

**UNIVERSIDADE ABERTA**



**DEPARTAMENTO DE CIÊNCIAS E TECNOLOGIA**

**Mestrado em Estatística, Matemática e Computação**

**Máquinas de Vetores Suporte para Classificação do *Onset* em  
dados Temporais de Eletromiografia**

**Luís Miguel Domingues Ferreira Silva**

Lisboa, 2014



**UNIVERSIDADE ABERTA**



**DEPARTAMENTO DE CIÊNCIAS E TECNOLOGIA**

**Mestrado em Estatística, Matemática e Computação**

**Máquinas de Vetores Suporte para Classificação do *Onset* em  
dados Temporais de Eletromiografia**

Dissertação apresentada para obtenção do grau de Mestre em Estatística,  
Matemática e Computação

**Luís Miguel Domingues Ferreira Silva**

Orientador: Professor Doutor Pedro Serranho

Co-orientador: Professor Doutor Pedro Pezarat Correia

Lisboa, 2014



## Resumo

Os estudos sobre parâmetros temporais em eletromiografia (EMG) focam a sua análise tendencialmente no *onset*, existindo uma escassez quanto à descrição e discussão dos fenômenos temporais. A dependência nos parâmetros dos algoritmos de detecção do *onset* e os diferentes métodos comprometem a reprodutibilidade de resultados. O objetivo deste trabalho é assim testar a performance de diferentes *features* no domínio do tempo na construção de modelos de Máquinas de Vetor Suporte (SVM) quanto à localização do *onset*. Sinal EMG de superfície foi recolhido durante o *swing* do golfe de 12 músculos (tronco e membro inferior) de 12 golfistas, 6 de handicap (Hc) baixo ( $Hc = 1.4 \pm 2.5 < 5$ ) e 6 de handicap alto ( $Hc = 24.6 \pm 4.2 > 18$ ). O sinal foi segmentado com janelas de 200 ms de 5 em 5 ms e depois foram extraídas as seguintes *features* no domínio do tempo: Valor Médio Absoluto, Comprimento do Formato da Onda, Diferença Absoluta do Desvio Padrão, Variância do EMG, Integral EMG e Detetor Logarítmico. As *features* foram selecionadas e ordenadas quanto à importância sendo construídos três conjuntos de 2, 4 e 6 *features* (F2, F4 e F6) para cada modelo. Após a realização de pesquisa de rede (*grid-search*), os melhores parâmetros quanto à precisão da classificação pelo modelo *radial basis function* (RBF) – SVM foram selecionados por *cross-validation*. O teste de Friedman foi aplicado para comparar os parâmetros ( $C, \gamma$ ) nos três conjuntos de *features* e a ANOVA mista para comparar a classificação e vetores suporte entre os grupos de *features* e grupos de handicap (alto Hc, baixo Hc e total). Verificamos que os grupos alto, baixo e total Hc apresentaram uma precisão de classificação de  $90.3 \pm 4\%$  (média $\pm$ desvio-padrão),  $90.8 \pm 4.9\%$  e  $89.4 \pm 3.7\%$  para F2,  $94.9 \pm 2.5\%$ ,  $95.0 \pm 3.3\%$ ,  $93.5\% \pm 3.2\%$  para F4 e  $95.2 \pm 2.4\%$ ,  $95.1 \pm 3.2\%$  e  $93.6 \pm 3.3\%$  para F6. Os valores dos parâmetros RBF, a classificação e o número de vetores suporte tende a ser similar entre F4 e F6, variando no entanto em relação a F2. Concluímos assim que quatro *features* garantem uma precisão na classificação superior a 90% em relação aos instantes de tempo classificados como antes e depois do *onset* podendo servir de base de construção de modelos SVM.

Palavras-chave: *Onset*, SVM, EMG, *features* domínio do tempo, aprendizagem

## Abstract

Studies on temporal parameters in electromyography (EMG) focus their analysis on onset. However, the description and discussion of temporal phenomena themselves is scarce and the results reproducibility is hard due to different parameters and methods. Thus, the aim of this work is to test the performance of different time-domain features building Support Vector Machines (SVM) models for onset detection. Surface EMG was collected from 12 muscles (trunk and lower limb) during the golf swing. Twelve golfers of two handicap (Hc) groups were recruited (6 low Hc =  $1.4 \pm 2.5 < 5$ ; 6 high Hc =  $24.6 \pm 4.2 > 18$ ). The signal was segmented with 200 ms windows, with a lag between windows of 5 ms followed by time-domain features extraction: Mean Absolute Value, Waveform Length, Difference Absolute Standard Deviation Value, Variance, Integrated EMG, and Integral logarithmic detector. The features were selected and ranked by relevance on three sets of 2, 4 and 6 features (F2, F4 and F6). After conducting grid-search for radial basis function (RBF) - SVM, the best parameters were selected for each model using cross-validation. The Friedman test was used to compare the parameters ( $C, \gamma$ ) of different models. A mixed ANOVA was performed to compare the support vector classification and interaction between features model and handicap groups (high Hc, Hc and low total). The high, low, and total Hc groups showed a classification accuracy of  $90.3 \pm 4\%$  (mean $\pm$ standard deviation),  $90.8\pm 4.9\%$  and  $89.4\pm 3.7\%$  for F2,  $94.9\pm 2.5\%$ ,  $95.0\pm 3.3\%$ ,  $93.5\%\pm 3.2\%$  for F4 and  $95.2\pm 2.4$ ,  $95.1\pm 3.2\%$  and  $93.6\pm 3.3\%$  to F6. RBF values of the parameters, classification and number of support vectors tends to be similar between F4 and F6, though varying in relation to F2. We conclude therefore that four features ensure an accuracy rate exceeding 90% in relation to the time classification as before and after the onset. Time-domain features could be a basis for constructing SVM classification models.

Keywords: Onset, SVM, EMG, time-domain features, learning

## **Agradecimentos**

O sentido agradecimento ao Professor Doutor Pedro Serranho pela orientação e acompanhamento neste caminho de aprendizagem que se concretizou neste trabalho. Obrigado pelas preciosas recomendações e conselhos.

Ao Professor Doutor Pedro Pezarat Correia pela disponibilidade, confiança e amizade com que tem acompanhado o meu percurso académico. Muito obrigado.

Ao Professor Doutor Jan Cabri que se preocupou em incluir recursos de projeto de investigação tornando possível a evolução deste trajeto académico.

A todos os professores do Mestrado em Estatística, Matemática e Computação da Universidade Aberta pelo excelente trabalho desempenhado nas diversas unidades curriculares, que igualmente culminam num produtivo ciclo de aprendizagem.

Aos golfistas que disponibilizaram o seu tempo deslocando-se ao laboratório de Comportamento Motor da Faculdade de Motricidade Humana participando em recolhas que são por si morosas e em condições laboratoriais exigentes.

Aos colegas do Laboratório de Comportamento Motor da Faculdade de Motricidade Humana conhecidos como o “grupo do golfe”, Dr. João Vaz, Professora Doutora Maria António Castro, Professor Doutor Orlando Fernandes, Mestre Sérgio Marta e Gustavo Reinaldo.

À Faculdade de Motricidade Humana, Universidade de Lisboa por cedência dos dados recolhidos no Laboratório de Comportamento Motor.

À Coordenação do Mestrado em Estatística, Matemática e Computação e Serviços Administrativos da Universidade Aberta pela rápida resposta sempre que esta foi necessária.

À Vilma, aos meus pais e demais familiares pela contribuição neste caminho de aprendizagem sem palavras que a possam quantificar.

*Deo Gratias*

# Índice

Resumo.....	iii
Abstract .....	iv
Agradecimentos .....	v
Índice.....	vii
Definições Operacionais .....	xi
Lista de Abreviaturas .....	xii
Índice Figuras .....	xiii
Índice Tabelas.....	xv
1. Introdução.....	1
2. Revisão de Literatura .....	5
2.1. Teoria Estatística da Aprendizagem .....	5
2.2. Considerações sobre a Aprendizagem Automática Supervisionada.....	11
2.3. Máquinas de Vetores Suporte.....	14
2.3.1. SVM com dados separáveis .....	15
2.3.2. SVM com dados não separáveis.....	21
2.3.2.1. SVM de margens suaves .....	22
2.3.2.2. SVM não linear e “kernel trick” .....	25
2.3.2.3. Alternativas e extensões SVM .....	28
2.4. A Eletromiografia .....	30
2.4.1. Conceito e caracterização.....	30
2.4.2. Fatores que influenciam o EMG.....	31

2.4.3.	Parâmetros temporais em EMG.....	33
2.5.	Reconhecimento de Padrões EMG através de SVM .....	35
2.5.1.	Etapas no reconhecimento de padrões EMG.....	35
2.5.2.	Segmentação dos dados .....	37
2.5.3.	Extração de <i>features</i> .....	38
2.5.3.1.	Domínio do tempo .....	39
2.5.3.2.	Domínio da frequência.....	40
2.5.3.3.	Domínio do tempo-frequência e outras aplicações.....	41
2.5.4.	Aplicação SVM em EMG .....	43
2.5.4.1.	Controlo mio-elétrico.....	43
2.5.4.2.	Diagnóstico de desordens neuromusculares.....	44
2.5.4.3.	Análise cinesiológica .....	46
3.	Método .....	49
3.1.	Apresentação do Problema.....	49
3.2.	Objetivos do Estudo.....	50
3.3.	Hipóteses .....	51
3.4.	Sujeitos.....	52
3.5.	Instrumentos .....	54
3.5.1.	Processamento da eletromiografia.....	54
3.5.2.	Processamento vídeo.....	55
3.6.	Segmentação e Extração de <i>Features</i> .....	55
3.7.	Máquinas de Vetores Suporte.....	60
3.8.	Tratamento Estatístico Complementar.....	61
3.9.	Desenho de Investigação.....	62

3.10. Limitações ao Estudo .....	63
4. Resultados .....	65
4.1. Seleção de <i>Features</i> .....	65
4.2. Parâmetros Custo $C$ e <i>Kernel</i> $\gamma$ .....	66
4.3. Classificação e Detecção do <i>Onset</i> .....	74
4.3.1. Determinação com <i>onset burst</i> igual a <i>onset peak</i> .....	76
4.3.2. Determinação do <i>onset burst</i> .....	78
4.4. Percentagem Vetores Suporte.....	81
5. Discussão.....	85
5.1. Conjunto de <i>Features</i> no Domínio do Tempo.....	86
5.2. Parâmetros SVM.....	88
5.3. Detecção do <i>Onset</i> e Precisão de Classificação .....	89
6. Conclusões e Recomendações.....	91
Referências .....	95
Apêndices .....	103
Consentimento Informado .....	103
Informação Para os Participantes .....	105
Caracterização dos Participantes.....	108
<i>Outputs</i> .....	110
Normalidade F2 por grupos.....	110
Normalidade F4 por grupos.....	111
Normalidade F6 por grupos.....	112
MANOVA não Paramétrica .....	113
Teste de Friedman para o parâmetro $C$ nos conjuntos de <i>features</i> alto handicap..	114

Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de <i>features</i> alto handicap..	115
Teste de Friedman para o parâmetro $C$ nos conjuntos de <i>features</i> baixo handicap.....	116
Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de <i>features</i> baixo handicap.....	117
Teste de Friedman para o parâmetro $C$ nos conjuntos de <i>features</i> ambos handicap.....	118
Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de <i>features</i> ambos handicap.....	119
Teste de Friedman para o parâmetro $C$ nos conjuntos de <i>features</i> todos modelos.....	120
Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de <i>features</i> todos modelos.....	121
ANOVA mista precisão SVM e comparações múltiplas .....	122
ANOVA mista vetores suporte e comparações múltiplas.....	125
Correlações de Pearson classificação vs vetores suporte.....	128

## Definições Operacionais

**Feature** – Termo atribuído a determinada entrada característica de um atributo cujos componentes vão ser catalogados por um vetor de rótulo  $\{-1,1\}$ . Usualmente é um vetor coluna com elementos que irão ser assinalados através de outro vetor que irá associar cada componente como pertencente a uma das classes.

**Onset** – Refere-se ao início de ativação do sinal EMG, ou seja, corresponde ao momento em que o músculo ativa.

**Onset burst** – Definição tradicional de *onset*, corresponde ao momento em que o músculo ativa mediante a primeira propagação dos potenciais de ação que é registado pelo EMG, independentemente da atividade muscular posterior.

**Onset peak** – Corresponde a qualquer início de ativação no registo EMG que antecede um período relevante de atividade. Pode corresponder ao início da atividade de maior relevância, a um instante *off / on* seguido de um pico de atividade muscular, ou ao próprio *onset burst* quando se verifica que antecede um pico de atividade bem definido.

## Lista de Abreviaturas

**BF** - Bicípite Femoral.

**CFS** - *Correlation-based Feature Selection*.

**DASDV** - *Difference Absolute Standard Deviation Value* (Diferença Absoluta do Desvio Padrão).

**EMG** - Eletromiografia.

**FS** - *Fisher Score*.

**GG** - Grande Glúteo.

**Hc** - Handicap.

**IEMG** - *Integrated EMG* (Integral do EMG).

**LOG** - *Logarithmic Detector* (Detetor Logarítmico).

**MAV** - *Mean Absolute Value* (Valor Absoluto Médio).

**MC** - Massa Comum.

**OE** - Oblíquo Externo.

**RBF** - *Radial Basis Function* (Função Base Radial).

**RF** - Reto Femoral.

**ST** - Semitendinoso.

**SVM** - *Support Vector Machines* (Máquinas de Vetor Suporte).

**VAR** - *Variance of EMG* (Variância do EMG).

**VE** - Vasto Externo.

**VI** - Vasto Interno.

**WL** - *Waveform Length* (Comprimento do Formato da Onda).

## Índice Figuras

Figura 1 – Classes de algoritmos de aprendizagem. ....	12
Figura 2 – Maximização das margens do hiperplano e a função $(x)$ . ....	17
Figura 3 – Maximização das margens do hiperplano com introdução variáveis de folga para o caso linear não separável. ....	23
Figura 4 – Fatores que influenciam o EMG de superfície. ....	32
Figura 5 – Fases de tratamento de EMG no reconhecimento de padrões ....	36
Figura 6 – Fluxograma do desenho de investigação sobre o tratamento EMG. ....	62
Figura 7 – Peso de cada feature na classificação segundo o Fisher Score. ....	65
Figura 8 – Relação do número de vezes que uma feature foi incluída no grupo mais robusto. ....	66
Figura 9 – Parâmetro $C$ segundo handicap e número de features. ....	70
Figura 10 – Parâmetro $\gamma$ segundo handicap e número de features. ....	70
Figura 11 – Pesquisa de grelha dos parâmetros $C$ e $\gamma$ com 4 features para o reto femoral esquerdo. ....	71
Figura 12 – Pesquisa de grelha dos parâmetros $C$ e $\gamma$ com 6 features para o reto femoral esquerdo. ....	72
Figura 13 – Pesquisa de grelha dos parâmetros $C$ e $\gamma$ com 4 features para a massa comum lado direito. ....	73
Figura 14 – Pesquisa de grelha dos parâmetros $C$ e $\gamma$ com 6 features para a massa comum lado direito. ....	73
Figura 15 – Classificação por grupo vs modelos. ....	75
Figura 16 – Determinação do onset para o Vasto Interno com modelo SVM-F4L para sujeitos de baixo handicap. ....	76
Figura 17 – Determinação do onset para o reto femoral esquerdo com modelo SVM-F4L para sujeitos de baixo handicap. ....	77
Figura 18 – Determinação do onset para o oblíquo externo esquerdo com modelo SVM-F6H para sujeitos de baixo handicap. ....	79

Figura 19 – Determinação do onset peak para a massa comum com modelo SVMT-F4T para sujeitos de baixo handicap .....	80
Figura 20 – Percentagem de Vetores Suporte necessários para cada modelo.....	81
Figura 21 – Diagrama de dispersão da classificação vs vetores suporte para F2.....	82
Figura 22 – Diagrama de dispersão da classificação vs vetores suporte para F4.....	83
Figura 23 – Diagrama de dispersão da classificação vs vetores suporte para F6.....	84

## Índice Tabelas

Tabela 1 – Caracterização dos participantes .....	53
Tabela 2 – Parâmetro $C$ para os modelos SVM.....	67
Tabela 3 – Parâmetro $\gamma$ para os modelos SVM.....	67
Tabela 4 – Teste Friedman e comparações múltiplas para o parâmetro $C$ .....	68
Tabela 5 – Teste de Friedman e comparações múltiplas para o parâmetro $\gamma$ .....	69







# 1. Introdução

Este estudo surge de um paradigma de investigação observado no estudo de parâmetros temporais em eletromiografia (EMG). Com o estudo dos parâmetros temporais procura-se conhecer quando o músculo ativa, estabelecendo através do registo EMG o tempo de atividade através de três variáveis de base, o *onset* (quando a atividade muscular é iniciada), o instante do pico máximo e o *offset* (quando o músculo finaliza atividade). O início e fim de atividade elétrica é relacionada com a contração do músculo, seja esta isométrica ou dinâmica. Ao estabelecer uma relação de quando o músculo “liga” e “desliga”, a mesma pode abranger uma perspetiva intramuscular, isto é, em relação ao próprio músculo, ou uma perspetiva intermuscular onde existe uma descrição sequencial do comportamento de vários músculos ao longo de determinado movimento. No entanto, existe uma carência de análises sobre como o Sistema Nervoso Central estabelece programas motores e daí organiza temporalmente o movimento, recorrendo-se à imagem sincronizada com o registo EMG (McGill et al., 2010).

O estudo de parâmetros temporais da atividade muscular tende a centrar-se sobre métodos de deteção do *onset* (ex.: Allison, 2003; Van Boxtel, 1993; Solnik, Rider, & Steinweg, 2010; Staude, Flachenecker, Daumer, & Wolf, 2001), os quais se distinguem em duas categorias de deteção: inspeção visual e algoritmos de deteção (Vaisman, Zariffa, & Popovic, 2010). Este é o ponto do qual emerge o supracitado paradigma. Para que um estudo sobre o *onset* seja considerado como aceitável deverá recorrer a algoritmos automáticos de deteção, mas a validade dos mesmos é aferida por inspeção visual, pois não existe um método de referência. Ainda, dentro das várias propostas de algoritmos para a deteção do *onset*, existe uma dependência nos parâmetros incluídos em relação às características do sinal (Hodges & Bui, 1996; Silva et al., 2013), como por exemplo, o desvio-padrão considerado em algoritmos de limiar e o rácio sinal/ruído. Mesmo com a realização de inspeção visual, o tipo de fenómeno fisiológico analisado não costuma ser descrito nem discutido em relação ao movimento. Este conjunto de limitações pesa na reprodutibilidade de resultados, a qual se encontra comprometida pelo tipo de algoritmo e parâmetros usados em cada método. Outra informação sobre fenómenos temporais que pode estar incluída no registo EMG tende a ser ignorada. Esta

problemática levou à introdução de um conceito mais lato de *onset*, o *onset peak* (Silva et al., 2013).

A conjugação dos dois tipos de dimensões de deteção, inspeção visual e algorítmica, transporta ao conceito de aprendizagem supervisionada, em que existe intervenção do investigador a dada altura do processo. Por sua vez, a aprendizagem supervisionada é um campo da Teoria Estatística da Aprendizagem (TEA), a qual está estreitamente associada ao reconhecimento de padrões em tarefas de classificação. Das várias técnicas utilizadas na TEA, as máquinas de vetores suporte (SVM – *Support Vector Machines*) têm apresentado uma boa performance, nomeadamente no reconhecimento de padrões no controlo mio-elétrico (Oskoei & Hu, 2008). As SVM são algoritmos de otimização matemática tendo sido inicialmente apresentadas no início da década de 90 (Boser, Guyon, & Vapnik, 1992) como um algoritmo de treino que maximiza a margem entre diferentes classes segundo padrões de treino. As SVM têm como base a TEA, uma vez que se pretende conseguir o melhor classificador possível, este que corresponderá aquele que apresenta menor risco empírico e satisfaça as respetivas restrições. Posteriormente foram apresentadas extensões a dados não separáveis (Cortes & Vapnik, 1995), ajuste em geral necessário para aplicações reais no processamento e tratamento de dados, pois os dados, por regra, não são linearmente separáveis.

O presente estudo procurou verificar o contributo da aprendizagem automática na deteção do *onset* em relação à atividade muscular de maior relevância no registo EMG, este recolhido durante o *swing* do golfe. Os modelos SVM tiveram a sua construção segundo a quantificação de entradas características (*features*) no domínio do tempo. Pretendeu-se conhecer a performance da aprendizagem automática na deteção do *onset* recorrendo a *features* no domínio do tempo.

Após esta introdução, o enquadramento teórico que suporta esta dissertação é apresentado. Introduce-se o conceito de TEA com uma retrospectiva histórica, passando pelo conceito de aprendizagem supervisionada até às SVM. A explanação matemática deste algoritmo de otimização está organizada em dois grandes grupos: SVM para dados separáveis e a sua extensão para dados não separáveis. Neste capítulo ainda são apresentados conceitos e limitações associados à EMG em geral, especificando posteriormente, a literatura sobre a deteção do *onset* e a aplicabilidade das SVM à

EMG. Neste último ponto, introduzem-se como grandes áreas de aplicação, o controlo mio-elétrico, diagnóstico de desordens neuromusculares e a análise cinesiológica.

O capítulo 3 que se segue apresenta o método, começando por introduzir o problema de estudo que consiste em quantificar o poder de *features* no domínio do tempo na classificação SVM para a deteção do *onset*. Serão dadas a conhecer as características dos participantes, assim como, todo o apetrechamento que foi realizado ao enquadramento teórico, desde procedimentos de recolha, processamento de dados e demais tratamento estatístico.

O capítulo 4 refere-se aos resultados deste estudo, estando subdividido em seleção de *features* onde são apresentados os conjuntos elaborados, parâmetros custo  $C$  e *kernel*  $\gamma$  baseados na pesquisa de grelha, classificação e deteção do *onset* propriamente dita e número relativo de vetores suporte. Quanto à classificação e deteção do *onset*, é realizada a apresentação dos fenómenos *onset* identificados. A classificação e o número de vetores suporte ainda incluem a correlação entre ambos face ao tipo de conjunto de *features* (F2, F4 e F6) em análise.

Após os resultados segue-se o capítulo que os discute à luz da literatura da especialidade. Apesar deste tipo de metodologia em que se combina SVM com a deteção do *onset* ser aqui introduzida, considerou-se a demais literatura sobre *features* no domínio do tempo aplicadas a EMG, neste âmbito quais os parâmetros *radial basis function* (RBF) e as considerações a tomar face à precisão deste tipo de algoritmos.

Termina-se com as conclusões que se retiram do presente estudo e implicações práticas sobre como aplicar o modelo e recomendações para futuros estudos.



## 2. Revisão de Literatura

### 2.1. Teoria Estatística da Aprendizagem

A Teoria Estatística da Aprendizagem (TEA) está intimamente associada ao reconhecimento de padrões e a funções discriminantes. Na pretensão de entender a gênese desta tarefa matemática que se baseia em classificar e discriminar, torna-se necessário recuar à década de 30 com o trabalho de Ronald Aylmer Fisher (Fisher, 1936). Neste trabalho, intitulado *The use of multiple measurements in taxonomic problems*, Fisher desenvolve os procedimentos para classificar espécies de plantas (*Iris setosa* e *Iris versicolor*) face a quatro medidas sobre o comprimento e largura das pétalas e sépalas. Se consideramos as quatro medidas em  $n$  medidas, tem-se como pressuposto conhecer a função linear que serve de base aos problemas de regressão:

$$X = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n$$

com  $n = 4$ , pretendendo-se conhecer a função linear dessas medidas que corresponde à função que irá *maximize the ratio of the difference between the specific means to the standard deviations within species* (Fisher, 1936, p.179, *ipsis verbis*). O conceito base consiste em prever determinada variável com base em outras medidas. Por sua vez, a diferença entre as médias de  $X$  para as duas espécies seria dada por:

$$D = \beta_1 d_1 + \beta_2 d_2 + \dots + \beta_n d_n$$

com  $d_j$  a representar a diferença entre as duas espécies das médias da  $j$ -ésima medida e a variância de  $X$  proporcional a

$$S = \sum_{p=1}^n \sum_{q=1}^m \beta_p \beta_q S_{pq}$$

com  $p$  e  $q$  a corresponder a cada uma das  $n, m$  medidas e  $S_{pq}$  o somatório dos quadrados ou produto dos desvios de cada medida entre as duas espécies. A melhor função discriminatória seria dada por aquela que maximiza o rácio  $\frac{D^2}{S}$  em função dos parâmetros  $\beta_1, \beta_2, \dots$  e  $\beta_n$  de forma independente para cada  $\beta$  com:

$$\frac{1}{2} \cdot \frac{\partial D}{\partial \beta} = \frac{S}{D} \cdot \frac{\partial D}{\partial \beta}$$

Os resultados obtidos através do rácio entre a diferença das médias totais e o somatório dos quadrados também permitem quantificar a probabilidade de desvios e erros na predição obtida. No trabalho de Fisher ainda foi apresentada a possibilidade de estender a mais de duas populações.

Em 1943, McCulloch e Pitts introduziram o conceito de redes neuronais como máquinas de computação. Em 1949, Hebb postula a primeira regra de aprendizagem auto-organizada e em 1958 Rosenblatt propõe o método perceptrão como o primeiro modelo de aprendizagem supervisionada (Haykin, 2001). O perceptrão é a forma mais simples de rede neuronal para classificar padrões linearmente separáveis desenvolvendo o *teorema de convergência de perceptrão*.

Outra obra que devemos considerar com enorme contributo para a TEA e muito importante para as SVM é a *Theory of Reproducing kernels* de Aronszajn (1950), vindo mais tarde assumir um papel primordial na classificação não-linear. Considerando uma classe linear  $F$  de funções  $f(x)$  em que  $F$  é uma classe complexa, ou seja, admite o produto por constantes complexas e que  $f \in F$  é definida pela norma  $\|f\|$ . Tem-se então a forma quadrática hermitiana ( $f$ ) em que  $\|f\|^2 = Q(f)$ . Desta forma,  $Q(f)$  é denominada por função funcional quadrática hermitiana se para cada constante  $\beta_1$  e  $\beta_2$  e funções  $f_1$  e  $f_2$  for obtida a função:

$$Q(\beta_1 f_1 + \beta_2 f_2) = |\beta_1|^2 Q(f_1) + \beta_1 \beta_2 Q(f_1, f_2) + |\beta_2|^2 Q(f_2) + \beta_1 \beta_2 Q(f_2, f_1)$$

Consoante o conceito de predição evolui, outros dois termos estão associados a este processo: a inferência e a mineração de dados (*data mining*). No entanto, o conceito de TEA propriamente dito foi introduzido nos anos 60 com o contributo de importantes

trabalhos onde constam a interpretação geométrica dos produtos internos *kernel* no espaço característico (Aizerman, Braverman, & Rozonoer, 1964) e o desenvolvimento matemático do denominado *Generalized Portrait algorithm* (Vapnik & Chervonenkis, 1964). No ano seguinte é introduzido o Teorema de Cover (Cover, 1965) sendo debatida a separabilidade no espaço de entradas mediante a construção de hiperplanos e distância das respectivas margens face aos vetores que se encontram sobre os mesmos. Apesar da importância destes trabalhos, a TEA apenas ganha popularidade na década de 90 com a introdução das SVM, nomeadamente em 1992 (Boser et al., 1992), e depois com a extensão para as margens suaves em 1995 (Cortes & Vapnik, 1995). Após estes estudos, a aplicabilidade da aprendizagem automática assume uma preponderância no reconhecimento de padrões.

A aprendizagem automática estuda o processo de aprendizagem em abstrato, isto é, como pode um computador aprender tarefas específicas seguindo algoritmos que advêm de treino. A aprendizagem automática de máquinas tem raízes na inteligência artificial, na estatística e na computação, mas ao contrário da primeira, não procura explicar ou gerar uma capacidade intelectual na máquina. O principal objetivo passa por desenvolver mecanismos (algoritmos) segundo os quais determinadas tarefas possam ser aprendidas no âmbito da inferência indutiva e com capacidade de generalização (Luxburg & Schölkopf, 2009). O problema com maior ênfase de estudo é o da classificação binária onde encontramos dois tipos de espaços, as entradas  $\mathcal{X}$  e os rótulos de classificação  $\mathcal{Y}$ , que no caso binário apenas podem assumir dois valores.

Quando estamos perante um problema de classificação de padrões, existe um objeto que se pretende classificar numa de duas classes, estas rotuladas por “-1” e “1”. Para decidir a qual das classes esse objeto irá pertencer temos que assumir que possuímos um conjunto de medidas sobre as características do mesmo. Cada característica é então, representada por um número real, podendo, no entanto, existir o constrangimento de algumas destas características não estarem devidamente representadas por um número. São estes os pressupostos que servem de suporte à grande maioria de métodos da Teoria Estatística da Aprendizagem. O conjunto de  $n$  características de um objeto de valores  $x_1, x_2, \dots, x_n$  pode ser agrupado de forma a constituir um vetor característico  $\mathbf{x}$ , que será um vetor no espaço  $n$ -dimensional  $\mathbb{R}^n$  denominado por espaço característico ou *feature space* (Burges, 1998; Kulkarni & Harman, 2011; Schölkopf & Smola, 2002). Desta forma, o objetivo da TEA pode ser

definido dado um conjunto de dados  $x_i$  representados num espaço (vetorial) e a respetiva classificação por classe  $y_i$ , de forma a determinar um classificador capaz de indicar a classe de um novo elemento no espaço (vetorial) dos dados com a menor percentagem de erro de classificação possível. O modelo de estimativa da função na TEA comum a problemas como o reconhecimento de padrões, estimação por regressão e estimação da densidade pode ser desenvolvido em três componentes fundamentais (Vapnik, 1999):

- a) Gerador de vetores aleatórios  $\mathbf{x} \in \mathbb{R}^n$ , elaborado de forma independente a partir de uma função de probabilidade cumulativa  $P(\mathbf{x})$ , fixa e desconhecida;
- b) Supervisor que devolve um vetor de saída  $y$  para cada vetor de entrada  $\mathbf{x}$  tendo em conta uma determinada função distribuição condicional  $P(y|\mathbf{x})$ , igualmente fixa e desconhecida;
- c) Uma máquina de aprendizagem capaz de implementar um conjunto de funções  $y = f(\mathbf{x}, \alpha)$ ,  $\alpha \in \Lambda$ , com  $f$  a representar o classificador e  $\alpha$  os parâmetros da função a aprender,  $y$  a classe de classificação e  $\mathbf{x}$  o elemento no espaço dos dados, com  $\Lambda$  a representar o conjunto de parâmetros possíveis.

Cabe escolher, a partir de um conjunto de funções, a que prevê a melhor resposta, seleção esta que depende da definição de  $n$  observações aleatórias, independentes e identicamente distribuídas denominadas por conjunto de treino, em que  $P(\mathbf{x}, y) = P(\mathbf{x})P(y|\mathbf{x})$ . O par  $(\mathbf{x}, y)$  é uma amostra independente desta distribuição  $P$ , uma amostragem designada como amostragem *iid* (independentes e identicamente distribuídos). Não são realizados pressupostos sobre os espaços de  $\mathbf{x}$  e  $y$ , mas sim, na forma como são gerados os pontos de treino (Luxburg & Schölkopf, 2009):

a) *A não realização de pressupostos sobre  $P$* , podendo ser realizada qualquer distribuição. É assumido que a distribuição probabilidade pertence a uma certa família de distribuições com o objetivo de estimar o parâmetro desta distribuição.

b) *Rótulos não determinísticos devido a ruídos ou a classes sobrepostas*. A distribuição  $P$  refere-se às entradas  $\mathbf{x}$  e aos rótulos  $y$ , pelo que estes não correspondem a uma função determinística de  $x_i$  devido à possibilidade de ruído associado à errada

atribuição do rótulo e/ou pela possibilidade de existirem duas classes com o mesmo rótulo ou sobreposição de classes.

c) *Amostragem independente*. Depende do problema proposto, a amostragem deverá garantir representatividade populacional em relação ao rótulo de dados não induzindo numa determinada direção aquando a aplicação da aprendizagem. Existem algumas áreas em que este pressuposto tende a ser ignorado como a previsão de séries temporais onde as entradas são geradas a partir de sobreposição de janelas temporais.

d) *A distribuição  $P$  é fixa*, não é assumida qualquer ordenação particular dos exemplos de treino e a distribuição de probabilidade subjacente não se altera ao longo do tempo. No entanto, no caso de séries temporais podemos não ter este pressuposto como inteiramente verdadeiro.

e) *A distribuição  $P$  é desconhecida no momento da aprendizagem*. A aprendizagem acontece por não se conhecer  $P$ , o acesso à distribuição depende dos exemplos de treino.

Para a maioria das aplicações, a classe a que determinado objeto pertence não é definida apenas pelo vetor característico (*feature vector*) devido essencialmente a duas razões (Kulkarni & Harman, 2011): 1. geralmente os recursos medidos não capturam todas as propriedades do objeto que são importantes para a classificação; 2. devido ao ruído existente nos valores das entradas características que dependem da aplicação e das medidas. Daí a necessidade de formulação estatística para o problema de reconhecimento de padrões.

Dentro de várias possibilidades de regras de decisão para a classificação em classes binárias “-1” ou “1”, interessa escolher aquela que melhor realiza a tarefa de classificação dos valores do vetor das entradas  $\mathbf{x}$ . Esta tarefa irá corresponder ao mapeamento  $f : R^n \rightarrow \{-1,1\}$  para que  $f(\mathbf{x})$  represente o rótulo  $y$ .

O conceito de problema de aprendizagem corresponde à função de aproximação que explica a forma como determinada amostra é mapeada, levando à seguinte definição:

**Definição 1** (Problema de Aprendizagem):

*O problema de aprendizagem é encontrar uma relação funcional desconhecida com  $f : \mathbb{R}^n \rightarrow \{-1,1\}$  entre os objetos  $\mathbf{x} \in \mathcal{X}$  e os rótulos de saída  $y \in \mathcal{Y}$  baseado unicamente numa amostra  $(\mathbf{x}, y) = ((x_i, y_i), \dots, (x_n, y_n)) \in (\mathcal{X}, \mathcal{Y})^n$  de dimensão  $n \in \mathbb{N}$  de dados independentemente e identicamente distribuídos (i.i.d.) a partir de uma distribuição  $P(\mathbf{x}, y)$  desconhecida. Se o espaço de saída  $\mathcal{Y}$  contém um número finito  $|\mathcal{Y}|$  de elementos, a tarefa é chamada um problema de aprendizagem de classificação.*

(Herbrich, 2002, p. 18)

Outra definição presente refere-se às entradas características (*features*) e espaço das entradas características (*feature space*), nomeadamente para modelar a semelhança entre objetos através de uma função do produto interno.

**Definição 2** (entradas e espaço característico)

*Segundo um classificador representado pela função  $f : \mathbb{R}^n \rightarrow \{-1,1\}$  que mapeia cada componente dos objetos  $\mathbf{x} \in \mathcal{X}$ , cada  $x_i$  é chamado de entrada característica (*feature*). As várias *features* vão construir um espaço característico de dimensionalidade  $n$ , pelo que o espaço euclidiano característico será representado por  $\mathcal{X} \subset \mathbb{R}^n$ . Existindo a necessidade de recorrer a produtos internos num espaço de dimensão superior com mapeamento  $\Phi: \mathcal{X} \rightarrow \mathcal{H}$  (*kernel trick*), então  $\mathcal{H}$  é denominado por espaço característico.*

(adaptado Herbrich, 2002, p. 20)

Para facilitar a distinção entre componentes de um vetor e dos vários vetores que representam *features*, a partir deste ponto procede-se a uma correção terminológica, em que  $x_i$  corresponderá a cada componente do vetor  $\mathbf{x}$ , sendo que  $\mathbf{x}$  será a representação vetorial de  $\mathbf{x}$  no espaço característico (*feature space*). Desta forma, ao inserir dados através de  $\Phi$  para o espaço  $\mathcal{H}$ , temos o mapeamento  $\Phi: \mathcal{X} \rightarrow \mathcal{H}$  com  $x \mapsto \mathbf{x}$ .

Ao compreender-se os procedimentos de base para a construção de um algoritmo de aprendizagem pretende-se que o mesmo possua a capacidade de inferir através de um conjunto de dados que podem ser vistos como um exemplo de treino. Basicamente, parte-se do pressuposto que a robustez do algoritmo dependerá da capacidade do mesmo detetar as regularidades intrínsecas dos dados (Scholkopf, Burges, & Vapnik, 1995). Será esta capacidade que determinará o sucesso da aprendizagem automática.

## **2.2. Considerações sobre a Aprendizagem Automática Supervisionada**

A habilidade de aprender é fundamental para que o desempenho dos algoritmos seja robusto. Porém, o processo de aprendizagem está dependente dos critérios paradoxais onde se inclui, ou seja, da relação ambiente e adaptação. A aprendizagem automática como ramo da inteligência artificial centra-se no objetivo de desenvolver algoritmos que possibilitem as máquinas de realizar tarefas cognitivas. Haykin (2001) refere que um sistema de inteligência artificial deverá abarcar três capacidades primordiais: (1) armazenar conhecimento, (2) aplicar esse conhecimento na resolução de problemas, e (3) adquirir novo conhecimento através da experiência. Para atingir os objetivos da aprendizagem automática, duas grandes dimensões de problemas de aprendizagem podem ser distinguidas na literatura (Haykin, 2001; Herbrich, 2002; Luxburg & Schölkopf, 2009; Rojas, 1996): a aprendizagem não supervisionada e a aprendizagem supervisionada, conforme ilustrado na figura 1.

