



Padrões de multimorbilidade em doentes internados com cancro da próstata em Portugal: Uma abordagem de análise de clusters

PATRÍCIA LOPES CARVALHO

Julho de 2023

PADRÕES DE MULTIMORBILIDADE EM DOENTES INTERNADOS COM CANCRO DA PRÓSTATA EM PORTUGAL: UMA ABORDAGEM DE ANÁLISE DE CLUSTERS

Patrícia Lopes Carvalho

2023

Instituto Superior de Engenharia do Porto

Departamento de Física

isen

P.PORTO

PADRÕES DE MULTIMORBILIDADE EM DOENTES INTERNADOS COM CANCRO DA PRÓSTATA EM PORTUGAL: UMA ABORDAGEM DE ANÁLISE DE CLUSTERS

Patrícia Lopes Carvalho

Estudante n.º 1210161

Dissertação apresentada ao Instituto Superior de Engenharia do Porto para cumprimento dos requisitos necessários à obtenção do grau de Mestre em Engenharia Biomédica, realizada sob a orientação do Doutor Julio Botelho Souza e coorientação da Doutora Maria Goreti Marreiros e do Doutor José Alberto da Silva Freitas.

2023

Instituto Superior de Engenharia do Porto

Departamento de Física

isen

P.PORTO

AGRADECIMENTOS

Ao meu orientador, Doutor Julio Botelho Souza, cuja orientação tornou este trabalho possível. Obrigada pela inestimável ajuda, dedicação e constante presença ao longo desta jornada.

À Professora Doutora Maria Goreti Marreiros e Professor Doutor José Alberto da Silva Freitas, obrigada pelos contributos e disponibilidade.

Ao Doutor Francisco Botelho, agradeço a partilha de conhecimentos e disponibilidade.

À equipa do Cintesis, em particular à Carolina, à Glória e ao João Viana, pelo caloroso acolhimento, foi um prazer colaborar lado a lado de profissionais exemplares assim.

Ao João Nuno, o meu amigo de longa data, obrigada por acreditar incondicionalmente em mim, pela valiosa ajuda prestada em todos os momentos. A nossa amizade permanecerá para sempre.

Aos meus pais e avós, por tudo o que sempre fizeram por mim, sem eles nada seria possível.

Ao Ricardo, o meu companheiro de vida, um agradecimento muito especial pela força, paciência e, principalmente, por todo o amor que me dá. Mal posso esperar para ver o que o futuro nos reserva.

Por fim, a todos os meus restantes familiares e amigos que, apesar de não mencionados, de alguma forma contribuíram para a concretização de mais uma importante etapa da minha vida.

página propositadamente em branco

RESUMO

A multimorbilidade pode ser definida como a coocorrência de duas ou mais condições crónicas. Este é um problema comum entre pacientes com cancro, que aumenta o *outcome* de resultados negativos, nomeadamente, morte prematura, complicações graves e baixa qualidade de vida, e resulta numa maior complexidade dos cuidados de saúde. A complexidade associada à coexistência de múltiplas condições crónicas, além da doença oncológica, requer abordagens de cuidados de saúde adaptadas e integradas. Isso representa um desafio significativo na vida dos pacientes, dos médicos e nos serviços de saúde como um todo. Este estudo tem como objetivo utilizar a análise de clusters para identificar e caracterizar padrões de multimorbilidade em pacientes com cancro da próstata usando dados hospitalares codificados clinicamente. Foram considerados os dados de internamentos com diagnóstico de cancro da próstata ocorridos em todos os hospitais públicos de Portugal Continental durante o período 2011-2017. Algoritmos de clustering de partição, nomeadamente, k-modes, PAM (*Partitioning Around Medoids*) e clustering hierárquico, foram empregues para identificar clusters de multimorbilidade. Os resultados obtidos a partir das diferentes abordagens de clustering foram comparados e avaliados quanto à relevância clínica. Um total de 10.394 episódios de internamentos foram analisados, com 6091 (58.6%) relatando multimorbilidade. Clusters semelhantes foram obtidos através dos diferentes algoritmos, com o método PAM apresentando alta estabilidade e melhores resultados em termos de coeficiente de silhueta média. A análise de 6 clusters obtidos com o PAM indica um padrão de diabetes coocorrendo com hipertensão e uma alta coocorrência de comorbilidades únicas, ou seja, hipertensão, doença pulmonar crónica, obesidade e arritmia concomitantes com o próprio cancro da próstata. A análise de clusters foi uma abordagem útil para detetar e caracterizar os diferentes padrões e perfis de multimorbilidade entre as hospitalizações por cancro da próstata em Portugal. Uma maior integração entre o cuidado do cancro e das comorbilidades deve ser reforçada para atender às necessidades dos pacientes com diversas doenças crónicas.

PALAVRAS-CHAVE

Análise de clusters, Machine Learning, Pacientes internados, Cancro da próstata, Multimorbilidade

página propositadamente em branco

ABSTRACT

Multimorbidity can be defined as the co-occurrence of two or more chronic conditions. This is a common issue among cancer patients, which increases the risk of negative outcomes, such as premature death, serious complications and poor quality of life, and results in greater complexity of care. The complexity associated with the coexistence of multiple chronic conditions, in addition to cancer, requires adapted and integrated health care approaches. This represents a significant challenge in the lives of patients, physicians and healthcare services as a whole. This study aims to use clustering analysis to identify and characterize multimorbidity patterns among prostate cancer patients using clinically coded hospital data. Data on hospital admissions with a diagnosis of prostate cancer occurring in all public hospitals in mainland Portugal during 2011-2017 were considered. Partitioning clustering algorithms, namely K-modes, PAM (Partitioning Around Medoids), and hierarchical clustering, have been employed to identify multimorbidity clusters. Results obtained from the different clustering approaches were compared and assessed regarding clinical relevance. A total of 10.394 inpatient episodes were analyzed, with 6091 (58.6%) reporting multimorbidity. Similar clusters were obtained through the different approaches, with PAM presenting high stability and better results in terms of average silhouette. The analysis of 6 clusters obtained with PAM indicates a pattern of diabetes co-occurring with hypertension and a high co-occurrence of single comorbidities, namely hypertension, chronic pulmonary disease, obesity, and arrhythmia with prostate cancer itself. Clustering analysis was a useful approach to detect and characterize the different patterns and profiles of multimorbidity among prostate cancer hospitalizations in Portugal. A greater integration between cancer and comorbidity care should be reinforced in order to meet the multimorbid patients' need.

KEYWORDS

Clustering, Machine Learning, Inpatient, Prostate Cancer, Multimorbidity

página propositadamente em branco

ÍNDICE

ÍNDICE DE FIGURAS.....	IX
ÍNDICE DE TABELAS	XI
LISTAS DE SIGLAS E SÍMBOLOS.....	XV
1 INTRODUÇÃO	1
1.1 Enquadramento e pertinência	1
1.1.1 Cancro da Próstata	1
1.1.2 Detalhando o conceito de Multimorbilidade.....	9
1.1.3 Indicadores de Multimorbilidade.....	11
1.1.4 Machine learning	13
1.1.5 Análise de clusters	14
1.2 Questão e objetivos de investigação.....	15
1.3 Opções metodológicas	16
1.4 Estrutura do trabalho	16
2 REVISÃO BIBLIOGRÁFICA.....	18
2.1 Metodologia de pesquisa	18
2.2 Análise dos artigos selecionados.....	19
3 MÉTODOS E APLICAÇÃO.....	21
3.1 Base de dados, população do estudo e definição de variáveis.....	21
3.2 Medida de Multimorbilidade	23
3.3 Software	23
3.4 Algoritmos de Clustering	24
3.4.1 Algoritmos de Partição.....	25
3.4.2 Métodos Hierárquicos.....	27
3.5 Medidas de Distância e Similaridade	28
3.6 Determinação do número ideal de clusters.....	31
3.7 Critérios de Validação.....	32
3.8 Árvores de decisão	32
3.9 Análise estatística.....	33
4 RESULTADOS E DISCUSSÃO	34
4.1 Apresentação de resultados.....	34
4.1.1 Número ideal de clusters e validação interna	37
4.1.2 Clusters de Multimorbilidade – Resultados e comparação entre os métodos.....	43
4.1.3 Caracterização e interpretação dos Clusters de Multimorbilidade obtidos pelo Método de PAM	49
4.1.4 Avaliação e interpretação dos clusters	54
4.2 Discussão de resultados	55
5 CONCLUSÃO	58

5.1 Conclusões finais	58
5.2 Limitações e investigação futura.....	58
REFERÊNCIAS BIBLIOGRÁFICAS	59
APÊNDICE A.....	67
APÊNDICE B.....	77
APÊNDICE C.....	80

página propositadamente em branco

ÍNDICE DE FIGURAS

Figura 1- a) Zonas anatómicas da próstata; b) Lobos prostáticos [6].	2
Figura 2 - Taxa de incidência e mortalidade do cancro da próstata por cada 100 mil habitantes, em 2020, consoante o continente [12].	4
Figura 3 – Fluxograma da revisão bibliográfica.	18
Figura 4 - Esquema do processo metodológico utilizado.	21
Figura 5 - Abordagem do algoritmo de partição [86].	25
Figura 6 - Esquema representativo do processo do algoritmo k-modes.	26
Figura 7 - Esquema representativo do processo do algoritmo PAM.	27
Figura 8 - Dendrograma do algoritmo de clustering hierárquico [86].	27
Figura 9 - Esquema representativo do processo do algoritmo hierárquico – abordagem aglomerativa.	28
Figura 10 - Representação dos métodos a) single linkage; b) complete linkage; c) average linkage [86].	28
Figura 11 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes	37
Figura 12 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o método PAM	37
Figura 13 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico	38
Figura 14 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes	39
Figura 15 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o método PAM	39
Figura 16 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico	40
Figura 17 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes	41
Figura 18 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o método PAM	41
Figura 19 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método <i>Average Silhouette</i> , da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico	42
Figura 20 – Visualização dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico.	45
Figura 21 - Visualização dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, para prostatectomia radical, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico.	47
Figura 22 - Visualização dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia ou prostatectomia radical, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico	49
Figura 23 - Estrutura da árvore de decisão dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, utilizando o algoritmo PAM.	55

página propositadamente em branco

ÍNDICE DE TABELAS

Tabela 1 - Mortalidade por Tumor Maligno da Próstata, em Portugal 2011 a 2017 [13].....	4
Tabela 2 - Escala de Gleason [19].....	5
Tabela 3 - Estadiamento Clínico do Tumor Primário (T) do Cancro da Próstata [19].....	6
Tabela 4 - Categorias do cancro da próstata com respetivas definições [2],[19].	7
Tabela 5 - Opções de tratamento para o cancro de próstata de baixo risco [2],[19].	8
Tabela 6 - Opções de tratamento para cancro da próstata de risco moderado [2].	9
Tabela 7 - Opções de tratamento para cancro da próstata de risco alto e de risco muito alto [2].	9
Tabela 8 - Comparação dos índices de multimorbilidade [34],[38],[40],[45],[46].....	12
Tabela 9 – Pacotes (<i>packages</i>) da Linguagem R utilizados.	24
Tabela 10 - Estatísticas descritivas do conjunto de dados analisado.....	34
Tabela 11 - Estatísticas descritivas do conjunto de dados de Braquiterapia e de Prostatectomia Radical	35
Tabela 12 - Frequência e frequência relativa das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante a faixa etária.....	36
Tabela 13 - Mediana e intervalo interquartil das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante o tempo de internamento.....	36
Tabela 14 - Frequência e frequência relativa das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante o número de óbitos.....	36
Tabela 15 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, <i>Average Silhouette</i> e o coeficiente de silhueta médio.....	38
Tabela 16 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, <i>Average Silhouette</i> e o coeficiente de silhueta médio.....	40
Tabela 17 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, <i>Average Silhouette</i> e o coeficiente de silhueta médio.....	42
Tabela 18 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia, utilizando diferentes abordagens de clustering.....	44
Tabela 19 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para prostatectomia radical, utilizando diferentes abordagens de clustering.....	46
Tabela 20 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia ou prostatectomia radical, utilizando diferentes abordagens de clustering.....	48
Tabela 21 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade.	50
Tabela 22 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar.....	51
Tabela 23 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade.....	51
Tabela 24 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade.....	52

Tabela 25 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares	52
Tabela 26 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital	53
Tabela 27 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser	54
Tabela 28 – Estatísticas segundo cada cluster de multimorbilidade obtido com o algoritmo PAM.....	55
 Tabela A 1 -Revisão da literatura	67
 Tabela B 1 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade, para o conjunto de dados de braquiterapia	77
Tabela B 2- Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar, para o conjunto de dados de braquiterapia.....	77
Tabela B 3 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade, para o conjunto de dados de braquiterapia	78
Tabela B 4 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade, para o conjunto de dados de braquiterapia.....	78
Tabela B 5 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares, para o conjunto de dados de braquiterapia.	78
Tabela B 6 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital, para o conjunto de dados de braquiterapia.....	79
Tabela B 7 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser, para o conjunto de dados de braquiterapia.	79
 Tabela C 1 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade, para o conjunto de dados de prostatectomia radical.	80
Tabela C 2 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar, para o conjunto de dados de prostatectomia radical.....	80
Tabela C 3 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade, para o conjunto de dados de prostatectomia radical.	81
Tabela C 4 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade, para o conjunto de dados de prostatectomia radical.....	81
Tabela C 5 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares, para o conjunto de dados de prostatectomia radical.	81
Tabela C 6 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital, para o conjunto de dados de prostatectomia radical.....	82
Tabela C 7 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser, para o conjunto de dados de prostatectomia radical.	82

página propositadamente em branco

LISTA DE SIGLAS

Lista de Siglas

ACG	Adjusted Clinical Groups
CCI	Chronic Condition Indicator
CCS	Clinical Classification Software
CCSR	Clinical Classifications Software Refined
CHU	Centro Hospitalar Universitário
CID	Classificação Internacional de Doenças
CID-9-MC	Classificação Internacional de Doenças da 9ª Revisão da Modificação Clínica
CIRS	Cumulative Index Illness Rating Scale
CLARA	Clustering Large Applications
DPOC	Doença pulmonar obstrutiva crónica
DRE	Digital Rectal Examination
EBRT	Radioterapia Externa
HBP	Hipertrofia benigna da próstata
HDRB	Braquiterapia de alta taxa de dose
HIFU	Ultrassom focalizado de alta intensidade
IA	Inteligência Artificial
ICD-9-CM	International Classification of Diseases, 9th revision, clinical modification
ICD-10-CM	International Classification of Diseases, 10th revision, clinical modification
ICD-10-PCS	International Classification of Diseases, 10th revision, procedure classification system
IIQ	Intervalo Interquartil
KDD	Knowledge Discovery in Databases
k-NN	k-Nearest Neighbors
LDRB	Braquiterapia de baixa taxa de dose
NUTSII	Nomenclatura das Unidades Territoriais para Fim Estatísticos
PAM	Partitioning Around Medoids
PCA	Análise de Componentes Principais
PSA	Antigénio específico da próstata
SIGIC	Sistema integrado de gestão de inscritos para cirurgia
SVM	Máquinas de Vetores de Suporte
TPA	Terapia de privação androgénica
TRUS	Transrectal ultrasound scan
WSS	Within-cluster sum of square

página propositadamente em branco

1 INTRODUÇÃO

Esta dissertação aborda um método de *machine learning* não supervisionado, nomeadamente a análise de clusters, para detetar padrões de multimorbilidade em pacientes internados com diagnóstico de cancro da próstata. O cancro da próstata é o tipo de cancro mais comum nos homens e a segunda causa mais comum de morte por cancro no mundo. Adicionalmente ao diagnóstico de cancro da próstata, uma percentagem substancial desses pacientes apresenta, pelo menos, uma comorbilidade adicional (também denominada de condição crónica), como diabetes, hipertensão e doenças cardiovasculares, que, por sua vez, repercute num risco superior de morte devido a outras causas além do cancro de próstata. A presença de múltiplas doenças crónicas, em simultâneo, no mesmo indivíduo é um problema de saúde reconhecido, que se denomina Multimorbilidade. Em relação ao tratamento do cancro, a multimorbilidade é uma preocupação particular, uma vez que a maioria dos pacientes com cancro apresentam doenças crónicas subjacentes em adição ao próprio cancro. As comorbilidades adicionais não só aumentam as hipóteses de *outcomes* negativos, como morte prematura e a utilização de serviços hospitalares não agendados, como aumentam a complexidade do atendimento. Este capítulo destina-se a introduzir o contexto da temática sobre o qual este trabalho se debruça, nomeadamente, abordando vários aspetos relacionados com o cancro da próstata e o conceito de multimorbilidade, para além de fundamentar a aplicação de *machine learning* na área da saúde e a análise de clusters. Segue uma especificação dos objetivos propostos e apresentação das opções metodológicas, finalizando-se com a estrutura da dissertação.

1.1 Enquadramento e pertinência

1.1.1 Cancro da Próstata

Anatomia

A próstata é uma glândula exócrina que integra o aparelho reprodutor masculino, cuja função é a produção de líquido seminal [1]. É um órgão sólido com estrutura ovoide, envolvido por uma cápsula fibrosa e composto por elementos fibrosos, glandulares e musculares, nomeadamente $\frac{2}{3}$ elementos glandulares e $\frac{1}{3}$ estroma fibromuscular. Numa pessoa adulta, em condições normais, tem aproximadamente 4cm de largura, 2cm de espessura e 3cm de comprimento, pesando cerca de 20 gramas [1],[2],[3]. O crescimento normal da próstata relaciona-se com o avanço da idade do homem.

A próstata encontra-se localizada na cavidade pélvica, atrás da face posterior da sínfise púbica, na base da bexiga, onde envolve a uretra prostática e os dois canais ejaculadores, superior à lâmina superior da fáscia do diafragma da pélvis e anterior ao reto [1],[4]. No seu interior, situa-se a parte prostática da uretra, os ductos ejaculatórios e o utrículo prostático, alojado numa loca fibrofascial, a loca prostática [1].

Segundo a análise histopatológica, a próstata apresenta quatro regiões ou zonas anatómicas distintas, sendo que a zona periférica representa aproximadamente 70% do volume da próstata (Figura 1-a). Esta zona rodeia a uretra distal, onde estão compreendidas a generalidade das glândulas prostáticas e, mais frequentemente, se originam os adenocarcinomas da próstata. A zona central é uma área em forma de cone que inclui a base da próstata e rodeia os ductos ejaculatórios, por onde o sémen transita para a uretra, compreende cerca de 25% do órgão, sendo 5% dos cancros da próstata decorrentes desta zona. A zona de transição, que se origina de ductos na uretra proximal, é o menor componente da próstata normal correspondendo apenas a 5 a 10% do volume da próstata, porém, geralmente, desenvolve um

aumento benigno ou hiperplasia benigna da próstata (HPB) [3],[4],[5]. Por último, a zona do estroma fibromuscular anterior consiste na região não glandular que representa um terço do volume da próstata. Localiza-se na face anterior da próstata entre as zonas periféricas [3].

Existe outro tipo de divisão da próstata, que a divide em cinco lobos prostáticos (Figura 1-b). O lobo anterior é o mais pequeno dos cinco e encontra-se localizado em frente da uretra, sendo bastante rico em estroma conjuntivo-muscular. O lobo posterior situa-se posteriormente à uretra, inferiormente aos ductos ejaculadores. O lobo médio localiza-se entre a uretra, abaixo da base da bexiga e acima dos ductos ejaculadores e do utrículo prostático. Os lobos laterais, esquerdo e direito, são os mais desenvolvidos e envolvem a uretra, os ductos ejaculadores e o utrículo prostático. Quando aumentam de volume, as relações diretas com o reto provocam alterações por toque rectal [1].

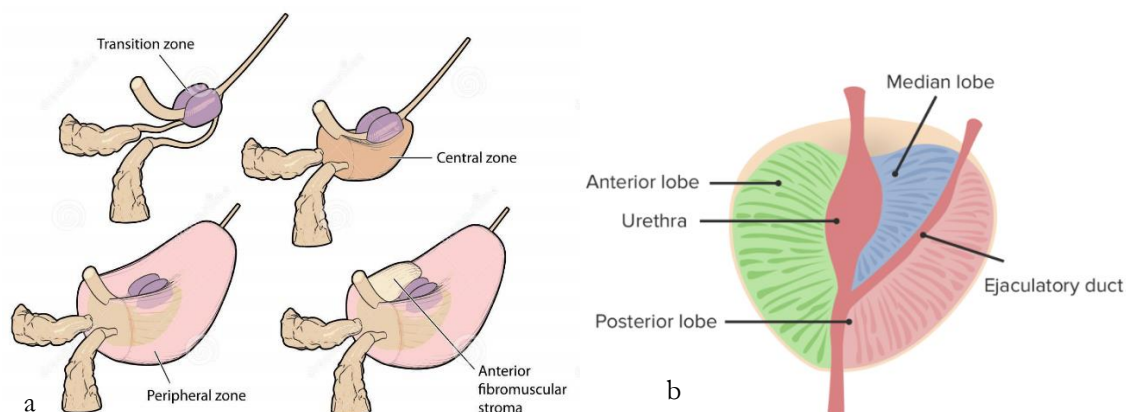


Figura 1- a) Zonas anatómicas da próstata; b) Lobos prostáticos [6].

O suprimento sanguíneo primário da próstata é fornecido pela artéria ilíaca interna e a drenagem venosa é realizada por uma rede venosa em torno da próstata, designada de plexo venoso prostático, a partir do qual o sangue é drenado para a veia ilíaca interna [5].

A drenagem linfática da próstata ocorre para os gânglios linfáticos ilíacos internos, ilíacos externos, obturadores e pré-sacrais, com drenagem ocasional direta para os gânglios linfáticos ilíacos comuns [2], [5]. Cerca de 85% da disseminação do cancro da próstata para os gânglios linfáticos ocorre nos gânglios linfáticos obturadores, ilíacos internos e externos [5].

A próstata é innervada por feixes neurovasculares ao longo da borda póstero-lateral da próstata [2]. Os gânglios autónomos estão agrupados perto dos feixes neurovasculares e pequenos troncos nervosos que surgem neste local penetram na cápsula para formar uma extensa rede dentro da glândula.

Etiologia

Existem alguns fatores de risco que parecem contribuir para o desenvolvimento do cancro da próstata. Apesar de diversos artigos referenciarem vários fatores ambientais, assim como relacionados com hospedeiro, associados a este cancro, ainda se desconhece quais os fatores que iniciam ou promovem diretamente a carcinogénese da próstata [5].

A idade, predisposição familiar, raça/etnia, mutações BRCA1/2 e síndrome de Lynch são considerados os fatores de risco mais prováveis para o diagnóstico deste cancro. Fatores como o tabagismo e obesidade estão diminutamente relacionados com o aparecimento desta doença, mas estão positivamente associados à sua mortalidade [7]. No entanto, consoante a literatura, são expostos outros fatores possíveis de contribuir para o aumento do risco [3].

O cancro da próstata é uma doença severamente impactada pelo envelhecimento, uma vez que a probabilidade do seu desenvolvimento aumenta consideravelmente com idade. Esta doença é rara em homens com menos de 45 anos e a sua incidência aumenta exponencialmente a cada década de vida, sendo que a maioria dos homens com cancro da próstata têm mais de 65 anos [8],[9]. A senescência do sistema imunológico, níveis mais elevados de stress oxidativo, exposição prolongada a agentes carcinogénicos e resposta inadequada a danos no DNA podem ser fatores explicativos para este acontecimento [5].

Diversos investigadores observaram clusters de famílias com indivíduos que possuíam cancro da próstata e sugeriram que a constituição genética pode aumentar a suscetibilidade à doença [5]. Estima-se que até 5% a 15% dos casos deste cancro são hereditários e o risco de desenvolver aumenta em duas vezes se existir um familiar em primeiro grau com cancro da próstata. Este risco aumenta entre cinco a onze vezes se existirem dois ou mais familiares em primeiro grau afetados [3],[7].

A variação na incidência e mortalidade do cancro de próstata em diferentes áreas geográficas é caracterizada pela acentuada diferença racial/étnica que depende de fatores geográficos [5],[7]. O carcinoma da próstata é mais frequente na raça negra, sendo que são mais frequentemente diagnosticados em idade mais precoce e parecem ter um prognóstico da doença pior comparativamente a caucasianos [5],[10].

No entanto, existe uma complexa interação entre raça/etnia, a dieta e meio ambiente desenvolvimento desta patologia. A incidência do cancro da próstata num determinado grupo racial/étnico tende a alterar, ficando equivalente à população presente na área geográfica para a qual esse grupo imigra, à medida que o grupo assimila o seu novo ambiente e cultura [5],[10].

A síndrome de Lynch resulta de uma mutação nos genes que reparam erros do material genético (MMR). A investigação de registos de cancro hereditário nos Estados Unidos da América revelou que homens com mutações nos genes MMR, característico da síndrome de Lynch, apresentam um risco geral duas vezes superior de cancro da próstata quando comparado com a população em geral, risco que é quase 2,5 vezes maior para homens com menos de 60 anos [3].

É conhecido que portadores de mutações nos genes supressores de tumor BRCA1 e BRCA2 têm um risco aumentado de desenvolver cancros do ovário e da mama comparativamente à população geral. No entanto, existem fortes evidências que mutações nestes genes também predispõem o cancro da próstata. Os portadores de uma mutação BRCA2 apresentam um risco até cinco vezes superior de desenvolver este tumor, aumentando para mais de sete vezes em portadores com história familiar de carcinoma da próstata de início precoce (<65 anos). Além disso, esta patologia tende a ser mais agressiva em indivíduos com esta mutação. Os portadores da mutação BRCA1 com menos de 65 anos apresentam um risco 3,8 vezes maior [3].

Todavia, vários indivíduos do sexo masculino com fatores de risco conhecidos não sofrem de cancro da próstata e, por outro lado, muitos dos que sofrem desta doença não apresentam qualquer destes fatores de risco (para além de terem mais de 65 anos).

Epidemiologia e Patologia

A taxa de incidência de cancro da próstata varia substancialmente entre as regiões e populações, sendo a neoplasia diagnosticada com mais frequência em 112 países, nos homens [7],[11]. Segundo os dados fornecidos pela GLOBOCAN, em 2020, existiram 1.41 milhões novos casos de cancro da próstata estimados em todo o mundo e 6.759 novos casos em Portugal, sendo a neoplasia maligna mais frequente nos homens neste país [11],[12]. As estimativas indicam que, em 2040, esse número ascenda a 8.216, o que corresponderá a um aumento de 21,6% [12]. Relativamente à mortalidade estima-se que, nesse

mesmo ano, tenham ocorrido cerca de 375.304 mortes em todo o mundo devido a este cancro, representando a 5ª causa de morte por cancro no homem [11],[12].

A saúde está diretamente relacionada com o nível de desenvolvimento social, económico e humano de um país. Apesar do cancro da próstata ser amplamente detetado globalmente, a sua incidência é mais comum na Europa, América do Norte e Oceânia. Pelo contrário, os países da Ásia e de África têm taxas de incidência menores para esta patologia, uma vez que os diagnósticos não são realizados com a mesma frequência que nos países mais desenvolvidos, de modo que os indivíduos acabam por falecer sem ter conhecimento da doença (Figura 2) [11],[12].

