i>Random Forest</i> (baseado em (Guo et al., 2011, Misra & Li, 2019))	32
Figura 30 <i>Logistic Regression</i> (baseado em (Dreiseitl & Ohno-Machado, 2002, Kambria, 2019))	33
Figura 31 Arquitetura de uma <i>CNN</i> (baseado em (Srinivas et al., 2019))	34
Figura 32 <i>Max e Average Pooling</i> (baseado em (Albawi et al., 2017, Lai et al., 2018))	35
Figura 33 Célula <i>LSTM</i> (baseado em (Hochreiter, 1997))	36

## Lista de Figuras

Figura 34 Importância das variáveis baseadas nos valores da estatística F na <i>ANOVA</i> e nos valores do coeficiente tau-b de <i>Kendall</i> .....	39
Figura 35 Matrizes de correlação de <i>Pearson</i> entre variáveis de marcha no estado <i>ON</i> e <i>OFF</i> .....	40
Figura 36 Dendrograma do agrupamento obtido pelo método <i>Hierarchical</i> para 14 pacientes com base na variabilidade da duração de um ciclo em ambos os estados <i>OFF</i> e <i>ON</i> (GCT_SD_OFF+ GCT_SD_ON) .....	42
Figura 37 Resultado da aplicação do método do cotovelo .....	42
Figura 38 Resultado do agrupamento obtido pelo método de <i>K-means</i> para 14 pacientes com base na variabilidade da duração de um ciclo em ambos os estados <i>OFF</i> e <i>ON</i> (GCT_SD_OFF+ GCT_SD_ON) .....	43
Figura 39 Ilustração dos valores da variabilidade da duração de um ciclo (GCT_SD) nos estados <i>OFF</i> e <i>ON</i> nos grupos 2 e 4 obtidos através da análise de clusters .....	43
Figura 40 Dendrograma do agrupamento obtido pelo método <i>Hierarchical</i> para 14 pacientes com base no comprimento de um ciclo (série temporal) em ambos os estados <i>OFF</i> e <i>ON</i> (SL_ON + SL_OFF) .....	44
Figura 41 Séries temporais do comprimento de um ciclo nos estados <i>OFF</i> e <i>ON</i> dos pacientes agrupados em cada um dos sete grupos ilustrados na Figura 40. ....	45
Figura 42 Valores sintéticos da variabilidade da duração de um ciclo associados ao paciente 3 .....	46
Figura 43 Valores sintéticos da variabilidade da duração de um ciclo e da média do comprimento de um ciclo associados ao paciente 3 (GCT_SD+SL) .....	47
Figura 44 Séries temporais sintéticas do comprimento de um ciclo associada ao paciente 1 .....	48
Figura 45 Séries temporais originais do comprimento de um ciclo nos estados <i>OFF</i> e <i>ON</i> dos pacientes do grupo 7 .....	55
Figura 46 Séries temporais originais do comprimento de um ciclo (à esquerda) e da duração de um ciclo (à direita) nos estados <i>OFF</i> e <i>ON</i> dos pacientes do grupo 7 .....	56
Figura 47 Dados sintéticos (GCT_SD) .....	71
Figura 48 Dados sintéticos (GCT_SD+SL) .....	72
Figura 49 Séries temporais sintéticas (SL) .....	73

## Lista de Tabelas

Tabela 1 Dados antropométricos do grupo de pacientes com Parkinson .....	10
Tabela 2 Matriz de Confusão .....	37
Tabela 3 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 2 com base na <i>GCT_SD</i> .....	49
Tabela 4 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 4 com base na <i>GCT_SD</i> .....	49
Tabela 5 Valores médios das métricas de desempenho dos algoritmos de classificação para o grupo 2 e 4 com base na <i>GCT_SD</i> .....	51
Tabela 6 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 2 com base na <i>GCT_SD</i> e <i>SL</i> .....	52
Tabela 7 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 4 com base na <i>GCT_SD</i> e <i>SL</i> .....	53
Tabela 8 Valores médios das métricas de desempenho dos algoritmos de classificação para o grupo 2 e 4 com base na <i>GCT_SD</i> e <i>SL</i> .....	54
Tabela 9 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 7 com base nas séries temporais do comprimento de um ciclo .....	55
Tabela 10 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 7 com base nas séries temporais do comprimento de um ciclo e da duração de um ciclo .....	57
Tabela 11 Modelos de regressão linear resultantes para as variáveis de marcha .....	74

# Acrónimos e Siglas

<b>Acrónimo/Sigla</b>	<b>Significado</b>
AIC	<i>Akaike Information Criterion</i>
ANN	<i>Artificial Neural Network</i>
ANOVA	<i>Analysis of Variance</i>
CNN	<i>Convolutional Neural Network</i>
DP	Doença de Parkinson
DPI	Doença de Parkinson Idiopática
DPV	Doença de Parkinson Vascular
DT	<i>Decision Tree</i>
DTW	<i>Dynamic Time Warping</i>
ECG	Eletrocardiograma
FN	Falso Negativo
FP	Falso Positivo
FPR	<i>False Positive Rate</i>
GCT	Duração de um ciclo (" <i>Gait Cycle Time</i> ")
GCT_SD	Variabilidade (desvio padrão) da duração de um ciclo
GCT_SD_OFF	Variabilidade (desvio padrão) da duração de um ciclo sem o efeito da medicação (estado <i>OFF</i> )
GCT_SD_ON	Variabilidade (desvio padrão) da duração de um ciclo sob o efeito da medicação (estado <i>ON</i> )
IMU	<i>Inertial Measurement Unit</i>
L-dopa	<i>Levodopa</i>
LR	<i>Logistic Regression</i>
ML	<i>Machine Learning</i>
PC	<i>Principal Component</i>

## Acrónimos e Siglas

PCA	<i>Principal Component Analysis</i>
RF	<i>Random Forest</i>
RNN	<i>Recurrent Neural Network</i>
ROC Curve	<i>Receive Operating Characteristic Curve</i>
SD	Desvio Padrão
SL	Comprimento do Ciclo
SVM	<i>Support Vector Machine</i>
TN	Verdadeiro Negativo
TP	Verdadeiro Positivo
TPR	Sensibilidade ou <i>True Positive Rate</i>

# Capítulo 1

## Introdução

Neste capítulo apresenta-se as principais motivações, um breve enquadramento, objetivos e resultados do trabalho realizado e respetivo impacto esperado. Por fim é descrita a estrutura da dissertação.

### 1.1 Motivação

Atualmente a Doença de Parkinson é uma doença sem cura que afeta um bom setor da sociedade, tal como diz o doutor Luís Abreu no episódio 3 da websérie SINOPSE (dezembro de 2020), “*A Doença de Parkinson (DP) é a segunda doença neurodegenerativa mais prevalente na população mundial, logo a seguir à doença de Alzheimer*”. Afeta milhões de pessoas e, em Portugal, um estudo publicado em 2017 estima uma prevalência de pelo menos 180 casos por cada 100 mil habitantes, portanto, existem cerca de 18 a 20 mil pessoas com DP no nosso país” (L., 2020). Esta doença é um distúrbio neurológico do movimento, progressivo e degenerativo que à medida que vai progredindo retira bastante qualidade de vida ao doente. Contudo, apesar de atualmente ainda não existir uma cura para esta doença, já existe medicação bastante eficaz no controlo dos sintomas possibilitando ao doente uma vida muito próxima da normalidade. No entanto, esta medicação precisa de ser constantemente ajustada, pois ou o doente cria imunidade ao medicamento ou a DP progride para uma nova fase e a dose da medicação prescrita não é a mais adequada. A levodopa (L-dopa) é o medicamento mais prescrito para a DP e o tratamento adequado com este fármaco pode restaurar a função motora dos pacientes a um nível mais próximo do normal, promovendo melhor qualidade de vida a estes doentes. A avaliação da marcha tem-se revelado como uma útil ferramenta clínica complementar no diagnóstico diferencial (Fernandes et al., 2021), assim como na avaliação mais objetiva do benefício da levodopa no perfil da marcha (Fernandes et al., 2019), (Gago et al., 2017). O padrão de marcha pode ser medido de forma rápida e a baixo custo utilizando tecnologia não-invasiva, como os sensores portáteis e recentemente os *smartwatches*. Neste contexto, a principal motivação do estudo desenvolvido no âmbito desta dissertação assenta na possibilidade de contribuir para um acompanhamento mais personalizado e permanente dos doentes com a Doença de Parkinson através da monitorização personalizada da toma da medicação. Um sistema

inteligente baseado na avaliação da marcha, e com recurso à tecnologia existente, poderá ser usado para monitorizar se o paciente tomou a medicação na hora que deveria tomar e ao mesmo tempo avaliar o efeito da levodopa na mobilidade do paciente. Um alerta que o paciente não tomou a medicação quando o paciente na verdade tomou, pode significar que a medicação está a deixar de fazer efeito e desta forma alertar o médico da possibilidade de ser necessário ajustar a dose de levodopa antes que os sintomas se intensifiquem (Apda, 2021, cuf, 2021).

## 1.2 Enquadramento

Os dispositivos móveis como o *smartwatch* têm já sensores de marcha incorporados que poderão ser utilizados num futuro próximo para monitorizar remotamente os pacientes com a doença de Parkinson (Erdem et al., 2019). Neste sentido, é importante perceber que características da marcha possíveis de serem avaliadas pelo *smartwatch* serão úteis para monitorizar o estado do paciente em termos de mobilidade.

Por outro lado, as aplicações recentes de técnicas de *machine learning* (ML) em dados clínicos de rotina têm-se mostrado poderosas na seleção e extração de biomarcadores importantes que podem permitir previsões mais precisas e compreender melhor as doenças (Dadashi et al., 2013, Fernandes et al., 2021, Rovini et al., 2018, Tahir & Manap, 2012, Wahid et al., 2015).

Neste trabalho, um conjunto de dados de marcha de pacientes com Doença de Parkinson, obtidos com recurso a sensores portáteis no Hospital da Senhora de Oliveira, Guimarães, foi usado para investigar, recorrendo a diferentes técnicas de *machine learning*, quais as variáveis da marcha possíveis de serem avaliadas por um *smartwatch* são mais importantes na monitorização do estado do paciente (sob vs. sem o efeito da medicação levodopa) (Figura 1).

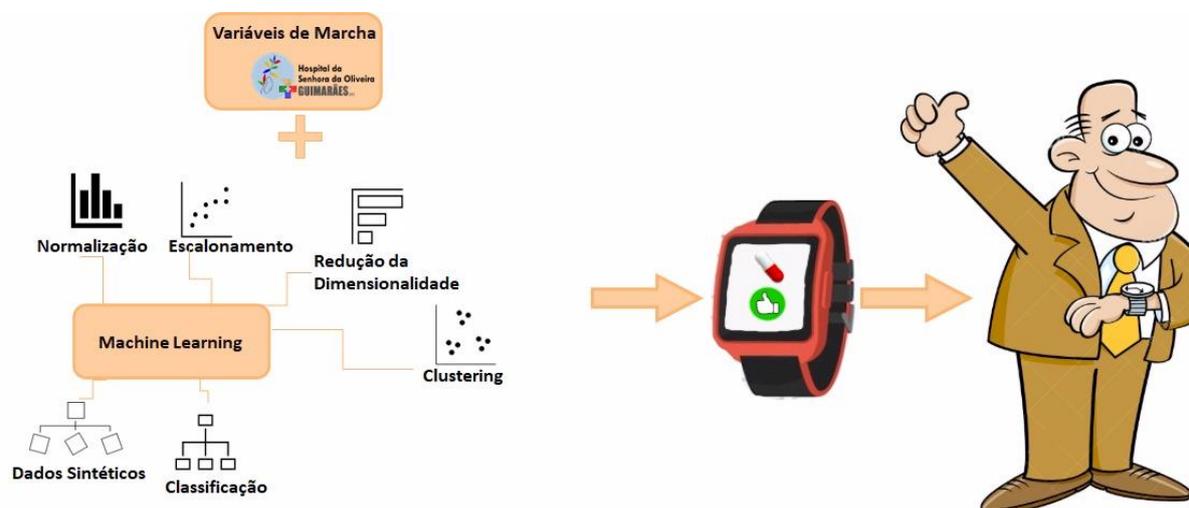


Figura 1 Esquema Geral do Projeto (ilustração *cartoon* à direita retirada de (B. K., 2019))

## 1.3 Objetivos e Resultados

Esta dissertação teve como objetivo geral investigar a possibilidade de monitorizar a toma da medicação levodopa com base na avaliação de marcha usando um *smartwatch* e aplicação de diferentes técnicas de *machine learning*. Mais precisamente, este trabalho teve os seguintes objetivos:

- Identificar quais as características da marcha possíveis de serem avaliadas usando um *smartwatch* atualmente disponível no mercado poderão ser úteis numa monitorização objetiva da toma da medicação levodopa.
- Avaliar a eficácia de diferentes técnicas de *machine learning* com base em dados de marcha na classificação do estado de pacientes com Doença de Parkinson (sob vs. sem o efeito da medicação levodopa).

Como resultado final, foi possível identificar um conjunto de características de marcha, recolhidos por um *smartwatch*, e algoritmos de *machine learning* mais indicados para desenvolver um sistema de monitorização remota da toma da medicação levodopa. Como impacto, espera-se assim contribuir para a evolução de ferramentas de monitorização remota, que permitirão um acompanhamento mais personalizado e contínuo dos pacientes com Doença de Parkinson.

## 1.4 Organização e Estrutura da Dissertação

No capítulo 2 é descrito o estado da arte, onde é feito um levantamento e uma breve apreciação crítica de outros trabalhos de investigação relacionados com o trabalho desenvolvido nesta dissertação.

No capítulo 3 é explicada a base de dados utilizada assim como os métodos estatísticos e de *machine learning* aplicados.

No capítulo 4 são apresentados e analisados os resultados obtidos da aplicação dos diferentes métodos, incluindo os resultados da aplicação de técnicas de seleção das variáveis mais relevantes e os resultados da aplicação de diferentes algoritmos de classificação do estado do paciente (sob vs. sem o efeito da medicação).

No capítulo 5, são apresentadas as principais conclusões, as limitações do trabalho desenvolvido e sugestões para trabalho futuro.

## Capítulo 2

### Estado da Arte

Neste capítulo é descrito o estado da arte no contexto do tema em estudo, onde são abordados três pontos importantes:

- Os dispositivos vestíveis para avaliação da marcha, em particular para medição de variáveis espaço-temporais da marcha;
- O uso da avaliação da marcha como ferramenta auxiliar de diagnóstico e monitorização de doentes com Parkinson;
- Os algoritmos de *machine learning* na análise da marcha.

#### 2.1 Dispositivos vestíveis para avaliação da marcha

Os dispositivos vestíveis podem ser utilizados para enfrentar alguns dos desafios relacionados com a deteção e gestão de condições de saúde adversas nas populações. Os dados dos sensores vestíveis ajudam a rastrear importantes estatísticas e/ou sintomas de doenças, e especialmente de doenças crónicas (Site et al., 2021). Entre os vários sensores que podem ser utilizados para monitorizar os sinais vitais de uma doença estão incluídos os sensores que permitem avaliar a marcha de um indivíduo. Os sensores inerciais (por exemplo acelerómetro, giroscópio, magnetómetros) são cada vez mais adotados em muitas aplicações práticas, e amplamente incorporados em dispositivos móveis/vestíveis (por exemplo, *smartphones*, *smartwatches*, *smartbands*). Esta tecnologia oferece uma nova abordagem de reconhecimento de marcha, na qual, a amostragem de dados de sinal por sensores inerciais ligados ao corpo do usuário (enquanto ele está a caminhar) é usada como fonte de dados para o reconhecimento da mesma (Sprager & Juric, 2015). As características dos dispositivos vestíveis, tais como flexibilidade, usabilidade, ausência de fios e conforto, fez com que esta abordagem rapidamente recebesse um esforço de pesquisa significativo, e resultados promissores foram alcançados nomeadamente na identificação de padrões de marcha característicos na doença de Parkinson (Brognara et al., 2019), (Schlachetzki et al., 2017). As variáveis espaço-temporais da marcha, tais como velocidade e comprimento de um ciclo, têm

sido usadas para avaliar a marcha tanto em populações saudáveis como em populações com patologias (Bugané et al., 2012, Muthukrishnan et al., 2020). Essas medidas têm-se revelado importantes na caracterização de dificuldades na marcha (Fernandes et al., 2019, 2018), assim como, servir como um biomarcador para monitorizar a progressão da doença ao longo do tempo (Esser et al., 2011), prever quedas (Gray & Hildebrand, 2000) e desenvolver programas direcionados de reabilitação da marcha (Studer et al., 2017). Diferentes algoritmos têm sido usados para calcular as variáveis espaço-temporais a partir de dados derivados de sensores de unidade de medição inercial (IMU) (Muthukrishnan et al., 2020). Sensores de IMU são uma combinação de vários sensores, como giroscópio, acelerómetro e magnetómetro. Sistemas baseados em IMU, como GAITRite® (GAITRite, 2021), RehaWatch® (Hasomed, 2021) e GaitUp® (Gaitup, 2020), incluem um software que processa os dados recolhidos e fornece as estatísticas (e.g. média, desvio padrão) de cada variável da marcha (e.g. velocidade, duração de um ciclo), assim como as séries temporais de cada variável. Neste trabalho, foram usados dados obtidos pelo sistema GaitUp®. O sistema é composto por dois sensores *Physilog*® e um software para computador. Para avaliar a marcha os dois sensores são colocados na parte superior dos pés com uma cinta elástica. *Physilog*®, é um sensor de movimento da *Gait Up* com sincronização sem fios, e que inclui um acelerómetro triaxial, um giroscópio triaxial, um magnetómetro triaxial, e um sensor de pressão barométrica. Em ambiente hospitalar, normalmente é solicitado ao doente para percorrer um percurso contínuo enquanto os sensores estão adquirindo os dados (Ferreira et al., 2019), (Belloni et al., 2021). Usando o software fornecido pela *Gait Up* com base num algoritmo de fusão patenteado na deteção de eventos de marcha e modelagem biomecânica, os valores das variáveis espaço-temporais de cada ciclo da marcha são obtidos conforme descrito e validado em estudos anteriores (Dadashi et al., 2013, Mariani et al., 2010). Diferentes variáveis espaço-temporais, incluindo a velocidade, comprimento de um ciclo da marcha, a duração de um ciclo, o tempo de apoio/balanço, a cadência, a elevação máxima/mínima do pé, podem ser recolhidas usando o sistema da *Gait Up*.

Sensores, tais como acelerómetro e giroscópio, incorporados nos *smartwatches* permitem extrair algumas variáveis espaço-temporais, tais como a velocidade, o comprimento da passada, e o tempo em duplo suporte (Erdem et al., 2019). Estudos recentes têm investigado o uso dos *smartwatches* na monitorização do estado de saúde dos utilizadores. Em (Gwak et al., 2018), um *smartwatch* e um oxímetro de pulso sem fio foram usados para detectar perda de capacidade cognitiva. Musale et al. (Musale et al., 2019) avaliaram a capacidade de um *smartwatch* comercial detetar padrões distintos da marcha do utilizador. Os resultados experimentais revelaram uma boa taxa de acerto de autenticação, com uma taxa de erro média de cerca de 8.2%.

Olhando ao mercado atual temos uma enorme variedade de *smartwatches* à nossa disposição, das mais variadas gamas, tendo como principal foco os *smartwatches* que possuem funcionalidades de "mobility metrics" tais como: *Apple Watch Series 6*, *Samsung Galaxy Watch Active 2*, *Garmin Fenix 6*, entre outros. Algumas variáveis espaço-temporais fazem parte do conjunto de "mobility metrics" possíveis de serem avaliados usando um *smartwatch*, sendo as mais comuns (Garmin, 2021, S., 2021) (ver Figura 2):

- a velocidade de um ciclo ("*stride velocity*"),
- o comprimento de um ciclo ("*stride length*"),
- a duração de um ciclo ("*gait cycle time*") e
- tempo em duplo suporte, ou seja, o instante em que ambos os pés se encontram em contacto no solo ("*double support time*"),
- cadência da passada (número de passos por minuto).



Figura 2 Variáveis espaço-temporais (baseado em (up, 2018))

## 2.2 Avaliação da marcha nos doentes com Parkinson

A avaliação da marcha tem sido usada nos últimos anos como uma ferramenta complementar de diagnóstico e monitorização da progressão da doença em doentes com Parkinson (Del Din et al., 2019). As síndromes de Parkinson manifestam-se por bradicinesia, rigidez, dificuldade de marcha e instabilidade postural, com etiologias heterogêneas, desde a doença idiopática de Parkinson (DPI) a síndromes atípicas de Parkinson, tais como o Parkinson Vascular (VaP) (Obeso et al., 2017), (Ferreira et al., 2019). As opções terapêuticas são limitadas, essencialmente por estimulação do sistema dopaminérgico através do fármaco levopoda (L-Dopa). Esta opção farmacológica é dúbia, não só pela raridade de ensaios clínicos, como também pela escassez de ferramentas de avaliação, que persistem essencialmente qualitativas

(escalas clínicas). Não existem diretrizes claras quanto à quantidade de levodopa que deve ser administrada, com a maioria dos autores a defenderem um tratamento individualizado baseado na avaliação clínica. Um estudo recente (Ferreira et al., 2019) revelou que a avaliação da marcha é uma potencial ferramenta clínica de personalização terapêutica em doentes com Parkinson. Diferentes circuitos neuronais de controlo da marcha (locomoção e estabilidade postural), com envolvimento dopaminérgico ou não dopaminérgico, poderão estar representados em diferentes variáveis da marcha. Como tal, a análise da marcha pode permitir identificar objetivamente o benefício da levodopa nos vários domínios da marcha e auxiliar na monitorização mais personalizada de cada paciente.

Vários estudos têm analisado o uso da avaliação da marcha tanto para o diagnóstico da doença de Parkinson (Fernandes et al., 2019, 2018) como para uma avaliação objetiva do efeito da levodopa (Bryant et al., 2011), (Bryant et al., 2016), (Son et al., 2021), (Ferreira et al., 2019). Um dos resultados destes estudos é que a levodopa pode diminuir a variabilidade em algumas das variáveis da marcha (duração de um ciclo, tempo de balanço, comprimento de um ciclo e velocidade de um ciclo) (Bryant et al., 2011, 2016), (Fernandes et al., 2019). Assim, avaliar a variabilidade da marcha poderá ser uma forma objetiva de monitorizar a toma da medicação levodopa e ao mesmo tempo avaliar o seu efeito.

## **2.3 Algoritmos de *machine learning* na análise da marcha**

Os algoritmos de *machine learning* consistem em vários tipos de algoritmos de classificação, agrupamento e previsão e algoritmos de redes neuronais utilizados para a análise de diferentes tipos de dados. Incluem também vários algoritmos de seleção de variáveis, tais como métodos baseados em *wrapper* e métodos baseados em filtros. Os dados obtidos a partir de sensores e dispositivos vestíveis podem ser extraídos para parâmetros importantes relacionados com a saúde e estes dados podem ser utilizados para criar vários modelos de classificação, agrupamento e previsão para a monitorização de doenças (Parkinson por exemplo), início de sintomas, etc. Estes dados digitais de saúde podem ser em formato estruturado ou semiestruturado. Diferentes estruturas de cuidados de saúde os dados necessitam de diferentes tipos de métodos de pré-processamento e análise para extrair os biomarcadores importantes a partir de dados em bruto. Dependendo dos dados e da informação a extrair dos dados, são utilizados vários algoritmos de *machine learning* para análise e pré-processamento (Braga et al., 2020, Caramia et al., 2018, Fernandes et al., 2018, Rehman et al., 2019).

Trabalhos recentes, e.g. (Braga et al., 2020), (Rehman et al., 2019), mostram que um subconjunto de características de marcha selecionadas utilizando *Random Forest (RF)*, *Support Vector Machine*

(*SVM*), Regressão Logística (*LR*) e Convolutional Neural Network (*CNN*), melhora a taxa de acerto da classificação da doença de Parkinson.

Diferentes técnicas de *machine learning* (ML) têm sido utilizadas para selecionar a melhor combinação de características de marcha relevantes para classificação (Caramia et al., 2017, Fernandes et al., 2019), (Rehman et al., 2019). Num estudo recente (Rehman et al., 2019) verificou-se que um subconjunto de características de marcha selecionadas utilizando *Random Forest* com ganho de informação e eliminação recursiva (RFE) com base nos algoritmos de classificação *Support Vector Machine* (SVM) e *Logistic Regression* melhora a taxa de acerto da classificação da doença de Parkinson. Os modelos largamente utilizados em *machine learning* para a classificação da marcha são SVM (Aich et al., 2018, Braga et al., 2020, Mannini et al., 2016, Pradhan et al., 2015, Wahid et al., 2015), *Random Forest* (RF) (Aich et al., 2018, Braga et al., 2020, Wahid et al., 2015) e *K-Nearest Neighbor* (KNN) (Pradhan et al., 2015, Wahid et al., 2015). Os resultados destes estudos mostram uma boa taxa de acerto de desempenho na classificação da marcha patológica.

Recentemente, foi também adotada uma técnica de *deep learning* para a tarefa de classificação de marcha (El Maachi et al., 2020, Fernandes et al., 2021). Em (El Maachi et al., 2020), as séries temporais de marcha foram utilizadas como sinais de entrada para o reconhecimento automático da marcha de Parkinson. Foi alcançada uma taxa de acerto de 98,7% utilizando uma abordagem baseada em *Convolution Neural Networks* (CNNs). Em (Fernandes et al., 2021), as séries temporais da marcha foram usadas como entrada para diferenciação de dois tipos de síndromes Parkinsonianas (doença de Parkinson idiopática vs parkinsonismo vascular). Antes de aplicar os algoritmos de classificação baseados em CNNs os valores das séries temporais da marcha foram normalizados de acordo com as características físicas dos pacientes usando um método de regressão linear. De seguida, o algoritmo de eliminação recursiva (RFE) baseada em diferentes algoritmos de classificação, *Support Vector Machine* (SVM), *lasso* e *random forest* foram usados para selecionar as variáveis mais importantes. Os melhores classificadores com base em CNNs alcançaram precisões de 79,33%, 82,33% e 86,00% usando os dados recolhidos sem o efeito da medicação (*OFF* estado), usando os dados recolhidos sob o efeito da medicação levodopa (*ON* estado), e usando os dados tanto recolhidos sem como sob o efeito da medicação (*OFF* e *ON* estado), respetivamente.

Para melhorar a performance dos modelos de *machine learning* também existem algumas técnicas de geração de dados de sintéticos que são aplicadas quando existe uma escassez de dados. Esta é uma abordagem que tem-se revelado eficaz na melhoria da robustez do modelo de *machine learning*. Nesta abordagem, a quantidade de dados é aumentada várias vezes através da introdução de ruído ou da

modificação apropriada dos dados originais (LeCun et al., 1998), (Shorten & Khoshgoftaar). Com mais dados de treino, podemos reduzir o *overfitting*, e treinar um modelo de forma mais robusta. Devido à sua eficácia, o aumento de dados tem sido amplamente adotado em várias tarefas (Tran & Choi, 2020). O aumento de dados alavanca dados limitados através da transformação das amostras existentes para criar novas amostras. Um desafio chave para o aumento de dados é gerar novos dados que mantenham o rótulo ("*label*") correto, o que normalmente requer conhecimento no domínio. No entanto, não é óbvio como levar a cabo o aumento dos dados e ter em consideração a conservação de rótulos em alguns domínios, por exemplo, dados de sensores vestíveis (Taewoong Um et al., 2017). As técnicas de aumentos de dados ("*data augmentation*") mais comuns (Lopez-Nava et al., 2020, Taewoong Um et al., 2017) são *jittering* (multiplica um valor aleatório a cada elemento de cada sinal), *scaling* (altera a magnitude de os dados numa janela multiplicando por um escalar aleatório), *downsampling* (seleciona de séries de entrada aleatórias e faz subamostragens uniformes para reduzir a série) e *cutting* (reduz em igual número a série de entrada).

## Capítulo 3

### Base de Dados e Métodos

Neste capítulo é, inicialmente, descrito de forma detalhada a base de dados utilizada. De seguida, todos os métodos estatísticos e de *machine learning*, incluindo técnicas de normalização dos dados, de seleção de variáveis, de agrupamento (*clustering*), de geração de dados sintéticos, e de classificação, são apresentados pela ordem que foram aplicados.