Na aprendizagem não supervisionada, as regras de decisão usadas pelo algoritmo não dependem da intervenção de um professor ou do investigador ao longo do processo, ou seja, não existem exemplos rotulados e previamente aprendidos. Usualmente, o investigador nem sabe quantas classes ou componentes discriminatórias vão ser produzidas após a utilização do algoritmo não supervisionado. Como exemplos de aprendizagem não supervisionada podem ser destacados a classificação por *clusters* e os modelos de variáveis latentes como a análise factorial.

A aprendizagem supervisionada, também designada como aprendizagem com professor, é constituída por um processo de treino através de exemplos de entrada-saída (rótulos) retirados de um ambiente desconhecido. Neste caso, os parâmetros são ajustados mediante a resposta desejada e o erro associado. Apesar do ilustrado na figura

1, é possível encontrar na literatura a aprendizagem por reforço associada à aprendizagem sem professor (Haykin, 2001). Este paradigma depende se a aprendizagem sem reforço consiste num mapeamento de entrada e saída relacionado com a interação contínua do ambiente, onde determinado sistema tem em conta uma sequência temporal de estímulos, ou se, pelo facto de existir um reforço é assumida a presença de professor, não sendo, no entanto, obtidas as respostas previamente. A categorização da aprendizagem por reforço dentro da aprendizagem supervisionada deve-se por ser utilizada após a apresentação do exemplos de treino, podendo ser produzido o resultado pretendido ou não. Na aprendizagem com correção considera-se o erro em conjunto com o vetor de entrada, sendo determinada a magnitude de correção desse erro (Rojas, 1996).

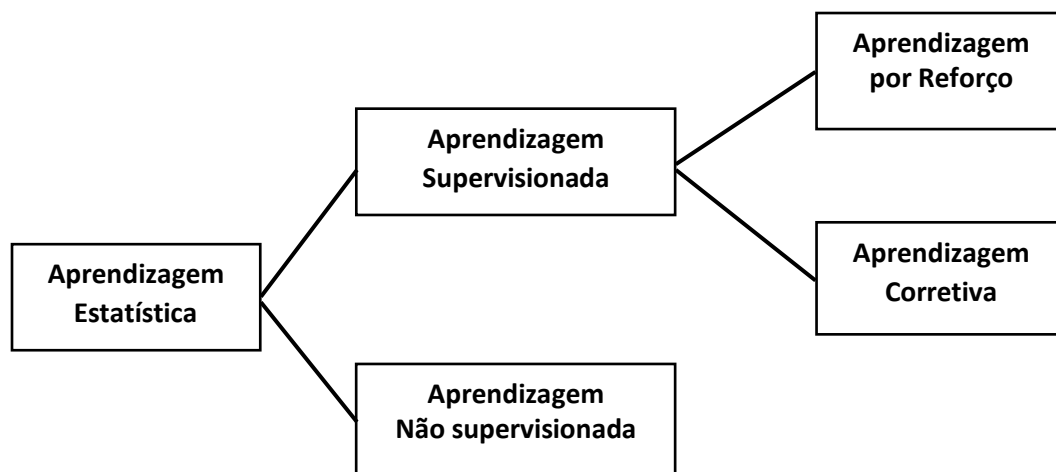


Figura 1 – Classes de algoritmos de aprendizagem (Rojas, 1996, p.79).

Na aprendizagem supervisionada existe um par que consiste no objeto de entrada e um valor rótulo de saída pertencente a uma classe. É esperado que após o algoritmo aprender através de exemplos consiga classificar corretamente novos dados. O investigador vai atuar em determinada altura do processo.

Quando nos reportamos às SVM como aprendizagem automática supervisionada, o esquema apresentado na figura 1 pode ser considerado, visto contemplar abordagens que dependem da separabilidade ou não separabilidade dos dados e demais extensões, como se passará a abordar. A probabilidade  $P(\mathbf{x}, y)$  descreve a relação entre os dados e os rótulos procurando a função  $f(\mathbf{x}, \alpha)$  que se traduza no

menor erro possível, ou seja, que forneça o menor risco expresso pelo risco esperado  $R(\alpha)$  (Burges, 1998; Scholkopf et al., 1995)

$$R(\alpha) = \int \frac{1}{2} |y - f(\mathbf{x}, \alpha)| dP(\mathbf{x}, \alpha) \quad (1)$$

Como  $P(\mathbf{x}, \alpha)$  é desconhecido utiliza-se o princípio da indução para inferir uma função  $f(\mathbf{x}, \alpha)$  para minimizar o erro, isto é, procede-se à minimização do risco empírico  $R_{emp}(\alpha)$  considerando  $l$  exemplos de treino.

$$R_{emp}(\alpha) = \frac{1}{2l} \sum_{i=1}^l |y_i - f(\mathbf{x}_i, \alpha)| \quad (2)$$

Nesta expressão não aparece a distribuição de probabilidade, mas sim a escolha de determinado  $\alpha$  para o par  $(\mathbf{x}_i, \alpha)$ . A parcela  $\frac{1}{2} |y - f(\mathbf{x}, \alpha)|$  é denominada função custo que neste caso apenas assume os valores 0 e 1. Tendo como base o princípio da minimização do risco estrutural, opta-se por um  $\eta \in [0,1]$  com a probabilidade de pelo menos  $1 - \eta$ , sendo obtido o seguinte limite (Burges, 1998; Scholkopf et al., 1995)

$$R(\alpha) \leq R_{emp}(\alpha) + \sqrt{\frac{h \left( \log \left( \frac{2l}{h} \right) + 1 \right) - \log \left( \frac{\eta}{4} \right)}{l}} \quad (3)$$

onde  $h$  é um valor inteiro não negativo denominado por dimensão Vapnik Chervonenkis (VC). A dimensão VC é uma propriedade de um conjunto de funções  $\{f(\alpha)\}$  sendo definida pelas classes que a integram, correspondendo ao número máximo de pontos de treino que podem ser representados por  $\{f(\alpha)\}$ . Desta forma, podem ser destacados o seguinte Teorema e respetivo corolário (Burges, 1998):

**Theorem 1:** Consider some set of  $m$  points in  $\mathbb{R}^n$ . Choose any one of the points as origin. Then the  $m$  points can be shattered by oriented hyperplanes if and only if the position vectors of the remaining points are linearly independent.

**Corollary 1:** The VC dimension of the set of oriented hyperplanes in  $\mathbb{R}^n$  is  $n+1$ , since we can always choose  $n + 1$  points, and then choose one of the points as origin, such that the position vectors of the remaining  $n$  points are linearly independent, but can never choose  $n + 2$  such points (since no  $n + 1$  vectors in  $\mathbb{R}^n$  can be linearly independent).

(Burges, 1998, p.4, *ipsis verbis*)

### 2.3. Máquinas de Vetores Suporte

As máquinas de vetores suporte (SVM – *Support Vector Machines*) são algoritmos de otimização matemática baseados na aprendizagem supervisionada. Esta aplicação foi apresentada em 1992 (Boser et al., 1992) como um algoritmo de treino que maximiza a margem entre os padrões de treino de diferentes classes. As SVM são inicialmente apresentadas como uma técnica de classificação para o reconhecimento de padrões, baseando-se na minimização do erro esperado de generalização através do método *leave-one-out* e através da dimensão VC que avalia a capacidade de classificação de um algoritmo. Basicamente, como as SVM têm como base a TEA, pretende-se conseguir o melhor classificador possível que corresponderá àquele que apresenta menor risco empírico e satisfaça as respectivas restrições com uma dimensão VC pequena. Na mesma década são apresentadas extensões a dados não separáveis (Cortes & Vapnik, 1995) ajustando mais à realidade no processamento e tratamento de dados.

### 2.3.1. SVM com dados separáveis

As SVM lineares com dados linearmente separáveis também são designadas por margens rígidas. O caso mais simples de exemplificar é o caso de margens linearmente separáveis para duas classes por se tratar meramente de uma dicotomia, pelo que começaremos por considerar este contexto. Considerando um conjunto de treino com  $n$  objetos  $\mathbf{x}_i \in \mathcal{X}$  com os rótulos de dados  $y_i \in \mathcal{Y}$ , em que  $\mathbf{x}_i \in \mathbb{R}^n$  e  $y_i \in \{-1,1\}$ , para  $i=1,2,\dots,n$ , o conjunto de treino é linearmente separável por um hiperplano que distinga as duas classes A e B (positiva e negativa, respetivamente). O conjunto de treino terá de considerar que para  $(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \dots, (\mathbf{x}_n, y_n)$ , tem-se que (Boser et al., 1992):

$$\begin{cases} y_i = +1 & \text{se } \mathbf{x}_i \in \textit{classe } \mathcal{X}(A) \\ y_i = -1 & \text{se } \mathbf{x}_i \in \textit{classe } \mathcal{X}(B) \end{cases} \quad (4)$$

Basicamente o que é pretendido é conhecer qual o melhor hiperplano que separa linearmente as duas classes, denominado por hiperplano ótimo. Este hiperplano será aquele cujas margens separadoras são maiores no seu espaço geométrico, pelo que se coloca um problema de maximização das margens. Os pontos  $\mathbf{x}$  que se encontram sobre o hiperplano terão que satisfazer a condição (Boser et al., 1992; Burges, 1998; Cortes & Vapnik, 1995; Müller, Mika, Rätsch, Tsuda, & Schölkopf, 2001):

$$\mathbf{w} \cdot \mathbf{x} + b = 0 \quad (5)$$

onde  $\mathbf{w}$  corresponde ao vetor normal ao hiperplano,  $|b|/\|\mathbf{w}\|$  representa a distância perpendicular do hiperplano à origem,  $\|\mathbf{w}\|$  é a norma euclidiana de  $\mathbf{w}$ , e  $\mathbf{w} \cdot \mathbf{x}$  é o produto escalar entre os vetores  $\mathbf{w}$  e  $\mathbf{x}$ . No caso de dados linearmente separáveis, o algoritmo do vetor de suporte escolhe o **hiperplano de separação com maior margem** (Burges, 1998; Cristianini & Shawe-Taylor, 2000) que satisfaz:

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq +1 & \text{para } y_i = +1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1 & \text{para } y_i = -1 \end{aligned} \quad (6)$$

assim, sendo  $\mathbf{w}$  o vetor ortogonal ao hiperplano, tendo o hiperplano definido por  $(\mathbf{w}, b)$  que separar os dados de treino, o classificador a obter pode ser representado pela função decisão  $D(x)$ :

$$D(x) = \text{sign}(\mathbf{w} \cdot \mathbf{x} + b). \quad (7)$$

Interessa o hiperplano com maior margem, pelo que é preciso conhecer o vetor  $\mathbf{w}$  que maximiza a margem. Sejam os hiperplanos que definem as classes dados por

$$H_{-1} = \{x \in R^n: \mathbf{w} \cdot \mathbf{x} + b = -1\}, \quad H_1 = \{x \in R^n: \mathbf{w} \cdot \mathbf{x} + b = 1\}, \quad (8)$$

então a distância entre os hiperplanos é dada por  $\frac{2}{\|\mathbf{w}\|}$ . Ao minimizar  $\|\mathbf{w}\|^2$  podemos encontrar o par de hiperplanos com a máxima margem sujeita às restrições (Burges, 1998; Cristianini & Shawe-Taylor, 2000):

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0 \quad \forall_i \quad (9)$$

A figura 2 ilustra a discriminação de dados separáveis pela maximização das margens e a função decisão  $D(x) = \mathbf{w} \cdot \mathbf{x} + b$ . Os vetores a negro que se encontram sobre os hiperplanos  $H_{-1}$  e  $H_1$  que formam as margens maximizadas são os vetores suporte.

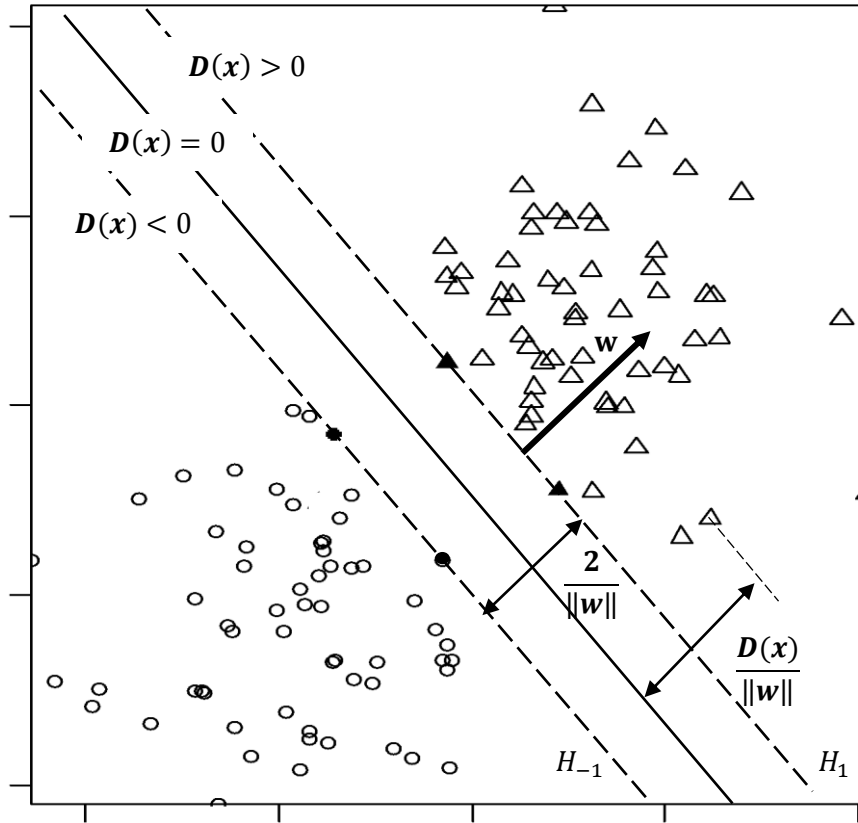


Figura 2 – Maximização das margens do hiperplano e a função  $D(x)$ .

Supondo que obtemos para os hiperplanos  $H_{-1}$  e  $H_1$  os vetores suporte  $x_{-1}$  e  $x_1$ , respectivamente. A projeção da distância entre os hiperplanos  $H_{-1}$  e  $H_1$  pode ser representada pela seguinte equação (Lorena & Carvalho, 2007):

$$(x_{-1} - x_1) \left( \frac{\mathbf{w}}{\|\mathbf{w}\|} \cdot \frac{x_{-1} - x_1}{\|x_{-1} - x_1\|} \right) \quad (10)$$

Como pretendemos a diferença  $x_{-1} - x_1$ , face à condição que  $H_{-1} : \mathbf{w} \cdot x_{-1} + b = 1$  e  $H_1 : \mathbf{w} \cdot x_1 + b = -1$ , a diferença será  $\mathbf{w} \cdot (x_{-1} - x_1) = 2$ , obtendo-se:

$$\frac{2(x_{-1} - x_1)}{\|\mathbf{w}\| \|x_{-1} - x_1\|} \equiv \frac{2}{\|\mathbf{w}\|} \quad (11)$$

Reforçando o já referido, o hiperplano ótimo será aquele que deriva das desigualdades  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0$  em que a norma  $\|\mathbf{w}\|$  é mínima, ou seja, um problema de otimização com as condições :

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 \quad (12)$$

com as restrições

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0,$$

que por sua vez, pode ser convertido num problema de otimização quadrática pela introdução de multiplicadores de Lagrange positivos (Burgess, 1998; Müller et al., 2001). A introdução dos multiplicadores de Lagrange deve-se a duas razões (Burgess, 1998): 1. porque as restrições enunciadas anteriormente vão ser substituídas por restrições nos multiplicadores de Lagrange que permitem maior simplicidade de manuseamento para o algoritmo de otimização; 2. o formato de produtos internos entre os vetores determinados pelos dados de treino é fundamental para a generalização, no caso de dados não lineares, como veremos adiante. Então, torna-se necessário associar às restrições o multiplicador  $\alpha_i$ , obtendo-se:

$$\alpha_i(y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) > 0 \quad (13)$$

Como a restrição se refere a um conjunto de restrições consoante os diferentes pontos no espaço  $\mathcal{X}$ , realiza-se a soma dos mesmos:

$$\sum_{i=1}^n \alpha_i(y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1) \quad (14)$$

Este termo, por sua vez, é agregado à função objetivo, originando a função Lagrangiana na sua formulação primordial (Burges, 1998; Cristianini & Shawe-Taylor, 2000; Müller et al., 2001):

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1) \quad (15)$$

Sendo um problema de minimização torna-se necessário minimizar  $\mathcal{L}(\mathbf{w}, b, \alpha)$  em relação a  $\mathbf{w}$  e  $b$  que significa as derivadas parciais de  $\mathcal{L}(\mathbf{w}, b, \alpha)$  serem igualadas a zero, tendo presentes as restrições  $\alpha_i \geq 0$ . Se a restrição  $y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq 0$  é violada, então  $y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 < 0$ , neste caso  $\mathcal{L}(\mathbf{w}, b, \alpha)$  pode aumentar pelo aumento do parâmetro  $\alpha_i$ . Ao mesmo tempo,  $\mathbf{w}$  e  $b$  terá de ser escolhido de forma a diminuir  $\mathcal{L}(\mathbf{w}, b, \alpha)$  e evita-se que esta função se torne num número arbitrariamente elevado e negativo (Schölkopf & Smola, 2002), obtendo-se um “ponto sela” segundo:

$$\frac{\partial}{\partial b} \mathcal{L}(\mathbf{w}, b, \alpha) = 0 \quad e \quad \frac{\partial}{\partial \mathbf{w}} \mathcal{L}(\mathbf{w}, b, \alpha) = 0 \quad (16)$$

Na formulação dual pretende-se minimizar uma nova função de custo. Com vista à formulação dual, começamos por impor (Boser et al., 1992; Burges, 1998; Chen, Lin, & Schölkopf, 2005; Lorena & Carvalho, 2007; Schölkopf & Smola, 2002):

$$\sum_{i=1}^n \alpha_i y_i = 0 \quad (17)$$

$$\mathbf{w} = \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i \quad (18)$$

Substituindo estas equações em  $L(\mathbf{w}, b, \alpha)$  obtém-se o problema de otimização com a seguinte formulação dual:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (19)$$

com as restrições:

$$\begin{cases} \alpha_i \geq 0, & \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (20)$$

Trocando os sinais, será o mesmo que obter:

$$\min_{\alpha} \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) - \sum_{i=1}^n \alpha_i \quad (21)$$

Pelas condições de Karush-Kuhn-Tucker (KKT) (Kuhn & Tucker, 1951) vamos obter o ponto sela que deriva de  $\mathcal{L}$  em respeito às variáveis primais por minimização determinando  $\mathbf{w}$  em que o vetor suporte será uma expansão do subconjunto de treino, em que  $\alpha_i$  é diferente de zero, ou seja, os dados que possuem  $\alpha_i > 0$  determinam as margens  $H_{-1}$  e  $H_1$  do hiperplano por se encontrarem sobre os mesmos. Os restantes dados da equação  $\alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0$  consideram  $\alpha_i = 0$  não sendo utilizados para o cálculo de  $\mathbf{w}$  (Burges, 1997; Chen, Lin, & Schölkopf, 2005; Hofmann, Schölkopf, & Smola, 2008; Schölkopf & Smola, 2002). O valor de  $b$  é calculado através dos vetores suporte expressos por  $\alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1) = 0$  (Lorena & Carvalho, 2007; Schölkopf & Smola, 2002):

$$b = \frac{1}{nSV} \sum_{\mathbf{x}_i \in SV} \frac{1}{y_j} - \mathbf{w} \cdot \mathbf{x}_j \quad (22)$$

em que  $nSV$  corresponde ao número de vetores suporte, podendo a equação ser expandida pela substituição de  $\mathbf{w}$  obtendo-se:

$$b = \frac{1}{nSV} \sum_{\mathbf{x}_i \in SV} \left( \frac{1}{y_j} - \sum_{\mathbf{x}_i \in SV} \alpha_i y_i \mathbf{x}_i \cdot \mathbf{x}_j \right) \quad (23)$$

Pode-se referir que é obtida a função linear classificadora  $g(x)$  que representa o hiperplano que separa os dados pela maior margem ao ter em conta aquele com maior capacidade de generalização, característica esta que diferencia as SVM lineares de margens rígidas das redes neurais perceptrão (Chen, Lin, & Schölkopf, 2005; Lorena & Carvalho, 2007; Schölkopf & Smola, 2002).

$$g(x) = \text{sgn}(f(x)) = \text{sgn} \left( \sum_{\mathbf{x}_i \in SV} \alpha_i y_i (\mathbf{x}_i \cdot \mathbf{x}_j) + b \right) \quad (24)$$

No entanto, a existência de dados linearmente separáveis não é comum em tratamentos reais, onde a existência de diversas fontes de ruído dificulta a aplicação deste método com as restrições impostas.

### 2.3.2. SVM com dados não separáveis

O algoritmo dicotómico de dados separáveis é fundamental para a compreensão das SVM, pois pode ser considerado como ponto de partida. No entanto, a realidade dos dados raramente está próxima da ideal, pelo que a aplicação a dados não separáveis torna-se útil. Para estender este algoritmo a dados não separáveis torna-se necessário

atribuir alguma folga às restrições. Esta folga é realizada pela introdução de um determinado custo através da introdução de variáveis não negativas  $\xi_i \geq 0, i = 1, \dots, l$  (Burges, 1998; Cortes & Vapnik, 1995). O hiperplano encontrado denomina-se por hiperplano de margens suaves para dados lineares não separáveis.

### 2.3.2.1. SVM de margens suaves

Considerando que os dados de treino não podem ser separados (para já, por um hiperplano) sem a existência de erro de classificação, torna-se necessário a introdução das variáveis de folga, mas minimizando o somatório dos erros obtidos, conforme exposto na figura 3. As restrições passam a ter a seguinte configuração (Burges, 1998; Cristianini & Shawe-Taylor, 2000):

$$\begin{aligned} \mathbf{w} \cdot \mathbf{x}_i + b &\geq +1 - \xi_i & \text{para } y_i = +1 \\ \mathbf{w} \cdot \mathbf{x}_i + b &\leq -1 + \xi_i & \text{para } y_i = -1 \end{aligned} \quad (25)$$

ou seja,

$$y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 \geq -\xi_i \quad (26)$$

para  $\xi_i \geq 0$ .

A função custo passa a ser adicionada à função objetivo, mas ao invés de minimizar  $\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2$ , passa-se a (Cortes & Vapnik, 1995):

$$\min_{w,b} \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i^\sigma \quad (27)$$

e acrescenta-se as restrições  $0 \leq \alpha_i \leq C$ .

O valor de  $C$  corresponde a uma escolha que por sua vez quantifica a penalidade para os erros. Um  $C$  maior corresponde à atribuição de uma maior penalidade para os erros (Borges, 1998).

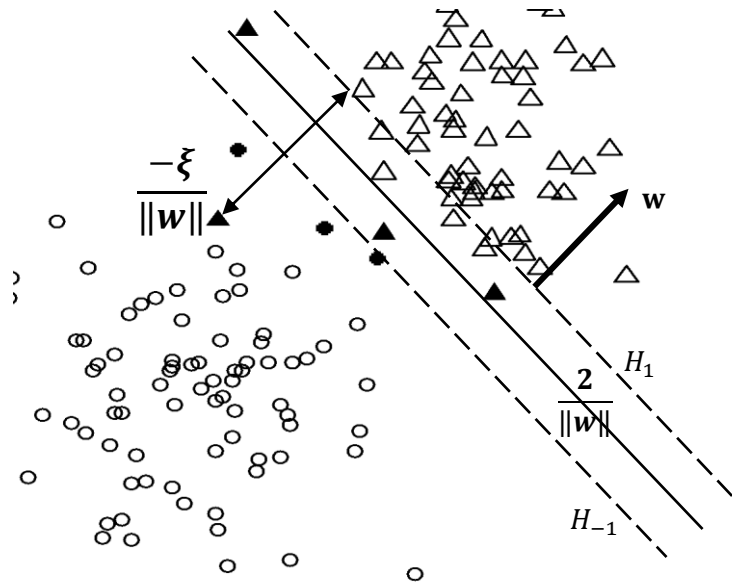


Figura 3 - Maximização das margens do hiperplano com introdução variáveis de folga para o caso linear não separável.

O valor de  $\sigma > 0$  na equação (27) será igual a um sendo o menor valor possível para que o problema de programação quadrática expresso em (19) tenha solução única (Cortes & Vapnik, 1995), permitindo a vantagem em relação aos multiplicadores de Lagrange, pelo que o problema dual é expresso da seguinte forma:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\mathbf{x}_i \cdot \mathbf{x}_j) \quad (28)$$

sujeito às restrições:

$$\begin{cases} 0 \leq \alpha_i \leq C, & \forall i = 1, \dots, n \\ \sum_{i=1}^n \alpha_i y_i = 0 \end{cases} \quad (29)$$

com a solução:

$$\mathbf{w} = \sum_{\mathbf{x}_i \in SV} \alpha_i y_i \mathbf{x}_i \quad (30)$$

onde  $SV$  corresponde ao número de vetores suporte. Utilizando novamente as condições Karush-Kuhn-Tucker obtém-se o seguinte problema segundo a função primal de Lagrange (Burges, 1998; Hofmann et al., 2008):

$$\mathcal{L}(\mathbf{w}, b, \alpha) = \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^n \xi_i - \sum_{i=1}^n \alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i) - \sum_{i=1}^n \eta_i \xi_i \quad (31)$$

onde  $\alpha_i, \eta_i \geq 0 \forall i \in n$ , pois  $\alpha_i, \eta_i$  são multiplicadores de Lagrange introduzidos para reforçar a positividade de  $\xi_i$ . Para processar a função dual de  $\mathcal{L}(\mathbf{w}, b, \alpha)$  torna-se necessário identificar as condições de primeira ordem em  $(\mathbf{w}, b)$ , através das seguintes restrições:

$$\frac{\partial}{\partial \mathbf{w}} \mathcal{L} = \mathbf{w} - \sum_{i=1}^n \alpha_i y_i \mathbf{x}_i = 0 \quad (32)$$

$$\frac{\partial}{\partial b} \mathcal{L} = - \sum_{i=1}^n \alpha_i y_i = 0 \quad (33)$$

$$\frac{\partial}{\partial \xi_i} \mathcal{L} = C - \alpha_i + \eta_i = 0 \quad (34)$$

$$(y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i) = 0 \quad (35)$$

$$\alpha_i (y_i (\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i) = 0 \quad (36)$$

$$\eta_i \xi_i = 0 \quad (37)$$

$$\alpha_i \in [0, C], \forall i \in n$$

Volta-se a considerar as condições Karush-Kuhn-Tucker para determinar o valor de  $b$ , provindo este de  $\alpha$ , pois para  $0 < \alpha_i < C$  basta considerar  $\alpha_i(y_i(\mathbf{w} \cdot \mathbf{x}_i + b) - 1 + \xi_i) = 0$  com  $\xi_i = 0$ . Apesar da introdução de variáveis que permitem uma folga suavizando as margens permitir uma aplicação mais vasta que o caso das margens rígidas, esta abordagem também apresenta limitações por ser linear. O seu uso tende a ser limitado solicitando decisões de fronteira não lineares.

### 2.3.2.2. SVM não linear e “kernel trick”

Os dois casos anteriores partem do pressuposto de um hiperplano de separação. No entanto, uma ideia engenhosa chamada de “kernel trick”, permite considerar um espaço de dimensão superior onde é feita uma separação linear, sendo que no espaço de *features* a superfície de separação pode não ser linear. Assim, considerando um mapeamento  $\Phi$  dos dados para um espaço de dimensão superior, espaço euclidiano  $\mathcal{H}$  (Burgess, 1998; Müller et al., 2001)

$$\Phi: \mathbb{R}^d \rightarrow \mathcal{H}$$

$$\mathbf{x} \mapsto \Phi(\mathbf{x})$$

Poder-se-ia pensar que um aumento da dimensionalidade significa também uma maior complexidade, mas neste caso verifica-se exatamente o contrário, a aprendizagem em  $\mathcal{H}$  apresenta maior simplicidade. Os procedimentos que servem de base a esta transformação são suportados pelas capacidades de separação estabelecidas pelo Teorema de Cover (Cover, 1965). O problema de aprendizagem passa então a processar-se em  $\mathcal{H}$  com o exemplo de treino  $(\Phi(\mathbf{x}_1), y_1), (\Phi(\mathbf{x}_2), y_2) \dots, (\Phi(\mathbf{x}_d), y_d) \in \mathcal{H} \times \mathcal{Y}$ . Devido à dificuldade de trabalhar explicitamente com  $\Phi$ , recorre a um truque através das equações de *kernel* denominado por “kernel trick”, que consiste em substituir o produto interno  $\mathbf{x}_i \cdot \mathbf{x}_j$  por uma função  $K(\mathbf{x}_i, \mathbf{x}_j)$  no algoritmo de treino, em que o *kernel*  $k$  terá de satisfazer alguns pressupostos evidenciados pelo teorema de Mercer como a positividade definida e o ser simétrico: se  $k$  é uma função *kernel* contínua de um operador inteiro positivo, pode-se construir um mapeamento para um

espaço  $\mathcal{H}$  onde  $k$  age como um produto interno (Cristianini & Shawe-Taylor, 2000; Herbrich, 2002; Hofmann et al., 2008).

Podem-se destacar três grandes benefícios no mapeamento dos dados para o espaço característico  $\mathcal{H}$  através de  $\Phi$  (Chen et al., 2005; Schölkopf & Smola, 2002):

1. Similaridade nas medidas do produto interno em que

$$K(\mathbf{x}_i, \mathbf{x}_j) := \mathbf{x}_i \cdot \mathbf{x}_j := \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)$$

2. Permite lidar com os padrões geometricamente ao usar álgebra linear e geometria analítica;
3. Possibilita o desenvolvimento de alguma variedade em algoritmos de treino e similaridade entre medidas devido a alguma liberdade para escolher o mapeamento  $\Phi$ .

Se considerarmos a transferência de dados de  $\mathbb{R}^2$  para  $\mathbb{R}^3$ , o conjunto de dados não lineares em  $\mathbb{R}^2$  para ser linearmente separável em  $\mathbb{R}^3$  (Burges, 1998; Hofmann et al., 2008; Lorena & Carvalho, 2003; 2007; Müller et al., 2001):

$$\Phi: \mathbb{R}^2 \rightarrow \mathbb{R}^3 \tag{38}$$

$$\Phi(\mathbf{x}) = (x_1, x_2) \mapsto (z_1, z_2, z_3) := (x_1^2, \sqrt{2}x_1x_2, x_2^2)$$

$$\mathbf{w} \cdot \Phi(\mathbf{x}) + b = w_1x_1^2 + w_2\sqrt{2}x_1x_2 + w_3x_2^2 + b = 0$$

Sendo agora precisa a aplicação de um hiperplano linear. Sobre as mesmas restrições já mencionadas para o caso de margens suaves lineares, o problema de otimização assume agora o formato:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j (\Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j)) \tag{39}$$

com o classificador:

$$g(x) = \text{sgn}(f(x)) = \text{sgn}\left(\sum_{\mathbf{x}_i \in SV} \alpha_i y_i \Phi(\mathbf{x}_i) \cdot \Phi(\mathbf{x}_j) + b\right) \quad (40)$$

Porém, problemas de maior dimensão apresentariam algumas dificuldades de controlar, daí a aplicação do “*kernel trick*” que permite recorrer ao produto interno entre dois espaços característicos  $\mathbf{x}$  e  $\mathbf{y}$ . Obtém-se:

$$\begin{aligned} (\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) &= (x_1^2, \sqrt{2}x_1x_2, x_2^2)(y_1^2, \sqrt{2}y_1y_2, y_2^2)^T \\ &= ((x_1, x_2)(y_1, y_2)^T)^2 \\ &= (\mathbf{x} \cdot \mathbf{y})^2 \\ &:= \mathbf{k}(\mathbf{x}, \mathbf{y}) \end{aligned}$$

Pode-se generalizar à função (Müller et al., 2001; Scholkopf, Smola, & Muller, 1996; Schölkopf & Smola, 2002):

$$\mathbf{k}(\mathbf{x}, \mathbf{y}) = (\mathbf{x} \cdot \mathbf{y})^d \quad (41)$$

Dando continuidade ao descrito pela notação anterior, considera-se que  $k(x_i, x_j)$  ao invés de  $k(\mathbf{x}, \mathbf{y})$  na forma dual do problema que irá assumir o formato:

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n \alpha_i \alpha_j y_i y_j \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) \quad (42)$$

sujeito às restrições já mencionadas para o caso não separável e com a solução (Begg, Palaniswami, Member, & Owen, 2005):

$$g(x) = \text{sgn}(f(x)) = \text{sgn}\left(\sum_{\mathbf{x}_i \in SV} \alpha_i y_i \mathbf{k}(\mathbf{x}_i, \mathbf{x}_j) + b\right) \quad (43)$$

Notamos que apesar de conceitualmente ser necessária a função de mapeamento  $\Phi$ , ela não aparece na formulação dual, pelo que não é necessário definir essa função. Na realidade, se tivermos uma função *kernel* que satisfaça as condições do teorema de Mercer, podemos aplicar o “*kernel trick*” e obter uma superfície não-linear de separação para o problema em estudo. Na literatura, costumam ser consideradas quatro funções *kernel* básicas (Hsu, Chang, & Lin, 2010):

- Linear:  $K(\mathbf{x}_i, \mathbf{x}_j) = \mathbf{x}_i^T \mathbf{x}_j$ ;
- Polinomial:  $K(\mathbf{x}_i, \mathbf{x}_j) = (\gamma \mathbf{x}_i^T \mathbf{x}_j + r)^d, \gamma > 0$ ;
- RBF (*Radial Basis Function*) gaussiana:  $K(\mathbf{x}_i, \mathbf{x}_j) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}_j\|^2), \gamma > 0$ ;
- Sigmoidal:  $K(\mathbf{x}_i, \mathbf{x}_j) = \tanh(\gamma \mathbf{x}_i^T \mathbf{x}_j + r)$ .

em que  $\gamma$ ,  $r$  e  $d$  são parâmetros *kernel*. O número de vetores suporte necessários ao usar as funções *kernel* polinomial, RBF ou sigmoide não diverge muito, sendo a RBF aquela que apresenta menor número de vetores na construção no modelo de decisão (Scholkopf et al., 1995).

### 2.3.2.3. Alternativas e extensões SVM

As  $\nu$ -SVM foram propostas por Schölkopf, Smola, Williamson, & Bartlett (2000) denominadas por novas SVM, uma classe de algoritmos tanto para classificação como para regressão. O parâmetro  $\nu$  permite controlar o número de vetores suporte e eliminar o parâmetro  $C$  no caso da classificação e o parâmetro  $\varepsilon$  na regressão. O

desenvolvimento deste algoritmo segue a seguinte proposição que consiste na parametrização  $\nu$ , no qual o parâmetro  $C$  é substituído por  $\nu \in [0,1]$ , denominando-se por  $\nu$ -SVM.  $C$  desaparece na equação ficando os parâmetros  $\nu$  e  $\rho$  com o seguinte problema primal (Chen et al., 2005):

$$\min_{\mathbf{w}, \xi, \rho} \frac{1}{2} \|\mathbf{w}\|^2 - \nu\rho + \frac{1}{2} \sum_{i=1}^n \xi_i, \quad \mathbf{w} \in \mathcal{H}; \xi \in \mathbb{R}^n, \rho, b \in \mathbb{R} \quad (44)$$

Com as restrições

$$\begin{cases} y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq \rho - \xi_i \\ \xi_i \geq 0 \\ \rho \geq 0 \end{cases} \quad (45)$$

O parâmetro  $\rho$  está associado a  $y_i(\mathbf{w} \cdot \mathbf{x}_i + b) \geq \rho - \xi_i$  porque esta sujeição divide duas classes pela margem  $2\rho/\|\mathbf{w}\|$  quando  $\xi = 0$ . Para explicar a significância de  $\nu$ , Chen et al. (2005) introduzem o conceito de margem de erro (*margin error*), que correspondem aos pontos de treino que contêm erro ou estão dentro da margem, com  $\xi_i > 0$ . A fração da margem de erros é dada pela equação:

$$R_{emp}^\rho[g] := \frac{1}{n} |\{i: y_i g(x_i) < \rho\}| \quad (46)$$

onde

$$g(x) = \text{sgn} \sum_{i=1}^n y_i \alpha_i k(x, x_i) + b \quad (47)$$

Outra extensão refere-se ao conceito de multiclases, que pode ser realizado tanto através de separadores lineares como por *kernel*, ou seja, através da construção e combinação de vários classificadores binários, ou então, considerando todos os dados para a otimização (Hsu & Lin, 2002). Em relação a usar todos os dados para otimização vários métodos são propostos, como “*one-against-all*”, “*one-against-one*” e gráficos acíclicos direcionados de SVM (DAG - *Directed acyclic graph SVM*) e “*all-together*” (Müller et al., 2001). Estes métodos permitem a comparação de elementos supostamente pertencentes a uma classe em relação às demais. Aliadas a estas técnicas, outras extensões podem ser utilizadas para diminuir a complexidade na classificação. Li, Yang, Jiang, Liu e Cai (2012) recolheram sinal de 32 canais de *features*, logo com dimensionalidade  $\mathbb{R}^{32}$ , na construção de um algoritmo SVM para classificar 17 gestos dos dedos. Aplicaram as SVM multiclases “*one-against-one*” obtendo um total de 136 classificadores binários com  $k(k - 1)/2$ , onde  $k$  corresponde ao número de classes, aplicando posteriormente *majority voting*. Partindo de um classificador binário para dados linearmente separáveis, as SVM evoluíram para uma ampla diversidade de aplicações permitindo tarefas supervisionadas no reconhecimento de dados não lineares e com maior dimensão de classes a classificar.

