



Sistema de Apoio a Terapeutas no Processo de Reabilitação de Pacientes de AVC

FERNANDO JOSÉ RIBEIRO AFONSO

outubro de 2017

Sistema de Apoio a Terapeutas no Processo de Reabilitação de Pacientes de AVC

Fernando José Ribeiro Afonso

**Dissertação para obtenção do Grau de Mestre em
Engenharia Informática, Área de Especialização em
Tecnologias de Conhecimento e Decisão**

Orientador: Luiz Felipe Rocha de Faria

Co-orientador: Paulo Jorge Machado Oliveira

Dedicatória

«A todos que nunca deixaram de acreditar em mim.»

Resumo

Uma das vertentes de investigação do Centro de Estudo do Movimento e Atividade Humana (CEMAH) está direcionada para a terapia destinada a doentes que sofreram AVC (Acidente Vascular Cerebral) e que em sequência do acidente apresentam limitação de movimentos. No processo de recuperação dos pacientes são aplicados protocolos de terapia que têm por objetivo induzir a recuperação da marcha autónoma dos pacientes e conduzir a uma marcha com a melhor qualidade possível. A escolha dos protocolos de terapia depende da análise de um conjunto de variáveis preditoras cujos valores são obtidos em laboratório, a partir de leituras de sensores aplicados nos membros dos pacientes durante a ocorrência de diferentes eventos, como por exemplo, levantar, sentar e início de marcha. A partir desta análise, é possível escolher o protocolo de intervenção mais adequado para a recuperação da marcha. Análises recentes apontam para a necessidade de não só intervir no membro afetado pelo AVC, como também no membro simétrico não afetado. Atualmente, o trabalho de análise descrito é realizado de forma manual pelos terapeutas. Porém, com o trabalho desenvolvido estudou-se e implementou-se um protótipo que permita efetuar a análise automática dos dados recolhidos pelos sensores, possibilitando assim realizar a recomendação dos protocolos de intervenção mais adequados. Foi ainda criado um protótipo que permite fazer o acompanhamento da aplicação do protocolo de intervenção escolhido pelo terapeuta, de forma a validar a decisão tomada ou sugerir alterações ao protocolo a aplicar em situações futuras.

Palavras-chave: AVC, recuperação, prescrição, recomendação, mineração de dados, clustering

Abstract

One of the research aspects of the Centro de Estudos do Movimento e Atividade Humana [Center for the Study of Movement and Human Activity] (CEMAH) is directed to the therapy destined to patients who suffered a stroke, and that in the sequence of the accident, present limitation of movements. In the recovery process of patients, therapy protocols are applied which aim to induce recovery of patients' autonomous gait and lead to a gait with the best possible quality. The choice of the therapy protocols depends on the analysis of a set of predictor variables whose values are obtained in the laboratory, from sensor readings applied in the patients' members during the occurrence of different events, such as standing, sitting and the beginning of the march. From this analysis, it is possible to choose the most appropriate intervention protocol and to infer the probability of the patient recovering the gait. Recent studies point to the need not only to intervene in the limb affected by the stroke but also in the unaffected symmetric limb. Currently, the analysis work described is performed manually by therapists. The aim of this work is to study and implement a prototype that allows automatic analysis of the data collected by the sensors so that the most appropriate intervention protocols can be recommended. It is also intended that the prototype allows monitoring of the application of the intervention protocol chosen by the therapist to validate the decision taken or suggest changes to the protocol to be applied in future situations.

Keywords: stroke, recovery, gait, prescription, recommendation, data mining, clustering

Agradecimentos

Um agradecimento especial à minha namorada, Nádia Dias, por todo apoio e motivação. Que nos momentos mais difíceis permaneceu a meu lado, que nunca me deixou desistir e esteve sempre pronta para me ajudar. Foi, e continua a ser, uma forte presença que me guia no meu percurso académico, profissional e pessoal. Um especial obrigado.

Agradeço aos meus pais e irmã, que possibilitaram os meus estudos. A todos os esforços e sacrifícios que fizeram em prol da minha educação, um sincero e profundo obrigado.

Agradeço aos Professores Luiz Faria e Paulo Oliveira que estiveram sempre disponíveis e que me ajudaram em todas as etapas. Agradeço também aos Professores Rubim Santos e Andreia Sousa que foram incansáveis até ao último momento em tudo que precisei.

A todos os colegas e amigos que me motivaram e ajudaram no meu percurso, desde a licenciatura à presente dissertação, obrigado pelo vosso companheirismo e amizade.

Ao instituto que me acolheu e ao departamento que me ensinou, a todos os docentes e envolvidos, obrigado.

Índice

1	Introdução	1
1.1	Contexto e problema	1
1.2	Objetivos	2
1.3	Resultados esperados	4
1.4	Processos e intervenientes	4
1.5	Restrições existentes	5
1.6	Abordagem preconizada	5
2	Análise de valor	7
2.1	Modelo de desenvolvimento	8
2.2	Valor	9
2.3	Proposta de valor	10
2.4	Modelo de negócio	10
2.5	Rede de valor	13
3	Estado da arte	15
3.1	Conceitos	18
3.2	Identificação de variáveis relevantes	19
3.3	Clustering	23
3.4	Sistema de recomendação	25
3.5	Sistemas semelhantes	28
4	Avaliação de soluções	31
4.1	Identificação de músculos significativos e análise da dependência funcional do membro afetado e do membro simétrico	31
4.2	Perfis de pacientes	32
4.2.1	Análise multicritério pode ser aplicado na escolha de algoritmos de <i>clustering</i>	32
4.2.2	Algoritmo de <i>clustering K-means</i>	37
4.3	Prescrição de tratamentos	37
5	Sistema proposto	39
5.1	Design da solução para o problema	39
5.2	Arquitetura	39
5.3	Opções de desenvolvimento	44
6	Avaliação	47
6.1	Hipóteses	47
6.2	Metodologias e Grandezas de Avaliação	48

6.3	Grandezas de avaliação	49
7	Desenvolvimento	51
7.1	Dados	51
7.2	Análise de dados.....	51
7.2.1	Identificação de variáveis relevantes	52
7.2.2	Prova de levantar-sentar - forças de reação do solo (FRS)	59
7.2.3	Clustering.....	61
7.3	Sistema de apoio à decisão.....	78
7.3.1	Plataforma Web	78
7.3.2	Sistema de recomendação.....	91
8	Conclusão	99
8.1	Objetivos realizados	99
8.2	Limitações e trabalho futuro.....	100
8.3	Criação de artigo	100
8.4	Apreciação final	100
	Bibliografia	103
	Anexos.....	109
	Paper - Feature-Selection Analysis of Relevant Muscles in Stroke Patients with Impaired Walking Ability.....	109
	Método de encriptação e descriptação MD5	113

Lista de Figuras

Figura 1 – Modelo de <i>canvas</i>	12
Figura 2 – Rede de valor	13
Figura 3 – Passos da mineração de dados.....	16
Figura 4 – Identificação de grupos nos pontos resultantes da aplicação de clustering.....	23
Figura 5 – Arquitetura da <i>framework</i>	28
Figura 6 – Diagrama de componentes (UML)	40
Figura 7 – Diagrama de etapas.....	41
Figura 8 – Diagrama de casos de uso (UML)	43
Figura 9 – Características da máquina utilizada para a análise de dados	52
Figura 10 – Exemplo de ficheiro resultante da prova de marcha monitorizada	53
Figura 11 – Prova de marcha monitorizada, disponibilizada pelo CEMAH	53
Figura 12 – Matriz de confusão do processo de identificação de variáveis relevantes na prova EMG.....	55
Figura 13 – Precisão de cada variável no processo de identificação de variáveis relevantes na prova EMG.....	57
Figura 14 – Código extraído do RStudio no processo de <i>clustering</i>	62
Figura 15 – Desenho do valor de <i>withinss</i> por número de <i>clusters</i> , relativos aos dados da prova e do paciente.....	63
Figura 16 – Clustering dos dados do paciente e da prova EMG, com 4 clusters	64
Figura 17 – Clustering dos dados do paciente e da prova EMG, com 7 clusters	64
Figura 18 – Clustering dos dados do paciente e da prova EMG, com 10 clusters	65
Figura 19 – Árvore de decisão dos clusters encontrados para os dados do paciente e da prova EMG.....	65
Figura 20 – Importância das variáveis dos clusters encontrados para os dados do paciente e da prova EMG.....	66
Figura 21 – Erro relativo por tamanho da árvore dos clusters encontrados para os dados do paciente e da prova EMG	67
Figura 22 - Desenho do valor de <i>withinss</i> por número de <i>clusters</i> , relativos aos dados do paciente e o resumo das provas	67
Figura 23 – Clustering dos dados do paciente e do resumo da prova EMG, com 3 clusters	68
Figura 24 – Clustering dos dados do paciente e do resumo da prova EMG, com 5 clusters	69
Figura 25 – Clustering dos dados do paciente e do resumo da prova EMG, com 7 clusters	69
Figura 26 – Clustering dos dados do paciente e do resumo da prova EMG, com 8 clusters	70
Figura 27 – Árvore de decisão do <i>clustering</i> dos dados do paciente e do resumo da prova EMG	71
Figura 28 – Desenho do valor de <i>withinss</i> por número de <i>clusters</i> , relativos aos dados da FRS e do paciente.....	72
Figura 29 – Clustering dos dados do paciente e da prova FRS, com 2 clusters	72
Figura 30 – Clustering dos dados do paciente e da prova FRS, com 5 clusters	73
Figura 31 – Clustering dos dados do paciente e da prova FRS, com 7 clusters	73
Figura 32 – Clustering dos dados do paciente e da prova FRS, com 11 clusters	73
Figura 33 – Clustering dos dados do paciente e da prova FRS, com 14 clusters	74

Figura 34 – Esquema exemplificativo dos parâmetros CoPx e CoPy	74
Figura 35 – Exemplo do resultado exploratório Calinski & Harabasz aplicado às provas FRS	76
Figura 36 - Diferentes variações do centro de pressão no sentido médio-lateral (CoPx)	77
Figura 37 - Diferentes variações do centro de pressão no sentido ântero-posterior (CoPy)	77
Figura 38 – Análise do período inicial de ativação do centro de pressão no sentido médio-lateral (CoPx)	78
Figura 39 – Versão simplificada da arquitetura de três camadas	79
Figura 40 – Modelo de dados	80
Figura 41 – Classe de registo do log de atividades	81
Figura 42 – Exemplo da inserção de um log de criação	81
Figura 43 – Exemplo da inserção de um log de edição	82
Figura 44 – Exemplo de registo de log de erro	82
Figura 45 – Página de autenticação	82
Figura 46 – Exemplo de sucesso na autenticação	83
Figura 47 – Exemplo de falha na autenticação	83
Figura 48 – Validação de permissões	84
Figura 49 – Criação de um paciente	84
Figura 50 – Ficha de paciente e seus registos	85
Figura 51 – Consulta de uma ficha de paciente	85
Figura 52 – Teste unitário à criação de paciente	86
Figura 53 – Teste unitário à leitura de paciente	86
Figura 54 – Teste unitário à edição de paciente	87
Figura 55 – Teste unitário à eliminação de paciente	87
Figura 56 – Teste unitário à validação de permissões (Sucesso)	87
Figura 57 – Teste unitário à validação de permissões (Insucesso)	88
Figura 58 – Teste unitário à validação de informação	88
Figura 59 - Classificação de recomendações e nível de evidência	93
Figura 60 – Exemplo de regras definidas, no formato drl	94
Figura 61 – Carregamento das regras para o motor de inferência	94
Figura 62 – Pedido de recomendação a um paciente	94
Figura 63 – Introdução de dados para a prescrição de tratamentos	96
Figura 64 – Exemplo de recomendação e justificação gerada	97

Lista de Tabelas

Tabela 1 – Tabela de classificação.....	33
Tabela 2 – Matriz de comparação de critérios.....	33
Tabela 3 – Matriz de comparação de critérios.....	33
Tabela 4 – Normalização da matriz de comparação de critérios	34
Tabela 5 – Cálculo das prioridades relativas	34
Tabela 6 - Cálculo da prioridade relativa da qualidade.....	35
Tabela 7 - Cálculo da prioridade relativa da qualidade com vetor de prioridade.....	35
Tabela 8 - Cálculo da prioridade relativa do desempenho	35
Tabela 9 - Cálculo da prioridade relativa do desempenho com vetor de prioridade	35
Tabela 10 - Cálculo da prioridade relativa da adaptação a um número elevado de dados.....	36
Tabela 11 - Cálculo da prioridade relativa da adaptação a um número elevado de dados com vetor de prioridade	36
Tabela 12 – Avaliação dos algoritmos	36
Tabela 13 – Metodologias de avaliação	48
Tabela 14 – Análise de grandezas de avaliação	49
Tabela 15 – Relevância dos músculos para a realização da marcha	55
Tabela 16 – Relevância da ativação dos músculos para a realização da marcha.....	56
Tabela 17 – Variáveis mais representativas da marcha de ambos os membros	58
Tabela 18 – 10 variáveis mais representativas da marcha.....	59
Tabela 19 – Teste funcional – Criação de paciente.....	88
Tabela 20 – Teste Funcional – Gestão da ficha de paciente	89
Tabela 21 – Teste funcional – Inserção de registo de tratamento	89
Tabela 22 – Teste funcional – Requisição de recomendação de tratamento.....	90
Tabela 23 – Teste funcional – Consulta de ficha de paciente	90

Acrónimos e Símbolos

Lista de Acrónimos

AVC	Acidente Vascular Cerebral
CEMAH	Centro de Estudo do Movimento e Atividade Humana
ETL	Extract, transform, load
FRS	Forças de reação do solo
EMG	Sinais electromiográficos
AHP	Analytic Hierarchy Process
PL	Presentation layer
BLL	Business logic layer
DAL	Data access layer

1 Introdução

Neste capítulo é feita uma breve introdução ao problema que o trabalho descrito se propõe a resolver. O problema é apresentado e contextualizado, sendo de seguida apresentada uma síntese da abordagem tomada para a resolução do mesmo.

1.1 Contexto e problema

O Centro de Estudo do Movimento e Atividade Humana (CEMAH) “(...) iniciou a sua atividade em 2007, com a participação nas áreas de Neurofisiologia, Audiologia, Fisioterapia, Terapia Ocupacional, Cardiopneumologia, Ciências Funcionais e Física. Neste Centro procuram promover a qualificação de alto nível de profissionais de saúde, estimulando a produção e difusão de conhecimento no âmbito específico do seu objeto de conhecimento, num quadro de referência internacional assim como estimular a formação científico-tecnológica dos seus investigadores e docentes, em diferentes níveis e ramos de especialização, bem como dos estudantes, futuros profissionais.” [1]

Uma das vertentes de investigação do CEMAH está direcionada para a terapia destinada a doentes que sofreram um Acidente Vascular Cerebral (AVC) e que, em sequência do acidente, apresentam limitação de movimentos.

O acompanhamento da anomalia do paciente é um processo que apresenta alguma complexidade, dependendo da análise de um conjunto de variáveis preditoras. O valor destas é obtido em laboratório, a partir de leituras de sensores aplicados nos membros dos pacientes durante a ocorrência de diferentes eventos chave, sendo exemplos: levantar, sentar e início de marcha. Com base nestes valores é feita uma análise onde é possível criar um perfil sobre o qual se tenta extrair uma previsão sobre a condição do paciente e possíveis tratamentos.

No processo de recuperação dos pacientes são aplicados protocolos de terapia que têm por objetivo possibilitar a recuperação da marcha autónoma dos pacientes, com a melhor qualidade motora possível. A escolha do protocolo de intervenção mais adequado baseia-se nos dados recolhidos durante o acompanhamento e nos resultados de tratamentos prévios. Esta fase consiste na aplicação de uma sequência de tratamentos de forma iterativa.

O CEMAH identificou alguns problemas na execução dos processos referidos: existe uma falta de informatização, automatização e apoio do acompanhamento da recuperação e tratamento de pacientes.

Neste sentido, foi detetada uma necessidade de aferir quais os músculos intervenientes que desempenham um papel mais decisivo na recuperação de marcha.

Neste âmbito, o CEMAH realizou estudos que permitiram identificar os músculos que têm mais impacto na recuperação da marcha, recorrendo a informação relativa ao estado dos mesmos. No entanto, nenhum dos estudos realizados apresentou resultados conclusivos. Os resultados são de difícil obtenção uma vez que este processo requer a avaliação de um músculo, e esta tem que ser realizada em par com os restantes músculos, dado que o efeito da ação pode ser amplificado ou até anulado pela atividade de um ou vários músculos.

Foi de interesse estudar a hipótese de existir a dependência funcional entre o membro afetado e o seu membro oposto. Embora não existisse nenhum estudo que comprovasse esta sujeição, os peritos presentes no CEMAH inferiam esta possibilidade, a partir de dados recolhidos e dos casos analisados.

Atualmente, o processo de acompanhamento do paciente é realizado alguns meses após o AVC e de uma forma pouco adaptada ao paciente – normalmente o primeiro tratamento prescrito é sempre o mesmo e o plano de tratamentos adaptado à reação do paciente ao tratamento. É pretendido iniciar a recolha de dados na fase do internamento do paciente e que seja possível obter dados de forma mais regular, com o primeiro registo a ser realizado não mais do que três meses após o AVC.

Atualmente, o trabalho de análise da prescrição de tratamentos é realizado de forma manual e iterativa, dado não existir nenhum sistema utilizado atualmente pelos terapeutas. No fim de cada tratamento é realizada uma avaliação e adaptada a prescrição de tratamentos. Esta avaliação pode ser demorada e pouco precisa, visto que é feita de forma manual e não existe uma forma fácil de comparar a evolução do paciente e a resposta deste aos tratamentos a outros casos já tratados.

1.2 Objetivos

O intento principal do projeto é auxiliar o profissional médico e terapêutico na compreensão, identificação do AVC e prescrição de pacientes afetados pelo mesmo. Para tal, o projeto está dividido em três módulos: compreensão dos parâmetros relevantes à recuperação e marcha do paciente, identificação de perfis de pacientes e prescrição de protocolos. O sistema será utilizado pelos profissionais médicos e peritos do CEMAH.

Na fase de compreensão foi requerida a análise dos parâmetros dos testes realizados nos pacientes, mais concretamente os músculos dos membros inferiores, e ser validada a hipótese de que o membro oposto possui impacto no processo de recuperação do membro afetado.

Foi também objetivado a identificação dos perfis de pacientes. Estes perfis poderão futuramente auxiliar no processo de recuperação, utilizando os padrões para a escolha de

análises que o paciente terá que realizar, prescrição de tratamentos e fornecer dados como a probabilidade de recuperação.

Na fase de prescrição de tratamento, foi solicitado a criação de um sistema de apoio à decisão que auxilia o terapeuta na indicação dos melhores tratamentos. Este poderá, como indicado anteriormente, utilizar os perfis descobertos no módulo de identificação de padrões, embora tenha que ser realizado um estudo com objetivo de compreender se existe alguma dependência entre o perfil e os tratamentos mais eficazes.

As três fases descritas dão origem a duas implementações. A primeira fase e parte da segunda – a descoberta de conhecimento - ser implementadas dando origem a um sistema de análise de dados. Este tem como propósito a recolha, processamento e extração de conhecimento, e foi desenvolvido no âmbito deste projeto conjuntamente com os peritos do CEMAH, utilizando técnicas de mineração de dados (*data mining*).

Após recolhida a informação necessária, tornou-se possível o desenvolvimento de um sistema de informação e apoio à decisão. Este sistema engloba a parte interativa de classificação de pacientes da segunda fase e toda a terceira. Traduz-se por uma aplicação Web, capaz de fornecer dados sobre um paciente, mediante determinados padrões, e aconselhar prescrições de tratamentos ao mesmo. O sistema será utilizado por profissionais médicos e terapeutas, com foco no paciente. É pretendido fornecer informação crucial ao paciente, como a possibilidade de recuperação e auxiliar o mesmo no processo de recuperação e tratamento. Para tal, ao paciente serão realizados testes sobre o membro afetado, que irão permitir encontrar um padrão em que este se enquadra. Após o padrão ser encontrado e validado, o processo de tratamento pode ser iniciado. Dá-se assim início a um processo iterativo de tratamentos e avaliações que têm como foco a recuperação do paciente.

Em suma, existem quatro grandes objetivos que foram propostos atingir:

- Avaliar a hipótese da existência de uma dependência funcional do membro afetado e do seu membro simétrico e os membros que apresentam uma maior relevância na previsão da recuperação da marcha. Esta avaliação foi realizada com base na atividade dos músculos dos membros inferiores;
- Identificar perfis de pacientes que sofreram AVC, através da utilização de técnicas de descoberta de conhecimento. Pretendeu-se encontrar perfis que atualmente não fossem do conhecimento dos peritos e que, dessa forma, fossem úteis aos mesmos;
- Criação de um protótipo, que surge do design e da implementação de um mecanismo de recomendação inteligente de protocolos de intervenção. Este foi criado com base no conhecimento dos peritos, nos dados recolhidos a partir da leitura de sensores referentes a variáveis predictoras e nos modelos identificados através do objetivo anterior. Este protótipo permite, após a intervenção terapêutica, monitorizar a evolução das variáveis relevantes com o objetivo de avaliar os resultados da intervenção e permitir a eventual correção da mesma;

- Realização de um estudo que demonstre a validade da abordagem e a utilidade da ferramenta desenvolvida.

1.3 Resultados esperados

O protótipo visa proporcionar uma maior compreensão dos fatores que causam problemas na marcha de um paciente vítima de AVC, através da análise dos músculos de ambos os membros. A equipa do CEMAH também dispõe de evidências que indicam que a análise do membro não-afetado possa adicionar uma nova dimensão ao estudo dos pacientes. Este membro, embora aparentemente sem problemas motores, pode, de facto, apresentar lesões e/ou problemas musculares pouco evidentes que irá compensar utilizando funções motoras e atividade do seu membro simétrico. Caso se comprove esta dependência, poderá possibilitar um trabalho futuro, onde será analisada em maior detalhe a atividade e razão de dependência desse membro.

Com a mineração dos dados é pretendido categorizar o paciente. Esta categorização corresponde ao padrão da doença presente no paciente e visa auxiliar na previsão da recuperação total ou parcial do mesmo, recomendar as análises, estudos e tratamentos a realizar sobre um paciente.

A categorização do paciente irá ainda auxiliar o início da prescrição de tratamentos. O protótipo deverá ainda ser capaz de aconselhar o profissional (médico ou terapeuta) no processo iterativo de tratamentos, adaptando as recomendações às escolhas feitas.

1.4 Processos e intervenientes

Como apresentado na secção Objetivos, existem diferentes entidades que irão interagir com o sistema proposto em diferentes fases do mesmo. São de seguida descritos os intervenientes:

Peritos – Serão a principal fonte de conhecimento utilizado no sistema. Para além de recolherem e tratarem os dados utilizados na aplicação, também fornecem conhecimento empírico e teórico sobre a doença e os protocolos de intervenção. Será com o auxílio destes que os resultados serão validados e avaliados.

Profissionais médicos – Os profissionais podem ser médicos ou terapeutas. São os principais utilizadores da plataforma Web, com o objetivo de registar informação e ter acesso a informação que irá apoiar as suas decisões.

Pacientes – Utilizadores secundários da aplicação. Estes podem apenas consultar alguma da informação da sua ficha de paciente na plataforma Web. Será com base nos estudos realizados sobre os peritos ou os profissionais médicos sobre os mesmos que serão recolhidos os dados.

Os intervenientes têm interação com o sistema em fases diferentes.

Numa primeira fase, os peritos (CEMAH) colaboram no processo de recolha e tratamento de dados. Com os dados tratados é então possível ao analista de dados aplicar técnicas com o objetivo de extrair novo conhecimento dos dados. O conhecimento é então validado e avaliado pelos peritos. Caso os resultados se provem insatisfatórios, há a necessidade de voltar a aplicar técnicas de extração de conhecimento ou então até realizar uma recolha e tratamento de dados diferente. Após a validação do conhecimento é possível ainda incentivar trabalho futuro na área da investigação, que será desenvolvida pelos peritos.

Com a informação recolhida, é então criada uma plataforma Web para disponibilização da mesma. Os profissionais médicos e os pacientes têm acesso à utilização da plataforma. Os pacientes têm um acesso muito restrito, podendo apenas consultar de forma limitada a sua ficha. Os profissionais, por sua vez, podem aceder a toda a informação das fichas de paciente e ainda ao sistema de apoio à decisão. O sistema tem como propósito auxiliar na categorização do paciente e na prescrição de tratamentos.

1.5 Restrições existentes

Todas as conclusões têm base nos dados já recolhidos e disponibilizados pelo CEMAH. Uma maior quantidade de pacientes analisados irá permitir obter um rigor maior nas conclusões obtidas. A principal limitação do projeto é a quantidade de dados (de pacientes) a que há acesso, dado que o centro possui um número reduzido de amostras possíveis de analisar. Embora esteja planeada a recolha e disponibilização de mais dados no futuro, a atual falta de dados poderá levar a conclusões com nível de confiança reduzido.

1.6 Abordagem preconizada

O CEMAH disponibiliza um conjunto de amostras recolhidas de indivíduos saudáveis e de indivíduos padecentes de AVC. As amostras são feitas em ambos os membros inferiores e durante as atividades consideradas importantes para o diagnóstico da doença – como sentar e levantar – analisando um conjunto de músculos. Sobre estes dados têm que ser utilizadas técnicas de mineração de dados para analisar os músculos que possuem um papel mais determinante na marcha de um paciente que sofreu um AVC, categorizar os pacientes e ainda apoiar o profissional médico durante a recuperação do paciente.

O objetivo principal do projeto é a possibilidade de analisar um paciente perante diferentes perfis previamente criados e um sistema de apoio à decisão na prescrição de tratamentos. Conjuntamente com as funcionalidades desenvolvidas é ainda pretendido obter mais informação referente à atividade de um paciente que sofreu de AVC, concretamente sobre os músculos que desempenham um papel mais relevante.

O objetivo principal do projeto é a possibilidade de analisar um paciente perante diferentes perfis previamente criados pelas técnicas de *clustering* e um sistema de apoio à decisão na prescrição de tratamentos. Conjuntamente com as funcionalidades desenvolvidas é ainda

pretendido obter um conhecimento sobre os músculos responsáveis pela recuperação da marcha de uma vítima de AVC.

Para a prescrição de tratamentos terá que ser criado um sistema de recomendação. Este deve de ser desenvolvido com a informação disponibilizada pelos peritos do CEMAH e dos terapeutas físicos que trabalham com eles. O sistema será utilizado durante o processo de recuperação do paciente, estando todas as funcionalidades disponíveis numa plataforma Web, de forma a facilitar o acesso pelos médicos e terapeutas.

As conclusões obtidas a partir das análises realizadas devem que ser disponibilizadas num documento científico, para desenvolvimento de trabalho futuro.

2 Análise de valor

“A criação de valor é fundamental para qualquer negócio, e qualquer atividade de negócio é sobre trocar algum bem ou serviço tangível e/ou intangível e ter o seu valor aceito e recompensado por clientes, tanto na empresa como na rede colaborativa ou externamente.” [2] (traduzido da fonte original)

O AVC é uma doença de início súbito na qual o paciente pode apresentar paralisção ou dificuldade de movimentação dos membros, dificuldade na fala ou articulação das palavras e déficit visual súbito de uma parte do campo visual. Pode ainda evoluir para um coma ou até mesmo a morte do paciente [3]. O processo de reabilitação pode ser longo, dependendo das características do próprio AVC, da região afetada, da rapidez de atuação para minimizar os riscos e do apoio que o doente tiver. Esta doença pode afetar qualquer parte do sistema neuro central que inclui o cérebro, o tronco encefálico, o cerebelo e até a medula espinhal.

O AVC afeta cerca de 17 milhões de pessoas anualmente, havendo probabilidade de as vítimas não fazerem uma recuperação total das suas funcionalidades motoras. No mesmo período de tempo, estima-se que o AVC cause 6 milhões de mortes das pessoas afetadas. [4].

O CEMAH visa procurar abordagens mais eficazes para auxílio à recuperação de pacientes de AVC. Neste projeto procuram encontrar respostas a duas grandes questões sobre a recuperação: que músculos têm um papel mais ativo e se o membro simétrico ao afetado possui alguma influência na recuperação do paciente. O CEMAH pretende ainda uma aplicação que auxilie tanto médicos como terapeutas no processo de recuperação e prescrição do paciente.

A prática do CEMAH envolve o estudo de um conjunto de músculos, obtido pelo resultado de diversos ensaios. No entanto, como não existe uma forma sistemática de validar o impacto de um músculo, este conjunto varia ao longo do tempo, para que possam ser reavaliados. Também teorizam a importância da análise do membro simétrico ao afetado. Os dados recolhidos apontam para a possibilidade, mas não existe nenhum estudo que permita comprovar este fenómeno.

Relativamente aos dados recolhidos, não existe nenhuma forma automática de extrair os dados desejados, sendo necessário utilizar tratamentos e análises sem suporte de um sistema. Mesmo com os tratamentos e análises existentes não existe forma de prever a recuperação do paciente de forma precisa. Atualmente, o tratamento prescrito ao paciente é realizado por peritos, baseado no conhecimento que estes possuem.

O projeto descrito neste documento visa permitir definir com maior precisão os músculos que devem ser analisados, permitindo um estudo mais focado e com resultados com maior qualidade. Durante este processo será ainda validada a hipótese da relevância do membro simétrico. Com todos os dados recolhidos será ainda desenvolvido um sistema para encontrar perfis de pacientes, de forma a os classificar. Após a fase de análise de dados, será desenvolvida uma aplicação que permitirá identificar um paciente nos padrões previamente

encontrados e, dessa forma, auxiliar os utilizadores no processo de recuperação e prescrição de pacientes, na área de previsão e apoio na escolha de tratamentos de recuperação.

É de seguida exposta a análise de valor realizada à proposta de solução. É apresentada a proposta de valor, modelo de negócio e rede de valor presente na abordagem ao problema.

2.1 Modelo de desenvolvimento

A primeira fase do desenvolvimento de qualquer aplicação parte do desenvolvimento do seu modelo. Para o projeto apresentado utilizamos o modelo de Peter Koen.

Segundo este modelo, existem cinco elementos chaves do modelo *the new concept development model* (NCD) [5]. Estes cinco elementos estão presentes no projeto apresentado neste documento.

A primeira fase é a identificação da oportunidade [6]. Nesta fase, tipicamente, o principal fator são os objetivos empresariais. A oportunidade presente neste documento surgiu na identificação por parte do CEMAH de uma oportunidade de melhorar a compreensão e atuação por parte dos profissionais médicos no tratamento de pacientes que sofreram de um AVC.

A essência desta fase está nas fontes e métodos utilizados para identificar que oportunidades devem de ser seguidas [7]. Os principais métodos/técnicas e/ou ferramentas utilizados são *roadmapping*, análise e previsão de tendências tecnológicas e de consumo, análise de inteligência competitiva, pesquisas de mercado e simulação de cenários. No entanto, no caso apresentado, a identificação surgiu durante o estudo realizado no CEMAH de pacientes em recuperação, junto dos hospitais. Foi detetado que haveria possibilidade de informatizar o trabalho realizado, levando a um melhoramento dos resultados obtidos.

A segunda fase, em suma, centra-se na confirmação que a oportunidade anteriormente identificada tem valor necessário para ser estudada. É ainda necessário recolher mais informação através dos métodos referidos na fase anterior.

Para o CEMAH foi de fácil constatação que a oportunidade identificada possuía valor suficiente, dado que atualmente não existe nenhuma ferramenta para a utilização neste âmbito.

A geração e enriquecimento da ideia surge no momento em que o departamento de investigação, constituído pelo CEMAH, se reuniu com o departamento informático, constituído pelo autor do documento. Durante o processo de troca de conceitos, objetivos e metodologias, a oportunidade identificada transformou-se num produto. A identificação de várias soluções tecnológicas e respetiva aplicabilidade ao projeto permitiu o crescimento das ideias inicialmente formadas.

O passo seguinte foi a seleção da ideia. O objetivo neste passo foi a concretização das várias ideias numa ideia concretizável e que satisfaz os requisitos necessários, removendo ideias que não eram possíveis atingir numa fase inicial. Nesta fase foram discutidos vários cenários, com diferentes técnicas e ferramentas, de forma a selecionar a solução mais vantajosa.

Por fim, foi inicializada a definição do conceito. Ou seja, a definição concreta de objetivos e de um plano de trabalho. Como não existem muitos riscos associados a este projeto, o principal objetivo foi criar um plano de trabalho exequível no tempo disponível sem comprometer o

funcionamento final desejado.

2.2 Valor

Valor é, por definição, o valor monetário, material ou avaliado de um ativo, bem ou serviço. Na proposta apresentada o valor provém de um serviço. Este serviço será facultado aos investigadores e profissionais médicos na forma de uma plataforma Web que irá informar e auxiliar os mesmos durante o processo de recuperação e prescrição.

O valor para os utilizadores será a presença de benefícios durante o processo referido. É apresentada uma solução que permitira centralizar informação e auxiliar o utilizador, através de dados inferidos e de um sistema de apoio à decisão.

Como atualmente o processo é realizado de forma manual, a plataforma irá trazer um automatismo presentemente não existente, com todas as vantagens já referidas. A sua utilização irá permitir obter resultados fundamentados e sistemáticos, de forma intuitiva e centralizada.

No projeto apresentado neste documento, os clientes da aplicação (profissionais médicos) beneficiarão de acesso a mais informação sobre o AVC, previsões sobre o paciente e respetiva recuperação e, por fim, um conjunto de recomendações para a prescrição de tratamentos.