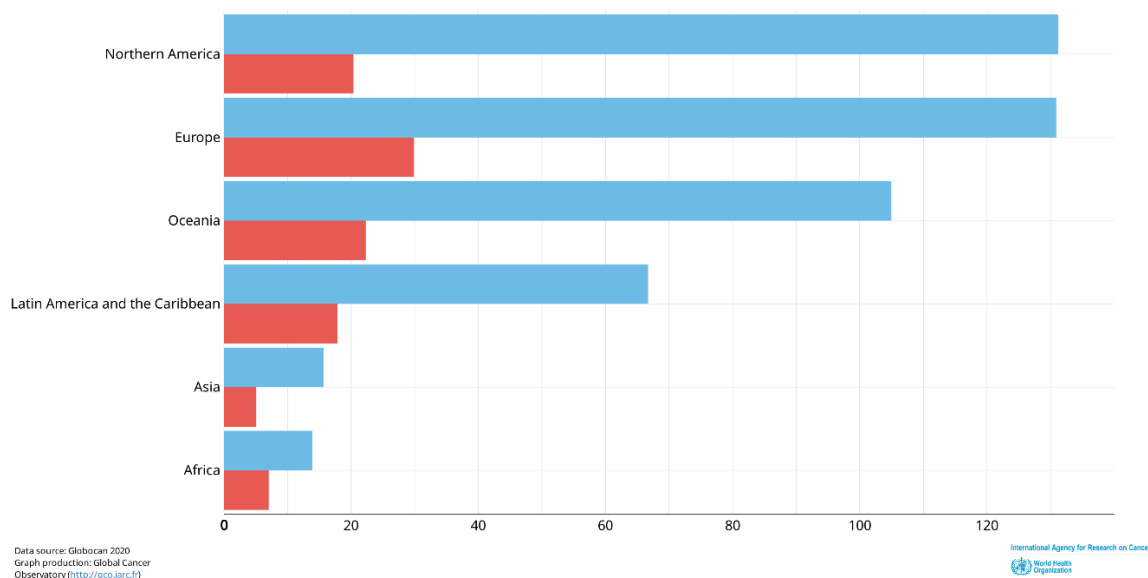


Figura 2 - Taxa de incidência e mortalidade do cancro da próstata por cada 100 mil habitantes, em 2020, consoante o continente [12].

Com base nos dados apresentados na Tabela 1, é possível concluir que, de 2011 a 2017, em Portugal, a mortalidade por neoplasia maligna da próstata tem-se mantido estável.

Tabela 1 - Mortalidade por Tumor Maligno da Próstata, em Portugal | 2011 a 2017 [13].

	2011	2012	2013	2014	2015	2016	2017
Número de Óbitos	1 815	1 806	1 714	1 787	1 723	1 836	1 794
Taxa Mortalidade Padronizada	22.3	22.1	20.3	20.8	19.4	20.4	19.4

Segundo um estudo realizado por Claudia Allemani et al em 2018, a taxa de sobrevivência ao cancro está a aumentar no mundo, sendo que, em Portugal, a taxa de sobrevivência de 5 anos para homens diagnosticados com esta patologia seja superior a 90% [14]. Contudo, apesar do aumento da taxa de sobrevivência graças ao progresso realizado pela ciência que nos permitiu ter uma melhor compreensão dos mecanismos moleculares e os fatores de risco envolvidos no cancro da próstata, este ainda permanece como a segunda causa principal de mortalidade por cancro nos homens em todo o Mundo [15].

Diagnóstico

Existirem inúmeros exames complementares que podem ser indicadores de cancro da próstata, sendo geralmente detetado através da combinação do toque retal (*Digital Rectal Examination- DRE*), níveis séricos de antígeno específico da próstata (PSA) e o exame transrectal por ultrassom (TRUS) da glândula da próstata. No entanto, o diagnóstico depende da confirmação histológica/ citológica no produto de biópsia ou na peça operatória [16].

Estudos sugerem que a taxa de sobrevida está intimamente relacionada com o estadio do tumor no momento do diagnóstico, apresentando 100% de sobrevida em cinco anos em pacientes diagnosticados com a doença em estadio inicial e menos de 33% de sobrevida em cinco anos, caso o diagnóstico seja de um estadio mais tardio, demonstrando assim a importância do diagnóstico precoce [17].

O toque retal é um método acessível, relativamente não invasivo e indolor, que demonstra mudanças na consistência da glândula. Em cerca de 18% dos casos, o carcinoma da próstata é detetado somente pelo toque retal suspeito, independentemente dos níveis de PSA [15]. Este exame tem valores estimados de 53,2% de sensibilidade, 83,6% de especificidade e um valor preditivo positivo de 17,8% [18].

O PSA é uma glicoproteína produzida quase exclusivamente pelas células do epitélio ductal da próstata. Este antígeno não é específico do cancro, mas órgão-específico, uma vez que pode estar elevado devido a hipertrofia prostática benigna (HBP), prostatite e outras condições não malignas. Porém, quando associado a outros exames e história clínica, é importante para definir grupos de risco para o cancro da próstata. Como variável independente, o PSA é um melhor preditor de cancro quando comparado com o DRE ou a TRUS [16].

A escala de Gleason é o sistema mais frequente na avaliação da graduação histológica, fundamentando-se na estrutura morfológica do cancro. A escala de classificação varia entre 1 (mais diferenciado) e 5 (menos diferenciado) e é a soma dos dois graus de Gleason mais dominantes numa amostra de tecido. Caso sejam verificados dois graus numa biópsia, o primeiro valor é o grau mais predominante e o segundo valor é o grau secundário, sendo que a pontuação final apresenta valores entre 2 e 10, sendo 10 o mais maligno. Nos dias de hoje, a escala de Gleason é dividida em cinco principais grupos de graus (Tabela 2) [19].

Tabela 2 - Escala de Gleason [19].

Grupo	Pontuação	Descrição
1	≤ 6	Composto apenas por glândulas discretas e bem formadas
2	$3 + 7 = 7$	Glândulas predominantemente bem formadas com menor componente de glândulas malformadas/fundidas
3	$4 + 3 = 7$	Glândulas predominantemente malformadas/fundidas com menor componente de glândulas bem formadas
4	$4 + 4 = 8$	Apenas glândulas malformadas/fundidas
	$3 + 5 = 8$	Glândulas predominantemente bem formadas e menos glândulas sem componentes
	$5 + 3 = 8$	Glândulas predominantemente ausentes e menor componente de glândulas bem formadas

5	9–10	Sem formação de glândulas (ou com necrose) com ou sem glândulas malformadas/fundidas
---	------	--

Estadiamento

O estadiamento é o processo pelo qual se verifica se ocorreu disseminação (metastização) do tumor para estruturas próximas ou para outras partes do corpo (metastização à distância). A informação obtida pelo processo de estadiamento determina a extensão (estádio) da doença, que é fundamental para planear o tratamento.

Para perceber a sua extensão, é necessário a análise histológica/ citológica dos tecidos recolhidos, assim como análises clínicas e exames imagiológicos. O estadiamento do carcinoma da próstata pode ser segundo a Tabela 3.

Tabela 3 - Estadiamento Clínico do Tumor Primário (T) do Cancro da Próstata [19].

Estadiamento do cancro	Definição
TX	Impossível avaliar o tumor primário
T0	Sem evidência de tumor primário
T1	Tumor não palpável, nem visualizável nos exames imagiológicos
T1a	Achado histológico incidental de tumor em $\leq 5\%$ do tecido ressecado
T1b	Achado histológico incidental de tumor em $> 5\%$ do tecido ressecado
T1c	Tumor identificado por biópsia por agulha, mas não palpável
T2	Tumor palpável e confinado dentro da próstata
T2a	Tumor não cresceu para mais de um dos lóbulos
T2b	Tumor envolve mais da metade de um lóbulo, mas não ambos
T2c	Tumor envolve ambos os lóbulos
T3	Tumor extraprostático, tendo atingido tecidos circundantes junto à próstata, vesículas seminais ou ambos, mas não envolvendo outros órgãos
T3a	Tumor não envolve as vesículas seminais
T3b	Tumor invade a(s) vesícula(s) seminal(is)
T4	Tumor invade estruturas adjacentes como esfíncter externo, reto, bexiga, músculos elevadores e/ou parede pélvica

Na estratificação por grupos de risco de recaída, a combinação de estágio clínico, escala de Gleason e valor do PSA permitem estratificar os doentes em categorias associadas a diferente risco de recorrência (Tabela 4).

Tabela 4 - Categorias do cancro da próstata com respetivas definições [2],[19].

Categorias	Definição
Baixo risco	cT1- cT2a PSA < 10 ng/ ml Escala de Gleason < 7
Risco moderado	cT2b–cT2c PSA 10-20 ng/ml Escala de Gleason 7
Risco elevado	cT3 Escala de Gleason 8-10 PSA > 20 ng/ml
Risco muito elevado	cT3b – cT4 Padrão primário de Gleason 5 2 ou 3 fatores de alto risco > 4 núcleos com grau de Grupo 4 ou 5

Tratamento

Nos últimos tempos, o desenvolvimento tecnológico tem permitido que a Medicina apresente uma evolução significativa, principalmente no diagnóstico e tratamento de doenças, como é o caso do cancro da próstata [20]. Não obstante, a prevenção da mesma apresenta um reduzido efeito na mortalidade relacionada à doença, sendo que o método ideal para o tratamento localizado do cancro da próstata ainda não se encontra bem definido [21]. Assim sendo, o uso de radioterapia, terapia hormonal, cirurgia e vigilância ativa permitem agir no procedimento terapêutico nos diferentes grupos de risco, isoladamente ou em combinação [21],[22]. Existem igualmente outras estratégias de tratamento, como quimioterapia, ultrassom focalizado de alta intensidade (HIFU) e criocirurgia, técnica pouco invasiva que congela o tecido cancerígeno, dependendo das condições clínicas. A escolha do tratamento depende do estadió da doença, do nível de PSA, da escala de Gleason, sinais clínicos, entre outros [21],[22],[19].

Uma das opções mais escolhidas pelos pacientes no início do tratamento é a prostatectomia radical. Esta permite a remoção da próstata, da cápsula circundante e das vesículas seminais, com ou sem linfadenectomia pélvica. Existem diferentes abordagens para esta técnica, como supra-púbica, perineal, transuretral e laparoscópica. A prostatectomia tem como consequência o risco de impotência e de, pelo menos, algum grau de incontinência urinária, sendo que existem técnicas que possibilitam a redução destes efeitos secundários. Contudo, a maior parte dos pacientes refere disfunção sexual significativa após esta técnica [19].

Seguidamente à prostatectomia radical, a radioterapia é classificada como a segunda maior opção terapêutica para o cancro de próstata, como a radioterapia externa (EBRT) e a braquiterapia. A radioterapia externa apresenta-se eficaz em casos sem metástases distantes e uma expectativa de vida de, pelo menos, 5 a 10 anos. A braquiterapia de baixa taxa de dose (LDRB) implica a inserção permanente de sementes radioativas com semi-vida de 60 dias sob a orientação de ultrassom. Estudos comprovam que o uso de prostatectomia radical e braquiterapia expressam resultados semelhantes no tratamento do cancro da próstata de baixo risco, mas diferentes sequelas de curto prazo. Este tratamento exige anestesia geral e um tempo de internamento de 2 a 3 dias. Na braquiterapia de alta taxa de dose (HDRB) existe uma inserção temporária de aplicadores na próstata, de forma a permitir uma alta dose de radiação na próstata com uma dose mínima na bexiga e no intestino. É considerada uma técnica de tratamento adequada para pacientes com doença mais avançada [21]. A maioria dos tratamentos de braquiterapia de alta taxa de dose podem ser realizados em ambulatório, ou seja, não necessitam de internamento.

As opções de tratamento existentes para o carcinoma da próstata segundo a estratificação por grupos de risco encontram-se descritas nas tabelas seguintes.

Tabela 5 - Opções de tratamento para o cancro de próstata de baixo risco [2],[19].

Opções de tratamento	Visão Geral	Prós	Contras
Vigilância ativa	Acompanhamento controlado com teste de PSA e biópsias repetidas Testes genómicos e biópsia guiada por ressonância magnética	Evita efeitos colaterais imediatos	Risco de progressão da doença
Prostatectomia radical	Aberta, laparoscópica ou robótica, com ou sem dissecação de gânglios linfáticos pélvicos	Remoção da totalidade do tumor/próstata Alívio de sintomas obstrutivos Obtenção do estadiamento patológico Evita a exposição à radiação	Risco operatório Aumento do risco de disfunção erétil (DE) e incontinência, comparativamente às opções não cirúrgicas
Radioterapia de intensidade modulada - Fracionamento convencional	74–80 Gy ao longo de 7–9 semanas	Procedimento <i>standard</i> , com acompanhamento de longo prazo	Extensa duração Potenciais efeitos secundários como cistite, proctite, disfunção erétil ou malignidade adicional
Radioterapia de intensidade modulada - Hipofracionamento moderado	60–70 Gy durante 4–5,5 semanas	Reduz o tempo de tratamento Custo reduzido em relação ao IMRT convencional	Tempo de tratamento moderadamente prolongado Potenciais efeitos secundários como cistite, proctite, DE ou malignidade adicional
Braquiterapia	Braquiterapia de baixa taxa de dose (LDR) com Iodo-125/Paládio-103 ou Braquiterapia de alta taxa de dose (HDR) com Irídio-192	Minimamente invasivo Acompanhamento a longo prazo disponível	Sintomas do trato urinário inferior (LUTS) agudo Potenciais efeitos secundários como cistite, proctite, DE ou malignidade adicional

Tabela 6 - Opções de tratamento para cancro da próstata de risco moderado [2].

Opções de tratamento
Vigilância ativa, caso a expectativa de vida seja inferior a 10 anos
EBRT ± TPA de curto prazo, ao longo de 4-6 meses
Braquiterapia
Prostatectomia radical: Considerar EBRT adjuvante para características adversas (+ margens, invasão das vesículas seminais, extensão extracapsular, PSA pós-operatório detetável)

Tabela 7 - Opções de tratamento para cancro da próstata de risco alto e de risco muito alto [2].

Opções de tratamento
EBRT ± TPA de longo prazo, ao longo de 2-3 anos
EBRT + Braquiterapia ± TPA de longo prazo
Prostatectomia radical (em pacientes selecionados): Considerar EBRT adjuvante para características adversas (+ margens, invasão das vesículas seminais, extensão extracapsular, PSA pós-operatório detetável). Se os gânglios linfáticos forem afetados, considerar TPA ± EBRT pélvica
Apenas TPA em pacientes que não são candidatos para terapia local

1.1.2 Detalhando o conceito de Multimorbidade

Com a evolução da medicina e com os notáveis desenvolvimentos na tecnologia médica, a esperança média de vida aumentou significativamente. À medida que a esperança média de vida expande, aumenta a incidência de pessoas com doenças crónicas complexas que interagem com fatores biológicos, psicológicos e sociais, e resultam num amplo espectro de manifestações e necessidades clínicas [23],[24]. A maioria dos doentes que sofrem de uma condição clínica crónica não possuem apenas uma, mas várias. Assim, surgiu o conceito de multimorbidade, comumente definida na literatura como a "coocorrência de duas ou mais doenças crónicas no mesmo indivíduo" [25]. Apesar da sua simples definição, é um problema relevante de saúde pública com uma padronização bastante complexa, devido à gama de doenças possíveis de serem mensuradas num indivíduo. Além disso, o seu conceito pode não se encontrar claramente diferenciado com comorbidade. Este último é definido como a existência ou ocorrência de qualquer entidade adicional distinta (doença ou fator de risco) adicional a uma doença em estudo – doença ou problema índice. No caso de multimorbidade não há, necessariamente, uma doença índice na sua caracterização [25]. Na atualidade, é amplamente reconhecido que a maioria das pessoas com alguma condição crónica tem maior probabilidade de apresentar multimorbidade do que ter apenas

uma condição, sendo a multimorbilidade, atualmente, a regra e não a exceção [26]. Na prática clínica, o termo é considerado um conceito mais centrado no paciente e que reflete mais adequadamente o *burden* da doença.

Estudos demonstraram que a multimorbilidade está associada a uma morte prematura do paciente, aumento da necessidade de recursos de saúde, complexidade do atendimento e redução da qualidade de vida do paciente [26],[27]. A multimorbilidade afeta a complexidade clínica associada à tomada de decisão. Essa complexidade resulta dos fatores ou elementos que têm de ser considerados, reconciliados e abordados aquando da tomada de decisões sobre uma estratégia terapêutica ou preventiva. Isso representa um desafio significativo na vida dos pacientes, no trabalho dos médicos de clínica geral e nos serviços de saúde em geral [28].

Estima-se que o número de pessoas com multimorbilidade deverá aumentar juntamente com o envelhecimento da população em >1% ao ano até 2030 [29]. A nível nacional, de acordo com os dados do Inquérito Nacional de Saúde, de 2005/2006, 36% da população tinha três ou mais doenças crónicas, sendo o sexo feminino mais afetado (42,5% vs 27,4%), tal como a população com 65 anos ou mais anos – (55,2% vs 14,8%) [3],[30]. Em 2011/2012, dados do SHARE (*Survey of Health, Ageing and Retirement in Europe*) demonstraram uma prevalência de multimorbilidade na população portuguesa, com 50 ou mais anos, de 39,4%. Este estudo também averiguou que o aumento do número de doenças crónicas se encontra associado a uma maior probabilidade de descrição do seu estado de saúde como razoável/desfavorável, depressão e capacidade funcional reduzida, sendo que esta associação se demonstrou bastante influente em Portugal. Outro estudo adicional realizado em 2013/2014, direcionado à população com 18 ou mais anos de idade que frequentava os cuidados de saúde primários em Portugal continental, revelou que 72,7% dessa população apresentava duas ou mais doenças crónicas. Além disso, constatou uma associação entre níveis educacionais baixos e rendimento económico insuficiente e maior prevalência de multimorbilidade. Estas percentagens seguem a mesma tendência no resto do mundo [23].

A incidência de cancro aumenta com a idade, assim como a incidência de multimorbilidade. Com o envelhecimento da população na maioria dos países desenvolvidos, as pessoas têm uma maior probabilidade de desenvolver cancro quando já apresentam outras condições crónicas. Para além disso, com os recentes avanços no diagnóstico e tratamento, as pessoas estão a viver mais tempo com cancro e, portanto, mais propensas a desenvolver outras doenças crónicas durante a vida, devido ao envelhecimento e aos tratamentos submetidos [26].

Pacientes com multimorbilidade e cancro podem enfrentar desafios adicionais no que diz respeito à qualidade dos cuidados oncológicos e necessitam uma maior atenção no atendimento médico, monitorização e incentivo dos cuidados centrados no paciente [31]. A presença da multimorbilidade impacta todas as etapas do tratamento, desde a prevenção e deteção precoce até ao tratamento no final da vida [26]. Relativamente à prevenção, esta é afetada pela multimorbilidade, na medida em que muitos dos fatores de risco, que podem ser alterados e encontram-se associados à prevenção do cancro como o, tabagismo, álcool, má alimentação, inatividade física, encontram-se igualmente associados a condições de longo prazo, como diabetes, doença pulmonar obstrutiva crónica (DPOC), doenças cardiovasculares, artrite e condições de saúde mental, cujos sintomas podem tornar a mudança de comportamento mais desafiante. Do mesmo modo, a deteção precoce do cancro pode ser influenciada pela multimorbilidade, positiva ou negativamente, sendo que o contato regular com os serviços de saúde por parte destes pacientes pode propiciar num diagnóstico precoce de cancro. No entanto, outros podem adquirir um diagnóstico tardio ao atribuírem a razão dos sintomas às comorbilidades [26]. Relativamente ao cancro da próstata, é estimado que 60% dos adultos mais velhos (> 65 anos) com esta patologia possuam multimorbilidade pré-existente [31].

1.1.3 Indicadores de Multimorbilidade

Com o crescente aumento do interesse na multimorbilidade, advém o aumento da literatura voltada para estudá-la nas suas diferentes dimensões: magnitude (e.g., contagem de diagnósticos), gravidade (e.g., índice de Charlson) ou padrão (e.g., agregados de diagnósticos) [32].

Existe uma elevada variabilidade relativamente à definição de multimorbilidade, o número de doenças consideradas e o perfil dessas doenças. Estas diferenças nas definições de terminologia apresentam-se como um desafio para encontrar uma forma de "medir" a multimorbilidade, isto é, encontrar um indicador a partir do qual seja possível prever a evolução do doente [33],[34],[35]. Além disso, as diferentes vertentes de "condições de saúde" contribuem para a proliferação de indicadores de multimorbilidade e para a incerteza na sua seleção, uma vez que alguns autores apenas avaliam condições crónicas e excluem fatores como a saúde mental, enquanto outros incorporam ambas condições, crónicas e agudas [33]. Existe igualmente bastante controvérsia sobre se o foco da multimorbilidade deveria ir além das condições crónicas para incluir fatores de risco e sintomas associados, como dislipidemia e incontinência, e fatores biopsicossociais, como o nível socioeconómico do paciente e as crenças e expectativas relativas à sua saúde [34]. Uma revisão sistemática, que englobou 74 estudos, aferiu que o número de comorbilidades consideradas variou de 4 a 147, com uma média aproximada de 20 doenças por estudo. As condições mais frequentemente consideradas são doenças cardiovasculares, diabetes, distúrbios músculo-esqueléticos e articulares, doenças respiratórias (doença pulmonar obstrutiva crónica (DPOC)), cancro e distúrbios de saúde mental [36].

Ademais ao número e tipos de condições incluídas, a sua avaliação depende em grande parte das fontes dos dados e dos métodos de recolha utilizados. As fontes de dados mais usuais incluem processos clínicos, base de dados administrativas, inquéritos de saúde, relatórios médicos e registos de saúde eletrónicos [33],[35].

Posto isto, como não existe uma medida universalmente estandardizada ou uma lista de doenças para definir multimorbilidade, vários índices foram desenvolvidos [37],[38]. Estes podem ser concebidos, por exemplo, para quantificar a multimorbilidade como uma covariável noutras análises (ex: cancro da próstata), para previsão de mortalidade ou para ajuste ao risco [37]. O ajuste ao risco é um conceito complexo que abrange fatores sociodemográficos do paciente, estabilidade clínica aguda, tendo sido elaborado como critério quantificador da gravidade da doença [39].

A medida mais comumente utilizada tem sido a contagem de doenças, definida como uma simples enumeração não ponderada do número de doenças por pessoa. Este método não reflete a experiência dos pacientes, a gravidade de cada condição e os efeitos resultantes da conjugação de diferentes doenças crónicas. No entanto, alguns índices combinam a contagem de doenças com a gravidade das condições médicas sobre um indivíduo, fatores fisiológicos ou demográficos, permitindo assim uma quantificação mais completa do *burden* das doenças. Os índices têm assim a vantagem de incluir vários parâmetros para quantificar diferentes aspetos da multimorbilidade e utilização de modelos sofisticados para previsão dos resultados [37].

Os índices mais utilizados e com evidências de validade mais significativas são o Índice de Charlson, Índice de Elixhauser, *Adjusted Clinical Groups (ACG) System* e *Cumulative Index Illness Rating Scale (CIRS)*. O índice de Charlson foi descrito pela primeira vez em 1987 e foi baseado em dados de mortalidade, num ano, de 607 pacientes admitidos num serviço hospitalar durante o período de um mês [40],[41],[42]. Este índice foi desenvolvido com base no risco relativo de morte e mede a gravidade da doença mediante o diferente impacto que as condições crónicas têm no prognóstico de um indivíduo [43]. O índice de Elixhauser foi desenvolvido em 1998 e funciona de forma semelhante ao índice de Charlson, distinguindo-se por incluir 30 comorbilidades (17 do índice de Charlson e 13 novas) [44],[43]. Em comparação com o índice anterior, no índice de Elixhauser, o peso é atribuído de igual forma a cada

condição, assumindo que todas as condições de saúde estão igualmente relacionadas com o desfecho [44]. Não obstante, estes dois índices só foram amplamente adotados a partir de 2005, quando o estudo de Quan et al permitiu que se calculassem estes índices com recurso a base de dados administrativas utilizando a Classificação Internacional de Doenças da 9ª Revisão da Modificação Clínica – CID-9-MC (classificação de diagnósticos e procedimentos que resulta da adaptação efetuada nos Estados Unidos da América da *International Classification of Diseases 9th Revision*, ICD 9 da OMS) ou do CID-10 [43].

Em 1968, foi criado um novo indicador, o *Cumulative Index Illness Rating Scale* (CIRS), que avalia 14 sistemas/órgãos, assim como o impedimento físico dos pacientes. Durante uma consulta médica ou revisão do processo clínico, os médicos classificam o “grau de gravidade” da condição para cada sistema do corpo (cardiovascular, respiratório, etc). As classificações são somadas e é obtida uma pontuação cumulativa da doença. Essa pontuação é útil para uma avaliação geral, contudo não apresenta especificidade clínica, uma vez que se concentra nos sistemas do corpo e não nas condições crónicas, especificamente [33].

O sistema ACG foi projetado pela *Johns Hopkins Bloomberg School of Public Health*, em 2009, para prever morbilidade futura e a utilização de recursos de saúde. A maioria dos estudos recorre a este sistema com modelos preditivos para aferir uma variedade de custos ou resultados de processos associados à multimorbilidade [40].

A Tabela 8 apresenta uma descrição dos índices de multimorbilidade mencionados.

Tabela 8 - Comparação dos índices de multimorbilidade [34],[38],[40],[45],[46].

	Descrição	Resultados	Limitações
Contagem de doenças	Baseado no número total de condições crónicas que um indivíduo possui; Recorre a registos médicos, diagnóstico clínico ou questionários e entrevistas.	Qualidade de vida.	Falta de uniformidade na lista de condições crónicas considerada; Não considera a gravidade de cada condição, fazendo com que a complexidade da multimorbilidade não seja adequadamente avaliada.
Índice de Charlson	19 condições crónicas. A cada condição é atribuída uma pontuação de 1, 2, 3 ou 6, dependendo do risco de morte a um ano associada a cada uma. A soma das respetivas pontuações dá-nos o risco final; Utiliza registos de saúde eletrónicos e questionários de autorrelato; Existem várias adaptações para dados administrativos.	Afere a presença de multimorbilidade (e.g., índice de três ou mais pontos significa grave comorbilidade) e prevê a mortalidade (e.g., índice de cinco ou mais pontos é preditor de morte a três anos, em 85% dos doentes).	Erros de codificação (particularmente quando se baseia em códigos ICD-9); Historial clínico incompleto.
Índice de Elixhauser	30 condições crónicas. Cada uma com peso de 1 ponto; Uso de dados administrativos.	Permite prever a mortalidade hospitalar, custo e duração do internamento, eventos adversos e altas hospitalares.	O maior número de condições crónicas dificulta o seu uso, onde o tempo é limitado.

<i>Adjusted Clinical Groups (ACG) System</i>	Sistema de <i>case-mix</i> que agrupa indivíduos com base no seu perfil de saúde, nomeadamente idade, sexo e diagnósticos clinicamente coerentes; Através de registos médicos ou dados de reivindicações de seguro.	Possibilita a obtenção do valor atual ou esperado de consumos dos serviços de saúde, tendo em consideração as morbilidades do indivíduo.	Não é <i>open source</i> ; Depende da fiabilidade dos registos ou dos dados do seguro.
<i>Cumulative Index Illness Rating Scale (CIRS)</i>	Baseia-se na apreciação global da situação do doente e não nos diagnósticos individuais, não necessitando de lista de condições crónicas pré-selecionada. Este índice é estruturado em 14 sistemas/órgãos, que são classificados (0–4) segundo a gravidade e somado para obter uma pontuação total (0–56); Utiliza registos médicos.	Prevê o “ <i>medical burden</i> ”	Os prognósticos variam consoante o tipo de condição, podendo ter melhorado desde que a criação do índice.