## 2.4. A Eletromiografia

### 2.4.1. Conceito e caracterização

A definição de eletromiografia está bem expressa na palavra em si quando a seccionamos, eletricidade (*electro*), músculo (*mio*) e registo (*grafia*). Desta forma, eletromiografia é o registo da atividade elétrica gerada no músculo durante a contração num determinado gesto motor. Com a eletromiografia obtemos a representação gráfica da atividade elétrica do músculo (Pezarat-Correia & Mil-Homens, 2004).

*Electromyography (EMG) is an experimental technique concerned with the development, recording and analysis of myoelectric signals. Myoelectric signals are formed by physiological variations in the state of muscle fiber membranes.*

(Basmajian & DeLuca, 1985 cit. Konrad, 2005, p. 4)

A EMG é um método de grande utilidade na compreensão da forma como o sistema nervoso central organiza grupos musculares temporalmente, permitindo inferir sobre aspectos como a sequência intermuscular e a intensidade de ativação, e fadiga. Por este motivo, a análise quantitativa de sinais EMG tende a ser realizada em torno de três tipos de parâmetros: de estrutura temporal, de amplitude e de frequência. As aplicações na investigação estão orientadas para estudos da função neuromuscular em diferentes habilidades motoras (desportivas, profissionais ou de reabilitação), na coordenação muscular, controlo motor e aprendizagem, fadiga, ergonomia e metodologia de treino (Pezarat-Correia & Mil-Homens, 2004).

A amplitude do sinal de EMG é estocástica podendo ser representada por uma distribuição Gaussiana. Esta amplitude do sinal pode variar de 0 a 10 mV (pico-a-pico) ou 0 a 1.5 mV (rms). As frequências do sinal EMG de superfície situam-se entre 0 e 500 Hz com elevado domínio entre os 50 e 150 Hz (De Luca, 1997). A eletromiografia de superfície tem como grande vantagem a facilidade de utilização, o que corresponde igualmente a um sério risco, pois *is too easy to use and consequently too easy to abuse* (De Luca, 1997, p.135), sendo necessário remover potenciais variáveis de influência.

#### **2.4.2. Fatores que influenciam o EMG**

As entradas extraídas do sinal EMG podem ser alvo de várias interferências que estão dependentes das condições experimentais e influenciam a qualidade do sinal obtido. Por este motivo, as publicações sobre esta área científica devem reportar todos os procedimentos desde a recolha até ao tratamento. Farina, Merletti e Enoka (2004) categorizam os fatores que influenciam o EMG em não fisiológicos e fisiológicos dividindo-se estes nas subcategorias ilustradas na figura 4.

Os fatores anatómicos estão relacionados com questões morfológicas dos sujeitos, pelo que são difíceis de controlar e comparar, motivo pelo qual se procede à normalização dos dados recolhidos. Como fatores anatómicos podem ser evidenciados a heterogeneidade dos tecidos subcutâneos, a distribuição e tamanho de áreas com unidade motoras, o comprimento das fibras musculares, ângulos de penação dos músculos e restante organização mio-tendinosa.

O sistema de detecção será mais eficiente quanto melhor for a qualidade do sinal que possibilita extrair face às suas características técnicas. A diminuição do ruído e da impedância no contacto entre pele e eléctrodo é um dos elementos fundamentais para a qualidade do sinal que justificam a devida preparação da pele e cuidados na colocação de eléctrodos.

Os fatores geométricos estão relacionados com os movimentos do músculo, como o encurtamento das fibras e inversões em relação às superfícies de detecção. Este aspeto traduz-se na falta de uniformidade ao longo do movimento quando realizadas em habilidade motoras dinâmicas.

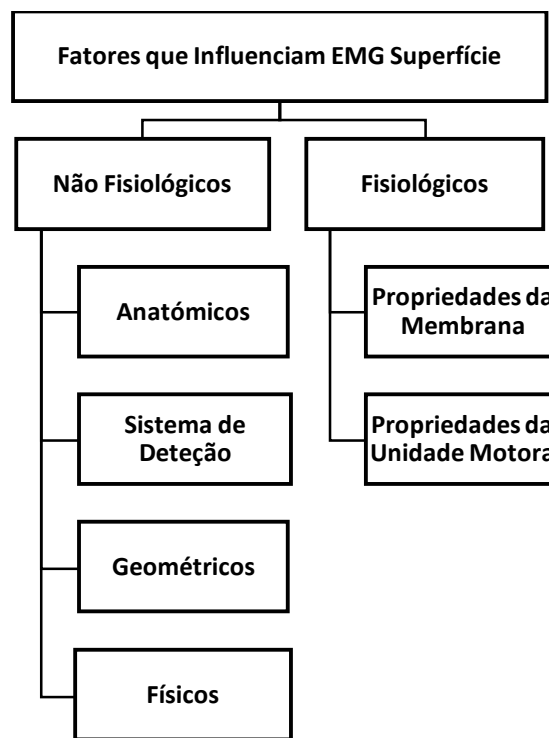


Figura 4 – Fatores que influenciam o EMG de superfície (adaptado Farina et al., 2004, p.1487).

A nível físico, o sinal EMG é influenciado pela condutividade dos tecidos e pela atividade de músculos vizinhos em relação aquele que está a ser analisado. A propagação da atividade elétrica de músculos próximos difundida no registo do sinal denomina-se por *crosstalk*.

Como fatores fisiológicos consideram-se as propriedades das fibras músculo-esqueléticas e das unidades motoras. Dependendo do tipo de fibras e das suas

características, verifica-se variação da velocidade média de condução dessa fibra e da propagação dos potenciais de ação no sarcolema.

As questões ambientais que levam à aquisição de ruído pelo registo EMG durante a propagação do sinal podem ser categorizadas da seguinte forma (Raez, Hussain, & Mohd-Yasin, 2006):

- a) Ruído inerente ao equipamento eletrónico: este tipo de ruído não pode ser eliminado, mas pode ser reduzido pela utilização de componentes eletrónicos de elevada qualidade;
- b) Ruído ambiente: deve-se à radiação eletromagnética à qual estamos expostos continuamente. O ruído ambiente pode ter amplitude superior ao sinal EMG numa ordem de um a três;
- c) Artefactos mecânicos: a introdução de artefactos mecânicos leva a distorções do sinal, podendo ser produzidos na 1) zona de contacto do eléctrodo e 2) nos cabos que ligam os eléctrodos ao restante equipamento;
- d) Ruído inerente à instabilidade de sinal: a amplitude do sinal EMG é de natureza aleatória, sendo influenciada pela capacidade de disparo das unidades motoras. Na sua maioria, como os artefactos mecânicos, situam-se dentro de uma frequência de 0 a 20 Hz, devendo ser removida.

Na determinação de parâmetros temporais, a quantificação face ao tipo de contração realizada não tem tanta importância como em estudos sobre intensidade. O que assume relevância é o rácio sinal-ruído e se a determinação do sinal de EMG tem origem no músculo que realmente se pretende estudar (De Luca, 1997). Torna-se fundamental prevenir influências como o *crosstalk*, eletrocardiograma e outras fontes de ruído que possam dificultar a precisão de algoritmos de deteção.

### **2.4.3. Parâmetros temporais em EMG**

Os principais parâmetros temporais considerados em EMG são o início de ativação muscular (*onset*), instante do pico máximo e o final de atividade muscular (*offset*). A ideia por detrás destes conceitos é a de quantificar quando o músculo liga, quando atinge o seu máximo de ativação, quando desliga, e durante quanto tempo

esteve em atividade (Konrad, 2005). No entanto, o parâmetro que tem sido alvo de estudo é o *onset* que por definição refere-se ao primeiro registo de atividade muscular dos potenciais de ação das unidades motoras (Solnik et al., 2010).

Vários métodos de deteção do *onset* têm sido propostos, mas a diversidade entre os mesmos e diferença na utilização de parâmetros dentro do mesmo tipo de algoritmos comprometem a reprodutibilidade de resultados (Morey-Klapsing et al., 2004). Estes métodos de deteção podem ser distinguidos em duas grandes categorias: inspeção visual e algoritmos de deteção (Vaisman et al., 2010). A inspeção visual requiere muito tempo e está dependente da sensibilidade dos observadores que determinam o parâmetro temporal, sendo tido como um processo subjetivo (Jöllenbeck, 2000). Por sua vez, os principais algoritmos de deteção são os que consideram um limiar, seja pelo valor médio e desvio padrão da amplitude da linha de base quando o músculo se encontra em repouso (Allison, 2003; Hodges & Bui, 1996; Silva et al., 2013), por intervalos de confiança (Van Boxtel et al., 1993), ou por limiar fixo percentual relativo ao pico EMG máximo (Konrad, 2005). Outra abordagem algorítmica é a que recorre a processos de otimização estatística (Micera, Sabatini, & Dario, 1998; Staude et al., 2001), como por exemplo, o método da máxima verosimilhança. Basicamente, pode-se afirmar que qualquer estudo sobre o *onset* deve recorrer a um algoritmo de deteção, porém, devido à inexistência de um procedimento de referência, a validade dos algoritmos é aferida por inspeção visual. Este aspeto é deveras paradigmático.

A relevância de detetar parâmetros temporais, tais como o *onset*, deve-se ao facto de permitir captar informação sobre a organização temporal e coordenação de grupos musculares durante determinada habilidade motora (De Luca, 1997). Apesar da importância deste conceito, alguns autores têm alertado para a necessidade de descrever os fenómenos temporais tendo em conta o significado fisiológico do registo EMG com as habilidades motoras estudadas (Silva et al., 2013; McGill et al., 2010; Tyler & Karst, 2004). Silva et al. (2013) introduziram o conceito de *onset peak*, que corresponde a qualquer momento onde é iniciado um pico de ativação significativo.

Três artigos estudaram o *onset* no *swing* do golfe (Cole & Grimshaw, 2008; Horton et al., 2001; Silva et al., 2013), todos recorrendo a métodos por limiar. Os dois primeiros incidiram em populações com e sem lombalgia. Silva et al. (2013) comparou dois métodos com linhas de base diferentes como referência, da própria repetição e o

repouso entre duas contrações voluntárias máximas. Nestes estudos os desvios-padrão considerados variaram entre 1 e 7.

## **2.5. Reconhecimento de Padrões EMG através de SVM**

### **2.5.1. Etapas no reconhecimento de padrões EMG**

O estudo do sinal EMG tende a estar associado a outras variáveis que ajudam a quantificar o mesmo, como por exemplo, alterações morfológicas, análise cinemática e o estudo da cinética. Com o cruzamento de outras variáveis com EMG torna-se possível retirar significado fisiológico da série temporal sobre o registo da ativação do músculo.

A utilização de SVM em EMG consistirá em caracterizar momentos inscritos no registo do sinal com relevância fisiológica, como por exemplo, intervalos entre momentos de repouso e atividade que seriam identificados por inspeção visual sejam automaticamente classificados com a maior robustez possível. Porém, o sinal EMG pode ser influenciado por diversos fatores que não estão diretamente relacionados com a atividade muscular (De Luca, 1997), o que terá o seu peso na classificação.

Para além de ruído que afeta a qualidade do sinal, a robustez no reconhecimento de padrões pode ser influenciada pela extração dos vetores característicos. Estes devem ser extraídos a partir de vários segmentos numa janela temporal para preservar a estrutura do sinal e não de amostras individuais, pois representaria uma perda significativa de informação (Hudgins, Parker, & Scott, 1993).

Oskoei & Hu (2007) referem que o reconhecimento de padrões através de controlo mio-elétrico, após os procedimentos de amplificação, filtragem e digitalização do sinal, seguem quatro passos fundamentais (Figura 5):

1. Segmentação dos dados: um segmento é uma sequência limitada por uma janela de temporal, esta é utilizada para estimar o vetor característico que corresponderá a um conjunto de entradas (*features*) desse sinal (no presente estudo segmenta-se em relação a um instante);
2. Extração de entradas (*features*): do segmento retiram-se as entradas que serão utilizadas para a classificação, ao invés do sinal bruto, das quais depende a

eficiência do reconhecimento de padrões. A etapa de extração de entradas é usada para aumentar a densidade de informação do sinal de EMG. A informação que discrimina a ativação muscular deve ser retida, enquanto outra informação não relevante deve ser descartada (Scheme & Englehart, 2011);

3. Classificação: utilização de uma técnica matemática que discrimine as entradas em classes mediante uma pré-categorização. Quanto maior for a capacidade de generalização, melhor será a eficiência do classificador;
4. Controlador: refere-se aos comandos de saída que têm como base os padrões retirados do sinal e sistemas de controlo. São colocados dois níveis de feedback: baixo nível quando entre o controlador e operador e alto nível quando entre o operador e cérebro (aprendizagem obtida de informação visual e centros coordenativos e motores).

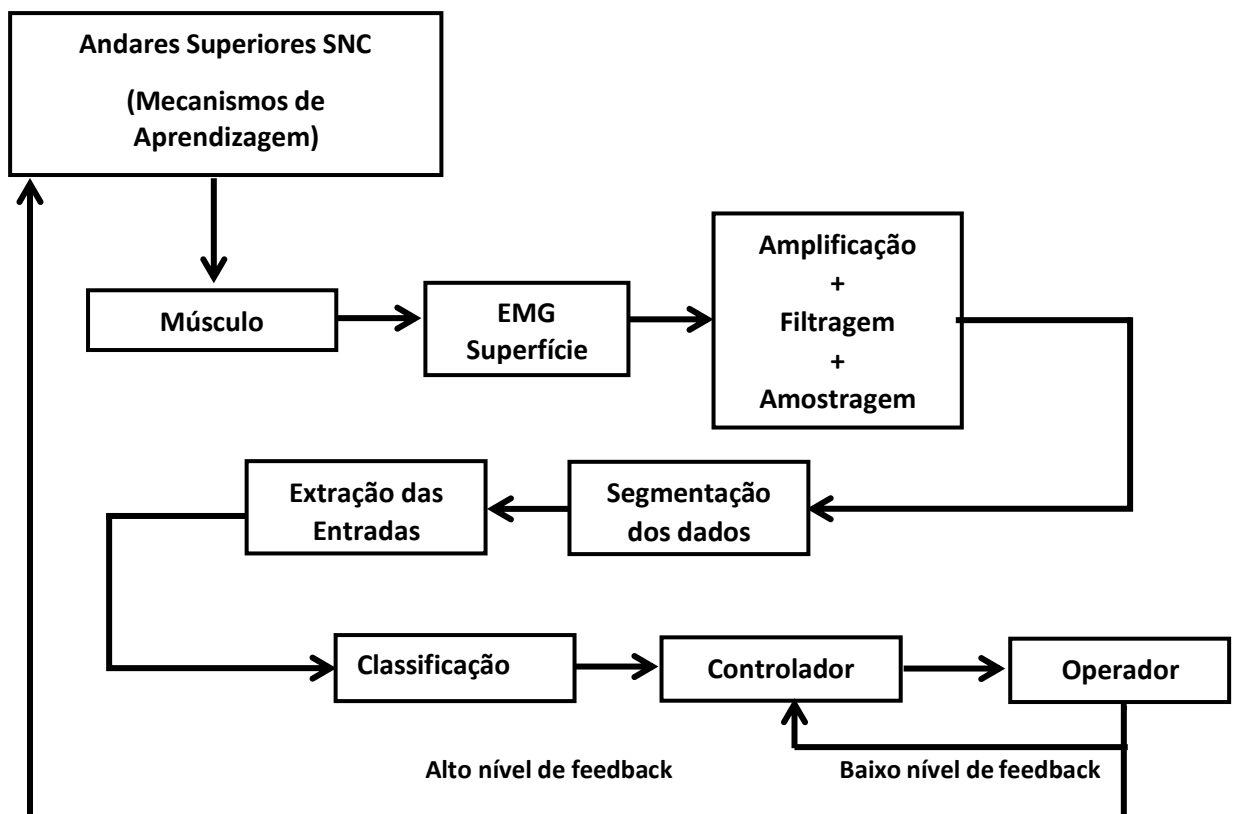


Figura 5 – Fases de tratamento de EMG no reconhecimento de padrões (adaptado de Oskoei & Hu, 2007, p. 277). SNC – Sistema nervoso central.

Quando aplicado qualquer tipo de segmentação coloca-se então a questão sobre a dimensão ótima dessa janela temporal. Um pequeno comprimento do segmento está

associado a desvios e variações na estimação da componente característica devido à perda de informação do sinal, um comprimento longo impõe uma carga computacional elevada, o que pode comprometer a execução da classificação em tempo real (Oskoei & Hu, 2008). Porém, as decisões conhecidas baseiam-se essencialmente no controlo mio-elétrico, não em relação à vizinhança de pontos que irá representar um determinado ponto central numa janela de segmentação.

### 2.5.2. Segmentação dos dados

Um segmento é uma janela temporal retirada da aquisição de dados, sendo um procedimento muito associado ao controlo mio-elétrico, onde será realizada a extração das entradas para classificação. A escolha do comprimento dos segmentos deve compreender que o sinal EMG apresenta dois estados: a) um estado transiente proveniente da despolarização das fibras, passando o músculo de um estado de repouso para ativação voluntária; b) um estado estacionário que corresponde ao equilíbrio durante uma contração voluntária que é mantido por determinado período de tempo (Oskoei & Hu, 2008). Ainda, estes autores referem que devido a restrições associadas a situações de tempo real, o comprimento do segmento adjacente e o tempo de processamento de classificação deve ser igual ou inferior a 300 ms. Englehart e Hudgins (2003) utilizaram janelas até 256 ms analisando dois parâmetros: o comprimento da janela ou segmento (*analysis window length*) e o atraso aceitável (*acceptable delay*). O segmento expressa a quantidade de dados consoante a taxa de amostragem (no caso deste estudo foi de 1000 Hz, logo 256 ms correspondem a 256 instantes da amostra), sendo esperado que uma maior quantidade de dados resulte em características com menor variância e melhor precisão. O atraso aceitável refere-se ao tempo de resposta do sistema de controlo (início da intenção mio-elétrica até ao momento que o sistema gera a classificação). O atraso aceitável ( $T_d$ ) é fornecido pela seguinte relação:

$$\tau \cdot m \leq T_d$$

onde  $\tau$  é o tempo de processamento e  $m$  o número de decisões de classificação realizadas. Foi verificado que o comprimento do segmento pode ser reduzido até 32 ms

não degradando a precisão e que o atraso aceitável é reduzido de 256 a 128 ms (Englehart & Hudgins, 2003). Outra questão a ter em conta em relação à escolha de uma janela temporal para segmentação, quando aplicada a estados transitórios de sinal EMG, são melhorias na capacidade de classificação das *features* extraídas 100 ms após o início de atividade, ou seja, após o *onset* quando o músculo se encontra em atividade (Hudgins, Parker & Scott, 1993).

Após a decisão sobre o comprimento dos dados torna-se necessário aplicar a técnica que irá realizar a segmentação dos dados, podendo ser executada por dois métodos principais (Englehart & Hudgins, 2003; Oskoei & Hu, 2007; Oskoei & Hu, 2008): segmentação sucessiva (*successive/adjacente/disjoint windowing*) e segmentação sobreposta (*overlapped segmentation*). Na segmentação sucessiva, segmentos seguidos um imediatamente ao adjacente com um comprimento predefinido são utilizados para a extração de entradas, surgindo depois de um atraso de processamento. O tempo de processamento é uma pequena porção do comprimento do segmento, o processador está inativo durante a restante duração do comprimento do segmento. Na segmentação sobreposta, um novo segmento desliza sobre o segmento que já está a ocorrer, com um incremento de tempo menor do que o comprimento do segmento. Deve ser maior do que o tempo de processamento, uma vez que o processador deve calcular o conjunto de entradas para classificação antes do segmento que se segue. A primeira técnica deverá ser aplicada para segmentos de comprimento igual ou inferior a 200 ms, enquanto a segmentação sobreposta quando a janela é superior a 200 ms (Oskoei & Hu, 2008).

Devido à natureza estocástica e não estacionária do sinal EMG, a decisão sobre o comprimento dos segmentos deve compreender a existência de estados de transição e estacionários, sendo estes últimos os mais utilizados. No entanto, para o cálculo de parâmetros temporais precisamos dos dois tipos de informação.

### **2.5.3. Extração de *features***

As componentes características, entradas ou *features* são valores numéricos caracterizadores de cada objeto a classificar (no caso da aplicação desta dissertação, o objeto a classificar é cada instante de tempo, que é depois classificado como sendo antes

do *onset* ou depois do *onset*). A sequência dessas entradas determina o vetor característico.

Existem três categorias para quantificar as entradas (Zecca, Micera, Carrozza, & Dario, 2002; Englehart & Hudgins, 2003; Oskoei & Hu, 2007; Oskoei & Hu, 2008; Phinyomark, Phukpattaranont, & Limsakul, 2012): domínio do tempo (*time domain*), domínio da frequência ou espectral (*frequency domain*) e o domínio tempo-frequência (*time-scale domain*). Desde a década de 70, o sinal EMG é modelado como amplitude que depende de ruído gaussiano cuja variância está relacionada com a força desenvolvida pelo músculo (Zecca & Micera, 2002), levando a diferentes abordagens que consigam ser sensíveis na quantificação de alterações do sinal. Oskoei & Hu (2007) expressam que as entradas podem ser avaliadas mediante duas abordagens, uma estrutural e outra fenomenológica. Na abordagem estrutural, as entradas são avaliadas através de modelos matemáticos que avaliam características como a variabilidade e sensibilidade ao ruído que estão associadas à constituição física e fisiológica do sinal. A abordagem fenomenológica realiza uma interpretação do sinal para além da sua estrutura, tendo em conta o desempenho e a robustez, sendo denominada por empírica.

*The success of any pattern classification system depends almost entirely on the choice of features used to represent the continuous time waveforms.*

(Hudgins, Parker, & Scott, 1993, p.86)

#### **2.5.3.1. Domínio do tempo**

A quantificação do sinal no domínio do tempo oferece maior simplicidade computacional baseando-se na amplitude do sinal EMG. A amplitude será a ordem de grandeza que expressa variação da atividade elétrica do músculo num determinado período, representando a sua intensidade (Pezarat-Correia & Mil-Homens, 2004). Ao considerar a atividade elétrica do músculo, em bruto, como um sinal estocástico com média zero, a amplitude pode ser definida como a variação do desvio-padrão do sinal ao longo do tempo, que é proporcional ao recrutamento e taxa de ativação de unidades motoras (Oskoei & Hu, 2007). A representação é realizada pela quantificação de entradas e a sua qualidade pode ser averiguada pelo rácio sinal/ruído que é definido pela

média das amostras que constituem um segmento pelo seu desvio-padrão. Quanto maior for este valor melhor será a qualidade da informação recolhida. No entanto, esta medida tende a ser utilizada em ações estáticas devido à grande variabilidade em habilidade motoras dinâmicas.

Na literatura mais direcionada para o estudo da quantificação das *features* do sinal EMG (e.g Clancy & Hogan 1997; Lucas, Gaufriau, Pascual, Doncarli, & Farina, 2008; Oskoei & Hu, 2008; Oskoei & Hu, 2007; Phinyomark, Limsakul, & Phukpattaranont, 2009; Phinyomark, Limsakul, & Phukpattaranont, 2011; Phinyomark et al., 2012; Yan, Wang, & Xie, 2008) encontram-se métodos como: raiz quadrada média ou RMS (*root mean square*), valor absoluto médio (MAV – *mean absolute value*), valor absoluto modificado tipo 1 e 2, declive do valor absoluto médio (*mean absolute value slope*), integral EMG (IEMG – *Integrated EMG*), variância EMG (VAR), integral quadrado simples (*simple square integral*), valor absoluto do momento temporal de 3<sup>a</sup>, 4<sup>a</sup> e 5<sup>a</sup> ordem (*absolute value of the 3rd, 4th and 5th temporal moment*), entrada de ordem *v* (*v-order*), detetor logarítmico (LOG – *logarithmic detector*), comprimento do formato da onda (WL – *waveform length*), alteração média da amplitude (*average amplitude change*), diferença do valor absoluto do desvio-padrão (DASDV – *difference absolute standart deviation value*), cruzamento em zero (*zero crossing*), amplitude Willison, taxa percentual mio-pulso (*myopulse percentage rate*), alteração de declive do sinal (*slope sign change*), janelas múltiplas hamming (*multiple hamming windows*) e janelas múltiplas trapezoidais (*multiple trapezoidal windows*). Por vezes associados a tempo, mas também a frequência podemos destacar os coeficientes autorregressivos e os coeficientes cepstrum.

Em estudos como Oskoei e Hu (2007; 2008) e Phinyomark et al. (2012) é possível encontrar explicações detalhadas sobre estas técnicas de extração de *features*.

### **2.5.3.2. Domínio da frequência**

No domínio da frequência podem-se destacar a média da frequência (*mean frequency*), mediana da frequência (*median frequency*), frequência máxima (*peak frequency*), potência média (*mean power*), potência total (*total power*), os 1<sup>o</sup>, 2<sup>o</sup> e 3<sup>o</sup> momentos espectrais (*1st, 2nd, and 3rd spectral moments*), rácio da frequência

(*frequency ratio*), rácio do espectro de frequência (*power spectrum ratio*) e a variância da frequência central (*variance of central frequency*). Com a aplicação de SVM, Oskoei e Hu (2008) aplicaram no domínio do tempo o espectro de potência (PS), coeficientes auto-regressivos de ordem 2 e 6 e a média e mediana das frequências do espectro EMG. Ainda se pode considerar um terceiro domínio, a escala de tempo, a qual apresenta maior poder em estados de transição mas impõe maior custo computacional.

### 2.5.3.3. *Domínio do tempo-frequência e outras aplicações*

O domínio do tempo-frequência refere-se à análise do sinal em determinada área permitindo uma incidência local através de transformação *wavelet* (*Wavelet transformation*). Este método decompõe séries temporais numa correspondência entre a escala e a frequência determinando o formato da variabilidade e como varia no tempo (Torrence & Compo, 1998). O principal benefício que estará associado à transformação *wavelet* é a produção de um subconjunto útil das componentes de frequência ou de escala do sinal interessado, enquanto as *features* no domínio do tempo traduzem-se por um vetor característico com todas as componentes do sinal (Phinyomark et al., 2011). A opção de utilizar transformação *wavelet* na análise espectral ao invés da transformada de Fourier, deve-se por esta perder informação do sinal no domínio do tempo como se determinado evento não tivesse ocorrido. Em sinais estacionários tal é aceitável, mas em características não estacionárias ou transitórias, a informação útil pode não ser considerada. Outras técnicas como a Transformada de Fourier de curta duração, já mapeiam o sinal para uma função bi-dimensional mas são limitadas pelo tamanho da janela (Oskoei & Hu, 2007).

Podem ser consideradas duas dimensões na análise *wavelet*: transformação *wavelet* contínua e transformação *wavelet* discreta. A primeira recorre aos produtos internos para medir a semelhança entre um sinal e uma função característica que é uma *wavelet*. Quando não se torna necessário decompor todo o sinal para obter a informação pretendida, pois torna-se redundante, a reconstrução do sinal não necessita de decomposição contínua, sendo a transformação *wavelet* discreta uma opção. A escala pode então ser aumentada alterando o espaçamento entre *wavelets* segundo determinada ordem que permitem criar sub-bandas. Igualmente, a transformação *wavelet* discreta projeta um sinal para um conjunto de funções de base que correspondem a versões que

são escaladas de uma função protótipo denominada por *mother wavelet* (Farina, do Nascimento, Lucas, & Doncarli, 2007; Lucas et al., 2008). Um método de aumentar a performance da transformação *wavelet* discreta é a transformada *wavelet* por pacote (*wavelet packet transformation*) considerando a igual relação tempo-frequência para todas as frequências (Conradsen et al., 2012). Existe uma correspondência entre o tempo e frequência numa análise *wavelet* onde uma escala baixa evidencia os detalhes que mudam rapidamente no sinal com frequência elevada, enquanto uma escala elevada mostra mudanças lentas grosseiras de baixa frequência (Oskoei & Hu, 2007).

A aplicação da transformada *wavelet* para extração de *features* de sinal EMG recorrendo a SVM para classificação tem sido utilizada na classificação de gestos dos dedos recorrendo apenas a um canal (Kumar, Poosapadi Arjunan, & Singh, 2013), posições da mão (Lucas et al., 2008), e na discriminação de diferentes níveis de esforço durante a corrida (Stirling, von Tscherner, Kugler, & Nigg, 2011). No diagnóstico de desordens neuromusculares também se verifica a utilização de transformação *wavelet*. Dobrowolski, Wierzbowski e Tomczykiewicz (2012) compararam várias famílias *wavelet*, como *Symlet*, *Daubechies* e *Coiflet* de diferentes ordens, atingindo sensibilidades superiores a 98% e especificidade de 100%. Conradsen et al. (2012) utilizou *wavelet* na construção de um sistema de deteção automático de ataques epiléticos.

Outra abordagem a ser considerada refere-se à análise de quantificação da recorrência (*Recurrence Quantification Analysis*), apresentando uma precisão que chega aos 98.28% na discriminação de desordens neuromusculares (Sultornsanee, Zeid, & Kamarthi, 2011). Este conceito parte do princípio que a EMG obedece a uma lei determinista e não estacionária significativa, o que justifica o estudo destes sinais através de um sistema não linear utilizando a análise de quantificação da recorrência. Sultornsanee, Zeid e Kamarthi (2011) classificaram as desordens neuromusculares tendo em conta três classes: saudável, neuropatia, miopatia, aplicando a análise de quantificação da recorrência como método de extração de características que serão alvo da aprendizagem automática nas SVM. Os parâmetros de extração de vetores característicos são descritos através da taxa de recorrência, o determinismo e a laminaridade.

Através da conjugação de vários métodos que incluem diversas características do sinal obtém-se uma análise fenomenológica.

#### 2.5.4. Aplicação SVM em EMG

No melhor do conhecimento, a aplicação de SVM para classificar padrões EMG desenvolveu-se em três domínios: controlo mio-elétrico, diagnóstico de distúrbios neuromusculares e análise cinésiofisiológica. A determinação do *onset* não tem sido considerada como no âmbito do reconhecimento de padrões (Oskoei & Hu, 2007).

##### 2.5.4.1. Controlo mio-elétrico

O objetivo dos estudos sobre controlo mio-elétrico consiste em reconhecer padrões neuromusculares que levem próteses ou robôs a realizarem determinados movimentos com a melhor precisão possível. Estes estudos incidem nos músculos do antebraço, pelo que os movimentos reconhecidos são essencialmente movimentos de preensão, podendo ser distinguidos em três categorias: movimentos grosseiros, reconhecimento de padrões dos músculos do pulso e braço; movimentos finos dos dedos; classificação de múltiplos movimentos dos dedos (Chen & Wang, 2013). Castellini e Smagt (2009) consideram três preocupações quando a classificação passa de movimentos grosseiros da mão para a especificidade dos dedos: a habilidade do paciente para produzir o gesto correto de preensão, a aplicação de força adequada ao movimento de preensão, e o feedback que é devolvido ao paciente.

EMG usa o tipo de controlo designado por *feed-forward*, além de precisão, exige capacidade adaptativa e de execução em tempo real, daí a importância da aprendizagem automática. Face à necessidade de extrair *features* com poder discriminativo, a sua extração requer métodos avançados de análise de sinais, a fim de tornar possível a descodificação das informações aí contidas (Merletti & Farina, 2008). O sucesso de um sistema de classificação depende da combinação de um conjunto de fatores (Crawford, Miller, Shenoy, & Rao, 2005): (1) cuidada seleção dos movimentos a classificar, (2) seleção dos locais nos músculos dos quais derivará o registo EMG, (3) simplicidade na representação das *features* em tempo real e (4) o estado da arte do método de classificação. O aumento de informação pelo adição de *features* pode conduzir a

performances mais elevadas, mas quando o reconhecimento de padrões tem de ser executado em tempo real é preciso considerar o custo computacional.

Outras propostas no reconhecimento de padrões são apresentadas tendo em conta *features* para além da quantificação do EMG. Huang, Zhang, Hargrove, Dou, Rogers e Englehart (2011) estudaram a musculatura a nível do membro inferior durante a marcha propondo o método de fusão mecânico-neuromuscular para o reconhecimento de padrões durante as fases da marcha. Para além dos aspetos mecânicos na manipulação de próteses, o refinamento mecânico de robôs surge como um desafio que recorre a métodos de classificação de padrões representando valias para ambos objetivos. Pode-se exemplificar como controlo de mão robótica (Yang et al., 2009), e de exosqueleto do pulso com controlo em tempo real (Khokhar, Xiao, & Menon, 2010). Porém, estes casos evidenciam grandes diferenças na precisão de classificação quando comparados conjuntos de treino e de teste.

Na literatura sobre controlo mio-elétrico, tende a existir uma preferência na aplicabilidade de EMG de superfície em relação a EMG de profundidade. Este aspeto deve-se a duas razões principais. Primeiro, porque o EMG de superfície é menos intrusivo que o de profundidade, depois por a informação contida no registo EMG de superfície se referir ao somatório do conjunto de potenciais de ação de todas as fibras ativas do músculo, e não apenas das fibras localizadas na área próxima onde a agulha (no caso da profundidade) está introduzida.

#### **2.5.4.2. Diagnóstico de desordens neuromusculares**

Os artigos que focam a pesquisa na discriminação de desordens neuromusculares dividem as classes de classificação tendo em conta três grupos: saudável, neuropatia e miopatia. As exceções encontradas focam epilepsia mio-clónica juvenil (Goker et al., 2012) e reconhecimento de ataques epiléticos (Conradsen et al., 2012). Ao contrário do controlo mio-elétrico encontramos tanto a utilização de eletromiografia de superfície como de profundidade, chegando a verificar-se a utilização de ambas as técnicas (Güler & Koçer, 2005). Estes autores verificaram que os pacientes com neuropatia apresentam valores elevados na densidade de espectro de frequência, enquanto nos miopáticos esses

valores são baixos. O valor da densidade do espectro de frequência para sujeitos saudáveis encontra-se entre os valores verificados para estas duas populações.

Dobrowolski, Wierzbowski e Tomczykiewicz (2012) reportam que sujeitos miopáticos apresentam um registo de EMG com potenciais multifásicos de baixa amplitude e duração denominada por *short-lasting*. Os potenciais de ação dos neuropáticos também são multifásicos, mas de alta amplitude e duração *long-lasting*. As desordens neuromusculares correspondem a falhas nos elementos que constituem as unidades motoras (moto neurónio alfa e todas as fibras que este inerva). O formato dos potenciais de ação das unidades motoras deve conter informação suficientemente discriminatória em relação à tipologia da patologia.

Na discriminação das características de uma população, a quantificação dos sinais EMG não tende a ser focado meramente em uma característica. Os estudos sobre doenças neuromusculares para além de usarem *features* no domínio do tempo ou de frequência também recorrem a outra abordagem que unifica a informação destes dois domínios. A maioria dos estudos recorre à quantificação das *features* que alimentam o classificador SVM no domínio do tempo-frequência, ou seja, decomposição *wavelet* (i.e. Conradsen et al., 2012; Dobrowolski et al., 2012; Subasi, 2012, 2013).

Os erros típicos na classificação deste tipo de populações derivam dos casos que se encontram no limiar entre classes, como situarem-se na fase inicial da doença ou terem uma desordem muito específica quanto à sua localização muscular (Conradsen et al., 2012). Estes autores, selecionaram o melhor grupo de *features* através de decomposição *wavelet*, realizando pós processamento SVM de forma a obter um único parâmetro de classificação, denominado de *Wavelet Index*.

A grande vantagem na aplicação de análise *wavelet* é a capacidade de concentrar a energia do sinal em poucos coeficientes. Subasi (2012) comparou diferentes tipos de métodos de aprendizagem automática na classificação de doenças neuromusculares. O denominado *fuzzy-SVM* associado à decomposição *wavelet* apresentou melhor performance, insensibilidade a sobre treino e maior fiabilidade que a análise discriminante linear (*linear discriminant analysis*), redes neurais artificiais (*artificial neural network*) e redes neurais RBF. A principal diferença entre SVM e *fuzzy-SVM* está relacionado com o custo  $C$ , uma vez que este parâmetro deriva do

produto entre o  $C$  do *fuzzy-SVM* e o membro *fuzzy*  $0 < s_i \leq 1$  relacionada com cada amostra de treino  $x_i$ .

Outras abordagens que têm propostas como a organização dos potenciais de ação das unidades motoras por *clusters* (Güler & Koçer, 2005), *multi-scale amplitude modulation-frequency* (Katsis, Goletsis, Likas, Fotiadis, & Sarmas, 2006), *scanning* do EMG (Goker et al., 2012), e análise da quantificação de recorrência (Sultornsanee et al., 2011). Este último estudo atingiu valores de classificação elevadíssimos no tibial anterior, com o grupo de miopatia a chegar aos 100%, com uma classificação média de 98.28%.