No entanto, para usufruir dos benefícios listados terão que aprender a utilizar o software disponibilizado e utilizar a informação fornecida nos processos de trabalho já utilizados. Isto levará a uma alteração de rotinas e aprendizagem de novas ferramentas. Há ainda o risco da desatualização das previsões realizadas, induzindo alguns erros no processo.

Os pacientes (cliente do cliente) terão como benefício uma avaliação mais rápida e melhor ponderada. Poderão ainda aceder a parte da informação do sistema. Este acesso levará também a um curto processo de aprendizagem e a uma alteração do paradigma atual de consulta de informação.

Durante o período prévio à aquisição e utilização do sistema, poderá existir algum descrédito da informação fornecida (por parte dos profissionais médicos). É desejado um software com capacidade de disponibilizar informação detalhada, acompanhada da fundamentação dos resultados apresentados.

A transição levará a um processo de aprendizagem e adaptação de rotinas. Este processo pode ser dispendioso, dada a complexidade do negócio e das rotinas neste imposto. No entanto, o valor obtido deste processo é objetivado superar os contratemplos lançados nesta fase.

Após a integração da aplicação e conhecimento nos processos é espectável a creditação parcial da informação providenciada, com sugestões de melhoramentos e adaptações a ambientes reais. Concretamente, poderá ser necessário adaptar a disposição da informação e respetiva justificação aos utilizadores. Este processo poderá ser alcançado através da adaptação dos textos aos termos utilizados habitualmente nos processos dos utilizadores, alterar o formato e apresentação da informação, entre outros.

Por fim, após a adaptação do *software* e dos seus utilizadores, é pretendido que estes utilizem o *software* no seu fluxo de trabalho. Futuramente, é ambicionada a inclusão total do *software* nos processos de trabalho já utilizados, apresentando as suas mais-valias. Existe ainda a possibilidade de alargamento da área de negócio ou integração com outros sistemas (atuais ou novos).

2.3 Proposta de valor

Plataforma Web que visa melhorar a compreensão atual do efeito da patologia AVC num paciente, tornando mais eficiente a forma como este recupera e é tratado. Esta ferramenta une o conhecimento de peritos na área da atividade humana com as novas tecnologias de informação, tornando-se um bem indispensável para profissionais médicos e terapêuticos que têm como foco o bem-estar e recuperação dos seus pacientes num ambiente comunicativo.

2.4 Modelo de negócio

Foi escolhido a representação do modelo de negócio através de um diagrama de *canvas* [8] Figura 1.

Na secção Proposta de valor é exposta a proposta de valor que é apresentado no diagrama de forma resumida. Neste tópico é importante referir que a disponibilização de informação se destina a ambos os segmentos apresentados, enquanto que os restantes pontos na proposta de valor não o são. Relativamente à segmentação de clientes podemos discriminar os profissionais médicos e terapêuticos como principal foco. Nesse sentido, as relações criadas com os nossos clientes serão realizadas, maioritariamente, em convenções e hospitais e centros terapêuticos adequados ao tema. Inicialmente será utilizada a plataforma Web como canal de comunicação havendo a possibilidade de integração com outros sistemas no futuro.

Para o arranque do projeto – plataforma Web, análise de dados e sistema de recomendação – as principais atividades serão o desenvolvimento de *software* e análise de dados. Para tal são necessários dados de pacientes e peritos na área. É de extrema importância colaborar com o CEMAH e, por consequente, hospitais e centros terapêuticos.

Para o desenvolvimento será necessário recolher informação e disponibilizar o *software* criado. Estas duas atividades terão um custo. Os ganhos poderão surgir de patrocinadores e, no futuro, de um sistema de subscrição à plataforma.

Para a aquisição inicial de clientes vão ser abordados possíveis clientes em exposições, certames e convenções da área, com valores de aquisição baixos – aquisição fundador. Desta forma vai ser possível criar uma carteira de clientes de forma mais rápida, transpondo mais segurança a futuros clientes. Posteriormente será tomada uma atitude mais proativa nas redes sociais e fóruns, com foco na aquisição de clientes através da apresentação de resultados e testemunhos dos clientes atuais e apelando ainda à conveniência, segurança e facilidade de utilização da plataforma Web.

De forma a reter os clientes vai ser disponibilizado um atendimento pessoal, com atenção na integração da aplicação nos atuais processos de cada cliente e implementação gradual de mais funcionalidades com um sistema grátis de atualização de versões.

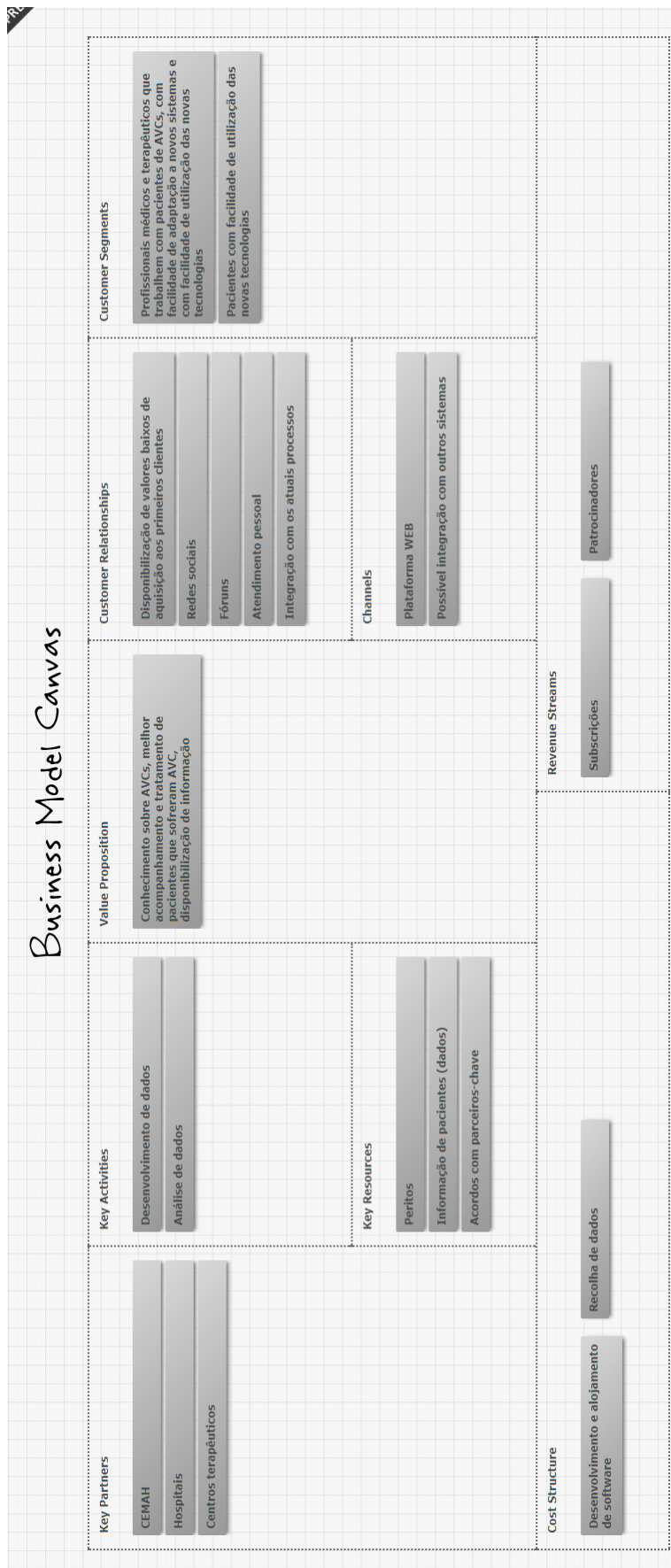


Figura 1 – Modelo de *canvas*

2.5 Rede de valor

Dado que a solução exposta é uma ferramenta a ser disponibilizada gratuitamente e a ser integrada com sistemas atuais, a melhor forma para a analisar será uma rede de valor, como a apresentada por Verna Allee [9] [10].

O valor do projeto está centrado na informatização de conhecimento atual, disponibilizando ainda novo conhecimento atualmente procurado, mas de difícil comprovação. Neste sentido, o projeto necessita de uma interação grande de diferentes departamentos. O departamento que possui o conhecimento (investigação) fornece dados de pacientes, métodos de análise e ainda hipóteses que gostariam de ver testadas.

O departamento de investigação tem de transmitir toda esta informação ao departamento de engenharia informática. Este está dividido em três módulos: análise de dados, sistema de recomendação e desenvolvimento da plataforma Web. O primeiro irá recolher e tratar a informação, testando teorias fornecidas, sistematizando conhecimento e gerando informação. Esta informação será utilizada pela plataforma Web, utilizando a mesma para auxiliar os profissionais médicos. Ainda nesta plataforma é necessário utilizar o sistema de recomendação para auxiliar os terapêuticos médicos na prescrição de tratamentos.

A Figura 2 apresenta a rede de valor do sistema, onde se pode ver a comunicação entre os diferentes departamentos. Podemos verificar que existem trocas intangíveis de conhecimento e dados, especialmente do departamento de investigação para o departamento de análise de dados. É ainda possível verificar as trocas tangíveis, com especial ênfase nas realizadas pela plataforma Web.

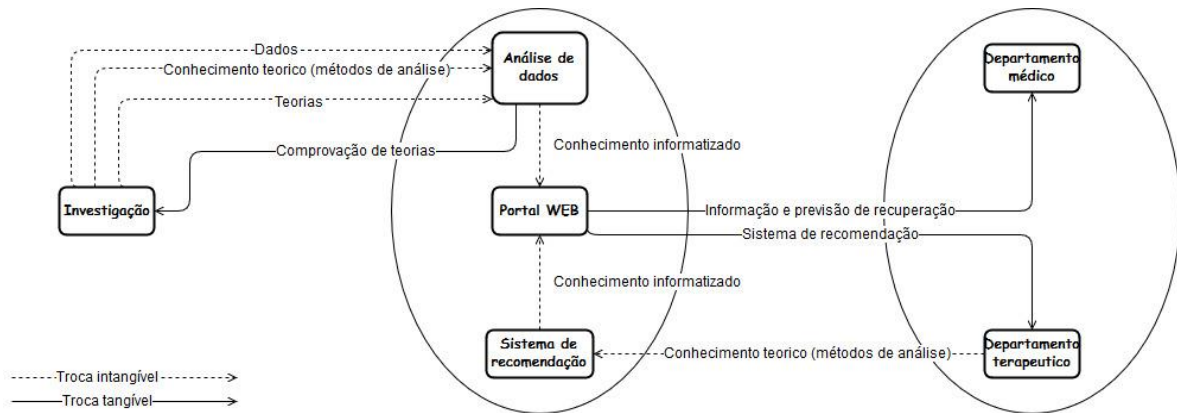


Figura 2 – Rede de valor

Sumariamente, é pretendido utilizar a informação atual dos departamentos de investigação para criar um sistema informatizado que auxilie, melhore e automatize parte do trabalho realizado pelo departamento médico.

3 Estado da arte

Atualmente, existem grandes quantidades de dados a serem recolhidas a todo o momento. No entanto, estes dados por si só não apresentam informação útil. Para tal, há uma necessidade de os analisar, transformar e apresentar como conhecimento. A este processo chama-se mineração de dados [11].

A necessidade de obter conhecimento a partir de grandes fontes de dados surgiu no final do século XX, com grandes empresas a armazenar dados sem terem a possibilidade de extrair conhecimento útil para o melhoramento do seu negócio [12]. Surge assim a mineração de dados.

A mineração de dados é um processo de análise que permite explorar dados à procura de conhecimento, e pode ser aplicado a diversas indústrias, como a medicina, o marketing, as vendas, etc. Este processo permite a descoberta de padrões não conhecidos pelos peritos do sistema (exemplo: a descoberta de padrões de consumo), gerar previsões (exemplo: previsão de aceitação de artigos por um mercado) e auxiliar a tomada de decisões fornecendo informação (exemplo: descobrir que genes são responsáveis por uma mutação específica).

A mineração de dados é um processo que inclui cinco elementos principais [13]:

1. Extrair, transformar e carregar dados: estes podem provir de repositórios de dados individuais (bases de dados, ficheiros, entre outros) ou um sistema integrado de repositórios de dados (armazém de dados).
2. Armazenamento e gestão de dados.
3. Fornecer informação a peritos e profissionais de tecnologias da informação.
4. Analisar os dados.
5. Apresentar os dados em um formato útil, como um gráfico ou tabela.

Como referido, o processo de descoberta de conhecimento é iterativo. Esta faz parte de um processo de transformação de dados em conhecimento. Os seus passos são apresentados na Figura 3 [14]:

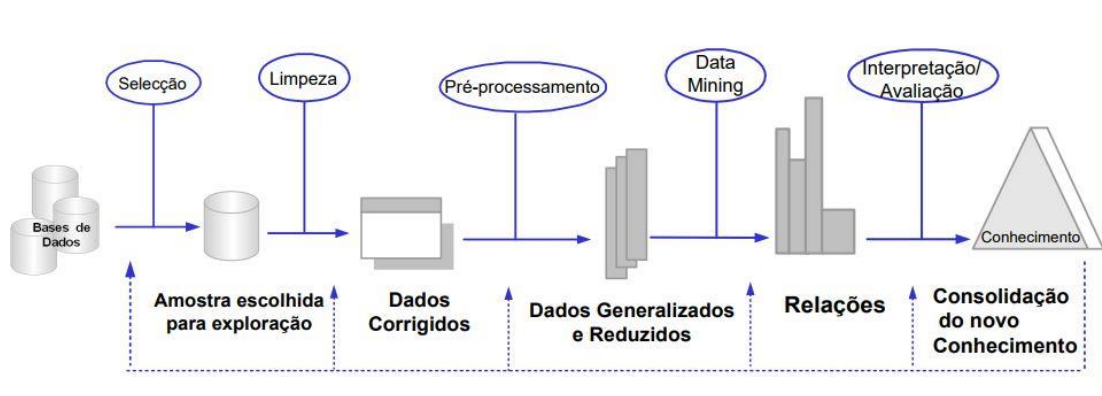


Figura 3 – Passos da mineração de dados

1. Seleção de dados: nesta fase são analisados os dados dos repositórios, de forma a compreender quais os dados relevantes ao estudo que vai ser realizado. O tamanho desta seleção varia mediante o repositório, o contexto e o estudo.
2. Pré-processamento: após a seleção dos dados é necessário realizar a limpeza e correção dos mesmos, garantindo a consistência dos dados. A limpeza de dados passa pela eliminação de ruído, duplicação, inconsistências e a eliminação de outros erros que possam existir nos dados.
3. Transformação: preparação de dados para a fase de mineração de dados. A preparação centra-se na formatação dos dados pré-processados.
4. Mineração de dados: este é o passo do processo que permite obter conhecimento a partir dos dados. Para tal, são aplicadas diversas técnicas de mineração de dados de forma a extrair conhecimento.
5. Interpretação e avaliação: o conhecimento obtido no passo anterior necessita de uma interpretação, de forma a dar ao mesmo valor. O resultado nem sempre é o pretendido, sendo sempre necessário avaliá-lo. Caso se prove insatisfatório, pode obrigar ao retorno a um dos passos anteriores, de forma a produzir melhor conhecimento.

Existem inúmeras técnicas aplicáveis nesta área. São apresentadas de seguida algumas das mais utilizadas:

- Classificação e previsão: As técnicas de classificação e previsão permitem a análise de dados e de comportamentos segundo certas regras. Com esta análise é possível atribuir uma classificação (categoria ou valor) a novos dados.
- *Clustering*: Com esta técnica é possível agrupar dados segundo a sua semelhança. Esta técnica é muito importante porque permite a criação de padrões/perfis que podem ser incorporados em outras técnicas.
- Análise de valores atípicos: a deteção de valores que fogem à distribuição normal dos dados pode transmitir um significado importante em alguns contextos, como anomalias ou comportamentos fraudulentos.
- Análise de associações: A descoberta de associações permite a compreensão mais profunda de comportamentos. Pode ser aplicável, por exemplo, em *marketing* e produção.
- Análise de evolução: esta técnica visa possibilitar a previsão de tendências de comportamento. As tendências de comportamento permitem prever o comportamento de entidades a certos ambientes, possibilitando adaptar o processo a diferentes cenários.

Embora o processo de mineração de dados apresente muitas vantagens, tem associadas algumas desvantagens. Elas são:

- Conhecimento pouco preciso ou errado: a mineração de dados tem como objetivo a descoberta de conhecimento previamente desconhecido ao sistema. Caso o processo de mineração não apresente resultados precisos ou que não sejam corretos pode levar a investimentos de negócio com efeitos desastrosos.
- Interpretação dos dados: caso exista uma interpretação errada dos dados pode levar a decisões erradas, como no ponto anterior.
- Tempo de processamento: quando a quantidade de dados a analisar é muito elevada o tempo de todo o processo aumenta drasticamente.
- Custo de implementação: desde da recolha de dados à geração de conhecimento, o processo pode demorar muito tempo, dada a iteratividade do mesmo. Como existe um custo associado a todo o processo, o custo total pode-se tornar bastante elevado.

Para a aplicação da mineração de dados existem várias ferramentas disponíveis no mercado. Entre elas encontram-se:

- WEKA (Waikato Environment for Knowledge Analysis) [15] – “(...) é uma coleção de algoritmos de aprendizagem de máquinas para tarefas de mineração de dados. Os algoritmos podem ser aplicados diretamente a um conjunto de dados ou chamados de seu próprio código Java. Weka contém ferramentas para pré-processamento de dados, classificação, regressão, agrupamento, regras de associação e visualização. Também é adequado para o desenvolvimento de novos esquemas de aprendizagem.”.
- Matlab [16] – “A plataforma MATLAB é otimizada para resolver problemas de engenharia e científicos. (...) Os gráficos incorporados facilitam a visualização e a obtenção de informações dos dados. Uma vasta biblioteca de caixas de ferramentas pré-construídas que permite que se comece imediatamente com algoritmos essenciais para o domínio.”.
- R [17] – “R é um ambiente de software grátis para computação estatística e gráfica. (...) R fornece uma grande variedade de estatística (modelagem linear e não-linear, testes estatísticos clássicos, análise de séries temporais, classificação, agrupamento, ...), técnicas gráficas e é altamente extensível.”.
- Oracle Data Mining (ODM) [18] – “(...) um componente da Oracle Advanced Analytics Database Option, fornece algoritmos de mineração de dados poderosos que permitem aos analistas de dados descobrir informações, fazer previsões e alavancar seus dados e investimentos Oracle.”.
- Microsoft SQL Server [19]– “O SQL Server tem sido líder em análises preditivas desde a versão lançada em 2000, fornecendo mineração de dados em Analysis Services. A combinação de Serviços de Integração, Reporting Services e SQL Server Data Mining fornece uma plataforma integrada para análises preditivas que engloba a limpeza e preparação de dados, a aprendizagem e reporting.”.

3.1 Conceitos

Nesta secção são apresentados alguns dos conceitos utilizados no documento. Estes são relacionados à análise de músculos, presente na fase de análise de dados.

- **AVC** – vulgarmente designado de derrame cerebral, é caracterizado pela perda rápida da função neurológica. Este derrame acontece devido ao rompimento ou entupimento dos vasos sanguíneos do cérebro. É uma doença súbita em que o doente apresenta paralisção ou dificuldade em movimentar os membros de um só lado do corpo, dificuldade em falar ou articular palavras e défice visual súbito de uma parte do campo visual.

- **Membro afetado** – membro do paciente que, após este ter sofrido de um AVC, apresenta problemas na realização de movimentos. Neste contexto, será sempre um membro inferior.
- **Membro simétrico** – membro oposto ao referido. Neste contexto, será sempre o membro inferior oposto ao membro afetado.
- **Músculo** - O músculo é um tecido mole responsável pelo movimento.
- **Atividade basal** – atividade de um músculo num estado não ativo, manifestada pela existência de uma pequena corrente elétrica.
- **Ativação de um músculo** - A ativação de um músculo está associada ao seu recrutamento para uma certa tarefa. Genericamente, os músculos estão sempre a ser percorridos por pequenas correntes elétricas, que resultam numa atividade basal. Quando o nível dessa atividade aumenta dizemos que o músculo foi ativado. Consideramos o músculo ativo quando o valor médio da sua atividade ultrapassa o valor médio basal mais de dois desvios padrão.

3.2 Identificação de variáveis relevantes

Na área da descoberta do conhecimento, existem técnicas que podem ser aplicadas para compreender a importância dos elementos constituintes dos modelos que compõe o negócio. À expansão e crescimento das áreas de negócio está, por consequente, associado um aumento de tamanho e complexidade dos modelos que nele atuam. Cada modelo por sua vez é constituído por numerosas variáveis que o descrevem. No entanto, nem todas essas variáveis apresentam um valor significativo para o propósito geral do negócio.

A análise de dados de um negócio complexo pode ser um processo demorado e com resultados menos precisos que o pretendido, caso não exista primeiramente um estudo e compreensão dos modelos. Este estudo começa pela seleção dos modelos e das suas variáveis que apresentam uma maior relevância. O conceito de relevância pode variar, no mesmo negócio, mediante o estudo que esteja a ser realizado. A relevância pode ser classificada como forte, fraca ou redundante. Existem várias aplicações destas técnicas, sendo alguns exemplos:

- *Feature Selection for fault detection systems: application to the Tennessee Eastman Process* [20]

Objetivo – Compreensão e seleção das variáveis relevantes aos sistemas de deteção de erro.

Método – Utilização do algoritmo STRASS (Strong Relevant Algorithm of Subset Selection). Este algoritmo analisa a relação entre as variáveis de forma a obter um modelo simplificado e com melhor performance, mas igualmente representativo.

Conclusão – O algoritmo desenvolvido tem como base *feature selection*. Este foi aplicado sobre Tennessee Eastman Process (TEP) para validação, obtendo resultados mais precisos (menos 1,4% classificações mal catalogadas face à técnica *1-nearest neighbor*). Foram ainda testados face a outras técnicas de *feature selection* onde se apresentaram resultados iguais ou superiores a nível de precisão.

- *A comparative study of feature selection in text categorization* [21]

Objetivo – Análise do resultado de diferentes algoritmos de *feature selection* para a categorização de texto.

Método – Os métodos avaliados foram *Document frequency thresholding (DF)*, *Information Gain (IG)*, *Mutual information (MI)*, *χ^2 Statistic (CHI)* e *Term Strength (TS)*.

Conclusão – As técnicas permitiram remover cerca de 98% dos termos únicos (exceto palavras vazias). As técnicas IG e CHI apresentaram a melhor eficiência sem perder a precisão de categorização.

- *Support Vector Machines based Arabic Language Text Classification System: Feature Selection Comparative Study* [22]

Objetivo – Utilização de técnicas de *feature selection* para a categorização de texto árabe.

Método – Para além dos métodos habituais utilizados em FS (*DF, IG, MI, CHI, MI*) foram ainda aplicados os métodos utilizados na análise e foram também utilizados os métodos Ng-Goh-Low (NGL), Galavotti-Sebastiani-Simi (GSS) e Odd Ratio (OR).

Conclusão – O algoritmo que apresentou o melhor resultado foi o CHI. É proposto o estudo futuro ainda de mais métodos, com o objetivo de obter resultados mais precisos em categorias mais reduzidas.

Para a identificação de variáveis significativas existem duas técnicas estudadas que são aplicáveis. Elas são: *Dimensionality reduction* e *feature selection* [23] [24].

A primeira permite a redução das variáveis num segmento representativo, sem perder a sua estrutura euclidiana. Esta técnica permite a criação de um novo modelo mais reduzido, mas igualmente representativo. As variáveis que constituem este modelo podem coincidir com as variáveis já existentes ou a combinação total ou parcial das mesmas.

A segunda permite selecionar as variáveis mais significativas, sem reduzir a análise. Tal como na primeira técnica será criado um novo modelo igualmente representativo. No entanto, as

variáveis que constituem este serão sempre as variáveis presentes no modelo original salvo as que não são consideradas significativas. Este é a técnica que será aplicada, dado que o modelo existente é constituído apenas pelos músculos e os dados pessoais do paciente.

Estas técnicas podem ser aplicadas em diferentes áreas, tal como a médica, financeira e de gestão. No enquadramento do projeto apresentado neste documento, seguem alguns exemplos aplicados na área da medicina:

- *Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes* [25]

Objetivo – Classificação de anomalias, segundo três tipos de tumor e nove linhas celulares a serem detetadas. São utilizadas técnicas de *feature selection* para essa deteção.

Método – Aplicação de técnicas de *feature selection* presentes no *software RankGene: Information Gain, Gini Index, Twoing rule, Max Minority e Sum of Variances*.

Conclusão – Os métodos foram aplicados de forma robusta e precisa sobre os conjuntos de dados examinados. Os resultados estão de acordo com o conhecimento médico. O estudo confirma dependência entre a classificação e os modelos utilizados.

- *Genetic Algorithms as a Tool for Feature Selection in Machine Learning* [26]

Objetivo – Utilização de técnicas de *machine learning* em ambientes reais e complexos, através de técnicas de *feature selection* e *classification*.

Método – São utilizadas técnicas de análise em que as variáveis são analisadas individualmente face ao resultado pretendido e técnicas que, pelo contrário, escolhem um subconjunto de variáveis relacionadas. Os resultados são depois utilizados num algoritmo genético (GA) para aplicação.

Conclusão – Os resultados obtidos mostram as vantagens da utilização de técnicas de *feature selection* para a criação de regras. Os resultados apresentam ainda a possibilidade de produzir modelos mais reduzidos sem prejudicar o processo de classificação.

Existem diferentes métodos de aplicar a técnica *feature selection*. Eles são:

- **Filter** - neste método são analisadas as variáveis em função do valor a prever, suprimindo as variáveis menos significativas face ao resultado final. No entanto, não analisam as variáveis entre si, resultando várias vezes na redundância de valores.

- **Wrapper** - ao contrário do método *filter*, este analisa a interação entre as variáveis. Caso existam poucos resultados pode levar a resultados demasiados ajustados aos dados inseridos. Se, pelo contrário, existirem muitos dados pode levar a um tempo de processamento muito elevado.
- **Embedded** - por fim, este método propõe uma solução híbrida dos métodos anteriormente referidos. Este método visa a combinação das vantagens dos métodos anteriores, aplicando ainda classificação dos dados em simultâneo.

Existem diferentes algoritmos para a aplicação de *feature selection*. São exemplo [27]:

- **LVF (LAS VEGAS FILTER)** - Gera aleatoriamente subconjuntos e calcula a sua medida de avaliação;
- **LVI (LAS VEGAS INCREMENTAL)** - Baseia-se no facto de não ser necessário utilizar toda a amostra X para avaliar a medida Y;
- **Relief** - Escolhe aleatoriamente uma instância X de Y e determina a instância da mesma classificação mais próxima (*near hit*) e sua instância mais próxima com classificação diferente (*near miss*) em relação a Y;
- **SFG (SEQUENTIAL FORWARD GENERATION)** - Iterativamente adiciona características a um subconjunto inicial, de forma a melhorar uma dada medida X, tendo em conta as características já presentes na solução;
- **SFFS (SEQUENTIAL FLOATING FORWARD SEARCH)** - É um algoritmo de custo exponencial, que em cada passo de seleção realiza um passo em frente seguido de um número variável de passos para trás;
- **Focus** - Avalia cada conjunto de recurso singular e depois cada conjunto de dois recursos e assim por diante;
- **B&B (BRANCH & BOUND)** - Utilizado para fazer pesquisas ótimas, dado um limiar K, a pesquisa para em cada nó cuja avaliação é menor do que K, de modo que os ramos eferentes são suprimidos;
- **QBB (QUICK BRANCH AND BOUND)** – híbrido entre uma variável do B&B (*automatic branch and bound (ABB)*) e LFV. Utiliza o LVF para arranjar um bom início para aplicar o ABB.

3.3 Clustering

Clustering é uma técnica que permite dividir os dados em grupos lógicos [28]. Estes grupos têm associado um significado e podem, por si mesmos, apresentar informação ou servir como componente inicial de futuras análises. Os grupos obtidos partilham características comuns e permitem uma compressão maior dos dados em análise. A criação destes grupos pode ser aplicada em diversas áreas, como a medicina (na segmentação de uma doença), vendas (segmentação de clientes) e até na análise climática (procura de padrões na atmosfera).

Esta técnica tem como requisitos [29] [30]:

- **Escalabilidade**, dado o elevado número de dados a que se aplica o *clustering*, é necessário que este seja escalável.
- **Tem que ser capaz de lidar com dados em diferentes formatos**, como números, texto e datas.
- **Os padrões (*clusters*) encontrados podem tomar formas diferentes**. Os padrões encontrados podem ter formas para além da forma circular (Figura 4).
- **Capaz de lidar com “ruído”**. Deverá ser capaz de lidar com informação pouco relevante presente nos dados.
- **Os resultados devem de ser interpretáveis**, especialmente pelos peritos que serão as entidades com melhor capacidade de avaliação dos resultados.
- **Deve de ser capaz de lidar com dados altamente dimensionais**, para além de lidar com dados de baixa dimensão.

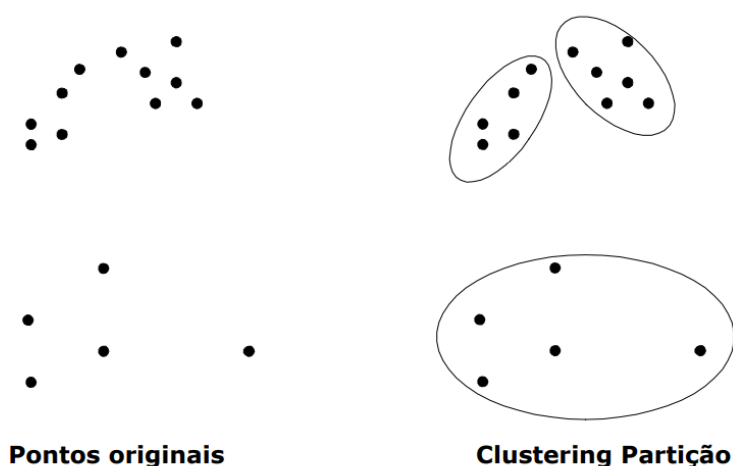


Figura 4 – Identificação de grupos nos pontos resultantes da aplicação de clustering

A aplicação da técnica de *clustering* varia na forma como aborda a divisão dos dados. A aplicação deve de se adaptar ao problema que visa resolver. São de seguida apresentados as abordagens mais comuns [31]:

- **Partitioning** - com estes métodos cada elemento fará parte de um grupo, e cada grupo terá pelo menos um elemento.
- **Hierarchical** - é feita uma decomposição hierárquica. Esta decomposição pode ser feita começando por cada elemento fazendo parte de um grupo (abordagem aglomeraria) ou, pelo contrário, começando por todos os elementos fazerem parte de um só grupo (abordagem divisiva).
- **Density-based** - com este grupo de métodos é utilizado o conceito de densidade, identificando conjuntos densos de pontos. Desta forma, são criados grupos de forma arbitrária e identificados os *outliers* nos dados.
- **Grid-Based** - com esta abordagem, os elementos são divididos numa grelha.
- **Model-Based** - estes modelos criam um modelo para cada cluster.
- **Constraint-based** - abordagem com base nas limitações da aplicação de onde os dados provêm/serão utilizados.

A aplicação da técnica de *clustering* é feita através da aplicação de algoritmos criados para este efeito. Alguns dos algoritmos aplicáveis são [32]:

- ***k-means clustering algorithm*** – Este algoritmo de *clustering* é um dos mais simples para a aprendizagem não-supervisionada, sendo necessário indicar quantos grupos (*k*) queremos encontrar. Os dados serão então agrupados em várias iterações num dos grupos, obtendo assim a classificação pretendida.
- ***Fuzzy c-means clustering algorithm*** - Funciona atribuindo a associação a cada ponto de dados correspondente a cada centro do grupo, com base na distância entre o centro do cluster e o ponto de dados. Quanto mais próximos os dados do centro do grupo maior é a sua adesão ao centro do grupo em particular
- ***Hierarchical clustering algorithm*** – Podendo ser aplicado de forma aglomerativa ou hierárquica, sendo uma o contrário da outra, este algoritmo funciona agrupando os dados um a um com base na medida de distância mais próxima de toda a distância entre os dados. Existem várias formas de calcular estas distâncias, tais como *single linkage*, *complete linkage*, *average linkage*, *centroid distance*, entre outras.

- *Gaussian(EM) clustering algorithm* – Assume, desde do princípio, que existem n grupos, tentando associar toda a informação nesses n grupos. Fá-lo prevendo as classes de todos os dados e, em seguida, maximizando a proximidade dos centros.
- *Quality threshold clustering algorithm* - Este algoritmo requer a especificação da distância *threshold* dentro dos grupos e o número mínimo de elementos em cada grupo. Os pontos são associados aos grupos que estão dentro do alcance da distância *treshold*.

3.4 Sistema de recomendação

Atualmente existe uma variedade de aplicações que possuem módulos especializados no auxílio da escolha por parte do utilizador. Estes sistemas são denominados por sistemas de recomendação. Estes sistemas são vastamente aplicados, especialmente em ambientes de comércio Web. Bons exemplos são os sistemas de recomendação da Amazon [33], Netflix [34] e do Spotify [35]. Quer no âmbito de compra ou consumo, os sistemas de recomendação são uma ferramenta essencial na utilização de sistemas, como um auxiliar de escolha para o utilizador.

A título de exemplo, 66% dos filmes vistos no Netflix são recomendados e 35% das vendas da Amazon surgem de recomendações. São de seguida apresentados alguns exemplos da aplicação destes sistemas:

- Sistema de recomendação da Amazon [36]
 Objetivo – Utilização de um sistema de recomendação de forma a personalizar a experiência de compra de cada utilizador, adaptando a loja ao perfil de cada utilizador.