#### 3.1 Base de Dados

A avaliação da marcha foi realizada a 15 pacientes com doença de Parkinson idiopática (*DPI*) e 14 pacientes com doença de Parkinson Vascular (*DPV*) (Fernandes et al., 2021) (Tabela 1). Para a seleção destes 29 pacientes com Parkinson foram utilizados os seguintes critérios de exclusão: existência de tremores, demência moderada e severa ( $CDR > 2$ ), doença esquelética e evidências de progressão clínica de doença desde o momento em que a mesma foi diagnosticada ( $Hoehn-Yahr > 3$ ) (Ferreira et al., 2020).

Tabela 1 Dados antropométricos do grupo de pacientes com Parkinson

	<b>Pacientes DPI</b>	<b>Pacientes DPV</b>
Idade (anos)	76.60 ± 4.29	80.53 ± 4.63
Peso (kg)	73.24 ± 12.53	66.17 ± 10.38
Altura (m)	1.67 ± 0.082	1.61 ± 0.085
Sexo (Feminino/Masculino)	4/11	6/8

Para a aquisição das variáveis de marcha foram atados dois sensores ao dorso de cada sapato através de duas bandas elásticas. De salientar que os sensores não exigiam quaisquer cuidados de alinhamentos ou calibração à priori. De seguida os pacientes participantes no estudo foram convidados a percorrer um percurso em linha reta de 30 metros, inverter o sentido da marcha e voltar a regressar ao ponto de partida a uma velocidade de marcha livre. Os pacientes foram avaliados em dois momentos distintos: no estado "OFF" (sem o efeito da medicação), após 12 horas sem qualquer toma de medicação; e no estado "ON" (sob o efeito da medicação), 60 minutos após tomar a medicação L-dopa (150% da dose matinal habitual). Os dados obtidos nos dois momentos foram convertidos em séries

temporais de marcha através de um software de análise de marcha ("*Gait Up*", *Switzerland*") (Fernandes et al., 2019). Diferentes variáveis da marcha foram avaliadas em cada ciclo de marcha (também conhecida como passada). Um ciclo de marcha começa quando o pé de referência entra em contacto com o chão e termina quando esse mesmo pé entra em contacto com o chão novamente. Este estudo focou-se em quatro variáveis de marcha passíveis de obter através do *smartwatch*: comprimento de um ciclo ("*stride length*"), cadência ("*cadence*"), percentagem de duração do ciclo em tempo de duplo suporte ("*double support*") e duração de um ciclo ("*gait cycle time*" ou "*stride time*"). Para cada paciente e momento (ON e OFF), após descartar os dois primeiros valores de cada uma das séries temporais, foram selecionados os valores dos primeiros 30 ciclos de marcha em linha reta (ver exemplo na Figura 3).

N	Paciente	Estado	Ciclos	Grupo	Duração de um Ciclo*	Cadência*	Tempo de Duplo Suporte*	Comprimento de um Ciclo*
691	N186	OFF	3	DPI	0,92711679	1,092929128	1,079374595	0,912658413
692	N186	OFF	4	DPI	0,98011124	1,033049602	1,179277357	0,922620782
693	N186	OFF	5	DPI	0,931741025	1,075435195	1,143853434	0,937445474
694	N186	OFF	6	DPI	0,920016992	1,091281141	1,142256739	0,905010053
695	N186	OFF	7	DPI	0,95285354	1,061594757	1,107579656	0,934862461
696	N186	OFF	8	DPI	0,93984361	1,079833925	1,142805057	0,88867967
697	N186	OFF	9	DPI	0,938680155	1,083521292	1,278444996	0,901016743
698	N186	OFF	10	DPI	0,920235418	1,102214472	1,171557236	0,848110087
699	N186	OFF	11	DPI	0,982186691	1,023108011	0,962848322	1,011805881
700	N186	OFF	12	DPI	1,074376319	0,940848174	1,064696529	1,04064021
701	N186	OFF	13	DPI	0,947914036	1,082631091	0,945145209	0,858887688
702	N186	OFF	14	DPI	0,924152107	1,10998655	1,209858917	0,873503791
703	N186	OFF	15	DPI	0,980898135	1,027483016	1,19063025	0,9386717
704	N186	OFF	16	DPI	0,923573367	1,098642417	1,164928931	0,902113736
705	N186	OFF	17	DPI	0,967325981	1,05939104	1,0773436	0,910082982
706	N186	OFF	18	DPI	0,971704237	1,045242702	1,195920712	0,920334435
707	N186	OFF	19	DPI	0,955486755	1,076466919	0,992488161	0,872568374
708	N186	OFF	20	DPI	0,957849823	1,072553474	0,916930671	0,890892805
709	N186	OFF	21	DPI	0,987166838	1,036514033	0,900677495	0,905644516
710	N186	OFF	22	DPI	0,968624057	1,044766153	1,048905505	0,946260103
711	N186	OFF	23	DPI	0,97659595	1,032070865	1,088391348	0,936816617
712	N186	OFF	24	DPI	0,934687094	1,084465857	1,14517887	0,899788083
713	N186	OFF	25	DPI	0,94938516	1,080185629	1,207241744	0,882554114
714	N186	OFF	26	DPI	0,962155769	1,061599113	0,985952194	0,891422799
715	N186	OFF	27	DPI	0,931219818	1,084753097	1,155790878	0,934256469
716	N186	OFF	28	DPI	0,942896282	1,082582638	1,20268315	0,865451977
717	N186	OFF	29	DPI	1,001360439	1,005549686	1,17489484	0,990228972
718	N186	OFF	30	DPI	0,956589922	1,054545626	1,100500525	0,928999531
719	N186	OFF	31	DPI	0,963586736	1,061240295	1,053162066	0,914571463
720	N186	OFF	32	DPI	0,966544326	1,064869819	1,162462226	0,887693321

Figura 3 Base de dados referente a um paciente em estado *OFF*. \*Valores normalizados de acordo com as características físicas (idade, género, peso e altura) e velocidade da passada (Fernandes et al., 2021) (ver Secção 3.2.1)

## 3.2 Machine Learning

*Machine Learning* é um conceito enorme no mundo da engenharia, que abrange tantos fatores que é praticamente impossível universalizar uma definição que consiga o descrever na totalidade. Segundo Aurélien Géron (Géron, 2019), *machine learning* é o campo de estudo que dá aos computadores a capacidade de aprender sem ser explicitamente programado. Alguns fatores chaves para o seu correto funcionamento são a velocidade de execução, a sua taxa de acerto e a sua segurança.

Os métodos de *machine learning* são tipicamente categorizados em aprendizagem não supervisionada ("*unsupervised*"), reforçada ("*reinforcement*") e supervisionada ("*supervised learning*") (Bishop, 2006). Algoritmos de *machine learning* supervisionados são os mais comuns quando falamos de diagnóstico, estes algoritmos necessitam que os dados estejam devidamente rotulados, sendo seu principal foco estabelecer uma relação entre os atributos dos dados de entrada (por exemplo tamanho da região do cérebro num MRI scan) e saída (por exemplo o diagnóstico arquivado) e através dessas relações faz a previsão das saídas para as novas entradas. Contudo, categorizar uma grande quantidade de dados pode ser um grande desafio para um *machine learning* supervisionado. Um algoritmo supervisionado também pode ser dividido em algoritmos de classificação e regressão. Classificação consiste, por exemplo, em prever o diagnóstico de arritmias com base no resultado de um exame ECG como entrada. Regressão já é algo mais matemático, em que o seu objetivo é prever o valor real de uma variável de entrada.

*Machine learning* não supervisionado, em contraste com supervisionado, não necessita de dados categorizados e são úteis para fazer *clustering* de dados, ou na redução da dimensão das bases de dados através de simplificações de dados complexos (Becht et al., 2019, Maaten & Hinton, 2008). Quando vocacionamos mais à saúde, os algoritmos de *clustering* focam em encontrar os endotipos dos pacientes, identificando padrões nos dados e agrupando áreas pelo seu grau de similaridade. Algoritmos de *clustering* não supervisionados também podem ser utilizados para fazer previsões.

À combinação dos algoritmos supervisionados e não supervisionados, chama-se semi-supervisionado ("*semi-supervised*") (Chapelle et al., 2009). Este método utiliza alguns dados rotulados juntamente com dados não rotulados, o que permite, por exemplo, o método *clustering (unsupervised)* melhorar a performance do método classificação (*supervised*). Em suma, este método é vantajoso quando a quantidades de dados é pequena. Por fim, o algoritmo de aprendizagem por reforçado ("*reinforcement learning*") (Watkins, 1989) funciona através de recompensas ou penalizações para se alcançar a saída pretendida.

Contudo quando se fala de *machine learning* é errado pensar que este se resume aos algoritmos de modelação do problema, os Classificadores, pois existe uma preparação dos dados muito importante antes da aplicação desses mesmos. Por exemplo, em muitos casos, é necessário primeiro uniformizar a escala de todas as variáveis para que estas tenham a mesma influência e peso no modelo que irá definir o nosso problema, salvaguardando más modelações por escalas de variáveis muito discrepantes (Escalonamento das Variáveis), seleccionar as variáveis que verdadeiramente irão ter influência na construção do modelo e eliminar as desnecessárias através de um estudo da correlação entre elas e da relevância dessas na modelação (Seleção das Variáveis), ou/e reduzir a dimensão da base de dados para diminuir a complexidade do modelo nos classificadores (Redução da Dimensão dos Dados).

A Figura 4 apresenta uma visão geral das etapas e técnicas de *machine learning* aplicadas neste trabalho, e descritas em detalhe de seguida.

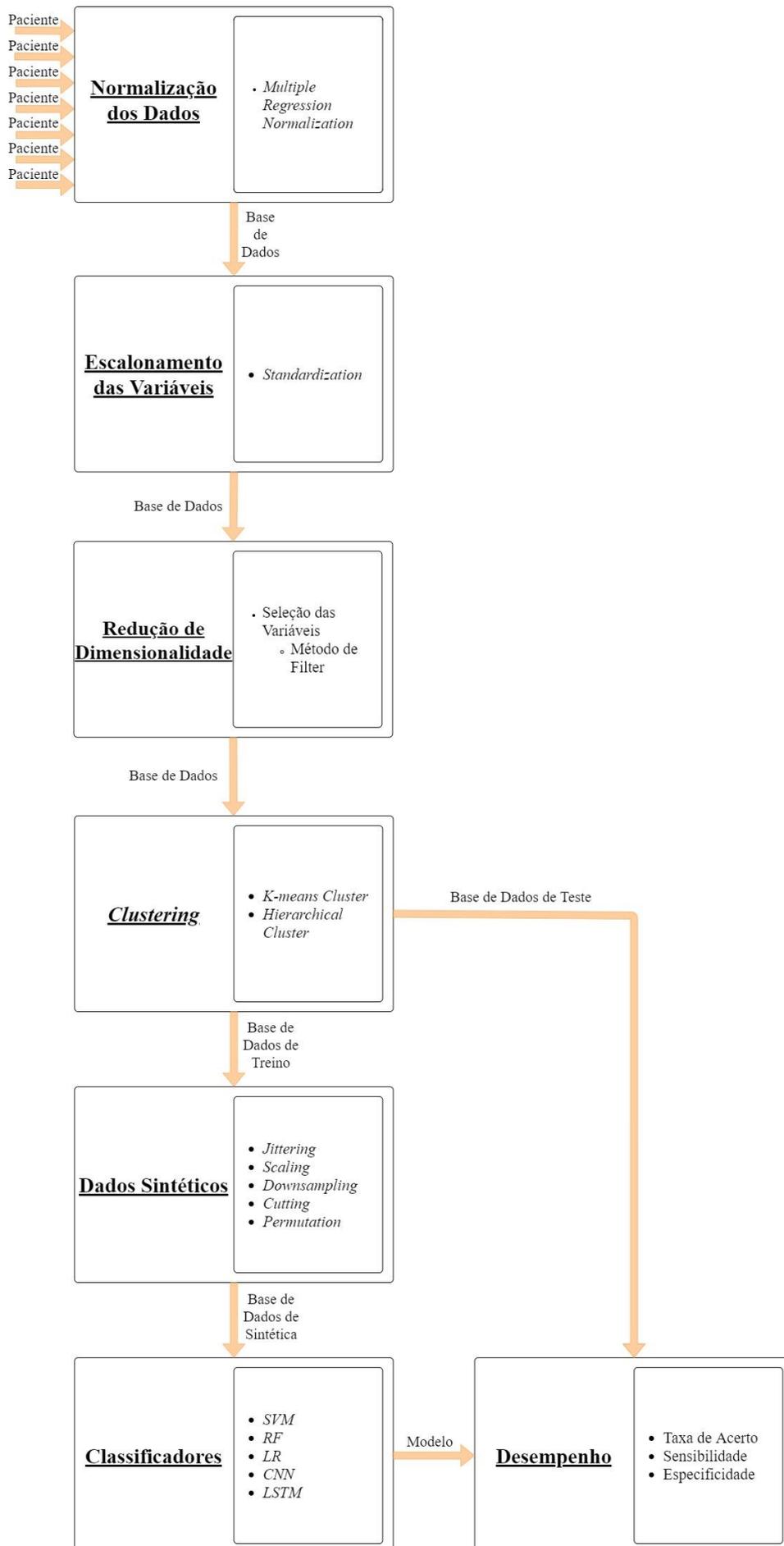


Figura 4 Esquema Geral das etapas e técnicas de *Machine Learning* aplicadas

### 3.2.1 Normalização dos Dados

A marcha de cada individuo é significativamente afetada pelas suas características físicas tais como altura, peso, idade e sexo, assim como pela velocidade da passada (Dadashi et al., 2013, Kirtley et al., 1985, Senden et al., 2012). Para atenuar o efeito das diferenças físicas e da velocidade da passada todas as séries temporais dos pacientes utilizadas neste estudo foram normalizadas usando o método *multiple regression normalization* em estudo anterior (Fernandes et al., 2021). Em (Wahid et al., 2016) melhores resultados na redução da interferência das propriedades físicas específicas do sujeito e velocidade da passada foram obtidos usando o método *Multiple Regression Normalization* em comparação com outros métodos tais como *dimensionless equations* e *detrending methods*. Cada ciclo da passada foi normalizado com base num modelo de regressão múltipla dado pela seguinte equação:

$$\hat{y}_i = \beta_0 + \sum_{j=1}^p \beta_j * x_{ij} + \varepsilon_i \quad (3.1)$$

onde  $\hat{y}_i$  representa a previsão para a variável de marcha dependente para a observação (ciclo da marcha)  $i$ ,  $\beta_0$  é termo constante,  $\beta_j$  corresponde ao coeficiente da característica física  $j$ ,  $x_{ij}$  é a variável independente  $j$  e, por fim,  $\varepsilon_i$  representa o erro residual para a observação  $i$ . As variáveis altura, peso, idade, sexo e velocidade da passada foram consideradas candidatas a variáveis independentes. Modelos de regressão linear foram calculados para todas as combinações possíveis de variáveis independentes linearmente correlacionadas com a variável dependente. Os coeficientes dos modelos foram estimados usando os valores médios das características físicas e da marcha de um grupo de 34 pessoas saudáveis. O critério de informação de Akaike (AIC) e a estatística  $R^2$  ajustado foram usados para selecionar o modelo de melhor ajuste. Para cada variável da marcha, o modelo de regressão de melhor ajuste foi usado para normalizar os valores de cada ciclo da marcha dessa variável dividindo o valor original  $y_i$  pelo valor previsto em (3.1)  $\hat{y}_i$  (Wahid et al., 2015):

$$y_i^n = \frac{y_i}{\hat{y}_i} \quad (3.2)$$

onde  $y_i^n$  representa o valor normalizado da observação  $i$ . Mais detalhes podem ser consultados em (Fernandes et al., 2021). Os modelos de regressão linear usados para normalizar os dados usados neste estudo estão apresentados na Tabela 11 em anexo.

### 3.2.2 Escalonamento de Variáveis

Uma das causas principais para um mau desempenho do algoritmo de *machine learning* é quando as escalas das variáveis são bastantes desniveladas. Variáveis com grandes escalas possuem um peso

muito superior menosprezando as variáveis com escalas mais pequenas. Para combater este problema existem técnicas de escalonamento de variáveis que consistem numa universalização das escalas de todas as variáveis. Entre o vasto número de técnicas de escalonamento as mais comuns são: *Standardization* e *Normalization* (Géron, 2019).

**Standardization** é uma técnica de escalonamento que centra os dados com média nula e desvio padrão unitário. Esta técnica é definida pela Equação (3.3) em que  $x'$  são os dados padronizados,  $\mu$  a média e  $\sigma$  o desvio padrão.

$$x' = \frac{x - \mu}{\sigma} \tag{3.3}$$

Na Figura 5 é demonstrada a transformação dos dados pela aplicação da *standardization*.

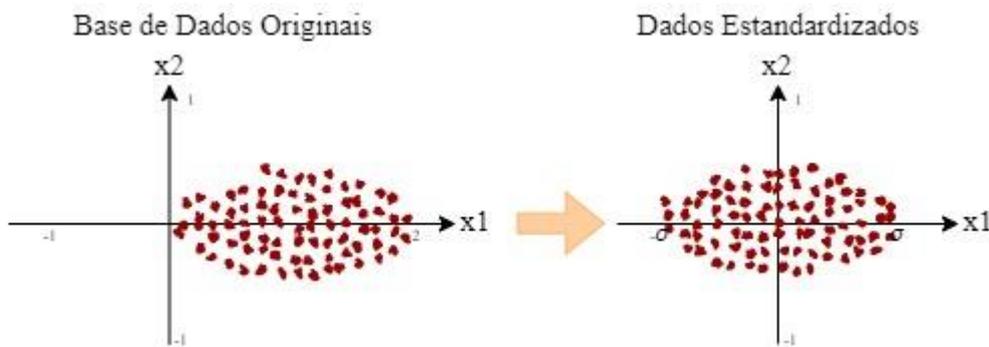


Figura 5 Exemplo de dados estandarizados (baseado em (Sugiyama, 2015))

**Normalization**, também conhecido como *Min-Max Scaling*, desloca e redimensiona os dados para um intervalo entre [0,1]. Este algoritmo é definido pela Equação (3.4) onde  $x'$  são os dados normalizados.

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)} \tag{3.4}$$

Na Figura 6 é ilustrado a transformação dos dados pela normalização.

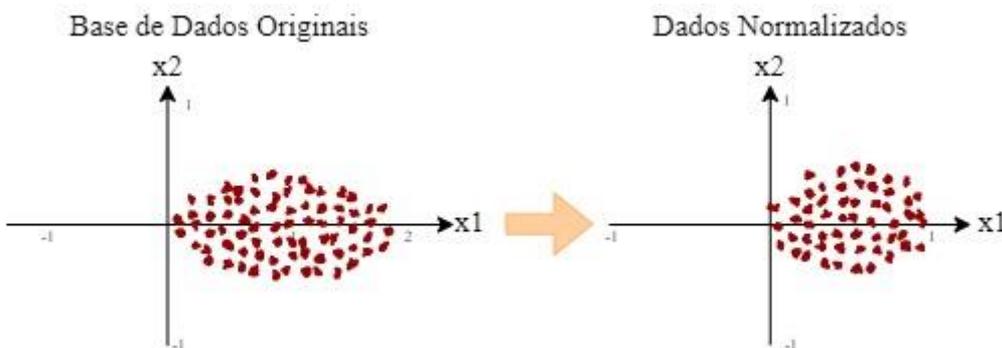


Figura 6 Exemplo de dados normalizados (baseado em (Sugiyama, 2015))

Como o problema desta dissertação é muito sensível a pacientes atípicos ("outliers") irei utilizar o método de *standardization* porque é menos sensível a *outliers* (Géron, 2019).

### 3.2.3 Redução de dimensionalidade

Quando se trabalha com conjuntos de dados grandes e complexos pode ser aconselhado utilizar técnicas de diminuição da dimensão para tornar mais acessível a computação dos modelos de *machine learning*. Outro benefício subjacente à redução de dimensionalidade é a eliminação ou redução de *overfitting* que acontece quando o modelo de *machine learning* se ajusta demais à base de dados de treino, e não consegue categorizar corretamente os dados de teste (que não foram usados no processo de treino). De notar, que um bom modelo de *machine learning* é generalizar bem a partir dos dados de treino de forma a permitir fazer previsões sobre dados que o modelo nunca viu. Assim, um bom modelo não pode sofrer de *overfitting* nem *underfitting*. *Underfitting* acontece quando o modelo não consegue capturar a tendência subjacente aos dados de treino impossibilitando uma previsão correta (Figura 13).

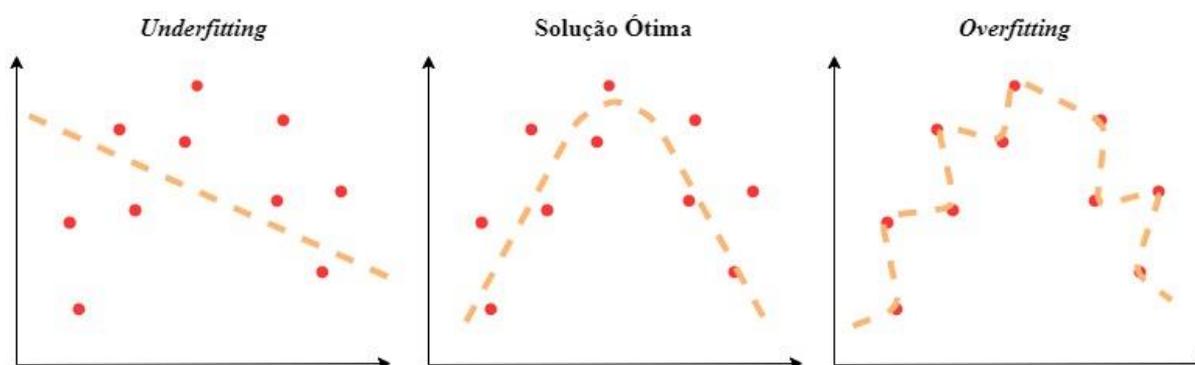


Figura 7 Avaliação do processo de treino (baseado em (Gollapudi, 2016))

A este balanceamento entre performance e dimensão de dados intitula-se de *The Curse of Dimensionality*, Figura 8.

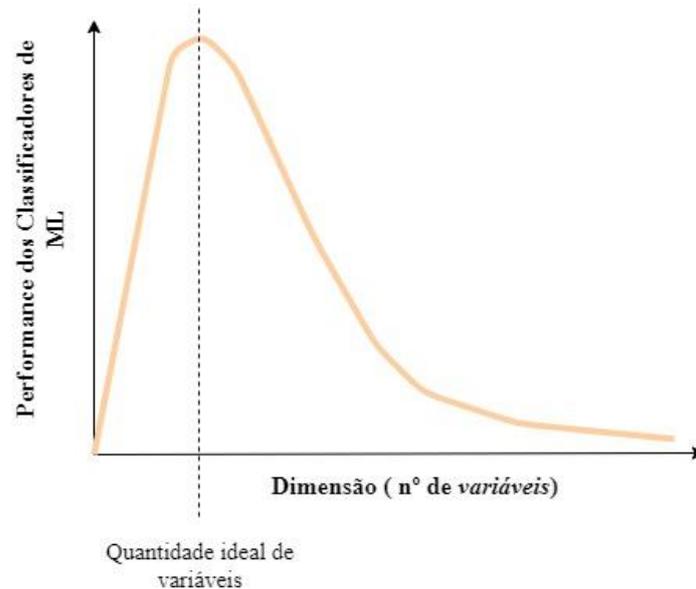


Figura 8 *The Curse of Dimensionality* (baseado em (Gollapudi, 2016))

A redução de dimensionalidade ("*Dimensionality Reduction*") pode ser obtida usando duas abordagens: seleção e/ou extração de variáveis (El Bouchefry & de Souza, 2020, Sugiyama, 2015).

### Seleção de variáveis

A seleção de variáveis consiste, de uma maneira sucinta, em selecionar as variáveis mais relevantes para um dado problema. Esta abordagem tem como objetivo remover variáveis de ruído, irrelevantes e redundantes do conjunto de dados original. Existem inúmeras vantagens em realizar uma seleção de variáveis antes de iniciar o nosso modelo de treino, tais como: modelos mais simples logo mais fáceis de interpretar, redução do tempo de treino, redução da probabilidade de *overfitting*, diminuição do poder computacional necessário, melhoramento do desempenho do nosso modelo.

Em termos de aplicação, os métodos mais usuais são: Método de *Filter*, Método de *Wrapper*, Método *Embedded* e Método de *Hybrid*.

**Método de *Filter*** - A característica principal deste método é que faz a seleção das melhores variáveis no conjunto de dados de forma totalmente independente de qualquer algoritmo de *machine learning*, ou seja, existem variáveis que são descartadas das bases de dados mesmo antes do processo de aprendizagem iniciar (Figura 9). Este método realiza a seleção das variáveis através de critérios de pontuação, atribuindo scores a cada variável e ordenando-as de forma decrescente. Os critérios de avaliação das variáveis baseiam-se em distância, informação, dependência e consistência. As suas principais vantagens são: suporta grande quantidade de dados, necessita de um baixo poder computacional, salvaguardando assim possíveis problemas de *overfitting*, e possui um processamento rápido e independente das técnicas de aprendizagem. No entanto umas das suas grandes desvantagens

é que ao descartar/ignorar por completo as técnicas de aprendizagem pode comprometer substancialmente a performance (Chandrashekar & Sahin, 2014).

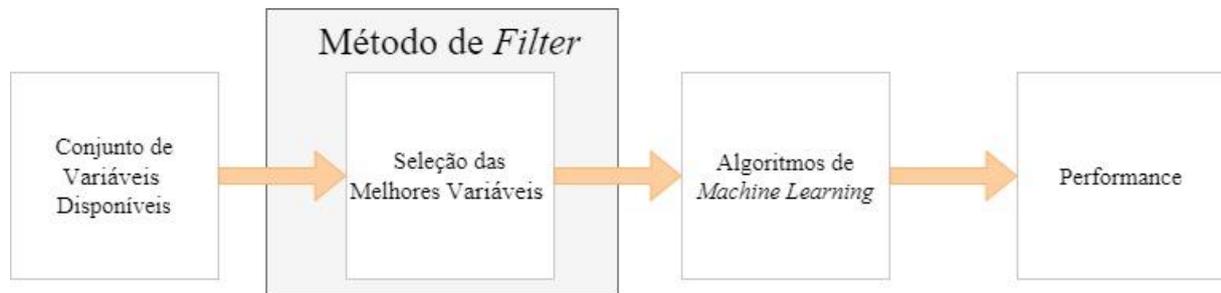


Figura 9 Fluxo dos Métodos de *Filter* (baseado em (Chandrashekar & Sahin, 2014))

Os métodos de *filter* podem-se dividir em *univariate* e *multivariate*. *Univariate* processa cada variável individualmente, no entanto, isso pode originar que sejam selecionadas variáveis redundantes (variáveis que não trazem qualquer informação adicional). Métodos de *filter multivariate*, por outro lado, têm em consideração todo o conjunto de variáveis, isto permite salvaguardar variáveis repetidas, redundantes e correlacionadas.

**Método de *Wrapper*** - Tal como foi visto anteriormente uma das principais desvantagens dos métodos de *filter* era ignorarem o efeito que o subconjunto das variáveis selecionado teria na performance do algoritmo de *machine learning*. O mesmo não se sucede no método de *Wrapper*, este procura avaliar o melhor subconjunto de variáveis através dos algoritmos, ou seja, analisa cada possibilidade de subconjuntos de dados com base na qualidade da performance num dado algoritmo. Contudo este método pode obrigar a um grande poder computacional e muitas vezes torna-se mesmo impraticável se as bases de dados forem muito grandes. Por outro lado, uma das suas vantagens é detetar as interações entre as variáveis com intuito de descobrir o melhor conjunto de variáveis para um determinado algoritmo. É importante salientar que, na aplicação deste método é necessário ser definido um critério de paragem, este é escolhido pelo próprio programador, sendo que existem alguns exemplos já universalizados: diminuição da performance do modelo, aumento da performance do modelo, número máximo de variáveis alcançado, etc. Caso este critério não seja bem definido o modelo corre alto risco de *overfitting*. Sobre o processo de procura do subconjunto de variáveis que melhor se adequa ao modelo existem alguns métodos, tais como: *Forward Feature Selection* (o modelo inicia o processo sem nenhuma variável e vai adicionando uma de cada vez), *Backward Feature Elimination* (o modelo começa com todas as variáveis e vai removendo uma de cada vez, sendo que é sempre removida aquela com menor peso), *Exhaustive Feature Selection* (o modelo experimenta todas as combinações possíveis de variáveis), *Recursive Feature Elimination* (este modelo cria vários modelos e guarda a melhor e pior variável de cada

iteração, ele faz esse processo exaustivamente até todas as serem testadas, no final constrói um ranking com base na ordem de eliminação), entre outros (Visalakshi & Radha, 2014).