1.1.4 Machine learning

A abundância de dados é uma realidade atual, em particular em ambiente hospitalar, visto que gera um grande e complexo volume de dados a partir de diversas fontes, como base de dados administrativas, processos clínicos, relatórios laboratoriais, registos de saúde eletrónicos, prescrições, aparelhos médicos, entre outros. Tanto a quantidade como a complexidade da informação implícita e distribuída em diferentes segmentos de dados dificultam a extração de conhecimento que permita apoiar a tomada de decisão [47],[48].

O processo *Knowledge Discovery in Databases* (KDD), em português designado por Descoberta de Conhecimento em Bases de Dados, integra teorias, métodos e algoritmos provenientes de diferentes áreas, para compreensão e extração de conhecimento útil a partir de grandes bases de dados. Este é composto por várias etapas operacionais, nomeadamente a seleção, pré-processamento, transformação, *data mining* e interpretação/avaliação [49].

Data Mining é o passo mais relevante de todo o processo de extração de conhecimento, sendo que a Descoberta de Conhecimento em Bases de Dados é, muitas vezes, denominada apenas por este conceito [49]. Foi definida, pela primeira vez, nos anos 90 como um processo de identificação matemática de padrões, geralmente através de grandes quantidades de dados [47],[50]. Apresenta assim o intuito de descobrir correlações, irregularidades e padrões, eventualmente implícitos ou desconhecidos, para previsão de resultados. *Data mining* recorre à modelação estatística para representar dados através de modelos matemáticos bem definidos e, por conseguinte, usar esse modelo para identificar relações e tendências de comportamentos entre as variáveis de dados [51].

As técnicas tradicionais de *data mining* demonstram-se insuficientes para lidar com os grandes volumes de dados atuais, seja devido à heterogeneidade e redundância dos dados provenientes de diversos registos

não padronizados ou a insuficiente qualidade destes. Desse modo, surgiu o desenvolvimento de novas técnicas de *data mining* baseadas em disciplinas como a Inteligência Artificial e *Machine Learning* [51].

A Inteligência Artificial (IA) é o domínio do conhecimento que objetiva o desenvolvimento de sistemas computacionais para solucionar problemas, oferecendo-lhes poderes cognitivos para efetuar tarefas que requerem inteligência humana [52]. *Machine learning* é uma das áreas da IA em que os sistemas de computador aprendem e melhoram automaticamente a partir da experiência, oferecendo a capacidade da IA para resolver problemas com base nos dados disponíveis [52],[53].

A inclusão de *machine learning* beneficia o processo de *data mining*, permitindo obter mais conhecimento sobre o processo de tomada de decisão, processar grandes *datasets* com maior precisão e eficiência e a capacidade de aprender e melhorar continuamente com novos dados [51]. Esta área tem alterado a maneira como as hipóteses para a investigação são formuladas, dado que, em vez de baseadas em teoria, a *machine learning* facilita a formulação de hipóteses baseadas em dados [54]. A aplicação de métodos de *machine learning* e *data mining* na pesquisa de multimorbilidade é uma abordagem fundamental para a utilização de grandes volumes de dados disponíveis para compreender de forma mais precisa as relações entre as doenças e, portanto, um tratamento mais direcionado e personalizado.

No desenvolvimento de um modelo de previsão utilizando *machine learning* é utilizado um conjunto de dados de treino para fazer previsões baseadas nos dados. É possível identificar três formas distintas de aprendizagem: aprendizagem supervisionada, onde os dados de treino dos algoritmos incluem as soluções desejadas, chamadas de labels, não supervisionada e aprendizagem por reforço [49].

Como o nome sugere, a aprendizagem supervisionada envolve a supervisão externa no treino de um modelo [51]. Neste tipo de aprendizagem as tarefas podem ser agrupadas em dois tipos: classificação, onde o algoritmo classifica uma previsão, ou seja, quando a saída é uma categoria; e regressão, onde o algoritmo tenta prever um resultado numérico. Os algoritmos considerados mais importantes e utilizados em problemas de aprendizagem supervisionada são: Árvores de decisão, Regressão Linear, Regressão logística, *k-Nearest Neighbors* (k-NN), Máquinas de Vetores de Suporte (SVM), *Random Forests* e Redes Neurais [49].

Por outro lado, os sistemas de aprendizagem de máquina não supervisionados são baseados num conjunto de dados desconhecidos, onde o algoritmo é responsável por descobrir padrões num novo conjunto de dados ou associações entre as variáveis [49],[51]. A utilização deste tipo de aprendizagem tem a vantagem de obter conhecimento sobre os dados dos quais não há qualquer tipo de informação antecipada.

O Clustering é uma técnica de *machine learning* não supervisionada que distribui dados em grupos consoante as suas semelhanças. Em comparação, a análise de componentes principais e a análise fatorial são ferramentas não supervisionadas de redução de dimensionalidade, que permitem simplificar e reduzir a complexidade dos dados através da diminuição do número de variáveis consideradas [55].

1.1.5 Análise de clusters

A análise de clusters, também denominada por Clustering, define-se como o processo de agrupar dados, com base nas características intrínsecas percebidas e atributos semelhantes, de modo que a obter grupos que sejam os mais homogêneos e bem separados possíveis [56],[57]. Um cluster é um conjunto de objetos similares entre si dentro do mesmo cluster e dissimilares em relação a objetos noutros clusters [58]. Webster (2008) descreve o clustering como “uma técnica de classificação estatística para identificar se os

indivíduos de uma população se enquadram em grupos diferentes, através de comparações quantitativas de múltiplas características” [59].

Segundo Ikotun, a análise de clusters apresenta como objetivo garantir maior similaridade intra-cluster e menor similaridade inter-clusters [57]. É uma técnica de aprendizagem não-supervisionada, uma vez que a classe que cada um dos objetos considerados integra não se encontra previamente atribuída, e de natureza exploratória ou descritiva, tendo como intuito compreender as características gerais e encontrar a estrutura subjacente aos dados através da descoberta de padrões e da exploração das relações escondidas nos dados em bruto. Como um algoritmo comum de *machine learning* não supervisionado, o clustering é amplamente utilizado em data mining e análise de dados [60],[61].

A análise de clusters foi introduzida por Driver e Kroebe, em 1932, no campo da antropologia, e mais tarde instituída na área da psicologia por Zubin, em 1938, e Tryon em 1939 [62], vivenciado uma constante evolução, nos tempos atuais, com a proposta e implementação de vários algoritmos para a resolução de problemas de agrupamento de dados. Relativamente à sua aplicabilidade, têm contribuído para diversas áreas de conhecimento, para obtenção de resultados com qualidade significativa, tais como a medicina, robótica, setor financeiro, inteligência artificial, marketing, etc [57].

No âmbito da multimorbidade, o clustering é relevante para a compreensão das combinações de condições crónicas, uma vez que afetam diferencialmente o *outcome* da saúde e da mortalidade do paciente. Assim, é possível identificar as mesmas patologias e fatores de risco comuns em grupos de indivíduos a fim de otimizar e agilizar um atendimento centrado no paciente, focado na multimorbidade. Estratégias para auxiliar os profissionais de saúde a priorizar a multimorbidade em pacientes com cancro são particularmente cruciais na medida em que a causa da morte entre doentes oncológicos é mais presumível ser ocasionada por condições não relacionadas ao cancro [63].

1.2 Questão e objetivos de investigação

O presente estudo permitirá, através da análise de clusters em pacientes internados com cancro da próstata nos hospitais portugueses, entre 2011 e 2017, identificar e caracterizar padrões de multimorbidade. Com esta informação pretende-se identificar os diferentes perfis de multimorbidade que caracterizam os internamentos por cancro da próstata a nível nacional, assim como identificar potenciais grupos de risco mais suscetíveis a *outcomes* negativos, como o óbito intra-hospitalar e o tempo de internamento. Procura-se também relacionar os diferentes perfis de multimorbidade identificados com outras variáveis clínicas e sociodemográficas, relacionadas com a gestão dos cuidados de saúde.

No contexto do cancro da próstata, o uso da análise de clusters poderá ser útil para identificar e caracterizar os diferentes perfis de doentes internados, para além de descobrir potenciais grupos de risco que requerem cuidados de saúde acrescidos. Dito isto, a informação obtida com esta análise poderá ser aplicada na gestão do tratamento para desenvolver ações de forma customizada, com o intuito de complementar as duas vias mais comuns de tratamento do cancro da próstata em Portugal (braquiterapia e prostatectomia radical), agregando uma manutenção mais eficaz da doença crónica e planeamento de cuidados de saúde e, consequentemente, auxiliar na definição de futuras intervenções orientadas para a promoção de comportamentos mais saudáveis e controlo de fatores de risco.

Para a prossecução do objetivo geral desta investigação, desenvolveram-se os seguintes objetivos específicos:

- Aplicar métodos de análise de clusters a dados reais de episódios de internamento sucedidos em hospitais públicos de Portugal, entre o ano 2011 e 2017, e comparar os resultados da aplicação de cada técnica estudada no processo de clustering (agrupamento);

- Apresentar métodos de validação interna que têm como objetivo avaliar a qualidade e estabilidade dos clusters obtidos pelos diferentes algoritmos de análise de clusters;
- Interpretar os clusters identificados e avaliar a relevância dos mesmos na perspetiva clínica e da gestão dos cuidados de saúde.

1.3 Opções metodológicas

Para se atingir os objetivos referidos anteriormente deste trabalho, foram consideradas algumas opções metodológicas. Estudos anteriores recorreram a métodos de *machine learning* não supervisionados, como a análise de clusters, para detetar padrões de multimorbilidade em pacientes com cancro utilizando dados hospitalares. Como referido anteriormente, a análise de clusters consiste no agrupamento de episódios de acordo com propriedades comuns. Este procedimento tem como propósito criar grupos homogêneos, nos quais a variabilidade dos episódios em relação às variáveis consideradas (as comorbilidades) seja mínima dentro de um cluster. Ao mesmo tempo, procura maximizar a variabilidade entre os clusters, ou seja, garantir que os grupos formados sejam distintos uns dos outros em termos de perfis clínicos. Essa abordagem visa, portanto, explorar as relações entre as comorbilidades dos pacientes com cancro da próstata, permitindo a identificação de grupos específicos com padrões de multimorbilidade semelhantes. Foram identificados métodos de clustering utilizados na literatura para encontrar padrões de multimorbilidade em dados hospitalares de pacientes com cancro, dos quais foram selecionados os mais aplicáveis a dados categóricos e adaptados a conjuntos de dados de grandes dimensões, especificamente os algoritmos: k-modes, clustering hierárquico e PAM (*Partitioning Around Medoids*).

1.4 Estrutura do trabalho

A estrutura do presente trabalho organiza-se em cinco capítulos e encontra-se definida a partir de um capítulo de «Introdução» (Capítulo 1), na qual é possível compreender o enquadramento teórico deste estudo. Inicialmente, é abordado o cancro da próstata, descrevendo sua anatomia, etiologia, epidemiologia, diagnóstico e tratamento. Na secção seguinte, é discutido o conceito de multimorbilidade, como este conceito tem sido definido na literatura, destacando-se a coexistência de múltiplas doenças crónicas num indivíduo e seu impacto no cuidado do cancro. Além disso, é caracterizada a aplicação de *machine learning* na área da saúde e a análise de clusters como uma técnica de agrupamento de dados. Este capítulo é constituído pela introdução, objetivo, opções metodológicas e a estrutura da presente dissertação.

O segundo capítulo concentra a «Revisão Bibliográfica» e é composto por uma síntese dos estudos mais pertinentes no que respeita o tema deste estudo. Este capítulo apresenta o levantamento da metodologia utilizada pelos artigos analisados.

O terceiro capítulo apresenta um elemento central desta dissertação, a descrição dos algoritmos de clustering, necessária para uma boa compreensão do trabalho apresentado ao longo deste estudo. Nas secções seguintes, são referidos quais os algoritmos adotados no contexto deste projeto e o motivo da sua seleção. São igualmente apresentadas algumas generalidades sobre as medidas de distância, os métodos eleitos para a seleção do número ideal de clusters e o critério de validação e avaliação da qualidade dos resultados obtidos.

No quarto capítulo, realiza-se uma análise dos resultados atingidos e avalia-se o desempenho dos métodos propostos, quando empregues aos conjuntos de dados em estudo. Além disso, são discutidos os resultados obtidos de acordo com o critério de validação selecionado, tal como a sua interpretação.

A dissertação finaliza com a apresentação das conclusões obtidas e das limitações sentidas ao longo da elaboração deste trabalho, assim como propostas para futuros desenvolvimentos.

2 REVISÃO BIBLIOGRÁFICA

A revisão da literatura teve em vista a recolha de informação sobre a investigação disponível e pertinente de diversas técnicas de análise de clusters, utilizadas para identificar padrões de multimorbilidade. Também é descrito o processo de revisão da literatura utilizado. Este capítulo revela, por um lado, quais as condições crónicas mais ponderadas para o estudo da multimorbilidade e os padrões de multimorbilidade mais identificados. Por outro lado, descreve as principais características metodológicas dos estudos e quais os algoritmos de cluster que estão a ser amplamente referenciados na literatura.

2.1 Metodologia de pesquisa

No Apêndice A encontra-se a tabela da revisão da literatura efetuada, com informações mais detalhadas acerca dos estudos analisados. Para tal, os critérios de inclusão dos artigos basearam-se nos seguintes parâmetros: artigos publicados a partir de 2013 (≤ 10 anos), com idioma em português ou inglês e que abordassem os tópicos “multimorbilidade”, “hospital” ou palavras derivadas e “grupos”, “subgrupos”, “padrões” ou “cluster”.

Foram identificados 609 artigos no total, após a verificação de todos os critérios de inclusão acima referidos, dos quais 93 permaneceram após a triagem dos títulos. Após a análise dos abstratos, foram revistos os textos completos de 20 artigos. A aplicação do último critério de exclusão resultou na inclusão final de 14 referências (Figura 3).

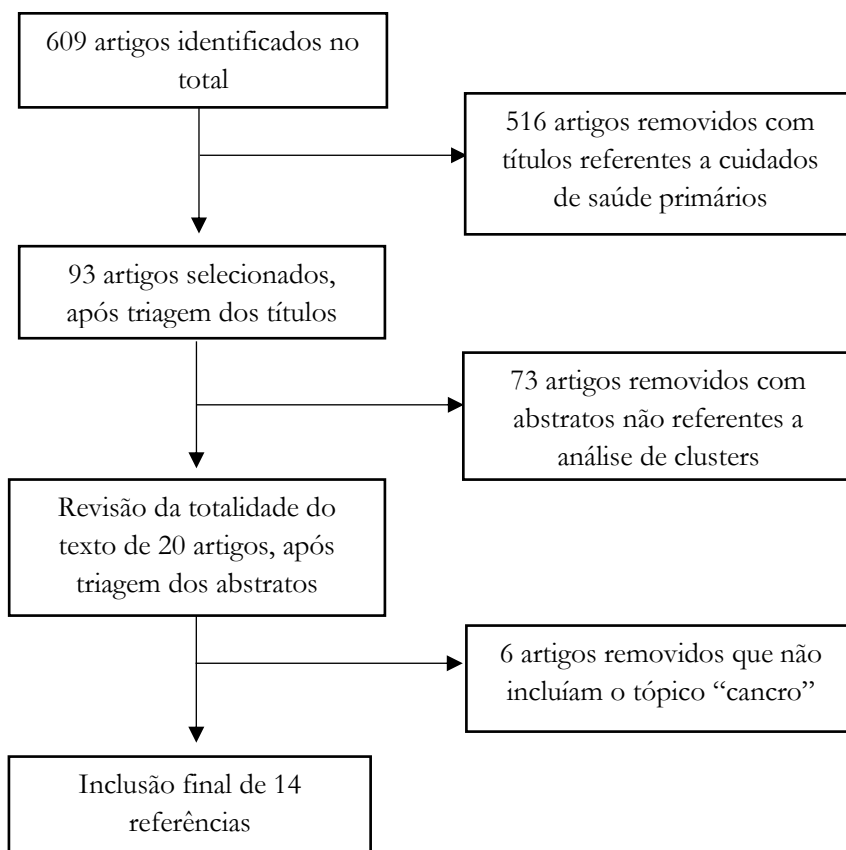


Figura 3 – Fluxograma da revisão bibliográfica

2.2 Análise dos artigos selecionados

No geral, os artigos exibiram um alto grau de heterogeneidade metodológica em termos de tamanho da amostra, idade e seleção dos participantes do estudo, fonte de dados, número de doenças de base consideradas e método de clustering utilizado.

Os números de participantes incluídos nos estudos variam entre 219 e 3.872.451. Em 10 dos 14 estudos, apenas participaram indivíduos maiores de idade (>18 anos), sendo que a maioria desses estudos apresentava apenas indivíduos com mais de 50 anos [64],[65],[66],[67],[68],[69]. Dos restantes artigos, somente 1 deles apresentava apenas participantes menores de idade (<18 anos) [70] e os 2 restantes estudos apresentava uma grande ambiguidade de idades, podendo possuir tanto adolescentes como idosos [71],[72].

Em termos de recrutamento da população do estudo, 10 estudos selecionaram os seus participantes a partir de registos médicos e clínicos, como registos de internamento e alta hospitalar, enquanto os restantes utilizaram dados administrativos para a recolha de participantes [63],[64],[68],[70].

O número de doenças incluídas nos estudos variou entre 3 e 308, sendo que um dos estudos não apresentou informações acerca do número de condições crónicas [66] e outro estudo considerou as condições definidas nos algoritmos *Agency for Healthcare Research and Quality Clinical Classification Software* (CCS) e *Chronic Condition Indicator* (CCI) [70].

Em relação à codificação, 12 dos 14 artigos codificaram as doenças com base na Classificação Internacional de Doenças, versões CID-9 ou CID-10 [63],[64],[65],[67],[68],[69],[70],[71],[72],[73],[74],[75]. Relativamente aos restantes, 1 artigo foi codificado de acordo com o algoritmo *phenotyping* [76] e outro não apresentou dados sobre a codificação [66].

Foram utilizados inúmeros algoritmos de clustering, como análise de correspondência múltipla [64],[72] análise de classes latentes [65],[70],[75], clustering hierárquico aglomerativo e k-modes, *Bernoulli Latent Block Model* [66], *Partitioning around Medoids* [68] e, por fim, *Network analysis* [76].

Por último, o número de clusters obtidos variou de 3 a 23 grupos distintos, sendo que 3 dos artigos não apresentaram dados relativos a esse parâmetro [63],[69],[76].

3 MÉTODOS E APLICAÇÃO

Previamente à apresentação dos resultados das experiências, será exposta a metodologia adotada para a obtenção desses resultados. A metodologia incorpora um conjunto de fases cujas etapas diferem de acordo com o algoritmo de clusters empregue. Para tal, é descrita a seleção das variáveis e as medidas de dissimilaridade, assim como o enquadramento teórico dos algoritmos de clusters utilizados. Por último, apresenta-se uma exposição dos métodos selecionados para a escolha do número ótimo de clusters e o critério de validação e avaliação da qualidade dos resultados obtidos.

3.1 Base de dados, população do estudo e definição de variáveis

Para a metodologia desenvolvida no âmbito deste trabalho de investigação, foram utilizados dados administrativos obtidos da Base de Dados Nacional de Morbilidade Hospitalar (BDMH), disponibilizados pela Administração Central do Sistema de Saúde (ACSS). A base de dados mencionada contém dados de todos os episódios de internamento e ambulatório sucedidos em hospitais públicos do Serviço Nacional de Saúde português. O universo de estudo foi restringido a episódios de internamento com diagnóstico principal de cancro da próstata e com alta hospitalar adquirida entre 1 de janeiro de 2011 e 31 de dezembro de 2017. Foram apenas extraídos os episódios rotulados na base de dados como estatisticamente válidos, isto é, episódios com tempo de permanência do utente no hospital de, pelo menos, 24h ou menor que 24h para pacientes que faleceram, saíram contra orientação médica ou foram transferidos para outro hospital. A partir desses dados, a população em estudo foi circunscrita a episódios de internamento por prostatectomia radical ou braquiterapia, as duas vias de tratamento principais para pacientes com cancro da próstata nos hospitais portugueses. Como os dados utilizados neste estudo foram previamente desidentificados pela ACSS, não foi necessária aprovação ética.

A arquitetura do processo metodológico encontra-se ilustrada na Figura 4.

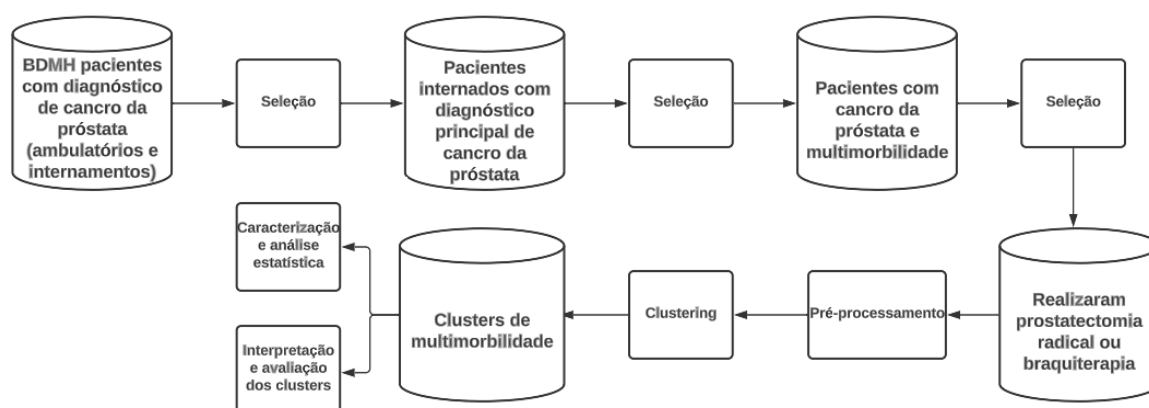


Figura 4 - Esquema do processo metodológico utilizado

Para o prosseguimento desta metodologia foi necessária a seleção das variáveis. Em adição às variáveis eleitas da base de dados mencionada anteriormente, foi produzida uma variável adicional para representar o diagnóstico principal de acordo com o *Clinical Classification Software* (CCS) [77], projetado para agrupar os códigos do *International Classification of Diseases, 9th revision, clinical modification* (ICD-9-CM), e com o *Clinical Classifications Software Refined* (CCSR) [78], projetado para agrupar os códigos do *International Classification of Diseases, 10th revision, clinical modification* (ICD-10-CM). Ademais, foram criadas

duas variáveis, uma para identificar o grupo hospitalar onde sucedeu o episódio de internamento e a outra a sua região. Assim como as anteriores, a variável “Índice de Elixhauser” não constava na base de dados utilizada. As variáveis seleccionadas para este estudo foram as seguintes:

- **Idade** – Variável quantitativa contínua - Idade do utente.
- **Dias de internamento** – Variável quantitativa discreta - Representa o número total de dias do tempo de permanência do utente no hospital.
- **CCS** – Corresponde a um número retirado de um sistema de classificação de episódios de internamento, sendo este denominado *Clinical Classification Software* para dados codificados em ICD-9-CM ou *Clinical Classification Software Refined* para dados em ICD-10-CM. Essa classificação foi desenvolvida pela *Agency for Healthcare Research and Quality* para agrupar os códigos da *International Classification of Diseases* em 275 grupos significativos e mutuamente exclusivos, relacionados com o diagnóstico principal. Neste enquadramento, foram utilizados o código ICD-10-CM = "C61" e código ICD-9-CM = "185", correspondentes ao diagnóstico de cancro da próstata.
- **Grupo Hospitalar** - Variável qualitativa nominal - Categoria do hospital, tal como definido pela Administração Central do Sistema de Saúde [23]. Esta categoria inclui instituições presentes no site de Benchmarking Hospitais da ACSS (Grupos B, C, D, E e F), que são semelhantes a nível da capacidade/complexidade.
- **Região** – Variável qualitativa nominal - Identifica a região hospitalar de acordo com a divisão NUTSII:
 - 1) Norte;
 - 2) Centro;
 - 3) Lisboa e Vale do Tejo;
 - 4) Alentejo;
 - 5) Algarve.
- **Tipo de admissão** - Variável qualitativa nominal - Natureza ou modo de admissão de um utente num estabelecimento de saúde:
 - 1) Programada;
 - 2) Urgente;
 - 3) Acesso;
 - 4) Peclec;
 - 5) Medicina Privada;
 - 6) SIGIC;
 - 7) Plano de Acesso à Cirurgia Oftalmológica;
 - 8) SIGIC Externo (desde 2012).

Os códigos 1,3,4,6 e 12 foram agrupados como “Programado”.

- **Peso Relativo** - Variável quantitativa contínua - Indica a complexidade da causa do internamento, representando uma medida do custo esperado desse episódio relativamente ao episódio médio nacional.
- **Severidade** – Variável qualitativa ordinal - Nível de severidade da doença:
 - 1) Menor;
 - 2) Moderado;
 - 3) Major;
 - 4) Extremo.
- **Mortalidade** – Variável qualitativa ordinal - Risco de mortalidade:

- 1) Menor;
 - 2) Moderado;
 - 3) Major;
 - 4) Extremo.
- **Procedimento médico** - Procedimento médico representado por um código da lista ICD-9-CM ou ICD-10-PCS (*International Classification of Diseases, 10th revision, procedure classification system*). Para este contexto, em particular, foram somente considerados episódios que realizaram braquiterapia (códigos ICD-9-CM: 92.20, 92.27, 92.28 ou 92.29 e códigos ICD-10: 0VH031Z ou DV10BYZ) ou prostatectomia radical (códigos ICD-9-CM: 605 e códigos ICD-10: 0VT00ZZ ou 0VT04ZZ).
 - **Comorbilidades** – Variável binária que identifica a presença ou ausência de 30 comorbilidades, definidas de acordo com o método de Elixhauser. As variáveis incluídas, referentes a cada comorbilidade de Elixhauser, foram as seguintes: Insuficiência cardíaca congestiva, Arritmias cardíacas, Doença vascular, Distúrbios de circulação pulmonar, Distúrbios vasculares periféricos, Hipertensão, Paralisia, Outros distúrbios neurológicos, Doença Pulmonar Obstrutiva Crônica, Diabetes Mellitus sem complicações, Diabetes Mellitus com complicações, Hipotireoidismo, Doença renal, Doença hepática, Úlcera péptica, SIDA, Linfoma, Tumor metastizado, Tumor sólido não metastizado, Artrite Reumatoide/Colagenoses, Coagulopatia, Obesidade, Perda de peso, Desequilíbrio hidroeletrólitos, Anemia provocada por perdas de sangue, Anemia por défice, Abuso de álcool, Abuso de drogas, Psicose, Depressão.
 - **Índice de Elixhauser** – O Índice de Elixhauser foi identificado de acordo com o trabalho de van Walraven C et al [79]. Neste é determinada a associação independente de cada comorbilidade de Elixhauser com a morte hospitalar, utilizando um modelo de regressão logística múltipla. As estimativas obtidas pelo modelo de regressão foram utilizadas para a atribuição de um peso para cada comorbilidade Elixhauser. Por exemplo, a associação de uma variável com 2 pontos é duas vezes mais “forte” que uma variável com 1 ponto, mas apresenta metade do impacto de uma variável com 4 pontos.