 Método – É utilizado um algoritmo próprio, denominado *Item-to-Item Collaborative Filtering*. Este algoritmo analisa os itens comprados e avaliados por cada utilizador a itens semelhantes. Para determinar a semelhança entre os itens são avaliadas as compras dos itens que são comprados conjuntamente.

 Conclusões – O algoritmo criado permite à Amazon a criação de uma loja totalmente adaptável, através de um sistema de recomendação baseado no perfil de consumo de cada utilizador e no perfil de venda de artigos.
- Sistema de recomendação da Netflix – Cinematch [37]
 Objetivo – Sistema de recomendação de itens de interesse e divertimento a utilizadores com base no seu perfil, denominado de Cinematch. São utilizados dados recolhidos de utilizadores com mais de 20 avaliações entre as datas de outubro de 1998 e dezembro de 2005.

Método – Sistema de recomendação baseado em perfis. A precisão do sistema é avaliada segundo a diferença dos valores previstos com os valores observados (*root-mean-square deviation* (RMSD)) [38].

Conclusões – 9,6% de melhoria em relação à média das previsões simples de filmes individuais.

A escolha pode ser realizada sobre um número elevado de itens, como livros, música, filmes ou até ideias. O sistema utiliza a atividade e perfil do utilizador para basear a sua recomendação. Se a recomendação surgir com base na atividade do utilizador o sistema funcionará como um *content-based system*. Nestes sistemas são analisadas as propriedades dos itens presentes na atividade do utilizador e são recomendados itens que possuam propriedades comuns. No entanto, caso a recomendação surja do perfil do utilizador é denominado de *Collaborative filtering system*. Nestes sistemas o comportamento do utilizador é analisado e criado um perfil. Este perfil pode ser então agrupado com outros perfis de utilizadores semelhantes (*user-based*). A recomendação surge com base nas escolhas do grupo cujo perfil do utilizador se enquadra. Pode ainda utilizar o perfil do utilizador para ver os itens em que ele mostrou interesse e sugerir itens semelhantes (*item-based*).

Existem ainda outros sistemas que utilizam a demografia e redes sociais de um utilizador para fazer sugestões. Por fim, há sempre combinações de dois ou mais sistemas apresentados, obtendo assim um híbrido que poderá variar na sua implementação.

Existe muita informação armazenada sobre pacientes na área médica. Esta informação contém os dados caracterizantes de um paciente, como idade e género, tal como consultas, exames e tratamentos. Com a evolução das tecnologias começam a ser implementados sistemas, como o sistema de recomendação, como um apoio aos profissionais médicos. Tais sistemas podem intervir em diferentes fases do processo, desde o diagnóstico à prescrição de tratamentos.

São agora apresentados alguns exemplos da aplicação destes sistemas na área médica:

- *Profiling basic health information of tourists: towards a recommendation system for the adaptive delivery of medical certified nutrition contents* [39]

Objetivo – Sistema de recomendação para a distribuição de dietas para turistas doentes, de forma a melhorar a qualidade de vida sem prejudicar as necessidades de lazer dos mesmos.

Método – O sistema proposto analisa as preferências de consumo e o estado de saúde do utilizador, combinando os dois para realizar uma sugestão nutricional. Todo este processo é realizado sobre uma plataforma Web.

Resultado – O sistema apresentado permite uma recomendação nutricional personalizada às preferências e estado de saúde de um utilizador.

- *Implementation of a hybrid approach for chronic disease risk assessment and recommendation system [40]*

Objetivo – O sistema proposto tem três objetivos: auxiliar os médicos a monitorar e prever a condição dos pacientes remotamente, ajudar os pacientes a terem acesso a profissionais de saúde, independentemente do tempo e local, e auxiliar os pacientes a controlar a condição das suas doenças. O propósito deste projeto é implementar um sistema de recomendação que seja capaz de gerir doenças crónicas e fornecer dados adequados aos pacientes.

Método – A abordagem proposta é um sistema híbrido de classificação e *Collaborative Filtering*. Para a recomendação são utilizadas regras de diagnóstico e conselhos médicos, conjuntamente com a classificação e perfil do paciente.

Resultado – O sistema é apresentado como uma ferramenta de apoio à decisão para os profissionais médicos no diagnóstico e gestão de pacientes com doenças crónicas.

Relativamente aos sistemas de recomendação foram analisadas duas diferentes abordagens, com possíveis aplicações. As abordagens são o sistema de fluxo de trabalho e o sistema de regras [41].

O sistema de fluxo de trabalho permite criar planos de tratamentos pré-definidos, com fluxos de tratamentos definidos. Os fluxos são construídos com apoio do conhecimento dos peritos. Com esta solução, o processo é iniciado com a seleção de um fluxo e, por consequente, de um tratamento. Após a realização do tratamento inicial é analisado o resultado e apresentados os tratamentos definidos na próxima etapa do fluxo em questão e cujos resultados obtidos satisfazem as suas condições.

O sistema de regras difere da solução anterior pelo facto de não haver fluxos definidos. É aconselhado e escolhido um tratamento inicial, mediante regras definidas. Como os fluxos, as regras são criadas com base no saber dos peritos. Após o primeiro passo, os dados são analisados novamente perante todas as regras e recomendado um novo tratamento, sem que haja um fluxo pré-definido.

O sistema de recomendação poderá ser complementado por um módulo de explicação. Este módulo permite apresentar a inferência realizada para obter a recomendação apresentada. A justificação é orientada a três tipos de entidades [42]:

- **Engenheiro do Conhecimento/Analista de dados** – Durante o desenvolvimento e validação do sistema a justificação pode auxiliar na deteção de problemas no conhecimento usado. Estes problemas podem centrar-se nos dados ou no conhecimento transmitido ou interpretação do mesmo.
- **Perito** – A justificação vai ajudar o perito a validar a recomendação, avaliando não só o resultado, mas também o raciocínio. Esta validação pode surgir da comparação do raciocínio da aplicação com o seu para além dos resultados.

- **Utilizador** – Para o utilizador pode ser igualmente interessante perceber o raciocínio da aplicação de forma aceitar a recomendação. A justificação pode ainda apoiar a decisão do utilizador dando alguns dados importantes para a formalização da escolha.

3.5 Sistemas semelhantes

Nesta seção é analisada uma solução semelhante à solução proposta neste documento, embora num contexto diferente.

O sistema analisado é o *Intelligence Based Recommendation System for Healthcare: a Patient centred Framework* [43]. A proposta deste documento é criar uma *framework* que permita melhorar a forma de prever, diagnosticar e tratar doentes. Para tal, são analisados diversos parâmetros de pacientes com o objetivo de encontrar os que representam um indicador mais relevante para a *framework*. Isto é atingido através de *feature selection*. Após a identificação dos parâmetros serão criados perfis de pacientes, mediante o objetivo (como, por exemplo, previsão de cancro ou tratamento de hipertensão). Ao analisar um paciente este será classificado mediante um dos perfis criados, através da técnica de *classification*. Estes modelos e os parâmetros foram uma base de conhecimento, utilizada para um sistema de recomendação, como é apresentado na Figura 5.

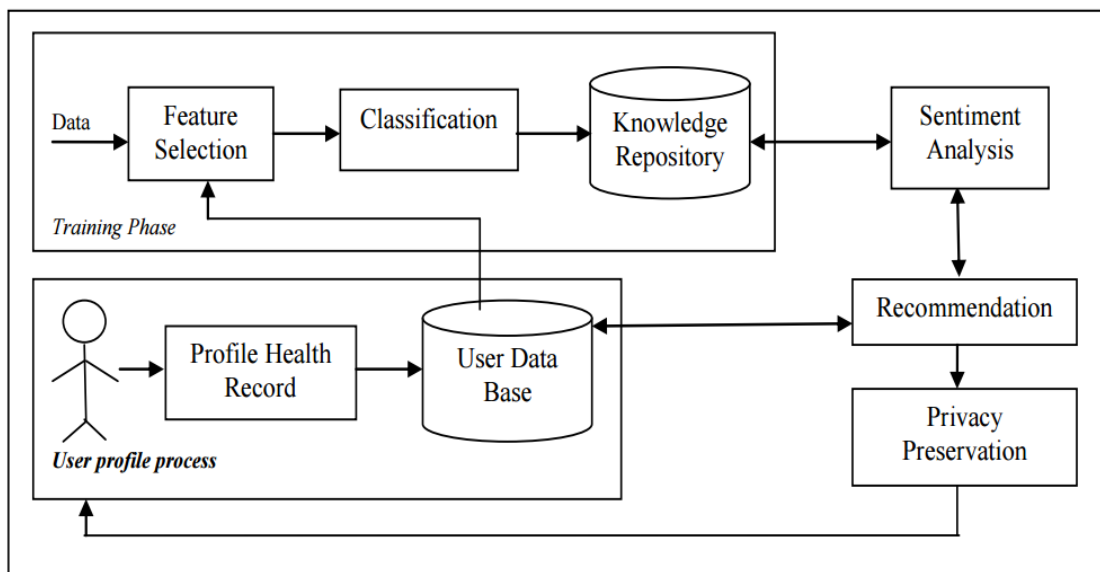


Figura 5 – Arquitetura da *framework*

Na componente de *feature selection* é usada a técnica *Principal Component Analysis* (PCA) para descobrir características relevantes dos dados. Na área de *classification* é usada a técnica *Support vector machines* (SVMs) para classificar um paciente segundo perfis previamente criados. Por fim, o sistema de recomendação utiliza um sistema combinado de regras e perfis de utilizadores para a sugestão de serviços médicos.

Esta estrutura será analisada e no capítulo Sistema proposto, onde uma arquitetura semelhante poderá ser utilizada, onde podem constar técnicas idênticas para a descoberta dos músculos que apresentam uma relevância maior na recuperação da marcha, técnicas de *clustering* para encontrar estes perfis e um sistema de recomendação para o auxílio na prescrição de tratamentos.

4 Avaliação de soluções

No estado da arte foram apresentadas técnicas de mineração de dados diferentes que poderão ser aplicadas na presente proposta. No presente capítulo são analisados os objetivos e discutidas abordagens à solução dos mesmos.

4.1 Identificação de músculos significativos e análise da dependência funcional do membro afetado e do membro simétrico

Dois dos objetivos propostos neste documento são a identificação de músculos significativos e análise da dependência funcional do membro afetado e do membro simétrico. Foram analisadas técnicas de identificação de variáveis relevantes que podem ser estudadas no presente capítulo para a identificação dos músculos mais representativos da marcha. Recordando, esta técnica permite descobrir que variáveis do modelo apresentam uma maior relevância.

Como para o objetivo da análise não é criar um novo modelo, mas sim remover as variáveis que representam os músculos menos relevantes, serão aplicadas técnicas de *feature selection*.

No caso de estudo deste documento, as técnicas de *feature selection* serão aplicadas sobre os músculos e a sua atividade - estes correspondem aos modelos das provas analisadas. Nem todos os músculos terão uma importância semelhante na análise da recuperação da marcha. Para compreender quais os músculos significativos, foram analisadas as atividades musculares do membro afetado durante algumas atividades representativas, tais como levantar, sentar e primeira marcha.

A análise efetuada aos músculos não será realizada individualmente, mas sim em conjunto com os restantes, já que a relevância de um músculo pode ser diminuída ou aumentada caso seja realizada em conjunto com um ou vários dos outros músculos.

No estudo da dependência funcional do membro afetado e do membro simétrico não serão necessariamente estes músculos que iremos analisar. Serão novamente analisados todos os músculos durante as atividades representativas e analisada a sua relevância. Caso existam músculos que apresentem um impacto no restauro da marcha será comprovada uma correlação entre os membros.

Como já referido, os músculos têm que ser analisados perante a retoma da funcionalidade motora do membro. No entanto, a sua atividade pode ser reduzida ou amplificada perante a atividade de outros membros. Nesse sentido, há a necessidade de analisar cada músculo em função da sua atividade e da atividade dos músculos que atuam no mesmo período. Por essa razão, a técnica de *feature selection* mais indicada a utilizar será *Embedded*.

4.2 Perfis de pacientes

Na procura de perfis de pacientes foram aplicadas técnicas de *clustering*. Esta técnica permitiu dividir os pacientes em grupos baseados nas semelhanças dos seus perfis e na sua atividade motora.

Foi objetivado que estes grupos representem padrões de recuperação e resposta a tratamentos semelhantes, de forma a melhorar o apoio ao paciente e a prescrição de tratamentos.

4.2.1 Análise multicritério pode ser aplicado na escolha de algoritmos de *clustering*

Nesta secção são analisados diferentes algoritmos de *clustering*, para serem aplicados na procura de perfis em pacientes. Para a realização da análise foi utilizado o método de Análise Hierárquica (AHP - Analytic Hierarchy Process) (T. SAATY, 1980. The Analytic Hierarchy Process, McGraw-Hill, New York), cujo o foco principal é auxiliar o processo de decisão dividindo o problema em níveis hierárquicos. Estes, por sua vez, são mais fáceis de compreender e avaliar.

O método AHP divide-se em 7 fases:

1. Construção da árvore hierárquica de decisão, que consiste em definir o problema e estruturar através de um diagrama hierárquico.
2. Comparação das alternativas e critérios, estabelecendo prioridades entre os elementos para cada nível da hierarquia, por meio de uma matriz de comparação.
3. Atribuir uma prioridade relativa a cada critério.
4. Avaliar a consistência das prioridades relativas.
5. Construção da matriz de comparação paritária para cada critério, considerando cada uma das alternativas selecionadas.
6. Obter a prioridade composta para as alternativas, utilizando a matriz de comparação do passo 5 com as prioridades relativas atribuídas no passo 3.
7. Escolha da alternativa mediante os resultados obtidos.

Os critérios utilizados para a decisão são: a qualidade do resultado, o desempenho do algoritmo e a adaptação deste a um número elevado de dados. Dada a grande quantidade de dados recolhidos a analisar (atividade de um músculo num espaço de tempo de milissegundos) é importante assegurar que o algoritmo a utilizar consegue produzir resultados satisfatórios com estes dados.

Para realizar a análise foram utilizados dados retirados do estudo *comparisons between data clustering algorithms* [44]. Este estudo compara quatro algoritmos diferentes na tarefa de clustering: K-Means, Hierarchical Clustering (HC), Self-Organization Map (SOM) e Expectation Maximization Clustering (EM). Os fatores utilizados para a análise foram o tamanho do conjunto de dados, número de clusters, tipo de dados e *software* utilizado. Com estes foi possível analisar o desempenho, qualidade e adaptação a um número elevado de dados dos algoritmos. O desempenho foi avaliado segundo a métrica de performance, a qualidade

segundo a métrica de exatidão e a adaptação a um número elevado de dados segundo a precisão com 36000 registos (600 linhas por 60 colunas). Cada parâmetro foi analisado utilizando um número de clusters desejado (no caso do algoritmo SOM esse número são os pontos por rede) – 8, 16, 32 e 64. Para a análise feita de seguida foi utilizada a média dos valores obtidos com os quatro números de grupos diferentes.

A Tabela 1 apresenta os resultados de aplicação dos algoritmos:

Tabela 1 – Tabela de classificação

Algoritmos	Qualidade	Desempenho	Adaptação a um número elevado de dados
A1 - K-Means	987,75	76,75	910
A2 - HC	915	82	850
A3 - SOM	875,25	72,25	830
A4 - EM	973,75	76	898

Considerando os critérios de avaliação referidos, estes são agora comparados entre si. O resultado é apresentado na Tabela 2:

Tabela 2 – Matriz de comparação de critérios

Crítérios	Qualidade	Desempenho	Adaptação a um número elevado de dados
Qualidade	1,00	4,00	3,00
Desempenho	0,25	1,00	2,00
Adaptação a um número elevado de dados	0,33	0,50	1,00

O passo seguinte é a normalização dos dados obtidos. O resultado deste processo está presente na Tabela 3 e na Tabela 4:

Tabela 3 – Matriz de comparação de critérios

Crítérios	Qualidade	Desempenho	Adaptação a um número elevado de dados
Qualidade	1,00	4,00	3,00
Desempenho	0,25	1,00	2,00

Adaptação a um número elevado de dados	0,33	0,50	1,00
Soma	1,58	5,50	6,00

Tabela 4 – Normalização da matriz de comparação de critérios

Critérios	Qualidade	Desempenho	Adaptação a um número elevado de dados
Qualidade	0,63	0,73	0,50
Desempenho	0,16	0,18	0,33
Adaptação a um número elevado de dados	0,21	0,09	0,17

Com a matriz normalizada é possível realizar o cálculo da prioridade relativa dos mesmos, como é apresentado na Tabela 5:

Tabela 5 – Cálculo das prioridades relativas

Critérios	Qualidade	Desempenho	Adaptação a um número elevado de dados	Prioridade relativa
Qualidade	0,63	0,73	0,50	0,620061373
Desempenho	0,16	0,18	0,33	0,224459788
Adaptação a um número elevado de dados	0,21	0,09	0,17	0,155478839

Com estes dados podemos verificar que o índice de consistência é de 0,05 e a razão de consistência de 0,09. Como o valor é inferior a 0,1 podemos afirmar que as prioridades relativas são consistentes.

As Tabela 6, Tabela 7, Tabela 8, Tabela 9, Tabela 10 e Tabela 11 apresentam o cálculo das prioridades relativas dos critérios de qualidade, desempenho e adaptação a um número elevado de dados.

Tabela 6 - Cálculo da prioridade relativa da qualidade

Qualidade	A1	A2	A3	A4
A1	1	5	9	3
A2	0,2	1	5	3
A3	0,11	0,2	1	5
A4	0,33	0,33	0,2	1

Tabela 7 - Cálculo da prioridade relativa da qualidade com vetor de prioridade

Qualidade	A1	A2	A3	A4	Vetor de Prioridade
A1	0,608108108	0,765306122	0,592105263	0,25	0,553879873
A2	0,121621622	0,153061224	0,328947368	0,25	0,213407554
A3	0,067567568	0,030612245	0,065789474	0,416666667	0,145158988
A4	0,202702703	0,051020408	0,013157895	0,083333333	0,087553585

Tabela 8 - Cálculo da prioridade relativa do desempenho

Desempenho	A1	A2	A3	A4
A1	1	0,5	3	1
A2	2	1	4	3
A3	0,33	0,25	1	0,5
A4	1	0,33	2	1

Tabela 9 - Cálculo da prioridade relativa do desempenho com vetor de prioridade

Desempenho	A1	A2	A3	A4	Vetor de Prioridade
A1	0,230769231	0,24	0,3	0,181818182	0,238146853
A2	0,461538462	0,48	0,4	0,545454545	0,471748252

A3	0,076923077	0,12	0,1	0,090909091	0,096958042
A4	0,230769231	0,16	0,2	0,181818182	0,193146853

Tabela 10 - Cálculo da prioridade relativa da adaptação a um número elevado de dados

Adaptação a um número elevado de dados	A1	A2	A3	A4
A1	1	5	9	7
A2	0,2	1	3	0,5
A3	0,11	0,33	1	0,3
A4	0,14	2	3,33	1

Tabela 11 - Cálculo da prioridade relativa da adaptação a um número elevado de dados com vetor de prioridade

Adaptação a um número elevado de dados	A1	A2	A3	A4	Vetor de Prioridade
A1	0,687772926	0,6	0,551020408	0,795454545	0,65856197
A2	0,137554585	0,12	0,183673469	0,056818182	0,124511559
A3	0,076419214	0,04	0,06122449	0,034090909	0,052933653
A4	0,098253275	0,24	0,204081633	0,113636364	0,163992818

Os resultados da avaliação dos algoritmos apresentados na tabela 12, segundo os parâmetros de qualidade, desempenho e adaptação a um número elevado de dados, o melhor algoritmo de *clustering* a aplicar é o *K-Means* (segundo a atribuição de importância dada aos parâmetros).

Tabela 12 – Avaliação dos algoritmos

Solução	Qualidade	Desempenho	Adaptação a um número elevado de dados	Avaliação
---------	-----------	------------	--	-----------

A1 - K-Means	0,553879873	0,238146853	0,65856197	0,499286358
A2 - HC	0,213407554	0,471748252	0,124511559	0,257573206
A3 - SOM	0,145158988	0,096958042	0,052933653	0,120000726
A4 - EM	0,087553585	0,193146853	0,163992818	0,123139711

4.2.2 Algoritmo de *clustering K-means*

O algoritmo então escolhido para utilizar no processo de *clustering* é o K-Means (MacQueen, 1967) [45] [46]. Este é um método de aprendizagem não supervisionada e que divide os dados em k grupos (clusters), colocando os dados no grupo mais próximo. O seu funcionamento pode ser sumariado em 4 passos:

1. Atribuição de centros de grupos (centroids), mediante os dados a analisar. São criados k centros,
2. Atribuir um grupo a cada registo dos dados, colocando-o no grupo mais próximo,
3. Quando todos os registos tiverem sido associados a um grupo os centros dos grupos são recalculados,
4. Os passos 2 e 3 são repetidos até os centros dos grupos deixarem de se moverem.

O funcionamento descrito pode ser aplicado aos dados relativos aos músculos. No entanto, no teste de levantar e sentar o objetivo prende-se na análise da evolução do centro de gravidade. Neste caso específico não é pretendido agrupar cada registo, mas sim agrupar conjuntos de registos – a variação do centro de gravidade. Existe uma implementação do K-Means denominada de KML [47] que será utilizada para o efeito.

O KML é uma implementação desenhada especificamente para permitir o *clustering* de dados longitudinais. Assim, torna-se possível analisar e agrupar as trajetórias do centro de gravidade ao longo da prova.

4.3 Prescrição de tratamentos

Para a prescrição de tratamentos a pacientes será criado um sistema de recomendação, baseado em regras. Este sistema facilita a captura e representação do conhecimento dos peritos, de forma clara e de fácil manutenção. Esta representação simplifica ainda a validação pela parte dos peritos do sistema. As regras serão criadas com base no conhecimento dos peritos – investigadores e terapeutas. O sistema visa automatizar e melhorar o atual processo de prescrição manual. Este processo habitualmente tem sempre o mesmo tratamento como início e, após este, são analisados os resultados e possíveis tratamentos consequentes. Será nesta fase que o sistema irá apoiar a decisão do responsável pela prescrição.

No tratamento de pacientes que sofreram de AVC não existem fluxos de tratamentos. O processo é iterativo, sendo analisado não só o historial e ordem dos tratamentos, mas também

o resultado do tratamento na prescrição do próximo tratamento. O facto de não haver um fluxo pré-definido invalida a utilização de um sistema de fluxo de trabalho. Por outro lado, o sistema de regras responde a todas as necessidades levantadas. Desse modo, foi escolhido este para utilização neste projeto.

5 Sistema proposto

O capítulo de design tem como objetivo apresentar a solução proposta. São apresentadas diferentes abordagens possíveis para solucionar os problemas apresentados, fazendo referência aos dados utilizados e à arquitetura adoptada. Por fim, são comparadas as alternativas, realçando as vantagens da arquitetura escolhida.

5.1 Design da solução para o problema

A solução proposta neste documento é, de facto, dividida em dois sistemas: um sistema de análise de dados e um sistema de informação e apoio à decisão.

O sistema de análise de dados foi realizado sobre um software para computação estatística e geração de gráficos. O *software* permite aceder a repositórios externos de dados e disponibilizar módulos de mineração de dados, com capacidade de analisar grandes quantidades de dados. Deverá ainda permitir a integração com outros sistemas, já que será integrado com a plataforma Web apresentada de seguida.

O sistema de informação e apoio à decisão é traduzido por uma plataforma Web, suportado por uma base de dados. Este sistema deverá ser capaz de comunicar com todos os módulos desenvolvidos, sendo o único que terá interatividade com o utilizador final. Devida a esta interação precisará ainda de uma interface de amigável e de fácil utilização. Esta aplicação contará com a possibilidade de os profissionais médicos adicionarem informação dos pacientes, consultarem informação apurada automaticamente (baseada no sistema anterior) e ainda de consultarem recomendações e registarem tratamentos, com os respetivos resultados. Nesta aplicação será ainda possível a interação do paciente, que poderá consultar alguma da informação da sua ficha clínica. Dada a diversidade de necessidades e pontos de acesso, a plataforma não deverá de necessitar de instalações ou configurações, podendo ser acessível a partir de qualquer plataforma com acesso à Internet, com o auxílio de um navegador de Internet.

5.2 Arquitetura

O utilizador comunica com o sistema através da interface (Figura 6). É esta ação que dará início a todo o processo de utilização da plataforma disponibilizada.

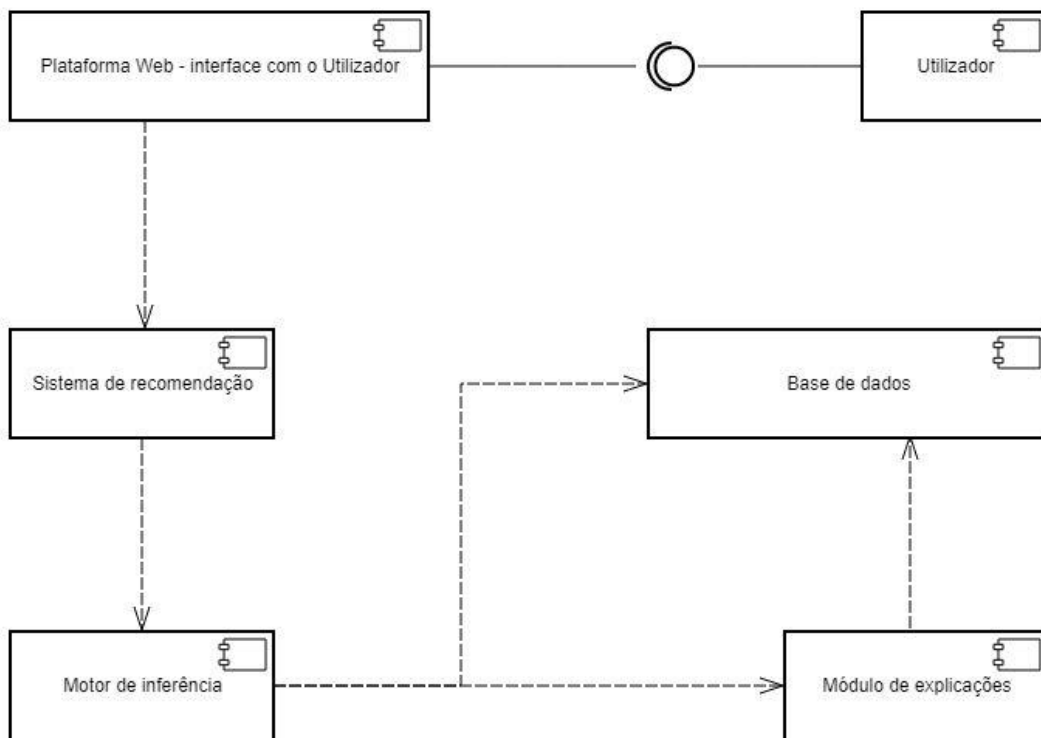


Figura 6 – Diagrama de componentes (UML)

Na camada de interface são processados todos os *inputs* e *outputs* de forma a serem apresentáveis ao utilizador. O utilizador inicia o processo autenticando-se no sistema. Mediante o tipo de utilizador (profissional médico ou paciente) é exposta informação e ações correspondentes ao nível de acesso. Os pacientes podem consultar a sua ficha de paciente, apenas com informação relativa aos exames e tratamentos que efetuaram. Os profissionais médicos podem consultar esta informação e ainda consultar o motor de inteligência disponível. Na interface são ainda filtrados os pedidos e redirecionados para o componente respetivo e é possível obter uma recomendação de tratamento a prescrever ao paciente, mediante os seus dados.

No sistema de classificação e previsão o profissional tem acesso a uma classificação de um determinado paciente, apresentando informação necessária à recuperação do mesmo. A informação permitirá ao profissional ter uma previsão da recuperação da marcha.

O sistema de recomendação irá auxiliar o profissional no processo de prescrição de tratamentos. Este irá recorrer ao sistema de classificação e previsão de forma a obter resultados com maior precisão. A sugestão é realizada com base num sistema pericial criado com o auxílio de terapeutas que transmitirão o conhecimento necessário para a construção da base de conhecimento. Este sistema é composto por motor de inferência e um módulo de explicações. O motor de inferência aplica as regras previamente criadas para obter uma recomendação. O motor de explicações utiliza o motor de inferência para apresentar ao utilizador o raciocínio que originou a recomendação.

Todo o processo é iniciado pela recolha de dados, como mostrado no diagrama da Figura 7. A recolha é realizada com o auxílio de detetores de atividade muscular e mapeadores de centros

de gravidade. Este ponto é acompanhado de um tratamento de dados, de acordo com as práticas no estudo da atividade humana utilizadas no CEMAH. O tratamento de dados consiste na remoção de ruído das leituras. Ambos os passos são realizados pelo CEMAH.

Através da aplicação de diferentes algoritmos é então possível determinar a existência de dependência funcional entre o membro afetado e o membro simétrico, detecção de padrões e identificação de músculos que apresentam um significado maior no restauro da marcha.

Através dos peritos é possível recolher e sistematizar informação de modo a criar um sistema de recomendação com base em regras.

Por fim, toda a informação gerada é centralizada num sistema de apoio à decisão, que toma a forma de uma plataforma Web.

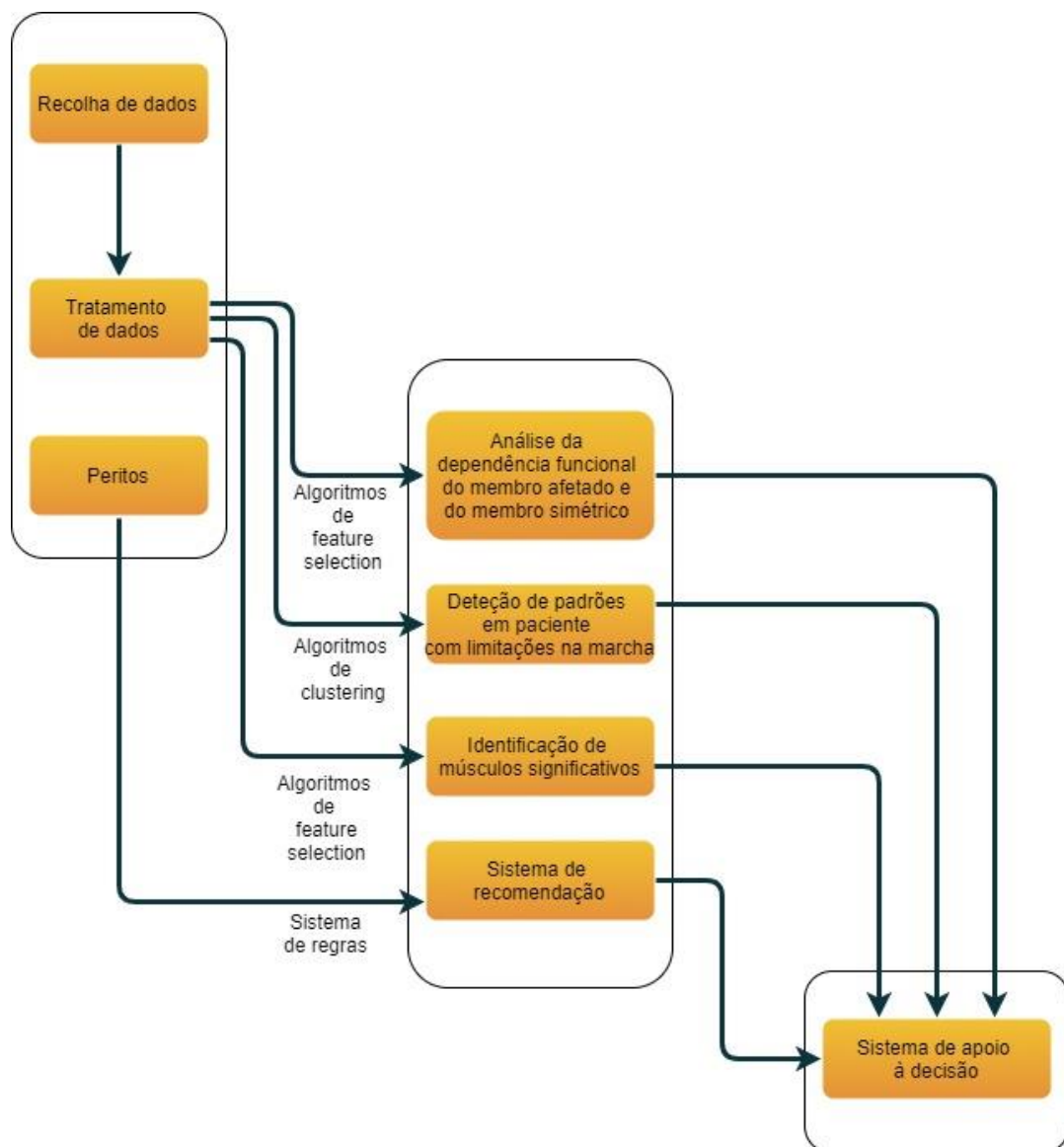


Figura 7 – Diagrama de etapas

A aplicação Web será constituída por um *Front-End* em *AngularJS* e um *Back-End* em C#. O sistema será suportado por uma base de dados em *Structured Query Language (SQL)*. O sistema irá seguir uma arquitetura *Model-View-Controller (MVC)*. O módulo de inteligência será realizado em R, com exceção ao sistema de recomendação que será implementado em C# e Drools, no ambiente de desenvolvimento Visual Studio. Os resultados integrados no *Back-End* da plataforma. Todas as ferramentas são de utilização gratuita.