Na Figura 10 é ilustrado um resumo do processo de funcionamento do Método de *Wrapper*.

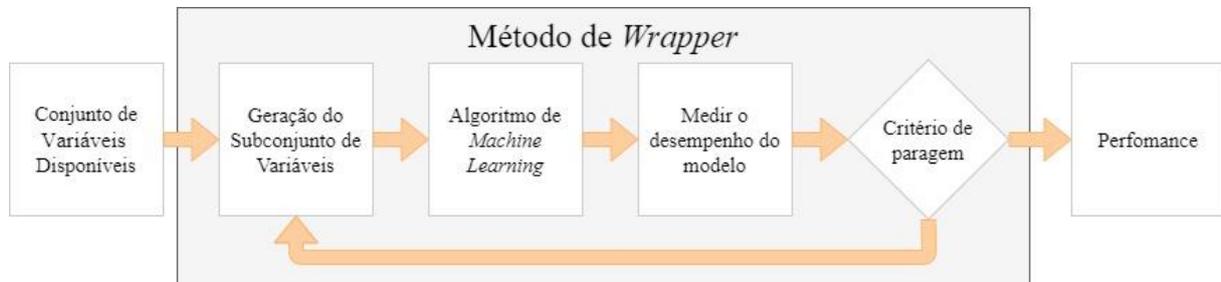


Figura 10 Fluxo do Método de *Wrapper* (baseado em (Visalakshi & Radha, 2014))

**Método de *Embedded*** - É uma combinação do Método de *Filter* com o *Wrapper*, ou seja, este método é uma junção das vantagens de ambos os métodos (Figura 11). Dos exemplos de aplicações mais comuns deste método é a regressão LASSO e RIDGE que consistem em funções de penalização para reduzir o *overfitting* (J., 2016).

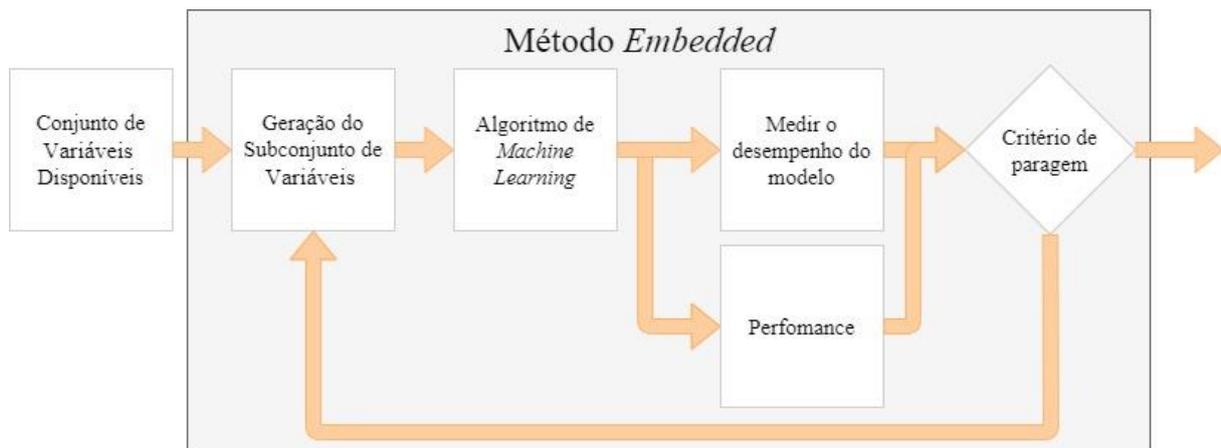


Figura 11 Fluxo do Método *Embedded* (baseado em (J., 2016))

**Método de *Hybrid*** - O método *hybrid* é uma combinação de vários métodos. Por exemplo, *Filter* com *Wrapper*, inicialmente utiliza-se um método de *Filter* para eliminar variáveis correlacionadas, e de seguida usa-se um método *Wrapper* para selecionar o melhor subconjunto de variáveis. Este método pode originar resultados bastantes promissores desde que aplicado corretamente. Outro exemplo de combinação mais utilizado é o Método *Embedded* e *Wrapper*.

Tendo em atenção o tamanho da base de dados utilizada e a relação entre as variáveis o método adotado, nesta dissertação, foi o Método de *Filter*. Para selecionar a técnica de *Filter* a aplicar é necessário analisar que tipo de variáveis tem a base de dados. Uma variável pode ser caracterizada em

numérica, que correspondem a variáveis matemáticas, ou categórica, que são os caracteres ou conjuntos de caracteres (Figura 12).

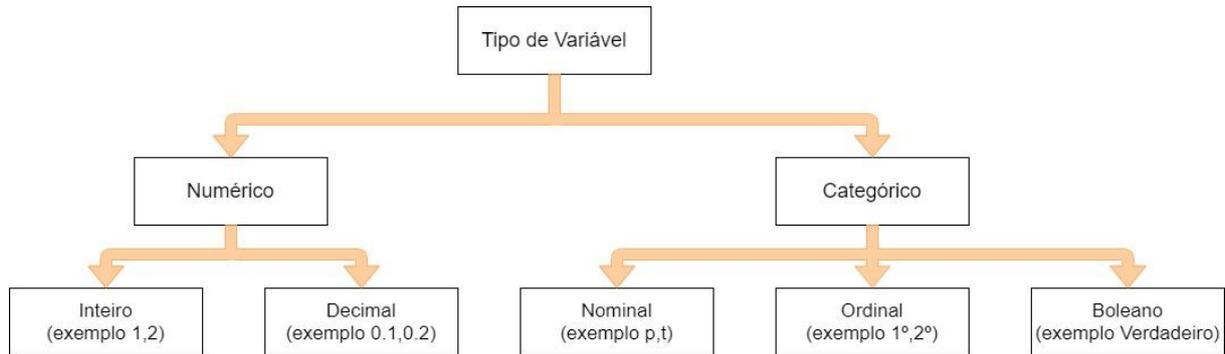


Figura 12 Tipos de Variáveis (baseado em (Kuhn & Johnson, 2013))

De acordo com o tipo de variável de entrada e de saída, uma técnica (e.g. Coeficiente de correlação de *Pearson*, *ANOVA*) pode ser selecionada para avaliar a importância da variável de entrada em relação à variável de saída de forma independente (Figura 13).

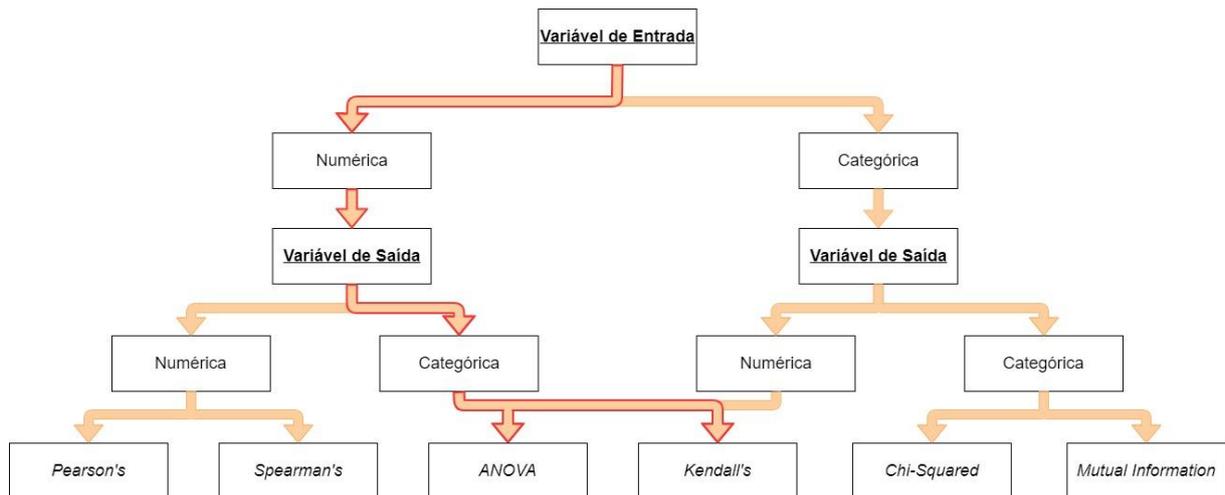


Figura 13 Seleção de uma técnica de *Filter* (baseado em (Kuhn & Johnson, 2013))

No caso do problema em estudo, a variável de saída é categórica porque é a variável referente ao estado do paciente: ON - está sobre o efeito da medicação e OFF - não está sobre o efeito da medicação. Por outro lado, as variáveis de entrada são numéricas porque se referem às variáveis da marcha. Assim dois possíveis métodos a aplicar são: teste ANOVA e correlação de tau de *Kendall*.

## **ANOVA**

*ANOVA* é o acrónimo de Análise de Variância (Ostertagová & Ostertag, 2013) e é um teste de hipótese estatístico paramétrico que no contexto em estudo permite analisar a variação de uma variável da marcha entre os dois momentos (ON vs. OFF). Na realização de um teste ANOVA uma estatística  $F$ , em homenagem a Sir Ronald Fisher, é calculada, baseada na razão de quadrados médios (razão entre

a soma de quadrados e os graus de liberdade) através da tabela de distribuição F, e usada para testar a hipótese nula ( $H_0$ : o valor médio é igual nos dois momentos) contra a hipótese alternativa ( $H_1$ : o valor médio nos dois momentos difere). A aplicação deste método de filtragem pressupõe algumas condições necessárias (Sauro & Lewis, 2016):

- Todas as observações devem ser independentes;
- As variáveis devem seguir uma distribuição, aproximadamente normal;
- As variâncias em cada grupo devem ser aproximadamente iguais;
- A variável dependente deve ser contínua.

No caso de as variáveis não apresentarem uma distribuição normal ou aproximadamente normal uma estratégia é transformar os dados para normalizar a distribuição e diminuir as variâncias. As principais transformações aplicadas a conjuntos de dados são: Logarítmica, Raiz quadrada e Arco-seno da raiz quadrada (Fraiman & Fraiman, 2018).

### **Coeficiente tau de *Kendall***

O coeficiente de correlação de *Kendall* é uma estatística usada para medir a força da dependência/correlação entre duas variáveis (Chen et al., 2002). Ele calcula uma pontuação normalizada para o número de classificações correspondentes ou concordantes entre duas amostras de dados. Este método não possui nenhuma suposição sobre a distribuição dos dados para ser aplicado.

A correlação de *Kendall* entre duas variáveis será elevada se as observações tiverem uma classificação semelhante (ou idêntica no caso de correlação igual a 1). A correlação de *Kendall* será baixa quando as observações tiverem uma classificação diferente quando comparadas as duas variáveis (ou completamente diferente no caso de correlação igual a -1) (Y., 2020).

### **Correlação de *Pearson***

Para avaliar a independência entre as variáveis de entrada a correlação de *Pearson* foi também aplicada. Duas variáveis de entrada fortemente correlacionadas têm o mesmo efeito sobre a variável de saída e então podem ser redundantes.

A correlação de *Pearson* estuda a correlação linear existentes entre duas variáveis. Diz-se que duas variáveis estão linearmente correlacionadas quando a relação entre estas se aproxima de uma função linear, tal como é possível ver na Figura 14. Correlação positiva ocorre quando há uma aglomeração dos

pontos com tendência crescente, ou seja, quando uma variável aumenta a outra também aumenta. Por outro lado, diz-se que existe correlação negativa quando os pontos se concentram em linha decrescente, à medida que uma variável aumenta a outra diminui. Por fim, diz-se correlação nula quando os pontos não seguem nenhuma tendência, nem positiva nem negativa, ou seja, há uma grande dispersão dos pontos (Sauro & Lewis, 2016), (Jo\st et al., 2017). Diz-se que duas variáveis estão moderadamente ou fortemente correlacionadas quando o valor absoluto do coeficiente de correlação é igual ou superior a 0.6 ( $|r| \geq 0.6$ ) (Akoglu, 2018).

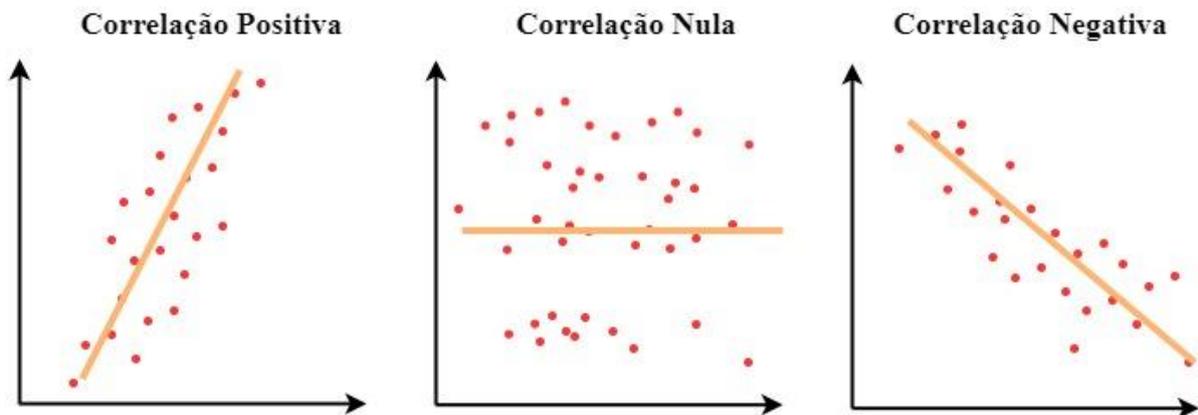


Figura 14 Correlação entre duas variáveis (baseado em (Akoglu, 2018))

Quando o estudo da correlação entre as variáveis excede as duas variáveis em simultâneo é comum a utilização de um diagrama de calor ("*heatmap*") para representar a correlação entre as variáveis, sendo a correlação das variáveis dada por uma escala de cores (Figura 15).

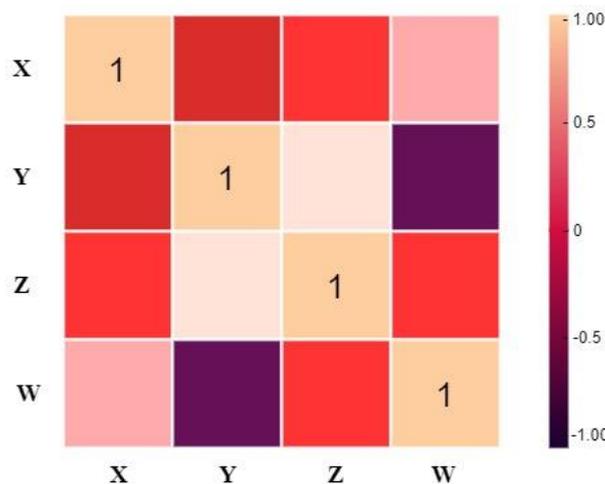


Figura 15 Mapa de Calor (baseado em (Akoglu, 2018))

### 3.2.4 Clustering

*Clustering* é um tipo de aprendizagem não supervisionada na qual são extraídos padrões do conjunto de dados de entrada sem as saídas estarem identificadas/rotuladas. Contudo, apesar de ser

um método de aprendizagem não supervisionado, este é também utilizado em aprendizagem supervisionada nomeadamente para classificar conjuntos de dados, pois agrupa os dados mais similares (Swarndeeep Saket & Pandya, 2016), (Bindra & Mishra, 2017).

Em suma, este algoritmo é basicamente uma ferramenta de agrupamento dos dados por grupos de similaridade e por consequência identificar e separar os dados mais distintos (Figura 16).

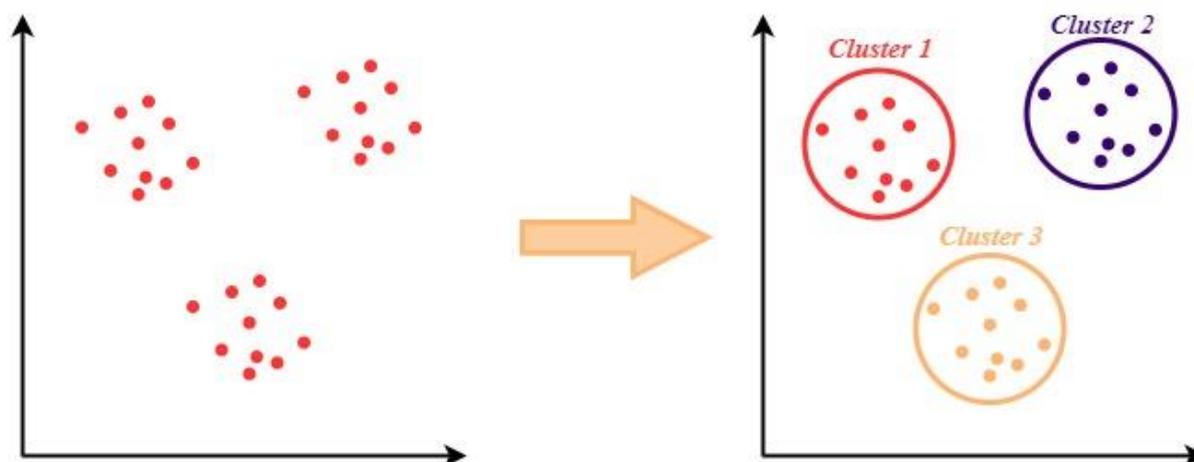


Figura 16 *Clustering* (baseado em (Swarndeeep Saket & Pandya, 2016))

Existem inúmeros algoritmos de *Clustering* (Bano & Khan, 2018) dois quais os dois mais usuais e que serão implementados neste trabalho são: *K-means* e *Hierarchical Clustering*. Ambos os métodos implementados realizam o agrupamento dos pacientes por similaridade através de uma análise das distâncias calculadas entre eles, estas distâncias, quando aplicadas a características de marcha, podem ser calculadas pela distância euclidiana, definida como a soma da raiz quadrada da diferença entre o paciente  $x$  e o  $y$  em suas respectivas dimensões. No caso das séries temporais o método *dynamic time warping (DTW)* é mais adequado do que a distância euclidiana para medir a similaridade entre duas séries (Berndt & Clifford, 1994).

### 3.2.4.1 K-means

O objetivo do *K-means* é agrupar pontos de dados similares e descobrir padrões subentendidos (Na et al., 2010). Para alcançar este objetivo, *K-means* precisa de definir um número  $k$  de clusters no conjunto de dados antes de iniciar. Essa limitação causa alguma ambiguidade no valor de  $k$  a definir. Para auxiliar nessa decisão utiliza-se frequentemente o método do cotovelo (Figura 17). Basicamente, através deste método, diz-se que o valor de  $k$  ideal é alcançado quando a variação da distorção do  $k$  para  $k + 1$  já não é significativa. Por exemplo, na Figura 17 o valor de  $k$  ideal seria 2, pois a variação da distorção (variação da distorção = distorção <sub>$k=3$</sub>  - distorção <sub>$k=2$</sub> ) já não é significativa quando comparada com a variação da distorção entre  $k=1$  e  $k=2$ .

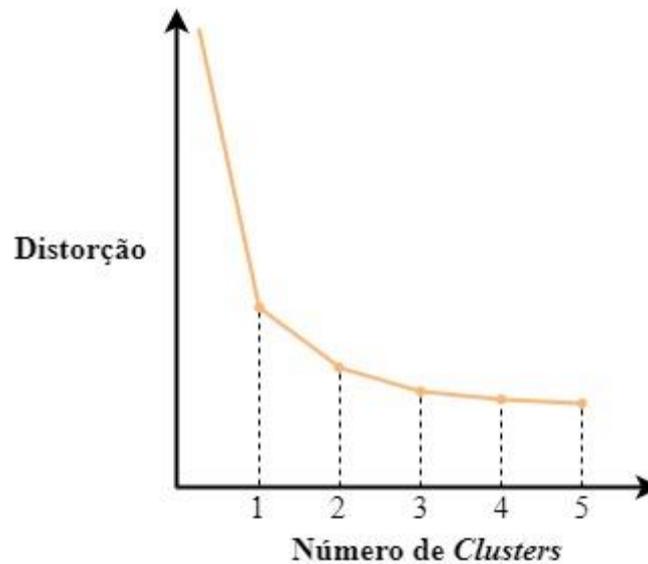


Figura 17 Método do Cotovelo (baseado em (Na et al., 2010))

Após definido o valor de  $k$  o algoritmo *K-means* está apto a executar. O número de centroides é definido pelo  $k$ , estes são o ponto de partida de cada *cluster*. Inicialmente a posição dos centroides é escolhida aleatoriamente. De seguida o algoritmo através de cálculos matemáticos irá ajustar a posição dos centroides, sendo que, este processo é repetido iterativamente até que matematicamente se atinja a solução ótima, quando isso se sucede significa que foram encontradas as posições dos centroides que melhor definem os dados (Figura 18).

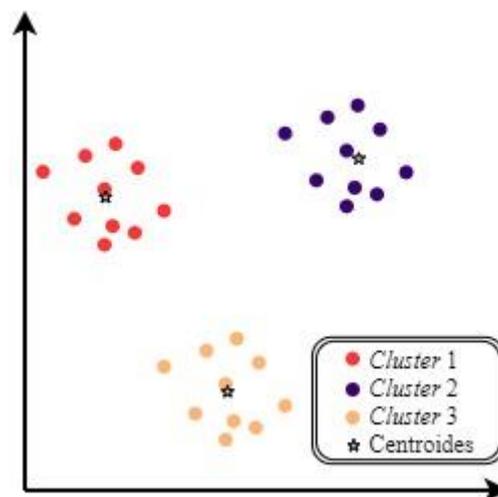


Figura 18 *K-means* (baseado em (Na et al., 2010))

### 3.2.4.2 Hierarchical Clustering

*Hierarchical Clustering*, tal como o nome sugere, é um algoritmo que constrói uma hierarquia de clusters (Nielsen, 2016). Este algoritmo pode ser dividido em duas abordagens: aglomeração (abordagem de baixo para cima) e divisiva (abordagem do topo para baixo). Na aglomeração este método inicia por considerar cada paciente como um cluster individual. De seguida agrega os dois clusters mais

próximos e assim sucessivamente até só restar um único cluster, sendo esta a abordagem mais popular. Na abordagem divisiva, o *hierarchical clustering* faz exatamente o oposto, inicialmente considera todos os pacientes como um cluster único e a cada iteração desagrupa os pacientes menos similares. O resultado final de ambas as alternativas podem ser observadas através de um dendrograma (Figura 19).

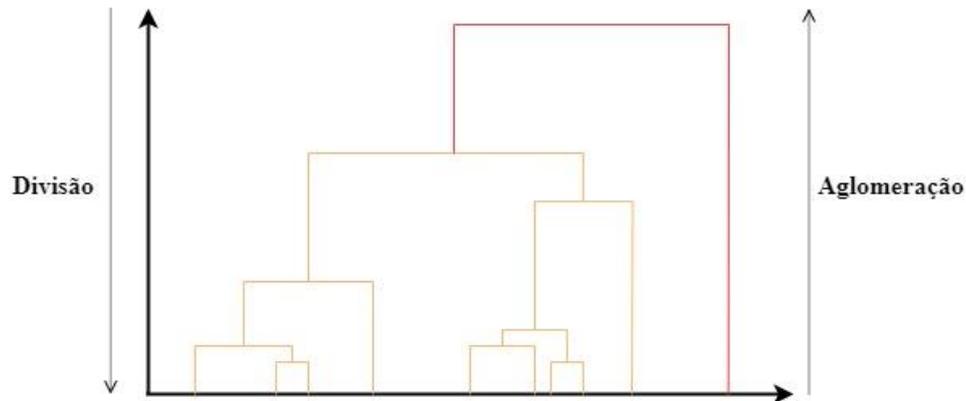


Figura 19 Exemplo de um dendrograma (baseado em (Nielsen, 2016))

### 3.2.5 Dados Sintéticos

Dados sintéticos são dados que são gerados artificialmente, em vez de obtidos por eventos reais. Estes dados normalmente são criados com ajuda de técnicas de aumento de dados ("*data augmentation*") e é usado numa grande variedade de processos, tal como validação de modelos e em processos de treino de modelos de Inteligência Artificial. Existem diversas técnicas de dados sintéticos, para o problema em causa foram selecionados cinco: *jittering*, *scaling*, *downsampling*, *cutting* e *permutation* (Lopez-Nava et al., 2020, Taewoong Um et al., 2017).

#### ***Jittering***

A técnica de *jittering* (Figura 20) é muito simples, basicamente, através de uma série temporal de uma variável, adiciona-se ruído gerado aleatoriamente por uma distribuição normal de magnitude *sigma* e média nula. De notar que para cada valor da série temporal é gerado um novo valor de ruído.

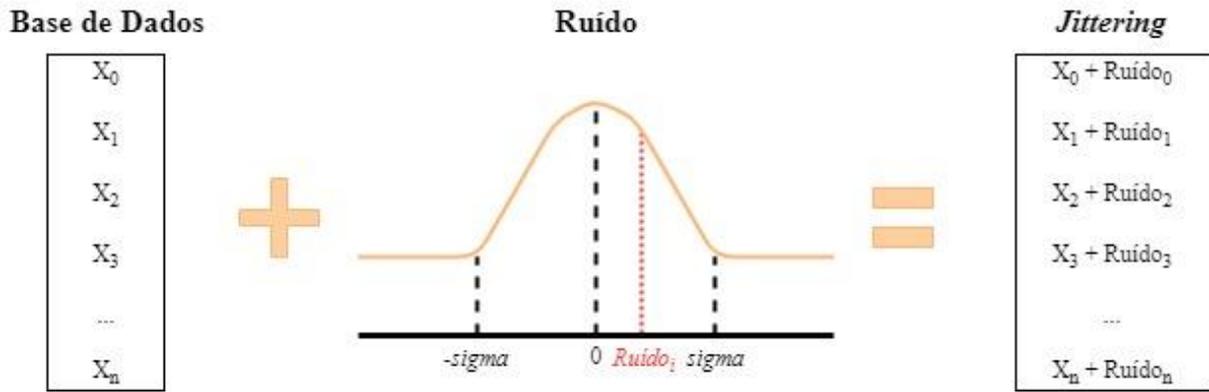


Figura 20 *Jittering*

### ***Scaling***

A técnica de *scaling* (Figura 21) é idêntica à de *jittering*, sendo que gera um ruído aleatório obtido por uma distribuição normal de magnitude sigma e média unitária, de seguida multiplica-se esse mesmo valor a toda a série temporal da variável.

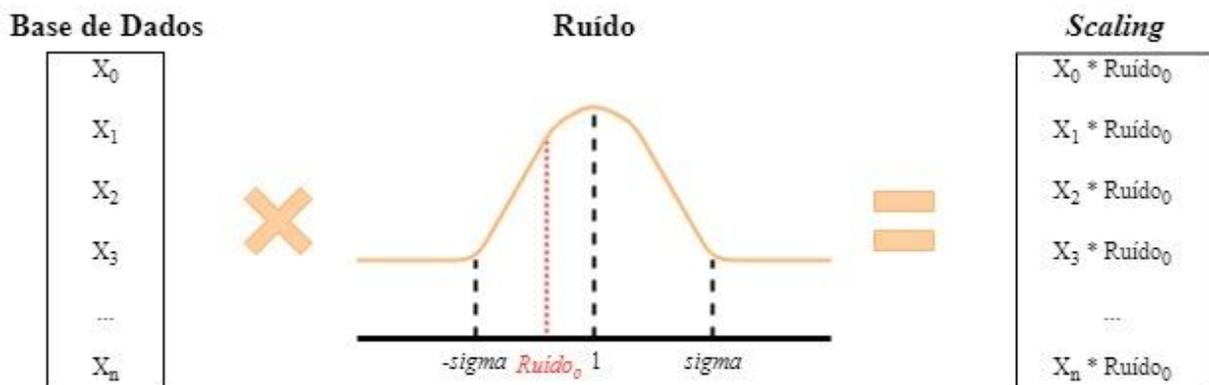


Figura 21 *Scaling*

### ***Downsampling***

*Downsampling* (Figura 22) é uma técnica de redução da série temporal, ou seja, esta técnica reduz o tamanho da série temporal subdividindo a série temporal em grupos mais pequenos e retornando a média desses valores tornando essa média o novo valor da série temporal. Esta técnica é muito usual quando se pretende remover/atenuar dados muito discrepantes à moda da série temporal.