#### **2.5.4.3. Análise cinesiológica**

Esta dimensão é muito semelhante ao descrito em relação ao controlo mio-elétrico na medida em que através de reconhecimento de padrões EMG pretende-se classificar determinadas habilidades motoras. A diferença reside no feedback para o sujeito e na necessidade de rápido reconhecimento em tempo real por parte do controlo mio-elétrico. Na análise cinesiológica, o pretendido é caracterizar o comportamento motor, como por exemplo, descrever como o sistema nervoso organiza os programas motores (quantificados pelo registo da atividade elétrica a nível intra e inter muscular) na execução de diferentes tarefas motoras. Este conhecimento permite melhorar a performance de atletas ou praticantes, evitar a incidência e prevalência de lesões, e servir de indicação clínica nos vários cenários de recuperação. Apesar de também se verificar uma grande tendência na musculatura do antebraço (Alkan & Günay, 2012; Futamata, Nagata, & Magatani, 2012; Tavakolan, Xiao, & Menon, 2011), já é possível verificar uma maior abrangência, como a nível dos membros inferiores (Stirling et al., 2011; Tolambiya, Thomas, Chiovetto, Berret, & Pozzo, 2011), músculos do tronco (Tolambiya et al., 2011), e no membro superior a atividade do bicípite braquial durante a flexão do cotovelo (Natarajan, Wininger, Kim, & Craelius, 2012).

Além de ser uma dimensão de maior abrangência quanto a variabilidade de músculos estudada, também o é em relação ao tipo de ações motoras. Os sinais recolhidos nos estudos das duas dimensões anteriores privilegiam ações discretas, enquanto agora os horizontes de pesquisa estendem-se a ações contínuas. Stirling, Von

Tscharner, Kugler e Nigg (2011) recolheram EMG durante a corrida com o intuito de identificar diferentes estratégias relacionadas com estágios de esforço, tendo confirmado a existência das mesmas. O papel das SVM como técnica classificadora é provido de sentido em estudos do controlo motor, pois permite identificar relações que podem estar subjacentes a uma tarefa face a um elevado número de variáveis que estão associadas à mesma (Tolambiya et al., 2011). O estudo realizado por estes autores demonstra bem essas relações entre variáveis ao compararem movimentos de alcançar com e sem constrangimentos associados às tarefas realizadas. As restrições consideradas foram do domínio postural (joelho estendido e redução de base de apoio) e domínio focal (trajetória com imposição do dedo em linha reta e imposição semicircular da trajetória do dedo). Estes autores descobriram melhor classificação na precisão de músculos da perna comparativamente com o tronco com um desenho de investigação que considera a totalidade do corpo e associa constrangimentos motores a tarefas como o apontar dedos a um alvo. As classificações que apresentaram uma classificação mais precisa foram a postural com extensão do joelho e a focal com a realização de trajetória com o dedo, com 97% e 91% de precisão, respetivamente.

A preocupação que haja maior dificuldade de construir modelos SVM em populações mais idosas comparativamente com jovens também está presente na literatura. Tavakolan et al. (2011) analisou a performance de SVM com *kernel* RBF na discriminação de oito movimentos da mão. A classificação dos movimentos na população jovem apresentou valores mais elevados comparativamente com a população envelhecida, com 97.6% e 90.6%, respetivamente. No entanto, os parâmetros  $C$  e  $\gamma$  não divergiram entre os dois grupos, com a grande maioria a apresentar um valor  $C$  igual a 10 para e o valor de  $\gamma$  a variar entre 0.2 e 2.4. Embora hajam alterações neurológicas e físicas significativas ao longo do processo de envelhecimento, a classificação de gestos da mão através de sinal EMG de sujeitos seniores recorrendo às SVM também apresenta uma boa robustez.



## 3. Método

### 3.1. Apresentação do Problema

A análise de padrões temporais do registo elétrico da atividade neuromuscular assume um papel primordial no entendimento de como o Sistema Nervoso Central organiza os programas motores nas diferentes tarefas. Dos parâmetros temporais estudados, o *onset* é aquele que se apresenta como maior objeto de estudo, nomeadamente, quanto a métodos de deteção. Apesar de existirem várias propostas, não existe concordância quanto ao melhor método a usar, mas sim que a reprodutibilidade dos diferentes métodos está comprometida. Um aspeto paradigmático deve-se ao facto de existir a necessidade de algoritmos automáticos para a deteção do *onset*, mas por outro lado a sua melhor ou pior fiabilidade depende meramente de inspeção visual, pois não é referenciado um método algorítmico como medida *standard*. Outra limitação é a medida exclusiva do momento de início de ativação muscular quando este pode ser confundido com pré-ativações que não são relevantes quanto a padrões motores relacionados com a habilidade motora em estudo. Outras definições operacionais de *onset* já foram introduzidas (Silva et al., 2013), mas a exploração do significado fisiológico de fenómenos temporais EMG é escassa.

Este estado da arte conduz à necessidade de algoritmos de deteção que privilegiem a discussão da variável temporal que estiver em análise que podem ser sistematizados segundo os seguintes critérios: (1) a performance dos algoritmos é estabelecida através de inspeção visual, (2) usualmente a informação é recolhida como sendo o momento inicial da ativação, (3) algoritmos de limiar aplicam janelas amostrais para evitar desvios na deteção devido a similaridades entre fenómenos relevantes e irrelevantes, (4) os resultados têm uma grande dependência inter algoritmo, (5) os resultados têm uma grande dependência intra algoritmo, (6) detetores *approximated generalized likelihood-ratio* são dependentes do teste do rácio de verossimilhança e procuram o máximo na função teste. Estes critérios conduzem à ideia que o ideal seria ensinar o computador sobre o fenómeno a estudar. Para tal, seria desejável para cada ponto saber como se situam os seus vizinhos num determinado intervalo quantificando

essa janela amostral. Se tivermos esses pontos classificados com rótulos de dados pertencentes em que  $y = -1$  ou  $y = 1$ , em que o primeiro rótulo corresponde aos pontos no registo de repouso e o segundo rótulo ao registo de atividade, determinar a passagem de  $y = -1$  para  $y = 1$  corresponderia ao instante temporal a detetar.

A ideia exposta no parágrafo anterior expressa a teoria da aprendizagem estatística, nomeadamente, existindo um “professor” ou intervenção do investigador que a determinada altura do processo ensina o computador, trata-se de aprendizagem supervisionada (Haykin, 2001; Herbrich, 2002; Luxburg & Schölkopf, 2009; Rojas, 1996).

As SVM têm sido uma técnica de otimização baseada na aprendizagem supervisionada introduzida em 1992 (Boser et al., 1992). Depois foram introduzidas as margens suaves ou dados não separáveis (Cortes & Vapnik, 1995), e a versão não linear com o “*kernel Trick*” tendo em conta a função  $(\Phi(\mathbf{x}) \cdot \Phi(\mathbf{y})) =: k(\mathbf{x} \cdot \mathbf{y})$  (Müller et al., 2001).

O conhecimento nesta área que recorre a SVM para reconhecimento de padrões abrange áreas como o controlo mio-elétrico, diagnóstico de desordens neuromusculares e análise cinesiológica. As *features* no domínio do tempo são das mais aplicadas por permitirem uma aplicação contínua e devido a menor custo computacional.

Surge o problema que sustenta o presente estudo, que pode ser exposto pela seguinte questão: em que medida *features* no domínio do tempo podem ser incluídas em modelos aprendizagem automática para a deteção do *onset* do registo de atividade de maior relevância e qual a precisão na classificação desses fenómenos?

### **3.2. Objetivos do Estudo**

Após a exposição do problema torna-se possível direcionar um conjunto de questões que daí derivam constituindo os seguintes objetivos:

- a) Caracterizar qual/ quais o(s) melhora(s) conjunto(s) de *feature(s)* para a classificação SVM de instantes antes e depois do *onset*. Saber qual o número necessário de *features* para uma boa classificação;

- b) Conhecer os intervalos onde se situam os parâmetros RBF e como estes variam entre os conjuntos de *features*;
- c) Visto o *onset* a detetar ser aquele que antecede a atividade mais relevante, pretende-se perceber a performance das SVM face a possíveis fenómenos que possam caracterizar diferentes tipos de *onset* (por exemplo a influência da linha de base);
- d) Identificar se os diferentes conjuntos de *features* e o nível técnico dos praticantes influenciam a performance do classificador;
- e) Perceber se o aumento da dimensão da amostra (ao considerar todos os handicap) altera a qualidade do classificador;
- f) Perceber a relação entre a precisão dos modelos SVM e o número de vetores suporte, ou seja, a maior ou menor facilidade com que o classificador conseguiu realizar o seu propósito.

### 3.3. Hipóteses

Para atingir os objetivos propostos têm de ser testadas as hipóteses que se passam a expressar:

*H<sub>1</sub>: Os grupos de handicap não diferem quanto aos diferentes parâmetros C e  $\gamma$  para cada conjunto de features.*

(ou pelo menos, a sua ligeira variância não afeta a precisão de classificação de forma determinante)

*H<sub>2</sub>: Os parâmetros C e  $\gamma$  não diferem entre os três tipos de conjuntos de features em todos os modelos usados.*

(ou pelo menos, a sua ligeira variância não afeta a precisão de classificação de forma determinante)

*H<sub>3</sub>: A classificação dos modelos SVM não difere entre os três tipos de conjuntos de features.*

*H<sub>4</sub>: A classificação dos modelos SVM não é influenciada pelo nível técnico dos praticantes.*

*H<sub>5</sub>: O número relativo de vetores suporte dos modelos SVM não difere entre os três tipos de conjuntos de features.*

*H<sub>6</sub>: O número relativo de vetores suporte dos modelos SVM não é influenciado pelo nível técnico dos praticantes, pelo que não existe interação .*

*H<sub>7</sub>: A performance dos modelos SVM não está correlacionada com o número relativo de vetores suporte em cada conjunto de features.*

### **3.4. Sujeitos**

Doze golfistas destros foram recrutados para análise do presente estudo, distribuídos por dois grupos segundo o handicap (*Hc*): seis sujeitos com handicap baixo ( $Hc = 1.4 \pm 2.5 < 5$ ) e seis sujeitos com handicap alto ( $Hc = 24.6 \pm 4.2 > 18$ ). A determinação do handicap foi baseada no sistema da European Golf Association (EGA, 2012). Os golfistas foram instruídos para realizarem alternadamente quatro *swings* com um taco *ferro 7*, comum para os golfistas independente do *Hc* (taco sete para distâncias  $> 150$  m). Como por repetição foram considerados 1001 instantes após a extração de *features*, totalizaram-se 24024 instantes (6 sujeitos x 1001 instantes por repetição x 4 repetições) a classificar por *Hc* e 48048 no total. As repetições foram realizadas sobre

tapete de relva artificial com características de absorção. Foi permitido aos sujeitos a realização de swings de aquecimento em conformidade com os hábitos de cada sujeito. Todos os sujeitos assinaram termo de consentimento informado (apêndice 1) dando a sua concordância em participar neste estudo tendo sido explicados todos os procedimentos (apêndice 2). A caracterização da amostra foi realizada com recurso à adaptação portuguesa de um questionário específico à prática de golfe em relação aos hábitos de atividade física e condição músculo-esquelética (Fox, Lindsay, & Vandervoort, 2002). A tabela 1 apresenta a caracterização da amostra tendo em conta as variáveis idade, experiência (tempo de prática), massa corporal, altura, e naturalmente, o nível técnico (handicap).

**Tabela 1 – Caracterização dos participantes**

	<b>Hc</b>	<b>Mínimo</b>	<b>Máximo</b>	<b>Média</b>	<b>Desvio-Padrão</b>
<b>Idade (anos)</b>	Alto	35.0	62.0	46.3	11.3
	Baixo	21.0	41.0	32.2	7.60
	Total	21.0	62.0	39.3	11.8
<b>Tempo Prática (anos)</b>	Alto	2.0	15.1	7.4	6.1
	Baixo	12.0	22.0	19.3	3.9
	Total	2.0	22.0	13.4	7.9
<b>Massa (kg)</b>	Alto	50.0	108.0	75.5	21.4
	Baixo	60.0	79.0	68.8	6.2
	Total	50.0	108.0	72.2	15.4
<b>Altura (m)</b>	Alto	1.50	1.83	1.68	0.12
	Baixo	1.68	1.82	1.71	0.05
	Total	1.50	1.83	1.70	0.09
<b>Handicap (Hc)</b>	Alto	18.0	29.0	24.6	4.2
	Baixo	-1	4.5	1.4	2.5
	Total	-1	29	13.0	12.6

## 3.5. Instrumentos

### 3.5.1. Processamento da eletromiografia

Os sinais de EMG foram recolhidos com elétrodos ativos (Plux, Lisboa, Portugal) e equipamento de telemetria bioPLUX® research 2010 (Plux, Lisboa, Portugal). Os dados EMG foram recolhidos a uma frequência de amostragem de 1000 Hz e amplificados a uma banda passante entre 10 e 500 Hz. Após armazenamento dos dados, estes foram filtrados digitalmente a uma banda passante de 10 – 490 Hz, procedendo-se à posterior retificação e suavização. A suavização foi realizada a 12 Hz, com filtro Butterworth de 4ª ordem. Os sinais EMG foram normalizados pela máxima amplitude do conjunto de repetições para cada sujeito. O processamento foi executado em MATLAB ® software V.R2010a (Mathworks Inc., Natick Massachusetts, USA). Todos os sinais EMG em bruto foram submetidos a inspeção visual antes de se iniciar o seu processamento.

Após a adequada preparação da pele (depilação, abrasão e limpeza com álcool), os elétrodos foram colocados na zona de maior proeminência do ventre muscular após contração nos seguintes músculos bilateralmente (Hermens et al., 1999; Horton et al., 2001):

- Bicípite femoral (BF): a 50% da linha que une a tuberosidade isquial e o epicôndilo lateral da tíbia;
- Semitendinoso (ST): a 50% da linha que une a tuberosidade isquial e o epicôndilo medial da tíbia;
- Grande Glúteo (GG): a 50% da linha que liga o sacro ao grande trocânter no fémur;
- Massa Comum (MC): 3 cm lateralmente do processo espinhoso da L3.

Foram ainda monitorizados unilateralmente do lado esquerdo:

- Reto femoral (RF): a 50% da linha que une a espinha ilíaca antero-superior ao bordo superior da patela;

- Vasto interno (VM): a 80% da linha que une espinha ilíaca antero-superior ao espaço articular frontal do bordo anterior do ligamento medial;
- Vasto externo (VL): a 2/3 da linha que liga a espinha ilíaca antero-superior ao bordo lateral da patela;
- Oblíquo externo (OE): 15 cm lateralmente ao umbigo.

O elétron terra foi colocado sobre o manúbrio.

### **3.5.2. Processamento vídeo**

Três câmaras de alta velocidade Basler A602fc (Basler Vision Technologies, Ahrensburg, Germany) de 100 Hz foram colocadas em localização anterior, posterior e superior oblíqua. Uma quarta câmara Casio Ex-FH20 camera (Casio, Tokyo, Japan) de 1000 Hz foi colocada frontalmente à bola com o intuito de determinar o momento de impacto. Nos tacos foram colocadas duas marcas refletoras (Horton et al., 2001) para permitir posterior identificação das três principais fases do swing (Bechler et al., 1995; Pink, Perry, & Jobe, 1993; Watkins et al., 1996). (1) Backswing – do início do swing até ao topo do swing, sendo a fase de preparação; (2) Downswing – do topo do backswing até ao momento do impacto, sendo a fase de execução; (3) Follow-Through – do impacto até ao final do swing, pelo que se caracteriza pela fase do resultado. Os vídeos das repetições realizadas foram captados e posteriormente processados com o software SIMI 3D Motion system (SIMI Reality Motion System GmbH, Unterschleissheim, Germany).

### **3.6. Segmentação e Extração de *Features***

A segmentação corresponde a um processo que remove de forma continuada janelas temporais que compreendem conjuntos de amostras do sinal EMG. A janela temporal contém 200 amostras que correspondem a 200 ms, visto a taxa de amostragem ser de 1000 ms. Esta dimensão foi suportada pela contribuição de Oskoei e Hu (2008)

que consideraram 200 ms como a fronteira entre utilização de segmentação contínua ou *overlapped*. Apesar da segmentação *overlapped* estar associada a valores superiores a 200 ms, foi considerado este valor por ser uma janela fronteira entre os dois tipos de segmentação, como também, devido ao objetivo deste estudo reportar-se à deteção do *onset*, ou seja de um momento específico. Desta janela temporal espera-se que esteja contida a informação que rodeia um ponto central dentro desse domínio quando aplicada a extração de *features*. O *lag* considerado entre cada janela foi de 5 ms.

Após a realização da segmentação do sinal procedeu-se à extração de *features* tendo sido estudadas seis tipos de *features* no domínio do tempo (Phinyomark, Phukpattaranont, & Limsakul, 2012): valor absoluto médio (MAV – *mean absolute value*), comprimento do formato da onda (WL – *waveform length*), diferença absoluta do desvio padrão (DASDV - *difference absolute standard deviation value*), variância do EMG (VAR – *variance of EMG*), integral do EMG, (IEMG - *integrated EMG*) e o detetor logarítmico (LOG - *Log detector*).

- Valor absoluto médio (MAV)

Supondo um conjunto de amostras  $\{x_1, x_2, \dots, x_N\}$  dentro de uma janela temporal representados de dimensão  $N$ , corresponde ao valor absoluto médio do sinal dessa janela, o MAV é fornecido:

$$MAV = \frac{1}{N} \sum_{i=1}^N |x_i| \quad (48)$$

Neste caso, a janela temporal definida corresponde determinada parcela de sinal em torno do instante considerado.

- Comprimento do formato da Onda (WL)

Corresponde ao comprimento cumulativo da forma de onda durante determinado segmento de tempo. O WL está relacionado com a amplitude da forma de onda, tendo

em conta a frequência e o tempo e pode ser visto como uma espécie de diferença finita de aproximação ao valor absoluto da primeira derivada (sem normalização pelo intervalo de tempo), sendo dada por:

$$WL = \sum_{i=1}^{N-1} |x_{i+1} - x_i| \quad (49)$$

- Diferença absoluta do desvio padrão (DASDV)

O desvio padrão é calculado com a diferença à média dada pela norma euclidiana das diferenças entre valores consecutivos, normalizando-as pelo comprimento do intervalo de tempo.

$$DASDV = \sqrt{\frac{1}{N} \sum_{i=1}^{N-1} |x_{i+1} - x_i|^2} \quad (50)$$

- Variância do EMG (VAR)

Foi uma das primeiras formas de quantificar o sinal EMG, tal como através do MAV. Usualmente, é definida como o valor médio do quadrado do desvio-padrão do sinal. Tendo em conta que a média do sinal EMG tende para zero, pode ser calculada recorrendo à soma dos quadrados dos valores do sinal a dividir pela dimensão do intervalo menos 1.

$$VAR = \frac{1}{N-1} \sum_{i=1}^N x_i^2 \quad (51)$$

- Integral EMG (IEMG)

O integral EMG é definido como a área que se encontra sob a curva do sinal de EMG retificado (valores absolutos), ou seja, consiste no integral do valor absoluto do sinal de EMG bruto. Consiste na determinação da área delimitada pela curva do sinal retificado (Pezarat-Correia & Mil-Homens, 2004).

$$IEMG = \sum_{i=1}^N |x_i| \quad (52)$$

- Detetor logarítmico (LOG)

É um detetor de não-linearidade que proporciona uma estimativa da força de contração do músculo com base no logaritmo.

$$LOG = e^{\frac{1}{N} \sum_{i=1}^N \log(|x_i|)} \quad (53)$$

Consideramos também vários conjuntos de *features*, de diferentes dimensões. A formação de conjuntos de *features* em grupo de dois (F2), quatro (F4) e seis (F6) foi baseada na ponderação pelo *Fisher Score* (Duda, Hart, & Stork, 2001) e no algoritmo *Correlation-based Feature Selection* (Hall, 1999), para ordenar as *features* quanto à sua importância para a classificação.

O algoritmo *Fisher Score* (FS) atribui uma ordenação a um conjunto de *features* consoante a sua adequação para a classificação, em função do índice de Fisher. Dado um vetor de rótulos  $y = \{y_1, y_2, \dots, y_c\}$  contendo  $c$  classes, o FS para cada *feature*  $i$  é definido por:

$$FS(f_i) = \frac{\sum_{j=1}^c n_j (\mu_{i,j} - \mu_i)^2}{\sum_{j=1}^c n_j \sigma_{i,j}^2} \quad (54)$$

onde  $\mu_i$  é a média da feature  $f_i$ ,  $n_j$  o número de amostras na  $j$ ésima classe,  $\mu_{i,j}$  e  $\sigma_{i,j}$  correspondem à média e à variância de  $f_i$  na classe  $j$ , respetivamente (Zhao et al., 2010).

O algoritmo *Correlation-based Feature Selection* (CFS), ao contrário do FS, considera o pressuposto que a seleção de *features* para aprendizagem automática deve-se processar tendo em conta a correlação entre as mesmas (Hall, 1999), de forma a beneficiar algoritmos comuns. Se a correlação entre cada um dos componentes de um teste em relação a determinada variável a classificar for conhecida, a inter-relação entre cada par de componentes para classificação pode ser estimada recorrendo a

$$M_s = \frac{k\bar{r}_{cf}}{\sqrt{k + k(k-1)r_{ff}}} \quad (55)$$

onde  $M_s$  é o mérito heurístico do conjunto de *features*  $S$ , dependendo da correlação entre a soma das componentes e a variável a classificar e  $k$  é o número de componentes.  $\bar{r}_{cf}$  é a média das correlações entre as componentes em relação à variável a classificar, e  $r_{ff}$  corresponde à média da inter-correlação entre as componentes. O numerador fornece informação sobre quanto certo conjunto de *features* discrimina determinada classe, o denominador caracteriza a redundância (Hall, 1999).

As *features* foram ponderadas segundo os dois algoritmos de seleção de *features* para cada músculo tendo sido realizado o seguinte ordenamento:

- 1) Ordem de ponderação segundo o FS;
- 2) Pontuação de cada *feature* segundo resiliência ao surgir como discriminatória no classificador CFS.

### 3.7. Máquinas de Vetores Suporte

Neste ponto não será incluída uma descrição exaustiva deste método matemático de aprendizagem automática visto a mesma já ter sido realizada na revisão de literatura deste estudo. Desta forma, serão dedicados os parâmetros específicos do algoritmo usado para classificação as classes que estão na base da determinação do *onset*. Foi selecionado para cada modelo os melhores parâmetros custo e *kernel* RBF:

$$K(x_i, x_j) = \exp\left(-\gamma\|x_i - x_j\|^2\right), \gamma > 0 \quad (56)$$

Para tal foi realizada uma pesquisa de grelha com *5-fold cross-validation* para determinar o erro de generalização para cada modelo, em duas fases: a pesquisa ampla e refinada. Dos 24024 instantes por *Hc* e dos 48048 para o total foi considerada uma relação de 80% para treino e 20% para teste. O custo  $C$  e *kernel*  $\gamma$  na pesquisa pelos melhores valores compreenderam os seguintes intervalos:

Pesquisa ampla:  $C = 2^{-2:1.25:8}$ ;  $\gamma = 2^{-7:1.25:3}$ .

Pesquisa refinada: dependendo dos valores obtidos ao melhor valor determinado na pesquisa ampla foi colocado um espaçamento  $h = 0.25$  tanto à esquerda como à direita desse valor.

Os procedimentos anteriormente referidos quando à escolha da função *kernel* e intervalos de pesquisa são baseados na literatura da especialidade (Hsu, Chang & Lin, 2010).

Cada modelo foi construído tendo como base 24024 amostras, tanto para alto como baixo handicap, e 48048 quando inclusão de ambos os grupos de handicap no mesmo modelo SVM.

Todo o processamento de segmentação, extração de *features*, treino e teste SVM, assim como, a confirmação da detecção do *onset* foi executado em MATLAB ® software V.R2010a (Mathworks Inc., Natick Massachusetts, USA).

### 3.8. Tratamento Estatístico Complementar

O tratamento estatístico tido como complementar refere-se à estatística de inferência, para além da apresentação descritiva. Os pressupostos de normalidade foram confirmados recorrendo ao teste Shapiro-Wilk.

A MANOVA não paramétrica foi aplicada para verificar diferenças entre os grupos Hc em relação aos parâmetros  $C$  e  $\gamma$  que servem de base para construir os modelos. Para este teste foi realizada uma transformação por ranks. A estatística de teste  $\chi^2$  para a MANOVA não paramétrica foi calculada recorrendo ao Traço Pillai, tendo sido corrigido o  $p\_value$ . O teste de Friedman foi aplicada aos parâmetros custo  $C$  para dados não separáveis SVM e ao *kernel*  $\gamma$  usados nos diferentes grupos de *features* (F2, F4 e F6), tendo sido realizadas as respetivas comparações múltiplas.

A ANOVA mista foi usada considerando como fatores de medidas repetidas os diferentes conjuntos de *features* e fatores independentes os grupos de handicap (alto, baixo e ambos os grupos de handicap) para a precisão de classificação e percentagem de vetores de suporte dos modelos SVM. A homogeneidade das variâncias foi testada pelo teste de Levene. As comparações múltiplas foram realizadas como teste de Bonferroni para as medidas repetidas e com o teste de Tukey para amostras independentes. A esfericidade foi verificada com o teste de Mauchly e quando esta não se verificou, os graus de liberdade foram corrigidos pelo Épsilon de Greenhouse-Geisser.

A existência de correlação entre a precisão de classificação e o número relativo de vetores suporte usados em cada modelo desenvolvido foi executada com recurso ao coeficiente de Pearson.

O nível de significância considerado foi de 5% sendo corrigido pelo número de ANOVAS quando necessário. Este tratamento estatístico foi realizado no programa IBM-SPSS 19.0 (IBM Corporation, New York, USA).

### 3.9. Desenho de Investigação

A figura 6 ilustra o fluxograma após recolha dos sinais EMG, incluído o processamento já mencionado anteriormente e sintetizando o desenho de investigação. Considera-se um estudo exploratório observacional analítico

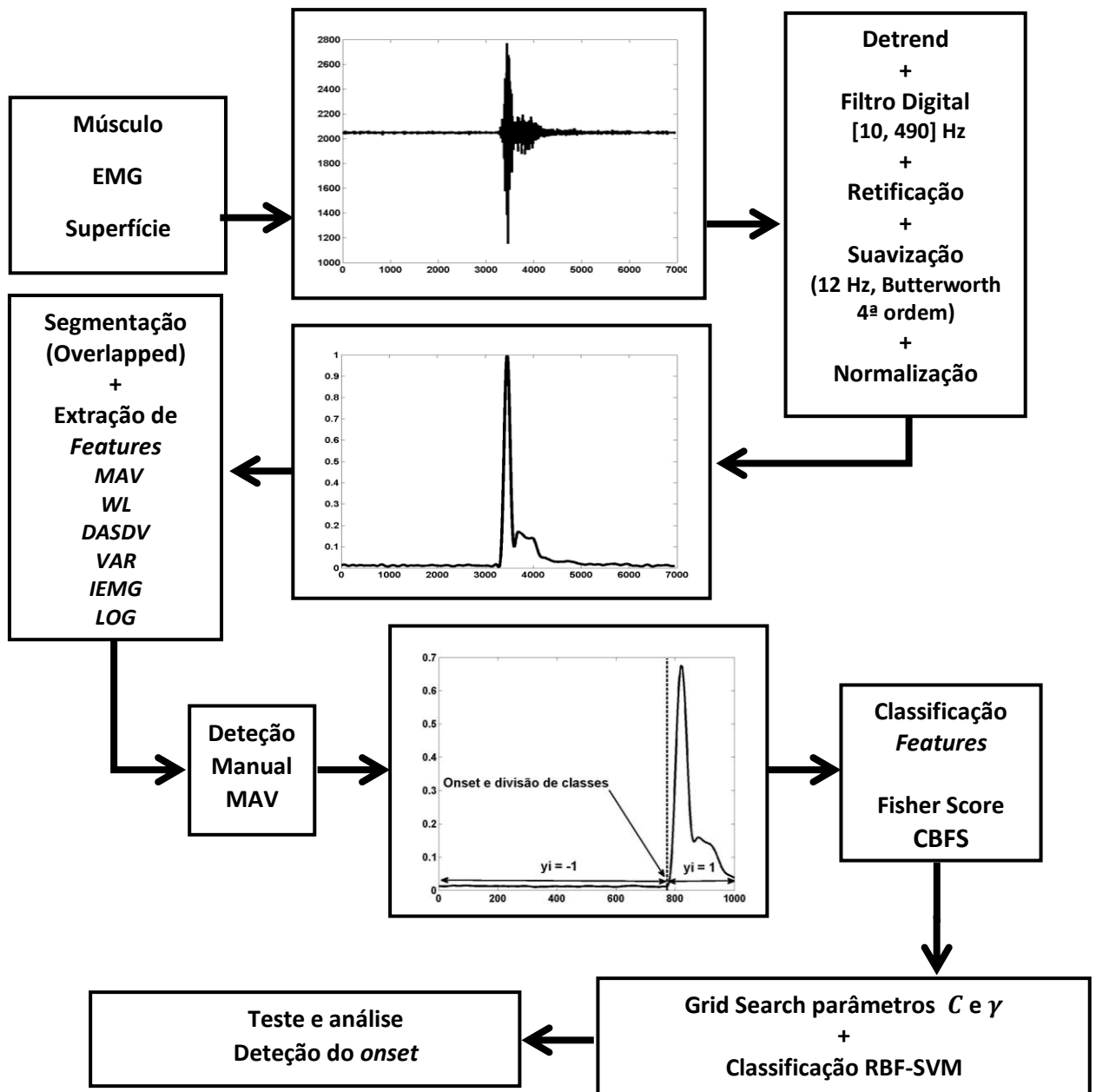


Figura 6 – Fluxograma do desenho de investigação sobre o tratamento EMG.

### 3.10. Limitações ao Estudo

Como limitações ao estudo podem ser indicados os fatores que influenciam a EMG expressos na revisão de literatura deste estudo, presentes em qualquer investigação que recorra a sinais bioelétricos como a EMG. Os mesmos foram devidamente controlados em concordância com a literatura da especialidade (De Luca, 1997; Konrad, 2005), de forma a procurar minimizar o seu efeito.

Estando controladas as limitações inerentes à recolha de EMG, as principais limitações encontradas estão associadas ao fator tempo. A decisão de considerar 1000 instantes após processamento de extração de *features* por repetição, existindo um processamento de corte posterior à segmentação e extração de *features*, esteve relacionada com o custo computacional no processo de construção dos modelos SVM. Optou-se por aumentar o custo computacional quando realizada a segmentação e extração de *features* para depois o evitar quando aplicadas as amostras ao processo de treino SVM. Apesar de este procedimento também apresentar vantagens, não se considera cada instante do sinal em bruto, mas uma representação do mesmo. A outra vantagem refere-se à quantidade relativa de cada classe para treino, pois temos maior número de amostras no repouso comparativamente com o período de atividade.

Independente do exposto no parágrafo anterior, a perda de dimensão também é verificada em relação ao número de repetições usada. Quando juntos os dois Hc obtêm-se 48048 instantes de amostras, chegando alguns modelos a demorarem entre 24 - 48 horas. Pode ser referido que este processo é realizado apenas uma vez, depois do modelo construído, a classificação de novas repetições é relativamente rápida.

Outra limitação que tende a ser uma máxima na maioria dos estudos é a dimensão da amostra, algo presente nos estudos de eletromiografia devido ao peso laboratorial envolvido. Igualmente, a divisão dos sujeitos por handicap representa também por si uma limitação, visto de uma recolha mais alargada não ser possível a utilização de todos os sujeitos.



## 4. Resultados

### 4.1. Seleção de *Features*

As *features* consideradas relevantes para a classificação pelo CFS foram a MAV, WL, DASDV e VAR, sendo as restantes *features* tidas como redundantes, análise esta consistente para todos os músculos. O FS reforçou estes resultados variando apenas em alguns músculos a ordem entre a WL e DASDV em relação à segunda e terceira classificadas. Através do FS, as *features* foram classificadas segundo variável categórica de 1 a 6 (1 melhor e 6 pior). O peso ilustrado na figura 7 foi calculado pelo rácio obtido entre somatório da classificação caso esta fosse sempre 6, pelo somatório dos resultados categóricos obtidos para essa *feature*. O objetivo foi facilitar a leitura quanto à capacidade de melhor generalização quantificando a performance dessa *feature*.

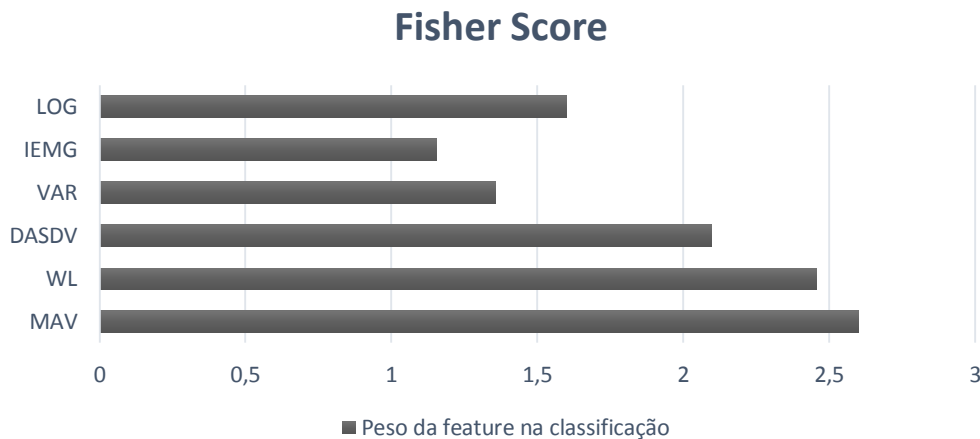


Figura 7 - Peso de cada *feature* na classificação segundo o Fisher Score.

A figura 8 ilustra o CFS quanto ao número de vezes que uma *feature* foi incluída num conjunto tido como o mais robusto. O conjunto obtido maior número de vezes foi [MAV, WL, DASDV, VAR].

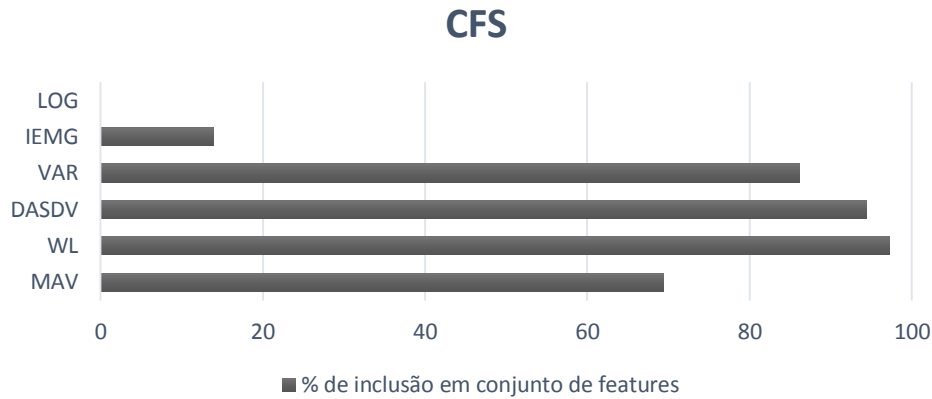


Figura 8 - Relação do número de vezes que uma feature foi incluída no grupo mais robusto.

Desta forma foram contruídos três conjuntos de *features* que serviram de vetores característicos ao modelo das SVM, tendo sido promovida a uniformidade entre os diferentes músculos para posterior comparação em relação aos diferentes grupos de handicap e dimensão da amostra:

**F2** – [MAV, WL], como o grupo das melhores classificadas, sendo a MAV a de maior relevância;

**F4** – [MAV, WL, DASDV, VAR], grupo mais relevante com a VAR a menos relevante;

**F6** – [MAV, WL, DASDV, VAR, IEMG, LOG], grupo que inclui as *features* mais relevantes e as redundantes.

## 4.2. Parâmetros Custo $C$ e Kernel $\gamma$

Para determinar os parâmetros custo  $C$  e *kernel*  $\gamma$  da função RBF foram realizadas pesquisas de grade (*grid search*) para cada músculo, para cada conjunto de *features* e para cada um dos grupos considerados neste estudo. Os resultados para estes parâmetros serão descritos segundo a mediana [mín, max]. No global dos músculos e ignorando qual o fenômeno identificado, se *onset burst* ou *onset peak*, após a pesquisa

refinada foram obtidos para os grupos de alto, baixo, ambos os Hc e total dos três tipos de modelo, os valores expressos nas tabelas 2 e 3.

**Tabela 2 – Parâmetro custo  $C$  para os modelos SVM**

$C$	F2	F4	F6
	Mediana [mín, máx]		
Alto	$2^{5.63} [2^{0.75}, 2^{8.05}]$	$2^{8.13} [2^{5.25}, 2^{8.75}]$	$2^{7.63} [2^{5.75}, 2^{8.75}]$
Baixo	$2^{6.02} [2^{-3}, 2^{8.75}]$	$2^8 [2^{6.25}, 2^{8.75}]$	$2^8 [2^{6.75}, 2^{8.75}]$
Total (ambos Hc)	$2^{1.77} [2^{-2.5}, 2^7]$	$2^{8.63} [2^{7.5}, 2^{8.75}]$	$2^{8.63} [2^{7.25}, 2^{8.75}]$
Total (modelos)	$2^{4.52} [2^{-3}, 2^{8.75}]$	$2^{8.34} [2^{5.25}, 2^{8.75}]$	$2^{8.25} [2^{5.75}, 2^{8.75}]$

Hc – handicap; total (ambos Hc) refere-se a modelos construídos tendo como base os dois Hc, tanto alto como baixo; o total (modelos) refere-se ao conjunto dos três grupos (alto, baixo, ambos Hc).