Análise de dados (R)

Toda a inteligência da aplicação Web e dos recursos utilizados pelos mesmos foram desenvolvidos na linguagem R. Esta linguagem responde a todas as necessidades atuais, não limitando a potencialidade de expansão futura, pela sua flexibilidade e grande comunidade que a mantém atualizada.

Esta ferramenta permite ainda uma fácil integração com o portal Web.

Plataforma Web

A plataforma Web é a aplicação criada e disponibilizada para utilização dos peritos do CEMAH. Nesta estão disponíveis funcionalidades que irão apoiar os profissionais médicos no processo de reabilitação de pacientes de AVC. O portal é constituído por três componentes diferentes:

- **Front-End** – camada de interação com o utilizador. Foi escolhido desenvolver o *Front-End* em *AngularJS* dada a facilidade de implementação do padrão MVC, permitir criar uma interface com design apelativo e de fácil utilização, com uma rápida resposta e um design atual e de fácil compreensão. O padrão MVC foi selecionado pela sua rápida estruturação em pequenos projetos, flexibilidade, encapsulamento e facilidade de apresentação dos modelos (objetos a representar) aos peritos da área;
- **Back-End** – camada de processamento de informação. A linguagem de *Back-End* escolhida foi o C#, na *framework* .NET. Esta linguagem permite uma boa escalabilidade, flexibilidade, utilização de bibliotecas para gestão de processos que não envolvem experiência de utilização. O C# permite ainda uma construção rápida de APIs que comunicam com o *Front-End*, não limitando o uso externo no futuro. Ainda no Back-End é utilizado Drools para a implementação do sistema de recomendação;
- **Base de dados** – onde os dados serão persistidos. Para o portal Web foi criada uma base de dados relacional, em SQL Server. Foi escolhido criar a base de dados neste sistema pela facilidade de crescimento e por ser facilmente apresentada e utilizada por colaboradores com conhecimento reduzido em informática, como é o caso dos investigadores do CEMAH.

A plataforma terá diferentes utilizadores, com módulos disponibilizados mediante o seu grupo: administrador, profissional médico e paciente. A aplicação é então desenhada para disponibilizar as seguintes funcionalidades, como pode ser consultada na Figura 8:

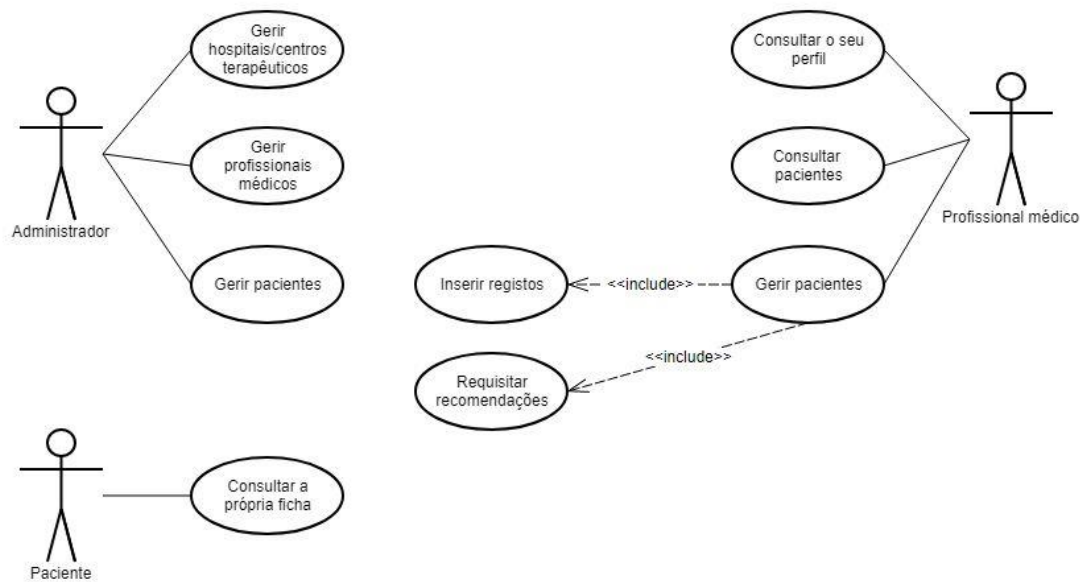


Figura 8 – Diagrama de casos de uso (UML)

- **Administrador do sistema** - Tem acesso à gestão de hospitais/centros terapêuticos, profissionais médicos e pacientes.
- **Profissional médico** - O profissional médico terá acesso a diferentes módulos após entrar no sistema. Tem acesso ao seu registo, onde constam algumas informações gerais, como o número de pacientes que tem em tratamento e os locais onde pratica. Tem acesso ainda ao registo dos seus pares, com informação limitada dos seus pacientes. Por fim, tem acesso à ficha do paciente, onde para além de consultar e registar informação poderá usufruir do módulo de recomendação. É ainda objetivado de ser estruturado o sistema de forma a possibilitar futuramente a gestão documental sobre um paciente, onde irão constar exames e análises pertinentes para o processo em questão.
- **Paciente** - O paciente após se autenticar terá acesso direto à sua ficha de paciente. Nesta ficha terá acesso aos seus dados de utente, informação disponibilizada sobre o AVC (como data de hospitalização e membro afetado) e dados sobre os seus exames e tratamentos.

5.3 Opções de desenvolvimento

Plataforma Web

Foi escolhido o desenvolvimento de uma plataforma WEB para interação com o utilizador. No entanto, a ideia inicial havia sido uma aplicação desktop.

Optou-se por uma plataforma Web pelos seguintes motivos:

- **Não é necessária instalação** – com a plataforma WEB deixa de ser necessário a instalação e até configuração por parte dos utilizadores nas suas máquinas, como com uma aplicação *desktop*. Considerada a funcionalidade dos pacientes visualizarem as suas fichas, este ponto torna-se ainda mais importante.
- **Pode ser visualizada em qualquer dispositivo com um navegador de Internet** – face à aplicação desktop a plataforma WEB pode ser acessível em diversos dispositivos, como telemóveis, tablets e computadores e diferentes sistemas operativos.
- **Disponibilização de uma API para futura integração** – devido à arquitetura utilizada será desenvolvida uma API que irá permitir futuras integrações com, por exemplo, aplicações móveis ou outros sistemas.

Foi colocada ainda a questão da dependência de acesso à Internet para a utilização da aplicação. No entanto, como há um grande foco na visualização e introdução de novos dados, seria sempre necessário o acesso à Internet. É um fator que, no contexto atual, não apresenta ser um obstáculo à maioria dos seus utilizadores.

Módulo de inteligência

O módulo de inteligência foi desenvolvido em R.

Uma opção alternativa teria sido MatLab. Embora o MatLab seja uma linguagem com uma forte afirmação na área de ciência de dados, optou-se pelo R. As principais razões foram as seguintes:

- **O R permite uma maior facilidade de manipulação de dados** – o principal foco do MathLab é a utilização de algoritmos e simuladores. Embora seja possível a manipulação de dados, o R apresenta uma maior versatilidade neste ponto.
- **Custo monetário e comunidade** – O R é uma ferramenta gratuita, suportada pela comunidade que a usa. A sua aderência tem aumentado, liderando como uma das linguagens de programação mais utilizadas. O uso da linguagem MatLab tem um custo elevado associado e é gerido por uma empresa privada. Para além do custo monetário, a comunidade representa um ponto importante dado que existe uma fonte maior de partilha de conhecimento.

Base de dados

Para o desenvolvimento da base de dados foi escolhido uma base de dados relacional. Não foi colocada a opção de um outro esquema, como por exemplo NO-SQL, dado o propósito da base de dados – gestão simples de relativamente poucos dados.

Existem várias opções para a implementação de uma base de dados com este esquema, tal como SQL Server e Oracle. Dada o conhecimento prévio da linguagem e a melhor integração com o sistema *Back-End* da plataforma (.NET), nomeadamente com os módulos desenvolvidos em C# e Drools, foi escolhida a utilização de SQL Server.

6 Avaliação

A qualidade dos resultados é um dos principais fatores na análise de qualquer solução. No entanto, dada a área de utilização do sistema e da natureza e minuciosidade dos dados utilizados, a avaliação da qualidade dos resultados toma um papel muito significativo neste documento.

As avaliações são feitas por módulo, de forma a aferir não só a qualidade do sistema, mas também da qualidade dos seus componentes. Os componentes podem, por sua vez, ser utilizados e estudados como elementos isolados.

Serão utilizadas duas abordagens diferentes à avaliação de todo o projeto:

- **Avaliação com base em métricas dos algoritmos utilizados** – existem formas de avaliação de resultados inerentes aos processos/métodos utilizados para obter os resultados. No projeto serão utilizadas métricas de avaliação desenvolvidas para validação da qualidade na mineração de dados.
- **Avaliação de peritos** – todos os resultados terão de ser apresentados e validados pelos peritos. Esta validação é crucial devido à natureza dos dados e do contexto do problema.

A avaliação foi realizada no fim de cada iteração do desenvolvimento do sistema. Caso os resultados não se apresentem satisfatórios – com base nas métricas ou no perito – haverá a necessidade de voltar a repetir a iteração ou iterações.

6.1 Hipóteses

Existem quatro tópicos abordados na solução e que têm que ser avaliados: identificação de músculos significativos, análise da dependência funcional do membro afetado e do membro simétrico, procura de perfis de pacientes e sistema de recomendação. Para proceder à avaliação dos resultados é necessário primeiramente definir os casos e as hipóteses que são propostas resolver:

- **H0: Identificação de músculos significativos**
H1: (Para um músculo) é representativo
H2: (Para um músculo) não é representativo
- **H0: Análise da dependência funcional do membro afetado e do membro simétrico**
H1: O membro simétrico afeta o resultado da recuperação

H2: O membro simétrico não afeta o resultado da recuperação

- **H0: Procura de perfis de pacientes**

H1: Existem perfis de pacientes

H2: Não existem perfis de pacientes

- **H0: Sistema de recomendação**

H1: Tem uma precisão superior a 85%

H2: Tem uma precisão inferior a 85%

6.2 Metodologias e Grandezas de Avaliação

Como referido, as análises e descobertas de padrões são módulos que irão influenciar a qualidade do conhecimento. Para validar o conhecimento apresentado esta terá que ser testada com grupos de teste de forma a validar a precisão da base do projeto (Tabela 13).

No sistema de recomendação é necessário também analisar a qualidade do conhecimento obtido, sendo utilizado a mesma metodologia de avaliação – grupos de teste. No entanto, a utilização da aplicação tem o intuito de auxiliar o utilizador na tomada de decisão. É importante avaliar a satisfação do mesmo, de forma a melhorar a interação com o sistema.

Tabela 13 – Metodologias de avaliação

Tópico	Metodologia
Análise do membro simétrico	Grupo de teste
Análise dos músculos representativos	Grupo de teste
Descoberta de padrões	Grupo de teste
Sistema de recomendação	Grupo de teste e resultados inquérito de satisfação

6.3 Grandezas de avaliação

As análises e descobertas de padrões são módulos que irão influenciar a qualidade da informação apresentada aos utilizadores e ao CEMAH. Serão então medidas a pureza, a entropia, a soma dos erros ao quadrado (Sum of Squared Error (SSE)) e o coeficiente da silhueta (Silhouette Coefficient) [48] dos resultados obtidos.

O sistema de recomendação é um sistema de apoio à decisão. Para tal, é importante que as recomendações sejam o mais corretas possíveis e, em simultâneo, sejam satisfatórias ao utilizador. A métrica sobre a qual este módulo será analisado é a raiz da média da precisão da recomendação (*mean average precision* [49]). As métricas enumeradas serão aplicadas sobre a aprovação do protocolo de tratamento por parte do utilizador. Este será complementado com a satisfação dos profissionais médicos que irão testar a aplicação. O grau de satisfação será analisado através de um inquérito de satisfação.

A identificação de músculos significativos e a análise da dependência funcional do membro afetado e do membro simétrico são realizados utilizando o mesmo conjunto de ferramentas. Deste modo, as metodologias para avaliar os mesmos serão idênticas. Para a procura de perfis em pacientes e para o sistema de recomendação serão utilizados métodos diferentes.

Para a avaliação dos tópicos foi decidido utilizar as grandezas apresentadas na Tabela 14.

Tabela 14 – Análise de grandezas de avaliação

Tópico	Grandeza
Análise do membro simétrico	Exatidão
Análise dos músculos representativos	Exatidão
Descoberta de padrões	Exatidão
Sistema de recomendação	Exatidão e satisfação do utilizador

Com os métodos referidos de avaliação, é objetivado então uma percentagem de precisão superior a 80% para a análise do membro simétrico e para a análise dos músculos representativos. Na descoberta de padrões pretende-se uma precisão mínima de 75%. O sistema de recomendação é avaliado segundo a precisão das recomendações de tratamento feitas e o tratamento escolhido. A precisão que se procura obter neste módulo é de 85% no final no mínimo de 100 casos de teste.

7 Desenvolvimento

Neste capítulo são apresentados os trabalhos de desenvolvimento propostos e analisados nos capítulos anteriores. O capítulo é dividido em duas secções: análise de dados e sistema de apoio à decisão. O sistema de apoio à decisão é dividido em duas secções: plataforma Web e sistema de recomendação. Cada secção é finalizada com a análise dos resultados obtidos.

A análise de dados permite obter confirmação de algumas suposições dos peritos e ainda descobrir conhecimento até ao momento desconhecido. As análises incidem sobre os músculos representativos da marcha e da trajetória do centro de gravidade durante a prova de levantar-sentar.

O sistema de recomendação irá auxiliar os peritos na prescrição de tratamentos aos pacientes, mediante as regras já utilizadas, de uma forma fácil e acessível. Este sistema terá ainda um módulo de explicações para fundamentar as recomendações feitas.

Por fim, a plataforma Web disponibiliza a informação obtida na análise de dados, o acesso ao sistema de recomendação e ainda os dados dos pacientes num sistema único, acessível a partir de qualquer navegador de Internet.

7.1 Dados

Os dados utilizados para a análise de dados foram fornecidos pelo CEMAH, constando neles valores da atividade eletromiográfica (EMG) do membro direito e esquerdo. A atividade recolhida recai sobre 6 músculos, denominados por Bicipete Femoral (BF), Gastrocnemio Medial (GM), Reto Femoral (RF), Solear (SOL), Tibial Anterior (TA) e Vasto Medial (VM). Existem dados de pacientes que sofreram de AVC e de voluntários sem problemas na marcha. Os dados começaram a ser recolhidos em 2011, e no caso dos pacientes estes foram recolhidos 6 meses após estes sofrerem um AVC.

Os dados foram tratados previamente pelo CEMAH, sendo normalizados e depois aplicado um filtro passa baixo (10Hz) *Butterworth* de 4ª ordem, utilizando um *zero-phase lag*, para a redução do ruído captado.

A recolha destes dados foi realizada com o auxílio do dispositivo médico bioPlux clinical [50].

7.2 Análise de dados

De seguida são apresentadas as experiências e resultados obtidos na análise de dados. Nesta secção o foco não é o de desenvolver qualquer ferramenta, mas de extrair conhecimento relevante dos dados disponibilizados. Esta informação visa servir como ferramenta aos peritos do CEMAH, no processo de recuperação dos pacientes vítimas de AVC.

Os dados utilizados para as análises realizadas foram obtidos por provas realizadas por vítimas de AVC e por voluntários supervisionados pelos peritos do CEMAH. Por cada prova realizada é gerado um ficheiro de texto com um volume de registos muito elevado – uma média de 5 MB por ficheiro, perfazendo cerca de 2 GB de informação no total. Para ser possível processar os dados foi necessário recorrer a uma máquina com características adaptadas. A ficha técnica da máquina utilizada (Figura 9 – Características da máquina utilizada para a análise de dados) incluem um sistema operativo de 64 bits e uma capacidade elevada de memória RAM. Estas duas configurações tornaram possível a execução de diversos algoritmos com o volume de dados existente em R.

Sistema	
Processador:	Intel(R) Xeon(R) CPU X5550 @ 2.67GHz 2.67 GHz (2 processadores)
Memória instalada (RAM):	48,0 GB
Tipo de sistema:	Sistema Operativo de 64 bits, processador baseado em x64
Caneta e Toque:	Não está disponível Introdução por Caneta ou Toque para este Ecrã

Figura 9 – Características da máquina utilizada para a análise de dados

7.2.1 Identificação de variáveis relevantes

Um dos primeiros pedidos dos peritos foi a deteção, caso existam, dos músculos que apresentam uma influência maior na análise da marcha. Para tal, foram aplicadas técnicas de identificação de variáveis relevantes sobre os dados obtidos em testes elaborados à marcha de pacientes saudáveis e de pacientes que haviam sofrido de AVC. Os resultados obtidos podem influenciar futuros exames com a exclusão dos músculos menos representativos e a inclusão de outros não considerados até então.

7.2.1.1 Extract Transform Load (ETL)

O processo de ETL [51] é o processo no qual os dados fornecidos são lidos, transformados e exportados para poderem ser analisados. A fase de leitura dos dados permite a leitura de diferentes fontes e ainda escolher apenas os dados relevantes às análises a realizar. A transformação dos dados permite que os dados já lidos possam ser trabalhados aplicando regras, combinações e outras técnicas de alteração e eliminação de dados. Por fim, a exportação disponibiliza os dados lidos e transformados num único ficheiro, base de dados ou outro formato necessário.

Como indicado, os resultados das provas são exportados em ficheiros de texto. Estes ficheiros possuem informação que não é relevante para a análise de dados, como os canais e frequência da aquisição de dados, a data de aquisição e a identificação do dispositivo.

Numa primeira fase, o processamento de dados passa pela leitura e remoção do cabeçalho de todas as provas de marcha monitorizada (análise de sinais electromiográficos - EMG) (Figura 10 – Exemplo de ficheiro resultante da prova de marcha monitorizada) (Figura 11 – Prova de marcha monitorizada, disponibilizada pelo CEMAH).

```
# bioPlux Text File Format
# Version: 1
# StartDateTime: 2011-06-14 10:21:33
# SamplingFrequency: 1000
# SampledChannels: 1 2 3 4 5 6 7 8
# SamplingResolution: 12
# AcquiringDevice: 00:07:80:93:4D:FF
# EndOfHeader
0      0      0      2056    2056    2063    2055    2057    2053    0      0
1      0      0      2053    2055    2058    2054    2054    2049    0      0
2      0      0      2050    2052    2047    2053    2053    2050    0      0
3      0      0      2053    2054    2047    2052    2054    2053    0      0
4      0      0      2055    2052    2052    2054    2054    2054    0      0
5      0      0      2052    2049    2043    2051    2052    2053    0      0
6      0      0      2056    2052    2040    2052    2053    2054    0      0
7      0      0      2054    2054    2047    2057    2056    2057    0      0
8      0      0      2050    2052    2038    2051    2052    2055    0      0
9      0      0      2054    2055    2036    2054    2050    2056    0      0
10     0      0      2052    2052    2039    2054    2056    2055    0      0
```

Figura 10 – Exemplo de ficheiro resultante da prova de marcha monitorizada



Figura 11 – Prova de marcha monitorizada, disponibilizada pelo CEMAH

As colunas têm o significado que se segue:

- Coluna 1: número de sequência
- Coluna 2: valor binário da porta digital de entrada
- Coluna 3: valor binário da porta digital de saída
- Coluna 4: sinal electromiográfico do músculo RF

- Coluna 5: sinal electromiográfico do músculo GM
- Coluna 6: sinal electromiográfico do músculo TA
- Coluna 7: sinal electromiográfico do músculo SOL
- Coluna 8: sinal electromiográfico do músculo BF
- Coluna 9: sinal electromiográfico do músculo VM.

Em cada prova é ainda eliminada a primeira coluna, que não apresenta significado. De seguida, são removidas todas as linhas até à deteção do valor binário de entrada que indica que a prova foi iniciada. Posteriormente, as colunas 2 e 3 também são removidas. Por fim, todos os dados são agregados num único ficheiro. Este ficheiro base será utilizado em outras fases de ETL de cada análise.

7.2.1.2 Random forests

Como indicado na secção 4.1, o algoritmo utilizado para a identificação dos músculos mais relevantes foi Embedded Feature-Selection. A implementação utilizada nesta análise foi a random forests, disponibilizada em R por Breiman and Cutler's Random Forests for Classification and Regression [52] [53]. Na utilização do algoritmo foram ajustados os seguintes parâmetros [54]:

- `x`, formula – Representa a formula com que se vai fazer a identificação. Para tal, foram os músculos em análise.
- `data` – É o conjunto de dados a processar, onde se encontram as variáveis em análise na formula. Na aplicação foi fornecido o ficheiro processado com os registos das provas.
- `ntree` – São o número de arvores a criar. Valor colocado de 1500.
- `mtry` – Número de variáveis escolhidas aleatoriamente como candidatos em cada divisão. Como estão a ser analisadas 6 variáveis o valor por defeito é raiz de 6. Nas experiências realizadas o valor foi sendo ajustado entre raiz de 6 / 2 até 3 x raiz de 6, em incrementos de raiz de 6/2.
- `nodesize` – Número de de terminais em cada árvore. Valor colocado a 1
- `importance` – Se a importância dos precedores deve de ser calculada. Colocado a verdadeiro

7.2.1.3 Músculos do membro afetado

Nesta etapa inicial é realizada uma análise dos músculos da perna afetada tendo por base de comparação voluntários saudáveis. É objetivado conseguir obter dados sobre quais os músculos que representam um maior impacto no sucesso da marcha.

7.2.1.3.1 ETL

O processo de ETL aplicado a esta fase utiliza o ficheiro de dados base já gerado. Foram tidos em conta os 6 músculos (RF, GM, TA, SOL, BF e VM) da prova de marcha monitorizada (EMG) e se o examinado era um paciente. Para os pacientes foram analisados os dados obtidos sobre a

perna afetada. Para os voluntários foram utilizados dados obtidos sobre uma das pernas, de forma a manter a proporção de membros inferiores direitos e esquerdos dos pacientes e dos voluntários. Para esta fase estão disponíveis 80 provas, com cerca de 6000000 registos.

Numa última fase é realizada uma normalização dos dados através da fórmula seguinte:

$$\frac{\text{Valor do Músculo} - 2048}{4096} * 5$$

Esta normalização faz parte de um processo aplicado pelo CEMAH e que vai de encontro aos valores aceites pela comunidade académica da área.

7.2.1.3.2 Interpretação de resultados

Os resultados obtidos permitem verificar a relevância de cada um dos músculos e a matriz de confusão da classificação (Figura 12 – Matriz de confusão). Com base nesta informação é possível calcular a percentagem de precisão obtida. Nesta análise, o valor obtido foi de 85%. Este valor satisfaz o valor mínimo de 80% de certeza requerido.

```

OOB estimate of error rate: 8.88%
Confusion matrix:
      0      1 class.error
0 1194665 116505 0.08885575
1 118495 1217607 0.08868709
    
```

Figura 12 – Matriz de confusão do processo de identificação de variáveis relevantes na prova EMG

Perante os dados apresentados na Tabela 15 – Relevância dos músculos para a realização da marcha, os músculos que possuem maior relevância são o VM, TA e BF. O menos representativo é o GM, que detém menos de metade da relevância dos músculos enumerados.

Tabela 15 – Relevância dos músculos para a realização da marcha

Músculo	Relevância (Accuracy)
VM	24
TA	17
BF	13
RF	9
SOL	7
GM	5

Perante os resultados, a opinião dos peritos do CEMAH valida alguns dos valores obtidos, especialmente nos músculos com maior relevância. Esta opinião serve para comprovar os resultados obtidos.

7.2.1.3.3 Estado de ativação dos músculos da perna afetada

Embora os resultados obtidos já respondam ao pedido pelos peritos, a análise realizada pode ser melhorada. Para além de apresentar tempos de execução altos também não indica a razão pela qual os músculos têm mais ou menos importância numa marcha normal. É então proposta uma nova análise que se foca se o músculo está ativo ou inativo em vez da força por este exercida. Cada registo irá então representar o estado de cada músculo num determinado momento da prova. É assim objetivado a obtenção de resultados mais precisos e de mais fácil explicação.

7.2.1.3.3.1 ETL

Para a identificação deste evento, é necessário medir a atividade desse músculo quando ele não está "ativado" (denominado por atividade basal), obtendo um valor médio e o desvio padrão. Considera-se o músculo ativo quando a sua atividade ultrapassa o valor médio basal mais de 2 desvios padrão.

O processo de ETL aplicado nesta fase é então uma extensão do referido em ETL, substituindo apenas o valor da força exercida pelo músculo pela sua ativação.

7.2.1.3.3.2 Interpretação dos resultados

Embora o estado de ativação de um músculo seja um valor de referência para a prova em análise, este estado provou não ser relevante dado que o erro aumentou de 8,9% para 25%. O erro obtido é muito superior ao valor objetivado e mostrando-se menos relevante que a análise anterior.

Tabela 16 – Relevância da ativação dos músculos para a realização da marcha

Músculo	Relevância (Accuracy)
BF	5.65
RF	4.98
VM	2.55
TA	1.77
GM	1.71
SOL	1.27

7.2.1.3.4 Resumo da prova (perna afetada)

Sendo os resultados atingidos na análise do estado dos músculos ainda não satisfatórios, foram investigados outros valores calculados que representassem os dados e que fossem passíveis de serem interpretados. Com a ajuda dos peritos foi possível resumir a atividade de um músculo no decorrer da prova em três variáveis: número de ativações, média de tempo ativo e média do seu valor enquanto ativo.

7.2.1.3.4.1 ETL

O processo de ETL é iniciado com os dados resultantes de ETL. Após esse tratamento inicial foi possível resumir a prova obtendo as 3 métricas referidas por músculo, calculados com base no estado do músculo, referenciado no ponto ETL Estado de ativação dos músculos da perna afetada.

O volume de dados é então reduzido drasticamente, permitindo uma exploração mais rápida e de verificação mais fácil pelos peritos.

7.2.1.3.4.2 Interpretação dos resultados

Os resultados obtidos apresentam um excelente valor de precisão – 92%. Estes valores são equivalentes aos valores obtidos na primeira prova e que já respondiam ao requerido. Com as métricas aplicadas é possível realizar a análise numa fração do tempo anterior e com resultados com uma interpretação mais acessível.

Como é possível verificar, os dados são concordantes com a primeira análise. Perante os peritos também foi possível validar que os valores obtidos são relevantes e que comprovam algumas das expectativas que existiam (Figura 13 – Precisão de cada variável no processo de identificação de variáveis relevantes na prova EMG). Os parâmetros mais importantes são então sobre os músculos VM, TA e BF, em específico o número de ativações e o valor médio de ativação. Estes dados permitem ainda validar a fraca relevância que o estado do músculo representa.

	MeanDecreaseAccuracy
RF_nAtivacoes	13.484200
GM_nAtivacoes	10.882843
TA_nAtivacoes	25.669422
SOL_nAtivacoes	9.832365
BF_nAtivacoes	22.167521
VM_nAtivacoes	31.653617
RF_pAtivacao	8.070488
GM_pAtivacao	14.623064
TA_pAtivacao	17.654225
SOL_pAtivacao	6.669429
BF_pAtivacao	16.092656
VM_pAtivacao	12.436032
RF_mAtivacao	17.941179
GM_mAtivacao	9.830063
TA_mAtivacao	25.135181
SOL_mAtivacao	10.336663
BF_mAtivacao	11.501974
VM_mAtivacao	26.611618

Figura 13 – Precisão de cada variável no processo de identificação de variáveis relevantes na prova EMG

Os dados obtidos nesta são então ideais para futuras análises pelos peritos, baseado na sua precisão, facilidade de interpretação e replicação futura sobre novas provas executadas.

7.2.1.4 Análise do impacto da perna não-afetada na marcha

O CEMAH, através dos exames efetuados ao longo do tempo, detetou a possibilidade da atividade dos músculos da perna com menos limitações afetar a marcha. Para averiguar esta possibilidade será repetido o processo de *feature-selection* nos dados das provas de marcha de ambas as pernas. Desta forma torna-se possível averiguar, entre a totalidade dos músculos dos membros inferiores avaliados, os músculos mais relevantes.

7.2.1.4.1 ETL

Como os dados continuam a ser os mesmos aplicados no estudo dos Músculos do membro afetado, o processo de ETL é semelhante ao já aplicado, com a única exceção da aplicação não a apenas ao membro afetado mas a ambos os membros inferiores.

7.2.1.4.2 Interpretação dos resultados

Os resultados obtidos apresentam uma percentagem de precisão acima dos 80% requeridos, com um resultado ligeiramente superior obtido na primeira análise feita em Músculos do membro afetado (obtida uma precisão de 92%). Os resultados suportam ainda a importância dos músculos VM e TA da perna afetada, mas apontando também para os músculos TA, VM e BF da perna simétrica (Tabela 17). Mediante esta informação é possível concluir que as suspeitas levantadas pelo CEMAH de facto se comprovam. Esta descoberta permite um estudo mais detalhado desta prova, quer pela inclusão dos músculos da perna simétrica quer pela exclusão dos músculos menos representativos. Abre ainda a possibilidade ao estudo de outros músculos não presentes na atual prova.

Tabela 17 – Variáveis mais representativas da marcha de ambos os membros

Membro/Músculo	Relevância (Accuracy)
Afetado/VM	21
Afetado/TA	18
Afetado/BF	15
Oposto/VM	14
Oposto/TA	12
Oposto/BF	9
Afetado/RF	7
Afetado/GM	6
Oposto/RF	5
Oposto/SOL	5
Afetado/SOL	4
Oposto/GM	3

7.2.1.4.3 Estado de ativação

Os resultados continuam a ser pouco representativos, dando mais certeza à possibilidade já levantada de que o estado de ativação de um músculo não é representativo da prestação dos músculos durante a prova de marcha monitorizada. Desta forma, os resultados obtidos nesta análise não serão alvo de futuros estudos.

7.2.1.4.4 Resumo da prova

Os resultados obtidos possuem uma percentagem de erro de 10,1% que, embora menos precisos que os atingidos na fase Resumo da prova (perna afetada), são aceitáveis a nível da precisão matemática objetivada. Os dados confirmam mais uma vez as suspeitas do CEMAH sobre a influência do membro simétrico no desempenho durante a marcha. Os resultados foram ainda apresentados e aprovados pelos peritos, sendo agora corroborados pelas análises realizadas neste projeto. Na Tabela 18 são apresentadas as 10 variáveis mais representativas da marcha, por ordem de relevância.

Tabela 18 – 10 variáveis mais representativas da marcha

Membro/Variável
Afetado/Número de ativações do VM
Afetado/Número de ativações do BF
Afetado/Número de ativações do TA
Oposto/Período médio de ativação do VM
Afetado/Período médio de ativação do VM
Afetado/Período médio de ativação do TA
Oposto /Período médio de ativação do TA
Oposto /Número de ativações do VM
Oposto/Número de ativações do BF
Afetado/Período médio de ativação do BF

Dada que a informação obtida nesta fase é a mais completa (dado que envolve ambos os membros) e possui uma compreensão fácil, vai ser utilizada nas análises realizadas no seguimento deste documento. Este resumo da prova da marcha monitorizada representa ainda uma vantagem a nível de volume de dados em análise, permitindo obter resultados mais rapidamente, por consequente, permitindo ajustes mais rápidos aos dados e aos algoritmos utilizados.

7.2.2 Prova de levantar-sentar - forças de reação do solo (FRS)

Na identificação de variáveis relevantes foram analisadas as provas de marcha monitorizada. No entanto, existe ao nosso dispor dados de uma outra prova realizada pelos peritos do CEMAH – prova de levantar e sentar. Numa fase ainda exploratória, houve um processo de ETL e uma exploração de dados. Posteriormente as provas servem para analisar padrões nesta prova.

7.2.2.1 ETL

Como nas provas da atividade electromiográfica previamente analisados, os resultados são exportados em ficheiros de texto. Existem 13 colunas com seguinte significado:

- Coluna 1: Fx_ plataforma 10
- Coluna 2: Fy_ plataforma 10
- Coluna 3: Fz_ plataforma 10
- Coluna 4: Mx_ plataforma 10
- Coluna 5: My_ plataforma 10
- Coluna 6: Mz_ plataforma 10
- Coluna 7: Fx_ plataforma 8
- Coluna 8: Fy_ plataforma 8
- Coluna 9: Fz_ plataforma 8
- Coluna 10: Mx_ plataforma 8
- Coluna 11: My_ plataforma 8
- Coluna 12: Mz_ plataforma 8
- Coluna 13: Trigger (nos ficheiros da FRS assume um valor médio de 369 quando está ativo e aproximadamente 0 quando está inativo)

As plataformas 8 e 10 são as plataformas de recolha de dados, sobre as quais os pacientes realizaram as provas.

Após a leitura dos ficheiros é necessário remover todos os registos até obter-se um valor positivo na coluna 13 – Trigger. Embora existam também duas plataformas só é analisada uma em cada prova, levando à exclusão de valores obtidos na plataforma não utilizada. Tal como na prova anterior, todos os dados são agregados num único ficheiro que será utilizado nas análises realizadas.

7.2.2.2 Identificação de variáveis relevantes

Numa fase exploratória foi decidido analisar a possibilidade de, à semelhança dos músculos, existirem valores que seriam representativos da prova. Para tal, foram aplicadas técnicas de *feature-selection*, através do algoritmo *random forest* utilizado anteriormente.

O resultado obtido foi de baixa precisão, com cerca de 65%. Após consultar os peritos foi comprovado que os resultados no formato em que são obtidos não fornecem informação suficiente para análise, já que a precisão obtida é insatisfatória. É então necessário processar os dados novamente para análise, com intuito de observar a evolução das variáveis ao longo do tempo. Como a análise tem que ser feita ao longo do tempo, justifica novamente a fraca precisão obtida, dado que *feature-selection* incide sobre a análise de cada registo em vez da progressão no tempo.