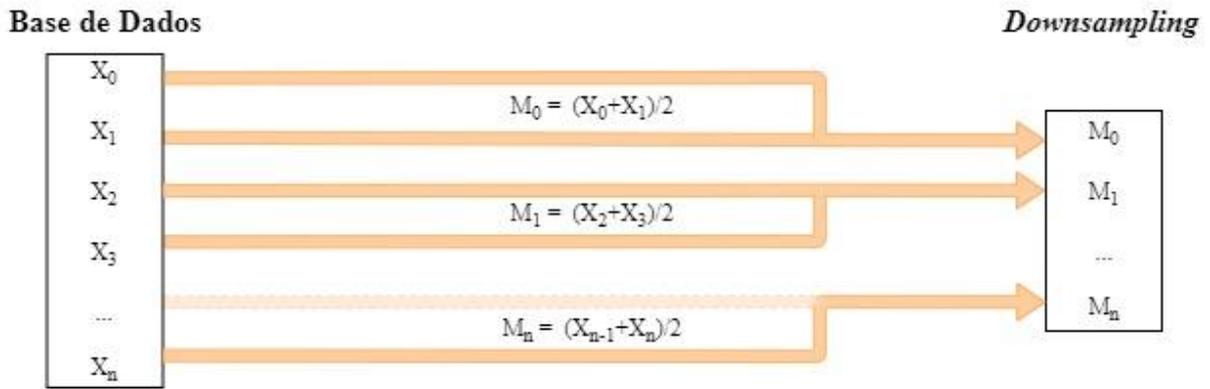


Figura 22 Downsampling

### Cutting

A técnica de *cutting* (Figura 23), também é uma técnica de redução do tamanho da série temporal do paciente, mas ao invés de técnica de *downsampling*, o *cutting* remove a mesma quantidade de dados ao início e ao fim da série temporal.

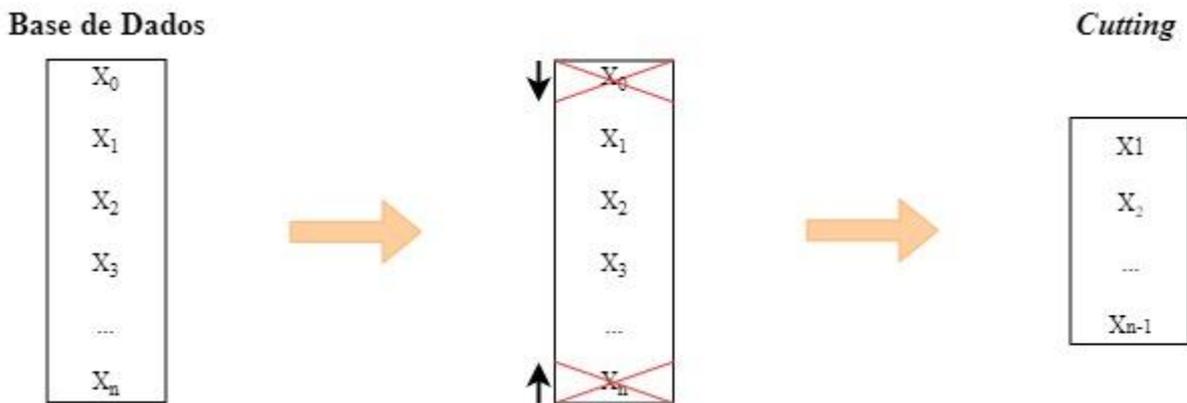
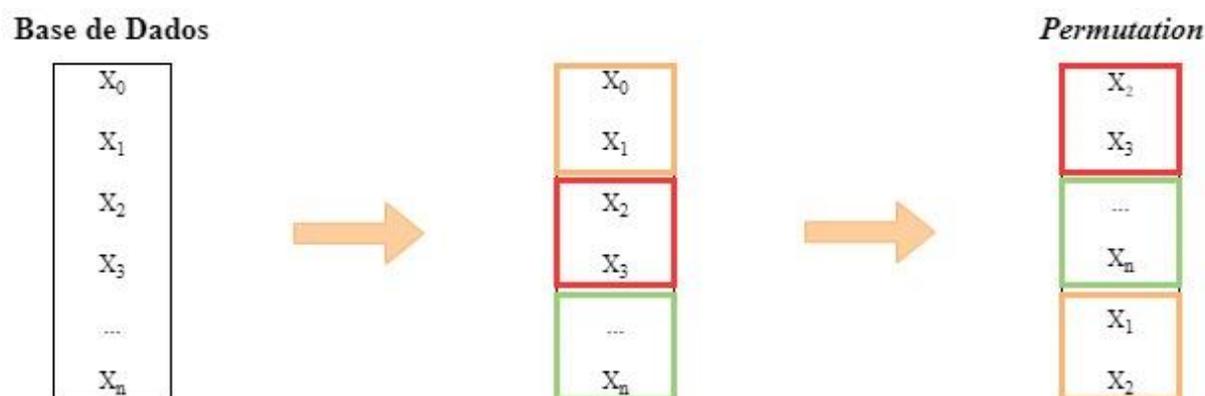


Figura 23 Cutting

### Permutation

A técnica de *permutation* (Figura 24) consiste em segmentar a série temporal e alterar a ordem desses segmentos, alterando a série temporal. De notar, que após aplicar esta técnica a média e o desvio padrão da série resultante é igual ao da série inicial. Assim, esta técnica não foi aplicada na geração de dados sintéticos para obtenção de novos valores de média e desvio padrão.

Figura 24 *Permutation*

### 3.2.6 Classificadores

Classificadores de *machine learning* são processos de previsão de classes de dados, com o intuito de aproximar uma função de mapeamento  $f$  de variáveis de entrada  $X$  para variáveis de saída discretas  $y$  (Géron, 2019, Gollapudi, 2016).

A classificação ("*classifiers*") enquadra-se na categoria de aprendizagem supervisionada ("*supervised learning*") (Bishop, 2006). Esta categoria pode ser subdivida em dois tipos de aprendizagem ("*learners*"): aprendizagem preguiçosa ("*lazy learners*") (Bontempi et al., 2002), (Homayouni et al., 2010) e aprendizagem ansiosa ("*eager learners*") (Niranjanamurthy et al., 2016).

Na aprendizagem preguiçosa no momento em que os dados de treino são rececionados a única operação que é realizada é o armazenamento destes na memória e só quando os dados de teste são rececionados é que é criado o modelo de previsão para prever a saída dos dados de teste. Isto faz com que o tempo de treino seja reduzido, mas o tempo de teste seja muito dispendioso.

Contrariamente na aprendizagem ansiosa no momento da receção dos dados de treino é criado logo o modelo de previsão através desses mesmos dados, criando hipóteses de saídas, possibilitando uma rápida previsão no momento da receção dos dados de teste, ou seja, mais tempo gasto no treino do que no teste. Este método é o mais comum pois normalmente o utilizador prefere gastar mais tempo no treino do seu algoritmo, mas depois no momento da execução do mesmo, quando for necessário fazer uma previsão este seja rápido e eficaz. Para o estudo em causa desta dissertação é mais benéfico métodos de aprendizagem ansiosa. Logo os algoritmos de classificação binários escolhidos foram: *support vector machine*, *random forest classifier* e *logistic regression*.

### 3.2.6.1 Support Vector Machine

*Support Vector Machine (SVM)* pode ser usado para classificação binária e regressão. O objetivo deste algoritmo de *machine learning* é encontrar o hiperplano ("*hyperplane*") num campo de  $N$  dimensões, sendo  $N$  o número de variáveis, que conseguia dividir os dados em classificações distintas (Bhargava et al., 2020, Kumar et al., 2020).

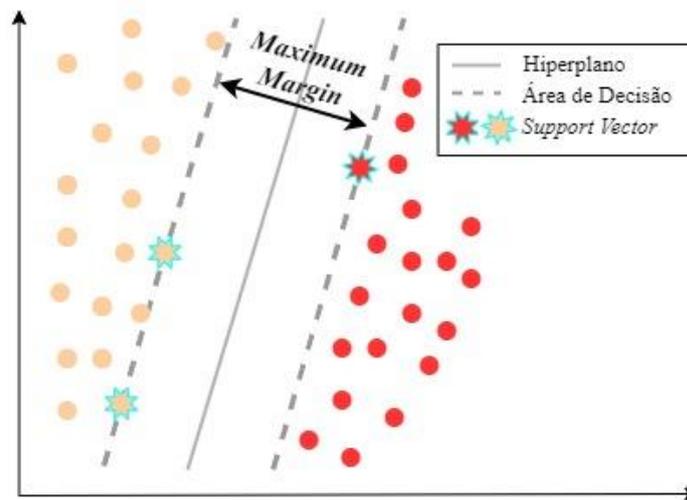


Figura 25 *Maximal Margin Classifier* (baseado em (Bhargava et al., 2020, Kumar et al., 2020))

Em simultâneo com o hiperplano é definido uma margem ("*Margin*"), quanto maior esta for mais precisas e fiáveis as previsões serão.

No entanto nem sempre é possível definir um hiperplano e uma margem tão intuitivamente como demonstrado na Figura 25, por vezes temos que permitir que existam classificações erradas ("*miss classifications*") para que seja possível traçar um hiperplano, tal como exemplificado na Figura 26.

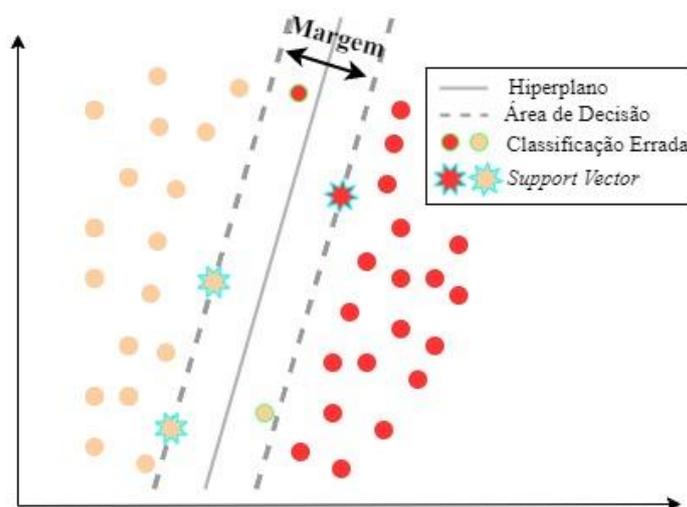


Figura 26 *Support Vector Machine Illustration* (baseado em (Bhargava et al., 2020, Kumar et al., 2020))

Contudo o algoritmo de *machine learning SVM* não é comum ser utilizado para exemplos simples como os ilustrados nas Figura 25 e Figura 26 mas sim exemplos em que não é possível linearmente definir-se um hiperplano na distribuição original dos dados. Por exemplo, se tivermos dados distribuídos num campo de 1 dimensão como representado na Figura 27, não existe qualquer possibilidade de se traçar um hiperplano que separe os dados.

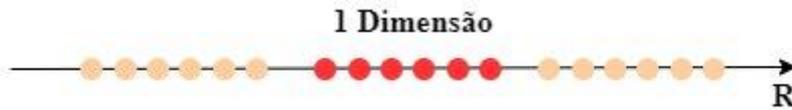


Figura 27 Exemplo a 1 dimensão (baseado em (Bhargava et al., 2020, Kumar et al., 2020))

Então o *SVM* através de técnicas de transformação de dados (*Polynomial Kernel*, *Radial Kernel*, *Radial Basis Function Kernel*, *Sigmoid Kernel* e *Linear Kernel*) aumenta a dimensão do campo dos dados de forma que seja possível traçar o hiperplano e a sua margem (Figura 28).

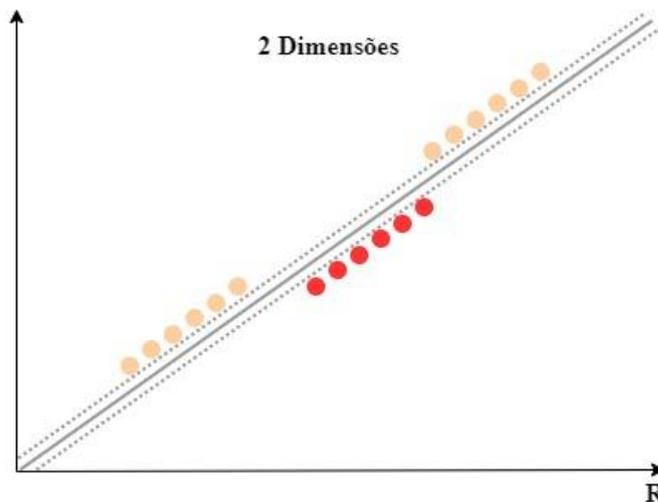


Figura 28 Exemplo a 2 dimensões (baseado em (Bhargava et al., 2020, Kumar et al., 2020))

### 3.2.6.2 Random Forest Classifier

*Random Forest* consiste numa combinação de várias árvores de decisão ("*Decision Tree*") para prever ou classificar o valor da variável de saída. Este método é capaz de suportar um grande número de variáveis e estimar a importância de cada uma delas na classificação sem necessidade de processos de eliminação de variáveis.

Com o intuito de aumentar a diversidade das árvores de decisão obtidas do conjunto de dados de treino, o classificador *random forest* utiliza técnicas de *bootstrap or bagging aggregation*. O método de *Bagging* cria novos conjuntos de dados através de repartições aleatórias da base de dados, sendo importante salientar que estes mesmos podem se repetir no mesmo conjunto de dados (Guo et al., 2011, Misra & Li, 2019).

Na Figura 29 está ilustrado o funcionamento geral do *random forest*, onde começa pela partição da base de dados em dois conjuntos, conjunto de treino e de teste, de seguida o conjunto de dados de treino sofre várias repartições aleatórias, *bootstrap aggregation*, desses subconjuntos de dados aleatórios é construída uma árvore de decisão para cada um que por sua vez cada um corresponde a um modelo. Por fim a previsão da saída é obtida pela aplicação da mesma entrada a todos os modelos e a saída mais obtida é considerada a saída final do modelo para essa entrada específica.

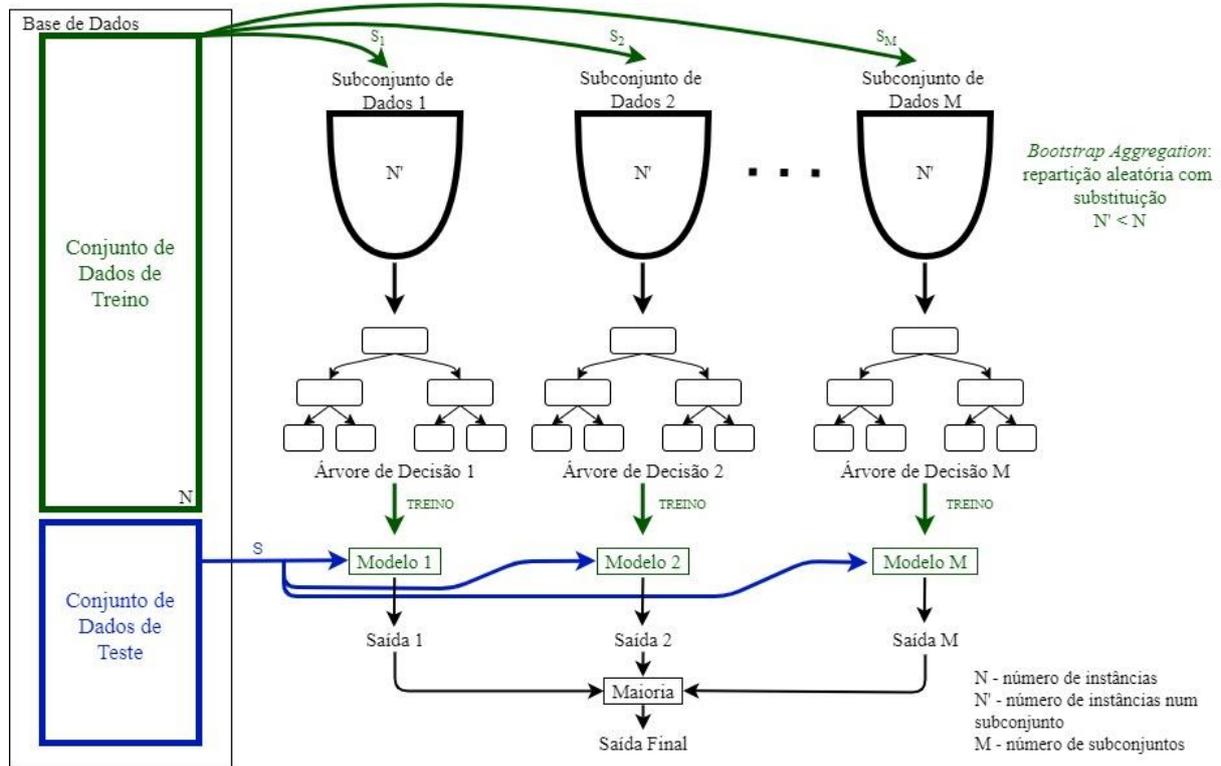


Figura 29 Funcionamento Geral do *Random Forest* (baseado em (Guo et al., 2011, Misra & Li, 2019))

O algoritmo também possui um método próprio de medição da performance do modelo através do *Out-of-Bag error* (Fuleky, 2019).

### 3.2.6.3 Logistic Regression

*Logistic Regression* (LR) é um método de *machine learning* de classificação binária. Este método assemelha-se a uma regressão linear, podendo ser descrito pela equação matemática:

$$y = \frac{e^{\beta_0 + \beta_1 * x}}{1 + e^{\beta_0 + \beta_1 * x}} \quad (3.5)$$

Em que os valores de entrada  $x$  são combinados linearmente com os valores dos pesos e dos coeficientes  $\beta$  para prever o valor de saída  $y$ .

Os coeficientes  $\beta$  do algoritmo de *logistic regression* são estimados através dos dados de treino. Estes são determinados pelo *maximum-likelihood estimation*.

*Maximum-likelihood estimation* é um algoritmo de aprendizagem comum utilizado por vários algoritmos de *machine learning*, embora faça suposições sobre a distribuição dos dados. Ou seja, este método é usado para otimizar os melhores valores dos coeficientes para os dados de treino.

Definidos os coeficientes iremos obter uma equação que graficamente se assemelhará a uma curva "S".

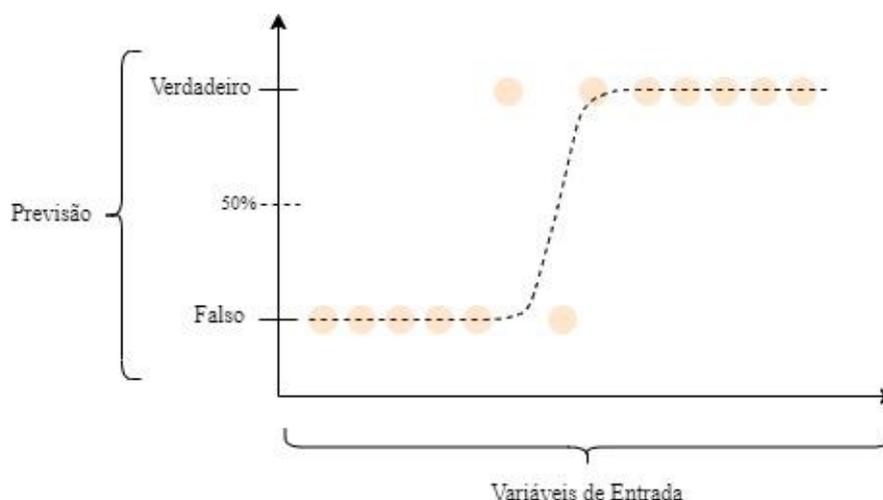


Figura 30 *Logistic Regression* (baseado em (Dreiseitl & Ohno-Machado, 2002, Kambria, 2019))

Por fim, ainda existe um método que permite avaliar se todas as variáveis de entrada têm importância na previsão. Se o efeito da variável na previsão for zero a variável não tem influência para a previsão da saída, ou seja, é desnecessária. Este método chama-se de *Wald's Test* (Dreiseitl & Ohno-Machado, 2002, Kambria, 2019).

#### 3.2.6.4 **Convolutional Neural Networks (CNN)**

*Convolution Neural Network (CNN)* é uma rede neuronal artificial (Albawi et al., 2017). Tal como o nome indica, *convolution*, significa uma função matemática derivada da integração de duas funções distintas. Esta arquitetura deriva das *Artificial Neural Networks (ANNs)* tradicionais e foi desenvolvida com o intuito de aumentar a performance do processo de inferência para dados de entrada de grandes dimensões. No entanto, existem outras vertentes em que o conceito pode ser aplicado, tais como:

- Dados de texto;
- Séries temporais de dados;
- Sequências de dados.

Tendo em vista o problema analisado nesta dissertação, o método que será aplicado são as séries temporais de entrada a uma dimensão. A arquitetura de uma *CNN* é simples (Srinivas et al., 2019),

sendo que pode ser dividida em três fases/estágios principais (*Convolution* com Função de Ativação, *Pooling* e *Fully-connected*), tal como é possível visualizar na Figura 31.

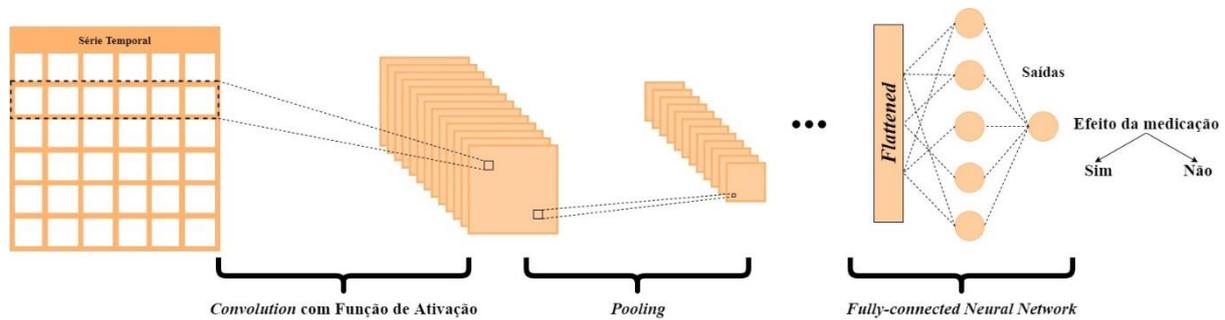


Figura 31 Arquitetura de uma *CNN* (baseado em (Srinivas et al., 2019))

Na camada convulsional é processada a operação de convolução dos dados de entrada. Isso é feito aplicando um filtro (ou *kernel*) sobre o *array* de entrada, calculando o valor da convolução e usando o resultado obtido como o valor de uma célula do *array* de saída, durante este processo o filtro é deslocado por uma janela deslizante predefinida por todo o *array* de entrada. Os parâmetros dos filtros são definidos durante o processo de treino à exceção da dimensão do filtro que é um hiper parâmetro definido previamente ao processo de treino. Por fim à saída da camada convulsional é lhe aplicada uma função não-linear, usualmente a função *ReLU*, que extrai o excesso de dados obtidos no *kernel*.

Semelhante à camada convulsional, a camada de *Pooling* é responsável por diminuir o tamanho do espaço ocupado pelas variáveis convulsionais. Existem duas técnicas de *Pooling*: *Max Pooling* e *Average Pooling*. No *max pooling* o algoritmo guarda o maior valor da matriz e no *average pooling* calcula a média da matriz e armazena essa mesma média (ver Figura 32). Contudo é preciso ter em consideração que alguns valores resultantes do *kernel* podem ser ruídos ou pouco significativos, e quando aplicado o método de *average pooling* esses valores têm peso na saída obtida, o mesmo não acontece com o *max pooling*, daí este método ser o mais usual e eficaz (Rawat & Wang, 2017). Como base de dados utilizada é muito sensível a ruído nesta dissertação também será aplicado o *Max Pooling* (Albawi et al., 2017, Lai et al., 2018).

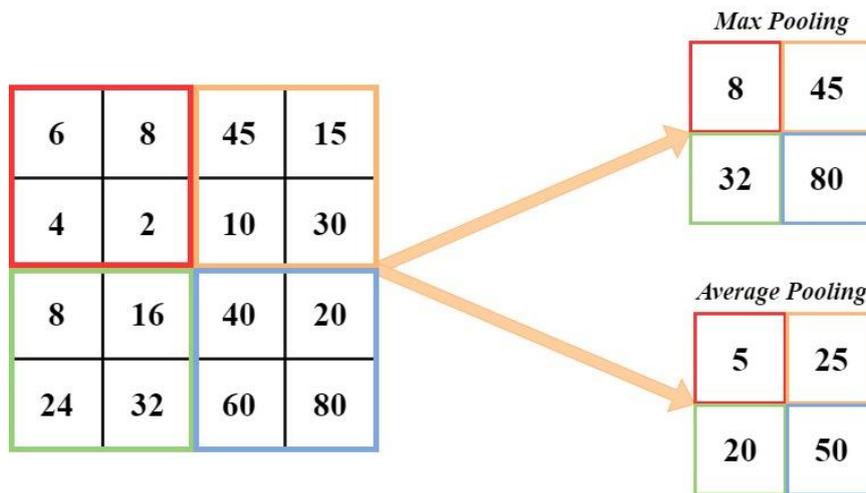


Figura 32 *Max e Average Pooling* (baseado em (Albawi et al., 2017, Lai et al., 2018))

O último estágio da rede é a *fully-connected*, que fornece a probabilidade de o paciente estar sob o efeito da medicação ou não. Uma vez mais, este estágio é naturalmente composto por uma *Artificial Neural Network (ANN)* com múltiplas camadas (mínimo 2, a de entrada e a de saída) (Gollapudi, 2016), (Gallo, 2015).

### 3.2.6.5 Long Short-Term Memory (LSTM)

Uma *Long Short-Term Memory* é um tipo especial de *recurrent neural network (RNN)*. As redes neuronais recorrentes apareceram como uma possível melhoria das redes neuronais tradicionais (*ANN*), isto porquê? Uma ANN inicia a sua iteração, finaliza-a e esquece o processo, ou seja, por outras palavras, não possui memória persistente. Esta lacuna em alguns modelos tornava-se algo bastante problemático e, para resolver esse problema, foram desenvolvidas as arquiteturas RNN, que são redes com *loops*, que permitem às informações persistirem (Elsaraiti & Merabet, 2021).

A LSTM é uma arquitetura de *recurrent neural network (RNN)* que “lembra” valores em intervalos arbitrários, sendo que a principal diferença de uma LSTM para uma RNN padrão é a sua insensibilidade ao comprimento do gap, ou seja, são capazes de aprender conexões de longo prazo.

Nesta estrutura a informação é retida pelas células e as manipulações de memória são feitas pelos portões (*gates*). Existem três portões (Hochreiter, 1997):

- A porta do esquecimento (*forget gate*), que decide quais partes do estado da célula (*cell state*) continuam importantes;
- a porta de entrada (*input gate*), que decide quais informações da memória a curto prazo devem ser adicionadas ao estado da célula;

- a porta de saída (*output gate*), que decide quais partes do estado da célula são importantes no instante atual para gerar a saída (*output*).

A memória de longo prazo é geralmente chamada de estado da célula. As setas em *loop* indicam a natureza recursiva da célula. Isso permite que as informações dos intervalos anteriores sejam armazenadas na célula *LSTM*. O estado da célula é modificado pela porta do esquecimento (*forget gate*) colocado abaixo do estado da célula e também ajustado pela porta de modulação de entrada. Da equação, o estado da célula anterior esquece, multiplica-se com a porta do esquecimento e adiciona novas informações através da saída das portas de entrada, ver Figura 33.