Considerando os resultados da MANOVA não paramétrica, não existem diferenças significativas entre os três grupos de Hc (alto, baixo, ambos Hc) que justifiquem a utilização de diferentes valores  $C$  e  $\gamma$  ( $\chi^2(12) = 16.415; N = 36; p = 0.17$ ) face à população estudada. A questão sobre a construção de modelos SVM para sinais EMG coloca-se então em relação ao número de *features* usadas.

**Tabela 3 – Parâmetro *kernel*  $\gamma$  para os modelos SVM**

$\gamma$	F2	F4	F6
	Mediana [mín, máx]		
Alto	$2^{-3.25} [2^{-5.25}, 2^3]$	$2^{-4.5} [2^{-5}, 2^{-3}]$	$2^{-4.5} [2^{-5}, 2^{-3}]$
Baixo	$2^{0.19} [2^{-6}, 2^{3.5}]$	$2^{-4.5} [2^{-5.5}, 2^{-3.5}]$	$2^{-4.5} [2^{-5.5}, 2^{-3.25}]$
Total (ambos Hc)	$2^{0.38} [2^{-5.5}, 2^3]$	$2^{-4.5} [2^{-5.5}, 2^{-3}]$	$2^{-4.62} [2^{-5.25}, 2^{-3}]$
Total (modelos)	$2^{-0.08} [2^{-6}, 2^{3.5}]$	$2^{-4.5} [2^{-5.5}, 2^{-3}]$	$2^{-4.5} [2^{-4.5}, 2^{-3}]$

Hc – handicap; total (ambos Hc) refere-se a modelos construídos tendo como base os dois Hc, tanto alto como baixo; o total (modelos) refere-se ao conjunto dos três grupos (alto, baixo, ambos Hc).

Na tabela 4 é possível verificar que a utilização de duas *features* difere sempre quando comparada com quatro e seis *features*, assim como, o parâmetro  $C$  não difere se utilizadas mais *features* que o conjunto F4 (ou seja, entre F4 e F6). Quando aplicado o teste de Friedman para analisar se os modelos diferem quanto à utilização de conjunto de *features* observa-se que existem diferenças significativas em relação ao parâmetro custo  $C$  ( $\chi^2(2) = 42.318; p < 0.001, N = 36$ ). As comparações múltiplas evidenciam que essas diferenças derivam da utilização de F2 tanto em relação a F4 como F6 ( $p < 0.001$ ). Os modelos com conjuntos F4 e F6 não diferem entre si ( $p = 1.0$ ).

**Tabela 4 – Teste Friedman e comparações múltiplas parâmetro  $C$**

$C$	<i>Features</i> (F2, F4, F6)		Comparações Múltiplas			
	Grupo	$\chi^2(gl)$	$p$	$(p\_ajustados)$		
				F2	F4	
Alto		9.911 (2)	0.007**	F4	0.013*	-
				F6	0.057	1.0
Baixo		14.683 (2)	0.001**	F4	0.005**	-
				F6	0.009**	1.0
Total (ambos Hc)		18.957 (2)	<0.001**	F4	0.002**	-
				F6	0.000**	1.0
Total (modelos)		42.318 (2)	<0.001**	F4	0.000**	-
				F6	0.000**	1.0

Hc – handicap; gl – graus de liberdade; total (ambos Hc) refere-se a modelos construídos tendo como base os dois Hc, tanto alto como baixo; o total (modelos) refere-se ao conjunto dos três grupos (alto, baixo, ambos Hc).

\*\*Significativo a um nível de significância de 1%.

\*Significativo a um nível de significância de 5%.

Na tabela 5 é apresentada a mesma análise para o parâmetro  $\gamma$ , verificando que apesar de não terem sido encontradas diferenças entre os grupos para os parâmetros *kernel*, quando os modelos construídos utilizaram maior número de amostras (neste caso o dobro), também são precisas mais *features*. Similar ao parâmetro custo  $C$ ,

também foram encontradas diferenças significativas para o *kernel*  $\gamma$  ( $\chi^2(2) = 19.528; p < 0.001, N = 36$ ), sendo novamente os modelos F2 os responsáveis por estas diferenças tanto em relação a F4 ( $p = 0.002$ ) como a F6 ( $p = 0.001$ ). Igualmente, verifica-se similaridade entre os conjuntos F4 e F6 ( $p = 1.0$ ).

**Tabela 5 – Teste de Friedman e comparações múltiplas parâmetro  $\gamma$**

$\gamma$	<i>Features</i> (F2, F4, F6)		Comparações Múltiplas		
	$\chi^2(gl)$	$p$	<i>(p_ajustados)</i>		
				<b>F2</b>	<b>F4</b>
Alto	2.390 (2)	0.303	<b>F4</b>	-	-
			<b>F6</b>	-	-
Baixo	7.860 (2)	0.020*	<b>F4</b>	0.057	-
			<b>F6</b>	0.074	1.0
Total (ambos Hc)	11.128 (2)	0.004 **	<b>F4</b>	0.032*	-
			<b>F6</b>	0.024*	1.0
Total (modelos)	19.528 (2)	<0.001**	<b>F4</b>	0.002**	-
			<b>F6</b>	0.001**	1.0

Hc – handicap; gl – graus de liberdade; total (ambos Hc) refere-se a modelos construídos tendo como base os dois Hc, tanto alto como baixo; o total (modelos) refere-se ao conjunto dos três grupos (alto, baixo, ambos Hc).

\*\*Significativo a um nível de significância de 1%.

\*Significativo a um nível de significância de 5%.

Estes resultados evidenciam que para um total de modelos (aumento da dimensão da amostra) torna-se extremamente necessário aumentar o número de *features* de dois para pelo menos quatro, já que é acentuada a diferença em relação a F2, mas não entre F4 e F6. As figuras 9 e 10 apresentam os diagramas em caixa para os parâmetros  $C$  e  $\gamma$  contabilizando o total dos dois handicaps para os músculos estudados e tendo em conta o handicap e conjuntos de *features*. Ou seja, pode ser visualizado quando a dimensão  $n$  do modelo aumentou de 24024 amostras para 48048, portanto para o dobro, construindo modelos SVM mistos ao incluir os dois grupos de handicap. No domínio

utilizado na pesquisa com espaçamento  $h = 0.25$  para  $\gamma$ , tanto a utilização de F4 como F6 apresentam expoentes negativos evidenciando similaridade entre os dois conjuntos.

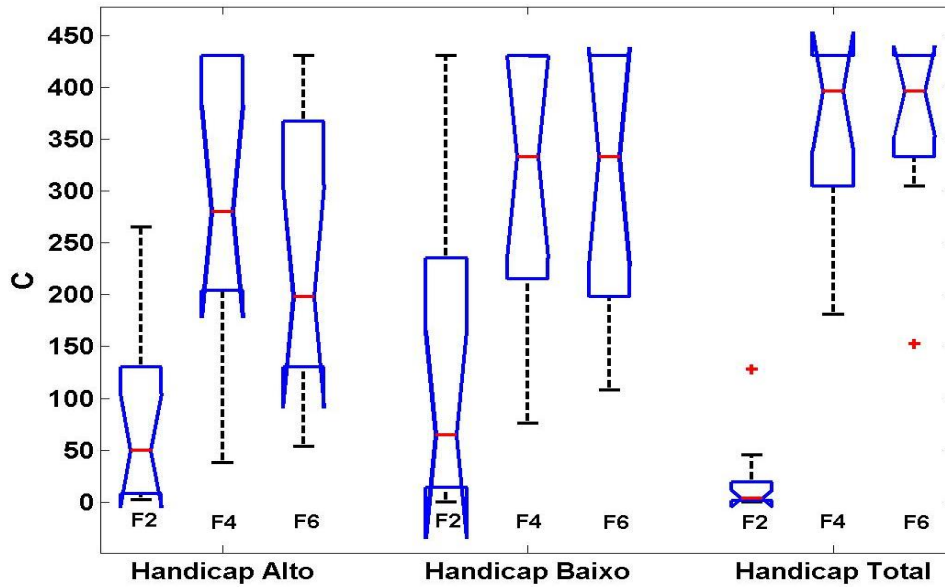


Figura 9 – Parâmetro C segundo handicap e número de features.

Na figura seguinte é bem visível a instabilidade para o parâmetro  $\gamma$  encontrada com o uso de duas *features* apenas, apesar de menor dispersão quando se usa modelos de maior dimensão de amostras.

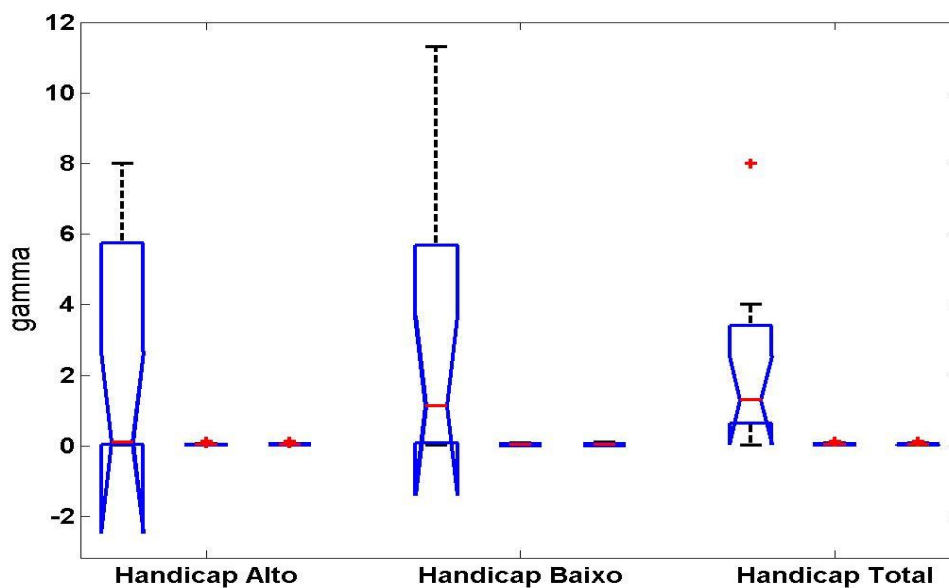


Figura 10 – Parâmetro  $\gamma$  segundo handicap e número de features.

Especificando a análise à realidade de dados para cada músculo, procede-se à apresentação dos resultados tendo em conta as pesquisas de procura dos melhores parâmetros  $C$  e  $\gamma$ . Visto a definição operacional depender da atividade mais relevante, podem ser encontradas diferentes características na deteção do *onset*, nomeadamente duas: a primeira quando a deteção tem como base o *onset burst* coincidindo este com o principal pico de atividade. A segunda refere-se a uma determinação do *onset* quando existe mais ruído, o *onset peak* quando o músculo já se encontra com atividade considerável. Neste caso o *onset* refere-se a um momento “desliga/liga” desse músculo.

As figuras 11 e 12 representam a pesquisa de grelha ampla para o RF esquerdo, ou seja, com um espaçamento entre expoentes  $h = 1.25$ . Neste caso temos apenas o início de ativação, mas este precede um pico máximo de atividade EMG, pode-se referir que o *onset burst* coincide com o *onset peak*. Verifica-se uma melhor performance do algoritmo quando aplicado a sujeitos baixo Hc, ou seja, aqueles que apresentam melhor nível técnico. Tem-se maior número de valores de  $C$  e  $\gamma$  a garantirem uma classificação elevada, tanto com a utilização de F4 como F6.

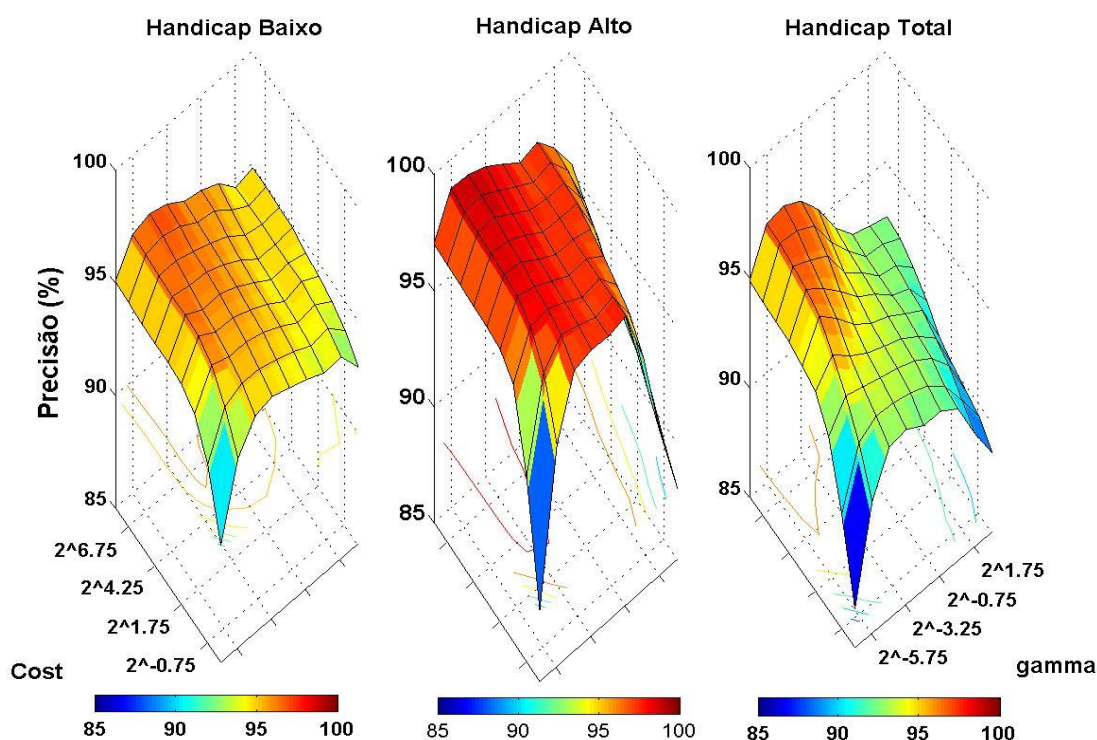


Figura 11 – Pesquisa de grelha dos parâmetros  $C$  e  $\gamma$  com 4 features para o reto femoral esquerdo.

As pesquisas de grelha amplas ilustradas pelas figuras 11 e 12 demonstram igualmente a similaridade em usar conjunto de quatro e seis *features* em conformidade com o exposto anteriormente em relação aos valores de  $C$  e  $\gamma$ .

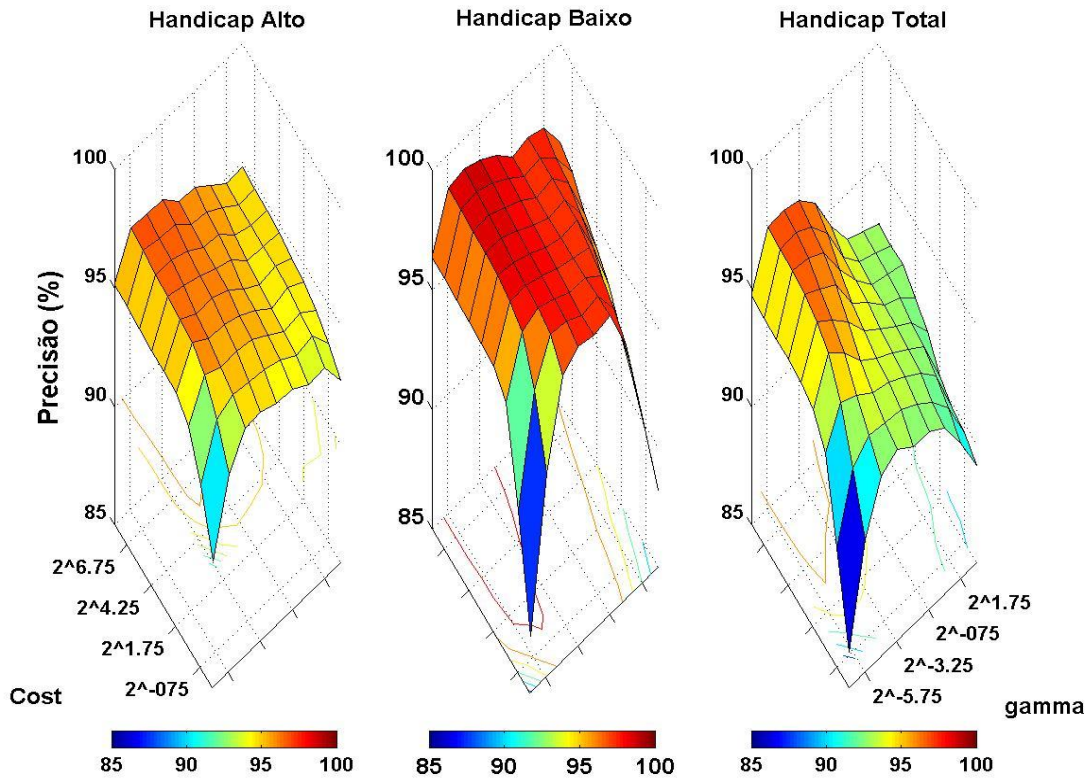


Figura 12 – Pesquisa de grelha dos parâmetros  $C$  e  $\gamma$  com 6 *features* para o reto femoral esquerdo.

O algoritmo para o RF esquerdo atingiu entre 99.02% e 99.05% de precisão para o mesmo valor  $C = 2^{8.75}$ , passando o parâmetro  $\gamma$  por valores entre  $2^{-5.25}$  e  $2^{-5}$  em relação às classificações mencionadas. Observa-se facilmente que para F4 e F6 tem-se tendencialmente um  $C$  elevado e um  $\gamma$  reduzido.

As figuras 13 e 14 ilustram resultados para a MC do lado direito. Este músculo tem a característica de já se encontrar com uma atividade muscular relativa em relação ao momento segundo o qual foi realizada a detecção do *onset*. Como anteriormente referido, para a MC, o *onset* considerado foi um momento de “liga/desliga” padrão na maioria das repetições, ou seja, um *onset peak* quando o músculo já está em atividade.

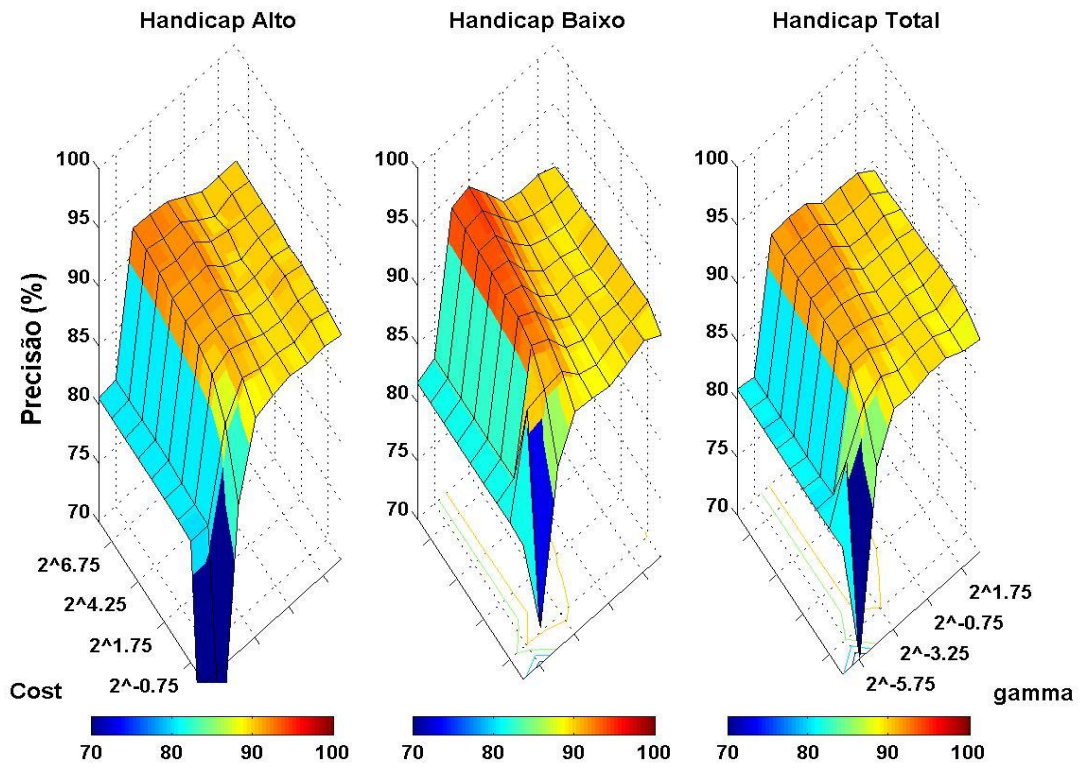


Figura 13 – Pesquisa de grelha dos parâmetros  $C$  e  $\gamma$  com 4 features para a massa comum lado direito.

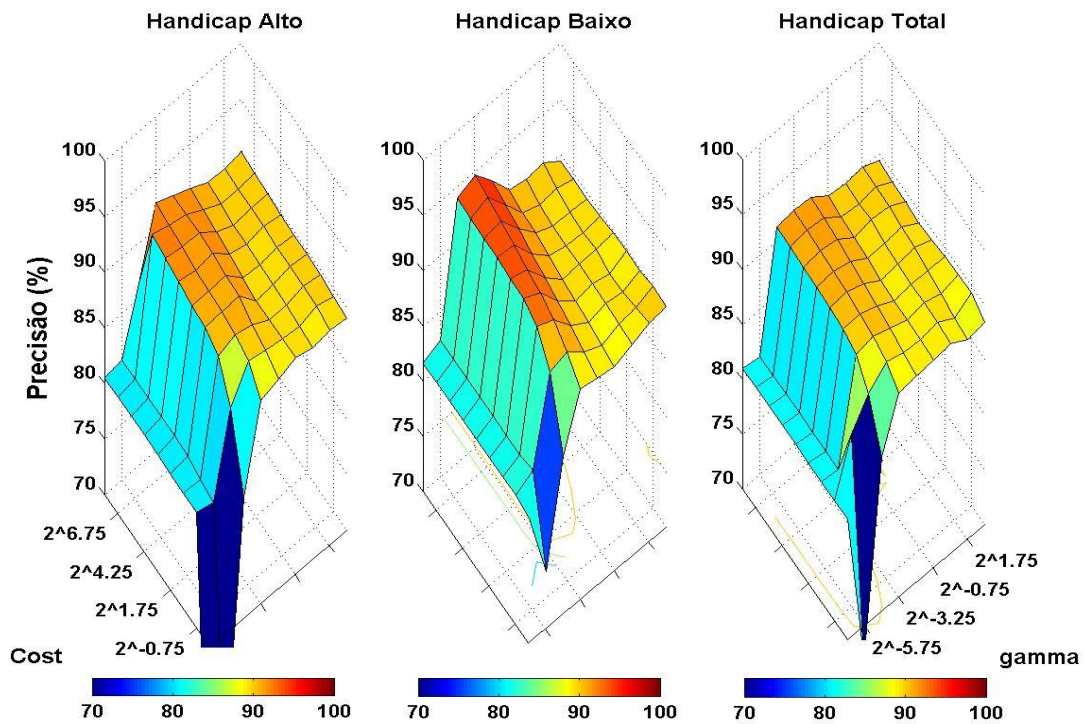


Figura 14 – Pesquisa de grelha dos parâmetros  $C$  e  $\gamma$  com 6 features para a massa comum lado direito.

### 4.3. Classificação e Detecção do *Onset*

Após o conhecimento de quais os valores ideais para o C-SVM RBF procedeu-se à utilização dos mesmos passando por dois pontos essenciais, a precisão quanto ao número de pontos classificados e a verificação regional de detecção do *onset*. Neste último, em conformidade com o fenómeno estudado procedeu-se à elaboração de um algoritmo de retroversão para o sinal em milissegundos.

Considerando todos grupos sem diferenciação do fenómeno temporal determinado pelo algoritmo verifica-se uma classificação para a utilização de F2, para os grupos de alto, baixo e total handicap, de  $90.3 \pm 4.0\%$ ,  $90.8 \pm 4.9\%$  e  $89.4 \pm 3.7\%$ , respetivamente. Quando aplicadas F4 e F6, os valores médios sobem na ordem dos 5%. Com F4, a precisão média da classificação foi de  $94.9 \pm 2.5\%$  para alto handicap,  $95.0 \pm 3.3\%$  para baixo handicap, e quando aumentada a dimensão pela junção dos dois grupos os modelos SVM apresentaram uma classificação média de  $93.5\% \pm 3.2\%$ . Quando aumentado o número de *features* para F6, os valores foram similares com  $95.2 \pm 2.4\%$ ,  $95.1 \pm 3.2\%$  e  $93.6 \pm 3.3\%$ , para os grupos alto, baixo e total handicap. A figura 15 apresenta o diagrama em caixa para estes resultados.

Apesar da similaridade entre a utilização de F4 e F6 e face à redundância verificada pelas técnicas de extração de *features* expostas no método do presente trabalho, a ANOVA mista apresentou diferenças significativas entre o tipo de conjunto de *features* utilizadas ( $F_{(1,030,33,984)} = 213.328; p < 0.001; \eta^2 = 0.866; \pi = 1.0$ ) em todos os pares. Através das comparações múltiplas confirma-se a não existência de diferenças significativas entre F4 e F6 ( $p = 0.11$ ), mas com F2 a diferir em relação aos demais conjuntos ( $p < 0.001$ ). Já a interação com os grupos definidos não apresentou um efeito significativo ( $F_{(2,060,33,984)} = 0.435; p = 0.657; \eta^2 = 0.026; \pi = 0.116$ ). Igualmente, não se verificaram diferenças significativas entre os grupos estudados ( $F_{(2,33)} = 0.700; p = 0.504; \eta^2 = 0.041; \pi = 0.158$ ).

Os músculos que evidenciam maior precisão na classificação foram tendencialmente do lado esquerdo, nomeadamente, o ST, VI, RF, VE e OE. Estes músculos evidenciaram uma performance na determinação do *onset* bastante elevada

mesmo para a aplicação de F2 com  $94.0 \pm 1.8\%$ , variando entre 98.0 e 91.5% de precisão.

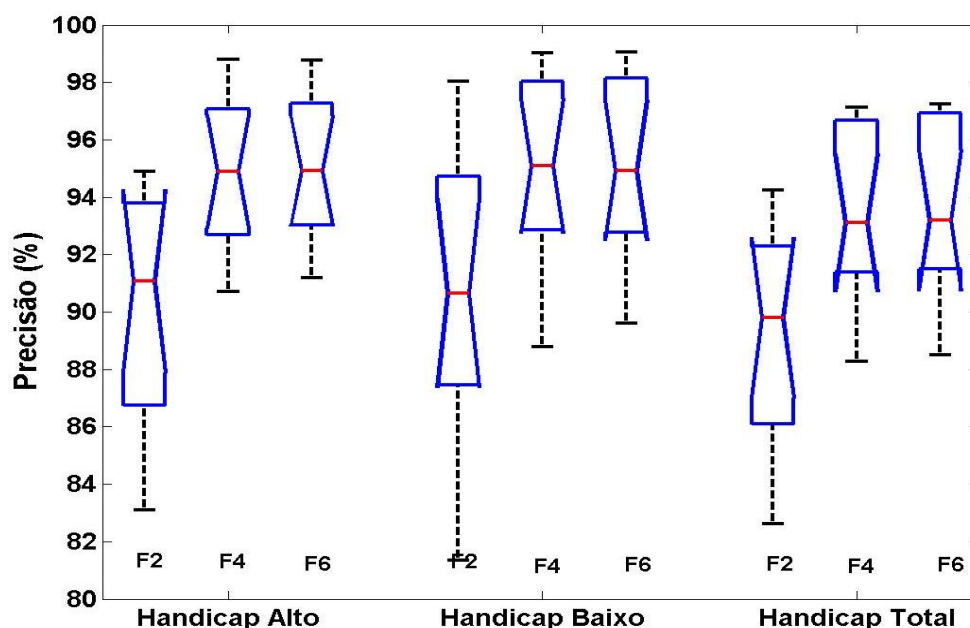


Figura 15 – Classificação por grupo vs modelos.

Quando utilizados os conjuntos F4 e F6, a precisão subiu para  $97.5 \pm 0.9\%$  e  $97.6 \pm 0.9\%$ , respetivamente. O valor mínimo obtido foi de 96.1% atingindo uma precisão máxima de 99.04%. Com exceção do OE, todos estes músculos apresentam o *onset burst* coincidente com o *onset peak*, variando apenas na inclinação de crescimento da onda ou por existir alguma pré ativação. No OE esquerdo não foi considerado nenhum pico principal de atividade tendo sido determinado sempre o *onset burst*. Para os restantes músculos, a utilização de apenas duas *features* já evidenciou um menor poder discriminativo, alcançando no entanto  $87.4 \pm 2.9\%$  de precisão. Nos conjuntos F4 e F6 verificou-se uma precisão de classificação de  $92.3 \pm 2.0\%$  para o primeiro e  $92.5 \pm 1.9\%$  quando utilizados modelos F6. Deste grupo de músculos, apenas o ST direito e o GG direito rondaram os 88% quando aplicados modelos baseados em 48048 amostras (ambos os grupos). Os restantes modelos apresentaram classificação com precisão superior a 90%.

Passa-se a expor os diferentes fenómenos *onset* que foram detetados ao considerar o *onset* como aquele que inicia a atividade com maior relevância.

### 4.3.1. Determinação com *onset burst* igual a *onset peak*

As figuras 16 e 17 ilustram a determinação do *onset burst* quando coincidente com o pico máximo de atividade, denominado por *onset peak*. Também são evidenciados, para além do *onset*, pontos mal classificados durante o período em que ocorre a linha de base. Notamos que nesta abordagem consideramos um instante de *onset*, qualquer instante de tempo em que a classificação passe de -1 a 1.

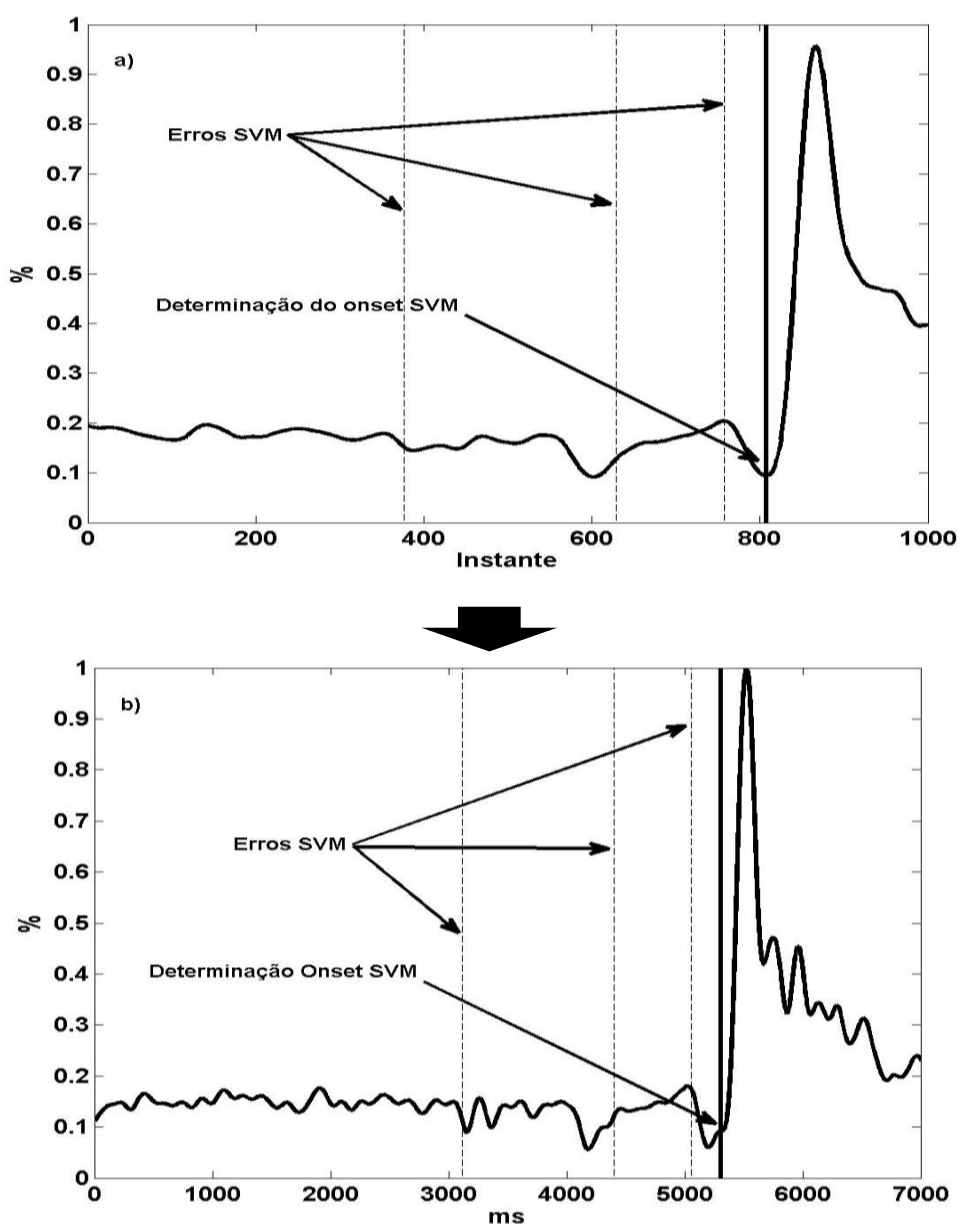


Figura 16 – Determinação do onset para o vasto interno com modelo SVM-F4L para sujeitos de baixo handicap a) Marcação após segmentação e extração de features; b) Transposição sinal retificado.

Assim, existem mais instantes que são considerados erradamente como de *onset*, o que pressupõe a necessidade de mais *features* ou de algoritmos de pós-processamento, que considerem por exemplo, a distância de cada instante ao pico de sinal máximo. De relevar também, que o instante de *onset* é sempre bem classificado em todos os casos, apesar de existirem outros falsos positivos. Na figura 17 é possível observar um sinal com uma linha de base bastante limpa com elevada precisão na determinação do *onset*.

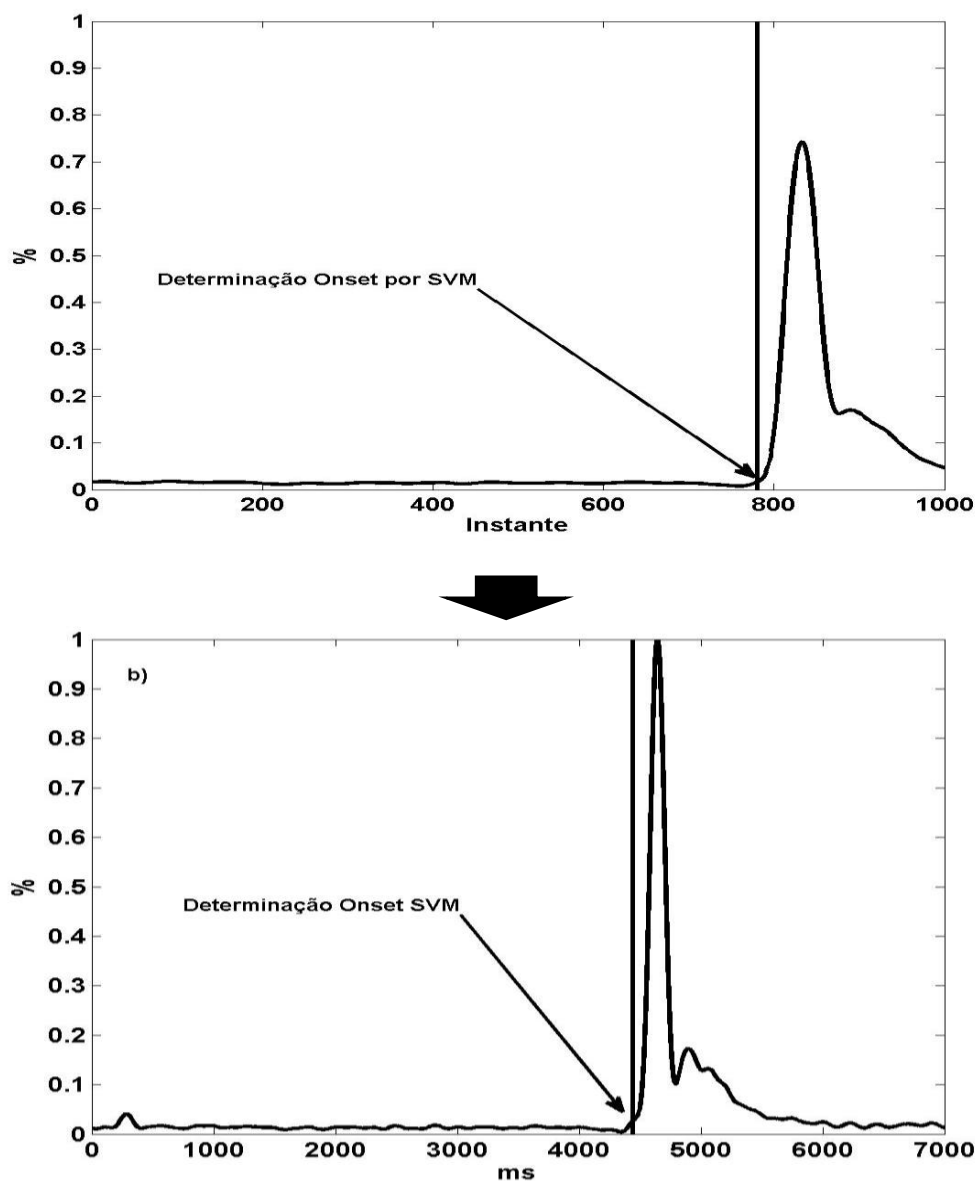


Figura 17 – Determinação do onset para o reto femoral esquerdo com modelo SVM-F4L para sujeitos de baixo handicap a) Marcação após segmentação e extração de features; b) Transposição sinal retificado.