7.2.3 Clustering

Um ramo pouco explorado pelos peritos do CEMAH em pacientes que sofreram de AVC é a procura e identificação de padrões, quer nas provas quer nos dados dos próprios pacientes. No entanto, foi expressado um interesse na procura desta nova informação. Para tal foram aplicados métodos de *clustering* nos dados com o objetivo de encontrar padrões nos pacientes em ambas as provas. Caso sejam encontrados padrões que sejam relevantes e validados pelos peritos poderá servir de base a trabalho futuro onde será possível interligar esta informação com, por exemplo, o sistema de recomendação de tratamentos.

7.2.3.1 Marcha monitorizada - sinais electromiográficos

A primeira procura de padrões é então na marcha monitorizada, onde serão analisados os sinais electromiográficos dos 6 músculos já enumerados, assim como também na idade, o género, a altura e o peso das pacientes. Como foi verificado na secção Identificação de variáveis relevantes, não só os músculos da perna afetada mas também os da sua oposta afetam o desempenho da marcha. Desta forma serão analisados todos os músculos (12 músculos, 6 de cada perna).

7.2.3.1.1 ETL

Como serão analisados ambos os membros inferiores e os seus respetivos músculos, o processo de ETL parte da base construída no ponto Análise do impacto da perna não-afetada na marcha. Serão utilizados os dados das provas (dado que obtiveram o melhor resultado no processo de identificação de variáveis relevantes) e os seus resumos (que, à semelhança, também provaram serem representativas das mesmas).

A única adição a estes dados são algumas informações do paciente, que até agora não eram necessárias para realizar as análises: idade, género, altura e peso. Esta informação estava presente num ficheiro Excel fornecido separadamente e que foi, nesta fase, integrado com os dados existentes.

São então gerados dois ficheiros com toda a informação – um com os registos da prova e os dados do paciente e outro com os resumos da prova e os dados dos pacientes.

7.2.3.1.2 K-Means

Conforme o estudo feito na seção do Estado da arte e Avaliação de soluções, o algoritmo utilizado para realizar o *clustering* dos dados é o K-Means. A implementação disponibilizada em R [55] utiliza possibilita a utilização de vários parâmetros, dos quais os seguintes foram ajustados:

- x – Variáveis em análise (ficheiro produzido das provas realizadas)
- centres – Número de *clusters*

- `nstart` – Número de conjuntos aleatórios de inicialização dos centros - colocadoS 1500

Após aplicar o algoritmo é obtido um objeto com as seguintes características:

- `cluster` - Um vetor de inteiros (de 1: k) indicando o cluster ao qual cada ponto é alocado.
- `centers` - Matrix dos centros do *clusters*.
- `totss` - A soma total dos quadrados.
- `withinss` - Vector de soma dos quadrados dentro do cluster, um componente por cluster.
- `tot.withinss` - Total da soma dos quadrados dentro do cluster, isto é, soma (`withinss`).
- `betweenss` - A soma dos quadrados de quadrados, ou seja, `totss-tot.withinss`.
- `size` – Número de pontos em cada cluster.
- `iter` - O número de iterações (externas).
- `ifault` - integer: indicador de possíveis falhas na aplicação do algoritmo.

Para obter um número de clusters ideal é desenhado um gráfico com os valores retornados de *withinss* com diferentes clusters (Figura 14 – Código extraído do RStudio no processo de clustering). Para o efeito, estudou-se o mínimo de 1 cluster e no máximo de 20. O efeito é uma evolução dos valores por número de clusters (Figura 15 – Desenho do valor de *withinss* por número de clusters). Numa situação ideal, o desenho descreveria uma progressão constante onde seria possível detetar uma angulação no gráfico [56] que, visualmente, se parece com um cotovelo. Os valores indentados serão o número de clusters com melhor probabilidade de obter os clusters que melhor caracterizam os dados.

```
wss <- (nrow(data)-1)*sum(apply(data,2,var))
for (i in 2:20) wss[i] <- sum(kmeans(data, centers=i)$withinss)
plot(1:20, wss, type="b", xlab="Number of Clusters", ylab="within groups sum of squares", main="Scree plot")
```

Figura 14 – Código extraído do RStudio no processo de *clustering*

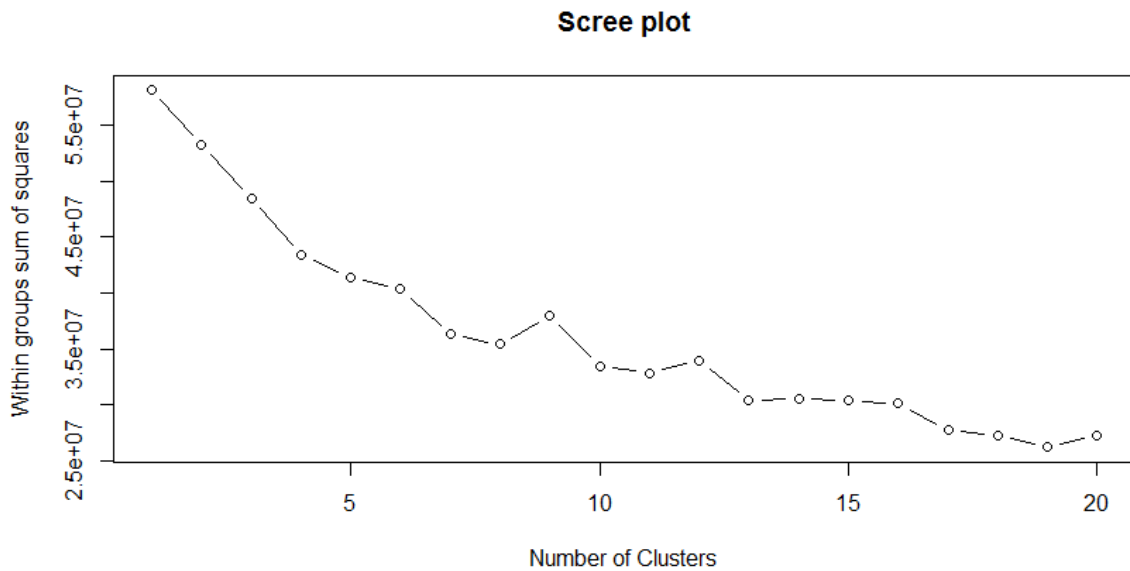


Figura 15 – Desenho do valor de *withinss* por número de *clusters*, relativos aos dados da prova e do paciente

7.2.3.1.3 Clustering

Para a aplicação do algoritmo K-Means é preciso indicar o valor de clusters que se procura. Para tal, como referido no ponto K-Means, é procurado o número de clusters possivelmente mais indicado (que produza o menor erro) pelo desenho do valor de *withinss* desde de 1 cluster até 20 clusters. Nesta fase é obtida a Figura 15 – Desenho do valor de *withinss* por número de clusters, relativos aos dados da prova e do paciente que mostra que não existe uma progressão ideal e que existem diversas opções de número de clusters a experimentar (representados pelos “vales” na progressão). No caso específico, os valores 4, 7 e 10 destacam-se. É então feito *clustering* aos dados com as três hipóteses e desenhado novamente um gráfico ilustrativo de como os dados se distribuem e os clusters a que pertencem (Figura 16 – Clustering dos dados do paciente e da prova EMG, com 4 clusters, Figura 17 – Clustering dos dados do paciente e da prova EMG, com 7 clusters, Figura 18 – Clustering dos dados do paciente e da prova EMG, com 10 clusters). Devido ao número elevado de registos e à capacidade da máquina para criar o gráfico, nas figuras referidas estão apenas exibidos pontos representativos dos centros dos clusters.

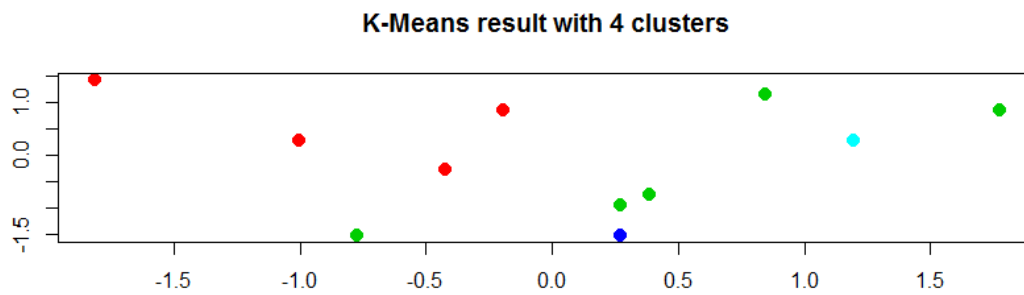


Figura 16 – Clustering dos dados do paciente e da prova EMG, com 4 clusters

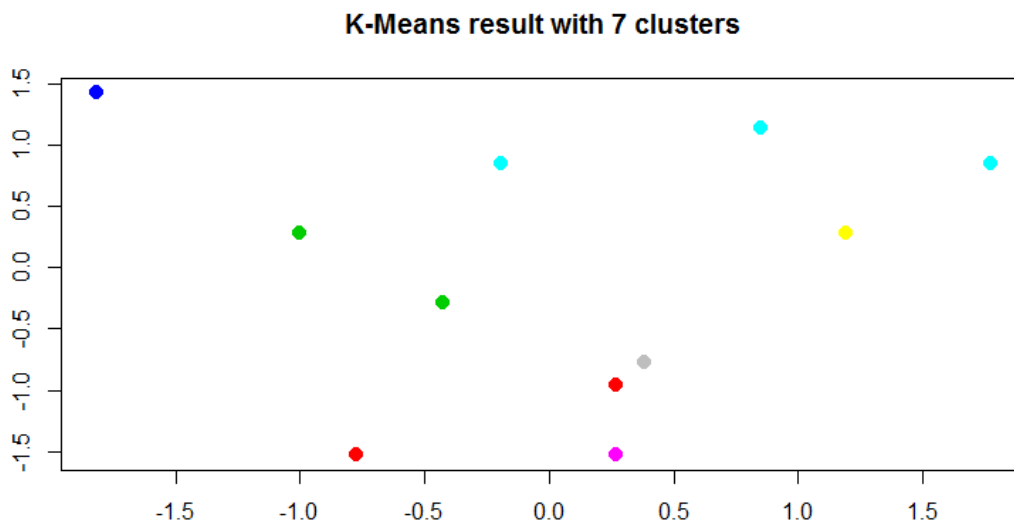


Figura 17 – Clustering dos dados do paciente e da prova EMG, com 7 clusters

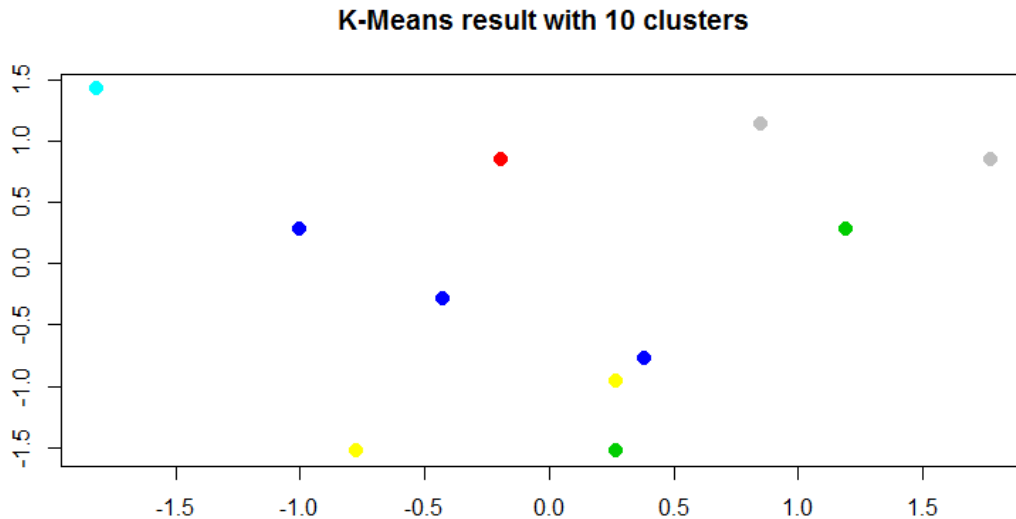


Figura 18 – Clustering dos dados do paciente e da prova EMG, com 10 clusters

Utilizando os resultados gráficos é possível verificar que a opção de 7 clusters apresenta clusters melhores definidos e menos dispersos.

No entanto, ao avaliar os resultados podemos verificar que perante os dados altura, idade, peso e sexo a prova apresenta pouco peso na avaliação final. Para analisar os resultados é utilizada a biblioteca *rpart* (Recursive Partitioning and Regression Trees) [57]. Esta biblioteca permite desenhar as árvores de decisão e descrever cada ramo. A árvore de decisão indica apenas o músculo TA da perna não-afetada como um possível classificador, no conjunto de dados obtidos durante os exames (Figura 19 – Árvore de decisão dos clusters encontrados para os dados do paciente e da prova EMG).

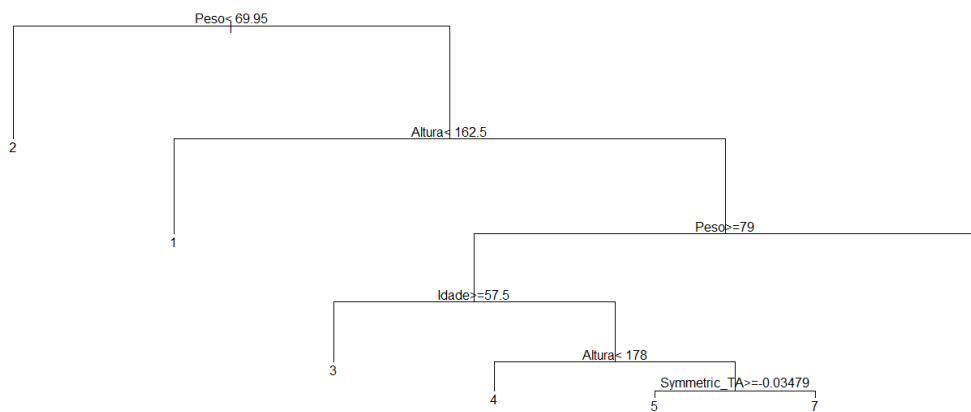


Figura 19 – Árvore de decisão dos clusters encontrados para os dados do paciente e da prova EMG

A biblioteca ainda permite verificar as variáveis mais relevantes para a definição dos clusters. Os resultados não refletem então as provas, mas sim o paciente (Figura 20 – Importância das variáveis dos clusters encontrados para os dados do paciente e da prova EMG).

```
variable importance
  Peso      Altura      Idade      Sexo
    29       22       17       12
Affected_RF Affected_VM Affected_GM Symmetric_TA
    9         7         2         1
Affected_TA
    1
```

Figura 20 – Importância das variáveis dos clusters encontrados para os dados do paciente e da prova EMG

Embora esta seja informação relevante já existem estudos mais aprofundados e com mais dados que os peritos podem consultar, incluindo mais pacientes e mais variáveis que os caracterizam. Os clusters encontrados têm um nível de precisão de 72% (Figura 21 – Erro relativo por tamanho da árvore dos clusters encontrados para os dados do paciente e da prova EMG) mas dado que os resultados incidem maioritariamente sobre os 4 dados do paciente, e não sobre a prova, o resultado seria mais relevante se apresentasse uma precisão bastante superior. Não excluindo os resultados obtidos, a exploração dos padrões continua com os resumos das provas com o objetivo de procurar padrões mais baseados nas provas dos pacientes.

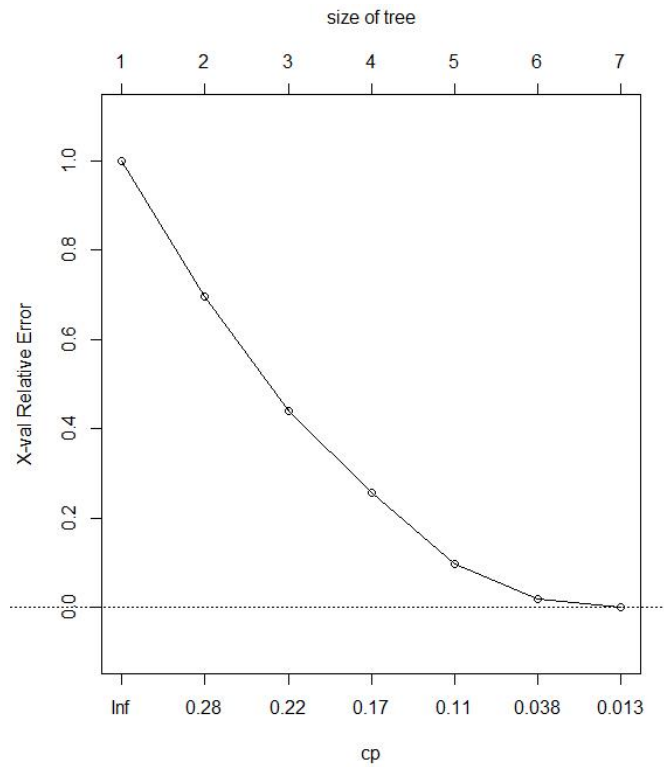


Figura 21 – Erro relativo por tamanho da árvore dos clusters encontrados para os dados do paciente e da prova EMG

7.2.3.1.4 Clustering do resumo das provas

Aplicando a técnica descrita na experiência anterior, foram analisados os possíveis clusters com os dados do paciente e os resumos das provas.

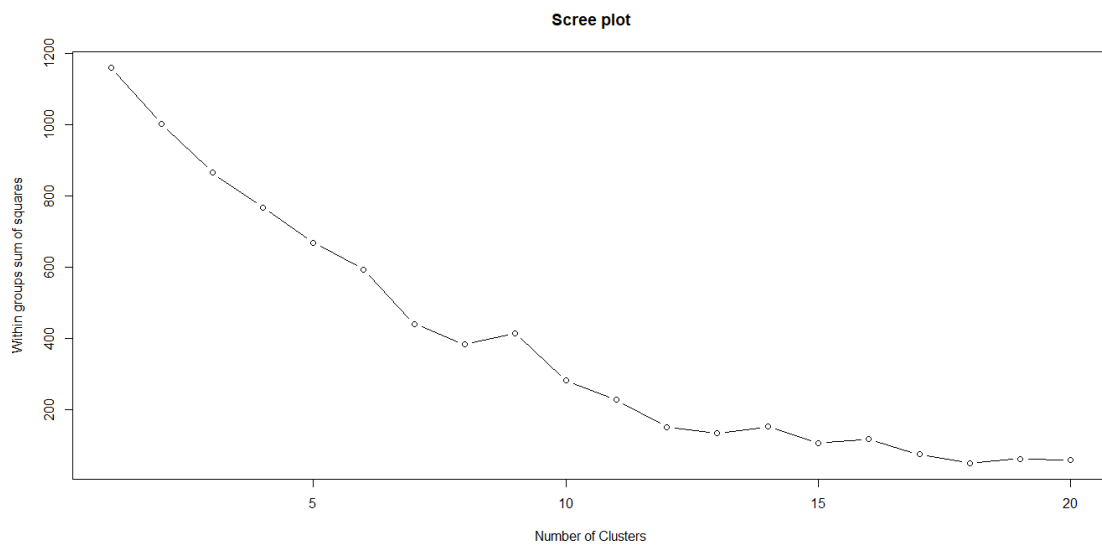


Figura 22 - Desenho do valor de *withinss* por número de *clusters*, relativos aos dados do paciente e o resumo das provas

Pelo gráfico obtido é visível uma progressão mais de acordo com o pretendido. No entanto ainda existem diversas possibilidades a estudar segundo o método do cotovelo, sendo elas 3, 5, 7 e 8. Nas figuras Figura 23, Figura 24, Figura 25 e Figura 26 é possível o resultado da aplicação do algoritmo *k-means* nos dados.

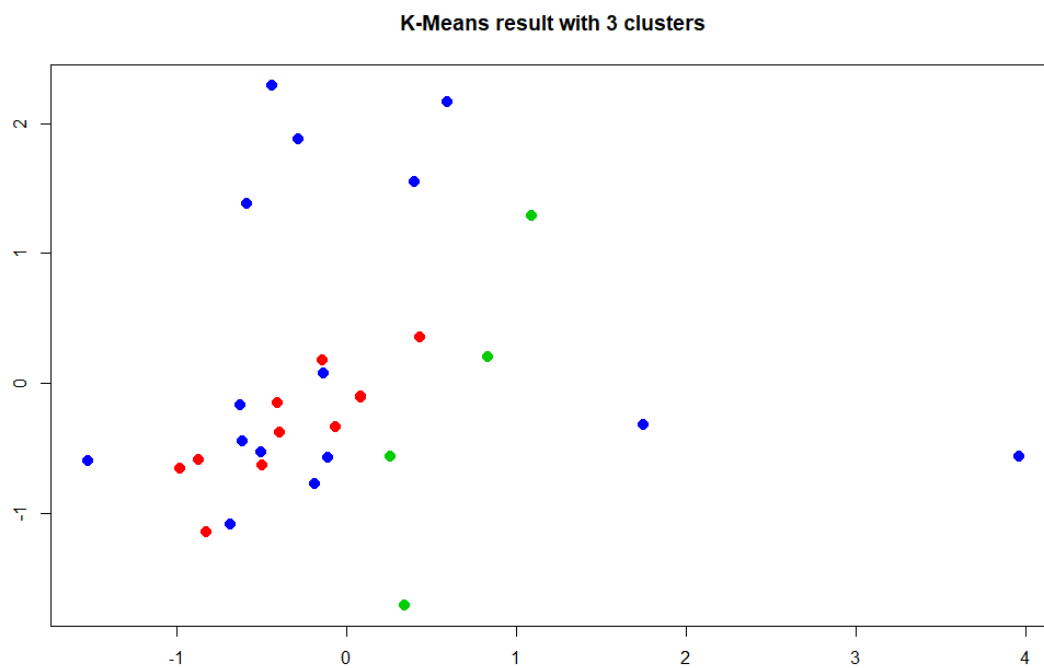


Figura 23 – Clustering dos dados do paciente e do resumo da prova EMG, com 3 clusters

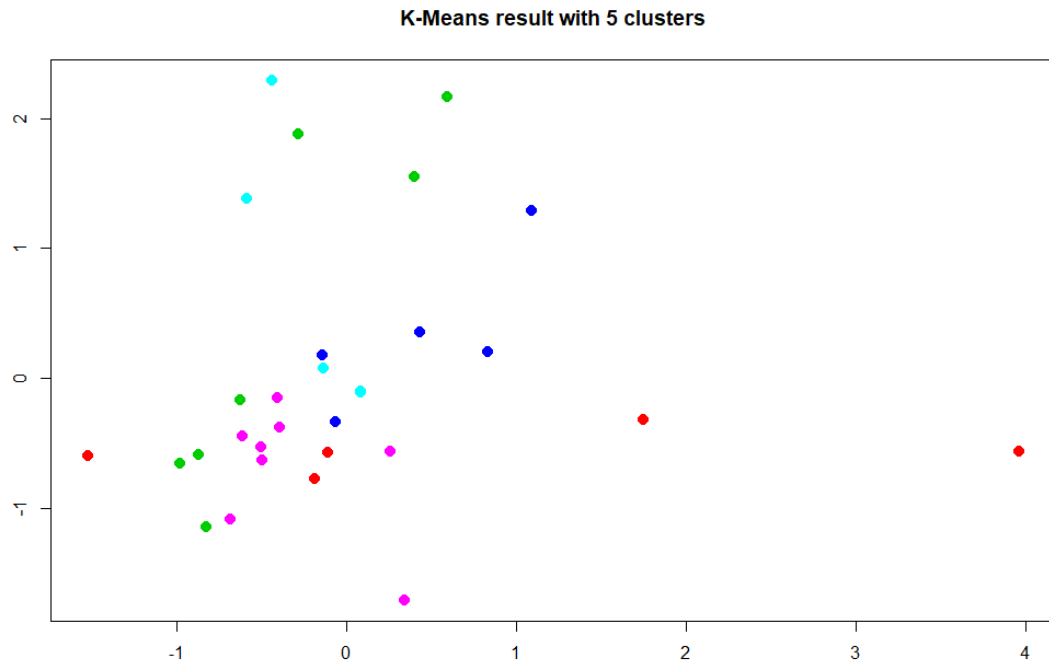


Figura 24 – Clustering dos dados do paciente e do resumo da prova EMG, com 5 clusters

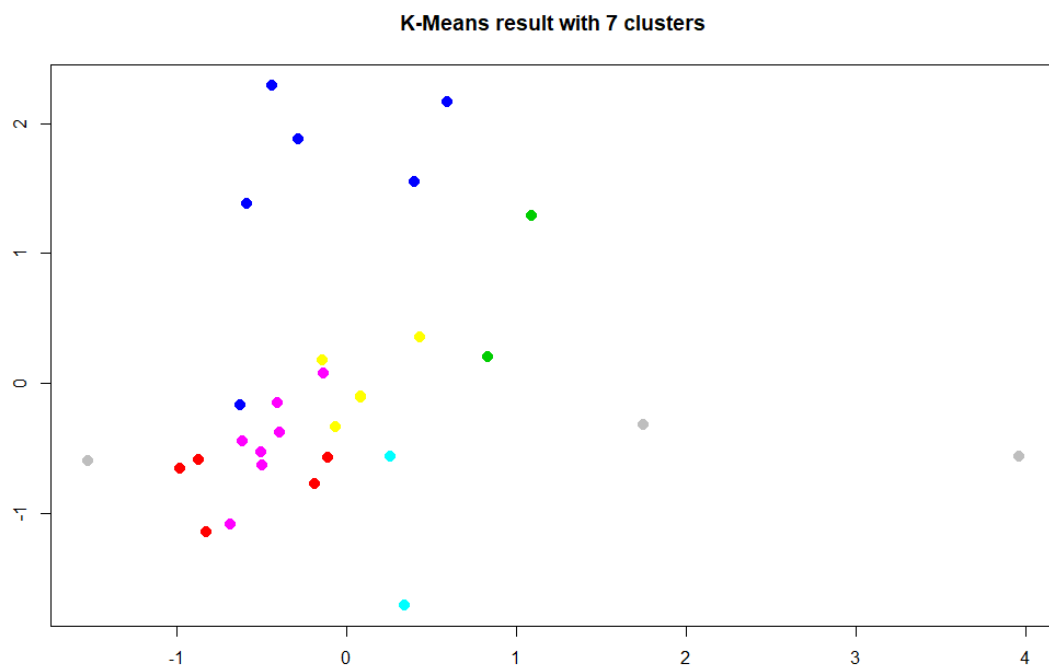


Figura 25 – Clustering dos dados do paciente e do resumo da prova EMG, com 7 clusters

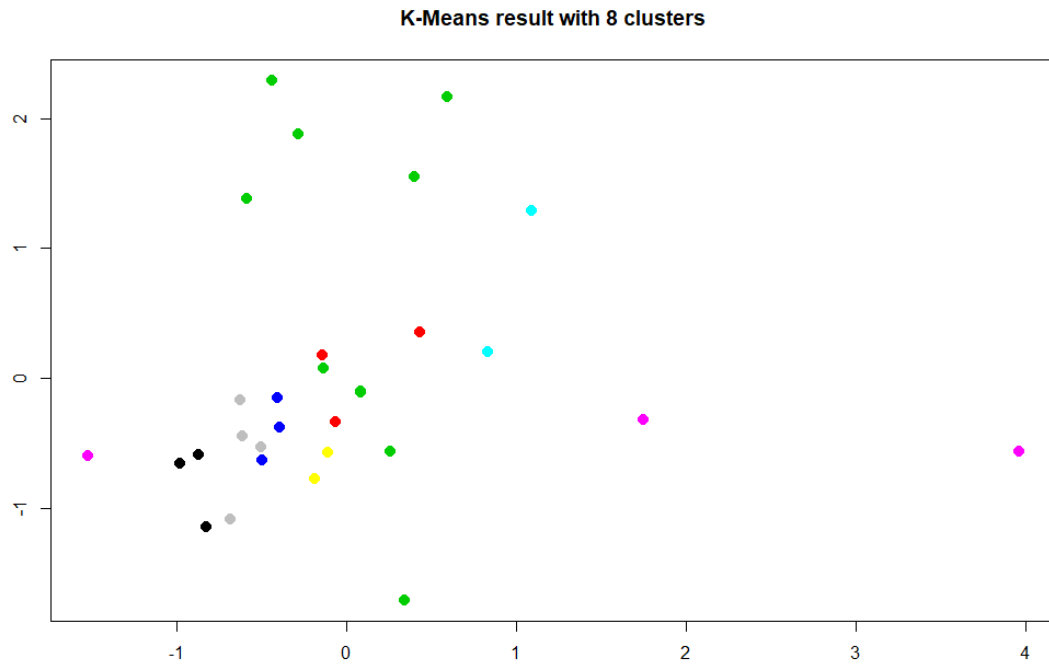


Figura 26 – Clustering dos dados do paciente e do resumo da prova EMG, com 8 clusters

O gráfico com os clusters melhor definidos foi criado com 7 clusters. Ao contrário da análise realizada com as provas, o *clustering* com o resumo das provas permite classificar os pacientes mediante os seus dados e a sua prestação na prova de marcha monitorizada (Figura 27 – Árvore de decisão do clustering dos dados do paciente e do resumo da prova EMG).

```

Node number 1: 30 observations, complexity param=0.35
predicted class=3 expected loss=0.6666667 P(node) =1
class counts: 3 3 10 7 3 2 2
probabilities: 0.100 0.100 0.333 0.233 0.100 0.067 0.067
left son=2 (14 obs) right son=3 (16 obs)
Primary splits:
Symmetric_GM_nAtivacoes < 1415 to the right, improve=6.223810, (0 missing)
Affected_BF_nAtivacoes < 1108.5 to the left, improve=5.848810, (0 missing)
Symmetric_SOL_nAtivacoes < 1170.5 to the right, improve=5.566667, (0 missing)
Affected_SOL_nAtivacoes < 1380.5 to the right, improve=5.080952, (0 missing)
Symmetric_GM_mAtivacao < 0.03148499 to the left, improve=4.737480, (0 missing)
Surrogate splits:
Symmetric_SOL_nAtivacoes < 1574 to the right, agree=0.900, adj=0.786, (0 split)
Peso < 71.2 to the right, agree=0.833, adj=0.643, (0 split)
Affected_SOL_nAtivacoes < 1380.5 to the right, agree=0.800, adj=0.571, (0 split)
Symmetric_GM_pAtivacao < 2.386294 to the right, agree=0.800, adj=0.571, (0 split)
Affected_GM_nAtivacoes < 1363 to the right, agree=0.767, adj=0.500, (0 split)

Node number 2: 14 observations
predicted class=3 expected loss=0.2857143 P(node) =0.4666667
class counts: 3 0 10 0 0 1 0
probabilities: 0.214 0.000 0.714 0.000 0.000 0.071 0.000

Node number 3: 16 observations
predicted class=4 expected loss=0.5625 P(node) =0.5333333
class counts: 0 3 0 7 3 1 2
probabilities: 0.000 0.188 0.000 0.438 0.188 0.062 0.125

```

Figura 27 – Árvore de decisão do *clustering* dos dados do paciente e do resumo da prova EMG

Os resultados foram ainda analisados pelos peritos do CEMAH que irão avaliar com futuros pacientes a validade dos resultados obtidos e se podem auxiliar o tratamento do paciente, associando tratamentos e respostas à classificação dada. Como trabalho futuro poderá também ser associado a recuperação parcial ou total da marcha à classificação, de forma a dar um dado indicativo da possibilidade de recuperação de marcha. Os resultados podem ainda ser expandidos com a aquisição de valores 3 meses após o AVC e introduzindo a distância temporal entre a prova e o AVC como um novo critério ou então analisando a progressão entre dois instantes pré-definidos, como 3 meses e 6 meses após o AVC.

7.2.3.2 Forças de reação do solo (FRS)

A segunda procura de padrões é na prova de levantar-sentar. Como anteriormente, para além dos dados da prova (forças de reação do solo) serão ainda analisados a idade, o género, a altura e o peso das pacientes.

7.2.3.2.1 ETL

Ao ficheiro previamente gerado em Prova de levantar-sentar - forças de reação do solo (FRS) é adicionada a informação do paciente.

7.2.3.2.2 Clustering

Seguindo os exemplos anteriores da aplicação K-Means deste documento, o primeiro passo é a exploração dos possíveis clusters que poderão ser obtidos com os dados (Figura 28 – Desenho do valor de *withinss* por número de clusters, relativos aos dados da FRS e do paciente).