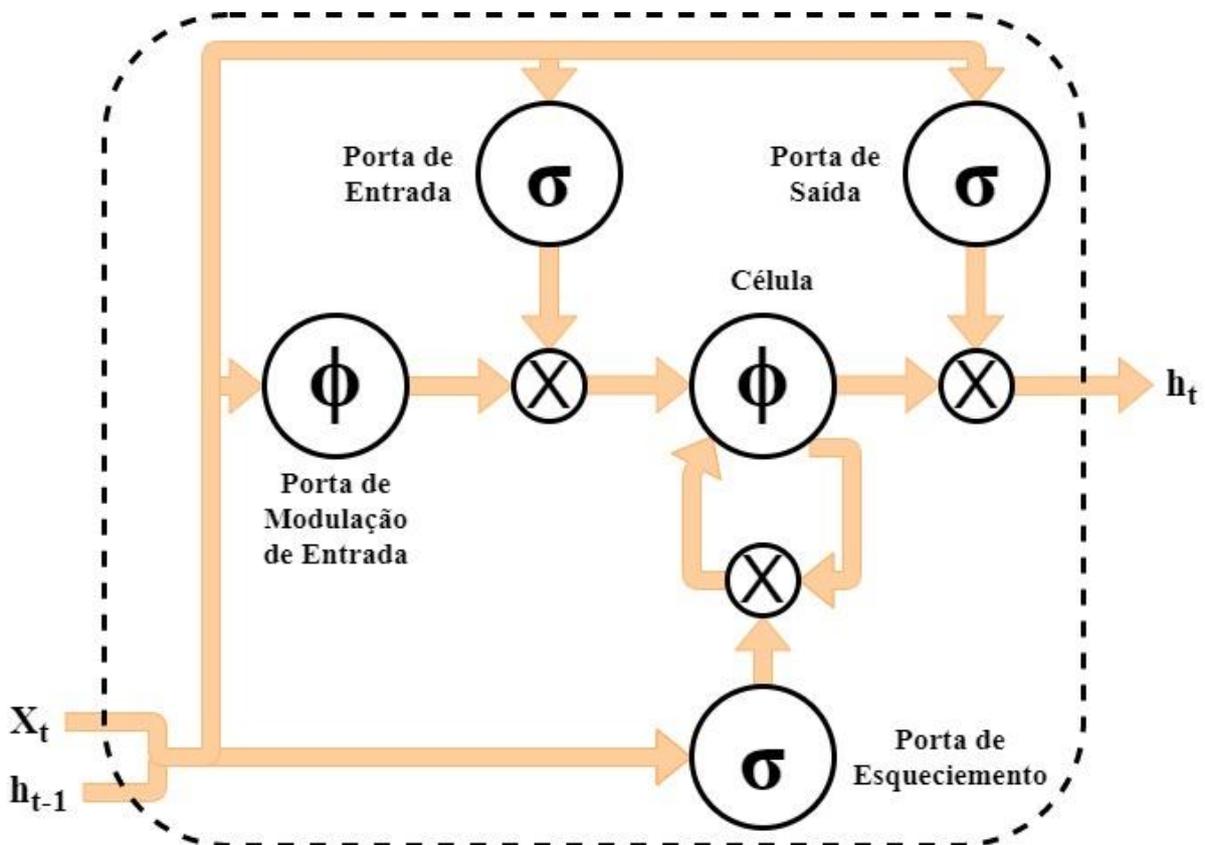


Figura 33 Célula *LSTM* (baseado em (Hochreiter, 1997))

### 3.2.7 Métricas para avaliar o desempenho do Classificador (“*Evaluate Classifier*”)

Para avaliar a performance dos modelos de classificação existem algumas métricas abaixo descritas que se baseiam nos valores da matriz de confusão (Tabela 2) (A., 2018), (Powers, 2020).

Tabela 2 Matriz de Confusão

		Previsão	
		Positivo	Negativo
Valor Real	Positivo	Verdadeiro Positivo (TP): Corresponde aos pacientes que foram corretamente previstos sob o efeito da medicação (estado <i>ON</i> )	Falso Negativo (FN): Corresponde aos pacientes que foram incorretamente previstos sem o efeito da medicação
	Negativo	Falso Positivo (FP): Corresponde aos pacientes que foram incorretamente previstos sob o efeito da medicação	Verdadeiro Negativo (TN): Corresponde aos pacientes que foram corretamente previstos sem o efeito da medicação (estado <i>OFF</i> )

Métricas:

**Taxa de acerto ("*Accuracy*")**: percentagem de previsões corretas

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (3.6)$$

**Sensibilidade ("*Recall*")** ou ("***Sensitivity***")

: percentagem de casos positivos corretamente previstos

$$Recall = \frac{TP}{TP + FN} \quad (3.7)$$

**Especificidade ("*Specificity*")**: percentagem dos casos negativos corretamente previstos

$$Specificity = \frac{TN}{TN + FP} \quad (3.8)$$

# Capítulo 4

## Resultados

Neste capítulo são apresentados e analisados os resultados obtidos da aplicação das diferentes técnicas de *machine learning* esquematizadas na Figura 4 apresentada no Capítulo 3. Inicialmente, é descrito o processo realizado para a seleção das variáveis de marcha mais relevantes. De seguida, são apresentados os resultados obtidos na análise de clusters com base em 4 diferentes conjuntos de dados. Para cada caso, os subgrupos encontrados com mais de 2 pacientes são usados para avaliar os modelos de classificação. Os dados sintéticos para treinar os modelos assim como os resultados obtidos na classificação são, por fim, apresentados juntamente com as suas limitações. Todas as técnicas de *machine learning* foram implementadas na linguagem de programação *python*, através de bibliotecas abertas, tais como, *Scikit-learn* (Kramer, 2016), *TensorFlow*, *Keras* (D. K., 2021, Keras, 2021), *Numpy* (NumPy, 2021), *Pandas* (Pandas, 2021), entre outras.

### 4.1 Seleção das variáveis de marcha mais relevantes

Tal como explicado na Secção 3.2.2, foram aplicados dois métodos estatísticos para seleção de variáveis, *ANOVA* e *Kendall*. Em ambos os métodos, *ANOVA* e *Kendall*, as variáveis mais importantes (com valor de estatística F superior a 1 e valor de coeficiente tau-b de *Kendall* superior a 0.5) para discriminar os pacientes em estado OFF (sem o efeito da medicação) versus em estado ON (sob o efeito da medicação) são: a variabilidade da duração de um ciclo, a variabilidade da cadência, a variabilidade do comprimento do ciclo, e a média do comprimento de um ciclo (Figura 34). Ou seja, podemos reduzir o leque inicial de variáveis a metade, excluindo a variabilidade e a média do tempo de duplo suporte, a média da cadência e a média da duração de um ciclo que são as variáveis estatisticamente menos relevantes para prever o estado do paciente.

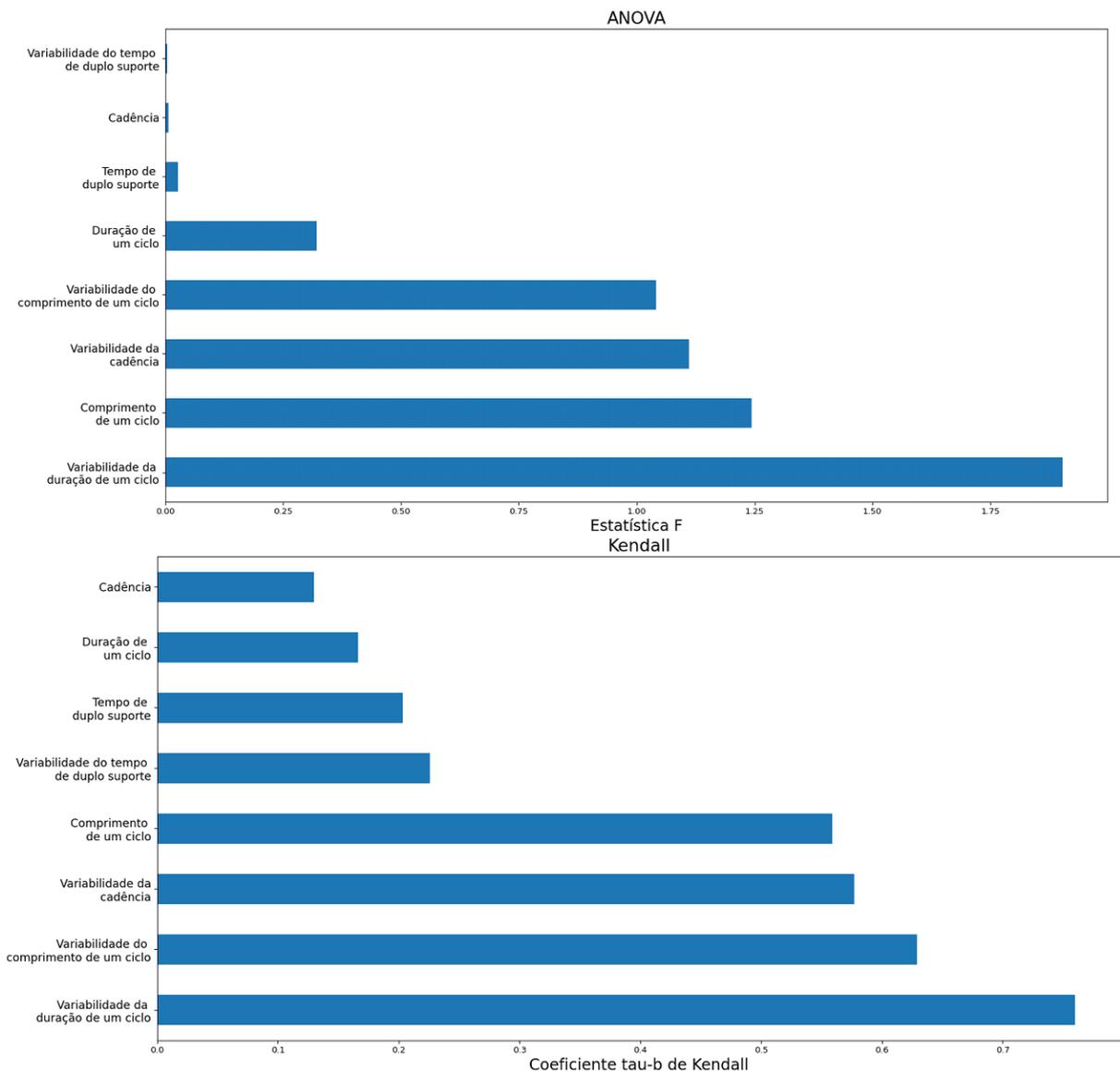


Figura 34 Importância das variáveis baseadas nos valores da estatística F na *ANOVA* e nos valores do coeficiente tau-b de *Kendall*

De notar que ambos os métodos são baseados na análise univariada, logo analisam o peso de cada variável de forma independente, o que pode originar que sejam selecionadas variáveis redundantes (variáveis que não trazem qualquer informação adicional). De forma a verificar quais as variáveis que estão fortemente correlacionadas e que conseqüentemente são redundantes determinou-se o coeficiente de correlação de *Pearson* entre as variáveis (Figura 35). Recordemos que duas variáveis estão fortemente correlacionadas quando o valor absoluto do coeficiente de correlação é igual ou superior a 0.6 ( $|r| \geq 0.6$ ).

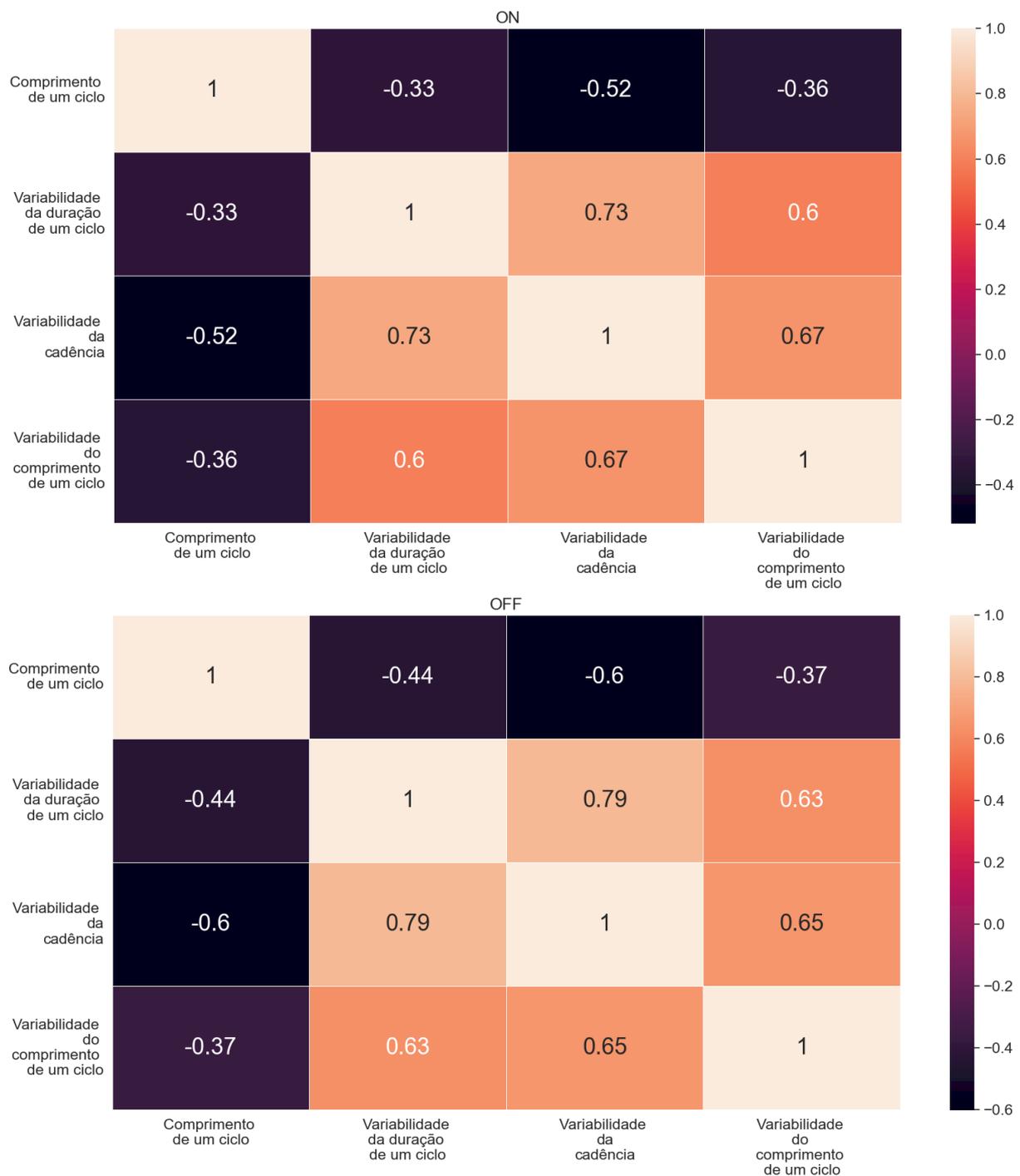


Figura 35 Matrizes de correlação de *Pearson* entre variáveis de marcha no estado *ON* e *OFF*

Através dos métodos *ANOVA* e *Kendall* a variabilidade da duração de um ciclo destacou-se como sendo a variável com maior valor de estatística *F*, ou seja, a mais importante. Uma vez que as variáveis da variabilidade da cadência e variabilidade do comprimento de um ciclo estão fortemente correlacionadas com a variabilidade da duração de um ciclo em ambos os estados (Figura 35), estas duas variáveis foram descartadas.

Assim as variáveis não fortemente correlacionadas mais relevantes para classificar o estado do paciente (*OFF* vs. *ON* medicação) são:

- variabilidade da duração de um ciclo (GCT\_SD);
- o comprimento de um ciclo (SL).

## 4.2 Análise de clusters

Tendo em consideração que a monitorização do estado de um paciente (*OFF* vs. *ON* medicação) faz sentido se o paciente apresentar uma melhoria significativa em resposta à medicação foram selecionados um subconjunto de pacientes da amostra inicial que apresentaram uma melhoria significativa na variabilidade da duração de um ciclo (variável que revelou ser mais relevante na discriminação dos dois estados). À semelhança do que é sugerido na literatura (Agostini et al., 2015), (Virhammar et al., 2012), é considerado como uma melhoria significativa na variabilidade da duração de um ciclo se o valor da variabilidade diminui do estado *OFF* para o estado *ON*, em pelo menos 10%. De acordo com este critério, foram selecionados 14 pacientes dos 29 pacientes considerados inicialmente.

### 4.2.1 Com base na variabilidade da duração de um ciclo (GCT\_SD)

Dois métodos de *Clustering*, Método de *K-means* (Secção 3.2.4.1) e o Método *Hierarchical* (Secção 3.2.4.2), foram aplicados com o objetivo de agrupar os pacientes por grau de similaridade com base nos valores da variabilidade da duração de um ciclo nos dois estados (*OFF* e *ON* medicação).

No método de *Hierarchical* foi aplicado a abordagem algomerativa, como explicado na Secção 3.2.4.2, tendo obtido o resultado apresentado na Figura 36. Por observação do dendrograma foram selecionados 4 clusters. O grupo 1, a verde, é constituído por um único paciente, o grupo 2, a vermelho, por quatro pacientes, o grupo 3, a roxo, por dois pacientes e por fim, o grupo 4, a azul, constituído por sete pacientes.

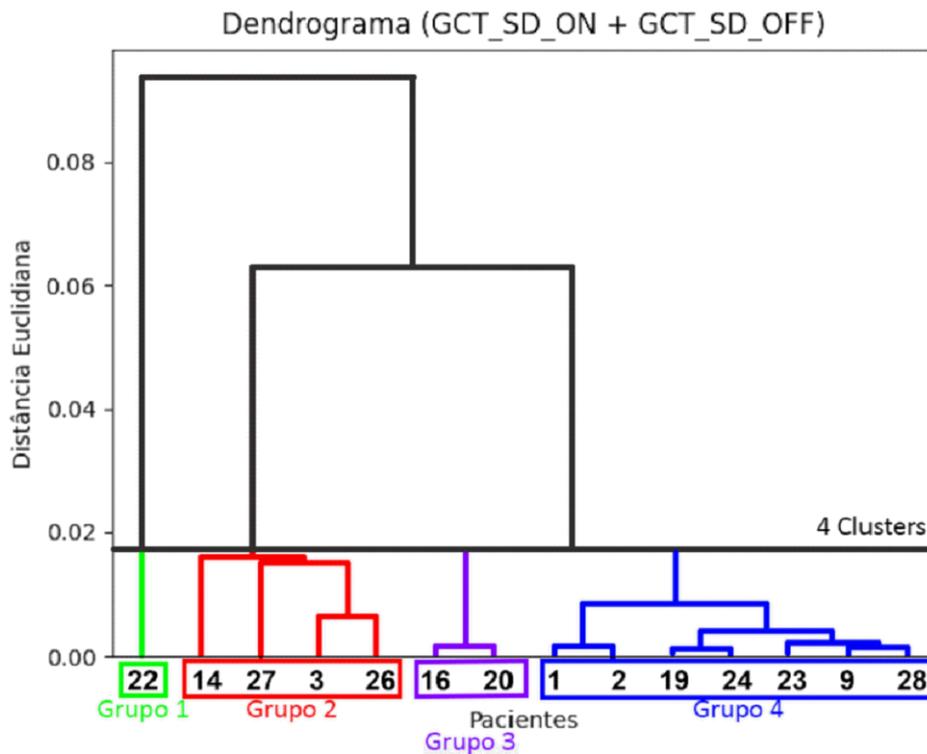


Figura 36 Dendrograma do agrupamento obtido pelo método *Hierarchical* para 14 pacientes com base na variabilidade da duração de um ciclo em ambos os estados *OFF* e *ON* (GCT\_SD\_OFF+ GCT\_SD\_ON)

Para a aplicação do método de *K-means* é necessário definir previamente o número de *clusters* pretendidos. Para tal é comum a utilização do método do cotovelo (Secção 3.2.4.1). Através deste método, neste caso, é sugerido 3 *clusters* (Figura 37). No entanto, de forma a comparar os resultados obtidos pelo *K-means* com os obtidos usando o *Hierarchical* usou-se  $k=4$  como o valor do número de *clusters*, resultando no agrupamento ilustrado na Figura 38.

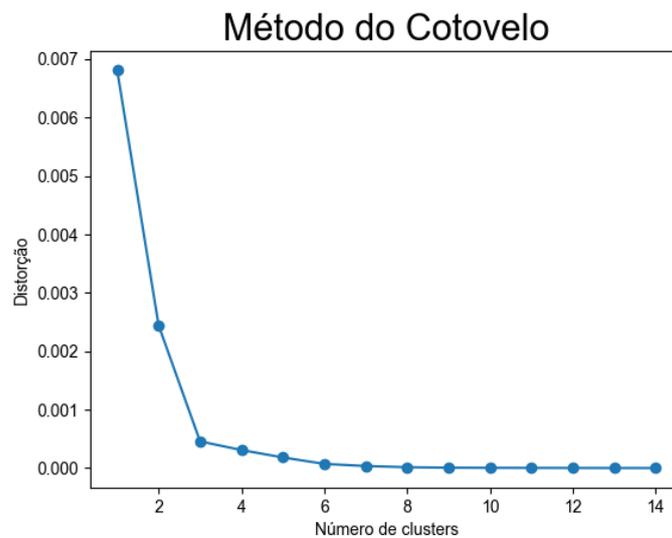


Figura 37 Resultado da aplicação do método do cotovelo

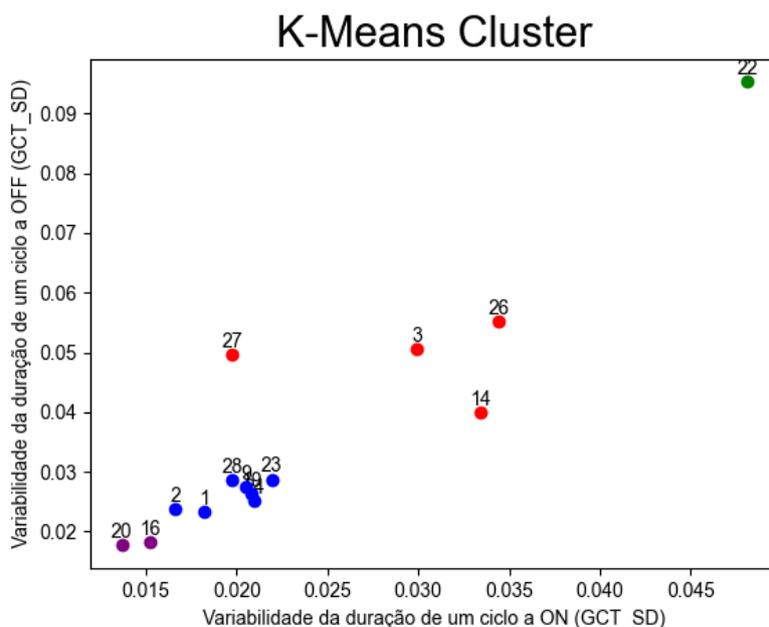


Figura 38 Resultado do agrupamento obtido pelo método de *K-means* para 14 pacientes com base na variabilidade da duração de um ciclo em ambos os estados *OFF* e *ON* (GCT\_SD\_OFF+ GCT\_SD\_ON)

Os 4 clusters obtidos pelo método *K-means* foram exatamente os mesmos que se obteve usando o método de *Hierarchical* (Figura 36). De salientar que o grupo 1 e 3 como não têm mais de 2 pacientes.

Na Figura 39 podemos visualizar os valores da variabilidade da duração de um ciclo (GCT\_SD) nos estados *OFF* e *ON* nos grupos 2 e 4, respetivamente. Podemos confirmar que em ambos os grupos existe uma clara separação dos valores de GCT\_SD dos pacientes a *ON* e *OFF*. Podemos notar também que o grupo 2 apresenta maiores valores de GCT\_SD em ambos os estados em relação ao grupo 4.

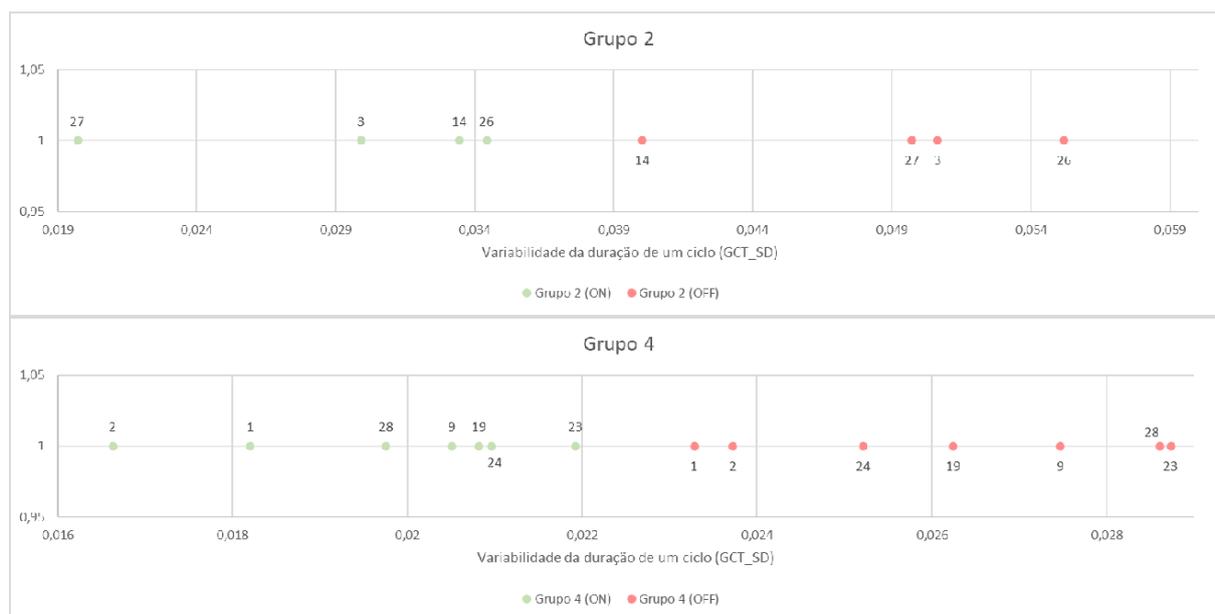


Figura 39 Ilustração dos valores da variabilidade da duração de um ciclo (GCT\_SD) nos estados *OFF* e *ON* nos grupos 2 e 4 obtidos através da análise de clusters

### 4.2.2 Com base no comprimento de um ciclo (série temporal)

Como também é pretendido neste estudo testar a classificação dos estados *OFF* e *ON* com base nas séries temporais das variáveis da marcha, a análise de clusters foi também aplicada ao mesmo subgrupo de 14 pacientes tendo em conta a série temporal do comprimento de um ciclo. Foi selecionado o comprimento de um ciclo, uma vez que, a média do comprimento de um ciclo e a variabilidade (desvio padrão) do comprimento de um ciclo são 2 das 4 variáveis da marcha que se revelaram mais relevantes na discriminação entre os estados *OFF* e *ON* (Secção 4.1). De notar que numa série temporal temos de forma implícita a informação sobre o valor da média e do desvio padrão da variável. Apesar de a variabilidade da duração de um ciclo (correlacionada com a variabilidade do comprimento de um ciclo) ter sido a variável que se revelou mais importante na predição do estado (*OFF* vs. *ON*), o valor da média da duração de um ciclo foi apontado como menos relevante.

Da aplicação do método *Hierarchical*, com base no cálculo das distâncias *dynamic time warping* (DTW) e seguindo uma abordagem aglomerativa obteve-se como resultado o agrupamento apresentado no dendrograma da Figura 40. Por observação do dendrograma foram selecionados 7 clusters em que apenas o grupo 7 tem mais de que dois pacientes.

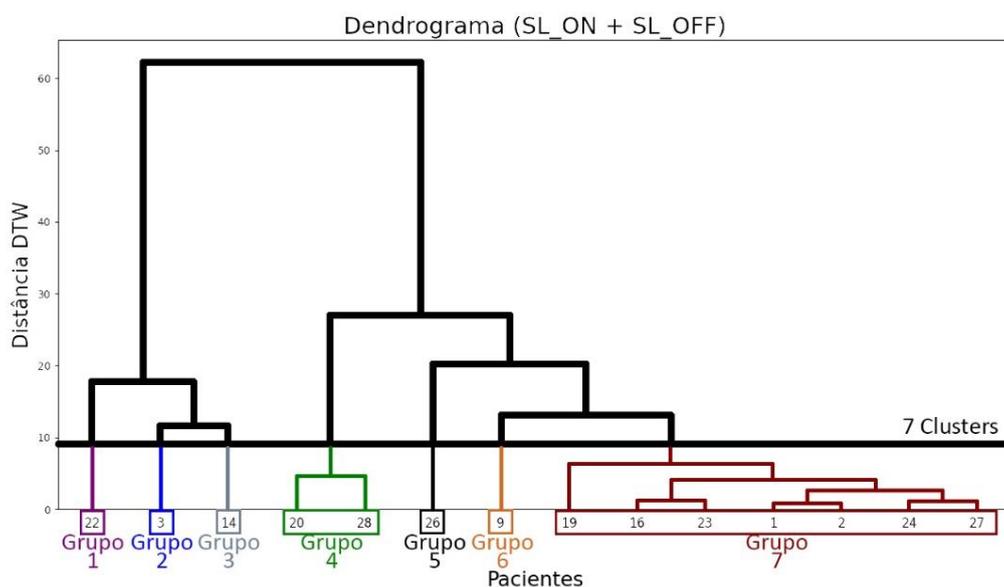


Figura 40 Dendrograma do agrupamento obtido pelo método *Hierarchical* para 14 pacientes com base no comprimento de um ciclo (série temporal) em ambos os estados *OFF* e *ON* (SL\_ON + SL\_OFF)

Escolhendo  $k=7$  como valor do número de clusters, os mesmos resultados foram obtidos a partir do método *K-means*. Na Figura 41 estão representadas as séries temporais do comprimento de um ciclo nos estados *OFF* e *ON* dos pacientes agrupados em cada um dos sete grupos. É possível observar que as séries temporais dos pacientes dos grupos 4 e 7 se sobrepõem dentro de cada grupo. Pode-se notar também, que o intervalo de valores do comprimento de um ciclo difere entre grupos.