A relação entre a linha de base e o sinal de atividade muscular para o RF esquerdo foi acentuada no geral das repetições o que justifica o elevado poder classificatório por parte dos modelos SVM. Neste tipo de sinais ou com ligeiras pré ativações a precisão do algoritmo é bastante elevada, tendo sido verificado que os pontos com erros de classificação tendem a ocorrer durante a linha de base com rácios sinal-ruído baixos.

#### 4.3.2. Determinação do *onset burst*

Neste ponto relembra-se que *onset burst* refere-se à definição tradicional de *onset*, quando o músculo ativa pela primeira vez em relação a determinada habilidade motora. Sempre que se verificaram repetições de teste com rácio sinal/ruído elevados, ou seja, uma linha de base com pouco ruído em relação ao período de atividade, como ilustrado na figura 18, os erros de classificação surgiram após o ponto considerado como *onset*. Esta situação é contrária ao encontrado anteriormente para o VI. A figura seguinte ilustra o OE esquerdo, um músculo que ativa muito próximo do início do movimento tendo apresentado uma das melhores precisões de classificação.

Os resultados obtidos em fenómenos onde o *onset* é bem definido, como o caso do OE onde se verificou a não existência de um pico definido, ou de músculos como o RF que apresentam o *onset burst* e *peak* como o mesmo fenómeno, ambos evidenciam uma precisão muito elevada. Mesmo com a utilização de apenas duas *features* o grupo de alto Hc obteve uma precisão de  $93.2 \pm 1.8\%$ , o grupo de baixo Hc apresentou  $94.7 \pm 2.4\%$ , e a junção de ambos os Hc chegou a uma precisão de  $92.3 \pm 1.7\%$ . Os resultados aumentam com a utilização de F4 e F6. Para a utilização de F4 encontram-se precisões de  $96.8 \pm 1.7\%$ ,  $97.7 \pm 1.6\%$  e  $96.7 \pm 1.4\%$ , para os grupos de alto, baixo e ambos Hc, respetivamente. Quando aplicadas F6 os resultados são aparentemente similares com  $97.8 \pm 1.5\%$  para o alto Hc,  $96.9 \pm 1.4\%$  para o baixo Hc, e para ambos a precisão foi de  $96.9 \pm 1.4\%$ . Assim, considerando apenas estes dois fenómenos foram obtidas as classificações de  $93.4 \pm 2.1\%$  para F2,  $96.7 \pm 1.6\%$  para F4 e  $96.1 \pm 2.2\%$  para F6 em relação ao total de modelos testados.

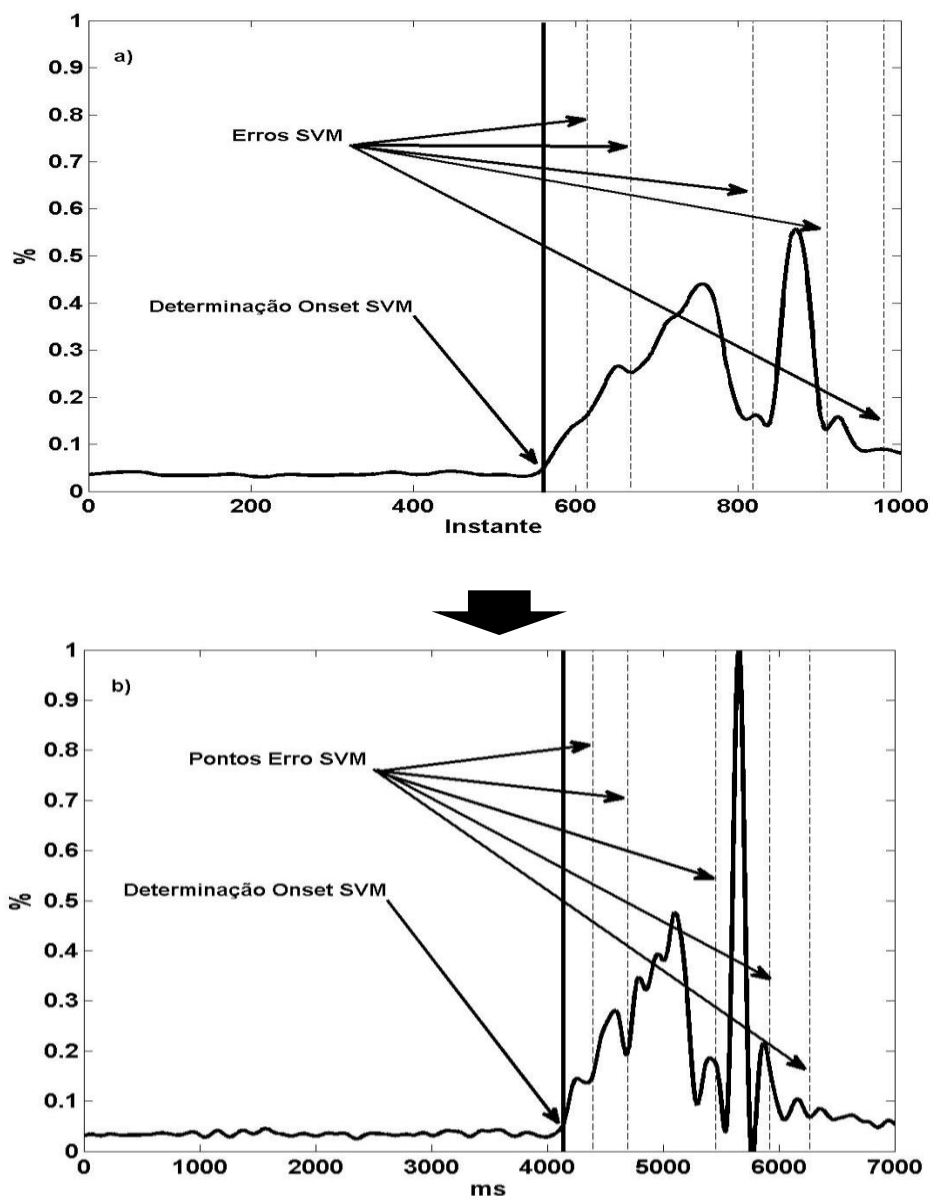


Figura 18 – Determinação do onset para o oblíquo externo esquerdo com modelo SVM-F6H para sujeitos de baixo handicap a) Marcação após segmentação e extração de features; b) Transposição sinal retificado.

Neste caso, não existe um pico definido e de curta duração, como observado anteriormente, mas uma atividade irregular ao longo do sinal e com uma dimensão temporal superior. No entanto, independente da performance classificativa observa-se precisão no ponto que corresponde ao *onset*.

### 4.3.3. Determinação do *onset peak* com atividade precedente

As repetições que apresentam maior número de erros são aquelas onde o rácio sinal/ ruído é menos elevado e cujo momento a determinar corresponde a um *onset peak* quando já existe atividade considerável precedente. Um bom exemplo refere-se à massa comum (figura 19), pois este grupo muscular já está com atividade considerável no início do movimento na maioria dos sujeitos sendo verificado um *onset peak* padrão.

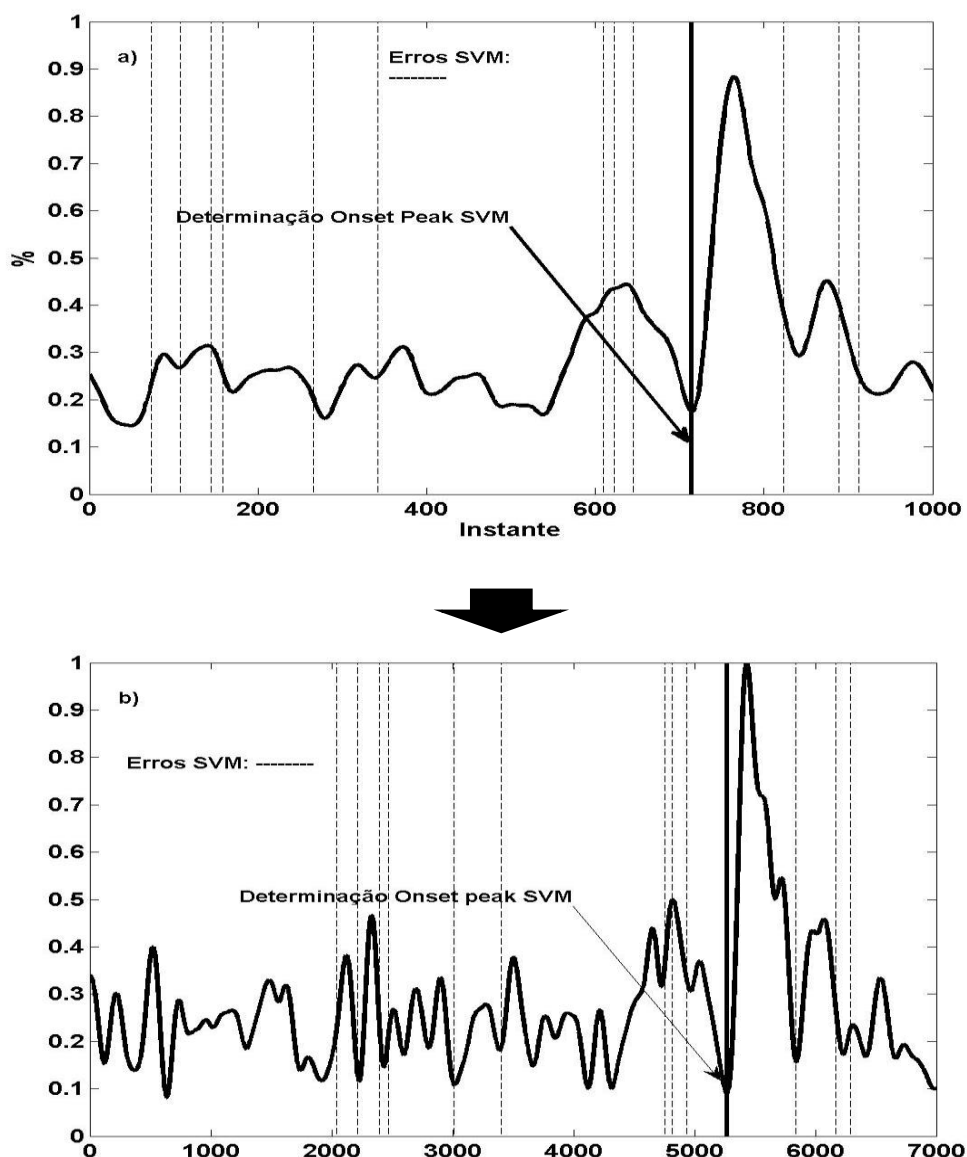


Figura 19 – Determinação do *onset peak* para a massa comum com modelo SVMT-F4T para sujeitos de baixo handicap a) Marcação após segmentação e extração de features; b) Transposição sinal retificado.

Em músculos como a MC de ambos os lados, GG direito e ST direito, já ocorreu ativação muscular associada a outra ação dentro da mesma habilidade motora que não a tida como principal, mas esse *onset peak* é detetado com precisão, apesar do elevado número de falsos positivos.

#### 4.4. Percentagem Vetores Suporte

Na percentagem de vetores suporte necessária para a classificação foram obtidos resultados similares ao verificado anteriormente para a precisão em relação às *features* e aos grupos de modelos. Os conjuntos de *features* diferem significativamente ( $F_{(1.0069,35.269)} = 29.523; p < 0.001; \eta^2 = 0.472; \pi = 1.0$ ) não havendo uma interação com os diferentes grupos que constituem os modelos ( $F_{(2.137,35.269)} = 0.410; p = 0.801; \eta^2 = 0.024; \pi = 0.139$ ). Quando aplicadas F2 foram requeridos  $35.8 \pm 11.0\%$ ,  $37.3 \pm 17.6\%$  e  $38.7 \pm 13.4\%$  para handicap alto, baixo e total, respetivamente. Quando aplicados os conjuntos F4 e F6, a percentagem de vetores suportes decresce cerca de 10% para alto handicap (F4 –  $26.3 \pm 7.6\%$ ; F6 –  $28.9 \pm 6.5\%$ ), para o baixo handicap (F4 –  $26.8 \pm 7.6\%$ ; F6 –  $26.5 \pm 9.6\%$ ) e para o total dos grupos (F4 –  $29.2 \pm 5.9\%$ ; F6 –  $29.4 \pm 6.2\%$ ).

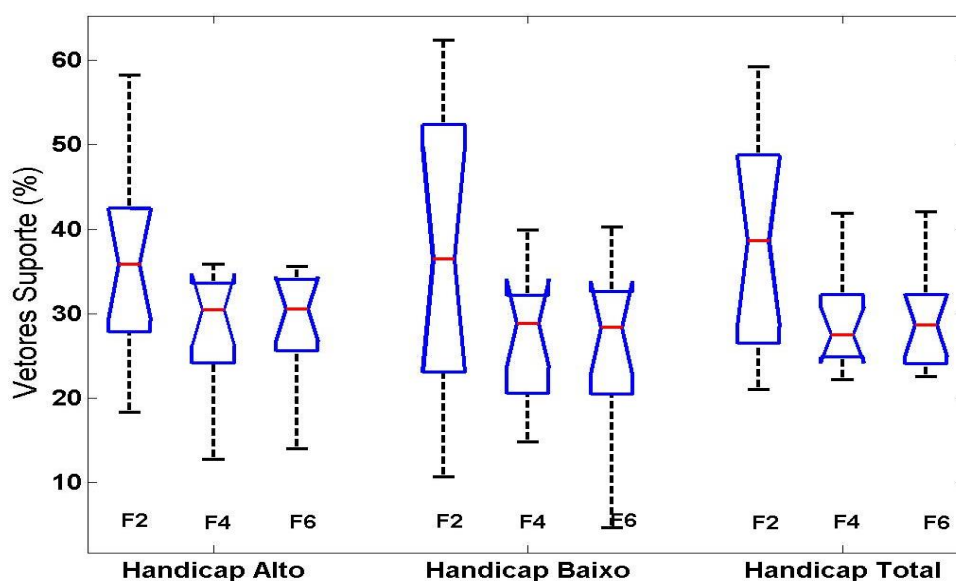


Figura 20 – Percentagem de vetores suporte necessários para cada modelo.

Através das comparações múltiplas podemos observar que a utilização de F6 não significa menor esforço por parte do algoritmo em relação a F4 ( $p = 1.0$ ), sendo a composição F2 que difere das demais ( $p < 0.001$ ). Tal como para a precisão de classificação, para além de não existir interação entre a utilização de diferentes grupos na construção de modelos, estes também não diferem entre si ( $F_{(2,33)} = 0.217; p = 0.806; \eta^2 = 0.013; \pi = 0.081$ ). As figuras 21, 22 e 23 são os diagramas de dispersão que relacionam as percentagens obtidas na classificação e a percentagem de dados que serviram como vetores suporte. Pode observar-se que existe uma tendência para menor necessidade quanto à quantidade de vetores suporte quando a precisão é maior, ou seja, quanto menor foi a precisão da classificação maior a quantidade relativa de vetores suporte. Esta tendência é mais acentuada com a utilização de F2 comparativamente com F4 e F6.

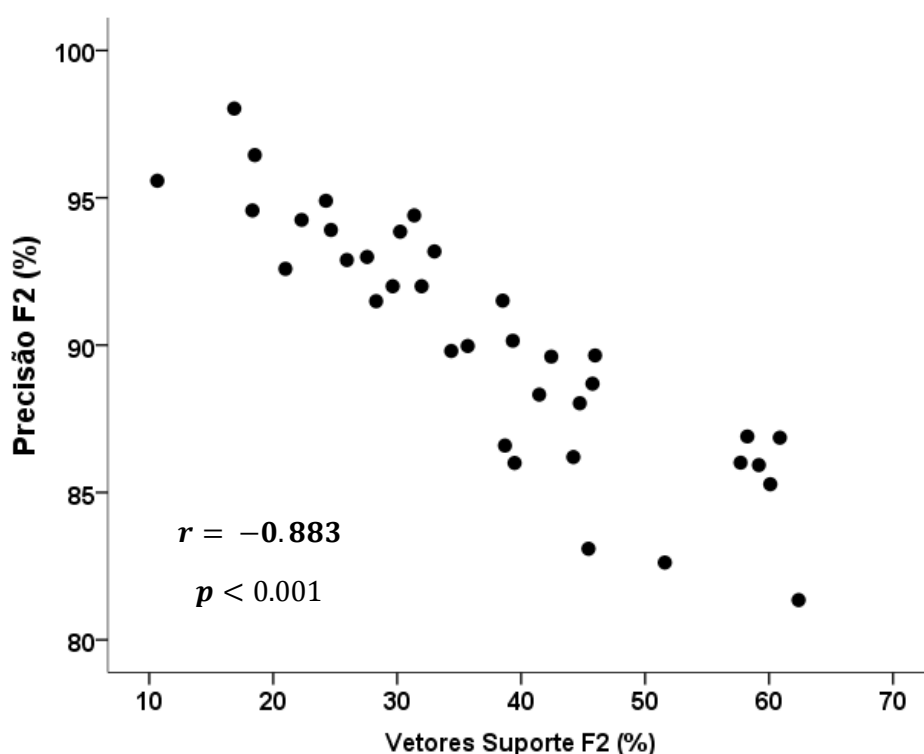


Figura 21 – Diagrama de dispersão da classificação vs vetores suporte para F2.

O músculo que utilizou menor número de vetores suporte no conjunto F2 foi o oblíquo esquerdo em sujeitos de baixo handicap recorrendo a 10.7% dos pontos de treino. Neste caso, como já foi possível observar o fenómeno determinado foi o *onset*

*burst*. O modelo F2 para o músculo ST do lado direito foi o que utilizou mais vetores suporte ao atingir 62.4%. O ST do lado esquerdo contraria totalmente estes resultados quando utilizados modelos F4 e F6, tendo sido aquele onde foi encontrado menor número de vetores suporte no primeiro caso, com 12.7%, e tendo 14% quando aplicado um modelo F6 em sujeitos com baixo handicap. Com o modelo F6, apenas são utilizados menos vetores suporte no OE do lado esquerdo ao necessitar apenas de 5% dos dados de treino.

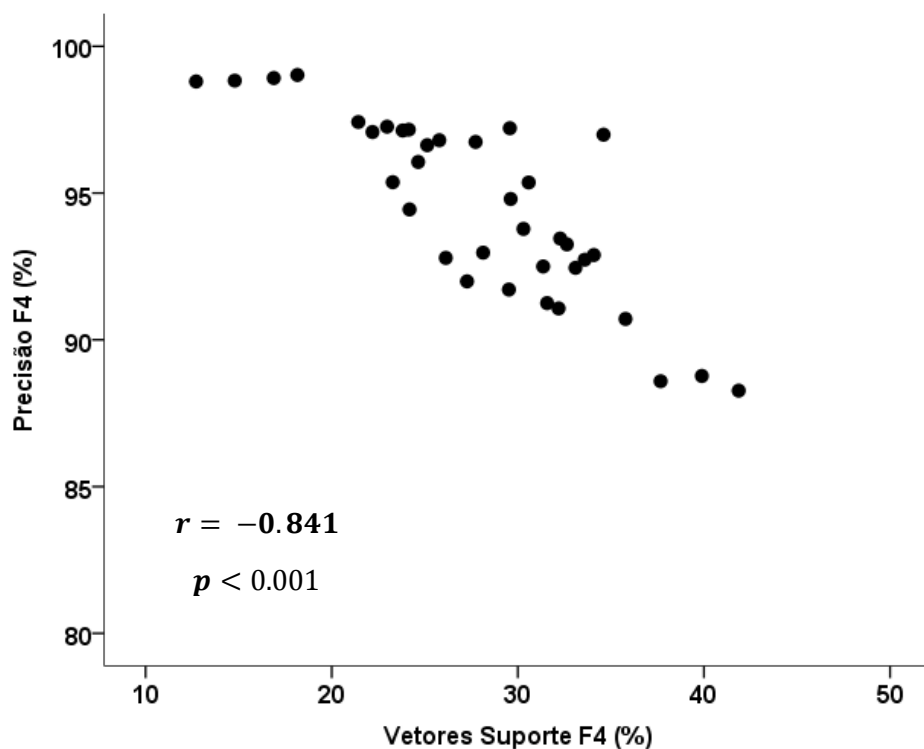


Figura 22 – Diagrama de dispersão da classificação vs vetores suporte para F4.

O OE é também o músculo com um número de vetores de suporte mais baixo conjuntamente com o ST esquerdo quando usados modelos F4, tendo sido usados 14.8% dos dados de treino. Para os modelos F4 e F6, a utilização máxima de vetores de suporte rondou os 42%. Os músculos que mais se aproximaram deste valor foram o ST nos sujeitos de baixo handicap e o BF no total das amostras, ambos do lado direito.

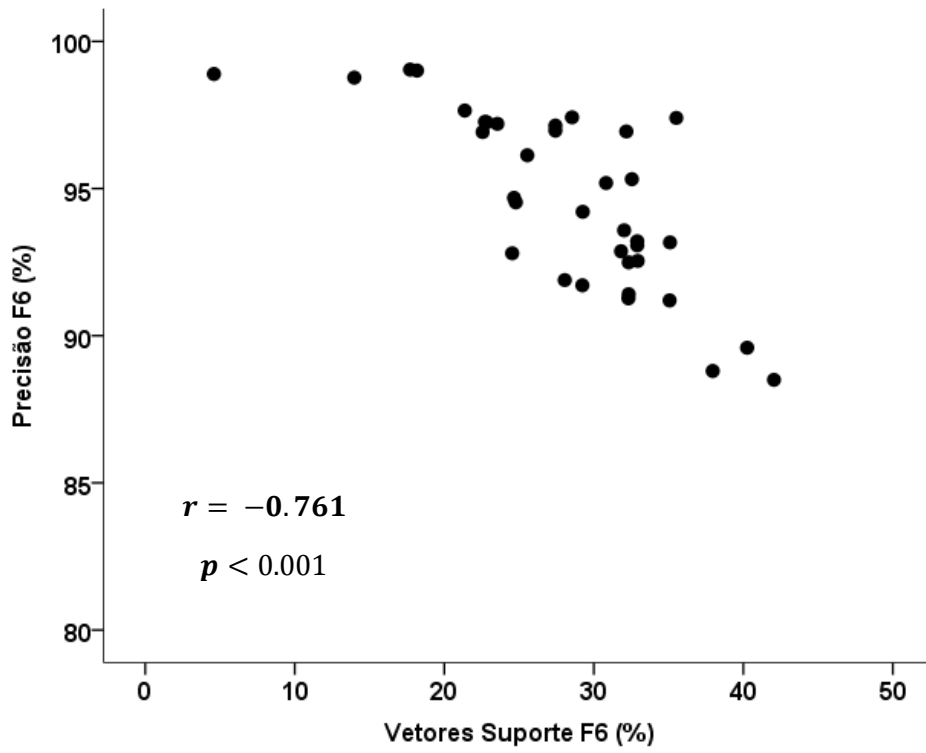


Figura 23 – Diagrama de dispersão da classificação vs vetores suporte para F6.

## 5. Discussão

O objetivo deste estudo consistiu em verificar se o reconhecimento de padrões temporais EMG através de SVM é fiável para a deteção do *onset* usando *features* no domínio do tempo. É pretendido ter sustentação teórica para a construção de modelos futuros que detetem o fenómeno temporal *onset burst* ou determinado *onset peak* permitindo discutir sequencialmente os comandos motores em diferentes tarefas. Neste caso particular, foi estudado o *swing* do golfe por ser uma habilidade motora dinâmica complexa que requer tanto precisão como velocidade de execução.

O primeiro ponto a realçar refere-se à falta de reprodutibilidade devido à utilização de diferentes algoritmos e/ou parâmetros dentro do mesmo tipo de algoritmos (Jöllenbeck, 2000; Morey-Klapsing et al., 2004; Silva et al., 2013). Dentro das várias propostas apresentadas têm-se discutido o automatismo e qualidade de deteção deixando de fora a análise do fenómeno fisiológico em si. Este aspeto está bem presente no estudo de McGill et al. (2010) sobre o fenómeno do duplo pico, sendo realizada uma apresentação do sinal sincronizado com a respetiva imagem do movimento. Quando se menciona o conceito de *onset*, este tende a referir-se apenas ao início da ativação muscular através do registo EMG, sendo por si redutor. Este aspeto deve-se por considerar uma habilidade motora como um todo, não traduzindo a totalidade de informação que pode estar contida no sinal, como também, devido à grande dependência do algoritmo de deteção e dos seus parâmetros de entrada. Ao considerar uma habilidade motora como um todo podem estar a ser ignoradas as diferentes ações motoras que aí estão contidas e que têm relevância para a análise do fenómeno fisiológico. Foi este o motivo que levou a outra definição, o *onset peak*, como um *onset* que ocorre antes de um pico de atividade EMG tido como relevante (Silva et al., 2013).

Outra situação que se verifica na literatura é uma tendência de estudar contrações musculares isométricas, de forma a evitar influência nefasta do movimento da pele e a alteração do local de deteção presente em ações dinâmicas (De Luca, 1997) e pela facilidade de deteção quando testado um algoritmo. Porém, o preço a pagar é a falta de inferência para ações dinâmicas, fundamentais no movimento desportivo e objetivo no treino funcional quando se o pretende ganhos clínicos. Face à necessidade

de recorrer à inspeção visual para validar a precisão dos diferentes algoritmos automáticos fundamenta-se a construção de modelos matemáticos de aprendizagem automática que já contenham a informação quanto ao fenómeno a estudar. Estando os diferentes fenómenos representados por classes é necessária a seleção de *features* que contenham informação representativa de cada classe a classificar.

### **5.1. Conjunto de *Features* no Domínio do Tempo**

Neste estudo utilizaram-se três conjuntos de *features* identificados como F2, F4 e F6, tendo a sua construção tido como base a ponderação obtida pelo *Fisher Score* (Duda, Hart, & Stork, 2001) e pelo algoritmo *Correlation-based Feature Selection* (Hall, 1999). Tratando-se da análise de uma habilidade motora dinâmica procura-se o conjunto que forneça maior estabilidade dentro da janela temporal que representa. F2 foi constituída pela MAV e pela WL, mas os conjuntos F4 e F6 apresentaram maior precisão. Face à similaridade de classificação entre F4 e F6, pode-se referir que basta a utilização de quatro *features*, quando as demais possam conter *features* redundantes.

Tkach, Huang, & Kuiken (2010) verificaram nos músculos bicípite e tricípite braquial combinações de diferentes *features* no domínio do tempo. As quatro *features* WL, *slope sign change*, LOG e coeficientes autorregressivos foram o conjunto mais estável quanto à variabilidade no nível de esforço exercido. Também um conjunto de quatro *features*, VAR, LOG, v-Order e o histograma EMG, demonstraram maior estabilidade quando a classificação foi sujeita a mudanças na localização de elétrodos. Para além da WL, VAR e LOG, apenas a MAV são comuns entre o presente estudo e o de Tkach, Huang e Kuiken (2010). A classificação obtida nas seleção de *features* antes da introdução no algoritmo SVM colocou como as quatro principais MAV, WL, DASDV e VAR, mas as diferença para LOG e IEMG em relação a VAR não foram muito acentuadas. Outras *features* foram estudadas por estes autores tendo apresentado um elevado poder de classificação, coeficientes autoregressivos e os coeficientes cepstrum. Porém, a inclusão deste tipo de *features* aumentam a dimensionalidade dos vetores característicos, o que por sua vez, aumenta a complexidade na classificação e o custo computacional. Esta foi a razão pela qual não foram incluídas este tipo de *features* no presente estudo. Comparativamente com o conjunto F2, visto ser aquele que possui

as duas primeiras *features* com melhor classificação, estes resultados estão em concordância com os resultados apresentados por Oskoei e Hu (2008). Estes autores verificaram que o conjunto MAV e WL apresentam um elevado poder discriminatório, sendo aquelas com maior estabilidade para diferentes janelas amostrais. O conjunto de quatro *features* que apresentaram melhor classificação foram MAV, WL, *zero crossing* e *slope sign change*, mas das *features* que constituíram F4, DASDV não foi estudada. Porém, quando sujeita a incrementos no número de repetições realizadas para classificar, a *feature* DASDV apresenta uma boa estabilidade e robustez em tarefas de classificação (Phinyomark et al., 2013).

Este estudo pretendeu conhecer a performance de *features* no domínio do tempo na classificação do sinal EMG para a deteção do *onset*. As *features* usadas apresentaram uma capacidade de classificação com um valor médio de 90.2%, 94.5% e 94.6%, para os conjuntos F2, F4 e F6, respetivamente. Visto, as *features* no domínio do tempo apresentarem melhor performance na classificação de sinal EMG quando comparadas com *features* no domínio da frequência (Oskoei & Hu, 2008; Phinyomark et al., 2013), podem servir de base para modelos de classificação, mesmo que sejam agregados outros domínios de *features*. A performance dos diferentes tipos de *features* pode, no entanto, variar face ao classificador utilizado, nomeadamente, *features* no domínio do tempo-frequência (Phinyomark et al., 2013).

A desvantagem das *features* no domínio do tempo está associada ao facto de serem calculadas através da amplitude do sinal sendo incluídas interferências (Phinyomark et al., 2009). Por exemplo, a *feature* IEMG apenas foi pontuada para F6, não estando nos anteriores grupos. Pezarat-Correia & Mil-Homens (2004) referem que este tipo de *feature* apresenta limitações pela reduzida informação descritiva dos componentes da curva. Por exemplo, sofre influência em relação ao tempo de recolha e tipo de contração realizada, pelo que, na comparação de dois sinais na mesma tarefa, pode refletir tanto a intensidade como a duração de ativação.

Chama-se a atenção para uma cautela quanto à escolha das *features*, os estudos acima discutidos incidiram essencialmente no controlo mio-elétrico. Isto é, o objetivo consistiu em classificar movimento com vetores construídos com quantificações de janelas temporais.

## 5.2. Parâmetros SVM

No melhor do nosso conhecimento, as SVM nunca foram utilizadas como instrumento do *onset*, visto este parâmetro ser considerado como um fenómeno que não está no âmbito do reconhecimento de padrões (Oskoei & Hu, 2007). Contrariando, parte-se do pressuposto que os momentos do sinal ao serem representados por *features* podem ser classificados como classe de atividade principal e classe da atividade que ocorre antes desta. Neste estudo foi utilizado o tratamento não linear com a equação *kernel* RBF, a qual tende a ser a equação mais utilizada nos estudos que recorrem às SVM para classificação, tanto a nível do controlo mio-elétrico (Bitzer & Smagt, 2006; Castellini, Smagt, Sandini, & Hirzinger, 2008; Castellini & Smagt, 2009; Chen, Zhu, & Zhang, 2010; Lucas et al., 2008; Yang et al., 2009), como na classificação de desordens neuromusculares (Christodoulou et al., 2012; Katsis et al., 2006; Subasi, 2012, 2013; Sultornsanee et al., 2011), estendendo a sua aplicação à classificação de posturas da mão tendo em conta o envelhecimento (Tavakolan et al., 2011).

A função RBF também é aconselhada em estudos de classificação em mapeamentos não lineares e quando se desconhece qual a precisão de outras equações (Chang & Lin, 2012; Hsu et al., 2010). A vantagem de usar uma função *kernel* segundo  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$  deve-se por permitir construir algoritmos em espaços de dimensão superior, sem ter de os construir efetivamente. Espera-se que uma classe *kernel* possa ser escrita no formato definido positivo para a equação  $k(x_i, x_j) = \langle \Phi(x_i), \Phi(x_j) \rangle$  seja satisfeita e mantida a relação primal-dual do algoritmo SVM (Hofmann et al., 2008; Lin & Lin, 2003). A proposição de Wendland garante que a *kernel* Gaussiana é definida positiva, que é o caso da RBF (Hofmann et al., 2008). Por outro lado, equações como a linear e a sigmoide, dentro de uma determinada amplitude, tendem a comportar-se como a RBF (Lin & Lin, 2003). A *kernel* polinomial usa mais hiperplanos que a RBF quando o número de hiperplanos influencia a complexidade na seleção dos modelos, com a RBF a desenvolver menores dificuldades numéricas (Chang & Lin, 2012). Acresce o facto da dimensão da amostra ser muito elevada e o número de *features* reduzido, quanto à possibilidade de aplicar a *kernel* linear.

A escolha dos parâmetros ( $C, \gamma$ ) depende da matriz de classificação obtida pela pesquisa de rede (*grid search*). Com exceção do grupo F2 que apresentou maior instabilidade entre modelos, verifica-se uma tendência com os melhores parâmetros  $C$  a variarem  $[2^{5.25}, 2^{8.75}]$  e  $\gamma$  do *kernel* RBF entre  $[2^{-6.25}, 2^{-1.25}]$ , tendo em conta espaçamento de 0.25. Verifica-se uma tendência para valores  $\gamma < 1$  com  $C$  a apresentar valores mais elevados rondando  $C = 2^{8.25}$ . Vários estudos que recorreram a classificação RBF-SVM com outros objetivos evidenciam esta tendência, variando essencialmente no parâmetro  $C$ . Por exemplo, no controlo mio-elétrico com EMG recolhido de músculos do antebraço verificam-se valores como  $C = 2^5, \gamma = 2^{-3}$  (Yang et al., 2009),  $C = 2^{[5.5, 6.5]}, \gamma = 2^{[-0.5, 0]}$  (Khokhar et al., 2010), na classificação de desordens neuromusculares com  $C = 2^{\approx 8.25}, \gamma = 2^{\approx -2.32}$  (Subasi, 2012), e na classificação de movimentos da mão comparando uma população idosa com jovens apresentando  $C = 2^{\approx 3.828}, \gamma = 2^{\approx -0.32}$  (Tavakolan et al., 2011) e  $C = 2^{6.25}; \gamma = 2^{-2.25}$  (Katsis et al., 2006). Outras abordagens têm sido apresentadas em que o parâmetro  $\gamma$  é calculado pelo inverso ou com a razão do número de atributos  $\gamma = \frac{1 \text{ ou } 12}{n}$  (Goker et al., 2012; Huang, Zhang, Hargrove, Dou, Rogers, Englehart, 2011; Khushaba, Kodagoda, Liu, & Dissanayake, 2013; Khushaba, Kodagoda, Takturi, & Dissanayake, 2012).

### 5.3. Deteção do *Onset* e Precisão de Classificação

Dos três artigos que exploram a deteção do *onset* no swing do golfe, dois incidem sobre a problemática da lombalgia (Cole & Grimshaw, 2008), o outro compara dois métodos que variam segundo a linha de base que serve de referência (Silva et al., 2013), mas todos aplicaram métodos algorítmicos de limiar. No presente estudo foram considerados diferentes fenómenos de *onset* ao basear a deteção na atividade com maior relevância. Este aspeto conduziu a que os rótulos das classes fossem separados segundo três critérios distintos: *onset burst* coincide com o *onset peak*, *onset burst* sem considerar nenhum pico de atividade em especial, *onset peak* com atividade precedente. O conceito de *onset peak* foi introduzido por Silva et al. (2013) ao considerar que o sinal EMG tem momentos de atividade relevantes que não devem ser descurados. Os artigos sobre o *onset* abordam essencialmente como melhorar a qualidade na sua

deteção, como por exemplo o operador Teager Kaiser (Solnik et al., 2010), a decomposição *wavelet* (Vannozzi, Conforto, & Alessio, 2010), e a análise por *singular spectrum-based change-point* (Vaisman et al., 2010). A exploração dos diferentes métodos está intimamente ligada ao rácio do sinal pelo ruído, pois este influencia a precisão dos resultados, nomeadamente no que se refere ao aparecimento de falsos positivos na classificação de instantes de *onset*. Desta forma, a atividade da linha de base é apontada como um dos principais fatores de distúrbio na deteção do *onset* por parte de algoritmos automáticos (Allison, 2003; Hodges & Bui, 1996; Silva et al., 2013). Então, a justificação para a maior ou menor facilidade de classificação dos modelos de classificação, tal como os demais algoritmos, estará associada ao ruído gaussiano branco que se verifica na linha de base (Lee, Cholewicki, & Reeves, 2007). Por sua vez, o poder de classificação está dependente da capacidade das *features* representarem cada uma das classes. Então, o erro de classificação das diferentes *features* que alimentam o classificador tende a diminuir consoante o rácio sinal ruído aumenta (Phinyomark et al., 2009). Esta situação está bem presente nos casos em que o fenómeno temporal a detetar foi o *onset peak* quando existe registo de atividade elétrica precedente. O ES foi o músculo onde este fenómeno aconteceu para todos os sujeitos sendo por isso considerado um padrão, em conformidade com investigação prévia (Silva et al., 2013). Em músculos cujos fenómenos detetados foram coincidentes com o *onset burst* e a linha de base não apresenta ruído gaussiano branco muito acentuado chega-se a obter para uma repetição apenas 3 a 5 erros de classificação, estando no entanto, igualmente incluído o verdadeiro *onset* como o momento de transição de uma classe para a outra.