3.2 Medida de Multimorbilidade

A multimorbilidade foi definida como os episódios com ≥ 2 condições crônicas, incluindo o próprio diagnóstico de cancro da próstata. Como referido anteriormente, um total de 30 condições crônicas foram identificadas através do método de Elixhauser [44]. Essas condições foram reconhecidas segundo uma lista predefinida de códigos ICD-9 e ICD-10, que foram pesquisados nos campos do diagnóstico secundário. A lista completa de códigos específicos ICD-9 e ICD-10 para as 30 condições de Elixhauser pode ser encontrada em Quan et al. (2005) [80].

3.3 Software

O presente estudo foi desenvolvido com recurso à linguagem e sistema de computação científica e estatística R 4.1.2, que é um ambiente de desenvolvimento de acesso livre para o armazenamento, processamento, cálculo, análise e representação de dados, e ao RStudio 2021.09.4+403, um IDE (*integrated development environment*) para R. A sua aplicação na área da saúde tem intensificado em função da simplicidade de utilização da técnica para a análise de dados hospitalares e clínicos. Toda a implementação foi auxiliada por funções e pacotes disponíveis neste software. Os principais pacotes R utilizados neste trabalho encontram-se apresentados na Tabela 9.

Tabela 9 – Pacotes (*packages*) da Linguagem R utilizados.

Pacote R	Objetivo
dplyr ¹	Manipular dados
icd ²	Identificar as condições de Elixhauser nos dados segundo Quan et al.
klaR ³	Implementar o método k-modes
cluster ⁴	Implementar algoritmos PAM e hierárquico
Factoextra ⁵	Visualizar dados
gtsummary ⁶	Criar tabelas analíticas e de resumo
caret ⁷	Treino de modelos de classificação e regressão

3.4 Algoritmos de Clustering

Diversos algoritmos têm sido utilizados para a construção de clusters. Eles diferem de acordo com vários critérios como o tipo de variáveis, a capacidade de processar grandes conjuntos de dados, a capacidade de lidar com *outliers*, a complexidade do algoritmo (tempo de computação) e o conhecimento precedente sobre os dados [58]. Os algoritmos de clustering mais amplamente usados são os métodos de partição e os métodos hierárquicos, emergindo outras categorias em conformidade com diferentes *datasets* [81],[60]. Estes algoritmos têm sido muito utilizados numa vasta gama de aplicações, principalmente devido à sua simplicidade e facilidade de implementação comparativamente a outros algoritmos. Não obstante à eficácia e eficiência dos métodos que integram esses dois grupos, estes encontram-se dependentes do fornecimento de conhecimento prévio ou informações do número exato de clusters para cada *dataset* a ser analisado. Todavia, ao lidar com *datasets* do “mundo real”, não é usual possuir qualquer informação prévia sobre o número de grupos que ocorrem naturalmente nos objetos de dados [81]. Estudos comprovaram assim que não existe um critério absoluto para determinar o melhor algoritmo de clustering. O resultado de um algoritmo pode ser mais adequado para um determinado *dataset*, mas não alcançar essa superioridade para outros conjuntos de dados [82].

Dado que o principal objetivo deste estudo é identificar padrões de multimorbidade nos internamentos devido a cancro da próstata em Portugal, incluindo a existência de grupos com perfis clínicos (comorbidades) distintos, possivelmente requerendo percursos de tratamento distintos e com diferentes riscos de desfechos indesejáveis, como morte, os métodos de clustering de partição podem constituir uma ferramenta viável para diferenciar esses pacientes. Como o *dataset* é composto apenas por variáveis categóricas, optou-se por aplicar inicialmente o método k-modes. Adicionalmente, foi testado o algoritmo de PAM que, assim como k-modes, consiste numa extensão do algoritmo de clustering k-means, todavia menos sensível a *outliers*. Além de se testar os métodos de partição mencionados, também se aplicou o método de clustering hierárquico, que é amplamente utilizado em vários campos, em virtude de permitir uma interpretação direta. Os algoritmos mencionados irão ser descritos nas secções seguintes.

¹ <https://www.rdocumentation.org/packages/dplyr/versions/0.7.8>

² <https://www.rdocumentation.org/packages/icd/versions/3.3>

³ <https://www.rdocumentation.org/packages/klaR/versions/1.7-2>

⁴ <https://www.rdocumentation.org/packages/cluster/versions/2.1.4>

⁵ <https://www.rdocumentation.org/packages/factoextra/versions/1.0.7>

⁶ <https://www.rdocumentation.org/packages/gtsummary/versions/1.6.3>

⁷ <https://www.rdocumentation.org/packages/caret/versions/6.0-94>

3.4.1 Algoritmos de Partição

Os algoritmos de partição visam particionar n observações em k clusters pré-determinados, otimizando uma função de critério definida e melhorando iterativamente a qualidade das partições [37],[58],[83]. Cada cluster é representado pelo centro (centróide) do cluster, p. ex. k-means, ou pela instância mais próxima do centro (medóide), como o *Partitioning Around Medoid* (PAM). O critério mais usualmente utilizado trata-se da minimização do critério de erro quadrado - soma das distâncias euclidianas ao quadrado a partir do cluster com o centróide mais próximo dos pontos [84]. A Figura 5 ilustra a abordagem dos algoritmos de partição.

Os métodos de partição geram uma única partição dos dados em oposição a uma estrutura de clusters, como o dendrograma produzido por métodos hierárquicos [58],[85]. Estes algoritmos são favoráveis para aplicações que compreendem grandes conjuntos de dados, para os quais a construção de um dendrograma é computacionalmente inexecutável [81]. Uma considerável desvantagem dos algoritmos de partição é o facto de existirem várias soluções possíveis [84].

Os algoritmos mais populares desta categoria incluem: k-means, k-modes, PAM (*Partitioning Around Medoids*), CLARA (*Clustering Large Applications*), Fuzzy c-means.

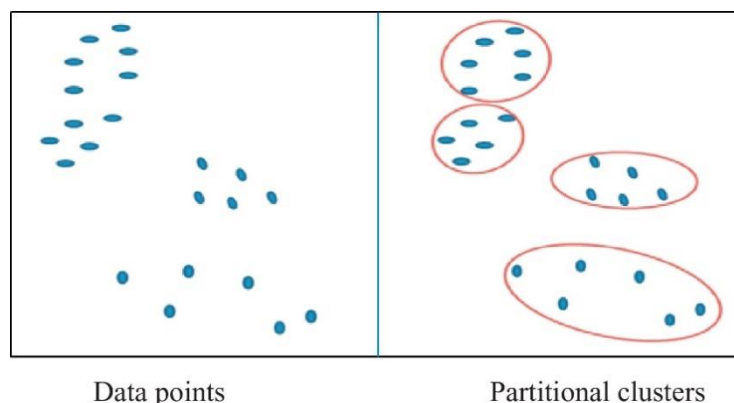


Figura 5 - Abordagem do algoritmo de partição [86].

K-Means

O algoritmo de clustering k-means, proposto pela primeira vez há mais de 50 anos, é um dos algoritmos mais utilizados na resolução de problemas de clustering [60],[86]. Este algoritmo atribui aleatoriamente todas as observações a um cluster e calcula um centróide para cada um, utilizando o valor médio de todos os objetos do cluster. Cada observação é, de seguida, reatribuída a um cluster cujo centro (centróide) se encontre mais próximo. A similaridade intra-cluster é otimizada cada vez que uma nova observação é atribuída a um cluster. Este processo é repetido inúmeras vezes até se alcançar a estabilidade [81].

As principais vantagens do algoritmo k-means são a sua simplicidade de implementação e baixa complexidade computacional, o que permite a sua execução em grandes *datasets* [81],[85]. Não obstante, este algoritmo apresenta adversidades preponderantes como a imposição da definição inicial do número de clusters, sensibilidade à escolha inicial do centróide e a *outliers* (objetos que se encontram distantes do centróide do cluster são impostos a esse cluster) [81],[84]. Em virtude do agrupamento preliminar aleatório, o algoritmo k-means não fornece o mesmo resultado a cada execução. A colocação de centróides iniciais é determinante, dado que locais diferentes ocasionam resultados discrepantes [85]. Para além disso, o uso de médias limita a aplicação do algoritmo a objetos de dados com variáveis

numéricas e é desconhecido o número de iterações necessárias para obtenção de uma distribuição estável [81].

K-modes

Em 1997, Huang et al. propuseram o algoritmo de clustering k-modes, uma extensão do popular método k-means, amplamente empregue no agrupamento de dados categóricos [87],[88]. Esta extensão abordou a limitação existente de agrupamento de dados somente numéricos imposta pelo algoritmo k-means, permitindo a implementação de dados categóricos de grande dimensão de bases de dados do mundo real [89]. O algoritmo k-modes substitui a média dos clusters pela moda e utiliza uma abordagem baseada em frequência para atualizar iterativamente a moda, melhorando a eficiência do processo de clustering [87],[88]. Afigura-se mais rápido que o algoritmo k-means, uma vez que necessita de menos iterações para convergir. No entanto, semelhante a esse algoritmo, apresenta certas desvantagens em virtude da estipulação que o número de clusters é conhecido antecipadamente. Apresenta, igualmente, disfunções quando os clusters são de diferente tamanho, densidade ou forma não circular [87]. A Figura 6 esquematiza o processo realizado por este algoritmo.

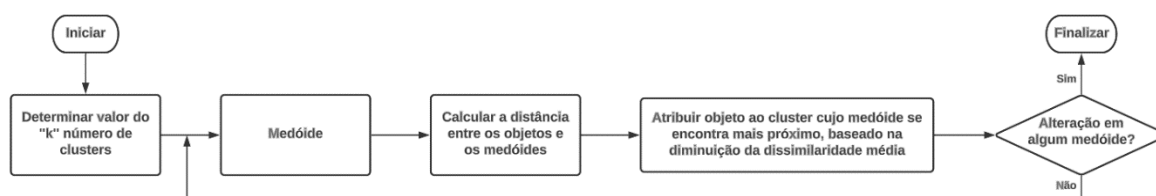


Figura 6 - Esquema representativo do processo do algoritmo k-modes.

Partitioning Around Medoids (PAM)

O método k-medóides, habitualmente designado por *Partitioning Around Medoid* (PAM), foi desenvolvido por Kaufman e Rousseeuw, em 1987 [60]. Este algoritmo funciona de forma semelhante ao algoritmo k-means, todavia, ao contrário deste, no algoritmo de PAM um cluster é representado pelo seu medóide, que é o objeto localizado mais centralmente no cluster, que comparativamente aos centróides são mais resistentes a *outliers* e ruído [84],[90].

O algoritmo de PAM inicia com a seleção aleatória de um objeto como medóide para cada um dos clusters. De seguida, é calculada a dissimilaridade média de todos os objetos não selecionados, para que sejam agrupados com o medóide mais semelhante. O PAM vai então substituindo iterativamente um dos medóides por um dos objetos não selecionados, até reduzir o valor da função objetivo. Este passo permite que o algoritmo obtenha uma melhor qualidade de agrupamento [84],[90].

Quando comparado com o algoritmo k-means, o PAM apresenta certas vantagens como a sua capacidade de lidar com todos os tipos de variáveis e todas as medidas de distância. Como este método minimiza a soma de dissimilaridades entre os objetos rotulados como sendo de um cluster e o medóide desse cluster, em detrimento da minimização da soma das distâncias euclidianas ao quadrado, notabiliza-se mais robusto [85]. Na Figura 7 apresenta-se um esquema do processo realizado por este algoritmo.

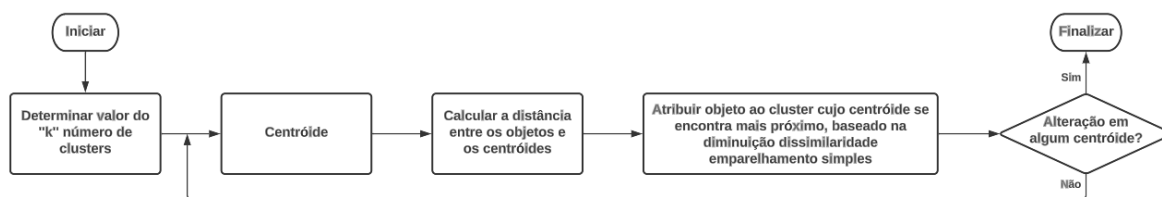


Figura 7 - Esquema representativo do processo do algoritmo PAM

Clustering for Large Applications (CLARA)

O algoritmo CLARA (*Clustering for Large Applications*) consiste numa extensão do algoritmo PAM que permite operar datasets que contém um grande número de dados (acima de vários milhares). Este algoritmo consiste na combinação de um processo de amostragem e o algoritmo de PAM, nomeadamente considera uma pequena amostra do conjunto de dados e aplica o algoritmo PAM para identificar os medóides da amostra [60],[90]. Fundamenta-se no conceito que, se a amostra for escolhida de modo suficientemente aleatório, o medóide representará corretamente todo o conjunto de dados [60]. A qualidade dos medóides definidos é medida em relação ao dataset completo, verificando a dissimilaridade média entre todos os objetos do conjunto de dados e os medóides [90].

3.4.2 Métodos Hierárquicos

O princípio dos métodos hierárquicos reside na construção de um relacionamento hierárquico entre os dados a agrupar, permitindo uma interpretação direta e transparente. Em oposição ao métodos de partição, os algoritmos hierárquicos produzem um dendrograma, composto por um único cluster completo no topo e clusters contendo apenas uma observação na parte inferior. Essa hierarquia pode ser instituída de cima para baixo (divisiva) ou de baixo para cima (aglomerativa), terminando quando for alcançado o número desejado de clusters (Figura 8) [84],[85]. São considerados métodos de clustering de boa qualidade, criando clusters de forma arbitrária e geralmente robustos ao ruído e com a vantagem de não impor a discriminação antecipada do número de clusters necessários.



Figura 8 - Dendrograma do algoritmo de clustering hierárquico [86].

Na abordagem aglomerativa, também denominada *bottom-up*, os dados são inicialmente distribuídos de modo que cada observação represente um cluster individualizado. Habitualmente, os dois clusters mais próximos são progressivamente congregados num novo cluster. Cada aglomeração ocorre a uma

distância maior entre os clusters do que a aglomeração anterior, podendo o algoritmo terminar quando os clusters estiverem muito distantes para serem agrupados (critério de distância) ou quando se atingir uma dada condição de paragem, como o critério do número de clusters (Figura 9) [85].

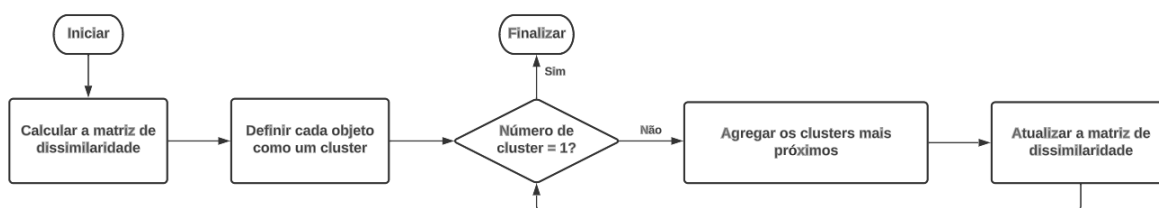


Figura 9 - Esquema representativo do processo do algoritmo hierárquico – abordagem aglomerativa.

A abordagem divisiva, também chamada *top-down*, faz o inverso da aglomerativa, iniciando com todas as observações num cluster e subdividindo em clusters menores, até que só exista uma observação por cluster ou que seja concretizada uma dada condição de paragem [86].

Os métodos de *clustering* hierárquico podem ser agrupados em cinco categorias de acordo com as medidas de dissimilaridade ou ligações, nomeadamente o do vizinho mais próximo (*single linkage*), do vizinho mais afastado (*complete linkage*), o critério de média (*average linkage*), o critério do centróide e o critério de Ward.

O método do vizinho mais próximo é um método no qual a ligação entre dois clusters é efetuada por um único par de elementos, ou seja, a distância entre os dois elementos (um em cada cluster) mais próximos (vizinhos). No método do vizinho mais distante ou *complete linkage*, a distância é estabelecida consoante a distância entre os elementos mais afastados dos clusters [60],[86]. Comparativamente ao algoritmo *complete linkage*, o *single linkage* apresenta-se mais versátil, possuindo diversas propriedades teóricas desejáveis. Todavia, de um ponto de vista pragmático, o algoritmo *complete linkage* produz hierarquias mais oportunas [85]. No método de ligação média, a distância entre dois clusters é determinada pela distância média de qualquer elemento de um cluster para qualquer elemento do outro cluster [60],[86]. O critério do centróide calcula a dissimilaridade entre o centróide de um cluster e o centróide de outro cluster. Por último, no método de variância mínima de Ward, os clusters são formados de modo a minimizar a variância total intra-cluster, onde o par de clusters com a mínima distância entre eles são agrupados.

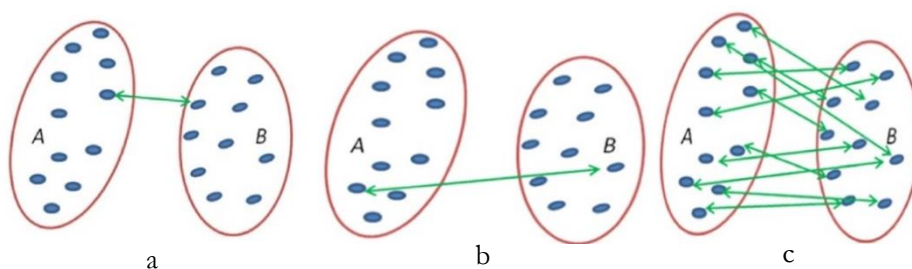


Figura 10 - Representação dos métodos a) *single linkage*; b) *complete linkage*; c) *average linkage* [86].

3.5 Medidas de Distância e Similaridade

Os métodos de clustering diferem na seleção da função objetivo, bem como na matriz de distância utilizada e na abordagem de elaboração da matriz de dissimilaridade [60]. A distância (dissimilaridade) e

a similaridade são a base para a construção de algoritmos de clustering [81],[91]. Esses algoritmos utilizam medidas de distância ou similaridade para agrupar observações (sujeitos ou objetos) semelhantes nos mesmos clusters, ao passo que observações diferentes ou distantes são colocadas em clusters distintos [91]. A distância ou a similaridade são geralmente mensuradas através de uma função de distância, na qual, quanto menor a distância, mais semelhantes dois objetos são considerados [92]. A escolha da medida entre os objetos a utilizar é muito importante pois a sua eficiência influencia diretamente os resultados do clustering [58],[93]. Essa escolha é comumente efetuada de forma arbitrária, sensível à representação dos objetos, processando todos os atributos equitativamente e pode depender da natureza dos dados e da especificidade da aplicação. A dissimilaridade é a medida elegida para reconhecer a relação entre os dados, quando se refere a dados quantitativos. Quanto a dados qualitativos, é priorizada a similaridade [81].

Medidas de distância para variáveis quantitativas

- **Distância de Minkowski:** A distância de Minkowski, também denominada de métrica de distância generalizada, constitui uma família de medidas de dissimilaridade estipulada por um parâmetro p , utilizadas em dados contínuos [85].

$$Dist_{XY} = \left(\sum_{k=1}^d |X_{ik} - X_{jk}|^{\frac{1}{p}} \right)^p \quad \text{Equação 1}$$

Onde:

- Distância de Manhattan quando $p=1$
- Distância Euclidiana quando $p=2$
- Distância Chebyshev quando $p \rightarrow \infty$
- **Distância Euclidiana:** Medida de distância mais destacada em vários problemas de *clustering*, que define o espaço cartesiano [58],[81]. Esta distância é considerada o padrão ou indicador mais comumente utilizado para dados numéricos. A distância Euclidiana entre dois valores é a diferença aritmética e impõe um peso progressivamente maior aos objetos que se encontram mais afastados [94].

$$Dist_{XY} = \sqrt{\sum_{k=1}^m (X_{ik} - X_{jk})^2} \quad \text{Equação 2}$$

- **Distância de Manhattan:** A distância de Manhattan mede a diferença absoluta entre as coordenadas de dois objetos em dados multivariados, sendo menos sensível ao ruído [58],[81].

$$d_1(x_i^{\rightarrow}, x_j^{\rightarrow}) = |x_i^{\rightarrow}, x_j^{\rightarrow}| = \sum_{k=1}^N |x_{ik} - x_{jk}| \quad \text{Equação 3}$$

- **Distância Chebyshev:** A distância Chebyshev determina a magnitude absoluta das diferenças entre as coordenadas de um par de objetos [81].

$$Dist_{XY} = \max_k |X_{ik} - X_{jk}|$$

Equação 4

- **Distância de Cosine:** A distância de Cosine é distância mais frequentemente aplicada em documentos. Mede o cosseno do ângulo entre dois pontos, onde θ é o ângulo entre dois vetores de dados e A e B são vetores n-dimensionais [81].

$$\theta = \frac{\arccos(A \cdot B)}{\|A\| \|B\|}$$

Equação 5

Medidas de distância e similaridade para variáveis nominais binárias

- **Distância de emparelhamento simples** (*simple matching distance*): Esta medida, também denominada de distância de Hamming, mensura a distância entre objetos contando o número de divergências. Quanto menor o número de incompatibilidades, mais semelhantes são os dois objetos. Pode ser representada matematicamente como [87]:

$$d(X, Y) = \sum_{j=1}^m \delta(x_j, y_j)$$

Equação 6

- **Coefficiente de Jaccard:** O coeficiente de Jaccard é uma medida comum para atributos binários, que mede a similaridade entre dois objetos de dados avaliando a interseção dividida pelas uniões desses objetos, como demonstrado na Equação 7 [81].

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

Equação 7

- **Coefficiente de Gower:** Métrica de distância para dados mistos (numéricos, categóricos, categóricos ordenais, entre outros) [95]. A similaridade entre dois objetos é comparada em cada atributo k, sendo atribuído a este um índice S_{ijk} [58].

$$S_{ij} = \sum_{k=1}^N S_{ijk} \delta_{ijk} \omega_k / \sum_{k=1}^N \delta_{ijk} \omega_k$$

Equação 8

Onde:

- $S_{ijk} = 0$ se os objetos se apresentarem totalmente distintos ou real positivo se possuírem algum nível de correspondência;
- $\delta_{ijk} = 1$ caso seja possível comparar i e j para o atributo k, e 0 caso contrário;
- ω_k - peso do atributo k.

O k-modes emprega a distância de emparelhamento simples para medir a dissimilaridade entre as instâncias. Por outro lado, o algoritmo de clustering de PAM e hierárquico foram executados numa matriz de distância de Gower calculada a partir dos dados, que, ao contrário de outras métricas de

distância comumente usadas como as distância Euclidiana e de Manhattan, é capaz de lidar com tipos de dados mistos.

3.6 Determinação do número ideal de clusters

Os métodos de partição, k-modes e PAM, requerem o número de clusters como parâmetro inicial, como foi referido anteriormente. O número ideal de clusters é algo subjetivo e depende do método utilizado para medir as similaridades e dos parâmetros de partição empregues [93]. A solução final e a sua qualidade dependem da seleção assumida nesta fase, podendo ser efetuada de modo automático ou fundamentada em métodos diretos ou métodos de cálculos estatísticos. Os métodos diretos consistem na otimização de um critério, como as somas de quadrados do cluster (método Elbow) ou a silhueta média (método *Average Silhouette*) [96], diversamente ao método *Gap statistic*, que é um método de cálculo estatístico.

Método Elbow: Dado que a ideia elementar por detrás do clustering é agrupar os dados de forma a minimizar a variação total intra-cluster (*total within-cluster sum of square* -WSS), o método Elbow considera a WSS total como uma função do número de clusters. O número apropriado de clusters verifica-se quando a adição de outro cluster não manifesta uma diminuição acentuada no valor da variação total intra-cluster [81].

O método Elbow executa o algoritmo diversas vezes com diferentes valores de k, elaborando um gráfico com o valor da variação em função de k, o k final é escolhido com base na localização do “cotovelo” (quando a curva de variação intra-cluster começa a ficar plana). No entanto, reconhecer que esse ponto depende da visualização manual do utilizador, sendo considerado uma abordagem ambígua.

O método Elbow pode ser definido da seguinte forma [81]:

- 1) Calcular o algoritmo de clustering para diferentes valores de k (1 a 10 clusters);
- 2) Para cada k, calcular a variação total intra-cluster (wss);
- 3) Traçar a curva de wss de acordo com o número de clusters k;
- 4) A localização de uma curva (cotovelo) no gráfico indica o número adequado de clusters.

Método *Average Silhouette*: Esta abordagem, além de poder ser utilizada para selecionar o número de clusters ideal, interpreta e avalia a consistência dos dados em cada cluster, ou seja, determina quão adequada foi a atribuição dos objetos ao grupo correspondente [96]. A análise da silhueta calcula um índice subtraindo a distância média entre cada instância e as restantes instâncias do mesmo cluster (distância intra-cluster) pela distância média do cluster mais próximo. Essa diferença é dividida pelo valor máximo entre as duas distâncias, como demonstrado na Equação 9 [81]:

$$s(p) = \frac{(b(i) - a(i))}{\max \{a(i), b(i)\}} \quad \text{Equação 9}$$

Esse índice produz uma medida normalizada, que pode assumir valores num intervalo [-1, 1]. As pontuações mais elevadas (mais próximas de 1) indicam uma qualidade superior dos clusters produzidos.

O método *Average Silhouette* é similar ao método Elbow, pode funcionar como um bom indicador do número ideal de clusters do seguinte modo:

- 1) Calcular o algoritmo de clustering para diferentes valores de k (1 a 10 clusters).
- 2) Para cada k, calcular a silhueta média das observações (avg.sil).

- 3) Desenhar a curva de *avg.sil* de acordo com o número de clusters k .
- 4) A localização do máximo é considerada como o número adequado de clusters.

Método *Gap statistic*: O método *Gap statistic* pode ser aplicado a qualquer método de clustering e compara a variação total intra-cluster, para diferentes valores de k , com os valores esperados, representados pela hipótese nula. O número ótimo do número de clusters é valor de k que maximiza a função *Gap*, que prenuncia que a estrutura do clustering está distante de uma distribuição uniforme aleatória de pontos [81].

Apesar do método *Elbow* e o método *Average Silhouette* fornecerem informações sobre parâmetros diferentes, ambos apresentam informações relevantes para a análise de clustering. Neste trabalho, foi realizada uma análise conjunta do método *Elbow* e do método *Average Silhouette* para orientação na eleição do número ótimo de clusters. Para esse propósito, foi verificado o número de clusters k versus o valor de *WSS* e o número de clusters k versus o coeficiente da silhueta médio, respetivamente, para todos os algoritmos excetuando o *k-modes*, que não é possível delinear o gráfico da *Average Silhouette*.

3.7 Critérios de Validação

Existem dois tipos de medidas de validação dos resultados de clustering distinguidos na literatura: a validação interna e a validação externa. Como o nome indica, no critério interno, a validação e avaliação da qualidade dos resultados obtidos por um dado algoritmo de clustering fundamenta-se em propriedades internas, como a compacidade e separação dos clusters. Em contrapartida, no critério externo, o clustering é avaliado comparando os clusters em relação a uma ou mais variáveis não utilizadas para o agrupamento como, por exemplo, o tempo de internamento. Cada uma destas medidas apresenta uma ampla variedade de critérios e medidas para efetuar a validação [97].

Neste estudo, optou-se pelo coeficiente de silhueta médio (ou *Silhouette Width*), um dos métodos de validação interna standard mais utilizados. A aplicação deste critério teve como principal objetivo a orientação na escolha do número ideal de clusters e do algoritmo de clustering mais apropriado para alcançar os objetivos deste trabalho.