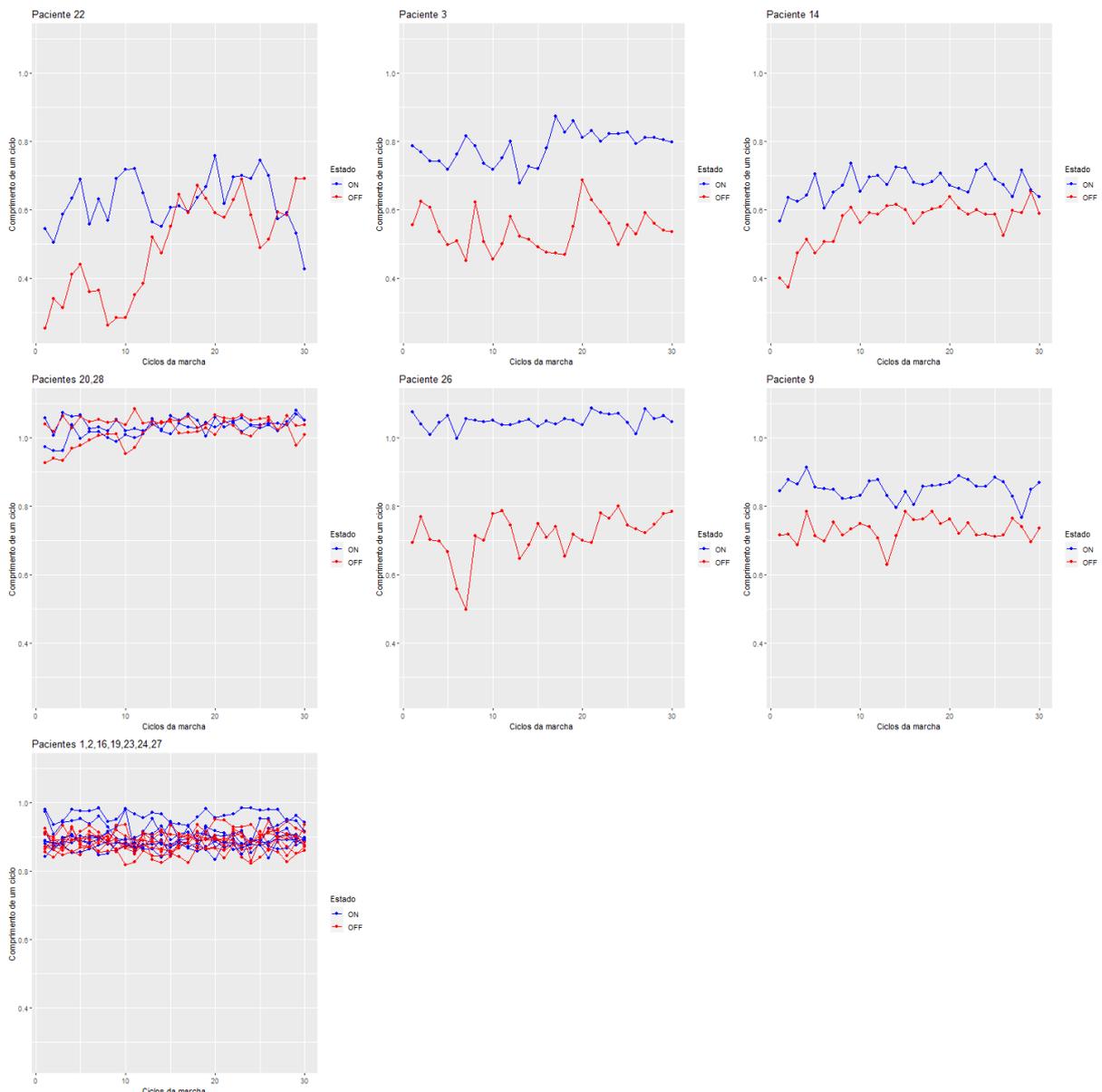


Figura 41 Séries temporais do comprimento de um ciclo nos estados *OFF* e *ON* dos pacientes agrupados em cada um dos sete grupos ilustrados na Figura 40.

### 4.3 Dados sintéticos

Tendo como base as séries temporais da duração de um ciclo e do comprimento de um ciclo em ambos os estados *OFF* e *ON* foram gerados dados sintéticos através da aplicação das técnicas descritas na Secção 3.2.5. Primeiramente, para cada paciente dos grupos 2 e 4 agrupados na análise de clusters com base na variabilidade da duração de um ciclo (Figura 36) foram geradas novas séries temporais relativas à duração de um ciclo e comprimento de um ciclo através da aplicação das técnicas *jittering*, *scaling*, *downsampling* e *cutting*. Concretamente, para cada paciente foram geradas 30 séries temporais da duração de um ciclo em estado *OFF*, 30 séries temporais da duração de um ciclo em estado *ON*, 30 séries temporais do comprimento de um ciclo em estado *OFF*, e 30 séries temporais comprimento de

um ciclo em estado *ON*, com base na respetiva série temporal desse paciente. De seguida, novos valores de variabilidade (desvio padrão) da duração de um ciclo foram calculados para cada série temporal da duração de um ciclo gerada e, também, foram determinados novos valores médios do comprimento de um ciclo calculados para cada série temporal do comprimento de um ciclo. Por exemplo, na Figura 42 podemos visualizar os valores sintéticos da variabilidade da duração de um ciclo gerados a partir das séries temporais da duração de um ciclo nos estados *OFF* e *ON* do paciente 3. Em anexo, Figura 47, poderão ser visualizados os valores sintéticos da variabilidade da duração de um ciclo associados aos restantes pacientes. Em todos os casos podemos visualizar uma clara distinção dos valores em estado *OFF* e em estado *ON*.

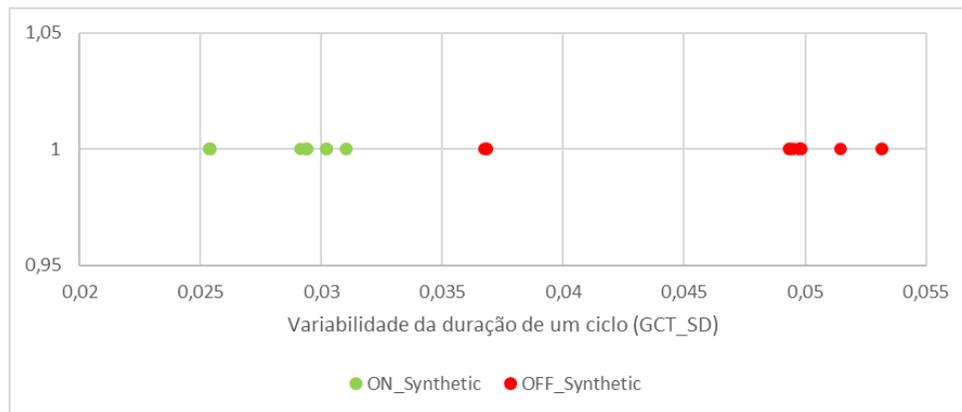


Figura 42 Valores sintéticos da variabilidade da duração de um ciclo associados ao paciente 3

Observando os valores sintéticos da variabilidade da duração de um ciclo e do comprimento de um ciclo em conjunto num gráfico 2D em que nas coordenadas  $x$  e  $y$  estão representados os valores de  $GCT\_SD$  e  $SL$  respetivamente (Figura 43 associados ao paciente 3, e em anexo Figura 48 associados aos restantes pacientes), podemos constatar que existe uma clara separação entre os pontos referentes ao estado *OFF* (pontos a vermelho) e os pontos referentes aos valores do estado *ON* (pontos a verde).

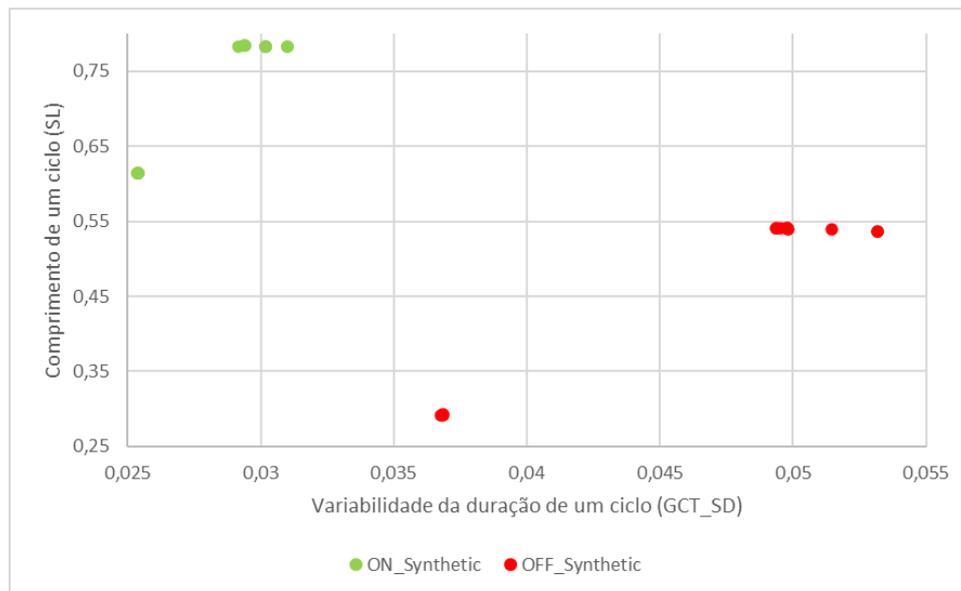


Figura 43 Valores sintéticos da variabilidade da duração de um ciclo e da média do comprimento de um ciclo associados ao paciente 3 (GCT\_SD+SL)

Numa segunda fase, foram gerados novas séries temporais relativas ao comprimento de um ciclo associadas aos pacientes do grupo 7 representado na Figura 40 através da aplicação das técnicas *jittering*, *scaling*, *downsampling*, *cutting* e adicionalmente a técnica *permutation*. Por exemplo, na Figura 44 podemos observar as séries temporais sintéticas do comprimento de um ciclo em estado *OFF* (a vermelho) e em estado *ON* (a azul) obtidas com base nas séries temporais do comprimento de um ciclo do paciente 1. Na Figura 49 em anexo, poderão ser visualizados as séries temporais do comprimento de um ciclo associados aos restantes pacientes. Pode-se notar que as séries temporais obtidas em estado *OFF* e *ON* em muitos dos casos, como no exemplo ilustrado na Figura 44, se sobrepõem, não havendo uma clara divisão nos valores das séries temporais em *OFF* e *ON*.

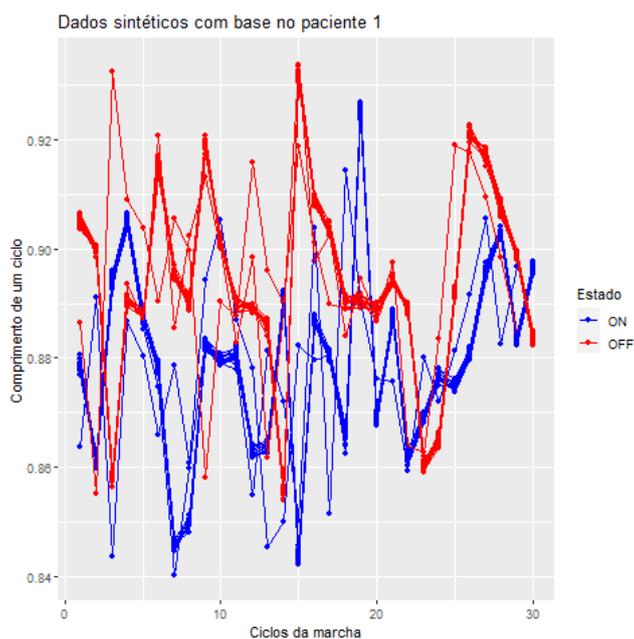


Figura 44 Séries temporais sintéticas do comprimento de um ciclo associada ao paciente 1

## 4.4 Classificação

Neste subcapítulo são apresentados os resultados obtidos pelos diferentes algoritmos de classificação com base em diferentes conjuntos de dados. De salientar que, em todos os casos, o treino e validação dos algoritmos de classificação foram realizados com um conjunto de dados sintéticos associados a um só paciente e o corresponde teste com os dados reais de um conjunto de pacientes diferentes ao usado para gerar os dados sintéticos. A performance dos algoritmos de classificação é avaliada segundo as métricas taxa de acerto, sensibilidade e especificidade (Secção 3.2.7).

### 4.4.1 Com base na variabilidade da duração de um ciclo (GCT\_SD)

O grupo 2 (ver Figura 36) é constituído por quatro pacientes (3, 14, 26 e 27), logo foram realizadas quatro modelações diferentes. Para cada base de dados sintética gerada foram aplicados três classificadores diferentes (*Support Vector Machine*, *Random Forest* e *Logistic Regression*), podendo os resultados obtidos serem consultados na Tabela 3. Com a exceção do caso em que os dados sintéticos foram gerados a partir do paciente 27, a taxa de acerto de teste do modelo foi sempre superior a 83,3%. Contudo, é importante entender que o teste efetuado neste grupo 2 só era constituído por três pacientes (três em estado *OFF* e três em estado *ON*), ou seja, cada previsão errada pelo modelo tinha um impacto de cerca de 16,7% na taxa de acerto de teste.

Tabela 3 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 2 com base na *GCT\_SD*

			Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 2 (3,14,26,27)	Dados Sintéticos: 3 Teste: 14,26,27	SVM	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	67%
		RF	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	67%
		LR	95%	90%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 14 Teste: 3,26,27	SVM	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	67%
		RF	100%	96.67%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	67%
		LR	96.67%	96.67%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 26 Teste: 3,14,27	SVM	100%	100%	Taxa de acerto	<b>83.33%</b>
					Sensibilidade	<b>67%</b>
					Especificidade	<b>100%</b>
		RF	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	67%
					Especificidade	100%
LR		100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>67%</b>	
				Especificidade	<b>100%</b>	
Dados Sintéticos: 27 Teste: 3,14,26	SVM	100%	100%	Taxa de acerto	66.67%	
				Sensibilidade	100%	
				Especificidade	33%	
	RF	100%	100%	Taxa de acerto	50%	
				Sensibilidade	100%	
				Especificidade	0%	
	LR	100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>100%</b>	
				Especificidade	<b>67%</b>	

Os resultados do grupo 4 (Tabela 4) foram igualmente promissores. De destacar os resultados obtidos pelo algoritmo de classificação *Logistic Regression* em que o valor da taxa de acerto de teste foi de 100% em 4 dos 7 modelos testados.

Tabela 4 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 4 com base na *GCT\_SD*

			Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 4 (1,2,9,19,23,24,28)	Dados Sintéticos: 1 Teste: 2,9,19,23,24,28	SVM	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
		RF	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
		LR	100%	100%	Taxa de acerto	<b>66.67%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>33%</b>

	Dados Sintéticos: 2 Teste: 1,9,19,23,24,28	SVM	100%	100%	Taxa de acerto	<b>58.33%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>17%</b>
		RF	100%	100%	Taxa de acerto	<b>58.33%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>17%</b>
		LR	100%	100%	Taxa de acerto	<b>58.33%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>17%</b>
	Dados Sintéticos: 9 Teste: 1,2,19,23,24,28	SVM	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		RF	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		LR	98.33%	93.33%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 19 Teste: 1,2,9,23,24,28	SVM	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		RF	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		LR	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
Dados Sintéticos: 23 Teste: 1,2,9,19,24,28	SVM	100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>67%</b>	
				Especificidade	<b>100%</b>	
	RF	100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>67%</b>	
				Especificidade	<b>100%</b>	
	LR	100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>67%</b>	
				Especificidade	<b>100%</b>	
Dados Sintéticos: 24 Teste: 1,2,9,19,23,28	SVM	100%	100%	Taxa de acerto	91.67%	
				Sensibilidade	100%	
				Especificidade	83%	
	RF	100%	100%	Taxa de acerto	91.67%	
				Sensibilidade	100%	
				Especificidade	83%	
	LR	100%	100%	Taxa de acerto	<b>100%</b>	
				Sensibilidade	<b>100%</b>	
				Especificidade	<b>100%</b>	
Dados Sintéticos: 28 Teste: 1,2,9,19,23,24	SVM	93.33%	100%	Taxa de acerto	50%	
				Sensibilidade	50%	
				Especificidade	50%	
	RF	96.67%	93.33%	Taxa de acerto	91.67%	
				Sensibilidade	100%	
				Especificidade	83%	
	LR	95%	93.33%	Taxa de acerto	<b>100%</b>	
				Sensibilidade	<b>100%</b>	
				Especificidade	<b>100%</b>	

Na Tabela 5 é apresentado os valores médios das métricas de desempenho obtidos em cada grupo pelos três algoritmos de classificação avaliados, em que o *Logistic Regression* apresentou melhores resultados com uma taxa de acerto média superior a 86,9%. Contudo os três classificadores apresentam

bons resultados sendo que o classificador com a pior taxa de acerto é o *Random Forest* com 75,0%. Um aspeto interessante é que o valor médio da sensibilidade foi sempre superior ou igual ao valor médio da especificidade, o que significa que o número de vezes em que um paciente sob o efeito da medicação foi corretamente classificado (estado *ON*) foi superior ou igual ao número de vezes em que um paciente sem o efeito da medicação foi corretamente classificado (estado *OFF*). Enquanto o valor médio da sensibilidade foi em todos os casos superior a 88,1% o valor médio da especificidade varia entre 58,5% e 91,8%.

Tabela 5 Valores médios das métricas de desempenho dos algoritmos de classificação para o grupo 2 e 4 com base na *GCT\_SD*

		Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 2 (3,14,26,27)	<i>SVM</i>	100.00%	100.00%	Taxa de acerto	79.17%
				Sensibilidade	91.75%
				Especificidade	66.75%
	<i>RF</i>	100.00%	99.17%	Taxa de acerto	75.00%
				Sensibilidade	91.75%
				Especificidade	58.50%
	<i>LR</i>	97.92%	96.67%	Taxa de acerto	<b>91.67%</b>
				Sensibilidade	<b>91.75%</b>
				Especificidade	<b>91.75%</b>
Grupo 4 (1,2,9,19,23,24,28)	<i>SVM</i>	99.05%	100.00%	Taxa de acerto	77.38%
				Sensibilidade	88.14%
				Especificidade	66.71%
	<i>RF</i>	99.52%	99.05%	Taxa de acerto	83.33%
				Sensibilidade	95.29%
				Especificidade	71.43%
	<i>LR</i>	99.05%	98.09%	Taxa de acerto	<b>86.90%</b>
				Sensibilidade	<b>95.29%</b>
				Especificidade	<b>78.57%</b>

#### 4.4.2 Com base na variabilidade da duração de um ciclo (*GCT\_SD*) e comprimento de um ciclo (*SL*)

Neste subcapítulo os 3 classificadores (*SVM*, *RF* e *LR*) serão testados novamente com base na variabilidade da duração de um ciclo juntamente com o comprimento de um ciclo (*GCT\_SD + SL*) para analisar se ao adicionar informação sobre o comprimento de um ciclo (*SL*) irá melhorar a performance dos modelos gerados.

O classificador *Logistic Regression* apresentou a melhor performance com uma taxa de acerto de 100% quando os dados sintéticos tiveram como base os pacientes 3 e 14, e uma taxa de acerto de 83,3% quando os dados sintéticos tiveram como base os pacientes 27 (Tabela 6). De notar, que estes mesmos valores de taxa de acerto foram obtidos com base apenas na variabilidade da duração de um ciclo (Tabela 3). Com os dados sintéticos obtidos com base no paciente 26, o melhor resultado foi obtido pelo

classificador *Support Vector Machine* com uma taxa de acerto de 100% (Tabela 6), sendo este valor superior aos valores de taxa de acerto obtidos com base apenas na variabilidade da duração de um ciclo (Tabela 3).

Tabela 6 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 2 com base na *GCT\_SD* e *SL*

			Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 2 (3,14,26,27)	Dados Sintéticos: 3 Teste: 14,26,27	<i>SVM</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	66.67%
					Especificidade	100%
		<i>RF</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	67%
		<i>LR</i>	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 14 Teste: 3,26,27	<i>SVM</i>	100%	100%	Taxa de acerto	50%
					Sensibilidade	100%
					Especificidade	0%
		<i>RF</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	66.67%
					Especificidade	100%
		<i>LR</i>	96.67%	96.67%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 26 Teste: 3,14,27	<i>SVM</i>	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		<i>RF</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	66.67%
					Especificidade	100%
<i>LR</i>		100%	100%	Taxa de acerto	83.33%	
				Sensibilidade	66.67%	
				Especificidade	100%	
Dados Sintéticos: 27 Teste: 3,14,26	<i>SVM</i>	100%	100%	Taxa de acerto	50%	
				Sensibilidade	100%	
				Especificidade	0%	
	<i>RF</i>	100%	100%	Taxa de acerto	66.67%	
				Sensibilidade	100%	
				Especificidade	33.33%	
	<i>LR</i>	100%	100%	Taxa de acerto	<b>83.33%</b>	
				Sensibilidade	<b>100%</b>	
				Especificidade	<b>66.67%</b>	

Na Tabela 7 estão descritos os resultados da aplicação dos três classificadores nos pacientes do grupo 4. Analisados os resultados de uma maneira geral é possível verificar que foram promissores, no entanto, e fazendo uma análise comparativa com os resultados obtidos com a utilização exclusiva da variabilidade da duração de um ciclo, *GCT\_SD*, (ver Tabela 4), observa-se que os resultados pioram, por exemplo, no paciente 9, em todos os classificadores (*RF*, *SVM* e *LR*) a taxa de acerto piorou com a adição do comprimento de um ciclo, *SL*, aos modelos. O mesmo aconteceu com o paciente 24 e 28 nos classificadores *RF/LR* e *RF/SVM* respetivamente, sendo que, a única situação que o acréscimo do *SL* apresentou melhorias foi no paciente 2 no classificador *SVM*.

Tabela 7 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 4 com base na *GCT\_SD* e *SL*

		Taxa de acerto de Treino	Taxa de acerto de Validação	Teste		
Grupo 4 (1,2,9,19,23,24,28)	Dados Sintéticos: 1 Teste: 2,9,19,23,24,28	<i>SVM</i>	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
		<i>RF</i>	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
		<i>LR</i>	100%	100%	Taxa de acerto	<b>66.67%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>33.33%</b>
	Dados Sintéticos: 2 Teste: 1,9,19,23,24,28	<i>SVM</i>	100%	100%	Taxa de acerto	<b>66.67%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>33.33%</b>
		<i>RF</i>	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
		<i>LR</i>	100%	100%	Taxa de acerto	58.33%
					Sensibilidade	100%
					Especificidade	17%
	Dados Sintéticos: 9 Teste: 1,2,19,23,24,28	<i>SVM</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	66.67%
					Especificidade	100%
		<i>RF</i>	100%	100%	Taxa de acerto	<b>91.67%</b>
					Sensibilidade	<b>83.33%</b>
					Especificidade	<b>100%</b>
		<i>LR</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	66.67%
					Especificidade	100%
	Dados Sintéticos: 19 Teste: 1,2,9,23,24,28	<i>SVM</i>	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		<i>RF</i>	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
		<i>LR</i>	100%	100%	Taxa de acerto	<b>100%</b>
					Sensibilidade	<b>100%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 23 Teste: 1,2,9,19,24,28	<i>SVM</i>	100%	100%	Taxa de acerto	<b>83.33%</b>
					Sensibilidade	<b>66.67%</b>
					Especificidade	<b>100%</b>
		<i>RF</i>	100%	100%	Taxa de acerto	<b>83.33%</b>
					Sensibilidade	<b>66.67%</b>
					Especificidade	<b>100%</b>
		<i>LR</i>	100%	100%	Taxa de acerto	<b>83.33%</b>
					Sensibilidade	<b>66.67%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 24 Teste: 1,2,9,19,23,28	<i>SVM</i>	100%	100%	Taxa de acerto	83.33%
					Sensibilidade	100%
					Especificidade	66.67%
<i>RF</i>		100%	100%	Taxa de acerto	58.33%	
				Sensibilidade	83.33%	
				Especificidade	66.67%	
<i>LR</i>		100%	100%	Taxa de acerto	<b>100%</b>	
				Sensibilidade	<b>100%</b>	
				Especificidade	<b>100%</b>	
Dados Sintéticos: 28 Teste: 1,2,9,19,23,24	<i>SVM</i>	100%	100%	Taxa de acerto	50%	
				Sensibilidade	100%	

	<i>RF</i>	100%	100%	Especificidade	0%
				Taxa de acerto	50%
				Sensibilidade	100%
	<i>LR</i>	100%	98.33%	Especificidade	0%
				Taxa de acerto	50%
				Sensibilidade	100%
				Especificidade	0%

Na Tabela 8 é apresentado os valores médios das métricas de desempenho obtidos em cada grupo pelos três algoritmos de classificação avaliados com base na *GCT\_SD* e *SL*. Na taxa de acerto de teste os resultados obtidos também foram positivos sendo que a confiança das previsões dos classificadores está compreendida entre 70% a 92%. De notar, que o valor médio da sensibilidade é superior ou igual ao valor médio da especificidade, ou seja, a proporção de pacientes sob o efeito da medicação (estado *ON*) corretamente classificados foi em média superior à proporção de pacientes sem o efeito da medicação (estado *OFF*) corretamente classificados.

Tabela 8 Valores médios das métricas de desempenho dos algoritmos de classificação para o grupo 2 e 4 com base na *GCT\_SD* e *SL*

		Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 2 (3,14,26,27)	<i>SVM</i>	100.00%	100.00%	Taxa de acerto	70.83%
				Sensibilidade	91.67%
				Especificidade	50.00%
	<i>RF</i>	100.00%	100.00%	Taxa de acerto	79.17%
				Sensibilidade	83.34%
				Especificidade	75.08%
	<i>LR</i>	99.17%	99.17%	Taxa de acerto	<b>91.67%</b>
				Sensibilidade	<b>91.67%</b>
				Especificidade	<b>91.67%</b>
Grupo 4 (1,2,9,19,23,24,28)	<i>SVM</i>	100.00%	100.00%	Taxa de acerto	79.17%
				Sensibilidade	<b>94.45%</b>
				Especificidade	66.71%
	<i>RF</i>	100.00%	100.00%	Taxa de acerto	<b>82.14%</b>
				Sensibilidade	78.57%
				Especificidade	67.95%
	<i>LR</i>	100.00%	100.00%	Taxa de acerto	79.76%
				Sensibilidade	76.19%
				Especificidade	<b>69.09%</b>

#### 4.4.3 Com base no comprimento de um ciclo (série temporal)

A partir da análise de *clustering* (ver Figura 40) o grupo 7 é o único grupo com mais de dois pacientes e que será usado para modelar e testar os algoritmos *Convolutional Neural Network (CNN)* e *Long Short-Term Memory (LSTM)*. Na Figura 45 estão impressas as séries temporais do comprimento de um ciclo dos sete pacientes do grupo 7 em ambos os estados, em que é possível observar as séries temporais referentes ao estado *ON* apresentam maiores valores de comprimento de um ciclo e menor variabilidade comparativamente às séries temporais no estado *OFF*.