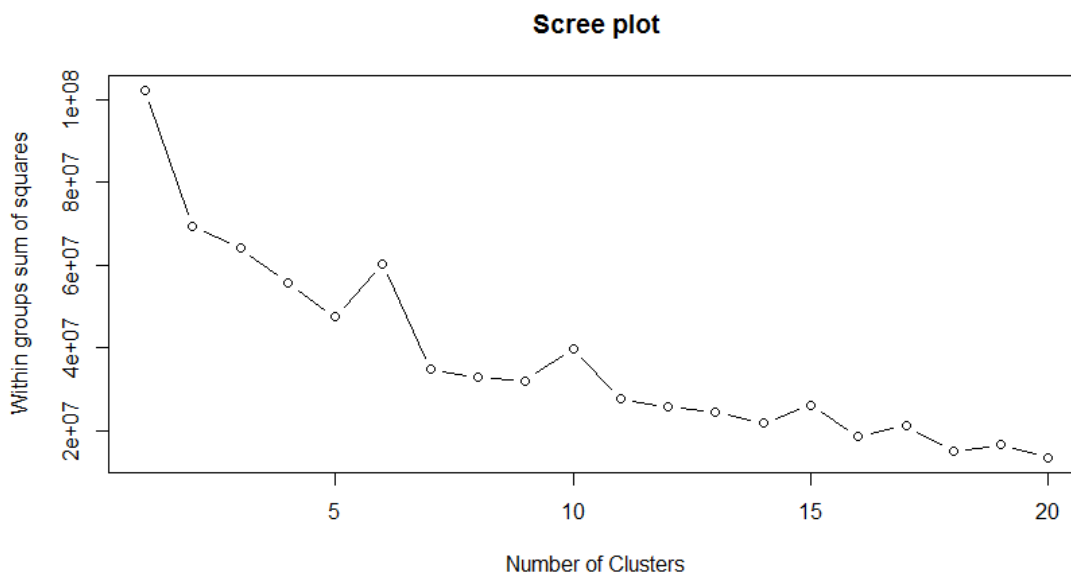


Figura 28 – Desenho do valor de *withinss* por número de clusters, relativos aos dados da FRS e do paciente

O resultado obtido não é de acordo com o pretendido, tornando a identificação das indentações mais difícil. São então feitas diversas experiências com 2, 5, 7, 11 e 14 clusters de forma a identificar o número de clusters que melhor representem os dados. As figuras Figura 29, Figura 30, Figura 31, Figura 32 e Figura 33 apresentam os estudos realizados.

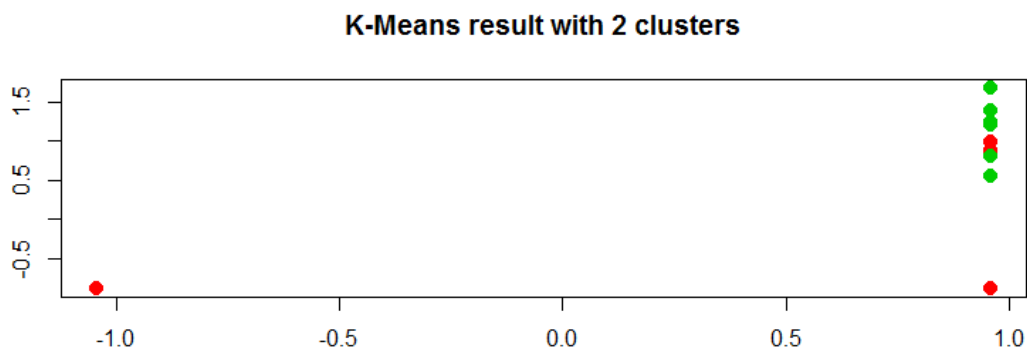


Figura 29 – Clustering dos dados do paciente e da prova FRS, com 2 clusters

K-Means result with 5 clusters

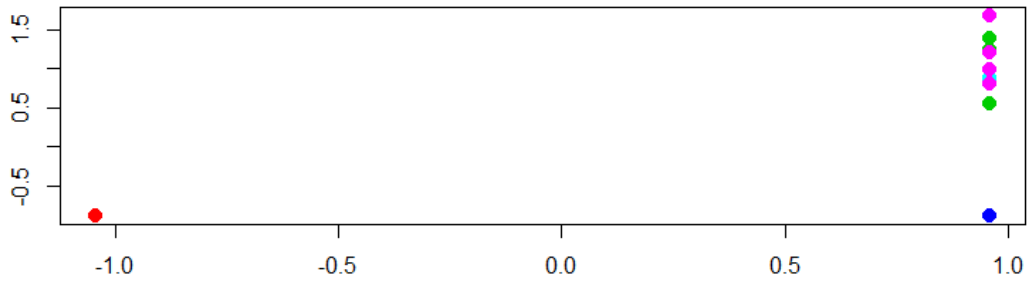


Figura 30 – Clustering dos dados do paciente e da prova FRS, com 5 clusters

K-Means result with 7 clusters

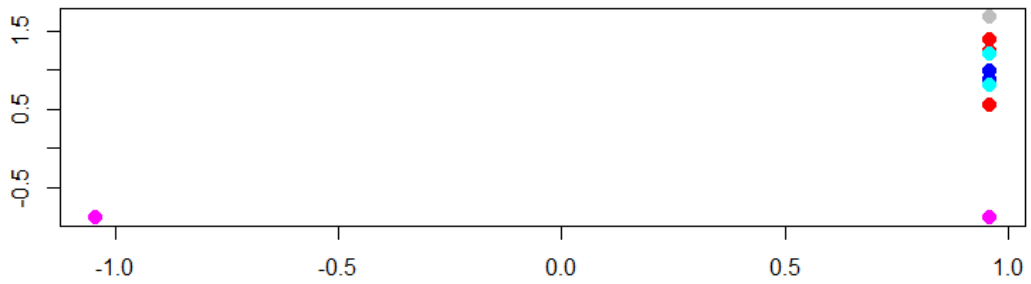


Figura 31 – Clustering dos dados do paciente e da prova FRS, com 7 clusters

K-Means result with 11 clusters

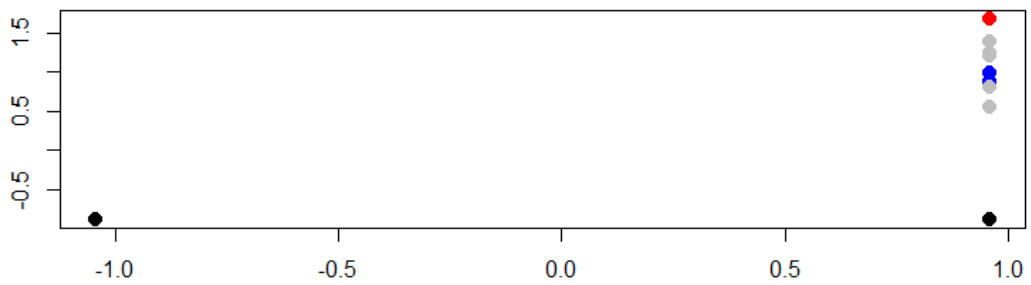


Figura 32 – Clustering dos dados do paciente e da prova FRS, com 11 clusters

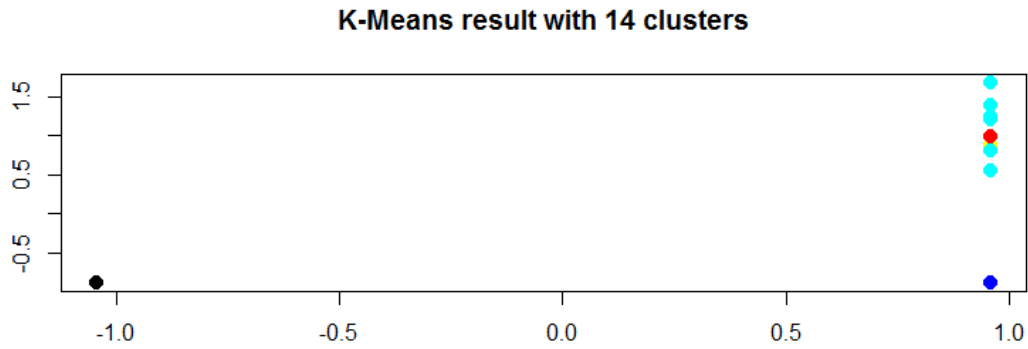


Figura 33 – Clustering dos dados do paciente e da prova FRS, com 14 clusters

Nenhum dos resultados apresenta centros bem definidos, havendo pouca distinção entre os clusters. Os resultados não apresentam relevância para a análise, levando ao estudo de outras possíveis abordagens.

7.2.3.2.3 Clustering longitudinal

Com a ajuda dos peritos, foi investigada uma nova abordagem ao problema. Como referido no capítulo Avaliação de soluções o que se procura estudar é o centro de gravidade. Os dados obtidos da plataforma estudada na seção anterior não representam corretamente o centro de gravidade, havendo a necessidade de serem transformados antes de se prosseguir na análise dos centros de gravidade. Para tal, foi necessário calcular a coordenada do Centro de Pressão no sentido médio-lateral (CoPx) e a coordenada do Centro de Pressão no sentido ântero-posterior (CoPy) (Figura 34 – Esquema exemplificativo dos parâmetros CoPx e CoPy) [58].

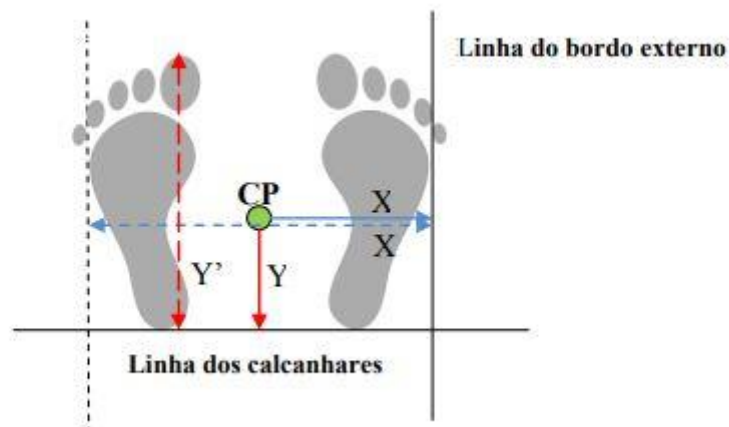


Figura 34 – Esquema exemplificativo dos parâmetros CoPx e CoPy

Nesta fase é realizado novamente um processo de ETL. Os dados carregados são os previamente trabalhados neste capítulo, sendo agora calculados os pontos CoPx e CoPy, segundo as seguintes formulas:

$$\text{CoPx} = (-h * Fx - My)/Fz$$

$$\text{CoPy} = (-h * Fy + Mx)/Fz$$

O h corresponde à altura de um objeto que seja colocado em cima da plataforma. Neste caso é 0.

São gerados dois novos ficheiros resultantes – um que apresenta os dados CoPx e outro com os dados CoPy. As linhas representam os pacientes enquanto que as colunas representam os momentos de captura de dados nas plataformas. Este novo formato surge como necessidade de avaliar a evolução de ambos os parâmetros ao longo do tempo.

No entanto, o algoritmo K-Means anteriormente utilizado não possui a capacidade de lidar com os dados neste novo formato. Enquanto que com a aplicação do K-Means se procura encaixar cada registo num cluster, o pretendido é comparar as variações do centro de gravidade. Para tal é necessário aplicar técnicas de *clustering* longitudinal. O K-Means possui uma implementação para análise longitudinal – *kml*. Este package utiliza os seguintes argumentos:

- `object` – Dados que o algoritmo irá analisar.
- `nbClusters` - Vector contendo o número de clusters com os quais *kml* deve funcionar. Por padrão, `nbClusters` é 2: 6, o que indica que *kml* deve procurar partições com, respectivamente, 2, depois 3, ... até 6 clusters. O número máximo de agrupamentos é 26.
- `nbRedrawing` - Define o número de vezes que *k-means* deve ser re-executado (com condições de partida diferentes) para cada número de clusters.
- `toPlot` - ou '`traj`' para plotar apenas as trajetórias, '`critério`' para plotar o critério sozinho, '`ambos`' para traçar ambos ou '`nenhum`' para não exibir nada (mais rápido).

Os parâmetros mais relevantes que foram ajustados:

- `object` – dados para realizar *clustering*.
- `nbRedrawing` – utilizado o valor standard (20).
- `toPlot` – foi optado pela opção “both”, de forma a ter acesso à visualização de *clustering*.

Previamente ao fornecimento dos dados através do parâmetro `object` é necessário os convertê-los para o formato `ClusterLongData`, utilizando o método `cld` da biblioteca *kml*. Este formato permite uma análise da matriz de dados ao longo da prova.

A procura do número de clusters ideal é feita através do critério Calinski & Harabasz [59]. O resultado pode ser concordante ou não concordante. A concordância é visualizada durante o período de clustering. Caso exista uma interseção de todas as linhas representativas dos clusters é considerado concordante. Se for não concordante, são experimentados os clusters onde exista interseção de algumas linhas representativas dos clusters. Na Figura 35 – Exemplo do resultado exploratório Calinski & Harabasz aplicado às provas FRS é visível os resultados não concordantes obtidos durante o período de clustering, onde é aconselhado experimentar 2, 3 e 4 clusters.

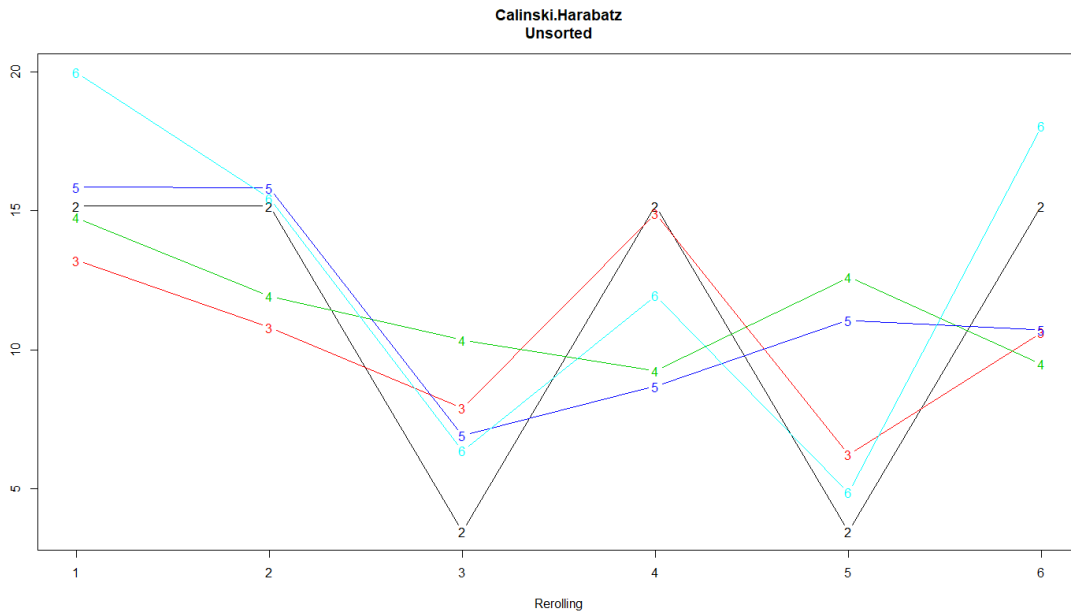


Figura 35 – Exemplo do resultado exploratório Calinski & Harabasz aplicado às provas FRS

7.2.3.2.4 Interpretação dos resultados

Foram analisados os valores do centro de pressão – CoPx e CoPy – e não são detetados padrões diferenciais ao longo da prova (Figura 36 - Diferentes variações do centro de pressão no sentido médio-lateral (CoPx) e Figura 37 - Diferentes variações do centro de pressão no sentido ântero-posterior (CoPy)).

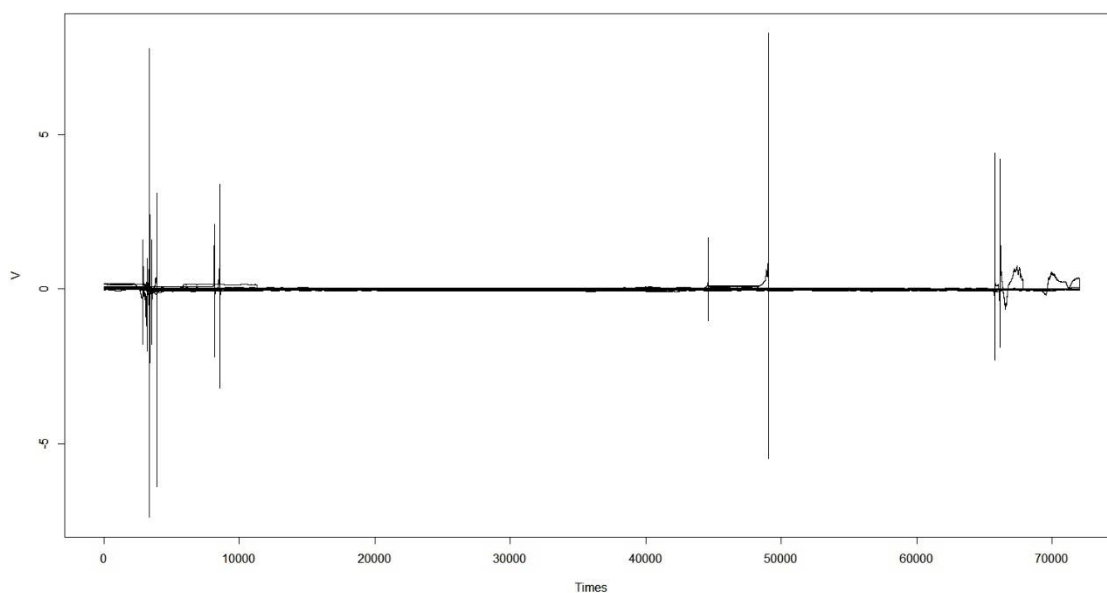


Figura 36 - Diferentes variações do centro de pressão no sentido médio-lateral (CoPx)

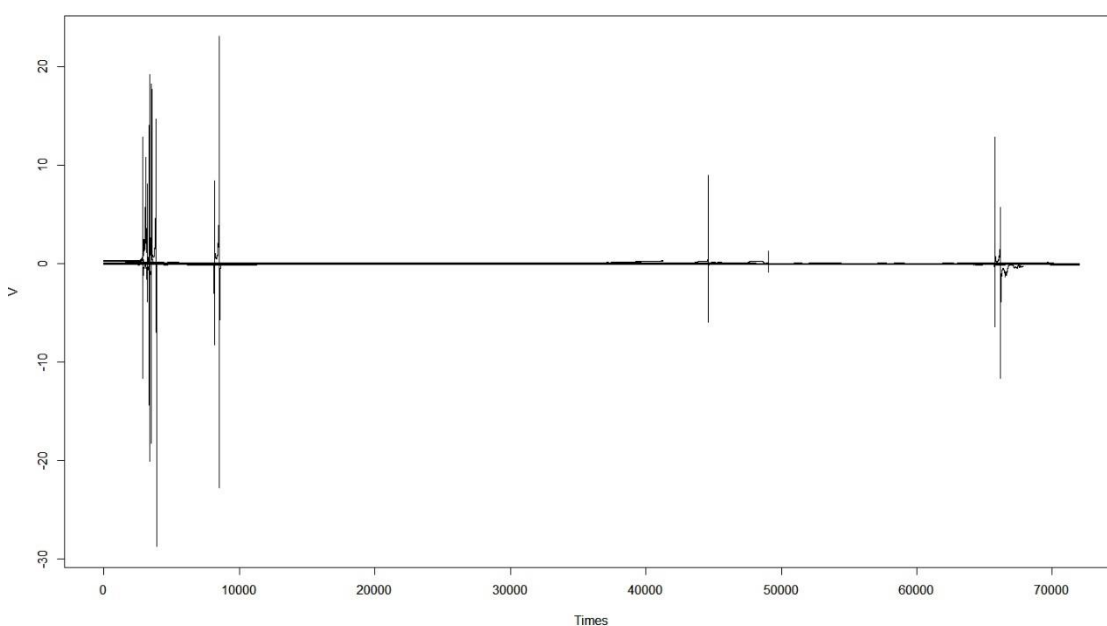


Figura 37 - Diferentes variações do centro de pressão no sentido ântero-posterior (CoPy)

Numa análise sumária, existem 3 momentos de variação do centro de pressão – ao início, após o meio e no final da prova. Estes momentos são sentidos em CoPx e CoPy. Conforme demonstrado na Figura 35 – Exemplo do resultado exploratório Calinski & Harabasz aplicado às provas FRS não existe concordância no número de padrões a analisar, tendo sido analisadas várias hipóteses. Ao analisar os períodos de variação em maior detalhe é possível visualizar que os clusters encontrados são divididos pelo momento de ativação, como a Figura 38 –

Análise do período inicial de ativação do centro de pressão no sentido médio-lateral (CoPx) demonstra.

Foram detetados 3 principais padrões em CoPx e CoPy, atendendo apenas ao momento da variação. Desta forma, os padrões encontrados não possuem significado relevante à análise dos pacientes de AVC, pois não existem variações diferentes do centro de pressão.

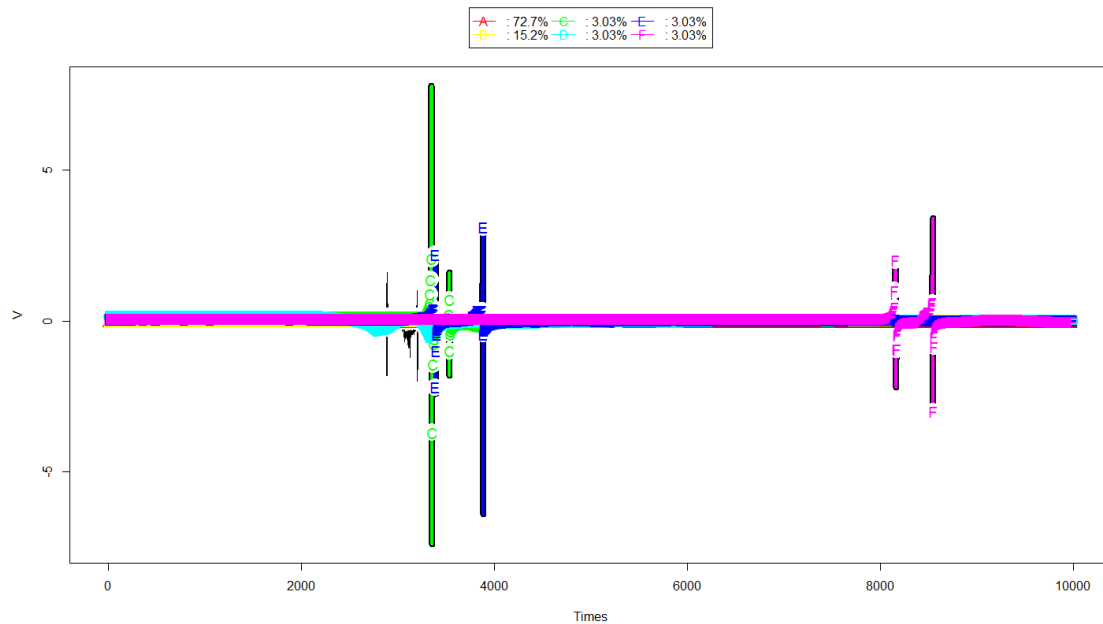


Figura 38 – Análise do período inicial de ativação do centro de pressão no sentido médio-lateral (CoPx)

7.3 Sistema de apoio à decisão

Foi criado um sistema de apoio à decisão para auxiliar os profissionais médicos no tratamento de pacientes de AVC. O sistema é disponibilizado sobre a forma de uma plataforma Web, onde está presente a informação encontrada na análise de dados, dados sobre o paciente e um sistema de recomendação de tratamentos. A informação obtida neste projeto e os dados do paciente – a ficha do paciente – permitem a centralização da informação e dos dados recolhidos.

O sistema de recomendação visa auxiliar no já atual processo de prescrição de tratamentos. O sistema recomenda tratamentos com base na ficha do paciente e nos tratamentos aplicados até ao momento. O sistema está ainda preparado para justificar as recomendações realizadas.

7.3.1 Plataforma Web

A plataforma Web é analisada nesta secção, iniciando pela apresentação do desenho e arquitetura do sistema. São expostas as decisões técnicas tomadas na implementação da

solução, fazendo referência ainda às entidades constituintes do sistema. Depois é apresentada a implementação da plataforma, páginas com que os utilizadores têm maior interação e código relevante. A secção é finalizada com a apresentação dos testes unitários e funcionais realizados.

7.3.1.1 Arquitetura

A plataforma Web foi construída atendendo à arquitetura de três camadas [60]. Esta arquitetura é constituída pelas camadas:

- **Presentation (PL)** – responsável pela interação com o utilizador
- **Business (BLL)** – responsável pela implementação das regras de negócio
- **Data (DAL)** – responsável pelo acesso a dados e serviços

Como a Figura 39 – Versão simplificada da arquitetura de três camadas [61] demonstra, a arquitetura em camadas dita que uma camada tem apenas acesso à camada abaixo. As camadas estão organizadas em PL > BLL > DAL.

A camada PL é constituída pelas páginas com que o utilizador interage, os inputs e outputs e o seu devido tratamento. São de especial menção os formulários da ficha de dados do paciente e do registo de tratamentos a pacientes.

A camada da logica de negócio (BLL) é constituída pelo sistema de recomendação, pelas validações de permissões/acessos e pela validação dos dados inseridos na ficha do paciente, do ponto de vista médico (como valores mínimos e máximos aceitáveis).

Na DAL é feita a comunicação com a base de dados SQL para a leitura e escrita de dados. Esta camada contém ainda os modelos que representam os objetos e entidades do sistema, os ViewModels [62] que representa a informação a apresentar e/ou recolher das páginas e o controlo de versões da aplicação através de Migrations [63].

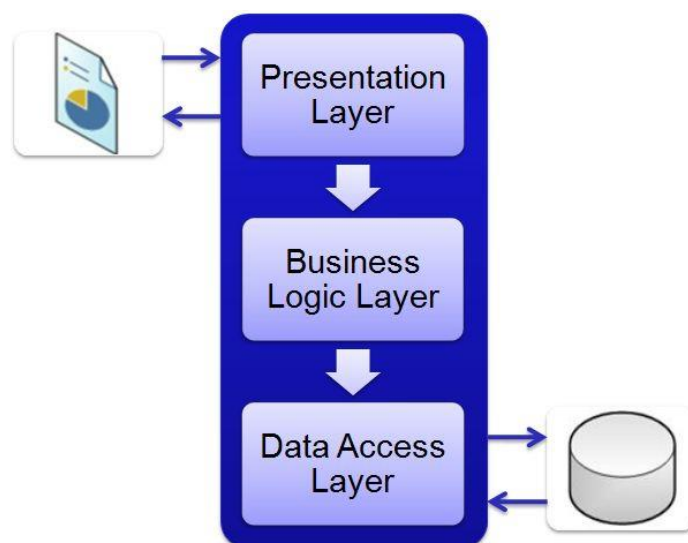


Figura 39 – Versão simplificada da arquitetura de três camadas

7.3.1.2 Modelo de dados

A Figura 40 – Modelo de dados representa a estrutura dos modelos existentes, assim como a relação entre os mesmos.

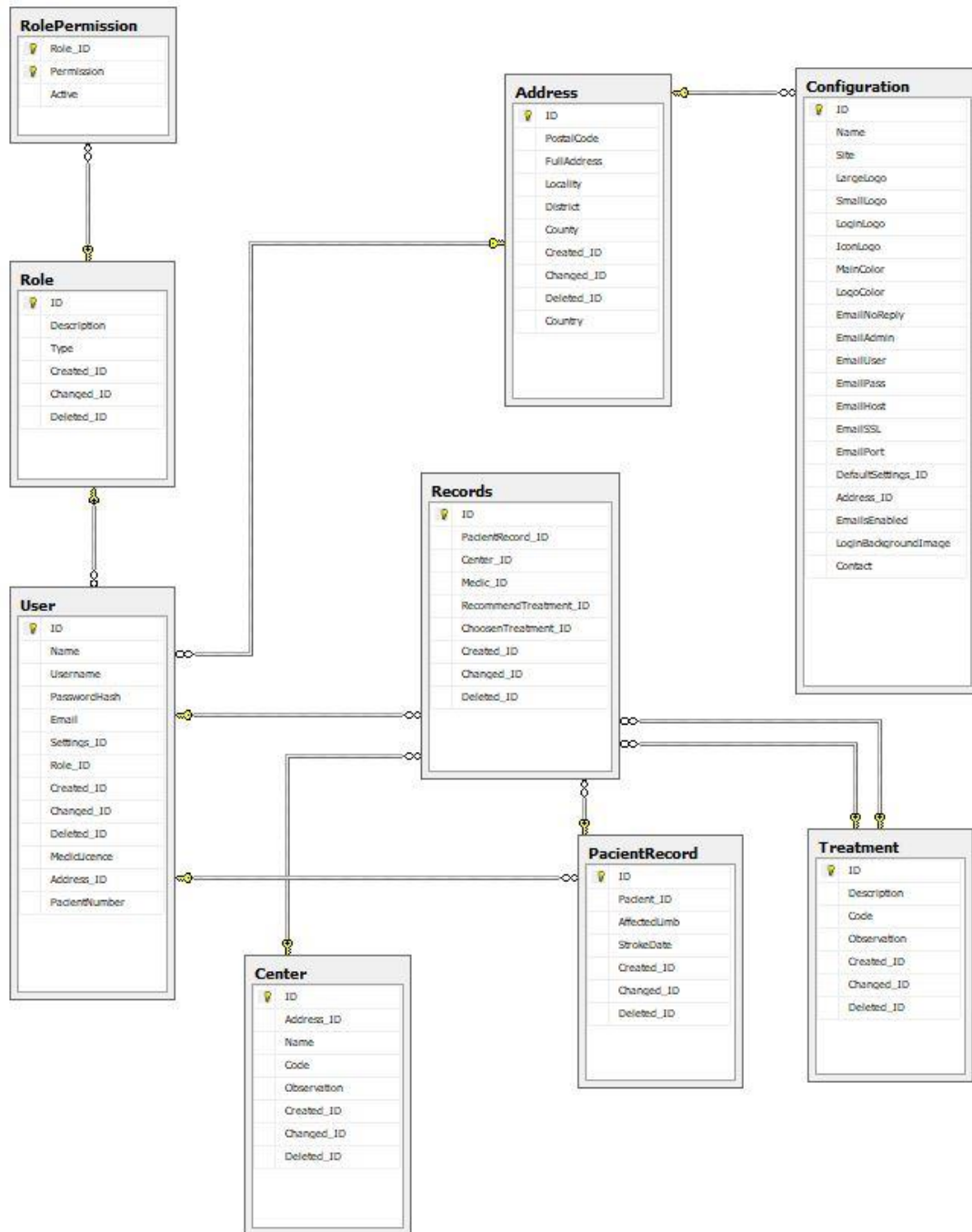


Figura 40 – Modelo de dados

De forma sumária, o modelo de dados reflete as seguintes entidades:

- **Configuration** – A tabela contém os dados relativos à aplicação de forma geral, como os logos, descrição, entre outros. Inclui ainda a morada, numa tabela diferente dado que é partilhado com outras entidades.
- **User** – Representa os utilizadores da aplicação.
- **Role e RolePermission** – A tabela *Role* representa as funções dos utilizadores. Exemplo são Administrador, profissional médico e paciente. A tabela *RolePermission* contém as permissões que cada função tem acesso.
- **Center** – Tabela representativo de hospitais e centros terapêuticos.
- **Treatments** – Tratamentos disponíveis para os pacientes.
- **PacientRecord e Records** - A ficha de paciente é disponibilizada na Tabela *PacientRecord*. Ela contém os dados da ficha do paciente, como o nome, o membro afetado e o número de utente. A tabela *Records* representa as diferentes interações que existem com a ficha de paciente, nomeadamente os tratamentos recomendados e os tratamentos escolhidos.

7.3.1.3 Implementação

Na presente secção são apresentados alguns detalhes da implementação da plataforma Web e as principais funcionalidades disponibilizadas aos utilizadores, sendo exibidas imagens da aplicação e um breve descritivo da funcionalidade analisada.

Para a implementação do front-end foi utilizado o template AdminLTE [64], que disponibiliza alguns componentes HTML e um layout para toda a aplicação.

Todas as ações sobre a aplicação despoletam um log de atividade, indicando quem realizou a ação, quando foi realizada e o que foi realizado (Figura 41 – Classe de registo do log de atividades, Figura 42 – Exemplo da inserção de um log de criação e Figura 43 – Exemplo da inserção de um log de edição). Estes logs permitem ao administrador do sistema ter acesso às origens de criações, edições e eliminações.

```
public class LogDAL : DBHandler
{
    public static object _lock = new object();
    public Log insert(int Object_ID, LogType Type, string Object_Type, int User_ID, DBHandler dbHandler)...
```

Figura 41 – Classe de registo do log de atividades

```
Log inserted = new LogDAL().insert(model.ID, LogType.Insert, ObjectType, User.ID, handler);
model.ID = new PacientRecordDAL(User).insert(model, inserted, handler);
```

Figura 42 – Exemplo da inserção de um log de criação

```
Log updated = new LogDAL().insert(model.ID, LogType.Update, ObjectType, User.ID, handler);  
int numberOfRowsChanged = new PacientRecordDAL(User).update(model, changed, handler);
```

Figura 43 – Exemplo da inserção de um log de edição

Existem ainda logs de erros que são despoletados na ocorrência de qualquer erro (Figura 44 – Exemplo de registo de log de erro). O log destes erros permite uma deteção e correção de erros mais eficiente.

```
int exception_ID = new ExceptionLogger().insert(User.ID, GetType().Name, System.Reflection.MethodBase.GetCurrentMethod().Name, e, handler.getParameters(), handler.getQueryFromCommand());  
throw new Exception("Não foi possível enviar o email");
```

Figura 44 – Exemplo de registo de log de erro

A funcionalidade comum entre todos os utilizadores é a autenticação perante o sistema (Figura 45 – Página de autenticação, Figura 46 – Exemplo de sucesso na autenticação e Figura 47 – Exemplo de falha na autenticação).