## 6. Conclusões e Recomendações

A utilização de quatro *features* do domínio do tempo é suficiente para determinar o início de ativação muscular com precisão superior a 95% na classificação dos pontos temporais. Desta forma, *features* no domínio do tempo como o valor absoluto da média do sinal, comprimento da onda, desvios padrão do sinal e variância do sinal podem servir de base na construção de algoritmos de aprendizagem automática na detecção do *onset*.

Quanto aos grupos de handicap poderem diferir em respeito aos parâmetros SVM, tal não se confirma. Aumentos amostrais, ou seja, a inclusão de mais exemplos de treino acentua as diferenças quando usados grupos de quatro ou seis *features* quando comparados com a utilização de apenas duas *features* no domínio do tempo. Os parâmetros custo e *kernel* apresentam diferenças entre a utilização de apenas duas *features*, sendo indiferente o uso de quatro ou seis *features* no domínio do tempo em relação aos parâmetros identificados na *grid-search*.

A classificação realizada pelos modelos SVM e o número relativo de vetores suporte não difere por serem realizados grupos específicos de handicap. Mais uma vez, as diferenças encontradas dizem respeito às *features* utilizadas. A utilização de duas *features* requer maior número de vetores de suporte e apresenta pior classificação do que a utilização de quatro ou seis *features*. Modelos com melhor precisão apresentam menor número relativo de vetores suporte, significando que quanto maior a facilidade de classificação melhor será a performance do algoritmo.

A definição de *onset* neste estudo foi o início de atividade mais relevante e padrão nas repetições estudadas. Tal procedimento conduziu à identificação de três fenômenos de *onset*, os quais se traduziram na maior ou menor dificuldade de classificação explanadas no parágrafo anterior. Qualquer *onset burst*, coincidente com o maior pico de atividade ou não, apresenta elevado poder de detecção. Quando detetado um *onset peak* já com atividade precedente, a precisão desce aumentando a dificuldade de detecção e os erros obtidos (apesar de ter sido identificado o *onset* pretendido). Desta forma, duas recomendações relacionadas com as máquinas de vetores suporte têm de ser

evidenciadas. A primeira refere-se ao tipo de fenómeno estudado, se o *onset burst* ou *peak*. Quando é pretendida a deteção do *onset burst*, a utilização de máquinas de vetor suporte pode ser binária classificando os pontos como pertencentes à classe repouso ou à classe atividade.

Quando o objetivo for a deteção de determinado *onset peak*, pode ser útil a utilização de máquinas de vetores suporte multiclases onde estão incluídas várias classes temporais ou *features* com outro tipo de informação. Podendo usar meramente o caso binário concorre uma perda de unicidade na classificação que pode corresponder à existência de maior número de erros na classificação, o que por sua vez, dificulta a construção de algoritmos que especificam quais dos pontos são o *onset* pretendido, mesmo que este também esteja marcado. A segunda recomendação surge no seguimento desta problemática de detetar o *onset peak*. Este estudo fornece referências para uma base de *features* sustentada no domínio do tempo, o que não quer dizer que seja única. Recomenda-se a realização de estudos que englobem para além de uma base de *features* no domínio do tempo também *features* de carácter fenomenológico, como a distância de cada ponto a determinada referência ou transformada *wavelet*.

Uma das grandes vantagens deste tipo de abordagem refere-se também à descrição do fenómeno a estudar, ao contrário de existir a mesma definição operacional para o *onset*, mas sem reprodutibilidade devido aos algoritmos de deteção utilizados. Pelo descrito, consideram-se como producentes os seguintes passos na aplicação deste método:

- (1) Descrever a definição operacional do fenómeno temporal a ser detetado (determinado *onset peak*, *onset burst*,...);
- (2) Decidir o tipo de segmentação (caso seja incluído), assim como, em caso de segmentação *overlapped* definir o valor do atraso incluído;
- (3) Fundamentar os métodos de extração de *features* (próprio sinal, domínio do tempo, fenomenológicas);
- (4) Verificar para o músculo em causa qual o momento de transição de classe que deve ser considerado como o *onset* desejado;

- (5) Selecionar as *features* relevantes e eliminar as que possam ser redundantes, ou utilizar a proposta neste estudo no conjunto de quatro *features* no domínio do tempo;
- (6) Realizar a pesquisa de rede (*grid-search*) e determinar o valor custo e parâmetro *kernel* mais vantajoso, ou então, usar os domínios expostos no presente estudo. Os mesmos deverão ser aceites caso se verifique uma boa pontuação classificativa;
- (7) Testar os modelos em relação a novas entradas de dados.

Recomenda-se o estudo de modelos que tendo como base *features* no domínio do tempo, visto garantirem uma performance acima de 90%, seja aumentado o número de *features* através de variáveis no domínio do tempo-frequência e fenomenológicas. Os resultados deste estudo conduzem ao pressuposto que estando garantido pelo menos um valor médio de 95% na precisão da classificação com *features* no domínio do tempo, caso sejam acrescentadas *features* com outro tipo de informação, como fenomenológicas, podem-se obter resultados mais próximos de 100%. Uma classificação de 100% evita a necessidade de algoritmos posteriores à classificação para extração do *onset*. Uma proposta de aplicação simples será utilizar *features* que incluam informação sobre a distância de cada instante a determinada referência (exemplo: distância de cada ponto ao instante do pico máximo).

Igualmente, recomenda-se a aplicação deste processamento diretamente ao sinal EMG com diferentes tipos de tratamento. A desvantagem será uma diminuição na forma de representar o sinal, a vantagem estará associada a menor custo computacional.



## Referências

- Aizerman, M. A., Braverman, E.M., & Rozonoer, L. I. (1964). Theoretical foundations of the potential function method in pattern recognition learning. *Automation and Remote Control*, 25, 821–837.
- Alkan, A., & Günay, M. (2012). Identification of EMG signals using discriminant analysis and SVM classifier. *Expert Systems with Applications*, 39(1), 44–47
- Allison, G. T. (2003). Trunk muscle onset detection technique for EMG signals with ECG artefact. *Journal of Electromyography and Kinesiology*, 13(3), 209–216.
- Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3), 337–404.
- Bechler, J.R., Jobe, F. W., Pink, M., Perry, J., & Ruwe, P. A. (1995). Electromyographic analysis of the hip and knee during the golf swing. *Clinical Journal of Sport Medicine*, 5(3), 162–166.
- Begg, R. K., Palaniswami, M., Member, S., & Owen, B. (2005). Support Vector Machines for Automated Gait Classification. *IEEE Transactions on Biomedical Engineering*, 52(5), 828–838.
- Bitzer, S., & Smagt, P. van der. (2006). Learning EMG control of a robotic hand: Towards Active Prostheses. In *Proceedings of the 2006 IEEE International Conference on Robotics and Automation* (pp. 2819–2823). Orlando, Florida.
- Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A Training Algorithm for Optimal Margin Classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory*. (pp. 144–152).
- Burges, C. (1998). A tutorial on support vector machines for pattern recognition. *Data mining and knowledge discovery*, 2, 121–167.
- Castellini, C., & Smagt, P. van der. (2009). Surface EMG in advanced hand prosthetics. *Biological Cybernetics*, 100(1), 35–47.
- Castellini, C., Smagt, P. van der, Sandini, G., & Hirzinger, G. (2008). Surface EMG for force control of mechanical hands. In *IEEE International Conference on Robotics and Automation*, Pasadena, CA, USA, May 19-23, 2008 (pp. 725–730). Pasadena.
- Chang, C., & Lin, C. (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2, 1–39. Retrieved in 2 May 2013 from <http://dl.acm.org/citation.cfm?id=1961199>.
- Chen, P.-H., Lin, C.-J., & Scholkopf, B. (2005). A tutorial on nu-support vector machines. *Applied Stochastic Models in Business and Industry*, 21(2), 111–136.
- Chen, X., & Wang, Z. J. (2013). Pattern recognition of number gestures based on a wireless surface EMG system. *Biomedical Signal Processing and Control*, 8(2), 184–192.
- Chen, X., Zhu, X., & Zhang, D. (2010). A discriminant bispectrum feature for surface electromyogram signal classification. *Medical Engineering & Physics*, 32(2), 126–35.

- Christodoulou, C. I., Kaplanis, P. A., Murray, V., Pattichis, M. S., Pattichis, C. S., & Kyriakides, T. (2012). Multi-scale AM–FM analysis for the classification of surface electromyographic signals. *Biomedical Signal Processing and Control*, 7(3), 265–269.
- Clancy, E., & Hogan, N. (1997). Theoretic and experimental comparison of root-mean-square and mean-absolute-value electromyogram amplitude detectors. In *19th International Conference - IEEE/EMBS Oct.30 - Nov.2* (Vol. 1267, pp. 1267–1270). Chicago.
- Cole, M., & Grimshaw, P. (2008). Trunk muscle onset and cessation in golfers with and without low back pain. *Journal of Biomechanics*, 41, 2829–2833.
- Conradsen, I., Beniczky, S., Wolf, P., Kjaer, T. W., Sams, T., & Sorensen, H. B. D. (2012). Automatic multi-modal intelligent seizure acquisition (MISA) system for detection of motor seizures from electromyographic data and motion data. *Computer Methods and Programs in Biomedicine*, 107(2), 97–110.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20(3), 273–297.
- Cover, T. M. (1965). Geometrical and statistical properties of systems of linear inequalities with applications in pattern recognition. *IEEE Transactions on Electronic Computers*, 326–334.
- Crawford, B., Miller, K., Shenoy, P., & Rao, R. (2005). Real-Time classification of electromyographic signals for robotic control. In *Proceeding of AAI* (pp. 523–528).
- Cristianini, N., & Shawe-Taylor. (2000). *An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge: Cambridge University Press.
- De Luca, C. J. (1997). The use of surface electromyography in biomechanics. *Journal of Applied Biomechanics*, 13, 135–163.
- Dobrowolski, A. P., Wierzbowski, M., & Tomczykiewicz, K. (2012). Multiresolution MUAPs decomposition and SVM-based analysis in the classification of neuromuscular disorders. *Computer Methods and Programs in Biomedicine*, 107(3), 393–403.
- Duda, R.O., Hart, P.E., & Stork, D.G. (2001). *Pattern Classification*. Second Edition. New York: Wiley, 2001.
- EGA. (2012). *Sistema de Handicap EGA 2012-2016*. Epalinges/Lausanne: Edição Federação Portuguesa de Golfe.
- Englehart, K., & Hudgins, B. (2003). A robust, real-time control scheme for multifunction myoelectric control. *IEEE Transaction on Biomedical Engineering*, 50(7), 848–854.
- Farina, D., do Nascimento, O. F., Lucas, M.-F., & Doncarli, C. (2007). Optimization of wavelets for classification of movement-related cortical potentials generated by variation of force-related parameters. *Journal of Neuroscience Methods*, 162(1-2), 357–363.
- Farina, D., Merletti, R., & Enoka, R. M. (2004). The extraction of neural strategies from the surface EMG. *Journal of Applied Physiology* (Bethesda, Md. : 1985), 96(4), 1486–1495.
- Fisher, R. A. (1936). The use of multiple measures in taxonomic problems. *Annals of Eugenics*, 7(2), 179–188.

- Fox, E., Lindsay, D., & Vandervoort, A. (2002). Musculoskeletal injury questionnaire for senior golfers. *Science and Golf IV. Proceedings of the World Scientific Congress of Golf* (pp. 88–99). St. Andrews, Scotland: Routledge Inc.
- Futamata, M., Nagata, K., & Magatani, K. (2012). The evaluation of the discriminant ability of multiclass SVM in a study of hand motion recognition by using SEMG. In *Conference proceedings : Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (Vol. 2012, pp. 5246–9).
- Goker, I., Osman, O., Ozekes, S., Baslo, M. B., Ertas, M., & Ulgen, Y. (2012). Classification of juvenile myoclonic epilepsy data acquired through scanning electromyography with machine learning algorithms. *Journal of Medical Systems*, 36(5), 2705–2711.
- Güler, N. F., & Koçer, S. (2005). Use of support vector machines and neural network in diagnosis of neuromuscular disorders. *Journal of Medical Systems*, 29(3), 271–284.
- Hall, M. A. (1999). *Correlation-based Feature Selection for Machine Learning*. University of Waikato.
- Haykin, S. (2001). *Redes Neurais. Princípios e Prática*. (2ª ed.). Santana: ARTMED.
- Herbrich, R. (2002). *Learning Kernel Classifiers. Theory and Algorithms*. Massachusetts: Massachusetts Institute of Technology.
- Hermens, H.J., Freriks, B., Merletti, R., Stegeman, D., Blok, J., Rau, G., Disselhorst-Klug, C., & Hâag, G., (1996). *Seniam: European Recommendations for Surface Electromyography* (1996-1999 vol. 1–9). The Netherlands: Roessingh Research and Development.
- Hodges, P., & Bui, B. (1996). A comparison of computer-based methods for the determination of onset of muscle contraction using electromyography. *Electroencephalography and Clinical Neurophysiology*, 101, 511–519.
- Hofmann, T., Schölkopf, B., & Smola, A. J. (2008). Kernel methods in machine learning. *The Annals of Statistics*, 36(3), 1171–1220.
- Horton, J. F., Lindsay, D. M., & Macintosh, B. R. (2001). Abdominal muscle activation of elite male golfers with chronic low back pain. *Medicine and Science in Sports and Exercise*, 33(10), 1647–54.
- Hsu, C., Chang, C., & Lin, C. (2010). A practical guide to support vector classification, 1(1), 1–16. Retrieved in 2 May 2013 from <https://www.cs.sfu.ca/people/Faculty/teaching/726/spring11/svmguide.pdf>
- Hsu, C.-W. & Lin, C.-J. (2002). A comparison of methods for multiclass support vector machines. *IEEE Transactions on Neural Networks*, 13(2), 415–425.
- Huang, H., Zhang, F., Hargrove, L. J., Dou, Z., Rogers, D. R., & Englehart, K. B. (2011). Continuous locomotion-mode identification for prosthetic legs based on neuromuscular–mechanical fusion. *IEEE Transaction on Biomedical Engineering*, 58(10), 2867–2875.
- Hudgins, B., Parker, P., & Scott, R. N. (1993). A new strategy for multifunction myoelectric control. *IEEE Transactions on Biomedical Engineering*, 40(1), 82–94.

- Jöllenbeck, T. (2000). Methodological limitations of EMG-based bio-mechanical motion analysis. *ISBS – Conference Proceedings Archive, 18 International Symposium on Biomechanics in Sports*.
- Katsis, C. D., Goletsis, Y., Likas, A., Fotiadis, D. I., & Sarmas, I. (2006). A novel method for automated EMG decomposition and MUAP classification. *Artificial Intelligence in Medicine, 37*(1), 55–64.
- Khokhar, Z. O., Xiao, Z. G., & Menon, C. (2010). Surface EMG pattern recognition for real-time control of a wrist exoskeleton. *Biomedical Engineering Online, 9*:41, 1–17.
- Khushaba, R. N., Kodagoda, S., Liu, D., & Dissanayake, G. (2013). Muscle computer interfaces for driver distraction reduction. *Computer Methods and Programs in Biomedicine, 110*(2), 137–149.
- Khushaba, R. N., Kodagoda, S., Takruri, M., & Dissanayake, G. (2012). Expert Systems with Applications Toward improved control of prosthetic fingers using surface electromyogram ( EMG ) signals. *Expert Systems with Applications, 39*(12), 10731–10738.
- Konrad, P. (2005). *The ABC of EMG: A Practical Introduction to Kinesiological Electromyography* (pp. 1–60). Noraxon INC. USA.
- Kuhn, H. W., & Tucker, A. W. (1951). Nonlinear programming. In *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability* (pp. 481–492).
- Kulkarni, S. R., & Harman, G. (2011). Statistical learning theory: a tutorial. *Wiley Interdisciplinary Reviews: Computational Statistics, 3*(6), 543–556.
- Kumar, D. K., Poosapadi Arjunan, S., & Singh, V. P. (2013). Towards identification of finger flexions using single channel surface electromyography--able bodied and amputee subjects. *Journal of Neuroengineering and Rehabilitation, 10*(50), 1–7.
- Lee, A. S., Cholewicki, J., & Reeves, N. P. (2007). The effect of background muscle activity on computerized detection of sEMG onset and offset. *Journal of Biomechanics, 40*, 3521–3526.
- Li, N., Yang, D., Jiang, L., Liu, H., & Cai, H. (2012). Combined use of FSR sensor array and SVM classifier for finger motion recognition based on pressure distribution map. *Journal of Bionic Engineering, 9*(1), 39–47.
- Lin, H., & Lin, C. (2003). A study on sigmoid kernels for SVM and the training of non-PSD kernels by SMO-type methods. Submitted to *Neural Computation*, 1–32. Retrieved in 2 May 2013 from <http://home.caltech.edu/~htlin/publication/doc/tanh.pdf>
- Lorena, A. C., & Carvalho, A. C. P. L. F. (2003). *Introdução às Máquinas de Vetores Suporte. Relatórios Técnicos do ICMC*.
- Lorena, A. C., & Carvalho, A. C. P. L. F. De. (2007). Uma introdução às support vector machines. *RITA, XIV*(2), 43–67.
- Lucas, M.-F., Gaufriau, A., Pascual, S., Doncarli, C., & Farina, D. (2008). Multi-channel surface EMG classification using support vector machines and signal-based wavelet optimization. *Biomedical Signal Processing and Control, 3*(2), 169–174.

- Luxburg, U. Von, & Schölkopf, B. (2009). Statistical Learning Theory: Models, concepts, and results. In D. M. Gabbay, S. Hartmann, & J. Woods (Eds.), *Handbook of the History of Logic*. (Vol. 10, pp. 651–706). Elsevier BV.
- McGill, S., Chaimberg, J., Frost, D., & Fenwick, C. (2010). Evidence of a double peak in muscle activation to enhance strike speed and force: an example with elite mixed martial arts fighters. *Journal of Strength and Conditioning Research*, 24(2), 348–357.
- Micera, S., Sabatini, A. M., & Dario, P. (1998). An algorithm for detecting the onset of muscle contraction by EMG signal processing. *Medical Engineering & Physics*, 20(3), 211–215.
- Morey-Klapsing, G., Arampatzis, A., & Brüggemann, G. P. (2004). Choosing EMG parameters: comparison of different onset determination algorithms and EMG integrals in a joint stability study. *Clinical Biomechanics* (Bristol, Avon), 19(2), 196–201.
- Müller, K.-R., Mika, S., Rätsch, G., Tsuda, K., & Schölkopf, B. (2001). An Introduction to Kernel-Based Learning Algorithms. *IEEE Transactions on Neural Networks*, 12(2), 181–201.
- Natarajan, G. S., Wininger, M., Kim, N. H., & Craelius, W. (2012). Relating biceps EMG to elbow kinematics during self-paced arm flexions. *Medical Engineering & Physics*, 34(5), 617–624.
- Oskoei, A. M., & Hu, H. (2007). Myoelectric control systems—A survey. *Biomedical Signal Processing and Control*, 2(4), 275–294.
- Oskoei, M. A., & Hu, H. (2008). Support vector machine-based classification scheme for myoelectric control applied to upper limb. *IEEE Transaction on Biomedical Engineering*, 55(8), 1956–1965.
- Pezarat-Correia, P. & Mil-Homens, P. (2004). *A Electromiografia no Estudo do Movimento Humano*. Cruz Quebrada: Edições FMH.
- Phinyomark, A., Limsakul, C., & Phukpattaranont, P. (2009). A novel feature extraction for robust EMG pattern recognition. *Journal of Computing*, 1(1), 2151–9617.
- Phinyomark, A., Limsakul, C., & Phukpattaranont, P. (2011). Application of wavelet analysis in EMG feature extraction for pattern classification. *Measurement Science Review*, 11(2), 45–52.
- Phinyomark, A., Phukpattaranont, P., & Limsakul, C. (2012). Feature reduction and selection for EMG signal classification. *Expert Systems with Applications*, 39(8), 7420–7431.
- Phinyomark, A., Quaine, F., Charbonnier, S., Serviere, C., Tarpin-Bernard, F., & Laurillau, Y. (2013). EMG feature evaluation for improving myoelectric pattern recognition robustness. *Expert Systems with Applications*, 40(12), 4832–4840.
- Pink, M., Perry, J., & Jobe, F. (1993). EMG analysis of the trunk in golfers. *American Journal of Sports Medicine*, 21(3):385–388.
- Raez, M. B. I., Hussain, M. S., & Mohd-Yasin, F. (2006). Techniques of EMG signal analysis: detection, processing, classification and applications. *Biological Procedures Online*, 8(1), 11–35.
- Rojas, R. (1996). *Neural Networks: A Systematic Introduction*. Berlin: Springer-Verlag.

- Scheme, E., & Englehart, K. (2011). Electromyogram pattern recognition for control of powered upper-limb prostheses: State of the art and challenges for clinical use. *The Journal of Rehabilitation Research and Development*, 48(6), 643–659.
- Schölkopf, B., & Smola, A. (2002). *Learning with Kernels*. Massachusetts: Massachusetts Institute of Technology.
- Schölkopf, B., Burges, C., & Vapnik, V. (1995). Extracting support data for a given task. In U. M. Fayyad & R. Uthurusamy (Eds.), *First International Conference on Knowledge Discovery & Data Mining*. (pp. 252–257). Menlo Park: AAAI Press.
- Schölkopf, B., Smola, A. J., Williamson, R. C., & Bartlett, P. L. (2000). New support vector algorithms. *Neural Computation*, 12, 1207–1245.
- Schölkopf, B., Smola, A., & Müller, K. R. (1996). Nonlinear Component Analysis as a Kernel Eigenvalue Problem (pp. 1–18). Tübingen.
- Silva, L., Marta, S., Vaz, J., Fernandes, O., Castro, M. A., Correia-Pezarat, P. (2013). Trunk muscle activation during golf swing: Baseline and threshold. *Journal of Electromyography and Kinesiology*, 23(5), 1174–1182.
- Solnik, S., Rider, P., & Steinweg, K. (2010). Teager–Kaiser energy operator signal conditioning improves EMG onset detection. *European Journal of Applied Physiology*, 2010(110), 489–498.
- Stauder, G., Flachenecker, C., Daumer, M., & Wolf, W. (2001). Onset detection in surface electromyographic signals: A systematic comparison of methods. *Journal on Applied Signal Processing*, 2001(2), 67–81.
- Stirling, L. M., von Tscharnner, V., Kugler, P. F., & Nigg, B. M. (2011). Classification of muscle activity based on effort level during constant pace running. *Journal of Electromyography and Kinesiology*, 21(4), 566–571.
- Subasi, A. (2012). Medical decision support system for diagnosis of neuromuscular disorders using DWT and fuzzy support vector machines. *Computers in Biology and Medicine*, 42(8), 806–815.
- Subasi, A. (2013). Classification of EMG signals using PSO optimized SVM for diagnosis of neuromuscular disorders. *Computers in Biology and Medicine*, 43(5), 576–586.
- Sultornasane, S., Zeid, I., & Kamarthi, S. (2011). Classification of Electromyogram Using Recurrence Quantification Analysis. *Procedia Computer Science*, 6, 375–380.
- Tavakolan, M., Xiao, Z. G., & Menon, C. (2011). A preliminary investigation assessing the viability of classifying hand postures in seniors. *Biomedical Engineering Online*, 10(1), 79.
- Tkach, D., Huang, H., & Kuiken, T. (2010). Study of stability of time-domain features for electromyographic pattern recognition. *Journal of Neuroengineering and Rehabilitation*, 7:21, 1–13.
- Tolambiyana, A., Thomas, E., Chiovetto, E., Berret, B., & Pozzo, T. (2011). An ensemble analysis of electromyographic activity during whole body pointing with the use of support vector machines. *PloS one*, 6(7), e20732.
- Torrence, C., & Compo, G. P. (1998). A Practical guide to wavelet analysis. *Bulletin of the American Meteorological Society*, 79(1), 61–78

- Tyler, A., & Karst, G. (2004). Timing of muscle activity during reaching while standing: systematic changes with target distance. *Gait & posture*, 20, 126–133.
- Vaisman, L., Zariffa, J., & Popovic, M. R. (2010). Application of singular spectrum-based change-point analysis to EMG-onset detection. *Journal of Electromyography and Kinesiology*, 20(4), 750–760.
- Van Boxtel, G. J. M., Geraars, L. H. D., Van Den Berg-Lenssen, M. M. C., & Brunia, C. H. M. (1993). Detection of EMG onset in ERP research. *Psychophysiology*, 1993, 405–412.
- Vannozzi, G., Conforto, S., & Alessio, T. D. (2010). Automatic detection of surface EMG activation timing using a wavelet transform based method. *Journal of Electromyography and Kinesiology*, 20(4), 767–772.
- Vapnik, V., & Chervonenkis, A. (1964). A note on one class of perceptrons. *Automation and Remote Control*, 25.
- Vapnik, V. N. (1999). An Overview of Statistical Learning Theory. *IEEE Transactions on Neural Networks*, 10(5), 988–999.
- Watkins, R. G., Uppal, G. S., Perry, J., Pink, M., & Dinsay, J. M. (1996). Dynamic electromyographic analysis of trunk musculature in professional golfers. *The American Journal of Sports Medicine*, 24(4), 535–538.
- Yan, Z., Wang, Z., & Xie, H. (2008). The application of mutual information-based feature selection and fuzzy LS-SVM-based classifier in motion classification. *Computer Methods and Programs in Biomedicine*, 90, 275–284.
- Yang, D., Zhao, J.-d., Gu, Y.-k., Wang, X.-q., Li, N., Jiang, L., Liu, H., Huang, H., & Zhao, D.-w.. (2009). An anthropomorphic robot hand developed based on underactuated mechanism and controlled by EMG signals. *Journal of Bionic Engineering*, 6(3), 255–263.
- Zecca, M., & Micera, S. (2002). Control of multifunctional prosthetic hands by processing the electromyographic signal. *Critical Reviews™ in Biomedical Engineering*, 30(4-6), 459–485.
- Zhao, Z., Morstatter, F., Sharma, S., Alelyani, S., Anand, A., & Liu, H. (2010). Advancing feature selection research-ASU feature selection repository, Arizona State University, Retrieved in 2 May 2013 from <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Advancing+Feature+Selection+Research+-+ASU+Feature+Selection+Repository#0>



# Apêndices

## Consentimento Informado

---

### Consentimento Informado

---

**Tema do Estudo:** Análise cinemática, cinética e EMG do *swing* no Golf

#### Objetivos do Estudo:

Analisar durante o *swing* efetuado com diferentes ferros (4, 7 e *pitch*) através do comportamento cinemático de todo o corpo, o comportamento cinético e o comportamento electromiográfico do tronco e membros inferiores.

Estudar o movimento que ocorre entre os vários segmentos corporais, durante um *swing* bem como as forças exercidas e também atividade muscular do membro inferior. Para tal será necessário colocar “sensores” e marcas em determinadas localizações anatómicas e realizar alguns testes, antes da recolha propriamente dita.

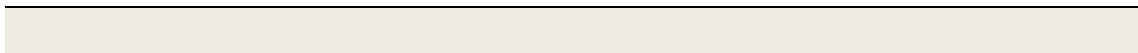
#### Procedimentos na Recolha:

- Informação ao participante sobre os procedimentos, esclarecimento de dúvidas, preenchimento questionário caracterização.
- Preparação da pele (depilação e limpeza) e colocação dos elétrodos nos músculos dos membros inferiores e tronco: Tibial anterior, Longo peroneal, Gémeos (*músculos das pernas*), Reto Femoral, Vasto Externo, Vasto Interno, Bicípite Femoral, Semitendinoso (*músculos das coxas*), Grande nadegueiro (*músculo da nádega*), Ereter da espinha (*músculo das costas*), Obliquo externo (*músculo da barriga*). (*análise da atividade muscular*).
- Colocação de marcadores reflexivos: Fita com 4 marcadores na cabeça, Ombros, Cotovelo lados, Punho, 7<sup>a</sup> vértebra cervical, 12<sup>a</sup> vértebra torácica, Pélvis posterior, Pélvis anterior, Anca, Joelho (lateral e medial), Parte anterior da tíbia, Tornozelos (lateral e medial), Calcânhar, Base dedo grande do pé, Base dedo mínimo do pé, Parte superior e inferior do taco. (*análise do movimento*).
- Colocação sobre o green e a plataforma de forças (*análise das forças*).

Os elétrodos e marcas reflexivas serão fixos à pele por fita adesiva apropriada.

A tarefa a realizar consistirá na realização do *swing* da forma mais natural possível tentando obter uma eficácia máxima no total de 30 tacadas efetuadas alternadamente com o ferro 4, 7 e *pitch*. O *swing* deverá ser efetuado.

As recolhas terão lugar no laboratório de Comportamento motor da Faculdade de motricidade humana.



Estou disposto(a) a colaborar como voluntário(a) no estudo Análise do Swing com diferentes ferros.

Foi-me explicado pelos investigadores a finalidade deste estudo, bem como os seus objetivos, princípios e procedimentos, sendo que os compreendi na totalidade e os aceito.

Sei também que os meus dados vão ser tratados confidencialmente e permito que sejam usados para a investigação, mantendo-se sempre o meu anonimato e privacidade. De igual forma, sei que sou livre de recusar a participação e posso desistir a qualquer momento, não resultando daqui qualquer consequência.

Assinatura do participante:

---

Data: \_\_\_\_|\_\_\_\_|\_\_\_\_

## Informação Para os Participantes

### Informação aos sujeitos

#### Objectivo do estudo

Com o intuito de efetuar a prevenção de lesões no Golfe pretendemos fazer uma análise do *swing*, estudando para isso o comportamento cinemático, cinético de todo o corpo e electro miográfico dos membros inferiores. Ou seja, pretendemos estudar o movimento que ocorre entre os vários segmentos corporais, durante um *swing* e também catividade muscular do membro inferior. Para tal será necessário colocar “sensores” e marcas em determinadas localizações anatómicas e realizar alguns testes, antes da recolha propriamente dita que consistirá em 24 *swings* com diferentes ferros (4, 7 e *pitch*)

#### Sequência dos Procedimentos

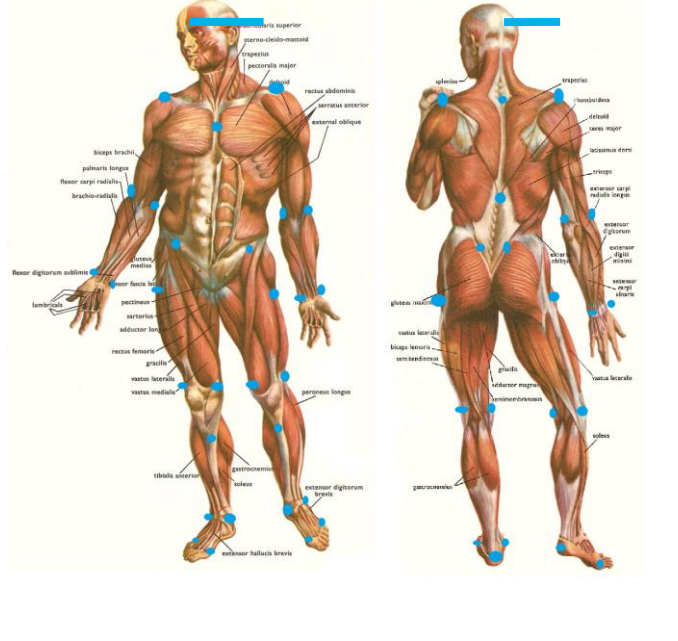
1º Depois do preenchimento de um questionário de caracterização do atleta, começamos por colocar os “sensores” de atividade muscular, que é precedida por depilação e limpeza da pele para que o sinal seja genuíno. Seguidamente precisamos saber qual a força máxima desses músculos para que seja possível efetuar comparações entre a amostra. Para isso vamos fazer-lhe os testes musculares abaixo descritos em que será aplicada resistência para contrariar o seu movimento com máxima força:

Músculo	Testes de força
	Sentado
Tibial anterior	Sentado na marquesa fletir o tornozelo trazendo o pé para si ao mesmo tempo que vira as plantas dos pés uma para a outra. Este teste é feito nos 2 pés ao mesmo tempo. Suportar resistência aplicada!
Peroneal longo	Deitado de lado (perna de baixo semifletida) colocar o pé de cima em bico do pé e depois levá-lo na direção do teto. Suportar resistência aplicada!
Vasto medial	Sentado na marquesa com a perna de fora fazer extensão do joelho Suportar resistência aplicada!
Reto femoral	
Vasto lateral	
	Deitado de barriga para cima
Oblíquo Externo	Deitado de barriga para cima, mãos atrás da nuca, tentar chegar com o ombro de um lado à anca do lado contrário. Suportar resistência aplicada!

Em pé	
Gêmeo medial	Em pé sobre um pé apoiado na parede, levantar o calcanhar do chão, mantendo o joelho esticado enquanto com os membros superiores resiste a esse movimento. Mais do que pressionar a ponta do pé deverá tentar levantar o calcanhar. Suportar resistência aplicada!
Gêmeo lateral	
Deitado de barriga para baixo	
Grande nadegueiro	Deitado de barriga para baixo rodar a anca para fora e levantá-la da marquesa na direção do teto. Suportar resistência aplicada!
Eretor da espinha	Deitado de barriga para baixo mãos debaixo da testa, levantar o tronco da marquesa. Suportar resistência aplicada!
Bicípíte femoral	Deitado de barriga para baixo fletir o joelho em teste. Suportar resistência aplicada!
Semitendinoso	

2º Para conhecermos a pressão efetuada em cada um dos pés colocaremos umas palmilhas dentro dos seus sapatos.

3º Seguidamente colocamos os “sensores” para a análise do movimento dos braços, pernas, cabeça e tronco, através da colocação de refletos esféricos nas referências anatómicas abaixo descritas:

<ol style="list-style-type: none"> <li>1. Fita com 4 marcadores na cabeça</li> <li>2. Ombros</li> <li>3. Cotovelo lados</li> <li>4. Punho</li> <li>5. 7ª vértebra cervical</li> <li>6. 12ª vértebra torácica</li> <li>7. Pélvis posterior</li> <li>8. Pélvis anterior</li> <li>9. Anca</li> <li>10. Joelho (lateral e medial)</li> <li>11. Parte anterior da tíbia</li> <li>12. Tornozelos (lateral e medial)</li> <li>13. Calcanhar</li> <li>14. Base dedo grande do pé</li> <li>15. Base dedo mínimo do pé</li> <li>16. Parte superior e inferior do taco</li> </ol>	
--	--

4º Iremos para o *green*, onde ligaremos as palmilhas ao aparelho e o deixamos efectuar o seu aquecimento habitual.

5º Por último, para percebermos a amplitude habitual do seu movimento, vamos gravar a sua execução dos seguintes movimentos:

De pé à posição de cócoras flectindo tronco e braços e depois para de pé fazendo o máximo de extensão do tronco e braços
Em Pé, com os joelhos esticados chegar com mãos ou chão (ou até onde conseguir)
Em Pé, inclinação lateral direita e esquerda
Em Pé, com os braços afastados e os cotovelos estendidos fazer rotação máxima dta e esq (rotação global dos tornozelos à coluna)

6º Finalmente procederemos às recolhas do movimento de *swing* no total de 24 tacadas efetuadas alternadamente com o ferro 4, 7 e *pitch*. O *swing* deverá ser efetuado da forma mais natural possível tentando obter uma eficácia máxima.

Pedimos-lhe que traga uns **calções confortáveis**, os seus sapatos de treino e os seus **tacos 4, 7 e *pitch***.

Prevemos que a totalidade dos procedimentos possa demorar cerca de 2h30m.

Estamos muito gratos pela sua colaboração e disponibilidade que desde já agradecemos!