3.8 Árvores de decisão

Um esquema de representação de clusters mais usualmente utilizado são as árvores de decisão. As árvores de decisão são um método de *machine learning* supervisionado que, como o nome sugere, desenvolve estruturas em forma de árvore para resolver problemas de classificação. Estes algoritmos baseiam-se na divisão contínua dos dados com base num determinado parâmetro, até que não seja necessário mais divisões, gerando assim uma folha [98]. A estrutura de uma árvore de decisão é constituída por nós, que representam os atributos a classificar, por ramos que apresentam os valores possíveis dos atributos e as folhas, que indicam as várias classes em que cada objeto pode ser classificado. A visualização gráfica das árvores de decisão apresenta como vantagem a sua facilidade de compreensão por parte dos utilizadores, fornecendo um meio intuitivo de analisar os resultados [99].

3.9 Análise estatística

As características descritivas da população incluída no estudo foram sumariadas pela mediana e pelo intervalo interquartil (IIQ) para variáveis contínuas e frequências (n) e percentagens para variáveis categóricas. Para descrever os clusters de multimorbilidade identificados foram expressas análises descritivas para cada cluster. Para além disso, as taxas de prevalência observado/esperado (O/E) relativas às comorbilidades de Elixhauser foram calculadas dividindo a prevalência de cada doença crónica do cluster pela prevalência correspondente dessa condição em toda a população com cancro da próstata incluída na análise. Seguindo a abordagem de Jansana et al. (2021) [94], foi destacada uma condição crónica de um cluster sempre que a sua razão de prevalência O/E foi de, pelo menos, 2.

4 RESULTADOS E DISCUSSÃO

Neste capítulo, é apresentado o tratamento estatístico efetuado com aplicação das técnicas referidas anteriormente, de forma a descrever o conjunto de dados utilizado, evidenciando as principais características do mesmo. Posteriormente, sucede-se a análise dos resultados obtidos pelo processo de clustering.

4.1 Apresentação de resultados

O conjunto de dados analisado inclui um total de 10.394 episódios de internamento compreendendo hospitalizações em função de um diagnóstico de cancro da próstata para realização de prostatectomia radical ($n=8.565$, 82,0%) ou braquiterapia ($n=1.830$, 18,0%) (Tabela 10). A maioria dos casos implicados relatou multimorbilidade ($n=6.091$, 58.6%), sendo as comorbilidades mais prevalentes: hipertensão ($n=4.596$, 44,0%), diabetes mellitus ($n=1.413$, 14%) e obesidade ($n=826$, 7,9%). Ademais, foi observada uma baixa frequência de internamentos que resultaram em morte ($n=31$, 0.3%) na amostra em estudo. A mediana da idade foi de 65 anos (IIQ: 60-69) e a mediana do tempo de internamento 5 dias (IIQ:4-7).

A cirurgia e a braquiterapia estão geralmente indicadas nas formas localizadas, nomeadamente para pacientes com cancro da próstata não metastizado. Verificando-se uma frequência atípica de metástases ($n=336$, 3.2%) nos pacientes desta amostra procedeu-se ao apuramento deste resultado. A faixa etária onde se exibiram mais pacientes com metástases foi dos 60 aos 74 anos (67%), com idade mediana dos pacientes de 66 anos (IIQ: 60-70). Quanto à duração do internamento, a mediana foi de 5 dias, tal como para a amostra total. Além disso, foi observada uma taxa de mortalidade de 7.1% desses pacientes. Dos 336 doentes metastizados, a maioria realizou prostatectomia radical (75%), sendo a braquiterapia o tratamento eleito para os restantes pacientes ($n = 85$, 25%). Estes doentes encontravam-se admitidos, maioritariamente, no grupo hospitalar E (44%), que concentra centros hospitalares universitários de grande porte, e localizam-se, principalmente, na região Norte (41%) e de Lisboa (41%). É essencial mencionar que nenhum dos pacientes metastáticos analisados possui outro tumor para além do carcinoma da próstata, indicando que as metástases não derivam de outro tipo de cancro.

Tabela 10 - Estatísticas descritivas do conjunto de dados analisado

Características	Total
Idade, mediana (IIQ)	65 (60-69)
Braquiterapia, n (%)	1830 (18)
Prostatectomia Radical, n (%)	8565 (82)
Metástases, n (%)	336 (3.2)
Tempo de internamento, mediana (IIQ)	5 (4-7)
Morte, n (%)	31 (0.3)
Multimorbilidade, n (%)	6091 (58.6)

Ao comparar os resultados obtidos da Tabela 11, podemos observar algumas discrepâncias nas características dos homens com carcinoma da próstata que realizaram braquiterapia ou prostatectomia radical. Em relação à idade, a mediana foi de 66 anos (IIQ: 61-70) para o conjunto da braquiterapia e de 64 anos (IIQ: 60-68) para os pacientes que efetuaram prostatectomia radical. Como seria de esperar, o tempo mediano de internamento dos doentes que realizaram cirurgia (6 dias) revelou-se mais longo comparativamente com a braquiterapia (2 dias). Para além disso, não obstante ao facto de ambos os

casos apresentarem um número reduzido de óbitos, verificou-se uma taxa de mortalidade mais elevada para os pacientes que realizaram braquiterapia (1.3%) comparativamente a uma taxa de 0.1% da prostatectomia radical. Ambos os conjuntos de dados manifestaram uma mediana de 1 comorbilidade (IIQ: 0-2), indicando uma similaridade no número de comorbilidades entre os pacientes alocados a ambos os braços de tratamento. Relativamente à distribuição dos grupos hospitalares, os pacientes cirúrgicos não revelam nenhuma predominância em específico, no entanto, observa-se uma elevada prevalência de pacientes que realizaram braquiterapia no grupo F, que concentra os hospitais especializados na área oncológica. Quanto à região do hospital, os episódios demonstraram-se heterogeneamente distribuídos, não relevando nenhum grupo particularmente expressivo. A presença de metástases foi observada em ambos os tratamentos, com 4.6% dos casos relativos a pacientes internados para braquiterapia e 2.9% dos casos de prostatectomia radical.

Diferenciando os pacientes metastizados consoante o tipo de tratamento que efetuaram, é possível verificar que os homens referenciados para braquiterapia apresentam uma idade superior (idade mediana de 71 anos) comparativamente aos doentes cirúrgicos (65 anos). O grupo da braquiterapia demonstrou-se igualmente irregular em relação à admissão típica de um paciente com carcinoma da próstata (programada), com 78% dos pacientes com metástases admitidos de forma urgente, requerendo uma intervenção imediata. Esse mesmo grupo revelou uma gravidade da doença bastante superior aos demais pacientes, com cerca de 50% desses doentes a exibir um nível de severidade e risco de mortalidade maior, ao passo que menos de 5% dos pacientes metastizados que realizam cirurgia apresentavam esse nível. Relativamente às restantes variáveis estudadas não se verificaram diferenças significativas.

Tabela 11 - Estatísticas descritivas do conjunto de dados de Braquiterapia e de Prostatectomia Radical

Braquiterapia		Prostatectomia Radical	
Características	Total	Características	Total
Idade, mediana (IIQ)	66 (61-70)	Idade, mediana (IIQ)	64 (60-68)
Metástases, n (%)	85 (4.6)	Metástases, n (%)	251 (2.9)
Tempo de internamento, mediana (IIQ)	2 (2-3)	Tempo de internamento, mediana (IIQ)	6 (4-7)
Morte, n (%)	24 (1.3)	Morte, n (%)	7 (<0.1)
Nºde comorbilidades, mediana (IIQ)	1 (0-2)	Nºde comorbilidades, mediana (IIQ)	1 (0-2)
Grupo hospitalar, n (%)		Grupo hospitalar, n (%)	
Grupo A	0 (0)	Grupo A	44 (0.5)
Grupo B	1 (<.01)	Grupo B	342 (4.0)
Grupo C	22 (1.2)	Grupo C	2150 (25)
Grupo D	24 (1.3)	Grupo D	1707 (20)
Grupo E	243 (13)	Grupo E	3075 (36)
Grupo F	1519 (84)	Grupo F	1159 (14)
Outros	21	Outros	88
Região do hospital, n (%)		Região do hospital, n (%)	
Norte	655 (36)	Norte	3560 (42)
Centro	739 (41)	Centro	1920 (23)
Lisboa	405 (22)	Lisboa	2645 (31)
Alentejo	5 (0.3)	Alentejo	187 (2.2)
Algarve	5 (0.3)	Algarve	168 (2.0)
		Outros	85

No que se refere à prevalência geral da multimorbilidade segundo a faixa etária dos pacientes incluídos no estudo (Tabela 12), verifica-se que a maioria dos doentes portugueses com cancro da próstata, entre 2011 e 2017, não apresentava nenhuma comorbilidade ($n = 4.303$) ou possuía apenas uma, para além do cancro da próstata ($n = 3.366$). Na presença de multimorbilidade, a generalidade dos pacientes com cancro da próstata possuía duas doenças crónicas para além do cancro da próstata ($n = 1.902$), compreendendo-se a maior proporção na faixa etária entre os 60 e os 74 anos (55%).

Tabela 12 - Frequência e frequência relativa das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante a faixa etária

	0 DC n=4303	1 DC n= 3366	2 DC n= 1902	3 DC n= 626	4 DC n= 150	5 DC n=31	≥6 DC n=16
0-49	122 (64.2%)	47 (24.7%)	16 (8.4%)	5 (2.6%)	0 (0%)	0 (0%)	0 (0%)
50-59	1083 (50.7%)	644 (30.1%)	288 (13.5%)	93 (4.4%)	22 (1.0%)	3 (0.1%)	3 (0,1%)
60-74	3007 (38.6%)	2573 (33.0%)	1555 (19.9%)	499 (6.4%)	125 (1.6%)	27 (0.3%)	11 (0,1%)
75+	91 (33.6%)	102 (37.6%)	43 (15.9%)	29 (10.7%)	3 (1.1%)	1 (0.4%)	2 (0.7%)

As Tabela 13 e Tabela 14 apresentam a prevalência do número de doenças crónicas, nos pacientes portugueses com carcinoma da próstata, consoante o tempo de internamento e o número de óbitos, respetivamente. Em ambos os casos é possível constatar que o número de comorbilidades não influenciou o desfecho dessas variáveis, sendo de importante relevância referir que o número de pacientes com 5 ou mais comorbilidades é bastante diminuto, não permitindo aferir conclusões relevantes.

Tabela 13 - Mediana e intervalo interquartil das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante o tempo de internamento

	0 DC n=4303	1 DC n= 3366	2 DC n= 1902	3 DC n= 626	4 DC n= 150	5 DC n=31	≥6 DC n=16
Tempo de internamento, mediana (IQR)	5 (4-7)	5 (4-7)	5 (4-7)	5 (4-7)	5 (3-7)	5 (3.5-7)	11 (4.75-23.8)

Tabela 14 - Frequência e frequência relativa das comorbilidades de Elixhauser (0 DC: 0 comorbilidades; 1 DC: 1 comorbilidade; 2 DC: 2 comorbilidades; 3 DC: 3 comorbilidades; 4 DC: 4 comorbilidades; 5 DC: 5 comorbilidades; >6 DC: 6 ou mais comorbilidades) em pacientes com cancro da próstata consoante o número de óbitos

	0 DC n=4303	1 DC n= 3366	2 DC n= 1902	3 DC n= 626	4 DC n= 150	5 DC n=31	≥6 DC n=16
Nºde óbitos, n(%)	6 (19.4)	5 (16.1)	11 (35.5)	5 (16.1)	1 (3.2)	0 (0)	3 (9.7)

4.1.1 Número ideal de clusters e validação interna

Como mencionado no Capítulo MÉTODOS E APLICAÇÃO, os algoritmos k-modes e PAM requerem que se predefina o número de clusters a produzir. Para estimar o número ideal de clusters, utilizou-se o método Elbow e o método *Average Silhouette*. Para o método Elbow, o número ideal de clusters foi identificado selecionando o valor k após o qual o valor da WSS não diminui significativamente, indicando que os clusters adicionais a esse valor possuem pouca relevância. No método *Average Silhouette*, o número ótimo de clusters foi aquele que maximizou a silhueta média no intervalo de valores possíveis para k. De seguida, são apresentados os gráficos obtidos pelos dois métodos referidos por cada algoritmo, para os 3 conjuntos de dados em análise.

Braquiterapia

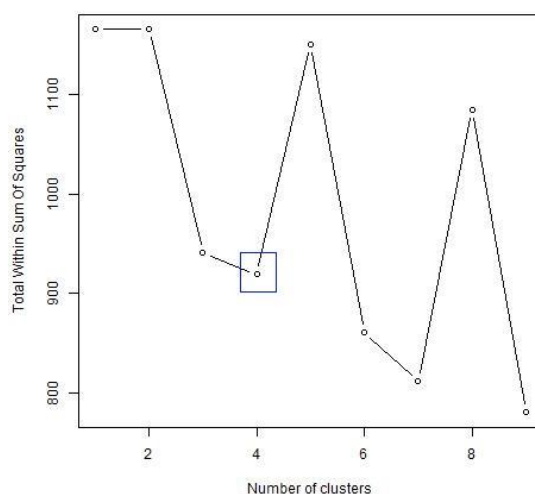


Figura 11 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes

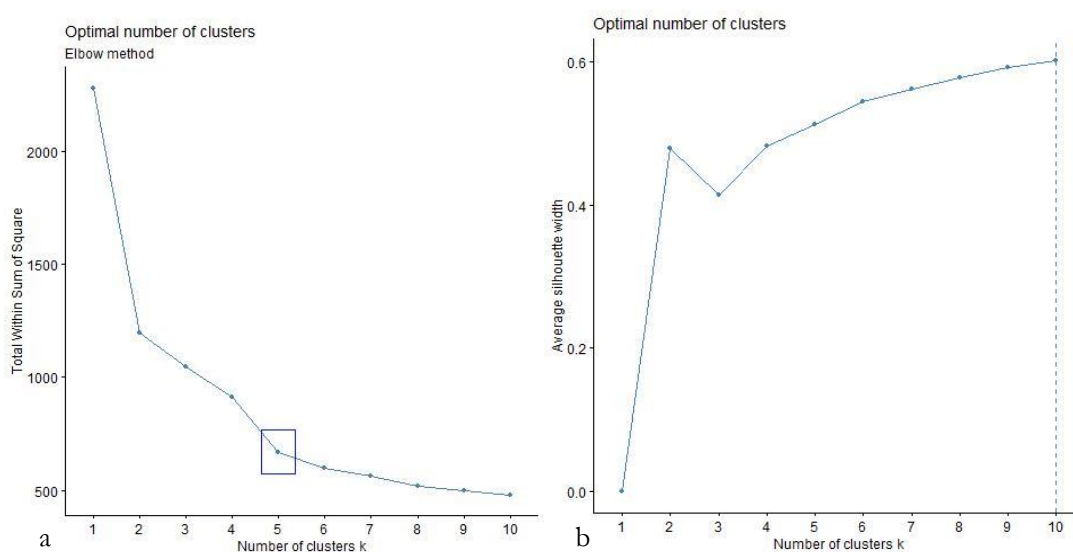


Figura 12 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o método PAM

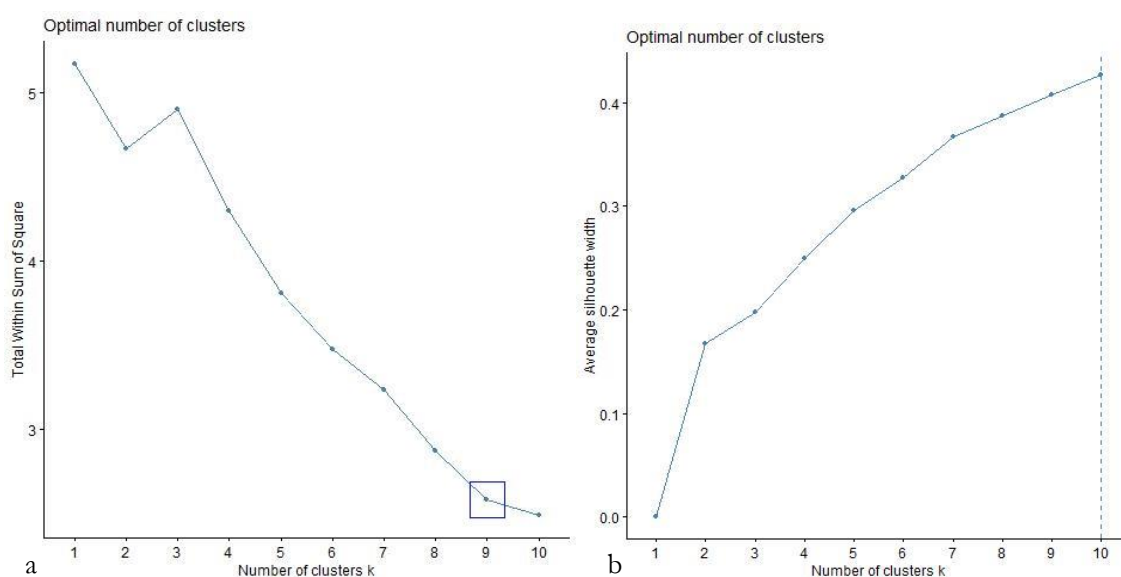


Figura 13 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico

Como, por vezes, o ponto em que se verifica o número ótimo de clusters não é facilmente identificado ou muito evidente, recorreu-se ao coeficiente de silhueta médio para auxílio na definição do melhor valor de k. Para a seleção do número ideal de clusters pelo coeficiente de silhueta médio foi verificado o seu valor médio, bem como o valor do coeficiente de silhueta individual dos clusters. A Tabela 15 apresenta uma comparação dos resultados obtidos pelos três processos, para cada algoritmo de clustering.

Tabela 15 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, *Average Silhouette* e o coeficiente de silhueta médio

	Método <i>Elbow</i>	Método <i>Average Silhouette</i>	Nº de clusters final
k-modes	5	-	5
PAM	5	10	4
Hierárquico	9	10	9

Atendendo à tabela anterior, os valores obtidos para ambos os processos diferiram na sua generalidade. No que diz respeito ao algoritmo k-modes, uma vez que não é possível a execução do gráfico pelo método *Average Silhouette*, o número ideal de clusters foi selecionado com base na análise do método *Elbow* (Figura 11), alcançando-se assim um valor final de 5 clusters (coeficiente de silhueta médio = 0.21). Em relação ao algoritmo de PAM, considerando, igualmente, o método de *Elbow* (Figura 12-a), o número ótimo de k corresponde a 5 clusters (coeficiente de silhueta médio = 0.418), contudo o cluster 1 apresenta um coeficiente de silhueta médio negativo (k=1: -0.11; k=2: 0.31; k=3: 0.10; k=4: 0.79; k=5: 0.32), o que significa que certas observações se encontram muito distante daquelas que se encontram no mesmo cluster. Verificando que o coeficiente de silhueta médio de cada cluster obtido, quando o número de cluster é 4, é positivo (k=1: 0.46; k=2: 0.43; k=3: 0.19; k=4: 0.41) e exibe uma média total semelhante

(coeficiente de silhueta médio = 0.416), elegeu-se o $k=4$ como o número ideal de clusters. Por fim, analisando o gráfico do método *Average Silhouette* (Figura 13-b) e o seu coeficiente, para o algoritmo de clustering hierárquico, verifica-se que o valor ótimo de k ocorre com o aumento constante do número de cluster. Assim sendo, tendo em conta que segundo o método Elbow (Figura 13-a), o número ótimo de clusters é 9 (coeficiente de silhueta médio = 0.44) e priorizando-se um menor número de cluster optou-se por esse valor.

Prostatectomia Radical

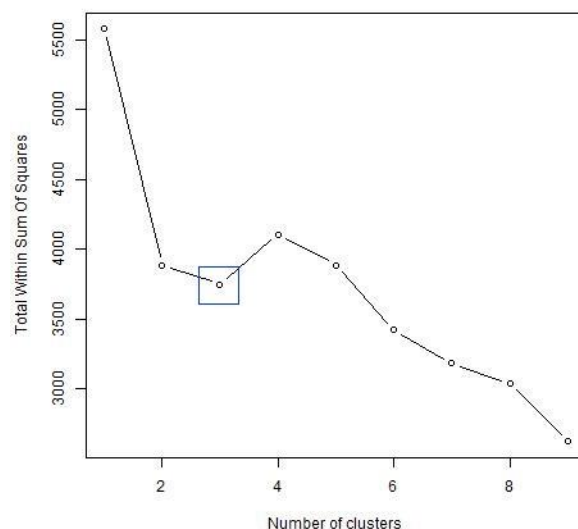


Figura 14 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes

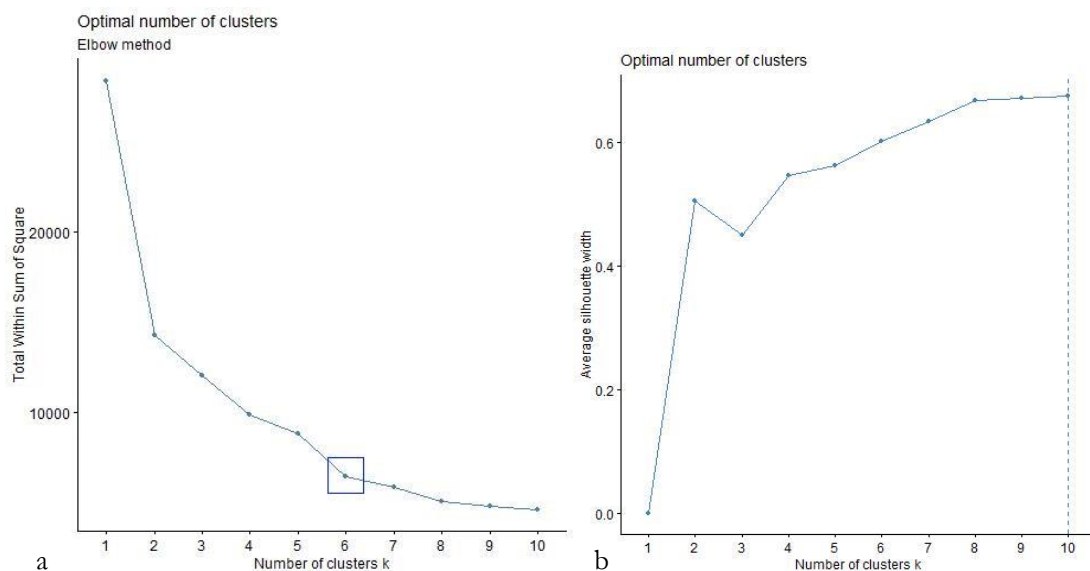


Figura 15 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o método PAM

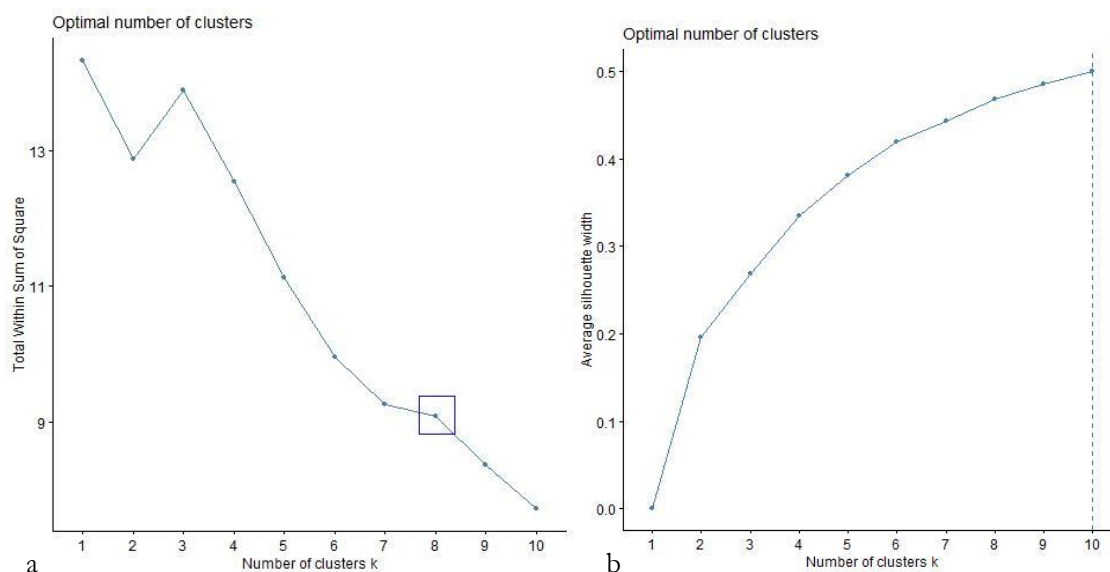


Figura 16 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico

A Tabela 16 contém os resultados do número ideal de clusters de cada processo, para os diferentes algoritmos de clustering.

Tabela 16 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, *Average Silhouette* e o coeficiente de silhueta médio

	Método Elbow	Método <i>Average Silhouette</i>	Nº de clusters final
k-modes	4	-	4
PAM	6	10	5
Hierárquico	8	10	8

Consultando a tabela anterior, o número ideal de clusters para o algoritmo k-modes é de 4 clusters (coeficiente de silhueta médio = 0.16) através do método Elbow (Figura 14). Como efetuado para o conjunto de dados anterior, o número ótimo de clusters escolhido para o algoritmo de PAM baseou-se na análise conjunta do método de Elbow (Figura 15-a) e do coeficiente de silhueta médio intra-cluster. Em k=5, a média do coeficiente de silhueta médio é 0.46, enquanto, individualmente, cada cluster apresenta um valor do coeficiente positivo (k=1: 0.46; k=2: 0.58; k=3: 0.25; k=4: 0.30; k=5: 0.59), sendo que, a partir deste valor, cada clustering apresenta, pelo menos, um coeficiente de silhueta médio individual negativo. Relativamente ao algoritmo de clustering hierárquico, o número ideal de clusters foi 8 (coeficiente de silhueta médio = 0.47).