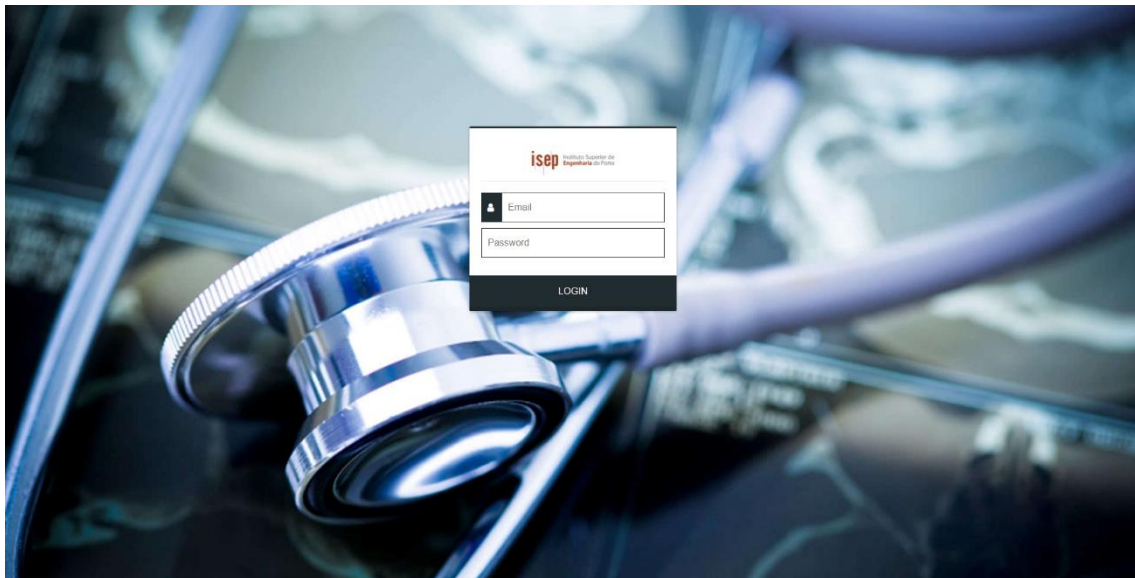


Figura 45 – Página de autenticação

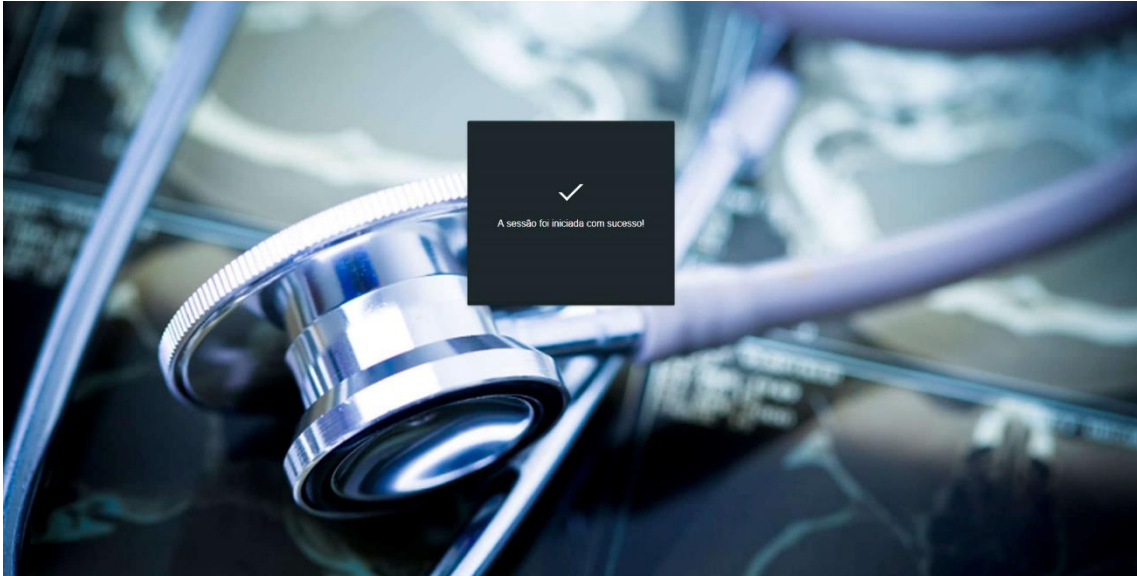


Figura 46 – Exemplo de sucesso na autenticação

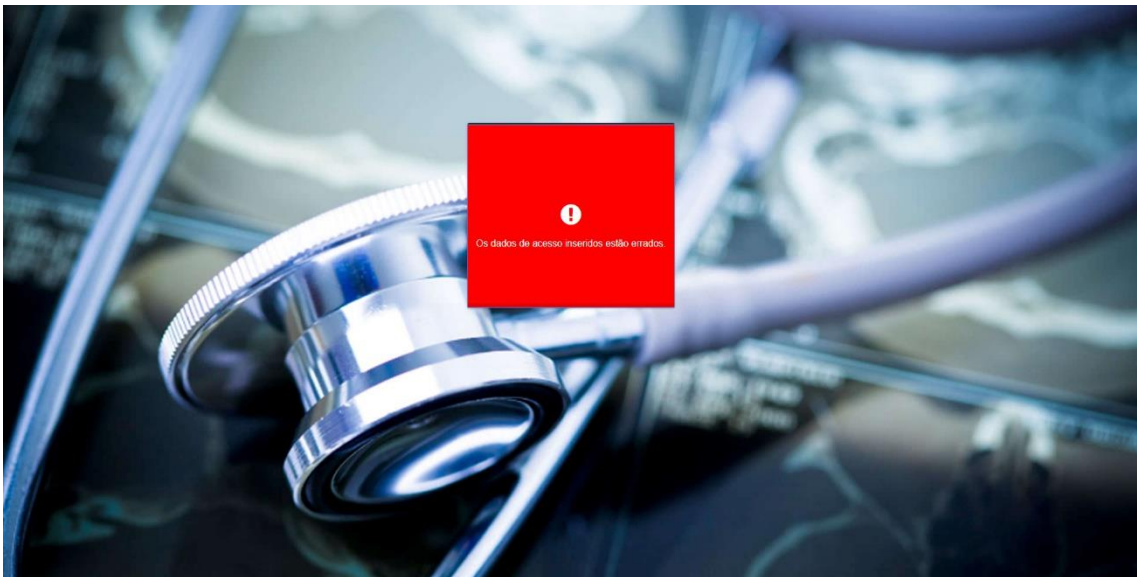


Figura 47 – Exemplo de falha na autenticação

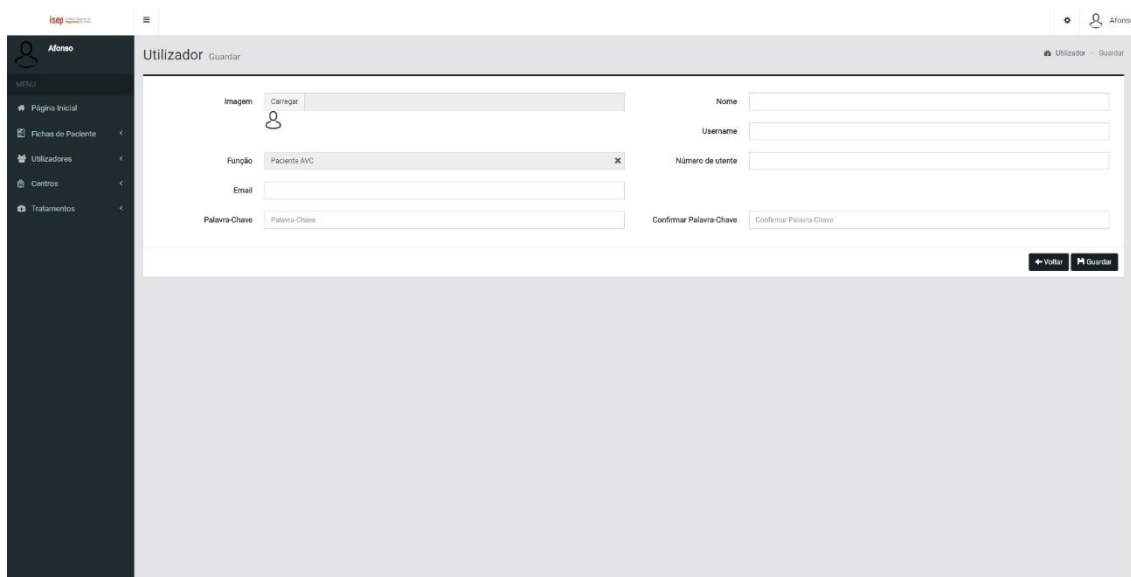
A password é encriptada e desencriptada com o algoritmo MD5 (

Anexos 2). Após o sucesso na autenticação, os dados do utilizador são persistidos em sessão [65]. Estes dados são posteriormente utilizados para validação de permissões (Figura 48 – Validação de permissões), acesso a informação do utilizador, entre outros.

```
LoggedUser user = Singleton.Instance.getUserInfo(User.Identity.GetUserId());
if (new RoleBLL().checkPermisson(user, PermissionKey.Users_Listing))
{
    Table table = new UserBLL(user).getAll(page);
    return Request.CreateResponse(HttpStatusCode.OK, table);
}
else return new HttpResponseMessage(HttpStatusCode.Forbidden);
```

Figura 48 – Validação de permissões

A gestão da ficha de paciente é a funcionalidade principal da aplicação. No processo de gestão, o primeiro passo é a criação da ficha. No processo de criação da ficha basta indicar informação sobre o paciente e o AVC (Figura 49 – Criação de um paciente).



The screenshot shows a web application interface for creating a patient record. The page is titled "Utilizador" and has a "Guardar" button. The form includes fields for "Imagem" (with a "Carregar" button), "Nome", "Username", "Número de utente", "Email", "Palavra Chave", and "Confirmar Palavra Chave". The "Função" field is set to "Paciente AVC".

Figura 49 – Criação de um paciente

Após a criação da ficha é possível indicar registos sobre o tratamento do paciente. Cada registo corresponde a uma consulta de prescrição de tratamento (Figura 50 – Ficha de paciente e seus registos).

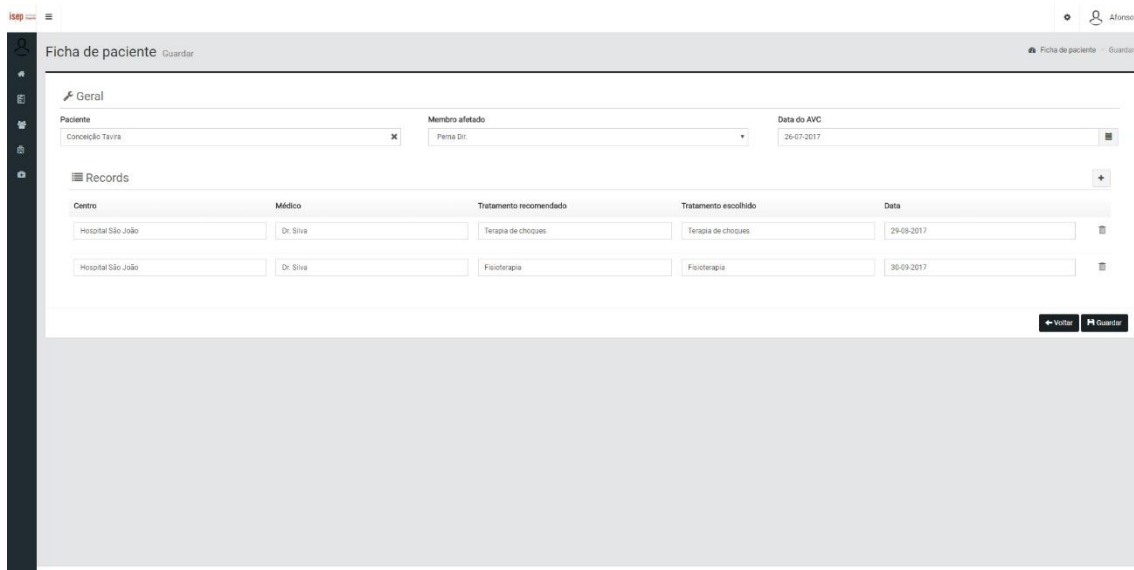


Figura 50 – Ficha de paciente e seus registos

No registo é possível pedir uma recomendação de tratamento, com a respetiva justificação. Esta recomendação pode ser aceite ou pode optar-se por um tratamento alternativo.

Os pacientes têm acesso a consultar a sua ficha, com apenas alguma informação disponível (Figura 51 – Consulta de uma ficha de paciente). O paciente não possui acesso ao sistema de recomendação, aos tratamentos recomendados e às observações inseridas pelo responsável médico.

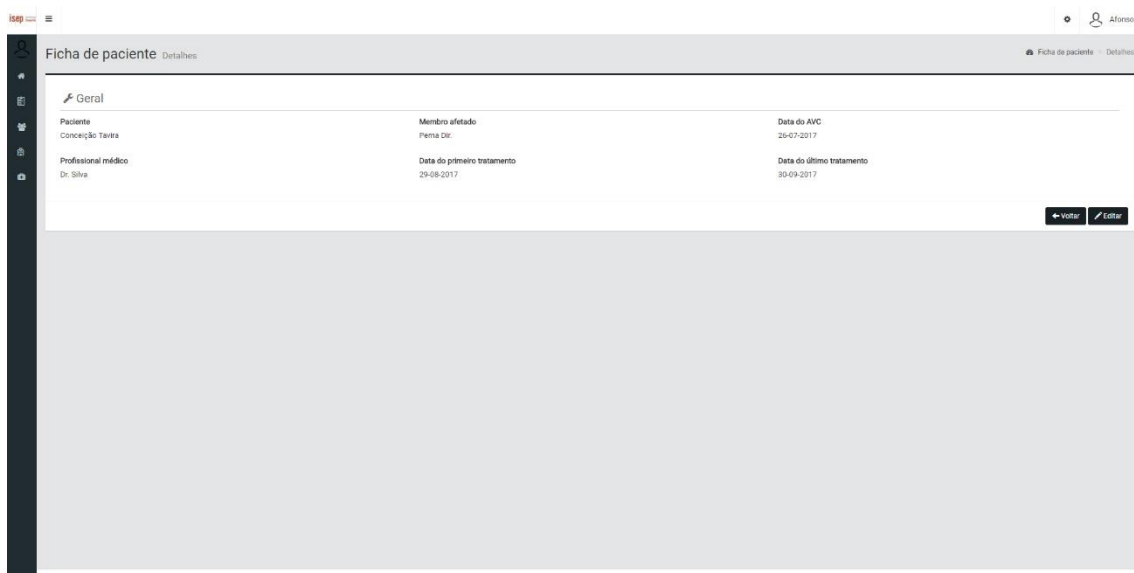


Figura 51 – Consulta de uma ficha de paciente

7.3.1.4 Testes unitários

A aplicação foi testada pelo método de testes unitários, a fim de verificar a inexistência de falhas nas funcionalidades disponibilizadas [66]. Para além das operações CRUD (Create, Read, Update, Delete) (Figura 52 – Teste unitário à criação de paciente, Figura 53 – Teste unitário à leitura de paciente, Figura 54 – Teste unitário à edição de paciente, Figura 55 – Teste unitário à eliminação de paciente) foram ainda criados testes para a validação de permissões (Figura 56 – Teste unitário à validação de permissões e Figura 57 – Teste unitário à validação de permissões) e validações de inserção de dados (Figura 58 – Teste unitário à validação de informação).

```
[TestMethod()]
0 references | 0 exceptions
public void createTest()
{
    PacientRecordVM vm = new PacientRecordVM();
    vm.Pacient = new User();
    vm.Pacient.Name = "Test name";
    vm.AffectedLimb = AffectedLimb.RightLeg;
    vm.StrokeDate = DateTime.Now;

    PacientRecordVM savedVM = new PacientRecordBLL(new LoggedUser().getAdmin()).save(vm);
    idTest = savedVM.ID;

    Assert.IsNotNull(savedVM);
}
```

Figura 52 – Teste unitário à criação de paciente

```
[TestMethod()]
0 references | 0 exceptions
public void getTest()
{
    PacientRecordVM vm = new PacientRecordBLL(new LoggedUser().getAdmin()).get(idTest);

    Assert.AreEqual(idTest, vm.ID);
}
```

Figura 53 – Teste unitário à leitura de paciente

```

[TestMethod()]
0 references | 0 exceptions
public void saveTest()
{
    PacientRecordVM vm = new PacientRecordBLL(new LoggedUser().getAdmin()).get(idTest);

    vm.Pacient.Name = "Test Pacient Name";
    new PacientRecordBLL(new LoggedUser()).save(vm);
    vm = new PacientRecordBLL(new LoggedUser()).get(idTest);

    Assert.AreEqual("Test Pacient Name", vm.Pacient.Name);
}

```

Figura 54 – Teste unitário à edição de paciente

```

[TestMethod()]
0 references | 0 exceptions
public void deleteTest()
{
    PacientRecordVM vm = new PacientRecordBLL(new LoggedUser().getAdmin()).get(idTest);

    new PacientRecordBLL(new LoggedUser()).delete(vm.ID);
    vm = new PacientRecordBLL(new LoggedUser()).get(idTest);

    Assert.AreEqual(0, vm.ID);
}

```

Figura 55 – Teste unitário à eliminação de paciente

```

[TestMethod()]
0 references | 0 exceptions
public void permissionFailedTest()
{
    LoggedUser userWithoutPermission = new LoggedUser().getUserWithoutPacientListingPermission();

    Singleton.Instance.logInUser(userWithoutPermission);
    new PacientRecordController().ListAllPacients();

    Assert.Fail();
}

```

Figura 56 – Teste unitário à validação de permissões (Sucesso)

```

[TestMethod()]
0 references | 0 exceptions
public void permissionSuccessTest()
{
    LoggedUser userWithPermission = new LoggedUser().getUserWithPacientListingPermission();

    Singleton.Instance.logInUser(userWithPermission);
    var result = new PacientRecordController().ListAllPacients();

    Assert.IsNotNull(result);
}

```

Figura 57 – Teste unitário à validação de permissões (Insucesso)

```

[TestMethod()]
0 references | 0 exceptions
public void strokeDateTest()
{
    PacientRecordVM vm = new PacientRecordBLL(new LoggedUser().getAdmin()).get(idTest);

    vm.StrokeDate = DateTime.Now.AddDays(1);
    new PacientRecordBLL(new LoggedUser()).save(vm);

    Assert.Fail();
}

```

Figura 58 – Teste unitário à validação de informação

7.3.1.5 Testes funcionais

São de seguida expostos os testes funcionais mais relevantes da plataforma Web.

Primeiramente é apresentado o teste à criação de um paciente, por um utilizador com permissão para realizar essa tarefa (Tabela 19 – Teste funcional – Criação de paciente).

Tabela 19 – Teste funcional – Criação de paciente

Funcionalidade	Criar paciente
<i>Pré-requisitos</i>	Conexão à internet Autenticação realizada Utilizador com permissão de criar pacientes
<i>Teste</i>	Aceder à opção “Utilizadores”, no menu lateral Selecionar a opção “Criar utilizador” No painel de criação, preencher corretamente os dados obrigatórios requeridos Selecionar o botão guardar
<i>Resultado obtido</i>	Após o preenchimento dos dados e a seleção do botão guardar o utilizador é redirecionado para a página de pacientes, onde está listado o novo paciente.
<i>Interpretação do resultado</i>	Teste executado e resultado em conformidade com os resultados esperados

O segundo teste é realizado à gestão da ficha do paciente, onde consta a sua informação pessoal e são registadas as consultas e tratamentos que lhe estão associadas (Tabela 20 – Teste Funcional – Gestão da ficha de paciente).

Tabela 20 – Teste Funcional – Gestão da ficha de paciente

Funcionalidade	Gerir ficha de paciente
<i>Pré-requisitos</i>	Conexão à internet Autenticação realizada Utilizador com permissão de edição de fichas de paciente Paciente já criado
<i>Teste</i>	Aceder à opção “Pacientes”, no menu lateral Selecionar a opção “Pacientes” Na página carregada, selecionar o paciente pretendido – podem ser utilizados filtros para esta seleção Abrir a ficha do paciente pretendido É aberto um novo painel com os dados e os registos do paciente, onde pode ser inserida ou alterada informação Selecionar o botão guardar
<i>Resultado obtido</i>	Após o preenchimento dos dados e a seleção do botão guardar o utilizador é redirecionado para a página de pacientes, onde está listado o paciente com a informação atualizada
<i>Interpretação do resultado</i>	Teste executado e resultado em conformidade com os resultados esperados

Na ficha do paciente é, como referido, possível inserir tratamentos. O teste seguinte tem como propósito validar essa acção (Tabela 21 – Teste funcional – Inserção de registo de tratamento).

Tabela 21 – Teste funcional – Inserção de registo de tratamento

Funcionalidade	Inserir registo de tratamento
<i>Pré-requisitos</i>	Conexão à internet Autenticação realizada Utilizador com permissão de edição de fichas de paciente Paciente já criado
<i>Teste</i>	Aceder à opção “Pacientes”, no menu lateral Selecionar a opção “Pacientes” Na página carregada, selecionar o paciente pretendido – podem ser utilizados filtros para esta seleção Abrir a ficha do paciente pretendido É aberto um novo painel com os dados e os registos do paciente. Na tabela de registos selecionar o botão de adicionar registos Inserir a informação do registo de tratamento, conforme pretendido Selecionar o botão guardar
<i>Resultado obtido</i>	Após o preenchimento dos dados e a seleção do botão guardar o utilizador é redirecionado para a página de pacientes, onde está listado o paciente com a informação atualizada

<i>Interpretação do resultado</i>	Teste executado e resultado em conformidade com os resultados esperados
-----------------------------------	---

No momento da prescrição de um tratamento é possível requerer uma recomendação. Essa recomendação é acompanhada de uma justificação. O teste apresentado na Tabela 22 – Teste funcional – Requisição de recomendação de tratamento valida esta funcionalidade.

Tabela 22 – Teste funcional – Requisição de recomendação de tratamento

Funcionalidade	Requerer recomendação de tratamento
<i>Pré-requisitos</i>	Conexão à internet Autenticação realizada Utilizador com permissão de edição de fichas de paciente Paciente já criado
<i>Teste</i>	Aceder à opção “Pacientes”, no menu lateral Selecionar a opção “Pacientes” Na página carregada, selecionar o paciente pretendido – podem ser utilizados filtros para esta seleção Abrir a ficha do paciente pretendido É aberto um novo painel com os dados e os registos do paciente. Na tabela de registos selecionar o botão de adicionar registos No campo de tratamento recomendado e no tratamento selecionado será carregada a recomendação Ao passar o cursor no tratamento recomendado é justificada a decisão O campo de tratamento selecionado é editável
<i>Resultado obtido</i>	Existe um tratamento recomendado, sendo acompanhado da sua justificação. O tratamento selecionado para o paciente pode ser diferente do tratamento recomendado.
<i>Interpretação do resultado</i>	Teste executado e resultado em conformidade com os resultados esperados

O paciente é também um utilizador do sistema. Na Tabela 23 – Teste funcional – Consulta de ficha de paciente é apresentado o teste funcional criado à consulta de uma ficha pelo próprio paciente.

Tabela 23 – Teste funcional – Consulta de ficha de paciente

Funcionalidade	Consulta de ficha de paciente, pelo mesmo
<i>Pré-requisitos</i>	Conexão à internet Paciente já criado Autenticação realizada pelo paciente Utilizador com permissão de consulta de fichas de paciente
<i>Teste</i>	Aceder à opção Ficha de Paciente no menu lateral Visualização da ficha de paciente, com apenas parte da informação disponível Inexistência da opção de editar

<i>Resultado obtido</i>	Ficha de paciente consultada, e sem opção de edição
<i>Interpretação do resultado</i>	Teste executado e resultado em conformidade com os resultados esperados

7.3.2 Sistema de recomendação

O sistema de recomendação faz parte do sistema de apoio à decisão criado no âmbito do projecto descrito neste documento. Nesta secção é apresentada a recolha e informatização da informação recolhida em forma de regras, o módulo de justificação das recomendações e a integração com a plataforma Web, apresentada na secção 7.3.1.

7.3.2.1 Regras

O primeiro passo necessário para a criação do sistema de recomendação é a recolha de informação e a criação de uma base de conhecimento que irá servir como o conhecimento utilizado pelo sistema.

A recolha de informação sobre os tratamentos prescritos às pacientes vítimas de AVC foi realizada junto dos peritos do CEMAH e de alguns documentos produzidos e disponibilizados por fontes que os peritos consideram fidedigna. Entre estes documentos encontram-se National clinical guideline for stroke por Royal College of Physicians, Stroke rehabilitation: long-term rehabilitation after stroke por National Institute for Health and Care Excellence e A Guideline for Healthcare Professionals por American Heart Association/American Stroke Association (AHS/ASS). O foco do sistema é a prescrição, a longo termo, de tratamentos para vítimas de AVC.

Foram passíveis de recolher regras sobre 15 tópicos, que abrangem desde do tratamento de um paciente com movimentação deblitada à prescrição de tratamentos para problemas oculares. Os tópicos são:

- Mobilidade
- Tratamento da incontinência intestinal e da bexiga
- Avaliação, prevenção e tratamento da dor no ombro hepélice
- Dor central após o acidente vascular cerebral
- Prevenção de quedas
- Convulsões
- Depressão, incluindo estado emocional e comportamental
- Incapacidade motora, atividade e mobilidade
- Distúrbios de comunicação
- Deteriorações sensoriais, incluindo toque, visão e audição
- Apraxia dos membros
- Afasia

- Falta de equilíbrio
- Tratamentos / Intervenções para deficiências visuais
- Perda de audição

Com o Guideline for Healthcare Professionals é possível não só fazer recomendação de tratamentos com base nos dados do paciente, mas também classificar o tratamento face ao seu benefício/risco e precisão. Ao analisar o tratamento face ao seu benefício e ao seu risco foram definidos 4 patamares:

- O benefício que o tratamento proporciona é muito superior ao risco, sendo assim recomendado pela sua utilidade e eficácia. Classificado como tratamentos de Class I.
- O benefício é superior ao risco, sendo favorável a recomendação do tratamento. Classificado como tratamentos de Class IIa.
- O benefício do tratamento é igual ou pouco superior ao risco, podendo ser considerado para a recomendação. Classificado como tratamentos de Class IIb.
- O risco é superior ao benefício, sendo prejudicial para o paciente. Estes casos não devem de ser recomendados. Classificado como tratamentos de Class III.
- Não definida – N/A

A precisão do tratamento é baseada nos casos de teste em que o tratamento foi aplicado e analisado. Existem 3 patamares definidos:

- A – Dados recolhidos de vários estudos clínicos ou meta-análises, sendo suficientes para uma recomendação fiável
- B – Dados recolhidos de apenas um estudo clínico ou meta-análise, sendo uma recomendação menos fiável.
- C – Recomendação realizada com base em censos, peritos ou tratamentos padrão.
- Não definida – N/A

A matrix obtida pelo cruzamento destes dados pode ser analisado no artefacto disponibilizado pelo AHS/ASS na Figura 59.

		SIZE OF TREATMENT EFFECT												
		CLASS I <i>Benefit >>> Risk</i> Procedure/Treatment SHOULD be performed/administered	CLASS IIa <i>Benefit >> Risk</i> Additional studies with <i>focused objectives</i> needed IT IS REASONABLE to perform procedure/administer treatment	CLASS IIb <i>Benefit ≥ Risk</i> Additional studies with <i>broad objectives</i> needed; <i>additional registry data</i> would be helpful Procedure/Treatment MAY BE CONSIDERED	CLASS III <i>No Benefit</i> or CLASS III <i>Harm</i>									
					<table border="1"> <thead> <tr> <th></th> <th>Procedure/Test</th> <th>Treatment</th> </tr> </thead> <tbody> <tr> <td>COR III: No benefit</td> <td>Not Helpful</td> <td>No Proven Benefit</td> </tr> <tr> <td>COR III: Harm</td> <td>Excess Cost w/o Benefit or Harmful</td> <td>Harmful to Patients</td> </tr> </tbody> </table>		Procedure/Test	Treatment	COR III: No benefit	Not Helpful	No Proven Benefit	COR III: Harm	Excess Cost w/o Benefit or Harmful	Harmful to Patients
	Procedure/Test	Treatment												
COR III: No benefit	Not Helpful	No Proven Benefit												
COR III: Harm	Excess Cost w/o Benefit or Harmful	Harmful to Patients												
ESTIMATE OF CERTAINTY (PRECISION) OF TREATMENT EFFECT	LEVEL A Multiple populations evaluated* Data derived from multiple randomized clinical trials or meta-analyses	<ul style="list-style-type: none"> Recommendation that procedure or treatment is useful/effective Sufficient evidence from multiple randomized trials or meta-analyses 	<ul style="list-style-type: none"> Recommendation in favor of treatment or procedure being useful/effective Some conflicting evidence from multiple randomized trials or meta-analyses 	<ul style="list-style-type: none"> Recommendation's usefulness/efficacy less well established Greater conflicting evidence from multiple randomized trials or meta-analyses 	<ul style="list-style-type: none"> Recommendation that procedure or treatment is not useful/effective and may be harmful Sufficient evidence from multiple randomized trials or meta-analyses 									
	LEVEL B Limited populations evaluated* Data derived from a single randomized trial or nonrandomized studies	<ul style="list-style-type: none"> Recommendation that procedure or treatment is useful/effective Evidence from single randomized trial or nonrandomized studies 	<ul style="list-style-type: none"> Recommendation in favor of treatment or procedure being useful/effective Some conflicting evidence from single randomized trial or nonrandomized studies 	<ul style="list-style-type: none"> Recommendation's usefulness/efficacy less well established Greater conflicting evidence from single randomized trial or nonrandomized studies 	<ul style="list-style-type: none"> Recommendation that procedure or treatment is not useful/effective and may be harmful Evidence from single randomized trial or nonrandomized studies 									
	LEVEL C Very limited populations evaluated* Only consensus opinion of experts, case studies, or standard of care	<ul style="list-style-type: none"> Recommendation that procedure or treatment is useful/effective Only expert opinion, case studies, or standard of care 	<ul style="list-style-type: none"> Recommendation in favor of treatment or procedure being useful/effective Only diverging expert opinion, case studies, or standard of care 	<ul style="list-style-type: none"> Recommendation's usefulness/efficacy less well established Only diverging expert opinion, case studies, or standard of care 	<ul style="list-style-type: none"> Recommendation that procedure or treatment is not useful/effective and may be harmful Only expert opinion, case studies, or standard of care 									

Figura 59 - Classificação de recomendações e nível de evidência

7.3.2.2 Implementação

A implementação do sistema de recomendação foi realizada utilizando Drools. Para que este sistema possa funcionar é preciso primeiramente criar uma base de conhecimento, que será utilizada para a realização de recomendações.

O conhecimento necessário para realizar as recomendações são os tratamentos, as classificações dos tratamentos e os dados do paciente. Como o sistema de recomendação foi concebido para ser integrado com a plataforma Web e, como analisado no ponto 7.3.1, a plataforma já possui estes modelos foram apenas inseridos os campos que estavam em falta. Os tratamentos já haviam sido modelados, ficando por incluir as suas classificações. Foi então criado os campos de classificação de precisão e de classificação de efeito. Nos pacientes faltam todos os campos que caracterizam o seu estado de saúde, que são fundamentais para a prescrição de tratamentos.

Com os campos necessários para a criação das regras criados foi então possível iniciar a criação das regras. O motor de inferência do Drools utiliza ficheiros drl que contêm as regras para a obtenção de resultados. A nomenclatura utilizada é de fácil leitura, permitindo que os peritos pudessem validar as mesmas e permitindo a gestão das regras de forma independente do sistema. Na Figura 60 – Exemplo de regras definidas é possível ver algumas regras definidas.

```

rule "BotulinumToxin"
  when
    custOrder : Patient( hasSevereHypertonicityInHemiplegicShoulderMuscles == true,
                        NeuropathicPainManifestedAsSensoryChange == false)
  then
    Result("Botulinum toxin injection");
  end

rule "NeuromodulatingPainMedication"
  when
    custOrder : Patient( hasSevereHypertonicityInHemiplegicShoulderMuscles == true,
                        NeuropathicPainManifestedAsSensoryChange == true)
  then
    Result("Neuromodulating pain medications");
  end
end

```

Figura 60 – Exemplo de regras definidas, no formato drl

Para a configuração do motor de inferência basta fazer o carregamento dos ficheiros (Figura 61).

```

private void StartDrools()
{
    PackageBuilder builder = new PackageBuilder();

    Stream stream = new FileStream(@"Rules.drl", FileMode.Open);
    builder.AddPackageFromDrl("Rules.drl", stream);

    Package pkg = builder.GetPackage();
    ruleBase = RuleBaseFactory.NewRuleBase();
    ruleBase.AddPackage(pkg);

    workingMemory = ruleBase.NewWorkingMemory();
    workingObj = new Patient();
    secondWorkingObj = null;
}

```

Figura 61 – Carregamento das regras para o motor de inferência

Para o pedido de inferência basta indicar o paciente ao motor e aplicar as regras previamente definidas, demonstrado na Figura 62.

```

this.patient = _thisForm.workingMemory.assertObject(workingObj);

_thisForm.workingMemory.fireAllRules();
Patient patient = (Patient)_thisForm.workingMemory.getObject(this.patient);

```

Figura 62 – Pedido de recomendação a um paciente

Ao aplicar as regras sobre o modelo é possível verificar que regras foram utilizadas e porquê, proporcionando assim uma justificação à recomendação efetuada. A recomendação e a sua explicação são então facultadas ao utilizador.

7.3.2.3 Integração com a plataforma WEB

O sistema de recomendação e da plataforma Web partilham, entre si, o modelo do paciente e do tratamento. Aliando este facto e que os dois módulos foram implementados na mesma linguagem – C# - foram ambos implementados no mesmo projeto. Desta forma a comunicação é realizada de uma forma direta.

Para o funcionamento correto do sistema de recomendação é necessário que a plataforma Web recolha e envie os dados do paciente para o motor de inferência. Exemplificando um caso de aplicação das regras da Figura 60, na Figura 63 é possível ver o formulário de introdução de dados.