## **Caracterização dos Participantes**

(Fox et al., 2002)

**Universidade de Lisboa**  
**Faculdade de Motricidade Humana**  
**Universidade Aberta**

### **Questionário sobre condições músculo-esqueléticas em jogadores de golfe adultos**

Em primeiro lugar, gostaríamos de lhe fazer algumas perguntas sobre si:

1. Qual é a sua data de nascimento? \_\_\_\_\_  
(dia)                      (mês)                      (ano)
2. Qual é a sua altura? \_\_\_\_\_ cm
3. Qual é o seu peso? \_\_\_\_\_ kg
4. Sexo masculino ou feminino? (deve escolher a opção adequada assinalando-a com um círculo no número correspondente à sua resposta)
  1. Masculino
  2. Feminino
5. Há quantos anos joga golfe? \_\_\_\_\_ anos/ meses (**riscar o que não interessa**)
6. É destro ou esquerdino? (Círculo)
  1. Destro
  2. Esquerdino

#### **Gostaríamos agora de o questionar acerca do seu histórico de doenças e lesões:**

7. O seu médico alguma vez lhe disse que tinha um problema cardíaco e que só deveria fazer actividade física recomendada por um médico? (Círculo)
  1. Sim
  2. Não
8. Sente dor no peito quando faz exercício físico? (círculo)
  1. Sim
  2. Não
9. No último mês, alguma vez teve dor no peito quando não estava a fazer actividade física? (Círculo)

1. Sim
2. Não

10. Alguma vez sentiu tonturas chegando a perder o equilíbrio ou alguma vez desmaiou?

1. Sim
2. Não

11. Tem algum problema ósseo ou articular que possa ser agravado pela prática de actividade física? (Círculo)

1. Sim
2. Não

12. Toma medicamentos (por exemplo, diuréticos) para a tensão arterial ou para problemas cardíacos? (Círculo)

1. Sim
2. Não

13. Conhece algum outro motivo que o possa impedir de fazer exercício físico? (círculo)

1. Sim
2. Não

Em caso afirmativo, especifique, por favor: \_\_\_\_\_

**De seguida, gostaríamos de fazer algumas perguntas sobre o seu jogo/prática de golfe:**

14. Em média, quantos metros bate com o seu driver? \_\_\_\_\_

15. Em média, quantos metros bate com o seu ferro 7? \_\_\_\_\_

16. Qual é o seu handicap? \_\_\_\_\_

(Refira o seu handicap aproximado, caso não tenha um oficial)

**Gostaríamos agora de o questionar sobre o seu swing:**

17. O seu swing é destro ou esquerdino? (Círculo)

1. Esquerdino
2. Destro

18. Os seus tacos de golfe estão adaptados (fitting) ao seu swing? (Círculo)

1. Sim
2. Não

## Outputs

### Normalidade F2 por grupos

		Tests of Normality					
Grupos		Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
C9x9_F2	High	.223	12	.102	.787	12	.007
	Low	.263	12	.022	.754	12	.003
	Total	.315	12	.002	.683	12	.001
g9x9_F2	High	.323	12	.001	.694	12	.001
	Low	.304	12	.003	.719	12	.001
	Total	.314	12	.002	.762	12	.004
ACC9x9_F2	High	.169	12	.200*	.914	12	.239
	Low	.117	12	.200*	.967	12	.874
	Total	.148	12	.200*	.946	12	.580
SV9x9_F2	High	.150	12	.200*	.935	12	.441
	Low	.142	12	.200*	.946	12	.578
	Total	.125	12	.200*	.945	12	.571
ST9x9_F2	High	.150	12	.200*	.935	12	.441
	Low	.142	12	.200*	.946	12	.578
	Total	.125	12	.200*	.945	12	.571
C5x5_F2	High	.206	12	.171	.852	12	.039
	Low	.214	12	.133	.840	12	.027
	Total	.336	12	.001	.582	12	.000
g5x5_F2	High	.364	12	.000	.693	12	.001
	Low	.262	12	.022	.766	12	.004
	Total	.309	12	.002	.763	12	.004
ACC5x5_F2	High	.168	12	.200*	.919	12	.276
	Low	.090	12	.200*	.980	12	.985
	Total	.144	12	.200*	.949	12	.619
SV5x5_F2	High	.119	12	.200*	.976	12	.963
	Low	.152	12	.200*	.939	12	.488
	Total	.113	12	.200*	.939	12	.483
ST5x5_F2	High	.152	12	.200*	.973	12	.938
	Low	.152	12	.200*	.939	12	.488
	Total	.113	12	.200*	.939	12	.483

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

## Normalidade F4 por grupos

Tests of Normality							
Grupos	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
C9x9_F4	High	.456	12	.000	.594	12	.000
	Low	.373	12	.000	.675	12	.000
	Total	.499	12	.000	.465	12	.000
g9x9_F4	High	.428	12	.000	.508	12	.000
	Low	.355	12	.000	.764	12	.004
	Total	.355	12	.000	.764	12	.004
ACC9x9_F4	High	.195	12	.200*	.947	12	.592
	Low	.135	12	.200*	.935	12	.431
	Total	.186	12	.200*	.900	12	.158
SV9x9_F4	High	.143	12	.200*	.937	12	.463
	Low	.147	12	.200*	.943	12	.537
	Total	.135	12	.200*	.986	12	.998
ST9x9_F4	High	.143	12	.200*	.937	12	.463
	Low	.147	12	.200*	.943	12	.537
	Total	.135	12	.200*	.986	12	.998
C5x5_F4	High	.252	12	.033	.861	12	.051
	Low	.245	12	.045	.861	12	.050
	Total	.288	12	.007	.810	12	.012
g5x5_F4	High	.353	12	.000	.681	12	.001
	Low	.210	12	.152	.889	12	.114
	Total	.260	12	.025	.832	12	.022
ACC5x5_F4	High	.187	12	.200*	.950	12	.630
	Low	.161	12	.200*	.934	12	.429
	Total	.205	12	.176	.892	12	.126
SV5x5_F4	High	.203	12	.186	.881	12	.089
	Low	.155	12	.200*	.958	12	.760
	Total	.182	12	.200*	.909	12	.205
ST5x5_F4	High	.203	12	.186	.881	12	.090
	Low	.155	12	.200*	.958	12	.760
	Total	.182	12	.200*	.909	12	.205

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

## Normalidade F6 por grupos

Tests of Normality							
Grupos	Kolmogorov-Smirnov <sup>a</sup>			Shapiro-Wilk			
	Statistic	df	Sig.	Statistic	df	Sig.	
C9x9_F6	High	.280	12	.010	.784	12	.006
	Low	.417	12	.000	.638	12	.000
	Total	.530	12	.000	.327	12	.000
g9x9_F6	High	.455	12	.000	.626	12	.000
	Low	.385	12	.000	.732	12	.002
	Total	.420	12	.000	.690	12	.001
ACC9x9_F6	High	.184	12	.200*	.963	12	.822
	Low	.170	12	.200*	.933	12	.409
	Total	.192	12	.200*	.882	12	.092
SV9x9_F6	High	.171	12	.200*	.965	12	.848
	Low	.155	12	.200*	.954	12	.689
	Total	.181	12	.200*	.930	12	.384
ST9x9_F6	High	.182	12	.200*	.939	12	.483
	Low	.155	12	.200*	.954	12	.689
	Total	.181	12	.200*	.930	12	.384
C5x5_F6	High	.174	12	.200*	.895	12	.135
	Low	.252	12	.033	.835	12	.024
	Total	.267	12	.018	.756	12	.003
g5x5_F6	High	.321	12	.001	.542	12	.000
	Low	.251	12	.035	.875	12	.076
	Total	.219	12	.116	.810	12	.012
ACC5x5_F6	High	.191	12	.200*	.947	12	.590
	Low	.166	12	.200*	.928	12	.362
	Total	.200	12	.200*	.886	12	.104
SV5x5_F6	High	.172	12	.200*	.888	12	.112
	Low	.174	12	.200*	.914	12	.243
	Total	.152	12	.200*	.913	12	.235
ST5x5_F6	High	.172	12	.200*	.888	12	.112
	Low	.174	12	.200*	.914	12	.243
	Total	.152	12	.200*	.913	12	.235

a. Lilliefors Significance Correction

\*. This is a lower bound of the true significance.

## MANOVA não Paramétrica

Multivariate Tests <sup>c</sup>						
Effect		Value	F	Hypothesis df	Error df	Sig.
Intercept	Pillai's Trace	.957	104.082 <sup>a</sup>	6.000	28.000	.000
	Wilks' Lambda	.043	104.082 <sup>a</sup>	6.000	28.000	.000
	Hotelling's Trace	22.303	104.082 <sup>a</sup>	6.000	28.000	.000
	Roy's Largest Root	22.303	104.082 <sup>a</sup>	6.000	28.000	.000
Grupos	Pillai's Trace	.464	1.462	12.000	58.000	.166
	Wilks' Lambda	.582	1.452 <sup>a</sup>	12.000	56.000	.171
	Hotelling's Trace	.640	1.439	12.000	54.000	.177
	Roy's Largest Root	.471	2.276 <sup>b</sup>	6.000	29.000	.064

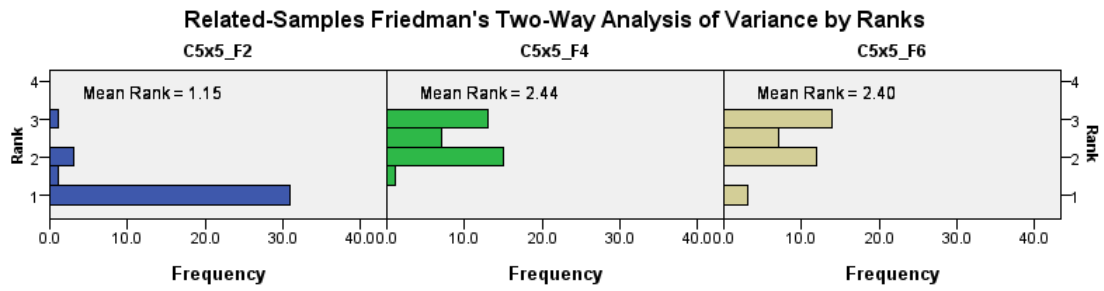
a. Exact statistic b. The statistic is an upper bound on F that yields a lower bound on the significance level. c. Design: Intercept + Grupos

Tests of Between-Subjects Effects						
Source	Dependent Variable	Type III Sum of Squares	df	Mean Square	F	Sig.
Corrected Model	Rank of C5x5_F2	874.625 <sup>a</sup>	2	437.312	4.800	.015
	Rank of g5x5_F2	107.042 <sup>b</sup>	2	53.521	.470	.629
	Rank of C5x5_F4	43.292 <sup>c</sup>	2	21.646	.191	.827
	Rank of g5x5_F4	50.167 <sup>d</sup>	2	25.083	.221	.803
	Rank of C5x5_F6	594.125 <sup>e</sup>	2	297.063	3.092	.059
	Rank of g5x5_F6	21.125 <sup>f</sup>	2	10.563	.092	.912
Intercept	Rank of C5x5_F2	12321.000	1	12321.000	135.244	.000
	Rank of g5x5_F2	12321.000	1	12321.000	108.123	.000
	Rank of C5x5_F4	12321.000	1	12321.000	108.622	.000
	Rank of g5x5_F4	12321.000	1	12321.000	108.560	.000
	Rank of C5x5_F6	12321.000	1	12321.000	128.248	.000
	Rank of g5x5_F6	12321.000	1	12321.000	107.114	.000
Grupos	Rank of C5x5_F2	874.625	2	437.313	4.800	.015
	Rank of g5x5_F2	107.042	2	53.521	.470	.629
	Rank of C5x5_F4	43.292	2	21.646	.191	.827
	Rank of g5x5_F4	50.167	2	25.083	.221	.803
	Rank of C5x5_F6	594.125	2	297.062	3.092	.059
	Rank of g5x5_F6	21.125	2	10.562	.092	.912
Error	Rank of C5x5_F2	3006.375	33	91.102		
	Rank of g5x5_F2	3760.458	33	113.953		
	Rank of C5x5_F4	3743.208	33	113.431		
	Rank of g5x5_F4	3745.333	33	113.495		
	Rank of C5x5_F6	3170.375	33	96.072		
	Rank of g5x5_F6	3795.875	33	115.027		
Total	Rank of C5x5_F2	16202.000	36			
	Rank of g5x5_F2	16188.500	36			
	Rank of C5x5_F4	16107.500	36			
	Rank of g5x5_F4	16116.500	36			
	Rank of C5x5_F6	16085.500	36			
	Rank of g5x5_F6	16138.000	36			
Corrected Total	Rank of C5x5_F2	3881.000	35			
	Rank of g5x5_F2	3867.500	35			
	Rank of C5x5_F4	3786.500	35			
	Rank of g5x5_F4	3795.500	35			
	Rank of C5x5_F6	3764.500	35			
	Rank of g5x5_F6	3817.000	35			

a. R Squared = .225 (Adjusted R Squared = .178) b. R Squared = .028 (Adjusted R Squared = -.031) c. R Squared = .011 (Adjusted R Squared = -.048)

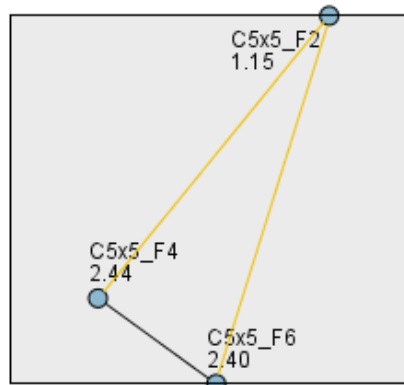
d. R Squared = .013 (Adjusted R Squared = -.047) e. R Squared = .158 (Adjusted R Squared = .107) f. R Squared = .006 (Adjusted R Squared = -.055)

## Teste de Friedman para o parâmetro *C* nos conjuntos de *features* alto handicap



<b>Total N</b>	36
<b>Test Statistic</b>	42.318
<b>Degrees of Freedom</b>	2
<b>Asymptotic Sig. (2-sided test)</b>	.000

### Pairwise Comparisons

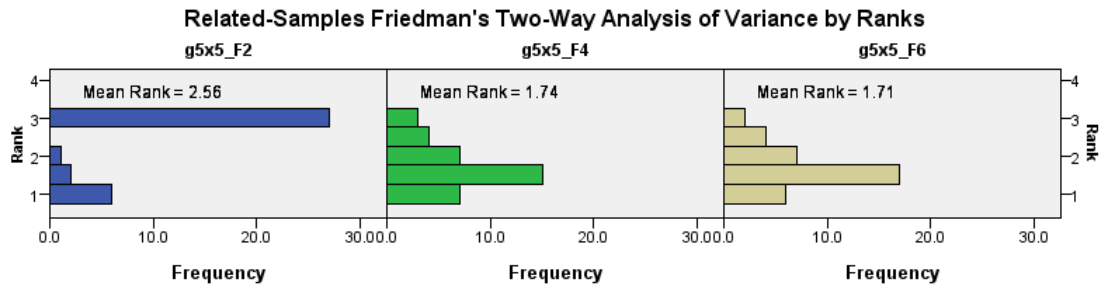


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C5x5_F2-C5x5_F6	-1.250	.236	-5.303	.000	.000
C5x5_F2-C5x5_F4	-1.292	.236	-5.480	.000	.000
C5x5_F6-C5x5_F4	.042	.236	.177	.860	1.000

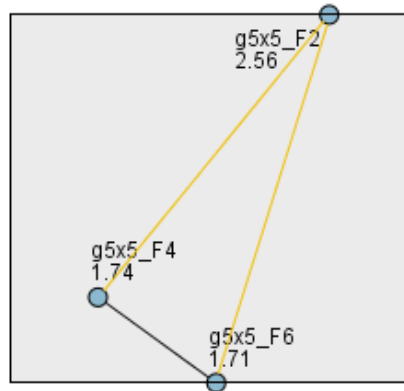
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de *features* alto handicap



Total N	36
Test Statistic	19,528
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

### Pairwise Comparisons

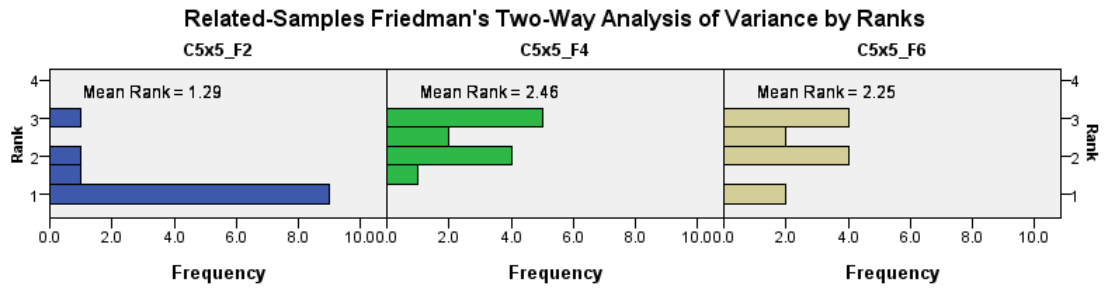


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
g5x5_F6-g5x5_F4	.028	.236	.118	.906	1.000
g5x5_F6-g5x5_F2	.847	.236	3.594	.000	.001
g5x5_F4-g5x5_F2	.819	.236	3.477	.001	.002

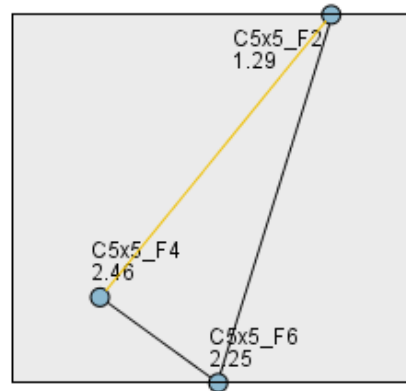
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## Teste de Friedman para o parâmetro *C* nos conjuntos de *features* baixo handicap



Total N	12
Test Statistic	9,911
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.007

### Pairwise Comparisons

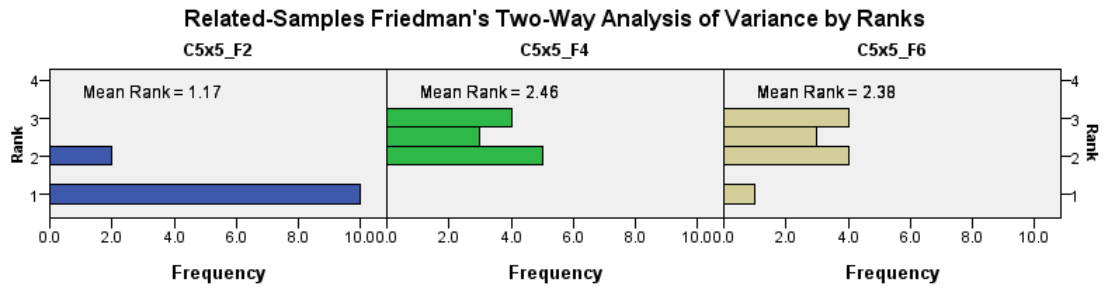


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C5x5_F2-C5x5_F6	-.958	.408	-2.347	.019	.057
C5x5_F2-C5x5_F4	-1.167	.408	-2.858	.004	.013
C5x5_F6-C5x5_F4	.208	.408	.510	.610	1.000

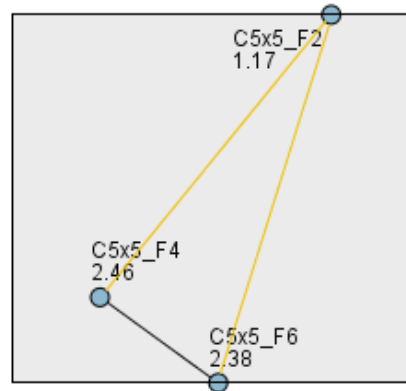
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de *features* baixo handicap



Total N	12
Test Statistic	14,683
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.001

### Pairwise Comparisons

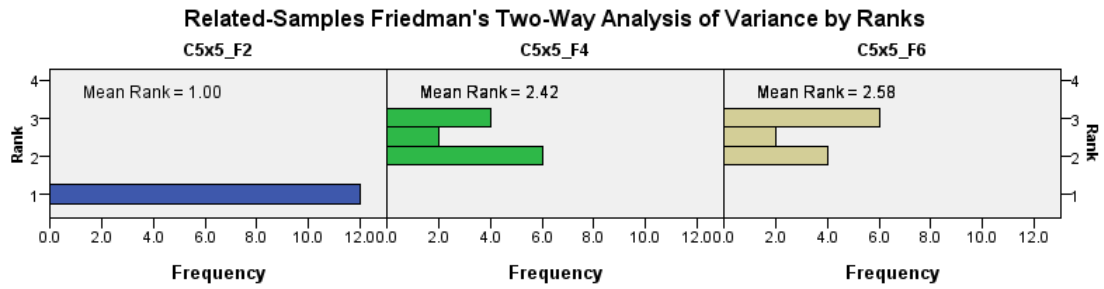


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C5x5_F2-C5x5_F6	-1.208	.408	-2.960	.003	.009
C5x5_F2-C5x5_F4	-1.292	.408	-3.164	.002	.005
C5x5_F6-C5x5_F4	.083	.408	.204	.838	1.000

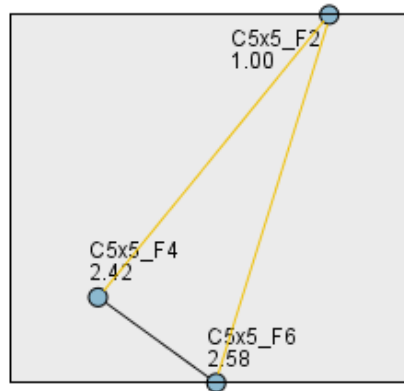
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## Teste de Friedman para o parâmetro *C* nos conjuntos de *features* ambos handicap



Total N	12
Test Statistic	18.957
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.000

### Pairwise Comparisons

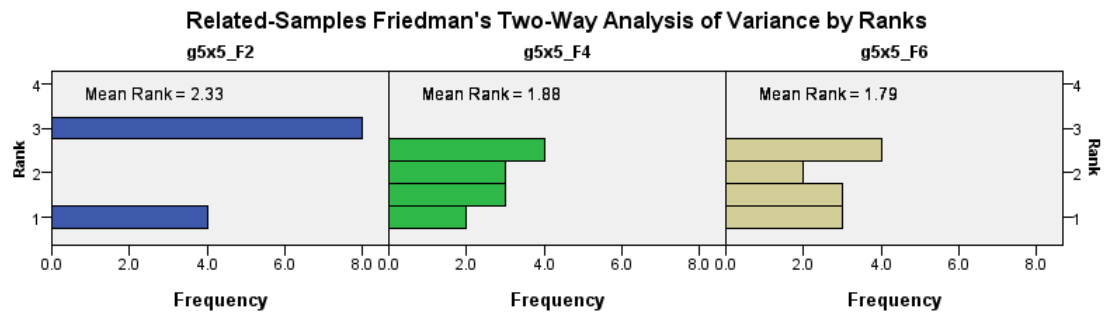


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
C5x5_F2-C5x5_F4	-1.417	.408	-3.470	.001	.002
C5x5_F2-C5x5_F6	-1.583	.408	-3.878	.000	.000
C5x5_F4-C5x5_F6	-.167	.408	-.408	.683	1.000

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

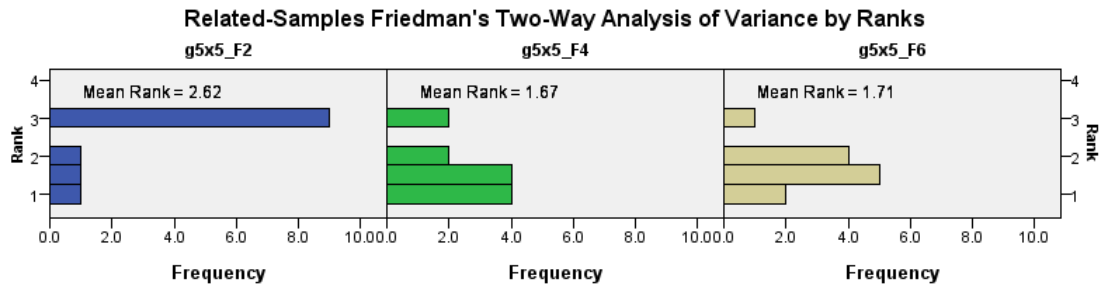
## Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de *features* ambos handicap



Total N	12
Test Statistic	2.390
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.303

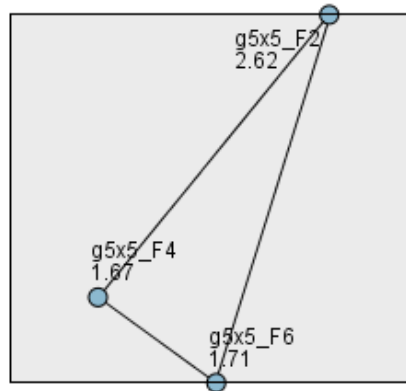
1. Multiple comparisons are not performed because the overall test retained the null hypothesis of no differences.

## Teste de Friedman para o parâmetro *C* nos conjuntos de *features* todos modelos



Total N	12
Test Statistic	7.860
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.020

### Pairwise Comparisons

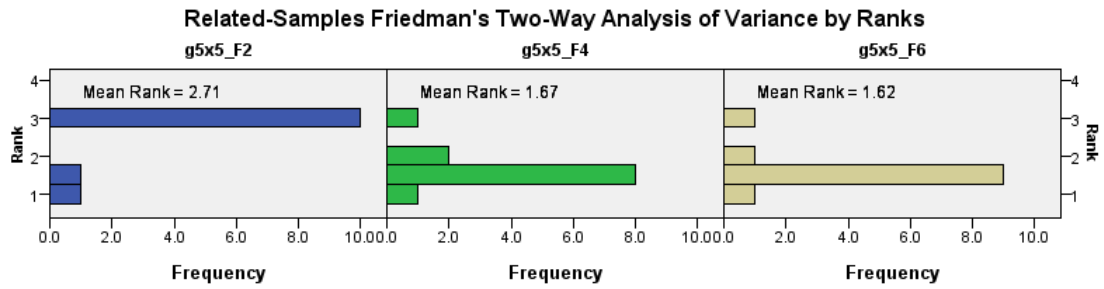


Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
g5x5_F4-g5x5_F6	-.042	.408	-.102	.919	1.000
g5x5_F4-g5x5_F2	.958	.408	2.347	.019	.057
g5x5_F6-g5x5_F2	.917	.408	2.245	.025	.074

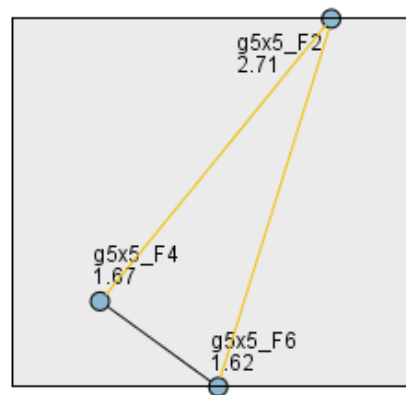
Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## Teste de Friedman para o parâmetro $\gamma$ nos conjuntos de *features* todos modelos



Total N	12
Test Statistic	11.128
Degrees of Freedom	2
Asymptotic Sig. (2-sided test)	.004

### Pairwise Comparisons



Each node shows the sample average rank.

Sample1-Sample2	Test Statistic	Std. Error	Std. Test Statistic	Sig.	Adj.Sig.
g5x5_F6-g5x5_F4	.042	.408	.102	.919	1.000
g5x5_F6-g5x5_F2	1.083	.408	2.654	.008	.024
g5x5_F4-g5x5_F2	1.042	.408	2.552	.011	.032

Each row tests the null hypothesis that the Sample 1 and Sample 2 distributions are the same. Asymptotic significances (2-sided tests) are displayed. The significance level is .05.

## ANOVA mista precisão SVM e comparações múltiplas

Descriptive Statistics				
	Grupos	Mean	Std. Deviation	N
ACC5x5_F2	High	90.2800	3.96007	12
	Low	90.7783	4.92698	12
	Total	89.4117	3.64891	12
	Total	90.1567	4.13166	36
ACC5x5_F4	High	94.9418	2.51643	12
	Low	95.0317	3.32435	12
	Total	93.4592	3.24987	12
	Total	94.4775	3.05247	36
ACC5x5_F6	High	95.1475	2.42186	12
	Low	95.1058	3.21831	12
	Total	93.5825	3.25399	12
	Total	94.6119	2.99529	36

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Features	.058	91.168	2	.000	.515	.549	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + Grupos

Within Subjects Design: Features

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Features	Sphericity Assumed	462.449	2	231.224	213.328	.000	.866	426.656	1.000
	Greenhouse-Geisser	462.449	1.030	449.061	213.328	.000	.866	219.688	1.000
	Huynh-Feldt	462.449	1.097	421.534	213.328	.000	.866	234.034	1.000
	Lower-bound	462.449	1.000	462.449	213.328	.000	.866	213.328	1.000
Features * Grupos	Sphericity Assumed	1.886	4	.471	.435	.783	.026	1.740	.146
	Greenhouse-Geisser	1.886	2.060	.916	.435	.657	.026	.896	.116
	Huynh-Feldt	1.886	2.194	.860	.435	.669	.026	.954	.118
	Lower-bound	1.886	2.000	.943	.435	.651	.026	.870	.115
Error(Features)	Sphericity Assumed	71.537	66	1.084					
	Greenhouse-Geisser	71.537	33.984	2.105					
	Huynh-Feldt	71.537	36.203	1.976					
	Lower-bound	71.537	33.000	2.168					

a. Computed using alpha = .05

Tests of Within-Subjects Contrasts

Measure:MEASURE\_1

Source	Features	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Features	Linear	357.291	1	357.291	215.263	.000	.867	215.263	1.000
	Quadratic	105.158	1	105.158	207.005	.000	.863	207.005	1.000
Features *	Linear	1.603	2	.801	.483	.621	.028	.966	.122
	Quadratic	.283	2	.142	.279	.759	.017	.557	.090
Error(Features)	Linear	54.773	33	1.660					
	Quadratic	16.764	33	.508					

a. Computed using alpha = .05

Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
ACC5x5_F2	.603	2	33	.553
ACC5x5_F4	.615	2	33	.547
ACC5x5_F6	.895	2	33	.418

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Grupos

Within Subjects Design: Features

Tests of Between-Subjects Effects

Measure:MEASURE\_1

Transformed Variable:Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Intercept	935740.873	1	935740.873	27650.500	.000	.999	27650.500	1.000
Grupos	47.396	2	23.698	.700	.504	.041	1.401	.158
Error	1116.777	33	33.842					

a. Computed using alpha = .05

Estimates

Measure:MEASURE\_1

Features	Mean	Std. Error	95% Confidence Interval	
			Lower Bound	Upper Bound
1	90.157	.702	88.728	91.586
2	94.478	.509	93.443	95.512
3	94.612	.498	93.598	95.626

Pairwise Comparisons

Measure:MEASURE\_1

(I) Features	(J) Features	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	-4.321*	.294	.000	-5.063	-3.579
	3	-4.455*	.304	.000	-5.221	-3.689
2	1	4.321*	.294	.000	3.579	5.063
	3	-.134*	.043	.011	-.243	-.026
3	1	4.455*	.304	.000	3.689	5.221
	2	.134*	.043	.011	.026	.243

Based on estimated marginal means

\*. The mean difference is significant at the .05 level.

a. Adjustment for multiple comparisons: Bonferroni.

2. Grupos \* Features

Measure:MEASURE\_1

Grupos	Features	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
High	1	90.280	1.216	87.805	92.755
	2	94.942	.881	93.149	96.734
	3	95.148	.863	93.392	96.903
Low	1	90.778	1.216	88.303	93.253
	2	95.032	.881	93.239	96.824
	3	95.106	.863	93.350	96.862
Total	1	89.412	1.216	86.937	91.887
	2	93.459	.881	91.667	95.252
	3	93.583	.863	91.827	95.338

MEASURE\_1

	Grupos	N	Subset	
			1	
Tukey HSD <sup>a,b,c</sup>	Total	12		92.1511
	High	12		93.4564
	Low	12		93.6386
	Sig.			.530
Scheffe <sup>a,b,c</sup>	Total	12		92.1511
	High	12		93.4564
	Low	12		93.6386
	Sig.			.561

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 11.281.

a. Uses Harmonic Mean Sample Size = 12.000.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = .05.

## ANOVA mista vetores suporte e comparações múltiplas

Descriptive Statistics				
	Grupos	Mean	Std. Deviation	N
SV5x5_F2	High	35.7856	11.00273	12
	Low	37.2646	17.63672	12
	Total	38.6639	13.41575	12
	Total	37.2380	13.92098	36
SV5x5_F4	High	28.1703	6.75105	12
	Low	26.7573	7.60280	12
	Total	29.2021	5.88837	12
	Total	28.0432	6.66489	36
SV5x5_F6	High	28.8497	6.48881	12
	Low	26.1425	9.40398	12
	Total	29.4033	6.18289	12
	Total	28.1318	7.42490	36

### Mauchly's Test of Sphericity<sup>b</sup>

Measure: MEASURE\_1

Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon <sup>a</sup>		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
Features	.129	65.620	2	.000	.534	.571	.500

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

b. Design: Intercept + Grupos

Within Subjects Design: Features

### Tests of Within-Subjects Effects

Measure: MEASURE\_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Features	Sphericity Assumed	2009.696	2	1004.848	29.523	.000	.472	59.045	1.000
	Greenhouse-Geisser	2009.696	1.069	1880.419	29.523	.000	.472	31.552	1.000
	Huynh-Feldt	2009.696	1.142	1759.346	29.523	.000	.472	33.724	1.000
	Lower-bound	2009.696	1.000	2009.696	29.523	.000	.472	29.523	1.000
Features *	Sphericity Assumed	55.772	4	13.943	.410	.801	.024	1.639	.139
	Greenhouse-Geisser	55.772	2.137	26.092	.410	.680	.024	.876	.113
	Huynh-Feldt	55.772	2.285	24.412	.410	.694	.024	.936	.115
	Lower-bound	55.772	2.000	27.886	.410	.667	.024	.819	.111
Error(Features)	Sphericity Assumed	2246.410	66	34.037					
	Greenhouse-Geisser	2246.410	35.269	63.694					
	Huynh-Feldt	2246.410	37.696	59.593					
	Lower-bound	2246.410	33.000	68.073					

a. Computed using alpha = .05

Tests of Within-Subjects Contrasts

Measure: MEASURE\_1

Source	Features	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Features	Linear	1492.609	1	1492.609	29.903	.000	.475	29.903	1.000
	Quadratic	517.087	1	517.087	28.478	.000	.463	28.478	.999
Features *	Linear	52.787	2	26.393	.529	.594	.031	1.058	.130
Grupos	Quadratic	2.985	2	1.492	.082	.921	.005	.164	.061
Error(Features)	Linear	1647.218	33	49.916					
	Quadratic	599.193	33	18.157					

a. Computed using alpha = .05

Levene's Test of Equality of Error Variances<sup>a</sup>

	F	df1	df2	Sig.
SV5x5_F2	1.643	2	33	.209
SV5x5_F4	.653	2	33	.527
SV5x5_F6	.950	2	33	.397

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept + Grupos Within Subjects Design: Features

Tests of Between-Subjects Effects

Measure: MEASURE\_1

Transformed Variable: Average

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared	Noncent. Parameter	Observed Power <sup>a</sup>
Intercept	104712.059	1	104712.059	439.538	.000	.930	439.538	1.000
Grupos	103.174	2	51.587	.217	.806	.013	.433	.081
Error	7861.666	33	238.232					

a. Computed using alpha = .05

Pairwise Comparisons

Measure: MEASURE\_1

(I) Features	(J) Features	Mean Difference (I-J)	Std. Error	Sig. <sup>a</sup>	95% Confidence Interval for Difference <sup>a</sup>	
					Lower Bound	Upper Bound
1	2	9.195*	1.665	.000	4.994	13.395
	3	9.106*	1.665	.000	4.906	13.306
2	1	-9.195*	1.665	.000	-13.395	-4.994
	3	-.089	.355	1.000	-.983	.806
3	1	-9.106*	1.665	.000	-13.306	-4.906
	2	.089	.355	1.000	-.806	.983

Based on estimated marginal means

\*. The mean difference is significant at the .05 level. a. Adjustment for multiple comparisons: Bonferroni.

2. Grupos \* Features

Measure: MEASURE\_1

Grupos	Features	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
High	1	35.786	4.123	27.396	44.175
	2	28.170	1.958	24.186	32.154
	3	28.850	2.165	24.445	33.255
Low	1	37.265	4.123	28.875	45.654
	2	26.757	1.958	22.773	30.741
	3	26.142	2.165	21.737	30.548
Total	1	38.664	4.123	30.275	47.053
	2	29.202	1.958	25.218	33.186
	3	29.403	2.165	24.998	33.808

## MEASURE\_1

	Grupos	N	Subset	
			1	
Tukey HSD <sup>a,b,c</sup>	Low	12	30.0548	
	High	12	30.9352	
	Total	12	32.4231	
	Sig.		.793	
Scheffe <sup>a,b,c</sup>	Low	12	30.0548	
	High	12	30.9352	
	Total	12	32.4231	
	Sig.		.810	

Means for groups in homogeneous subsets are displayed.

Based on observed means.

The error term is Mean Square(Error) = 79.411.

a. Uses Harmonic Mean Sample Size = 12.000.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

c. Alpha = .05.

## Correlações de Pearson classificação vs vetores suporte

		Correlations					
		ACC5x5_F2	SV5x5_F2	ACC5x5_F4	SV5x5_F4	ACC5x5_F6	SV5x5_F6
ACC5x5_F2	Pearson Correlation	1	-.883**	.927**	-.760**	.922**	-.702**
	Sig. (2-tailed)		.000	.000	.000	.000	.000
	N	36	36	36	36	36	36
SV5x5_F2	Pearson Correlation	-.883**	1	-.887**	.768**	-.882**	.734**
	Sig. (2-tailed)	.000		.000	.000	.000	.000
	N	36	36	36	36	36	36
ACC5x5_F4	Pearson Correlation	.927**	-.887**	1	-.841**	.997**	-.780**
	Sig. (2-tailed)	.000	.000		.000	.000	.000
	N	36	36	36	36	36	36
SV5x5_F4	Pearson Correlation	-.760**	.768**	-.841**	1	-.818**	.960**
	Sig. (2-tailed)	.000	.000	.000		.000	.000
	N	36	36	36	36	36	36
ACC5x5_F6	Pearson Correlation	.922**	-.882**	.997**	-.818**	1	-.761**
	Sig. (2-tailed)	.000	.000	.000	.000		.000
	N	36	36	36	36	36	36
SV5x5_F6	Pearson Correlation	-.702**	.734**	-.780**	.960**	-.761**	1
	Sig. (2-tailed)	.000	.000	.000	.000	.000	
	N	36	36	36	36	36	36

\*\* . Correlation is significant at the 0.01 level (2-tailed).