Braquiterapia + Prostatectomia Radical

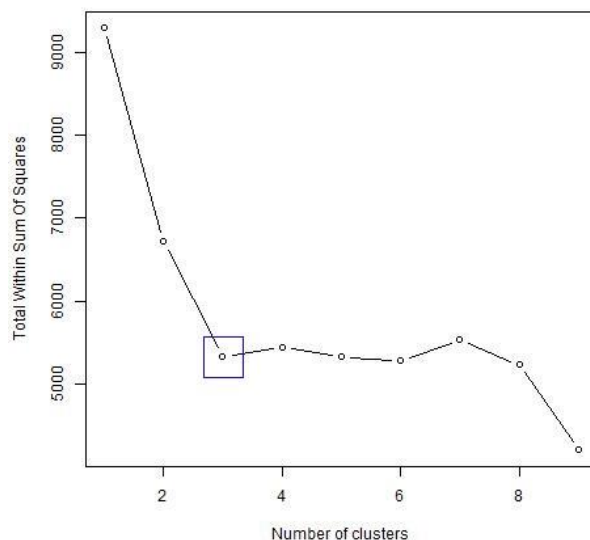


Figura 17 - Representação do método Elbow, da variação total intra-cluster em função do número k de clusters, para o método k-modes

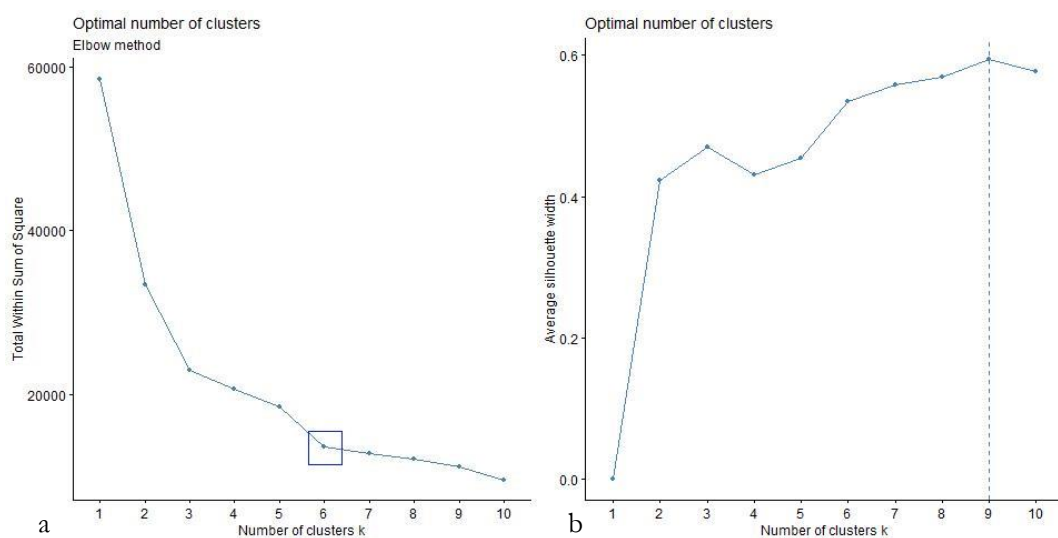


Figura 18 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o método PAM

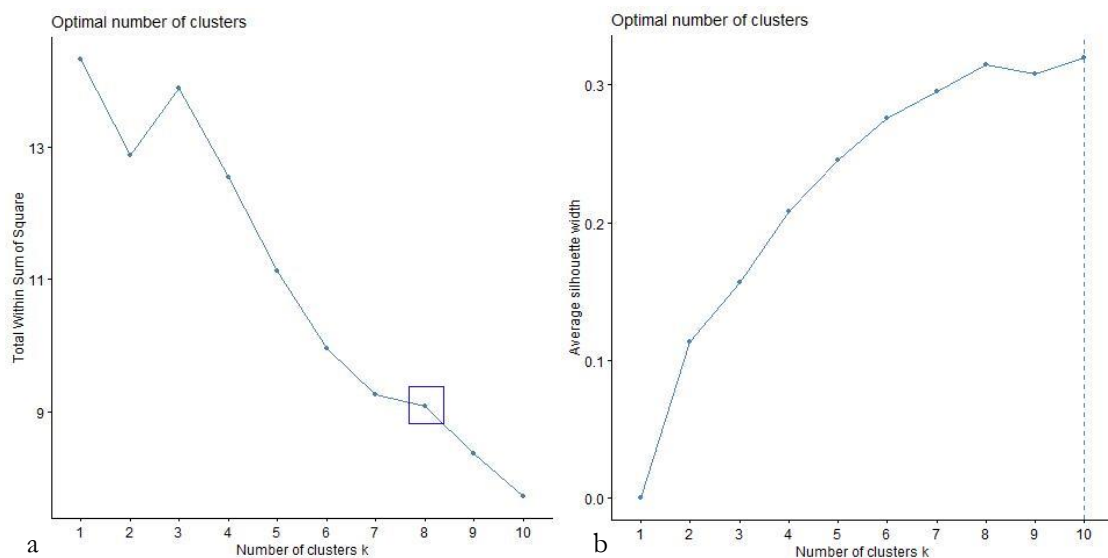


Figura 19 - a) Representação do método Elbow, da variação total intra-cluster em função do número k de clusters; b) Representação do método *Average Silhouette*, da variação silhueta média em função do número k de clusters, para o algoritmo de clustering hierárquico

Os resultados do número ótimo de clusters obtidos para os diferentes algoritmos de clustering encontram-se demonstrados na Tabela 17.

Tabela 17 - Resultados do número ideal de clusters para cada algoritmo de clustering, pelo método Elbow, *Average Silhouette* e o coeficiente de silhueta médio

	Método <i>Elbow</i>	Método <i>Average Silhouette</i>	Nº de clusters final
k-modes	4	-	4
PAM	6	9	6
Hierárquico	8	10	8

De acordo com a análise realizada utilizando o algoritmo k-modes, considerando o método Elbow (Figura 17) e o coeficiente de silhueta médio, foi determinado que o número ideal de clusters é 4. Nessa configuração, o coeficiente de silhueta médio é de -0.006. Assim sendo, verificou-se o valor do coeficiente para $k=5$ (coeficiente de silhueta médio = 0.065) e, averiguando que o valor não apresenta uma discrepância significativa relativamente ao número de cluster 4 e que o $k=4$ apresentou um cluster com valor intra-cluster (0.67) bastante superior ao valor máximo com $k=5$ (0.26), conclui-se um número ótimo de 4 clusters. Por outro lado, ao avaliar o algoritmo PAM, observou-se que o número ótimo de clusters é 5. Nesse caso, a média do coeficiente de silhueta médio é de 0.45 e cada cluster, individualmente, apresenta um valor positivo de coeficiente de silhueta ($k=1$: 0.46; $k=2$: 0.58; $k=3$: 0.25; $k=4$: 0.30; $k=5$: 0.59). Além disso, a partir desse valor, o coeficiente de silhueta médio começa a diminuir. Quanto ao algoritmo de clustering hierárquico, determinou-se que o número ideal de clusters é 8, obtendo um coeficiente de silhueta médio de 0.43.

4.1.2 Clusters de Multimorbilidade – Resultados e comparação entre os métodos

Um dos principais objetivos desta dissertação foi identificar e interpretar os padrões de multimorbilidade entre pacientes com cancro da próstata, através da análise de clusters. As tabelas seguintes apresentam os clusters obtidos com a aplicação de cada uma das abordagens de clustering, incluindo a métrica de qualidade selecionada, o coeficiente de silhueta médio. Cada cluster foi expresso de acordo com as comorbilidades mais prevalente ou tratamento realizado. Assim sendo, as doenças crónicas que surgem na designação dos clusters ocorreram em, pelo menos, 50% dos episódios dentro dos clusters. As cores foram utilizadas para destacar clusters comuns identificados pelas diferentes abordagens.

Braquiterapia

Os resultados da Tabela 18 mostram os resultados de diferentes algoritmos de clusters (k-modes, PAM e Hierárquico) aplicados ao conjunto de dados de pacientes com carcinoma da próstata que realizaram braquiterapia. Como é possível verificar, cada algoritmo gerou clusters diferentes, representados pelo número de pacientes pertencentes a cada cluster, as suas respetivas comorbilidades predominantes e, pelo coeficiente de silhueta médio, para avaliar a qualidade dos clusters formados por cada algoritmo. No caso do algoritmo k-modes, obteve-se resultados bastante discrepantes. Os Cluster 2 e 3 obtiveram um valor máximo de coeficiente de silhueta médio, indicando uma elevada qualidade de clustering, sendo associados a comorbilidades como a hipertensão e a hipertensão concomitante com diabetes mellitus, respetivamente. Por outro lado, os Clusters 1 e 5 revelaram uma ausência de características predominantes, apresentando um coeficiente de silhueta médio negativo, o que sugere uma reduzida qualidade de clustering. No algoritmo PAM, foram obtidos quatro clusters. Todos os clusters apresentaram coeficientes de silhueta médio positivos, indicando uma boa qualidade de clustering. O Cluster 1 foi associado principalmente a hipertensão, enquanto os restantes a combinações de hipertensão com outras características, como arritmia, álcool e diabetes mellitus. No algoritmo Hierárquico, alguns clusters revelaram números relativamente baixos no que se refere ao coeficiente de silhueta média, como é o caso do Cluster 1, indicando uma qualidade inferior no agrupamento de pacientes com características como paralisia e metástases. Os clusters da hipertensão e da hipertensão agrupada com diabetes mellitus foram encontrados por todos os algoritmos.

Em resumo, os resultados indicam que os algoritmos PAM e Hierárquico apresentaram uma melhor qualidade de clustering de pacientes com cancro da próstata, em comparação com o algoritmo k-modes. Os clusters formados pelo PAM e Hierárquico mostraram uma maior separação e associação com características específicas, como hipertensão, arritmia, diabetes mellitus e consumo de álcool. Por outro lado, o k-modes teve dificuldade em formar clusters distintos, resultando em coeficientes de silhueta médio negativos e uma falta de características predominantes em alguns clusters.

Tabela 18 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia, utilizando diferentes abordagens de clustering

Algoritmo	Clusters	Coefficiente de silhueta médio
K-modes	Cluster 1 (n= 139; Sem características predominantes)	-0.37
	Cluster 2 (n= 107; Hipertensão, Diabetes Mellitus)	1.0
	Cluster 3 (n= 386; Hipertensão)	1.0
	Cluster 4 (n= 177; Hipertensão)	-0.41
	Cluster 5 (n= 320; Sem características predominantes)	-0.41
PAM	Cluster 1 (n= 655; Hipertensão)	0.46
	Cluster 2 (n= 93; Hipertensão, Arritmia)	0.43
	Cluster 3 (n= 149; Hipertensão, Álcool)	0.19
	Cluster 4 (n= 232; Hipertensão, Diabetes Mellitus)	0.41
Hierárquico	Cluster 1 (n= 31; Paralisia, Metástases)	0.11
	Cluster 2 (n= 76; Hipertensão, Arritmia)	0.46
	Cluster 3 (n= 493; Hipertensão)	0.62
	Cluster 4 (n= 101; Hipertensão, Obesidade)	0.14
	Cluster 5 (n= 105; Hipertensão, Doença Pulmonar Obstrutiva Crónica)	0.16
	Cluster 6 (n= 189; Hipertensão, Diabetes Mellitus)	0.40
	Cluster 7 (n= 41; Depressão)	0.29
	Cluster 8 (n= 88; Hipertensão, Álcool)	0.36
	Cluster 9 (n= 5; Hipertensão, Depressão, Metástases, Morte)	0.13

Para visualizar os clusters obtidos, pelos três algoritmos aplicados, utilizamos a função `fviz_cluster()` (pacote `factoextra`). Esta função permite o desenho de um gráfico de dispersão com a coloração de cada ponto de dados, de acordo com sua atribuição de cluster. Também possibilitando realce de elipses de concentração em torno de cada cluster. Como o dataset em estudo possui diversas variáveis, o algoritmo de Análise de Componentes Principais (PCA) é empregue para reduzir a dimensionalidade dos dados, colocando os pontos de dados de acordo com as duas principais dimensões.

A Figura 20 exibe a visualização dos clusters obtidos pelos três algoritmos para o conjunto de dados de braquiterapia.

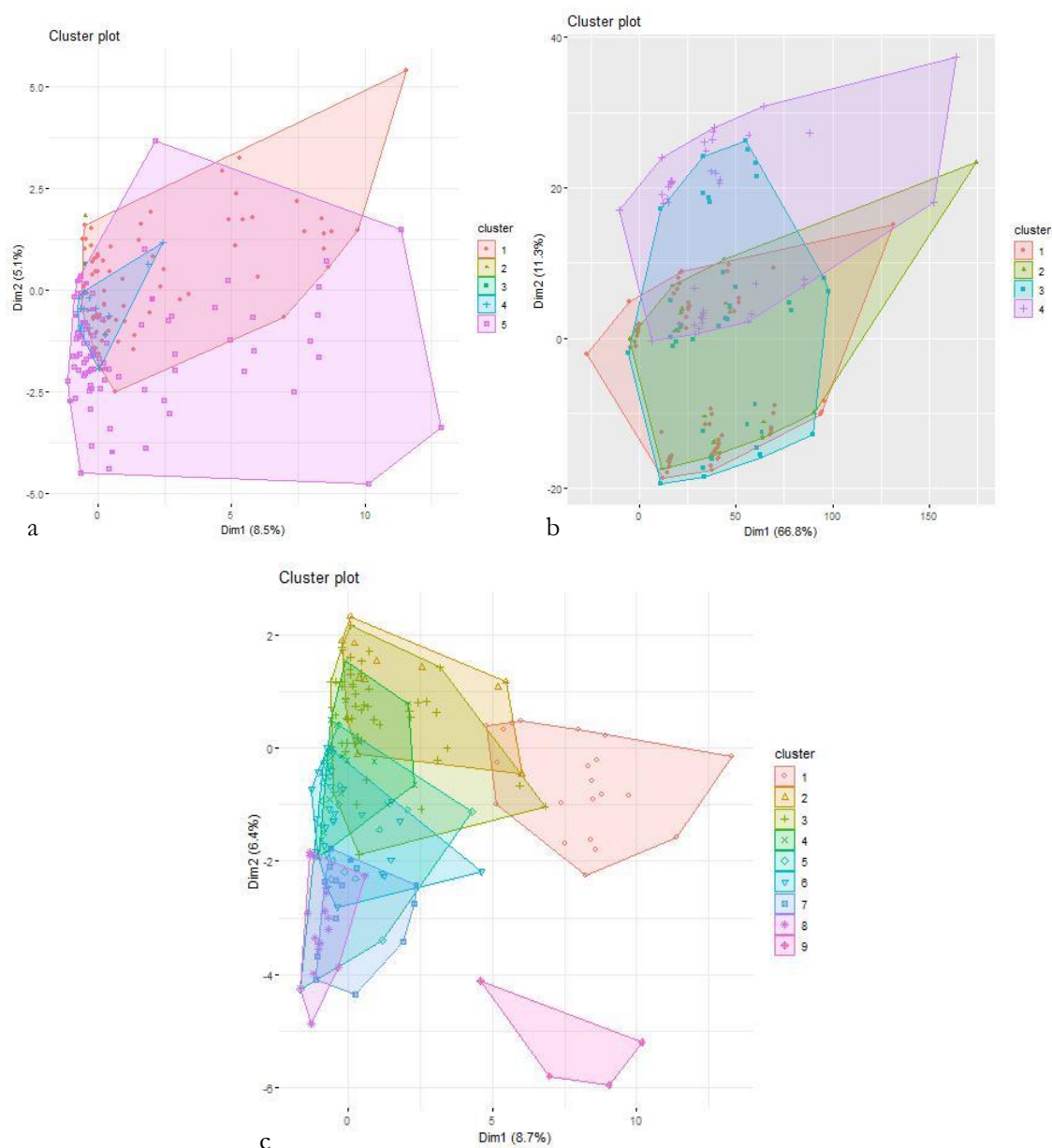


Figura 20 – Visualização dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico

Prostatectomia Radical

Relativamente ao conjunto de dados da prostatectomia radical, os resultados da Tabela 19 mostram a eficácia dos algoritmos de clustering aplicados. No caso do algoritmo k-modes, apenas o Cluster 4 expõem um coeficiente de silhueta médio positivo, o que sugere que o grupo têm características distintas e bem definidas. Por outro lado, os restantes clusters apresentam coeficientes de silhueta médio negativos, indicando uma sobreposição de características ou agrupamentos menos claros. Já o algoritmo PAM exibe coeficientes de silhueta médio positivos para todos os clusters, indicando que todos os grupos têm características distintas e são bem separados. Além disso, os clusters 2 e 5 apresentam coeficientes de silhueta médio mais altos, sugerindo agrupamentos particularmente bem definidos. Por fim, o algoritmo Hierárquico também gerou clusters com coeficientes de silhueta médio positivos, indicando uma boa qualidade de clustering. Assim como o dataset anterior, o cluster relativo à hipertensão foi detetado por todos os algoritmos e os clusters formados pelo PAM e Hierárquico

demonstraram uma maior separação e associação em características específicas, como a doença pulmonar obstrutiva crónica e a hipertensão concomitante com arritmia, diabetes mellitus e obesidade.

Tabela 19 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para prostatectomia radical, utilizando diferentes abordagens de clustering

Algoritmo	Clusters	Coefficiente de silhueta médio
K-modes	Cluster 1 (n= 1344; Sem características predominantes)	-0.23
	Cluster 2 (n= 1694; Hipertensão)	-0.42
	Cluster 3 (n= 99; Sem características predominantes)	-0.27
	Cluster 4 (n= 1825; Hipertensão)	1.0
PAM	Cluster 1 (n=2839; Hipertensão)	0.47
	Cluster 2 (n=921; Hipertensão, Diabetes Mellitus)	0.58
	Cluster 3 (n=366; Hipertensão, Arritmia)	0.25
	Cluster 4 (n=620; Hipertensão, Obesidade)	0.30
	Cluster 5 (n=216; Doença Pulmonar Obstrutiva Crónica)	0.59
Hierárquico	Cluster 1 (n= 2455; Hipertensão)	0.52
	Cluster 2 (n= 1008; Hipertensão, Diabetes Mellitus)	0.43
	Cluster 3 (n= 170; Depressão)	0.53
	Cluster 4 (n= 317; Hipertensão, Arritmia)	0.28
	Cluster 5 (n= 336; Doença Pulmonar Obstrutiva Crónica)	0.42
	Cluster 6 (n= 364; Hipertensão, Obesidade)	0.63
	Cluster 7 (n= 123; Tumor sólido não metastizado)	0.30
	Cluster 8 (n= 189; Álcool)	0.13

A Figura 21 apresenta a visualização dos clusters obtidos pelos três métodos aplicados, para o conjunto de dados relativo ao paciente que realizaram prostatectomia radical.

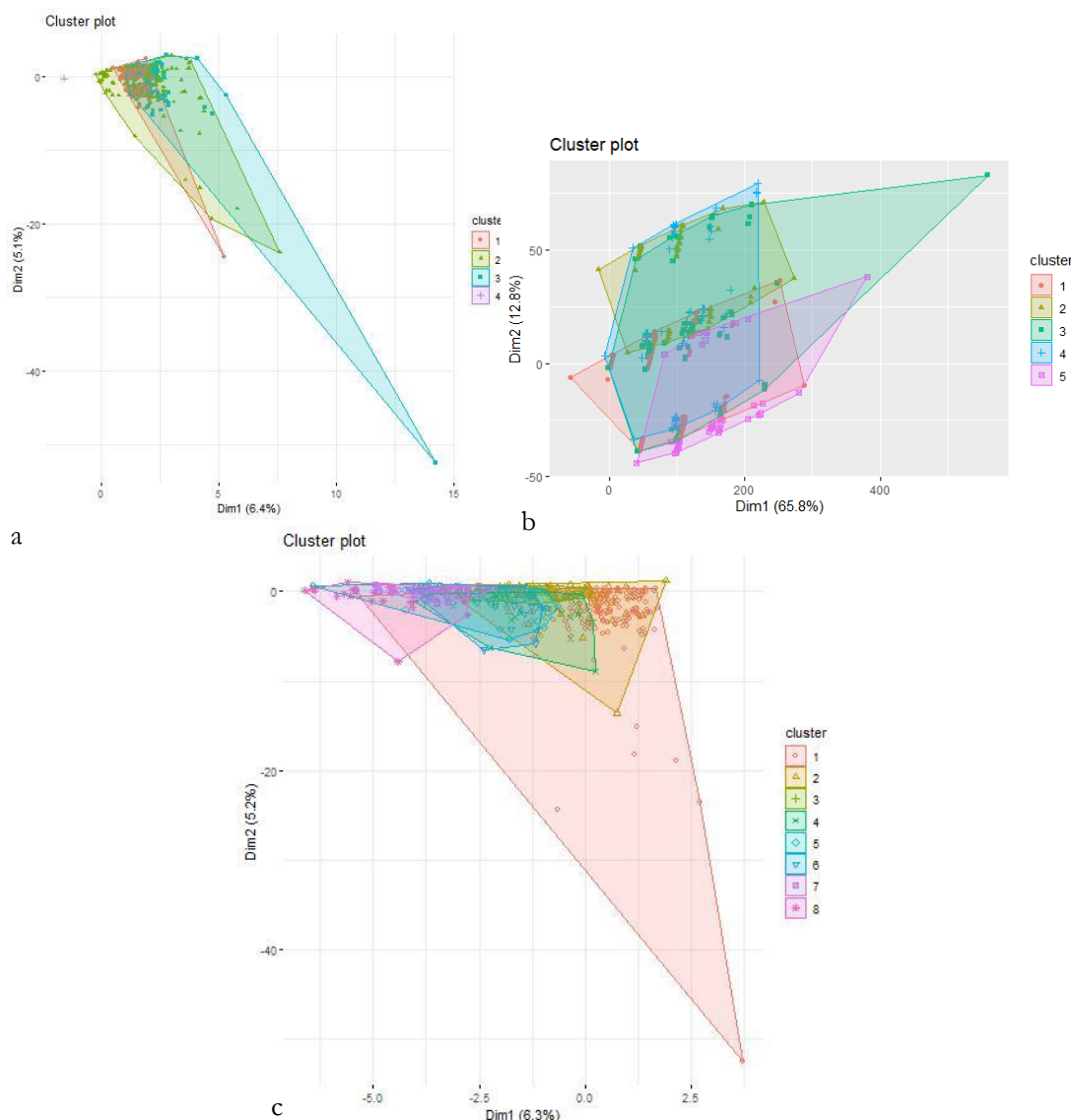


Figura 21 - Visualização dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, para prostatectomia radical, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico

Braquiterapia + Prostatectomia Radical

Por fim, para o conjunto de dados dos pacientes que realizaram braquiterapia ou prostatectomia radical foram identificados distintamente grupos de pacientes que apresentavam coocorrência comum de comorbilidades, nomeadamente, hipertensão, tanto em pacientes submetidos a braquiterapia como cirurgia, e hipertensão concomitante com diabetes, que foram encontradas usando as três abordagens de clustering (Tabela 20). Além disso, todos os métodos de clustering identificaram grupos semelhantes quanto às características principais, com exceção de dois clusters compostos por episódios com a maioria dos casos relatando Depressão/ Álcool e submetidos a cirurgia, obtidos apenas com Hierárquico (Cluster 3 e 7). O algoritmo k-modes apresentou clusters repetitivos, nomeadamente os Clusters 1 e 4, que não diferiam relativamente às características principais, dado que ambos os clusters apresentaram uma elevada prevalência de episódios de hipertensão e cirurgia.

Concluindo, os clusters obtidos pelos três métodos, para os conjuntos de dados analisados, são clinicamente relevantes pois identificam distintamente grupos de pacientes que apresentavam coocorrência comum de comorbilidades. Em termos de desempenho, o algoritmo k-modes demonstrou-

se mais instável em comparação com as outras abordagens, na medida em que tentativas repetidas geraram clusters distintos, além de, geralmente, apresentarem clusters redundantes. A aplicação de abordagens de clustering hierárquico e de PAM tendeu a produzir clusters mais estáveis, com grupos a apresentar características claramente definidas, sem redundância. Não obstante, como o coeficiente de silhueta médio evidenciou-se superior para os clusters obtidos pelo método de PAM, estes foram selecionados para avaliar e descrever os padrões de multimorbilidade nos internamentos por cancro da próstata em Portugal.

Tabela 20 - Clusters de multimorbilidade identificados em internamentos por cancro da próstata, para braquiterapia ou prostatectomia radical, utilizando diferentes abordagens de clustering

Algoritmo	Clusters	Coeficiente de silhueta médio
K-modes	Cluster 1 (n= 3690; Hipertensão, Cirurgia)	-0.09
	Cluster 2 (n= 628; Hipertensão, Diabetes Mellitus, Cirurgia)	0.67
	Cluster 3 (n= 882; Hipertensão, Braquiterapia)	-0.24
	Cluster 4 (n= 891; Hipertensão, Cirurgia)	0.08
PAM	Cluster 1 (n=2839; Hipertensão, Cirurgia)	0.49
	Cluster 2 (n=921; Hipertensão, Diabetes Mellitus, Cirurgia)	0.58
	Cluster 3 (n=366; Hipertensão, Arritmia, Cirurgia)	0.25
	Cluster 4 (n=620; Hipertensão, Obesidade, Cirurgia)	0.30
	Cluster 5 (n=265; Doença Pulmonar Obstrutiva Crónica, Cirurgia)	0.37
	Cluster 6 (n=1080; Hipertensão, Braquiterapia)	0.38
Hierárquico	Cluster 1 (n= 2577; Hipertensão, Cirurgia)	0.46
	Cluster 2 (n= 1008; Hipertensão, Diabetes Mellitus, Cirurgia)	0.44
	Cluster 3 (n= 170; Depressão, Cirurgia)	
	Cluster 4 (n= 319; Hipertensão, Arritmia, Cirurgia)	0.54
	Cluster 5 (n= 336; Doença Pulmonar Obstrutiva Crónica, Cirurgia)	0.27
	Cluster 6 (n= 364; Hipertensão, Obesidade, Cirurgia)	0.44
	Cluster 7 (n= 189; Álcool, Cirurgia)	
	Cluster 8 (n= 1128; Hipertensão, Braquiterapia)	0.64
		0.15
		0.33

A Figura 22 ilustra a representação visual dos clusters obtidos por meio dos três métodos empregues, aplicados ao conjunto de dados referente a pacientes submetidos a braquiterapia ou prostatectomia radical.

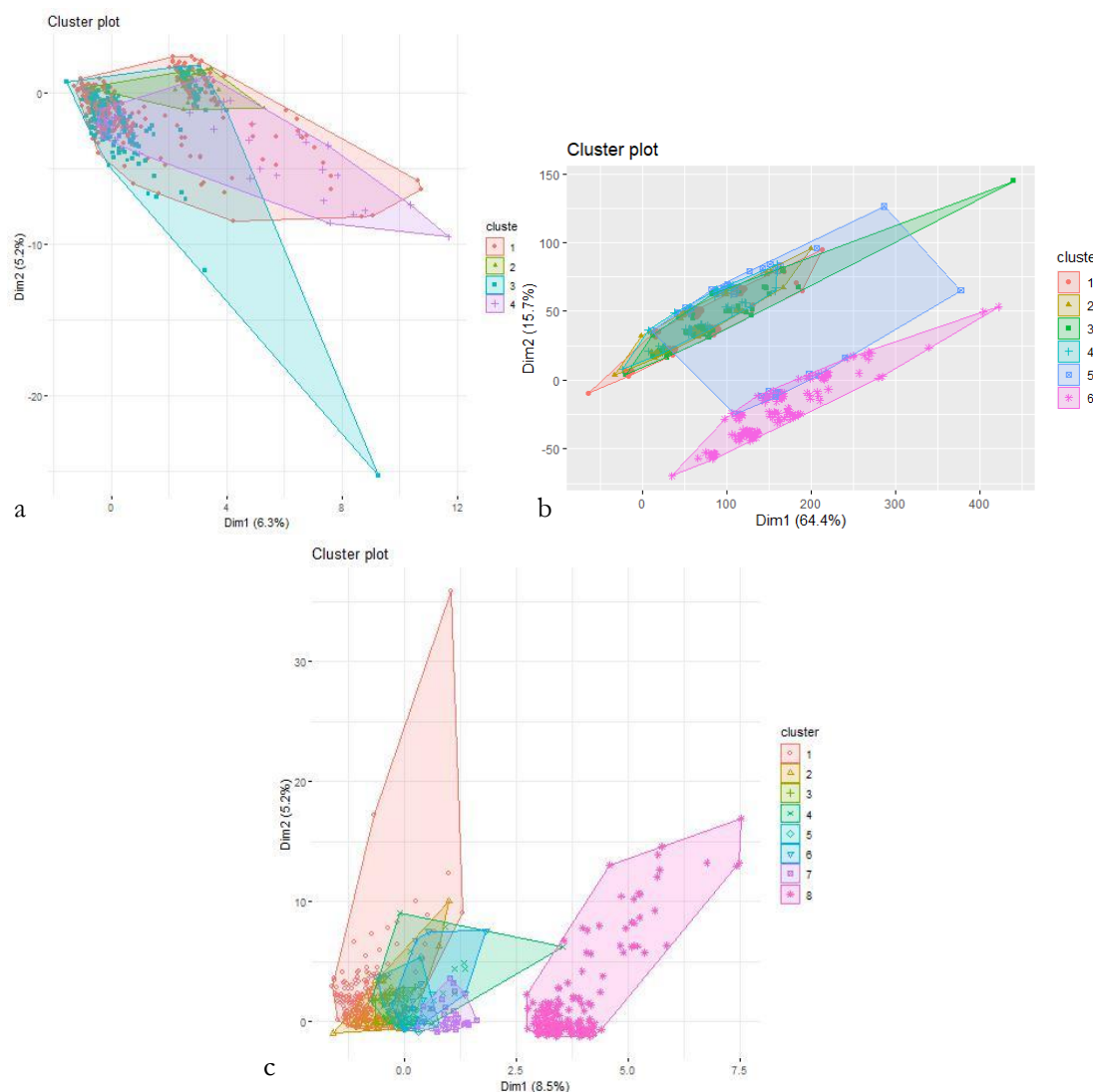


Figura 22 - Visualização dos clusters de multimorbidade identificados em internamentos por cancro da próstata, para braquiterapia ou prostatectomia radical, utilizando o algoritmo a) k-modes; b) PAM; c) hierárquico

4.1.3 Caracterização e interpretação dos Clusters de Multimorbidade obtidos pelo Método de PAM

A Tabela 21 apresenta a discriminação das características dos clusters obtidos pelo algoritmo de PAM, em comparação com o conjunto total de dados. Os resultados obtidos que englobam todos os internamentos demonstraram-se análogos e representativos relativamente a amostra da braquiterapia (Apêndice B) e da prostatectomia radical (Apêndice C). Somente as comorbilidades de Elixhauser com prevalência geral acima de 1% são manifestadas. As prevalências das comorbilidades destacadas são referentes a uma relação O/E maior que 2.