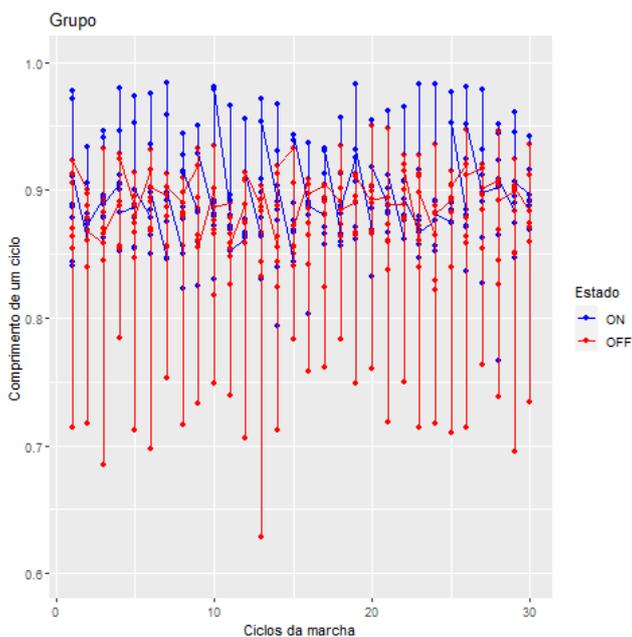


Figura 45 Séries temporais originais do comprimento de um ciclo nos estados *OFF* e *ON* dos pacientes do grupo 7

Na Tabela 9 estão apresentados os resultados obtidos no treino e teste dos diferentes modelos de avaliação com base nas séries temporais do comprimento de um ciclo. Obteve-se uma taxa de acerto média nas *CNN* de 50% e nas *LSTM* de 51.19%. O melhor desempenho com uma taxa de acerto de 75% foi obtido por ambos os classificadores quando os dados sintéticos foram gerados com base no paciente 1.

Tabela 9 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 7 com base nas séries temporais do comprimento de um ciclo

			Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 7 (1, 2, 16, 19, 23, 24, 27)	Dados Sintéticos: 1 Teste: 2, 16, 19, 23, 24, 27	<i>CNN</i>	100%	100%	Taxa de acerto	<b>75%</b>
					Sensibilidade	<b>50%</b>
					Especificidade	<b>100%</b>
		<i>LSTM</i>	100%	100%	Taxa de acerto	<b>75%</b>
					Sensibilidade	<b>50%</b>
					Especificidade	<b>100%</b>
	Dados Sintéticos: 2 Teste: 1, 16, 19, 23, 24, 27	<i>CNN</i>	100%	100%	Taxa de acerto	66.67%
					Sensibilidade	33.33%
					Especificidade	100%
		<i>LSTM</i>	89.58%	90.0%	Taxa de acerto	50.0%
					Sensibilidade	66.67%
					Especificidade	33.33%
	Dados Sintéticos: 16 Teste: 1, 2, 19, 23, 24, 27	<i>CNN</i>	88.75%	90.0%	Taxa de acerto	50%
					Sensibilidade	0%
					Especificidade	100%
		<i>LSTM</i>	50.0%	50.0%	Taxa de acerto	50%
Sensibilidade					0%	
Especificidade					100%	
Dados Sintéticos: 19 Teste: 1, 2, 16, 23, 24, 27	<i>CNN</i>	100.0%	100.0%	Taxa de acerto	50%	
				Sensibilidade	100%	
				Especificidade	0%	
	<i>LSTM</i>	100.0%	100.0%	Taxa de acerto	50%	
				Sensibilidade	100%	
				Especificidade	100%	

	Dados Sintéticos: 23 Teste: 1, 2, 16, 19, 24, 27	<i>CNN</i>	100.0%	100.0%	Especificidade	0%
					Taxa de acerto	33.33%
					Sensibilidade	66.67%
		<i>LSTM</i>	100.0%	100.0%	Especificidade	0%
					Taxa de acerto	33.33%
					Sensibilidade	66.67%
	Dados Sintéticos: 24 Teste: 1, 2, 16, 19, 23, 27	<i>CNN</i>	100.0%	100.0%	Especificidade	0%
					Taxa de acerto	41.67%
					Sensibilidade	83.33%
		<i>LSTM</i>	100.0%	100.0%	Especificidade	0%
					Taxa de acerto	41.67%
					Sensibilidade	33.33%
Dados Sintéticos: 27 Teste: 1, 2, 16, 19, 23, 24	<i>CNN</i>	100.0%	100.0%	Especificidade	50%	
				Taxa de acerto	58.33%	
				Sensibilidade	66.67%	
	<i>LSTM</i>	100.0%	100.0%	Especificidade	50%	
				Taxa de acerto	58.33%	
				Sensibilidade	66.67%	

#### 4.4.4 Com base no comprimento de um ciclo (série temporal) e duração de um ciclo (série temporal)

Nesta subsecção tanto a modelação como o teste dos classificadores *CNN* e *LSTM* foram executados tendo em consideração as séries temporais do comprimento de um ciclo e a duração de um ciclo (Figura 46) de forma a avaliar se a inclusão de mais uma variável melhoraria a performance dos algoritmos de classificação *CNN* e *LSTM*.

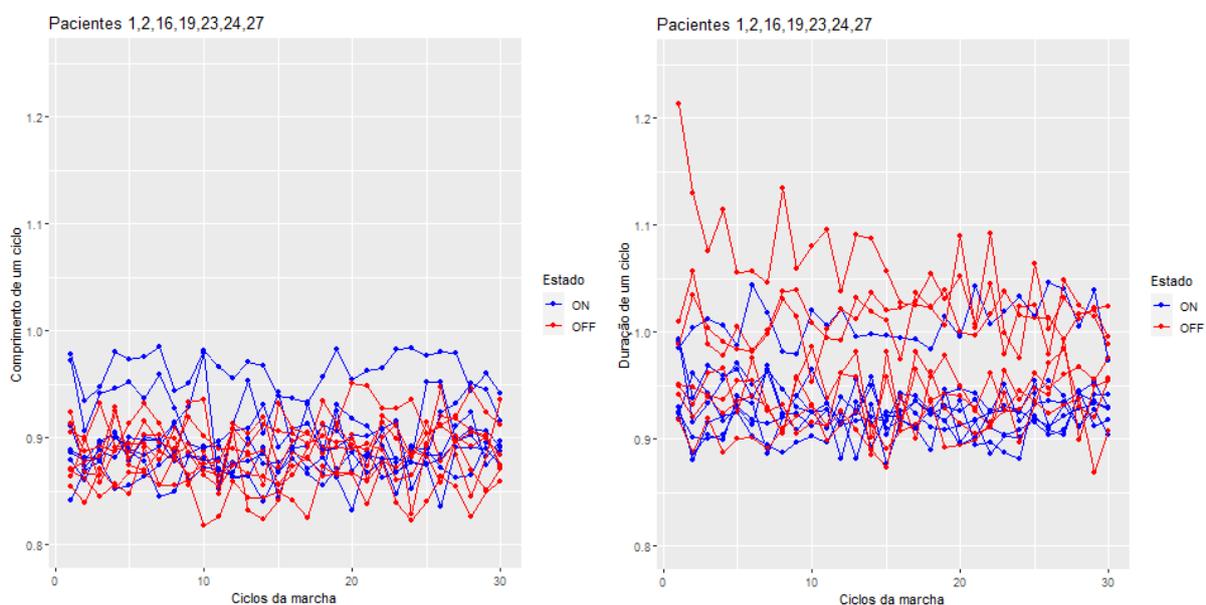


Figura 46 Séries temporais originais do comprimento de um ciclo (à esquerda) e da duração de um ciclo (à direita) nos estados *OFF* e *ON* dos pacientes do grupo 7

A partir dos resultados obtidos e apresentados na Tabela 10 podemos constatar que em geral a taxa de acerto obtida com base nas séries temporais referentes a duas variáveis, comprimento de um ciclo e duração de um ciclo, foi inferior à taxa de acerto obtida com base nas séries temporais referentes

a uma variável, comprimento de um ciclo (Tabela 9). Os valores médios da taxa de acerto foram de 45,3% e 40,5% para a *CNN* e a *LSTM*, respetivamente.

Tabela 10 Métricas de desempenho dos diferentes modelos de classificação aplicados ao grupo 7 com base nas séries temporais do comprimento de um ciclo e da duração de um ciclo

			Taxa de acerto de Treino	Taxa de acerto de Validação	Teste	
Grupo 7 (1, 2, 16, 19, 23, 24, 27)	Dados Sintéticos: 1 Teste: 2, 16, 19, 23, 24, 27	<i>CNN</i>	100%	100%	Taxa de acerto	50%
					Sensibilidade	83.33%
					Especificidade	16.67%
		<i>LSTM</i>	100%	100%	Taxa de acerto	33.33%
					Sensibilidade	0%
					Especificidade	66.67%
	Dados Sintéticos: 2 Teste: 1, 16, 19, 23, 24, 27	<i>CNN</i>	100%	100%	Taxa de acerto	41.67%
					Sensibilidade	83.33%
					Especificidade	66.67%
		<i>LSTM</i>	100%	100%	Taxa de acerto	33.33%
					Sensibilidade	0%
					Especificidade	66.67%
	Dados Sintéticos: 16 Teste: 1, 2, 19, 23, 24, 27	<i>CNN</i>	100%	100%	Taxa de acerto	66.67%
					Sensibilidade	100%
					Especificidade	33.33%
		<i>LSTM</i>	100%	100%	Taxa de acerto	75%
					Sensibilidade	100%
					Especificidade	50%
	Dados Sintéticos: 19 Teste: 1, 2, 16, 23, 24, 27	<i>CNN</i>	100.0%	100.0%	Taxa de acerto	33.33%
					Sensibilidade	33.33%
					Especificidade	33.33%
		<i>LSTM</i>	100.0%	100.0%	Taxa de acerto	50%
					Sensibilidade	33.33%
					Especificidade	66.67%
Dados Sintéticos: 23 Teste: 1, 2, 16, 19, 24, 27	<i>CNN</i>	100.0%	100.0%	Taxa de acerto	33.33%	
				Sensibilidade	0%	
				Especificidade	66.67%	
	<i>LSTM</i>	100.0%	100.0%	Taxa de acerto	33.33%	
				Sensibilidade	0%	
				Especificidade	66.67%	
Dados Sintéticos: 24 Teste: 1, 2, 16, 19, 23, 27	<i>CNN</i>	100%	100%	Taxa de acerto	50%	
				Sensibilidade	83.33%	
				Especificidade	16.67%	
	<i>LSTM</i>	100%	100%	Taxa de acerto	25%	
				Sensibilidade	0%	
				Especificidade	50%	
Dados Sintéticos: 27 Teste: 1, 2, 16, 19, 23, 24	<i>CNN</i>	100.0%	100.0%	Taxa de acerto	41.67%	
				Sensibilidade	16.67%	
				Especificidade	66.67%	
	<i>LSTM</i>	100.0%	100.0%	Taxa de acerto	33.33%	
				Sensibilidade	0%	
				Especificidade	66.67%	

## 4.5 Discussão

Em todos os testes realizados as precisões de treino e validação apresentaram valores muito próximos do perfeito (100%). Isto aconteceu porque os modelos foram treinados e validados com base nos dados sintéticos gerados a partir dos dados reais de um único paciente. Como o objetivo é avaliar os modelos numa classificação personalizada do estado de um paciente os dados gerados não devem ser muito distintos do original. Claramente, se o doente com Parkinson apresentar um padrão de marcha no estado *OFF* e um padrão de marcha diferente no estado *ON* e dentro de cada estado não houver grandes variações no padrão de marcha, com qualquer um dos modelos avaliados irá se obter uma boa classificação do estado do paciente. No entanto, isso não é tão linear quando o doente apresenta uma variação maior no padrão de marcha em cada estado. Daí, neste estudo se ter usado um conjunto de dados reais de outros pacientes para testar os modelos. Os pacientes foram agrupados com base nas características da marcha usadas nos modelos de classificação de forma que o conjunto de teste fosse constituído por dados de marcha que representasse possíveis observações do mesmo paciente.

O algoritmo que apresentou melhores valores de taxa de acerto de teste foi o *Logistic Regression*, sendo que, os melhores resultados foram obtidos quando aplicado unicamente a variabilidade da duração de um ciclo (taxa de acerto média superior a 86,9%). De salientar, que com base na variabilidade da duração de um ciclo os valores de taxa de acerto de teste obtidos pelos outros dois algoritmos, *Random Forest* and *Support Vector Machine*, também foram bons (taxa de acerto média superior a 75%). Com base na variabilidade da duração de um ciclo e do comprimento de um ciclo, apesar de os resultados também serem promissores (taxa de acerto média de teste superior a 70,8%) os valores de taxa de acerto não foram, na maioria dos casos, superiores aos valores de taxa de acerto obtidos com base apenas na variabilidade da duração de um ciclo. Em média os valores de sensibilidade foram superiores aos valores da especificidade, ou seja, o número de classificações corretas do estado *ON* (sob o efeito da medicação) foi superior ao número de classificações corretas do estado *OFF*.

Os resultados obtidos com base nas séries temporais não foram tão promissores. Em trabalhos anteriores (Fernandes et al., 2021), (Fernandes et al., 2018), a maior taxa de acerto na classificação de DPI e DPV, foi obtida usando *Convolution Neural Networks (CNNs)* com base em séries temporais da marcha (taxa de acerto superior a 79,3%) (Fernandes et al., 2021) em comparação com outros métodos convencionais, como *multiple layer perceptrons (MLPs)* e *deep belief networks (DBNs)*, com base em características da marcha (e.g. média do comprimento de um ciclo) (precisões de aproximadamente 63,9% e 72,8%) (Fernandes et al., 2018). As *CNNs* são capazes de extrair características dinâmicas de séries temporais (Wang et al., 2019) o que permitiu detetar com maior taxa de acerto as diferenças

subjacentes no padrão de marcha entre DPI e DPV. No entanto, neste trabalho os resultados obtidos com a aplicação de *CNNs* e *LSTMs* com base em séries temporais não foram superiores aos alcançados usando os métodos mais convencionais (*SVM*, *RF* e *LR*) com base nas características da marcha (valores da média e do desvio padrão como medida de variabilidade).

## Capítulo 5

### Conclusão, Limitações e Trabalho Futuro

Neste capítulo são apresentadas as principais conclusões e limitações do trabalho desenvolvido e termina-se com sugestões de trabalho futuro.

#### 5.1 Conclusão

Recentemente, a aplicação de técnicas de *machine learning* com base em dados de marcha têm-se revelado uma ferramenta complementar útil e eficaz no diagnóstico e avaliação da progressão da doença de Parkinson ((Brognara et al., 2019, Dadashi et al., 2013, Fernandes et al., 2021, 2018, Muthukrishnan et al., 2020, Schlachetzki et al., 2017, Wahid et al., 2015)). Tendo como inspiração estes estudos anteriores, o principal objetivo desta dissertação foi avaliar a eficácia de diferentes técnicas de *machine learning* na classificação do estado de um paciente com Parkinson (sob vs. sem o efeito da medicação levodopa, ou seja, ON vs. OFF) com base em dados de marcha possíveis de serem obtidos por um *smartwatch*. A partir dos resultados obtidos, podemos concluir que a utilização de técnicas de *machine learning* baseadas em variáveis de marcha, tais como a variabilidade da duração de um ciclo, que é possível aceder com um *smartwatch*, poderá ser um bom suporte para a monitorização do estado do paciente (ON vs. OFF). Note-se que um paciente classificado como "estado OFF" pode significar que ainda não tomou o medicamento ou que o medicamento já não está a fazer efeito. No entanto, é de notar que para que seja útil este tipo de monitorização é necessário que o paciente apresente alterações na marcha com a toma da medicação (por exemplo, uma melhoria positiva de pelo menos 10% nos valores de marcha sem a medição para com a medicação).

#### 5.2 Limitações

A principal limitação deste estudo prendeu-se com a base de dados disponível. Idealmente, os dados usados no treino, validação e teste dos algoritmos de classificação deveriam ser dados reais de diferentes avaliações da marcha referentes ao mesmo paciente. Contudo, devido às limitações da atualidade, covid-19, não foi possível recolher uma base de dados ideal para o problema. Com o intuito

de contornar/atenuar esta limitação foram aplicadas técnicas de *clustering*, onde se agrupou os pacientes tendo em conta o seu padrão de marcha e resposta à medicação, e à posteriori gerou-se uma base de dados sintética de forma a ter o número suficiente de dados para treinar os modelos.

### 5.3 Trabalho Futuro

Os bons resultados obtidos nesta dissertação justificam o alargamento da investigação aqui iniciada como trabalho futuro:

- Realizar o treino, validação e teste dos modelos de classificação com base em dados de marcha do mesmo paciente com e sem medicação;
- Estender o estudo a novas variáveis de marcha que com a evolução dos *smartwatches* começam a ser possíveis de obter (e.g. tempo de balanço (*swing time*));
- Analisar se a utilização do método de redução da dimensionalidade Análise de Componentes Principais (PCA) traria vantagens e melhorias no desempenho dos modelos;
- Estudar a possibilidade do treino e classificação dos modelos serem totalmente realizados no *smartwatch* ou com recurso a outro dispositivo (e.g. *computador*, *smartphone*). Analisar as vantagens e desvantagens, por exemplo, estudo do poder computacional necessário, estudo do risco de perda de informação na comunicação, atrasos de comunicação, entre outros;

## Referências

- [1] A. L., “Doença de Parkinson afeta entre 18 a 20 mil pessoas em Portugal: SPN - Sociedade Portuguesa de Neurologia,” 2020. [Online]. Available: <https://www.spneurologia.com/noticias/doenca-de-parkinson-afeta-entre-18-a-20-mil-pessoa/51>. [Accessed: 15-Jul-2021]
- [2] C. Fernandes, F. Ferreira, R. L. Lopes, E. Bicho, W. Erlhagen, N. Sousa, and M. F. Gago, “Discrimination of idiopathic Parkinson’s disease and vascular parkinsonism based on gait time series and the levodopa effect,” *Journal of Biomechanics*, p. 110214, 2021, doi: 10.1016/j.jbiomech.2020.110214.
- [3] C. Fernandes, F. Ferreira, M. Gago, O. Azevedo, N. Sousa, W. Erlhagen, and E. Bicho, “Gait classification of patients with Fabry’s disease based on normalized gait features obtained using multiple regression models,” in *2019 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2019, pp. 2288–2295, doi: 10.1109/BIBM47256.2019.8983241.
- [4] M. Gago, F. Ferreira, N. Mollaei, M. Rodrigues, N. Sousa, E. Bicho, and P. Rodrigues, “Gait analysis as a complementary tool in the levodopa dose decision in vascular Parkinson’s disease,” 2017.
- [5] Apda, “Education & Support: American Parkinson Disease Assoc.,” 2021. [Online]. Available: <https://www.apdaparkinson.org/resources-support/>. [Accessed: 15-Jul-2021]
- [6] cuf, “Doença de Parkinson,” 2021. [Online]. Available: <https://www.cuf.pt/saude-a-z/doenca-de-parkinson>. [Accessed: 15-Jul-2021]
- [7] N. S. Erdem, C. Ersoy, and C. Tunca, “Gait analysis using smartwatches,” in *2019 IEEE 30th International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC Workshops)*, 2019, pp. 1–6, doi: 10.1109/pimrcw.2019.8880821 .
- [8] F. Dadashi, B. Mariani, S. Rochat, C. J. Büla, B. Santos-Eggimann, and K. Aminian, “Gait and foot clearance parameters obtained using shoe-worn inertial sensors in a large-population sample of older adults.,” *Sensors (Basel)*, vol. 14, no. 1, pp. 443–57, 2013, doi: 10.3390/s140100443.
- [9] E. Rovini, C. Maremmani, A. Moschetti, D. Esposito, and F. Cavallo, “Comparative motor pre-clinical assessment in Parkinson’s disease using supervised machine learning approaches,” *Annals of biomedical engineering*, vol. 46, no. 12, pp. 2057–2068, 2018, doi: 10.1007/s10439-018-2104-9.
- [10] N. M. Tahir and H. H. Manap, “Parkinson disease gait classification based on machine learning approach,” *Journal of Applied Sciences(Faisalabad)*, vol. 12, no. 2, pp. 180–185, 2012, doi: 10.3923/jas.2012.180.185.
- [11] 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 journal of biomedical and health informatics*, vol. 19, no. 6, pp. 1794–1802, 2015, doi: 10.1109/jbhi.2015.2450232 .
- [12] B. K., “ilustrações cartoon,” 2019. [Online]. Available: <https://www.dreamstime.com/illustration/cartoon-man-looking-his-watch.html>. [Accessed: 22-

Nov-2021]

- [13] A. Site, J. Nurmi, and E. S. Lohan, "Systematic review on machine-learning algorithms used in wearable-based eHealth data analysis," *IEEE Access*, 2021, doi: 10.1109/ACCESS.2021.3103268.
- [14] S. Sprager and M. B. Juric, "Inertial sensor-based gait recognition: A review," *Sensors*, vol. 15, no. 9, pp. 22089–22127, 2015, doi: 10.3390/s150922089.
- [15] L. Brognara, P. Palumbo, B. Grimm, and L. Palmerini, "Assessing gait in Parkinson's disease using wearable motion sensors: a systematic review," *Diseases*, vol. 7, no. 1, p. 18, 2019, doi: 10.3390/diseases7010018.
- [16] J. C. Schlachetzki, J. Barth, F. Marxreiter, J. Gossler, Z. Kohl, S. Reinfelder, H. Gassner, K. Aminian, B. M. Eskofier, J. Winkler, and others, "Wearable sensors objectively measure gait parameters in Parkinson's disease," *PLoS one*, vol. 12, no. 10, p. e0183989, 2017, doi: 10.1371/journal.pone.0183989.
- [17] F. Bugané, M. Benedetti, G. Casadio, S. Attala, F. Biagi, M. Manca, and A. Leardini, "Estimation of spatial-temporal gait parameters in level walking based on a single accelerometer: Validation on normal subjects by standard gait analysis," *Computer methods and programs in biomedicine*, vol. 108, no. 1, pp. 129–137, 2012, doi: 10.1016/j.cmpb.2012.02.003.
- [18] N. Muthukrishnan, J. J. Abbas, and N. Krishnamurthi, "A wearable sensor system to measure step-based gait parameters for parkinson's disease rehabilitation," *Sensors*, vol. 20, no. 22, p. 6417, 2020, doi: 10.3390/s20226417.
- [19] C. Fernandes, L. Fonseca, F. Ferreira, M. Gago, L. Costa, N. Sousa, C. Ferreira, J. Gama, W. Erlhagen, and E. Bicho, "Artificial neural networks classification of patients with Parkinsonism based on gait," in *2018 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2018, pp. 2024–2030, doi: 10.1109/bibm.2018.8621466.
- [20] P. Esser, H. Dawes, J. Collett, M. G. Feltham, and K. Howells, "Assessment of spatio-temporal gait parameters using inertial measurement units in neurological populations," *Gait & posture*, vol. 34, no. 4, pp. 558–560, 2011, doi: 10.1016/j.gaitpost.2011.06.01.
- [21] P. Gray and K. Hildebrand, "Fall risk factors in Parkinson's disease," *Journal of Neuroscience Nursing*, vol. 32, no. 4, p. 222, 2000, doi: 10.1097/01376517-200008000-00006.
- [22] V. Studer, R. Maestri, I. Clerici, L. Spina, I. Zivi, D. Ferrazzoli, and G. Frazzitta, "Treadmill training with cues and feedback improves gait in people with more advanced Parkinson's disease," *Journal of Parkinson's disease*, vol. 7, no. 4, pp. 729–739, 2017, doi: 10.3233/jpd-171126 .
- [23] GAITrite, "World Leader in Temporospacial Gait Analysis," 2021. [Online]. Available: <https://www.gaitrite.com/>. [Accessed: 27-Oct-2021]
- [24] Hasomed, "RehaGait," 2021. [Online]. Available: <https://hasomed.de/en/products/rehagait/>. [Accessed: 27-Oct-2021]

- [25] Gaitup, "Our products: Easy and accurate motion analysis solutions," 2020. [Online]. Available: <https://research.gaitup.com/products/>. [Accessed: 27-Oct-2021]
- [26] F. Ferreira, M. F. Gago, E. Bicho, C. Carvalho, N. Mollaei, L. Rodrigues, N. Sousa, P. P. Rodrigues, C. Ferreira, and J. Gama, "Gait stride-to-stride variability and foot clearance pattern analysis in Idiopathic Parkinson's Disease and Vascular Parkinsonism," *Journal of biomechanics*, vol. 92, pp. 98–104, 2019, doi: 10.1016/j.jbiomech.2019.05.039.
- [27] G. Belloni, L. Seematter-Bagnoud, H. Krief, K. Aminian, and C. J. Büla, "Falls Efficacy Is Associated With Better Gait and Functional Outcomes After Rehabilitation in Older Patients," *Archives of Physical Medicine and Rehabilitation*, vol. 102, no. 6, pp. 1134–1139, 2021, doi: 10.1016/j.apmr.2020.12.017.
- [28] B. Mariani, C. Hoskovec, S. Rochat, C. Büla, J. Penders, and K. Aminian, "3D gait assessment in young and elderly subjects using foot-worn inertial sensors.," *J Biomech*, vol. 43, no. 15, pp. 2999–3006, 2010, doi: 10.1016/j.jbiomech.2010.07.003.
- [29] M. Gwak, M. Sarrafzadeh, and E. Woo, "Support for a clinical diagnosis of mild cognitive impairment using photoplethysmography and gait sensors," in *Proceedings, APSIPA Annual Summit and Conference*, 2018, vol. 2018, pp. 12–15.
- [30] P. Musale, D. Baek, N. Werellagama, S. S. Woo, and B. J. Choi, "You walk, we authenticate: lightweight seamless authentication based on gait in wearable IoT systems," *IEEE Access*, vol. 7, pp. 37883–37895, 2019, doi: 10.1109/access.2019.2906663.
- [31] Garmin, "Dinâmicas de corrida e Medições fisiológicas," 2021. [Online]. Available: <https://discover.garmin.com/pt-PT/performance-data/running/#ground-contact-time-balance>. [Accessed: 26-Jan-2021]
- [32] M. S., "Apple introduces new Mobility Metrics and other key HealthKit data with iOS 14 and watchOS 7," 2021. [Online]. Available: <https://www.myhealthyapple.com/apple-introduces-new-mobility-metrics-and-other-key-healthkit-data-with-ios-14-and-watchos-7/>. [Accessed: 26-Jan-2021]
- [33] G. up, *Gait Analysis Report*. 2018.
- [34] S. Del Din, M. Elshehabi, B. Galna, M. A. Hobert, E. Warmerdam, U. Suenkel, K. Brockmann, F. Metzger, C. Hansen, D. Berg, and others, "Gait analysis with wearables predicts conversion to parkinson disease," *Annals of neurology*, vol. 86, no. 3, pp. 357–367, 2019, doi: 10.1002/ana.25548.
- [35] J. Obeso, M. Stamelou, C. Goetz, W. Poewe, A. Lang, D. Weintraub, D. Burn, G. M. Halliday, E. Bezard, S. Przedborski, and others, "Past, present, and future of Parkinson's disease: A special essay on the 200th Anniversary of the Shaking Palsy," *Movement Disorders*, vol. 32, no. 9, pp. 1264–1310, 2017, doi: 10.1002/mds.27115.
- [36] M. S. Bryant, D. H. Rintala, J. G. Hou, A. L. Charness, A. L. Fernandez, R. L. Collins, J. Baker, E. C. Lai, and E. J. Protas, "Gait variability in Parkinson's disease: influence of walking speed and dopaminergic treatment," *Neurological research*, vol. 33, no. 9, pp. 959–964, 2011, doi:

- 10.1179/1743132811y.0000000004.
- [37] M. S. Bryant, D. H. Rintala, J.-G. Hou, R. L. Collins, and E. J. Protas, "Gait variability in Parkinson's disease: levodopa and walking direction," *Acta Neurologica Scandinavica*, vol. 134, no. 1, pp. 83–86, 2016, doi: 10.1111/ane.12505.
- [38] M. Son, S. H. Han, C. H. Lyoo, J. A. Lim, J. Jeon, K.-B. Hong, and H. Park, "The effect of levodopa on bilateral coordination and gait asymmetry in Parkinson's disease using inertial sensor," *npj Parkinson's Disease*, vol. 7, no. 1, pp. 1–7, 2021, doi: 10.1038/s41531-021-00186-7.
- [39] J. Braga, F. Ferreira, C. Fernandes, M. F. Gago, O. Azevedo, N. Sousa, W. Erlhagen, and E. Bicho, "Gait Characteristics and Their Discriminative Ability in Patients with Fabry Disease with and Without White-Matter Lesions," in *International Conference on Computational Science and Its Applications*, 2020, pp. 415–428, doi: 10.1007/978-3-030-58808-3.
- [40] C. Caramia, D. Torricelli, M. Schmid, A. Munoz-Gonzalez, J. Gonzalez-Vargas, F. Grandas, and J. L. Pons, "IMU-based classification of Parkinson's disease from gait: A sensitivity analysis on sensor location and feature selection," *IEEE journal of biomedical and health informatics*, vol. 22, no. 6, pp. 1765–1774, 2018, doi: 10.1109/jbhi.2018.2865218.
- [41] 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," *Scientific reports*, vol. 9, no. 1, pp. 1–12, 2019, doi: 10.1038/s41598-019-53656-7 .
- [42] C. Caramia, I. Bernabucci, C. D'Anna, C. De Marchis, and M. Schmid, "Gait parameters are differently affected by concurrent smartphone-based activities with scaled levels of cognitive effort," *PloS one*, vol. 12, no. 10, p. e0185825, 2017, doi: 10.1371/journal.pone.0185825.
- [43] S. Aich, K. Younga, K. L. Hui, A. A. Al-Absi, and M. Sain, "A nonlinear decision tree based classification approach to predict the Parkinson's disease using different feature sets of voice data," in *2018 20th International Conference on Advanced Communication Technology (ICACT)*, 2018, pp. 638–642, doi: 10.23919/icact.2018.8323864.
- [44] A. Mannini, D. Trojaniello, A. Cereatti, and A. M. Sabatini, "A machine learning framework for gait classification using inertial sensors: Application to elderly, post-stroke and huntington's disease patients," *Sensors*, vol. 16, no. 1, p. 134, 2016, doi: 10.3390/s16010134.
- [45] C. Pradhan, M. Wuehr, F. Akrami, M. Neuhaeusser, S. Huth, T. Brandt, K. Jahn, and R. Schniepp, "Automated classification of neurological disorders of gait using spatio-temporal gait parameters," *Journal of Electromyography and Kinesiology*, vol. 25, no. 2, pp. 413–422, 2015, doi: 10.1016/j.jelekin.2015.01.004.
- [46] I. El Maachi, G.-A. Bilodeau, and W. Bouachir, "Deep 1D-Convnet for accurate Parkinson disease detection and severity prediction from gait," *Expert Systems with Applications*, vol. 143, p. 113075, 2020, doi: 10.1016/j.eswa.2019.113075.
- [47] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2324, 1998, doi:

- 10.1109/5.726791.
- [48] C. Shorten and T. Khoshgoftaar, "A survey on image data augmentation for deep learning. J. Big Data 6 (1), 1-48 (2019)", doi: 10.1186/s40537-019-0197-0.
- [49] L. Tran and D. Choi, "Data augmentation for inertial sensor-based gait deep neural network," *IEEE Access*, vol. 8, pp. 12364–12378, 2020, doi: 10.1109/ACCESS.2020.2966142.
- [50] T. Taewoong Um, F. M. J. Pfister, D. Pichler, S. Endo, M. Lang, S. Hirche, U. Fietzek, and D. Kulić, "Data augmentation of wearable sensor data for parkinson's disease monitoring using convolutional neural networks," *arXiv e-prints*, p. arXiv-1706, 2017, doi: 10.1145/3136755.3136817.
- [51] I. H. Lopez-Nava, L. M. Valentín-Coronado, M. Garcia-Constantino, and J. Favela, "Gait activity classification on unbalanced data from inertial sensors using shallow and deep learning," *Sensors*, vol. 20, no. 17, p. 4756, 2020, doi: 10.3390/s20174756.
- [52] F. Ferreira, C. Fernandes, J. Braga, E. Bicho, W. Erlhagen, and M. F. Gago, "The effect of levodopa medication on stride time variability in patients with Parkinsonism," *Journal of Statistics on Health Decision*, vol. 2, no. 2, pp. 63–64, 2020, doi: 10.34624/jshd.v2i2.20109.
- [53] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems*. O'Reilly Media, 2019, ISBN: 978-1-492-03264-9.
- [54] C. M. Bishop, *Pattern recognition and machine learning*. springer, 2006.
- [55] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature biotechnology*, vol. 37, no. 1, pp. 38–44, 2019, doi: 10.1038/nbt.4314.
- [56] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *Journal of machine learning research*, vol. 9, no. Nov, pp. 2579–2605, 2008.
- [57] O. Chapelle, B. Scholkopf, and A. Zien, "Semi-supervised learning (chapelle, o. et al., eds.; 2006)[book reviews]," *IEEE Transactions on Neural Networks*, vol. 20, no. 3, pp. 542–542, 2009.
- [58] C. J. C. H. Watkins, "Learning from delayed rewards," 1989.
- [59] C. Kirtley, M. W. Whittle, and R. Jefferson, "Influence of walking speed on gait parameters," *Journal of biomedical engineering*, vol. 7, no. 4, pp. 282–288, 1985, doi: 10.1016/0141-5425(85)90055-x .
- [60] R. Senden, K. Meijer, I. C. Heyligers, H. H. C. M. Savelberg, and B. Grimm, "Importance of correcting for individual differences in the clinical diagnosis of gait disorders.," *Physiotherapy*, vol. 98, no. 4, pp. 320–4, 2012, doi: 10.1016/j.physio.2011.06.002.
- [61] F. Wahid, R. Begg, N. Lythgo, C. J. Hass, S. Halgamuge, and D. C. Ackland, "A multiple regression approach to normalization of spatiotemporal gait features," *Journal of applied*

- biomechanics*, vol. 32, no. 2, pp. 128–139, 2016, doi: 10.1123/jab.2015-0035.
- [62] M. Sugiyama, *Introduction to statistical machine learning*. Morgan Kaufmann, 2015, pp. 405–428, doi: 10.1016/b978-0-12-802121-7.00046-7 .
- [63] S. Gollapudi, *Practical machine learning*. Packt Publishing Ltd, 2016, ISBN: 978-1-78439-968-9.
- [64] K. El Bouchefry and R. S. de Souza, “Learning in Big Data: Introduction to Machine Learning,” in *Knowledge Discovery in Big Data from Astronomy and Earth Observation*, Elsevier, 2020, pp. 225–249, doi: 10.1016/b978-0-12-819154-5.00023-0 .
- [65] G. Chandrashekar and F. Sahin, “A survey on feature selection methods,” *Computers & Electrical Engineering*, vol. 40, no. 1, pp. 16–28, 2014, doi: 10.1016/j.compeleceng.2013.11.024.
- [66] S. Visalakshi and V. Radha, “A literature review of feature selection techniques and applications: Review of feature selection in data mining,” in *2014 IEEE International Conference on Computational Intelligence and Computing Research*, 2014, pp. 1–6, doi: 10.1109/ICCIC.2014.7238499.
- [67] A. J., “A Complete Tutorial on Ridge and Lasso Regression in Python,” 2016. [Online]. Available: <https://www.analyticsvidhya.com/blog/2016/01/ridge-lasso-regression-python-complete-tutorial/>. [Accessed: 28-Jan-2021]
- [68] M. Kuhn and K. Johnson, “Filter Methods,” in *Applied predictive modeling*, Springer, 2013, p. 499, doi: 10.1007/978-1-4614-6849-3.
- [69] E. Ostertagová and O. Ostertag, “Methodology and application of oneway ANOVA,” *American Journal of Mechanical Engineering*, vol. 1, no. 7, pp. 256–261, 2013, doi: 10.12691/ajme-1-7-21.
- [70] J. Sauro and J. Lewis, “An introduction to correlation, regression, and ANOVA,” in *Quantifying the User Experience*, Elsevier, 2016, pp. 277–320, doi: 10.1016/b978-0-12-802308-2.00010-2 .
- [71] D. Fraiman and R. Fraiman, “An ANOVA approach for statistical comparisons of brain networks,” *Scientific reports*, vol. 8, no. 1, pp. 1–14, 2018, doi: 10.1038/s41598-018-23152-5.
- [72] P. Y. Chen, M. Smithson, and P. M. Popovich, *Correlation: Parametric and nonparametric measures*, no. 139. Sage, 2002, ISBN: 0761922288 9780761922285.
- [73] C. Y., “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: 08-Jul-2021]
- [74] G. Jo\`vst, J. Huber, M. Heri\`vcko, and G. Polan\`vci\`vc, “Improving cognitive effectiveness of business process diagrams with opacity-driven graphical highlights,” *Decision Support Systems*,

- vol. 103, pp. 58–69, 2017, doi: 10.1016/b978-0-12-374892-8.00006-5 .
- [75] H. Akoglu, “User’s guide to correlation coefficients,” *Turkish journal of emergency medicine*, vol. 18, no. 3, pp. 91–93, 2018, doi: 10.1016/j.tjem.2018.08.001.
- [76] J. Swarndeept Saket and S. Pandya, “An overview of partitioning algorithms in clustering techniques,” *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 5, no. 6, pp. 1943–1946, 2016, doi: 10.3233/IDA-2007-11602.
- [77] K. Bindra and A. Mishra, “A detailed study of clustering algorithms,” in *2017 6th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions)(ICRITO)*, 2017, pp. 371–376, doi: 10.1109/icrito.2017.8342454 .
- [78] S. Bano and M. Khan, “A survey of data clustering methods,” *International Journal of Advanced Science and Technology*, vol. 113, pp. 133–142, 2018, doi: 10.14257/ijast.2018.113.14.
- [79] D. Berndt and J. Clifford, “Using dynamic time warping to find patterns in time series,” in *AAAI-94 workshop on knowledge discovery in databases*, 1994, vol. 2.
- [80] S. Na, L. Xumin, and G. Yong, “Research on k-means clustering algorithm: An improved k-means clustering algorithm,” in *2010 Third International Symposium on intelligent information technology and security informatics*, 2010, pp. 63–67, doi: 10.1109/iitsi.2010.74 .
- [81] F. Nielsen, “Hierarchical clustering,” in *Introduction to HPC with MPI for Data Science*, Springer, 2016, pp. 195–211, doi: 10.1007/978-3-319-21903-5\_8.
- [82] G. Bontempi, M. Birattari, and H. Bersini, “Lazy learning: a logical method for supervised learning,” in *New learning paradigms in soft computing*, Springer, 2002, pp. 97–136, doi: 10.1007/978-3-7908-1803-1\_4 .
- [83] H. Homayouni, S. Hashemi, and A. Hamzeh, “A lazy ensemble learning method to classification,” *International Journal of Computer Science Issues (IJCSI)*, vol. 7, no. 5, p. 344, 2010, ISBN: 1694-0814 .
- [84] M. Niranjanamurthy, M. Kiran, S. Anupama, and C. Dharmendra, “Comparative study on performance testing with JMeter,” *Int. J. Adv. Res. Comput. Commun. Eng.*, vol. 5, no. 2, pp. 70–76, 2016, doi: 10.17148/IJARCCE .
- [85] D. Bhargava, S. Vyas, and A. Bansal, “Comparative analysis of classification techniques for brain magnetic resonance imaging images,” in *Advances in Computational Techniques for Biomedical Image Analysis*, Elsevier, 2020, pp. 133–144, doi: 10.1016/B978-0-12-820024-7.00007-4.
- [86] A. Kumar, L. Bi, J. Kim, and D. D. Feng, “Machine learning in medical imaging,” in *Biomedical Information Technology*, Elsevier, 2020, pp. 167–196, doi: 10.1016/B978-0-12-816034-3.00005-5 .
- [87] L. Guo, N. Chehata, C. Mallet, and S. Boukir, “Relevance of airborne lidar and multispectral image data for urban scene classification using Random Forests,” *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 66, no. 1, pp. 56–66, 2011, doi:

- 10.1016/j.isprsjprs.2010.08.007.
- [88] S. Misra and H. Li, “Noninvasive fracture characterization based on the classification of sonic wave travel times,” *Machine Learning for Subsurface Characterization*, pp. 243–287, 2019, doi: 10.1016/B978-0-12-817736-5.00009-0.
- [89] P. Fuleky, *Macroeconomic forecasting in the era of big data: Theory and practice*, vol. 52. Springer, 2019, doi: 10.1007/978-3-030-31150-6 .
- [90] S. Dreiseitl and L. Ohno-Machado, “Logistic regression and artificial neural network classification models: a methodology review,” *Journal of biomedical informatics*, vol. 35, no. 5–6, pp. 352–359, 2002, doi: 10.1016/S1532-0464(03)00034-0.
- [91] Kambria, “Logistic Regression For Machine Learning and Classification,” 2019. [Online]. Available: <https://kambria.io/blog/logistic-regression-for-machine-learning/>. [Accessed: 29-Mar-2021]
- [92] S. Albawi, T. A. Mohammed, and S. Al-Zawi, “Understanding of a convolutional neural network,” in *2017 International Conference on Engineering and Technology (ICET)*, 2017, pp. 1–6, doi: 10.1109/icengtechnol.2017.8308186 .
- [93] K. Srinivas, B. Kavitha Rani, M. Varaprasad Rao, G. Madhukar, and B. Venkata Ramana, “Convolution Neural Networks for Binary Classification,” *Journal of Computational and Theoretical Nanoscience*, vol. 16, no. 11, pp. 4877–4882, 2019, doi: 10.1166/jctn.2019.8399.
- [94] W. Rawat and Z. Wang, “Deep convolutional neural networks for image classification: A comprehensive review,” *Neural computation*, vol. 29, no. 9, pp. 2352–2449, 2017, doi: 10.1162/NECO\_a\_00990.
- [95] L. Lai, N. Suda, and V. Chandra, “Cmsis-nn: Efficient neural network kernels for arm cortex-m cpus,” *arXiv preprint arXiv:1801.06601*, 2018.
- [96] C. Gallo, “Artificial neural networks tutorial,” in *Encyclopedia of Information Science and Technology, Third Edition*, IGI Global, 2015, pp. 6369–6378, doi: 10.4018/978-1-4666-5888-2.ch626.
- [97] M. Elsaraiti and A. Merabet, “Application of Long-Short-Term-Memory Recurrent Neural Networks to Forecast Wind Speed,” *Applied Sciences*, vol. 11, no. 5, p. 2387, 2021, doi: 10.3390/app11052387.
- [98] S. Hochreiter, “JA1 4 rgen Schmidhuber (1997). ‘Long Short-Term Memory,’” *Neural Computation*, vol. 9, no. 8, 1997, doi: 10.1162/neco.1997.9.8.1735.
- [99] M. A., “Metrics to Evaluate your Machine Learning Algorithm,” 2018. [Online]. Available: <https://towardsdatascience.com/metrics-to-evaluate-your-machine-learning-algorithm-f10ba6e38234>. [Accessed: 29-Jan-2021]

- 
- [100] D. M. Powers, "Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation," *arXiv preprint arXiv:2010.16061*, 2020.
- [101] O. Kramer, "Scikit-learn," in *Machine learning for evolution strategies*, Springer, 2016, pp. 45–53.
- [102] D. K., "Home - Keras Documentation," 2021. [Online]. Available: <https://faroit.com/keras-docs/1.2.0/>. [Accessed: 22-Apr-2021]
- [103] Keras, "Keras," 2021. [Online]. Available: <https://keras.io/>. [Accessed: 22-Apr-2021]
- [104] NumPy, "NumPy," 2021. [Online]. Available: <https://numpy.org/>. [Accessed: 15-Nov-2021]
- [105] Pandas, "pandas," 2021. [Online]. Available: <https://pandas.pydata.org/>. [Accessed: 15-Nov-2021]
- [106] V. Agostini, M. Lanotte, M. Carlone, M. Campagnoli, I. Azzolin, R. Scarafia, G. Massazza, and M. Knaflitz, "Instrumented gait analysis for an objective pre-/postassessment of tap test in normal pressure hydrocephalus," *Archives of physical medicine and rehabilitation*, vol. 96, no. 7, pp. 1235–1241, 2015, doi: 10.1016/j.apmr.2015.02.014.
- [107] J. Virhammar, K. G. Cesarini, and K. Laurell, "The CSF tap test in normal pressure hydrocephalus: evaluation time, reliability and the influence of pain.," *Eur. J. Neurol.*, vol. 19, no. 2, pp. 271–6, 2012, doi: 10.1111/j.1468-1331.2011.03486.x.
- [108] J. Wang, Y. Chen, S. Hao, X. Peng, and L. Hu, "Deep learning for sensor-based activity recognition: A survey," *Pattern Recognition Letters*, vol. 119, pp. 3–11, 2019.

# Anexos

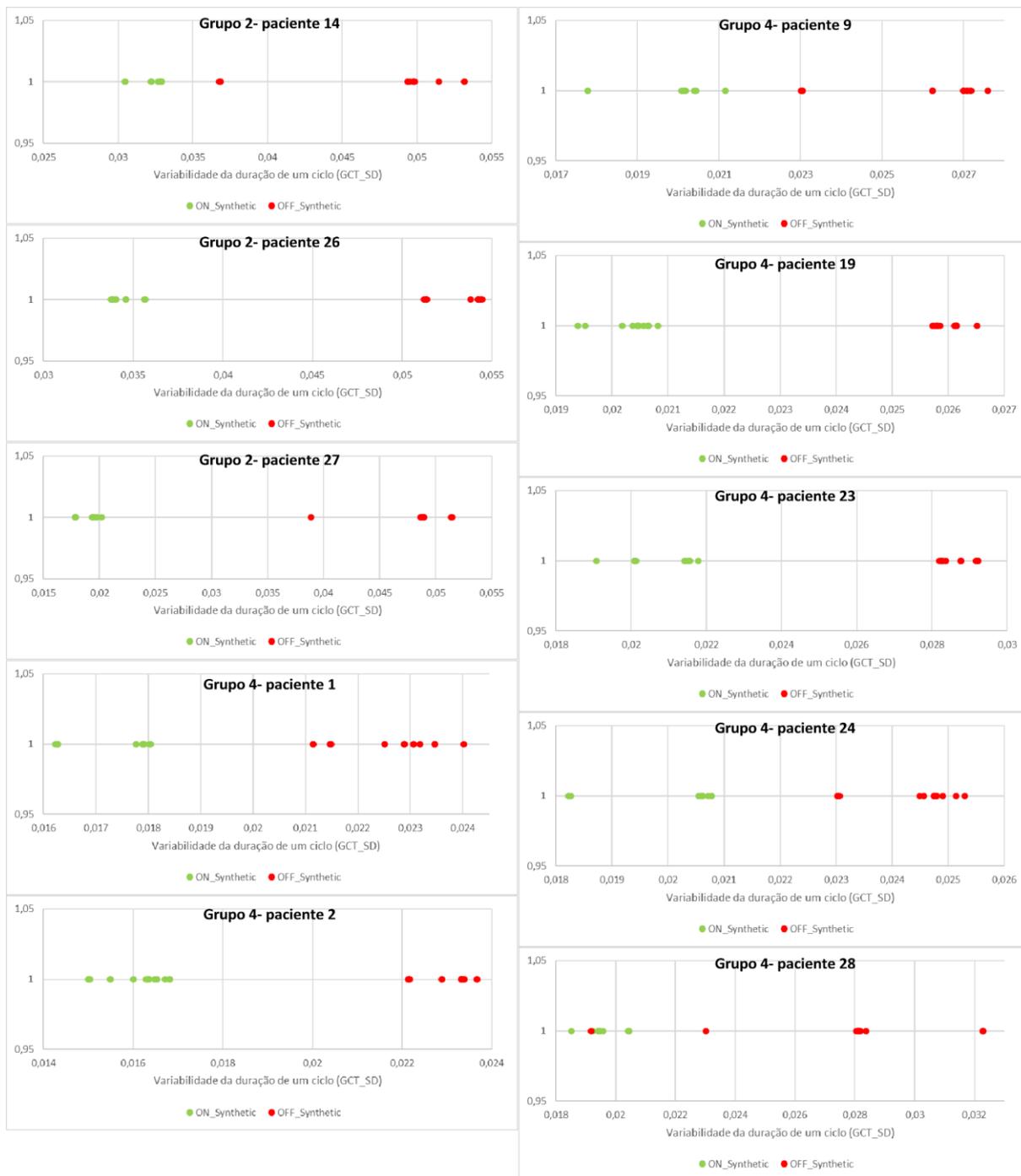


Figura 47 Dados sintéticos (GCT\_SD)

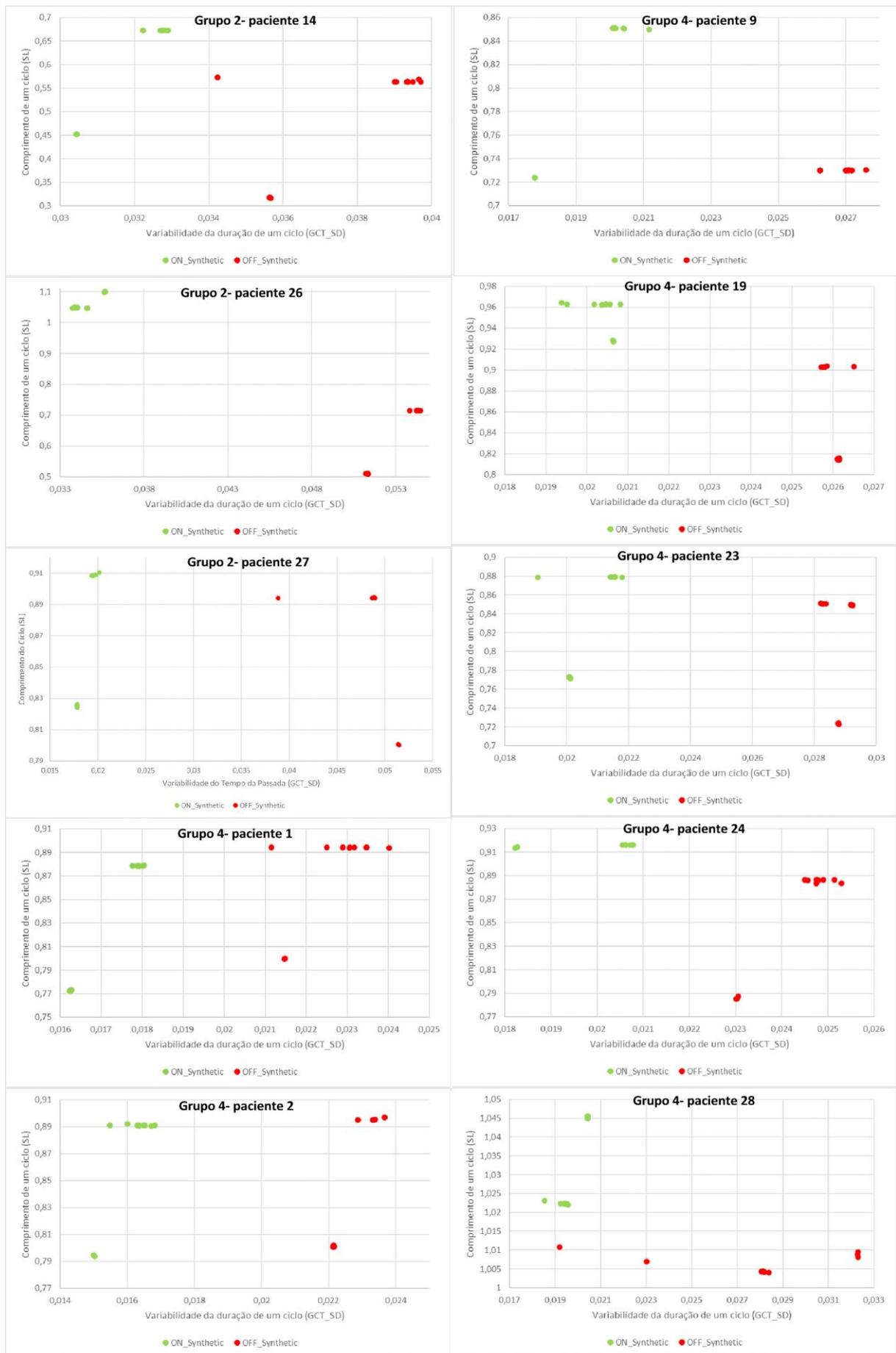


Figura 48 Dados sintéticos (GCT\_SD+SL)

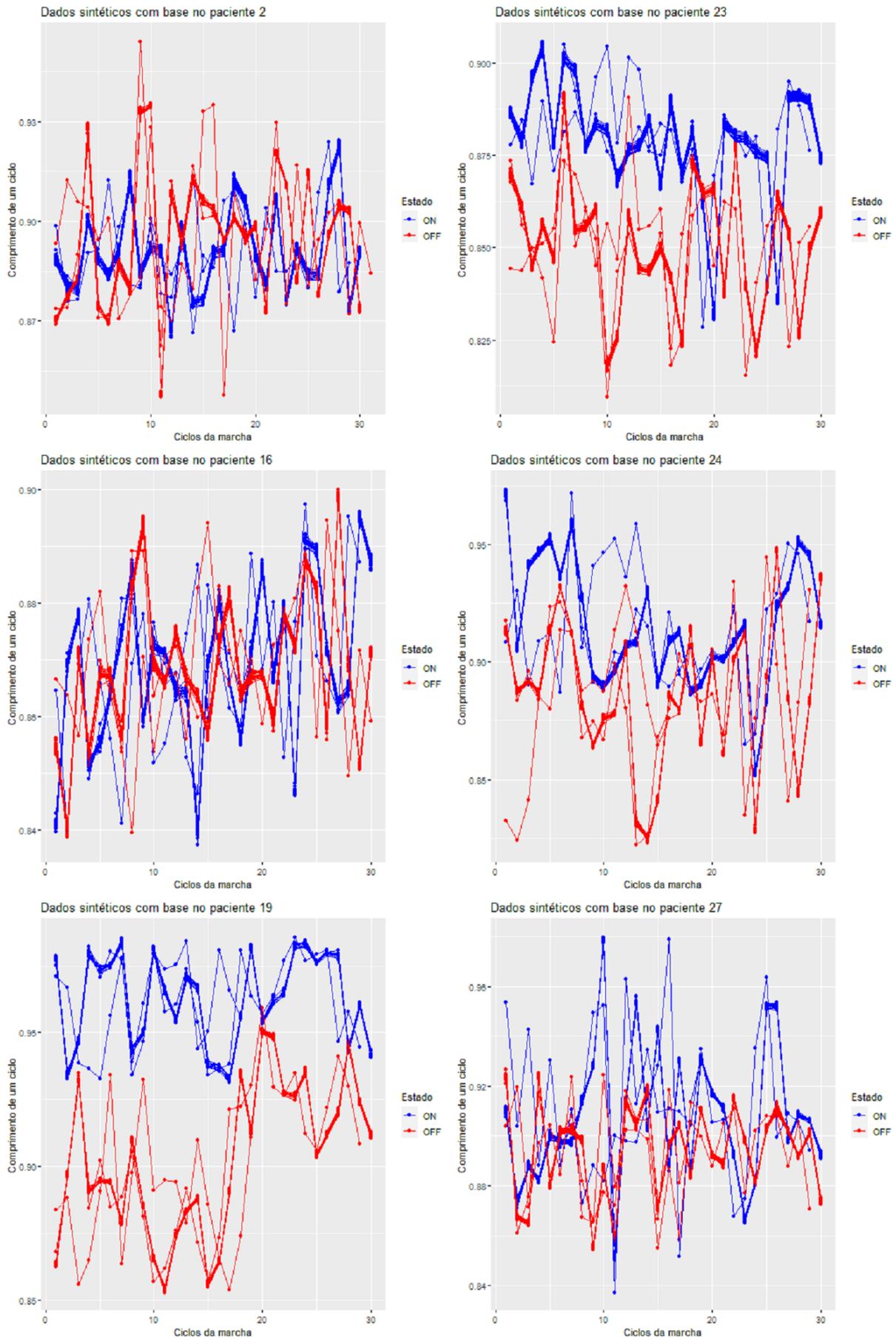


Figura 49 Séries temporais sintéticas (SL)

Tabela 11 Modelos de regressão linear resultantes para as variáveis de marcha

<b>Variáveis de Marcha</b>	<b>Modelo</b>	<b>Coefficientes Estimados (<math>\beta</math>)</b>	<b>AIC</b>	<b>R<sup>2</sup> Ajustado</b>
Duração de um Ciclo	$\beta_0 + \beta_1 * \text{idade} + \beta_2 * \text{altura} + \beta_3 * \text{velocidade}$	0.84, -0.0011, 0.372, -0.286	-143.71	0.771
Cadência	$\beta_0 + \beta_1 * \text{idade} + \beta_2 * \text{altura} + \beta_3 * \text{velocidade}$	137.86, 0.128, -42.44, 33.41	197.76	0.776
Tempo de Duplo Suporte	$\beta_0 + \beta_1 * \text{velocidade}$	35.78, -11.89	187.90	0.277
Comprimento de um Ciclo	$\beta_0 + \beta_1 * \text{idade} + \beta_2 * \text{altura} + \beta_3 * \text{velocidade}$	-0.23, -0.0014, 0.502, 0.611	-126.29	0.928