Centro	Médico	Tratamento recomendado	Tratamento escolhido	Data
Hospital São João	Dr. Santos	Pesquisar Pacientes	Pesquisar Pacientes	22-10-2017
<p>Mobilidade Tratamento da incontinência intestinal e da bexiga Dor central após o acidente vascular c</p> <ul style="list-style-type: none"> <input type="checkbox"/> Hiper tonicidade grave nos músculos hemiplegicos do ombro <input type="checkbox"/> Sintomas de dor neuropática <input type="checkbox"/> Lesão nos tecidos moles <input type="checkbox"/> Inflamação dos músculos <input type="checkbox"/> Hemiplegia e restrições na amplitude do movimento dos ombros 				
		Avaliação, prevenção e tratamento da dor no ombro hieplice		

Figura 63 – Introdução de dados para a prescrição de tratamentos

Com os dados inseridos pelo profissional médico sobre o paciente é então obtida uma recomendação de um tratamento para o paciente. Esta recomendação é acompanhada de uma justificação e das classificações do tratamento (Figura 64).

- Hipertonicidade grave nos músculos hemiplegicos do ombro
- Sintomas de dor neuropática
- Lesão nos tecidos moles
- Inflamação dos músculos
- Hemiplegia e restrições na amplitude do movimento dos ombros

Tratamento recomendado: Medicamentos contra a dor neuromodulante (Classe IIa A)

Tratamentos analisados:

- Botulinum toxin injection
- Medicamentos contra a dor neuromodulante
- Injeção de corticosteróide
- Tenotomia cirúrgica do peitoral maior

Figura 64 – Exemplo de recomendação e justificação gerada

8 Conclusão

O problema discutido neste documento surge como uma necessidade do CEMAH no estudo e prescrição de tratamentos a pacientes. O estudo feito e a solução desenvolvida visam satisfazer parcialmente o assunto complexo que aqui é abordado.

O estudo dos dados veio confirmar algumas teorias dos peritos e revelou alguma informação que poderá vir a tornar-se relevante mediante um estudo futuro. O facto de os dados parecerem indicar que a perna simétrica exerce influencia sobre a marcha dos pacientes vem dar um fundamento estruturado à hipótese já levantada pelo CEMAH. Os músculos com maior importância também foram detetados e salientados, havendo possibilidade de alterar o conjunto de músculos analisados. Foram ainda detetados padrões que, futuramente, podem revelar-se importantes no estudo destes casos. Toda a análise realizada apresenta-se como uma proposta de estudo para futuras explorações.

O sistema de apoio à decisão apresenta-se como uma solução para a centralização de dados e apoio na prescrição de tratamentos. A plataforma Web pode ser expandida e integrada com outros sistemas, mas, de momento, apresenta uma solução que responde às necessidades assinaladas pelos peritos com uma interface de fácil utilização. O sistema de recomendação é uma ferramenta útil, especialmente com o módulo de justificações. É pretendido que este sistema produza uma melhoria no processo de recuperação e tratamento dos pacientes de AVC.

8.1 Objetivos realizados

Quando o CEMAH propôs este projeto, procurava uma solução para retirar informação a partir dos dados que já havia recolhido, assim como uma ferramenta que o auxiliasse na prescrição de tratamentos aos seus pacientes. Com o trabalho efetuado foi possível responder a estas necessidades.

A informação que haviam recolhido nos exames aos pacientes permitiu que fossem analisados e obtidos dados que suportam a hipótese de que os músculos da perna simétrica à perna afetada têm impacto na marcha. Esta hipótese, lançada pelos peritos, tem neste trabalho um estudo introdutório que poderá ser desenvolvido com a realização de mais exames e de um foco diferente dos mesmos. Com os dados foi ainda possível identificar os músculos com maior relevância na marcha. Esta identificação responde a uma necessidade que os peritos nomearão como talvez a mais pertinente, já que, não tendo os dados sido estudados, o grupo de músculos em análise nas provas não era o ideal. Com esta informação existe agora um indicador que permite saber quais os músculos mais relevantes entre o grupo de músculos estudados. É assim possível o estudo de outros músculos que tenham uma área de atuação semelhante ou dependente.

Nas primeiras reuniões exploratórias com os peritos do CEMAH foi proposta a procura e identificação de padrões nas provas que realizavam. Embora não existissem dados que apontassem para esta hipótese, a ideia de poderem existir padrões nas provas e que seria, portanto, possível associar resultados a tratamentos aos mesmos agradou, em muito, os peritos. Nesse sentido, as provas de caminhar e levantar-sentar foram analisadas e foram encontrados padrões na prova EMG. Os padrões serão agora analisados e validados pelos peritos em futuras provas. Foi analisada e excluída a hipótese de haver padrões na prova FRS, o que coincidiu com outro estudo realizado em paralelo pelo CEMAH. Os peritos sublinham que a inexistência de padrões provém da existência de vários cenários iniciais, tais como a postura do paciente na cadeira, a distância até ao encosto, entre outras.

Foi também desenvolvido um sistema de apoio à decisão, que veio permitir a centralização de informação e dos processos. A base deste sistema é uma plataforma Web que pode ser acedida pelos peritos e os próprios pacientes, de forma simples e sem a necessidade de nenhum dispositivo ou localização particular. Este sistema apoia ainda os peritos na prescrição de tratamentos para uma grande variedade de pacientes de AVC, justificando as recomendações realizadas segundo os parâmetros atualmente utilizados.

8.2 Limitações e trabalho futuro

Como referido anteriormente, a grande limitação do projeto deveu-se à quantidade de dados. Idealmente existiram mais dados para realizar as análises, permitindo obter resultados com maior confiança.

Futuramente deverão ser analisados os padrões obtidos. Caso sejam significativos, é possível registar e associar a probabilidade de recuperação total ou parcial da marcha. Esta informação poderá provar-se crucial para o processo de recuperação dos pacientes.

8.3 Criação de artigo

Na sequência do estudo aos músculos mais relevantes apresentados neste documento foi criado um artigo, com o objetivo de partilhar as descobertas feitas e a metodologia utilizada (Anexo 1). O artigo foi submetido para validação no jornal *Human Movement Science*, tendo o artigo um foco maior nas descobertas realizadas sobre os músculos dos membros inferiores de pacientes de AVC que apresentam uma marcha com problemas e não nas tecnologias utilizadas para a realização dessa mesma descoberta.

8.4 Apreciação final

O projeto deu-me a possibilidade de aplicar os conhecimentos obtidos no mestrado num ambiente real, com todas as implicações que isso acarreta. Todos os impasses que encontrei – desde a qualidade dos dados às interações com os peritos – refletem as advertências feitas durante o curso.

Pessoalmente, foi gratificante trabalhar num projeto de uma dimensão e complexidade desafiante e num tema que, infelizmente, é uma realidade com que parentes meus lidam de momento.

Creio que este desafio foi uma forma de desenvolver um leque de conhecimento prático nas áreas onde pretendo um dia trabalhar. Foi uma experiência que, sem dúvidas, mostrou alguns desafios que a realidade coloca ao conhecimento teórico. Espero levar comigo todo o conhecimento e obstáculos vencidos no projeto para a minha carreira profissional.

Os peritos do CEMAH deram uma avaliação positiva ao projeto, tendo conhecimento de algumas descobertas realizadas e ficando de validar as restantes. Têm planeado realizar mais recolhas para que seja possível realizar estudos mais completos e variados no futuro.

Bibliografia

- [1] C. -. C. d. E. d. M. e. A. Humana. [Online]. Available: <https://support.office.com/pt-pt/article/Criar-uma-bibliografia-81b1ba4a-6d0b-4475-88ca-f150ed6f49a0>. [Acedido em 10 2017].
- [2] S. F. E. P. Nicola e J. J. P. Ferreira, "Conceptual model for decomposing the value for the customer," *EMS'12—3rd Industrial Engineering and Management Symposium. Universidade do Porto*, pp. 41-43, 2012.
- [3] P. Seguro. [Online]. Available: [<http://www.planoseguro.pt/identificar-um-avc>]. [Acedido em 10 2017].
- [4] P. O. P. Neto, "Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo," [Online]. Available: <http://www.fmrp.usp.br/acidente-vascular-cerebral-avc-e-a-segunda-principal-cao-de-morte-no-brasil>. [Acedido em 10 2017].
- [5] P. Koen, "Associate Professor Stevens Institute, Researcher, Consultant and FEI Curator," [Online]. Available: <http://frontendinnovation.com/fei/what-is-the-new-concept-development-ncd-model>. [Acedido em 10 2017].
- [6] P. Koen, "Providing clarity and a common language to the "fuzzy front end"," *Research-Technology Management 44.2* , pp. 46-55, 2001.
- [7] P. Koen, "Understanding the Front End: A Common Language and Structured Picture," 2004. [Online]. Available: http://www.kip.zcu.cz/kursy/imi/ERASMUS/KoenMay26UnderstandFrontEndPDMA_01.pdf. [Acedido em 10 2017].
- [8] Y. P. Alexander Ostewalder, *Business Model Generation: handbook for visionaries, game changers, and challengers.*, John Wiley & Sons., 2010.
- [9] V. Allee, "Understanding Value Networks," [Online]. Available: [www.gurteen.com/gurteen/gurteen.nsf/id/L001020/\\$File/understanding-value-networks.pdf](http://www.gurteen.com/gurteen/gurteen.nsf/id/L001020/$File/understanding-value-networks.pdf). [Acedido em 10 2017].
- [10] V. Allee, "Value Network Analysis and value conversion of tangible and intangible assets," *Journal of intellectual capital 9.1* , pp. 5-24, 2008.
- [11] C. A. A. & C. M. Grosan, *Swarm intelligence in data mining*, 2006.
- [12] F. Rodrigues, "Descoberta de Conhecimento," [Online]. Available: <http://www.dei.isep.ipp.pt/~anamadur/SIdocs/DWOLAPDM.pdf>. [Acedido em 10 2017].

- [13] B. Palace, "Data Mining: What is Data Mining?," [Online]. Available: <http://www.anderson.ucla.edu/faculty/jason.frand/teacher/technologies/palace/datamining.htm>. [Acedido em 10 2017].
- [14] P. Raj, "Data mining process," [Online]. Available: <https://pt.slideshare.net/prithviness/data-mining-process>. [Acedido em 10 2017].
- [15] "WECA," [Online]. Available: <http://www.cs.waikato.ac.nz/ml/weka/index.html>. [Acedido em 10 2017].
- [16] "Matlab," [Online]. Available: <https://www.mathworks.com/products/matlab.html>. [Acedido em 10 2017].
- [17] "R," [Online]. Available: <http://www.rdatamining.com/r>. [Acedido em 10 2017].
- [18] "Oracle Data Mining," [Online]. Available: <http://www.oracle.com/technetwork/database/options/advanced-analytics/odm/overview/index.html>. [Acedido em 10 2017].
- [19] "Microsoft SQL Server," [Online]. Available: <https://msdn.microsoft.com/en-us/library/bb510516.aspx>. [Acedido em 10 2017].
- [20] B. C.-M. M. D. a. N. Z. H. Senoussi, "Feature Selection for Fault Detection Systems: Application to the Tennessee Eastman Process," *Applied Intelligence* 44.1, pp. 111-122, 2016.
- [21] J. O. P. Yiming Yang, "A comparative study of feature selection in text categorization," *Icml. Vol. 97*, pp. 412-420, 1997.
- [22] A. M. Mesleh, "Support Vector Machines based Arabic Language Text Classification System: Feature Selection Comparative Study," *Advances in Computer and Information Sciences and Engineering. Springer, Dordrecht*, pp. 11-16, 2008.
- [23] D. W. R. L. B. Aha, "Feature selection for case-based classification of cloud types: An empirical comparison," *Proceedings of the AAAI-94 workshop on Case-Based Reasoning*, vol. 106, p. 112, 1994.
- [24] D. W. K. D. & A. M. K. Aha, "Instance-based learning algorithms," *Machine learning*, vol. 6, pp. 37-66.
- [25] S. A. Thanyaluk Jirapech-Umpai, "Feature selection and classification for microarray data analysis: Evolutionary methods for identifying predictive genes," *BMC bioinformatics* 6.1 , p. 148, 2005.
- [26] K. D. J. H Vafaie, "Genetic Algorithms as a Tool for Feature Selection in Machine Learning," *Tools with Artificial Intelligence, 1992. TAI'92, Proceedings., Fourth International Conference on. IEEE*, pp. 200-203, 1992.

- [27] L. B. À. N. Luis Carlos Molina, "Feature Selection Algorithms: A Survey and Experimental Evaluation," *Data Mining, 2002. ICDM 2003. Proceedings. 2002 IEEE International Conference on. IEEE,* pp. 306-313, 2002.
- [28] S. a. S. N. S. Sumathi, *Introduction to data mining and its applications. Vol 29*, Springer, 2006.
- [29] "Data Clustering Algorithms," [Online]. Available: <https://sites.google.com/site/dataclusteringalgorithms/>. [Acedido em 10 2017].
- [30] M. M. P. F. A.K. Jain, "Data Clustering: A Review," *ACM computing surveys (CSUR)*, pp. 264-323, 1999.
- [31] "Data mining - cluster analysis," [Online]. Available: https://www.tutorialspoint.com/data_mining/dm_cluster_analysis.htm. [Acedido em 10 2017].
- [32] A. R. Jure Leskovec, "Stanford University," [Online]. Available: <https://web.stanford.edu/class/cs345a/slides/12-clustering.pdf>. [Acedido em 10 2017].
- [33] "Amazon," [Online]. Available: www.amazon.com. [Acedido em 10 2017].
- [34] "Netflix," [Online]. Available: www.netflix.com. [Acedido em 10 2017].
- [35] "Spotify," [Online]. Available: www.spotify.com. [Acedido em 10 2017].
- [36] B. S. a. J. Y. Greg Linden, "Amazon.com Recommendations - Item-to-Item Collaborative Filtering," *IEEE Internet computing 7.1*, pp. 76-80, 2003.
- [37] S. L. James Bennett, "The Netflix Prize," *Proceedings of KDD cup and workshop. Vol. 2007*, p. 35, 2007.
- [38] S. Simon, *The Essentials of Biostatistics for Physicians, Nurses, and Clinicians*, 2013.
- [39] B. C. G Agapito, "Profiling basic health information of tourists: towards a recommendation system for the adaptive delivery of medical certified nutrition contents," *High Performance Computing & Simulation (HPCS)*, pp. 616-620, 2014.
- [40] A. H. W. O. M Ati, "Implementation of a hybrid approach for chronic disease risk assessment and recommendation system," em *International Conference on Innovative Technologies*, 2014.
- [41] M. a. R. K. e. Hofmann, *RapidMiner: Data mining use cases and business analytics applications*, CRC Press, 2013.
- [42] C. Ramos, "Arquitetura de um Sistema Pericial," [Online]. Available: http://www.dei.isep.ipp.pt/~csr/SP/FichadeDisciplina_SP_4ano_2003-2004_.doc. [Acedido em 10 2017].
- [43] K. Latha, "Intelligence Based Recommendation System for Healthcare: a Patient centred

Framework,” em *CAASR International Conference on Innovative Engineering and Technologies*, 2015.

- [44] O. A. Abbas, “Comparisons between data clustering algorithms,” *International Arab Journal of Information Technology (IAJIT)*, 2008.
- [45] C. C. S. R. S. S. Kiri Wagstaf, “Constrained K-means Clustering with Background Knowledge - Proceedings of the Eighteenth International Conference on Machine Learning, 2001, p. 577–584,” [Online]. Available: <https://web.cse.msu.edu/~cse802/notes/ConstrainedKmeans.pdf>. [Acedido em 10 2017].
- [46] I. e. B. (. Dipartimento di Elettronica, “A Tutorial on Clustering Algorithms,” Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), [Online]. Available: https://home.deib.polimi.it/matteucc/Clustering/tutorial_html/kmeans.html. [Acedido em 10 2017].
- [47] a. B. F. [. Christophe Genolini [cre, “kml: K-Means for Longitudinal Data,” [Online]. Available: <https://cran.r-project.org/web/packages/kml/index.html>. [Acedido em 10 2017].
- [48] K. S. University. [Online]. Available: www.cs.kent.edu/~jin/DM08/ClusterValidation.pdf. [Acedido em 2017].
- [49] Z. Z. [Online]. Available: <http://fastml.com/evaluating-recommender-systems/>. [Acedido em 10 2017].
- [50] ““Bioplux clinical” revoluciona fisioterapia,” [Online]. Available: http://www.plux.info/files/pluxindmedia/20101124_portugalglobaln28.pdf. [Acedido em 10 2017].
- [51] A. S. Panos Vassiliadis, “EXTRACTION, TRANSFORMATION, AND LOADING,” [Online]. Available: http://www.cs.uoi.gr/~pvassil/downloads/ETL/SHORT_DESCR/08SpringerEncyclopedia_draft.pdf. [Acedido em 10 2017].
- [52] R. p. b. A. L. a. M. W. Fortran original by Leo Breiman and Adele Cutler, “Breiman and Cutler's Random Forests for Classification and Regression,” [Online]. Available: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>. [Acedido em 10 2017].
- [53] R. p. b. A. L. a. M. W. Fortran original by Leo Breiman and Adele Cutler, “randomForest: Breiman and Cutler's Random Forests for Classification and Regression,” [Online]. Available: <https://cran.r-project.org/web/packages/randomForest/index.html>. [Acedido em 10 2017].
- [54] T. a. F. Hastie, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer-Verlag, 2010.
- [55] H. a. W. (1979), “K-Means Clustering,” [Online]. Available: <https://stat.ethz.ch/R-manual/R-devel/library/stats/html/kmeans.html>. [Acedido em 10 2017].

- [56] D. J. a. C. L. S. Ketchen Jr, "The application of cluster analysis in strategic management research: an analysis and critique," *Strategic management journal*, pp. 441-458, 1996.
- [57] F. J. H. O. R. A. a. S. C. J. (. C. a. R. T. W. Breiman L., "Recursive Partitioning and Regression Trees," [Online]. Available: <https://stat.ethz.ch/R-manual/R-devel/library/rpart/html/rpart.html>. [Acedido em 10 2017].
- [58] Â. N. M. Ribeiro, "Variabilidade dos limites de estabilidade na posição ortostática em utentes com Acidente Vascular Encefálico," Escola Superior de Tecnologia da Saúde do Porto, Instituto Politécnico do Porto, 2013.
- [59] X. A. M. S. C. A. Christophe Genolini, "kml and kml3d: R Packages to Cluster Longitudinal Data," *Journal of Statistical Software*, vol. 65, pp. 1-34, 2015.
- [60] I. Departamento de Engenharia Informática, "Arquitetura Layers," [Online]. Available: <http://www.dei.isep.ipp.pt/~psousa/aulas/EINF/EINF-2006-01.pdf>. [Acedido em 10 2017].
- [61] P. Patel, "Three Layer Architecture in C# .NET," [Online]. Available: <https://www.codeproject.com/Articles/36847/Three-Layer-Architecture-in-C-NET>. [Acedido em 10 2017].
- [62] J. Galloway, "Views and ViewModels," Microsoft, [Online]. Available: <https://docs.microsoft.com/en-us/aspnet/mvc/overview/older-versions/mvc-music-store/mvc-music-store-part-3>. [Acedido em 10 2017].
- [63] T. Dykstra, "Code First Migrations and Deployment with the Entity Framework in an ASP.NET MVC Application," Microsoft, [Online]. Available: <https://docs.microsoft.com/en-us/aspnet/mvc/overview/getting-started/getting-started-with-ef-using-mvc/migrations-and-deployment-with-the-entity-framework-in-an-asp-net-mvc-application>. [Acedido em 10 2017].
- [64] A. Studio, "AdminLTE," Almsaeed Studio, [Online]. Available: <https://adminlte.io/themes/AdminLTE/index2.html>. [Acedido em 10 2017].
- [65] Microsoft, "Como criar e acessar variáveis de sessão no ASP.NET MVC," Microsoft, [Online]. Available: <https://code.msdn.microsoft.com/How-to-create-and-access-447ada98>. [Acedido em 10 2017].
- [66] G. W. C. R. A. H. G. H. Saisang Cai, "Unit Test Basics," Microsoft, [Online]. Available: <https://docs.microsoft.com/pt-pt/visualstudio/test/unit-test-basics>. [Acedido em 10 2017].

Anexos

**Paper - Feature-Selection Analysis of Relevant Muscles in
Stroke Patients with Impaired Walking Ability**

Feature-Selection Analysis of Relevant Muscles in Stroke Patients with Impaired Walking Ability

Fernando Afonso
Informatics Master Degree in
Knowledge-based and Decision, ISEP
Portugal
Afonso91@gmail.com

Abstract— One of the main research lines of the Centro de Estudos do Movimento e Atividade Humana [Center for the Study of Movement and Human Activity] (CEMAH) focuses on the therapy destined for patients who suffered a stroke and who, in the sequence of the accident, present a more limited range of motion (Anon., s.d.). Recent studies indicate that there might be a need to intervene not only in the leg affected by the stroke but also in the seemingly unaffected opposite limb. The aim of this paper is the study this possibility by analyzing a set group of muscles in each limb and by learning which ones have the most impact on the patients' movement. The technique of feature-selection was used for this study, applied in R with random forests. It was then possible to obtain that the opposite leg of the affected one has an relevant impact on the ability to walk without impair with an accuracy of 91%, being also able to determine the muscle with more relevance to a normal march.

Keywords — Stroke, impaired movement, muscles, affected limb, opposite limb, feature-selection, R, random forests

I. INTRODUCTION

Monitoring the condition of a patient who suffered a stroke is a task that presents a great degree of complexity, since it's dependent on the analysis of a set of predictor variables. The value for each of these variables – i.e. muscles – is obtained in the laboratory, from sensors applied to the limbs of the patients during the occurrence of different key events, examples being: standing up, sitting and starting to walk. Based on these values, an analysis is made which is then used to create a profile. That profile is then itself used in an attempt to make a prediction about the patients' condition.

A need was identified to determine which of the muscles involved in walking played a more decisive role in recovering a normal gait. The hypothesis of functional dependence between affected limb and its opposite was also deemed to be an important avenue for study. Although there is no conclusive study that proves this link, there are medical experts who, based on cases analyzed, have inferred its possible existence.

II. DATA COLLECTION

The data used on this paper was collected during a monitored walk, where electromyographic (EMG) activity values for some of the patients' muscles were collected – Biceps Femoris (BF), Gastrocnemius Medialis (GM), Rectus

Femoris (RF), Soleus (SOL), Tibialis Anterior (TA) and Vastus Medialis (VM). The data were first normalized and then a low-pass filter (10Hz) 4th order Butterworth using zero-phase lag was applied.

These data were collected using the medical device bioPlux Clinical (Bioplux, s.d.), with help from the medical experts at CEMAH.

III. RELEVANT FEATURES

Regarding discovery of knowledge, there are techniques that can be applied to understand the importance of each of the elements of the models that make up an area of study (Rodrigues, s.d.). The expansion and growth of the areas of study is, therefore, linked to an increase in the size and complexity of the models that operate therein. Each model in its turn consists of values that describe it. However, not all of these variables are of significant importance to the overall purpose of the business. Analyzing data from a complex business can be a time-consuming process – with results at times not as accurate as would be desired –, especially if no preliminary study is made in order to understand the models.

In the case being studied in this paper, techniques of feature selection (Casimira, 2016) (Langley, 1997) (Guyon, 2003) will be applied to the muscles' activity – these correspond to the relevant models of the business. Not all muscles will be of similar importance in the analysis of gait recovery, and, in order to understand which muscles are of greater significance, the activities of the limb whose activity has been affected during walking will be analyzed. Muscles should not be analyzed on a case by case basis, but rather as a whole, since a muscle can have greater or lesser importance if performed in tandem with one or more other muscles.

In the study of functional dependence of the affected limb and its opposite, all the muscles will be analyzed again during the representative activities and studied for their relevance. If there are muscles of the symmetric limb that have an impact on the restoration of gait, a correlation between the limbs will be verified.

It is necessary to analyze each muscle in accordance with its value and the value of the muscles that act at the same time. For this reason, the most appropriate feature selection

technique to use is Embedded. This technique was applied in R and the package used was random forests (Cutler, s.d.).

IV. MUSCLES OF THE AFFECTED LIMB AND HOW THEY AFFECTED THE PATIENTS' GAIT

A. ETL (Oliveira, s.d.)

The first step upon understanding which of the muscles were more relevant to the patients' gait is to extract, transform and load the data collected into a single file. After loading all the tests performed and removing all the noise that was collected, it was necessary to normalize all the data. With that purpose in mind, the following formula was applied:

$$\frac{\text{Electromyographic value} - 2048}{4096} * 5$$

The final result was a matrix with more than 6,000,000 records, with data collected from patients who suffered a stroke and volunteers who didn't present any signs of impaired walking.

B. Muscle relevance

After loading all the treated data into R, it was possible to apply the random forest algorithm. In order to get the best results possible, a maximum number of terminal nodes trees in the forest of 1 was required, with 1500 trees and a number of variables randomly sampled as candidates at each split between $(\sqrt{6})/2$ and $3\sqrt{6}$ in increments of $(\sqrt{6})/2$.

The results were obtained with an error of 8.88%, calculated from the confusion matrix that was made available by the algorithm. The relevance of the studied muscles was as follows:

Table 1 – Results of feature-selection on the data collected from the affected limb

Muscle	Relevance Rank
VM	1
TA	2
BF	3
RF	4
SOL	5
GM	6

C. Summary of the collected data

Although the results obtained already met the criteria set by the medical experts, the analysis performed can be improved. In addition to presenting high run times, it also does not indicate why the muscles have more or less importance in a normal gait. A new analysis was then proposed that focuses on whether the muscle is active or inactive rather than on the force exerted by it.

With help from the medical experts at CEMAH, it was possible to summarize in three variables the activity of a muscle during testing: number of activations, mean activation time and average of its value as active. The volume of data is then drastically reduced, allowing for faster scanning and easier verification by the medical experts. The results obtained show an accuracy with an approximate value of 92%. These values are equivalent to the ones obtained in the first test and are already in line with the values deemed necessary by the experts. With the metrics applied it is possible to perform the analysis in a fraction of the previous time, and the results are also more easily interpreted.

The results obtained are in accordance with the first analysis. The most important parameters are then on the muscles VM, TA and BF, in particular the number of activations and the mean value of activation. The results are as follows:

Table 2 - Results of feature-selection on the summarized data collected from the affected limb

Variable	Relevance Rank
Number of activations of VM	1
Mean activation time of VM	2
Number of activations of TA	3
Mean activation time of TA	4
Number of activations of BF	5
Mean activation time of RF	6
Average activation value of TA	7
Average activation value of BF	8
Average activation value of GM	9
Number of activations of RF	10
Average activation value of VM	11
Mean activation time of BF	12
Number of activations of GM	13
Mean activation time of SOL	14
Mean activation time of GM	15
Number of activations of SOL	16
Average activation value of RF	17
Average activation value of SOL	18

V. AFFECTED AND OPPOSITE LIMB

CEMAH, through examinations carried out over time, detected the possibility that the activity of leg muscles with fewer injuries affects a patient's gait. To investigate this possibility, the feature-selection process was repeated in the test data of both legs, thus making it possible to ascertain, from

the totality of the lower limb muscles evaluated, the ones that are the most relevant.

Since the error obtained between the two sets on the study of only the affected leg is very similar, the relevance of the muscles was again analyzed but this time with the summarized data, seeing as it is faster to run the random forest algorithm on it and the results are also easier to interpret.

The results obtained present an accuracy of 91%, similar to the accuracy obtained in the previous sections. The results support the importance of the VM and TA muscles of the affected leg, but also point to the relevance of the VM and the TA of the opposite leg. Using this information, it is possible to conclude that the tests do in fact corroborate the suspicions of the CEMAH medical experts. Following are eight more relevant features from both limbs.

Table 3 – Top results of feature-selection on the summarized data collected from both limbs

Limb/Muscle	Relevance Rank
Affected/Number of activations of VM	1
Affected/Number of activations of BF	2
Affected/Number of activations of TA	3
Symmetric/Mean activation time of VM	4
Affected/Mean activation time of VM	5
Affected/Mean activation time of TA	6
Symmetric/Mean activation time of TA	7
Symmetric/Number of activations of VM	8

VI. CONCLUSION

During the work that led to the writing of this paper, we were able to gather information about the affected limbs of stroke patients and their opposites. In each limb, a predefined group of muscles was analyzed, and techniques of feature-selection were used to understand which ones bore the most relevance.

The study led to the discovery that the most relevant muscles from the group were the VM, TA and BF. This result was obtained with an error below the one initially defined. It

was also found that the monitored walking could be summarized, allowing for faster and easier analysis.

The study also supports the hypothesis that the opposite limb affects a patient's gait, pointing to the relevance of the muscles of this limb.

VII. REFERENCES

- Anon., s.d. *CEMAH*. [Online]
Available at: <http://paginas.estsp.ipp.pt/cemah/?q=node/1>
[Last access in 10 2017].
- Bioplux, s.d. "*Bioplux clinical*" *revolucionaria fisioterapia*. [Online]
Available at:
http://www.plux.info/files/pluxindmedia/20101124_portugal_obaln28.pdf
[Last access in 10 2017].
- Casimira, R. E. B. G. C. A. Y., 2016. *The use of features selection and nearest neighbors rule for faults diagnostic in induction motors*. Volume 19, Issue 2, March 2006, Pages 169-177 ed. s.l.:s.n.
- Cutler, B. a., s.d. *Random Forests*. [Online]
Available at: <https://cran.r-project.org/web/packages/randomForest/randomForest.pdf>
[Last access in 10 2017].
- Guyon, I. a. A. E., 2003. *An introduction to variable and feature selection*. The Journal of Machine Learning Research, March 2003, Pages 1157-1182 ed. s.l.:s.n.
- Langley, A. L. B. a. P., 1997. *Selection of relevant features and examples in machine learning*. Artificial Intelligence. Volume 97, Issues 1-2, December 1997, Pages 245-271 ed. s.l.:s.n.
- Oliveira, P., s.d. *ETL*. [Online]
Available at:
<https://moodleant.iscp.ipp.pt/file.php/235375/Teoricas/ETL/ETL.pdf>
[Last access in 10 2017].
- Rodrigues, F., s.d. *Estruturas de Informação*. [Online]
Available at:
<https://moodleant.iscp.ipp.pt/file.php/235376/DESCOT1-DCBD/DESCO-01-ProcessDCBD.pdf>
[Last access in 10 2017].

Método de encriptação e descriptação MD5

```
public static string Encrypt(string toEncrypt, bool useHashing)
{
    byte[] keyArray;
    byte[] toEncryptArray = UTF8Encoding.UTF8.GetBytes(toEncrypt);

    System.Configuration.AppSettingsReader settingsReader = new AppSettingsReader();

    // Get the key from config file
    string key = (string)settingsReader.GetValue("SecurityKey", typeof(String));

    //If hashing use get hashcode regards to your key
    if (useHashing)
    {
        MD5CryptoServiceProvider hashmd5 = new MD5CryptoServiceProvider();
        keyArray = hashmd5.ComputeHash(UTF8Encoding.UTF8.GetBytes(key));

        //Always release the resources and flush data
        // of the Cryptographic service provide. Best Practice
        hashmd5.Clear();
    }
    else
        keyArray = UTF8Encoding.UTF8.GetBytes(key);

    TripleDESCryptoServiceProvider tdes = new TripleDESCryptoServiceProvider();
    //set the secret key for the tripleDES algorithm
    tdes.Key = keyArray;
    //mode of operation. there are other 4 modes.
    //We choose ECB(Electronic code Book)
    tdes.Mode = CipherMode.ECB;
    //padding mode(if any extra byte added)

    tdes.Padding = PaddingMode.PKCS7;

    ICryptoTransform cTransform = tdes.CreateEncryptor();
    //transform the specified region of bytes array to resultArray
    byte[] resultArray =
        cTransform.TransformFinalBlock(toEncryptArray, 0,
            toEncryptArray.Length);
    //Release resources held by TripleDes Encryptor
    tdes.Clear();
    //Return the encrypted data into unreadable string format
    return Convert.ToBase64String(resultArray, 0, resultArray.Length);
}
```

```

public static string Decrypt(string cipherString, bool useHashing)
{
    byte[] keyArray;

    //get the byte code of the string
    byte[] toEncryptArray = Convert.FromBase64String(cipherString);

    System.Configuration.AppSettingsReader settingsReader = new AppSettingsReader();

    //Get your key from config file to open the lock!
    string key = (string)settingsReader.GetValue("SecurityKey", typeof(String));

    if (useHashing)
    {
        //if hashing was used get the hash code with regards to your key
        MD5CryptoServiceProvider hashmd5 = new MD5CryptoServiceProvider();
        keyArray = hashmd5.ComputeHash(UTF8Encoding.UTF8.GetBytes(key));
        //release any resource held by the MD5CryptoServiceProvider
        hashmd5.Clear();
    }
    else
    {
        //if hashing was not implemented get the byte code of the key
        keyArray = UTF8Encoding.UTF8.GetBytes(key);
    }

    TripleDESCryptoServiceProvider tdes = new TripleDESCryptoServiceProvider();
    //set the secret key for the tripleDES algorithm
    tdes.Key = keyArray;
    //mode of operation. there are other 4 modes.
    //We choose ECB(Electronic code Book)

    tdes.Mode = CipherMode.ECB;
    //padding mode(if any extra byte added)
    tdes.Padding = PaddingMode.PKCS7;

    ICryptoTransform cTransform = tdes.CreateDecryptor();
    byte[] resultArray = cTransform.TransformFinalBlock(
        toEncryptArray, 0, toEncryptArray.Length);
    //Release resources held by TripleDes Encryptor
    tdes.Clear();
    //return the Clear decrypted TEXT
    return UTF8Encoding.UTF8.GetString(resultArray);
}

```