O Cluster 6, exclusivamente, concentrou todos os episódios de braquiterapia, apresentando uma elevada prevalência de hipertensão em comparação com a população geral em estudo. Tal como o Cluster 6, o Cluster 1 relatou igualmente uma preponderância proeminente de episódios com hipertensão, no entanto em casos que realizaram prostatectomia radical. Os clusters cirúrgicos 2, 3 e 4 agruparam episódios dominados pela coocorrência do carcinoma da próstata com hipertensão e uma condição crónica

específica, nomeadamente diabetes mellitus, arritmia e obesidade, respetivamente. Ainda que a maioria dos clusters exibam uma prevalência muito alta de hipertensão, o Cluster 5 apresentou uma relação O/E de 0, não apresentando assim nenhum caso com hipertensão. Todos os episódios desse cluster reportaram doença pulmonar obstrutiva crónica. No que se refere ao peso relativo de internamento, o Cluster 6, que engloba o total de episódios de braquiterapia, demonstrou custo esperado dos episódios bastante superior relativamente aos internamentos por prostatectomia radical. Por fim, como esperado, os clusters referentes à prostatectomia radical manifestaram um maior tempo de internamento comparativamente aos episódios de braquiterapia. Não foram observadas diferenças significativas em relação à idade média, mortalidade e metástases nos seis grupos.

Tabela 21 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbilidade n=4303
Arritmia, n(%)	0 (0)	0 (0)	366 (100)	0 (0.0)	17 (6.4)	125 (12)	508 (4.9)	-
Hipertensão, n(%)	2377 (84)	688 (75)	242 (66)	445 (72)	0 (0.0)	844 (78)	4596 (44)	-
Doença Pulmonar Obstrutiva Crónica, n(%)	151 (5.3)	40 (4.3)	21 (5.7)	40 (6.5)	265 (100)	64 (5.9)	581 (5.6)	-
Diabetes Mellitus, n(%)	0 (0.0)	921 (100)	53 (14)	161 (26)	20 (7.5)	258 (24)	1413 (14)	-
Doença renal, n(%)	85 (3.0)	27 (2.9)	13 (3.6)	13 (2.1)	5 (1.9)	29 (2.7)	172 (1.7)	-
Úlcera péptica, n(%)	61 (2.1)	11 (1.2)	3 (0.8)	4 (0.6)	5 (1.9)	18 (1.7)	102 (1.0)	-
Tumor, n(%)	105 (3.7)	15 (1.6)	6 (1.6)	16 (2.6)	7 (2.6)	39 (3.6)	188 (1.8)	-
Obesidade, n(%)	0 (0)	0 (0)	45 (12)	620 (100)	28 (11)	133 (12)	826 (7.9)	-
Abuso de álcool, n(%)	141 (5.0)	29 (3.1)	13 (3.6)	38 (6.1)	23 (8.7)	140 (13)	384 (3.7)	-
Depressão, n(%)	169 (6.0)	20 (2.2)	11 (3.0)	22 (3.5)	11 (4.2)	55 (5.1)	288 (2.8)	-
Metástases, n(%)	78 (2.7)	29 (3.1)	9 (2.5)	17 (2.7)	8 (3.0)	59 (5.5)	336 (3.2)	136 (3.2)
Nºcomorbilidades, mediana (IQR)	1 (1-1)	2 (2-2)	2 (2-3)	2 (2-3)	1 (1-2)	1 (1-2)	1 (0-2)	-
Idade, mediana (IQR)	65 (61-69)	65 (62-69)	66 (62-70)	64 (60-68)	65 (60-68)	67 (62-71)	65 (60-69)	64 (59-68)

Tempo de internamento, mediana (IQR)	6 (4-7)	6 (4-7)	6 (4-7)	5 (4-7)	5 (3-7)	2 (2-3)	5 (4-7)	5 (4, 7)
Peso Relativo de Internamento, média (s)	1.13 (±0.74)	1.19 (±0.89)	1.30 (±0.86)	1.17 (±0.85)	2.57 (±2.95)	8.50 (±1.25)	6.36 (±1.71)	5.35 (±1.56)
Braquiterapia, n(%)	0 (0)	0 (0)	0 (0)	0 (0.0)	49 (18)	1080 (100)	1830 (18)	701 (16)
Cirurgia, n(%)	2,839 (100)	921 (100)	366 (100)	620 (100)	216 (82)	0 (0.0)	8565 (82)	3603 (84)
Morte, n(%)	1 (<0.1)	0 (0.0)	4 (1.1)	1 (0.2)	0 (0.0)	19 (1.8)	31 (0.3)	6 (0.1)

A Tabela 22 fornece uma visão mais detalhada do modo como os pacientes com carcinoma da próstata foram admitidos nos hospitais, segundo o cluster em que foram dispostos. Em geral, pode-se observar que a maioria dos pacientes em todos os clusters foram admitidos de forma programada, o que sugere que os tratamentos foram agendados com antecedência. A proporção de casos urgentes é reduzida na generalidade dos clusters à exceção do Cluster 6, onde se encontram os episódios de internamento que realizaram braquiterapia. A presença de urgência nesse cluster pode indicar complicações adicionais que exigiram uma abordagem mais célere. A categoria "Medicina Privada" apresenta um número irrelevante de casos em todos os clusters, registando apenas um único paciente, nessa categoria, no Cluster 6.

Tabela 22 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbilidade n=4303
Programada, n(%)	2,807 (99)	907 (98)	365 (100)	615 (99)	262 (99)	1,026 (95)	10234 (98)	4252 (99)
Urgente, n(%)	32 (1.1)	14 (1.5)	1 (0.3)	5 (0.8)	3 (1.1)	53 (4.9)	159 (1.5)	51 (1.2)

A análise da Tabela 23 manifesta informações relevantes acerca da distribuição dos pacientes e a severidade da sua multimorbilidade. A maior parte dos pacientes em todos os clusters apresentou uma severidade menor relativamente às doenças que possuem.

Tabela 23 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbilidade n=4303
Menor, n(%)	2657 (94)	843 (92)	293 (80)	573 (92)	240 (91)	972 (90)	9725 (94)	4147 (96)
Moderado, n(%)	164 (5.8)	71 (7.7)	60 (16)	43 (6.9)	22 (8.3)	75 (6.9)	581 (5.6)	146 (3.4)

Major, n(%)	14 (0.5)	7 (0.8)	10 (2.7)	3 (0.5)	3 (1.1)	33 (3.1)	80 (0.8)	10 (0.2)
Extremo, n(%)	4 (0.1)	0 (0)	3 (0.8)	1 (0.2)	0 (0)	0 (0)	8 (<0.1)	0 (0)

A probabilidade de ocorrer a morte do paciente, de acordo com o cluster em que se encontravam, manifestou diferenças consideráveis (Tabela 24). Os Clusters 1 e 2, que possuem o maior número de pacientes que realizaram prostatectomia radical, são caracterizados pela classificação "Menor" em relação ao risco de mortalidade, sugerindo que a maioria dos pacientes no estudo possui um risco reduzido de mortalidade. Por outro lado, os Clusters 4 e 6, relativos a episódios que realizaram prostatectomia radical com elevada prevalência de hipertensão e obesidade e os episódios que realizaram braquiterapia, respetivamente, exibiram que a generalidade desses casos apresenta uma probabilidade moderada de mortalidade, indicando um risco superior em comparação com a população geral em estudo. Por fim, os Clusters 3 e 5 revelaram uma prevalência semelhante de um risco de mortalidade reduzido e moderado.

Tabela 24 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbi- lidade n=4303
Menor, n(%)	2323 (82)	738 (80)	163 (45)	96 (15)	90 (34)	63 (5.8)	6,923 (67)	3450 (80)
Moderado, n(%)	477 (17)	159 (17)	178 (49)	505 (81)	155 (58)	833 (77)	3122 (30)	815 (19)
Major, n(%)	34 (1.2)	23 (2.5)	22 (6.0)	19 (3.1)	20 (7.5)	183 (17)	338 (3.3)	37 (0.9)
Extremo, n(%)	5 (0.2)	1 (0.1)	3 (0.8)	0 (0)	0 (0)	1 (<0.1)	11 (0.1)	1 (<0.1)

Os resultados da Tabela 25 refletem a distribuição do grupo hospitalar onde se encontravam os pacientes internados com cancro da próstata. Analisando os resultados, o Grupo E revelou uma maior predominância em todos os clusters à exceção do Cluster 6. Isso sugere que pode ser mais comum pacientes com carcinoma da próstata serem encaminhados para os hospitais associados a esse grupo, nomeadamente o Centro Hospitalar Universitário (CHU) do Porto, CHU de São João, CHU de Coimbra, CHUs de Lisboa Central, Ocidental e Norte. No Cluster 6, o Grupo F tem um número significativamente superior de pacientes em comparação com os outros grupos, podendo indicar que episódios de braquiterapia são tratados com maior frequência nos Institutos Portugueses de Oncologia.

Tabela 25 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbi- lidade n=4303
Grupo A, n(%)	18 (0.6)	6 (0.7)	1 (0.3)	10 (1.6)	1 (0.4)	0 (0)	44 (0.4)	8 (0.2)
Grupo B, n(%)	81 (2.9)	32 (3.5)	14 (3.9)	23 (3.7)	7 (2.6)	1 (<0.1)	343 (3.3)	185 (4.4)

Grupo C, n(%)	696 (25)	247 (27)	76 (21)	210 (34)	69 (26)	18 (1.7)	2172 (21)	856 (20)
Grupo D, n(%)	558 (20)	160 (17)	49 (14)	71 (12)	34 (13)	17 (1.6)	1730 (17)	841 (20)
Grupo E, n(%)	1048 (37)	364 (40)	148 (41)	229 (37)	78 (29)	90 (8.3)	3318 (32)	1361 (32)
Grupo F, n(%)	413 (15)	108 (12)	71 (20)	74 (12)	76 (29)	954 (88)	2678 (26)	982 (23)
Outros	25	4	7	3	0	0	109	70

Ao analisar a distribuição geográfica dos hospitais onde os pacientes de cada cluster estavam internados para o respetivo tratamento, observaram-se desigualdades significativas (Tabela 26). É possível verificar que algumas regiões hospitalares, como o Norte e Centro, têm uma presença mais relevante em vários clusters, indicando uma maior incidência de carcinoma da próstata e respetivas comorbilidades nessas áreas geográficas. O cluster 6 apresentou uma maior representatividade na região "Centro", enquanto o cluster 4 demonstrou uma distribuição mais equilibrada entre as diferentes regiões. Já os clusters restantes foram mais prevalentes na região "Norte". As regiões do Alentejo e do Algarve apresentaram proporções relativamente menores de pacientes em todos os clusters. Essas discrepâncias podem ser atribuídas a diversos fatores, incluindo densidade populacional e concentração de hospitais.

Tabela 26 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394	Sem Multimorbili dade n=4303
Norte, n(%)	1227 (44)	371 (40)	146 (41)	228 (37)	122 (46)	370 (34)	4215 (41)	1751 (41)
Centro, n(%)	670 (24)	228 (25)	95 (26)	168 (27)	63 (24)	572 (53)	2659 (26)	863 (20)
Lisboa, n(%)	824 (29)	287 (31)	114 (32)	211 (34)	72 (27)	132 (12)	3050 (30)	1410 (33)
Alentejo, n(%)	51 (1.8)	20 (2.2)	4 (1.1)	8 (1.3)	7 (2.6)	2 (0.2)	192 (1.9)	100 (2.4)
Algarve, n(%)	42 (1.5)	11 (1.2)	1 (0.3)	2 (0.3)	1 (0.4)	4 (0.4)	172 (1.7)	111 (2.6)
Outros	25	4	6	3	0	0	106	68

A Tabela 27 apresenta o peso para as comorbilidades Elixhauser presentes em cada cluster. Porém, apenas os pesos com uma preponderância geral acima de 1% se encontram evidenciados. Todos os clusters apresentaram uma frequência reduzida dos pesos mais prevalentes da população em estudo. O Cluster 4 apresenta a média mais baixa (5.79), indicando que os pacientes neste cluster tendem a ter um peso associado às comorbilidades de Elixhauser menor.

Tabela 27 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=265	Cluster 6 n=1080	Total n=10394
-4, n(%)	0 (0)	0 (0)	1 (0.3)	116 (19)	0 (0)	9 (0.8)	556 (5.3)
-3, n(%)	78 (2.7)	0 (0)	6 (1.6)	250 (40)	15 (5.7)	60 (5.6)	194 (1.9)
-1, n(%)	28 (1.0)	12 (1.3)	19 (5.2)	29 (4.7)	3 (1.1)	27 (2.5)	113 (1.1)
0, n(%)	19 (0.7)	3 (0.3)	6 (1.6)	3 (0.5)	0 (0)	15 (1.4)	7,625 (73)
3, n(%)	42 (1.5)	99 (11)	62 (17)	0 (0)	9 (3.4)	55 (5.1)	432 (4.2)
4, n(%)	10 (0.4)	17 (1.8)	11 (3.0)	0 (0)	2 (0.8)	11 (1.0)	140 (1.3)
5, n(%)	3 (0.1)	1 (0.1)	2 (0.5)	0 (0)	1 (0.4)	2 (0.2)	450 (4.3)
12, n(%)	14 (0.5)	1 (0.1)	0 (0)	0 (0)	3 (1.1)	6 (0.6)	244 (2.3)
média (s)	9.47 (±2.59)	10.38 (±2.55)	9.88 (±2.89)	5.79 (±2.98)	9.58 (±3.38)	9.66 (±3.46)	0.81 (±3.37)

4.1.4 Avaliação e interpretação dos clusters

O algoritmo de árvore de decisão foi empregue para compreender, de modo visual, quais os atributos (nós) que mais corretamente classificam os clusters obtidos pelo método de PAM (folhas). Para aplicação da metodologia de árvores de decisão foi aplicado o pacote CART. A árvore de decisão obtida é apresentada na Figura 23 e é constituída por dezassete nós, dos quais nove são folhas.

A análise em detalhe da árvore de decisão permite verificar que, no primeiro momento, no nível 1 é efetuada a separação total entre os clusters através da realização da braquiterapia, isolando o Cluster 6 dos restantes. Ademais, é indicado que 81% da população em estudo não realizou braquiterapia. De seguida, para o segundo nível da árvore de decisão, a variável Diabetes Mellitus permite separar os Cluster 1 e 2, sendo que no Cluster 2 todos os pacientes detinham diabete mellitus inversamente ao Cluster 1. Nos níveis 3 e 4, a divisão é realizada entre quem possui ou não obesidade e entre pacientes que possuem ou não arritmia, respetivamente. Em contrapartida, a ramificação no nó 16 é efetuada com a variável Doença pulmonar obstrutiva crónica e, quando os pacientes assumem esta condição, 53% dos episódios são considerados como pertencentes ao Cluster 5. Por fim, no nível 6 a variável Hipertensão reparte os Clusters 1 (possui esta comorbilidade) e 5 (não possui esta comorbilidade).

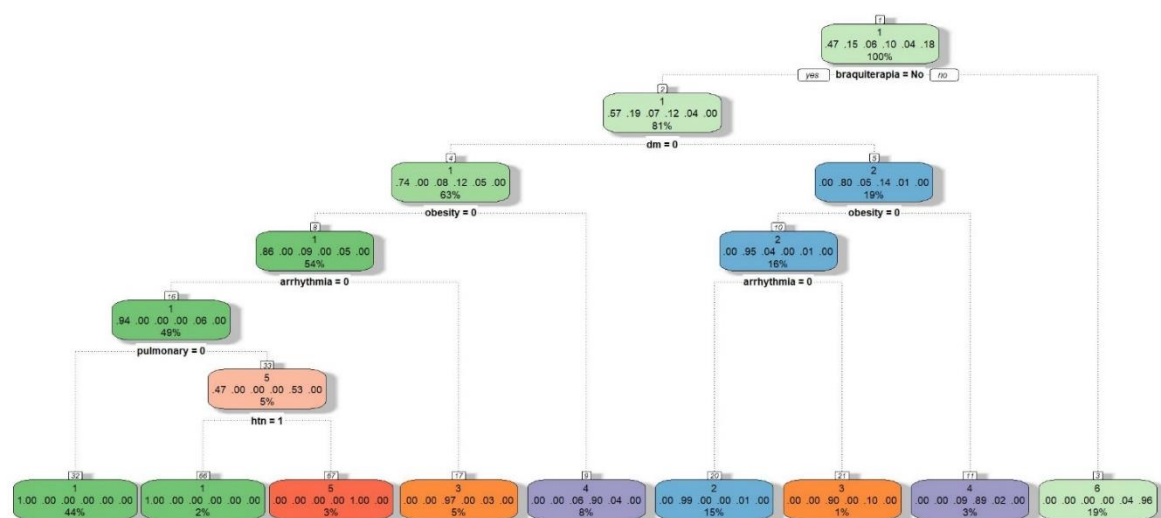


Figura 23 - Estrutura da árvore de decisão dos clusters de multimorbilidade identificados em internamentos por cancro da próstata, utilizando o algoritmo PAM

Relativamente às estatísticas gerais, a exatidão global (*accuracy*) da classificação, neste caso, foi de 98% com um intervalo de confiança de (0.97, 0.98), o que indica que os dados foram bem classificados. Outra medida de avaliação da exatidão global é o índice de concordância Kappa que pode ter valores entre -1 e 1 e, visto que se obteve um valor de 0.97, pode-se relatar uma concordância quase perfeita. A sensibilidade é a probabilidade de acerto na atribuição de um objeto a um dado cluster e a especificidade a probabilidade de acerto na previsão de um objeto não corresponder a um cluster. O poder preditivo do modelo é definido pela probabilidade de acerto nas previsões desse modelo. Os resultados da análise da sensibilidade, especificidade, valor preditivo e exatidão, para cada cluster, encontram-se declarados na Tabela 28.

Tabela 28 – Estatísticas segundo cada cluster de multimorbilidade obtido com o algoritmo PAM

	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Sensibilidade	1.0	0.99	0.96	0.90	1.0	0.96
Especificidade	1.0	1.0	0.99	1.0	0.98	1.0
Valor preditivo	1.0	1.0	0.88	1.0	0.64	1.0
Exatidão	1.0	1.0	0.98	0.94	0.99	0.98

4.2 Discussão de resultados

Nesta dissertação, utilizou-se análise de clusters para descrever padrões de multimorbilidade em internamentos por cancro da próstata que relataram braquiterapia ou prostatectomia radical. Na extensão do nosso conhecimento, este é o primeiro estudo a utilizar a análise de clusters para avaliar padrões de multimorbilidade nos internamentos por cancro da próstata em Portugal, tendo uma natureza quase nacional devido à dimensão amostral do estudo, onde o número de episódios incluídos foram obtidos ao longo de 6 anos e de todos os hospitais do Serviço Nacional de Saúde português.

Dos 10 394 pacientes admitidos no hospital com cancro da próstata, aproximadamente 58% (6091) apresentavam multimorbilidade, ou seja, a coexistência de múltiplas doenças crónicas. No entanto, não foi observado um impacto no número de óbitos em decorrência de um maior número de doenças crónicas presentes nos pacientes. De acordo com Rizzuto et al (2017), a relação entre multimorbilidade e mortalidade tem sido objeto de diversos estudos, contudo os resultados obtidos têm sido inconsistentes, com alguns estudos a relatarem uma associação negativa entre a presença de múltiplas condições crónicas e a sobrevida dos indivíduos, enquanto outros não encontraram essa associação [100]. Além disso, em contrapartida às evidências previamente documentadas na literatura [101], os resultados encontrados neste estudo não corroboram uma relação entre o número de comorbilidades e a duração do internamento hospitalar. Não obstante, ao analisar a duração do internamento hospitalar relativamente ao tipo de tratamento realizado, o tempo mediano para os pacientes submetidos à prostatectomia radical foi 6 dias, o que se revelou significativamente mais longo quando comparado aos pacientes que submetidos a braquiterapia (2 dias). É importante salientar que tanto a prostatectomia radical quanto a braquiterapia são tratamentos comumente indicados para casos de cancro da próstata em estadios inicial e não metastático. Posto isto, foi observada uma frequência incomum de casos de metástases em 336 pacientes. Nesses casos, a maioria realizou prostatectomia radical (75%), enquanto a braquiterapia foi o tratamento preferencial para os restantes pacientes (25%), ou seja, verifica-se um padrão de tratamento diferente do que se encontra estabelecido nas *guidelines* tradicionais. Estes últimos pacientes foram predominantemente admitidos no grupo hospitalar E, composto pelos principais hospitais públicos universitários de Portugal.

Diferentes abordagens de clustering foram aplicadas, de modo geral, produzindo clusters semelhantes em termos das principais características clínicas. A análise de clusters demonstrou-se uma abordagem útil e prática para detetar padrões de multimorbilidade, visto que diferenciou claramente os episódios de acordo com as duas principais vias de tratamento para pacientes hospitalizados em Portugal, além de revelar coocorrências prevalentes de doenças com potencial impacto na sobrevida oncológica.

Os seis clusters identificados mediante o algoritmo de clustering *Partitioning Around Medoids* (PAM) foram selecionados para posterior avaliação, de modo a alcançar o objetivo proposto. No que se refere ao número de comorbilidades, os vários clusters apresentaram um número semelhante, consistindo o número máximo em duas doenças crónicas para além do diagnóstico principal de cancro da próstata (Clusters 2, 3 e 4). Em relação aos principais padrões observados, dois clusters (Cluster 1 e Cluster 6) foram compostos exclusivamente por episódios com elevada prevalência de hipertensão. Esse padrão foi encontrado tanto para os episódios que relataram prostatectomia radical quanto para braquiterapia. Os clusters 2, 3 e 4 foram compostos por episódios cirúrgicos relatando diabetes mellitus, arritmia e obesidade, respetivamente, concomitantes a hipertensão e ao cancro da próstata. No Cluster 5, a DPOC coocorreu principalmente com o cancro da próstata isolado. A coexistência de condições clínicas de diferentes especialidades médicas no mesmo cluster constitui uma situação desafiadora para médicos e pacientes.

As nossas evidências foram alinhadas com a literatura anterior, na qual o diagnóstico de cancro da próstata coocorreu principalmente com comorbilidades como a hipertensão, arritmia, obesidade e diabetes mellitus [63],[102]. A ideia principal é que as doenças não surgem de forma aleatória e independente nos pacientes. Os clusters foram constituídos por episódios que exibiam condições que podem estar associadas a fatores de risco modificáveis, com impacto potencial no risco de desenvolver ou morrer de cancro da próstata, como uma dieta rica em gordura e tabagismo [103],[104]. Contudo, uma vez que não existem evidências suficientes a respeito de indicações claras de prevenção, além do diagnóstico precoce, para reduzir a mortalidade por cancro da próstata [7], o foco deve-se sublinhar no aumento dos esforços para controlar doenças crónicas, principalmente a hipertensão, durante o tratamento do cancro, o que inclui esforços para controlar comportamentos em todas as fases da

sobrevivência do cancro da próstata. Os clusters identificados são clinicamente relevantes na medida em que revelam potenciais grupos de risco que devem ser visados a fim de garantir maior sobrevida e evitar visitas não programadas aos serviços de saúde.

Este trabalho também permitiu efetuar inferências acerca da probabilidade de morte da população investigada. A generalidade dos pacientes pertencentes aos Clusters 4 e 6, que realizaram prostatectomia radical e braquiterapia, respetivamente, exibiram uma probabilidade moderada de mortalidade. Estes clusters demonstraram uma elevada prevalência de hipertensão e hipertensão em conjunto com obesidade, no caso do Cluster 4. Esses resultados podem indicar que pacientes com essas características clínicas específicas podem ter um risco acrescido de complicações e desfechos adversos. Essa associação entre comorbilidades e risco de mortalidade é consistente com estudos anteriores que identificaram que o excesso de gordura acumulada pode intensificar até 10% o risco de morte por carcinoma da próstata [101]. No entanto, a predominância de pacientes com risco de mortalidade considerado “Menor” sugere que, em geral, a mortalidade das comorbilidades relacionadas com o cancro da próstata na amostra em estudo é relativamente baixa.

A distribuição geográfica dos pacientes revelou-se diferente consoante os clusters. Regiões hospitalares como o Norte, o Centro e Lisboa foram mais representativas em vários clusters, indicando uma possível maior incidência de carcinoma da próstata e respetivas comorbilidades nessas áreas geográficas. Estas observações corroboram o RON - Registo Oncológico Nacional que identificaram variações regionais na incidência do cancro da próstata em Portugal, com um maior número de novos casos, no ano de 2018, na região de Lisboa (2 016), Norte (1 821) e Centro (1 125) [105]. Essa concordância entre as evidências deste estudo e os dados do RON reforça a validade das informações obtidas, salientando a necessidade de se considerar redes de referência oncológica nessas áreas demográficas, onde doentes oncológicos são observados, orientados e tratados por equipas multidisciplinares, de modo a disponibilizar tratamentos muito especializados. Em relação ao tipo de tratamento, verificou-se uma distribuição geográfica heterogénea e assimétrica, com o Cluster 6, ou seja, o grupo de pacientes que realizaram braquiterapia, a demonstrar uma prevalência mais exacerbada no centro do país, diferindo da prevalência verificada nos clusters cirúrgicos.

Por fim, é possível aferir que existem padrões de tratamento diferentes a nível nacional, com especial ênfase no grupo hospitalar F, o grupo constituído pelos Institutos Portugueses de Oncologia, onde se verifica que optam pela braquiterapia como tratamento padrão para os pacientes com carcinoma da próstata. Relativamente às restantes variáveis consideradas, não houve diferenças significativas entre pacientes que realizaram braquiterapia e os que realizaram prostatectomia radical.

5 CONCLUSÃO

Neste capítulo são apresentadas algumas considerações finais sobre o estudo realizado, que teve como objetivo identificar e caracterizar padrões de multimorbilidade em pacientes internados com cancro da próstata. Serão igualmente mencionadas as limitações sentidas nesta investigação e as perspetivas de desenvolvimento futuro do presente trabalho.

5.1 Conclusões finais

A análise de clustering é uma abordagem útil para detetar e caracterizar diferentes padrões e perfis de episódios de internamento por cancro da próstata em hospitais portugueses. A multimorbilidade encontra-se presente em 58% dos homens admitidos para tratamento de cancro da próstata, essas condições clínicas adicionais são potencialmente responsáveis pela causa da morte, destacando assim a importância de abordar o estado de comorbilidade, através da integração e coordenação do tratamento de doenças crónicas complementarmente ao tratamento oncológico. Em particular, este estudo revelou que a coocorrência de hipertensão e diabetes pode possivelmente incorrer num *burden* superior nas hospitalizações por motivo de cancro da próstata em Portugal, exigindo, portanto, uma maior atenção durante o tratamento. No entanto, não se verificou um impacto significativo no número de óbitos ou no tempo de internamento hospitalar com a presença de um maior número de doenças crónicas nos pacientes.

5.2 Limitações e investigação futura

A análise da multimorbilidade em pacientes hospitalizados permanece pouco explorada e, como tal, o nosso trabalho traz mais avanços nesse campo. Contudo, é importante indicar as limitações subjacentes a este estudo. Em primeiro lugar, os dados hospitalares utilizados para as análises foram obtidos de uma base de dados administrativa que, por sua vez, apresenta problemas de qualidade subjacentes, principalmente relacionados com a codificação clínica e o grau em que as comorbilidades são documentadas e codificadas [106]. Neste sentido, em Portugal, todos os procedimentos de codificação são realizados por médicos com formação específica e verificados por auditorias internas e externas. A precisão dos dados está intrinsecamente ligada à precisão da codificação clínica, e os hospitais podem apresentar variações significativas em termos de práticas de codificação clínica, o que pode impactar substancialmente a qualidade dos dados obtidos. Uma limitação adicional deste estudo é a exclusão da análise do terceiro tratamento padrão para pacientes com carcinoma da próstata, que é a radioterapia externa. Esta modalidade de tratamento é realizada em regime ambulatorio e é mais comumente selecionada para pacientes com maior número de comorbilidades e de idade avançada. Além disso, as abordagens de clustering consideradas neste estudo não consideram as mudanças temporais na condição dos pacientes e os episódios são restritos e assumidos como parte de um único cluster, ao invés da possibilidade de pertencer a mais de um cluster simultaneamente. Por último, como os dados foram desidentificados, os pacientes não puderam ser identificados nem acompanhados de forma longitudinal.

Para combater esta última limitação, a sequência lógica deste trabalho será a aplicação de técnicas de *record linkage*, uma abordagem valiosa para vincular e acompanhar os dados longitudinais dos pacientes ao longo do tempo, permitindo compreender a evolução das comorbilidades e avaliar a associação entre diferentes clusters e a sobrevivência. Outra proposta futura será replicar a abordagem desenvolvida em outros contextos, nomeadamente, outros tipos de cancro e regimes (ambulatorio).

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] J. A. E. Pina, *Anatomia Humana dos Órgãos (2ª Edição)*. Lidel, 2017.
- [2] M. C. Ward and R. D. Tendulkar, *Essentials of Clinical Radiation Oncology*. Springer Publishing Company, 2018. doi: 10.1891/9780826168559.
- [3] J. H. Mydlo and C. J. Godec, *Prostate Cancer: Science and Clinical Practice: Second Edition*. Elsevier Ltd, 2016. doi: 10.1016/C2013-0-09891-X.
- [4] T. Matos *et al.*, “Os Técnicos de Diagnóstico e Terapêutica no Cancro da Próstata,” *Newsletter TDT*, no. 21, p. 19, 2015.
- [5] L. L. Gunderson and J. E. Tepper, *Clinical Radiation Oncology: Third Edition*. Elsevier Saunders, 2012. doi: 10.1016/B978-1-4377-1637-5.00023-7.
- [6] A. Figueiredo, M. Pereira, P. Príncipe, R. Nogueira, T. Lopes, and T. Maricoto, *Sintomas do Trato Urinário Inferior (LUTS) no Homem*. Springer Healthcare Ibérica S.L., 2017.
- [7] G. Gandaglia *et al.*, “Epidemiology and Prevention of Prostate Cancer,” *Eur. Urol. Oncol.*, vol. 4, no. 6, pp. 877–892, 2021, doi: 10.1016/j.euo.2021.09.006.
- [8] American Cancer Society, “Cancer Facts & Figures,” Atlanta: American Cancer Society, 2015.
- [9] Liga Portuguesa Contra o Cancro, “Sinais de Alerta - Cancro da Próstata.” <https://www.ligacontracancro.pt/cancro-da-prostata-sinais-de-alerta/> (accessed Jul. 10, 2023).
- [10] C. Babb, M. Urban, D. Kielkowski, and P. Kellett, “Prostate Cancer in South Africa: Pathology Based National Cancer Registry Data (1986–2006) and Mortality Rates (1997–2009),” *Hindawi Publ. Corp.*, pp. 1–9, 2014, doi: 10.1155/2014/419801.
- [11] H. Sung *et al.*, “Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries,” *CA. Cancer J. Clin.*, vol. 71, pp. 209–249, 2021, doi: 10.3322/caac.21660.
- [12] J. Ferlay *et al.*, “Global Cancer Observatory: Cancer Today,” *International Agency for Research on Cancer*, 2020. <https://gco.iarc.fr/today/home> (accessed Jul. 10, 2023).
- [13] Direção-Geral da Saúde, “Programa Nacional para as Doenças Oncológicas,” 2019.
- [14] A. C *et al.*, “Global surveillance of trends in cancer survival: analysis of individual records for 37,513,025 patients diagnosed with one of 18 cancers during 2000–2014 from 322 population-based registries in 71 countries (CONCORD-3),” *Lancet*, vol. 391, no. 10125, pp. 1023–1075, 2018, doi: 10.1016/S0140-6736(17)33326-3.Global.
- [15] P. Rawla, “Epidemiology of Prostate Cancer,” *Elmer Press*, vol. 10, no. 2, pp. 63–89, 2019, doi: 10.1016/j.mednuc.2007.11.003.
- [16] N. Mottet, E. Briers, M. De Santis, S. Gillessen, and J. Grummet, “EAU - EANM - ESTRO - ESUR - ISUP - SIOG - Guidelines on Prostate Cancer,” *Eur. Assoc. Urol.*, 2023, [Online]. Available: http://www.uroweb.org/fileadmin/tx_eauguidelines/2005/Pocket/Prostate_Cancer.pdf
- [17] D. Jones, C. Friend, A. Dreher, V. Allgar, and U. Macleod, “The diagnostic test accuracy of rectal examination for prostate cancer diagnosis in symptomatic patients: A systematic review,” *BMC Fam. Pract.*, vol. 19, no. 1, pp. 1–6, 2018, doi: 10.1186/s12875-018-0765-y.
- [18] L. Naji *et al.*, “Digital rectal examination for prostate cancer screening in primary care: A systematic review and meta-analysis,” *Ann. Fam. Med.*, vol. 16, no. 2, pp. 149–154, 2018, doi: 10.1370/afm.2205.
- [19] P. W. Sperduto and J. P. Gibbons, *Khan’s Treatment Planning in Radiation Oncology*, Fifth edit. Wolters Kluwe, 2022.

- [20] Sociedade Portuguesa de Oncologia, “Revista Portuguesa de Oncologia - Consenso Nacional de Cancro da Próstata,” *Rev. Port. Oncol.*, 2019, [Online]. Available: <https://apurologia.pt/wp-content/uploads/2020/01/cons-nac-cancro-prostata-.pdf>
- [21] F. Z. Chen and X. K. Zhao, “Prostate cancer: Current treatment and prevention strategies,” *Iran. Red Crescent Med. J.*, vol. 15, no. 4, pp. 279–284, 2013, doi: 10.5812/ircmj.6499.
- [22] IAEA - International Atomic Energy Agency, “Strategies for the Management of Localized Prostate Cancer: A Guide for Radiation Oncologists,” pp. 1–55, 2014, [Online]. Available: <http://www.iaea.org/Publications/index.html>
- [23] G. Quinaz Romana, I. Kislaya, M. R. Salvador, S. Cunha Gonçalves, B. Nunes, and C. Dias, “Multimorbidity in Portugal: Results from the first national health examination survey,” *Acta Med. Port.*, vol. 32, no. 1, pp. 30–37, 2019, doi: 10.20344/amp.11227.
- [24] K. D. Boudoulas, F. Triposkiadis, R. Gumina, D. Addison, C. Iliescu, and H. Boudoulas, “Cardiovascular Disease, Cancer, and Multimorbidity Interactions: Clinical Implications,” *Cardiol.*, vol. 147, no. 2, pp. 196–206, 2022, doi: 10.1159/000521680.
- [25] C. Harrison, H. Britt, G. Miller, and J. Henderson, “Examining different measures of multimorbidity, using a large prospective cross-sectional study in Australian general practice,” *BMJ Open*, vol. 4, no. 7, pp. 1–9, 2014, doi: 10.1136/bmjopen-2013-004694.
- [26] D. N. Blane and M. Lewandowska, “Living with cancer and multimorbidity: The role of primary care,” *Curr. Opin. Support. Palliat. Care*, vol. 13, no. 3, pp. 213–219, 2019, doi: 10.1097/SPC.0000000000000454.
- [27] D. Rodrigues, “Multimorbilidade como desafio à decisão clínica,” *AIMGF Mag.*, pp. 23–25, 2019.
- [28] F. Prazeres and L. Santiago, “Prevalence of multimorbidity in the adult population attending primary care in Portugal: A cross-sectional study,” *BMJ Open*, vol. 5, no. 9, 2015, doi: 10.1136/bmjopen-2015-009287.
- [29] R. Ofori-Asenso, K. L. Chin, A. J. Curtis, E. Zomer, S. Zoungas, and D. Liew, “Recent Patterns of Multimorbidity Among Older Adults in High-Income Countries,” *Popul. Health Manag.*, vol. 22, no. 2, pp. 127–137, 2019, doi: 10.1089/pop.2018.0069.
- [30] P. Broeiro, “Multimorbilidade e comorbilidade: duas perspectivas da mesma realidade,” *Rev. Port. Clínica Geral*, vol. 31, no. 3, pp. 150–160, 2015, doi: 10.32385/rpmgf.v31i3.11520.
- [31] R. M. Fiano *et al.*, “Associations of multimorbidity and patient-reported experiences of care with conservative management among elderly patients with localized prostate cancer,” *Cancer Med.*, vol. 9, no. 16, pp. 6051–6061, 2020, doi: 10.1002/cam4.3274.
- [32] T. Haregu, B. Oldenburg, G. Setswe, and J. Elliott, “Perspectives, Constructs And Methods In The Measurement Of Multimorbidity And Comorbidity: A Critical Review,” *Internet J. Epidemiol.*, vol. 10, no. 2, pp. 1–9, 2012, doi: 10.5580/2cc3.
- [33] J. Suls *et al.*, “Measuring Multimorbidity: Selecting the Right Instrument for the Purpose and the Data Source,” *Med Care*, vol. 59, no. 8, pp. 743–756, 2021, doi: 10.1097/MLR.0000000000001566.Measuring.
- [34] L. E. Griffith *et al.*, “Key factors to consider when measuring multimorbidity,” *J. Comorbidity*, vol. 8, no. 1, 2018, doi: 10.1177/2235042x18795306.
- [35] I. S. S. Ho *et al.*, “Examining variation in the measurement of multimorbidity in research: a systematic review of 566 studies,” *Lancet Public Heal.*, vol. 6, no. 8, 2021, doi: 10.1016/S2468-2667(21)00107-9.
- [36] T. T. Makovski, S. Schmitz, M. P. Zeegers, S. Stranges, and M. van den Akker, “Multimorbidity and quality of life: Systematic literature review and meta-analysis,” *Ageing Res. Rev.*, vol. 53, 2019, doi: 10.1016/j.arr.2019.04.005.

- [37] L. E. Stirling, L. González-Saavedra, D. S. Mullin, C. W. Ritchie, G. Muniz-Terrera, and T. C. Russ, "Measuring multimorbidity beyond counting diseases: systematic review of community and population studies and guide to index choice," *BMJ*, 2020, doi: 10.1136/bmj.m127.
- [38] E. S. Lee *et al.*, "Systematic review on the instruments used for measuring the association of the level of multimorbidity and clinically important outcomes," *BMJ Open*, vol. 11, no. 5, pp. 1–21, 2021, doi: 10.1136/bmjopen-2020-041219.
- [39] P. Broeiro-Gonçalves, P. Nogueira, and P. Aguiar, "Multimorbidity and disease severity measured by the charlson index in portuguese hospitalised patients during the year 2015: A cross-sectional study," *Acta Med. Port.*, vol. 32, no. 1, pp. 38–46, 2019, doi: 10.20344/amp.9728.
- [40] A. L. Huntley, R. Johnson, S. Purdy, J. M. Valderas, and C. Salisbury, "Measures of multimorbidity and morbidity burden for use in primary care and community settings: A systematic review and guide," *Ann. Fam. Med.*, vol. 10, no. 2, pp. 134–141, 2012, doi: 10.1370/afm.1363.
- [41] M. Yurkovich, J. A. Avina-Zubieta, J. Thomas, M. Gorenchtein, and D. Lacaille, "A systematic review identifies valid comorbidity indices derived from administrative health data," *J. Clin. Epidemiol.*, vol. 68, no. 1, pp. 3–14, 2015, doi: 10.1016/j.jclinepi.2014.09.010.
- [42] M. Cai *et al.*, "Comparing the performance of charlson and elixhauser comorbidity indices to predict in-hospital mortality among a Chinese population," *Clin. Epidemiol.*, vol. 12, pp. 307–316, 2020, doi: 10.2147/CLEP.S241610.
- [43] L. Rodrigues, A. Maio, F. Freitas, J. M. Bastos, and E. Ramalheira, "Charlson Comorbidity Index Score and the odds of death in COVID-19 patients," *J. Stat. Heal. Decis.*, vol. 4, no. 1, pp. 23–24, 2022.
- [44] A. Elixhauser, C. Steiner, D. R. Harris, and R. M. Coffey, "Comorbidity Measures for Use with Administrative Data," *Med. Care*, vol. 36, no. 1, pp. 8–27, 1998, doi: 10.1097/00005650-199801000-00004.
- [45] M. Y. Wei, D. Ratz, and K. J. Mukamal, "Multimorbidity in Medicare beneficiaries: Performance of an ICD- coded multimorbidity-weighted index," *J Am Geriatr Soc*, vol. 68, no. 5, pp. 999–1006, 2020, doi: 10.1111/jgs.16310.Multimorbidity.
- [46] N. Sharma, R. Schwendimann, O. Endrich, D. Ausserhofer, and M. Simon, "Comparing Charlson and Elixhauser comorbidity indices with different weightings to predict in-hospital mortality: an analysis of national inpatient data," *BMC Health Serv. Res.*, vol. 21, no. 1, pp. 1–18, 2021, doi: 10.1186/s12913-020-05999-5.
- [47] N. Jothi, N. A. Rashid, and W. Husain, "Data Mining in Healthcare - A Review," *Procedia Comput. Sci.*, vol. 72, pp. 306–313, 2015, doi: 10.1016/j.procs.2015.12.145.
- [48] S. V. G. Subrahmanya *et al.*, "The role of data science in healthcare advancements: applications, benefits, and future prospects," *Ir. J. Med. Sci.*, vol. 191, no. 4, pp. 1473–1483, 2022, doi: 10.1007/s11845-021-02730-z.
- [49] I. Kavakiotis, O. Tsave, A. Salifoglou, N. Maglaveras, I. Vlahavas, and I. Chouvarda, "Machine Learning and Data Mining Methods in Diabetes Research," *Comput. Struct. Biotechnol. J.*, vol. 15, pp. 104–116, 2017, doi: 10.1016/j.csbj.2016.12.005.
- [50] U. Fayyad, G. Piatetsky-Shapiro, and P. Smyth, "From Data Mining to Knowledge Discovery in Databases," *AI Mag.*, vol. 17, no. 3, 1996, doi: 10.1007/978-3-319-18032-8_50.
- [51] A. Zia, M. Aziz, I. Popa, S. A. Khan, A. F. Hamedani, and A. R. Asif, "Artificial Intelligence-Based Medical Data Mining," *J. Pers. Med.*, vol. 12, no. 9, 2022, doi: 10.3390/jpm12091359.
- [52] W. Salah Alaloul and A. Hannan Qureshi, "Data Processing Using Artificial Neural Networks," in *Dynamic Data Assimilation - Beating the Uncertainties*, Intechopen, 2020. doi: 10.5772/intechopen.91935.
- [53] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, N. A. E. Mohamed, and H. Arshad,

- “State-of-the-art in artificial neural network applications: A survey,” *Heliyon*, vol. 4, no. 11, 2018, doi: 10.1016/j.heliyon.2018.e00938.
- [54] T. Taipalus, V. Isomöttönen, H. Erkkilä, and S. Äyrämö, “Data Analytics in Healthcare: A Tertiary Study,” *SN Comput. Sci.*, vol. 4, 2023, doi: 10.1007/s42979-022-01507-0.
- [55] C. M. Eckhardt *et al.*, “Unsupervised machine learning methods and emerging applications in healthcare,” *Knee Surgery, Sport. Traumatol. Arthrosc.*, vol. 31, no. 2, pp. 376–381, 2023, doi: 10.1007/s00167-022-07233-7.
- [56] P. Festa, “A biased random-key genetic algorithm for data clustering,” *Math. Biosci.*, vol. 245, no. 1, pp. 76–85, 2013, doi: 10.1016/j.mbs.2013.07.011.
- [57] A. M. Ikotun, A. E. Ezugwu, L. Abualigah, B. Abuhaija, and J. Heming, “K-means clustering algorithms: A comprehensive review, variants analysis, and advances in the era of big data,” *Inf. Sci. (Nij.)*, vol. 30, 2023, doi: 10.1016/j.ins.2022.11.139.
- [58] B. Ben Ali and Y. Massmoudi, “K-means clustering based on gower similarity coefficient: A comparative study,” in *5th International Conference on Modeling, Simulation and Applied Optimization, ICMSAO*, 2013. doi: 10.1109/ICMSAO.2013.6552669.
- [59] Merriam-Webster Online Dictionary, “Cluster analysis,” *Merriam-Webster Online Dict.*, vol. 366, no. 8317, p. 4, 2003, doi: 10.4324/9781003196617-7.
- [60] A. Nagpal, A. Jatain, and D. Gaur, “Review based on data clustering algorithms,” *IEEE Conf. Inf. Commun. Technol.*, 2013, doi: 10.1109/CICT.2013.6558109.
- [61] H. Hu, J. Liu, X. Zhang, and M. Fang, “An Effective and Adaptable K-means Algorithm for Big Data Cluster Analysis,” *Pattern Recognit.*, vol. 139, 2023, doi: 10.1016/j.patcog.2023.109404.
- [62] Y. Zhou, H. Wu, Q. Luo, and M. Abdel-Baset, “Automatic data clustering using nature-inspired symbiotic organism search algorithm,” *Knowledge-Based Syst.*, vol. 163, pp. 546–557, 2019, doi: 10.1016/j.knosys.2018.09.013.
- [63] A. P. Koné, D. Scharf, and A. Tan, “Multimorbidity and Complexity Among Patients with Cancer in Ontario: A Retrospective Cohort Study Exploring the Clustering of 17 Chronic Conditions with Cancer,” *Cancer Control*, vol. 30, pp. 1–16, 2023, doi: 10.1177/10732748221150393.
- [64] M. Ruiz, A. Bottle, S. Long, and P. Aylin, “Multi-morbidity in hospitalised older patients: Who are the complex elderly?,” *PLoS One*, vol. 10, no. 12, pp. 1–11, 2015, doi: 10.1371/journal.pone.0145372.
- [65] N. R. Fillmore *et al.*, “Defining Multimorbidity and Its Impact in Older United States Veterans Newly Treated for Multiple Myeloma,” *J. Natl. Cancer Inst.*, vol. 113, no. 8, pp. 1084–1093, 2021, doi: 10.1093/jnci/djab007.
- [66] M. Mahmoud *et al.*, “Patterns of Comorbidity and In-Hospital Mortality in Older Patients With COVID-19 Infection,” *Front. Med.*, vol. 8, no. September, 2021, doi: 10.3389/fmed.2021.726837.
- [67] M. Bucholtz *et al.*, “Identifying pre-existing conditions and multimorbidity patterns associated with in-hospital mortality in patients with COVID-19,” *Sci. Rep.*, vol. 12, 2022, doi: 10.1038/s41598-022-20176-w.
- [68] L. Robertson, R. Vieira, J. Butler, M. Johnston, S. Sawhney, and C. Black, “Identifying multimorbidity clusters in an unselected population of hospitalised patients,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–10, 2022, doi: 10.1038/s41598-022-08690-3.
- [69] Y. Zhang *et al.*, “Associations Among Multimorbid Conditions in Hospitalized Middle-aged and Older Adults in China: Statistical Analysis of Medical Records,” *JMIR Public Heal. Surveill.*, vol. 8, no. 11, 2022, doi: 10.2196/38182.
- [70] R. L. Harrington, D. M. Qato, J. W. Antoon, R. N. Caskey, G. T. Schumock, and T. A. Lee,

- “Impact of multimorbidity subgroups on the health care use of early pediatric cancer survivors,” *Cancer*, vol. 126, no. 3, pp. 649–658, 2020, doi: 10.1002/cncr.32201.
- [71] J. Carmona-Pérez *et al.*, “Multimorbidity clusters in patients with chronic obstructive airway diseases in the EpiChron Cohort,” *Sci. Rep.*, vol. 11, no. 1, 2021, doi: 10.1038/s41598-021-83964-w.
- [72] M. Matesanz-Fernández, T. Seoane-Pillado, I. Iñiguez-Vázquez, R. Suárez-Gil, S. Pérttega-Díaz, and E. Casariego-Vales, “Description of multimorbidity clusters of admitted patients in medical departments of a general hospital,” *Postgrad. Med. J.*, vol. 98, no. 1158, pp. 294–299, 2022, doi: 10.1136/postgradmedj-2020-139361.
- [73] F. T. T. Lai *et al.*, “Comparing Multimorbidity Patterns Among Discharged Middle-Aged and Older Inpatients Between Hong Kong and Zurich: A Hierarchical Agglomerative Clustering Analysis of Routine Hospital Records,” *Front. Med.*, vol. 8, 2021, doi: 10.3389/fmed.2021.651925.
- [74] G. Cezard, F. Sullivan, and K. Keenan, “Understanding multimorbidity trajectories in Scotland using sequence analysis,” *Sci. Rep.*, vol. 12, no. 1, pp. 1–15, 2022, doi: 10.1038/s41598-022-20546-4.
- [75] P. McLoone *et al.*, “Classification of long-term condition patterns in rheumatoid arthritis and associations with adverse health events: a UK Biobank cohort study,” *J. Multimorbidity Comorbidity*, vol. 13, p. 263355652211486, 2023, doi: 10.1177/26335565221148616.
- [76] V. Kuan *et al.*, “Identifying and visualising multimorbidity and comorbidity patterns in patients in the English National Health Service: a population-based study,” *Lancet Digit. Heal.*, vol. 5, no. 1, pp. e16–e27, 2023, doi: 10.1016/S2589-7500(22)00187-X.
- [77] Agency for Healthcare Research and Quality, “Clinical Classifications Software (CCS) for ICD-9-CM,” *Healthcare Cost and Utilization Project (HCUP)*, 2017. <https://hcup-us.ahrq.gov/toolssoftware/ccs/ccs.jsp> (accessed Jul. 10, 2023).
- [78] Agency for Healthcare Research and Quality, “Clinical Classifications Software Refined (CCSR),” *Healthcare Cost and Utilization Project (HCUP)*, 2022. https://hcup-us.ahrq.gov/toolssoftware/ccsr/ccs_refined.jsp (accessed Jul. 10, 2023).
- [79] C. Van Walraven, P. C. Austin, A. Jennings, H. Quan, and A. J. Forster, “A modification of the elixhauser comorbidity measures into a point system for hospital death using administrative data,” *Med. Care*, vol. 47, no. 6, pp. 626–633, 2009, doi: 10.1097/MLR.0b013e31819432e5.
- [80] H. Quan, V. Sundararajan, P. Halfon, and A. Fong, “Coding algorithms for defining comorbidities in,” *Med. Care*, vol. 43, no. 11, 2005, [Online]. Available: <http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Coding+Algorithms+for+Defining+Comorbidities+in#1%5Cnhttp://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Coding+algorithms+for+defining+comorbidities+in%231>
- [81] A. E. Ezugwu *et al.*, “A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects,” *Eng. Appl. Artif. Intell.*, vol. 110, 2022, doi: 10.1016/j.engappai.2022.104743.
- [82] Q. Huang, R. Gao, and H. Akhavan, “An ensemble hierarchical clustering algorithm based on merits at cluster and partition levels,” *Pattern Recognit.*, vol. 136, 2023, doi: 10.1016/j.patcog.2022.109255.
- [83] S. Ben Salem, S. Naouali, and Z. Chtourou, “A rough set based algorithm for updating the modes in categorical clustering,” *Int. J. Mach. Learn. Cybern.*, vol. 12, no. 7, pp. 2069–2090, 2021, doi: 10.1007/s13042-021-01293-w.
- [84] S. Deepti, S. Lokesh, S. Sisodia, and Khushboo Saxena., “Clustering Techniques : A Brief Survey of Different Clustering Algorithms,” *Int. J. Latest Trends Eng. Technol. Clust.*, vol. 1, no. 3, pp. 82–87, 2015, [Online]. Available: <https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.1037.1127&rep=rep1&type=pdf>

- [85] K. Igor and K. Matjaz, "Cluster Analysis," in *Machine learning and data mining: introduction to principles and algorithms*, Horwood Publishing, 2008. doi: 10.5860/choice.45-3834.
- [86] A. Saxena *et al.*, "A review of clustering techniques and developments," *Neurocomputing*, vol. 267, pp. 664–681, 2017, doi: 10.1016/j.neucom.2017.06.053.
- [87] S. S. Khan and A. Ahmad, "Cluster center initialization algorithm for K-modes clustering," *Expert Syst. Appl.*, vol. 40, no. 18, pp. 7444–7456, 2013, doi: 10.1016/j.eswa.2013.07.002.
- [88] H. Zhou, Y. Zhang, and Y. Liu, "A Global-Relationship Dissimilarity Measure for the k - Modes Clustering Algorithm," *Comput. Intell. Neurosci.*, 2017, doi: 10.1155/2017/3691316.
- [89] L. Bai and J. Liang, "The k-modes type clustering plus between-cluster information for categorical data," *Neurocomputing*, vol. 133, pp. 111–121, 2014, doi: 10.1016/j.neucom.2013.11.024.
- [90] S. Renjith, A. Sreekumar, and M. Jathavedan, "Evaluation of partitioning clustering algorithms for processing social media data in tourism domain," *2018 IEEE Recent Adv. Intell. Comput. Syst. RAICS 2018*, no. 1, pp. 127–131, 2019, doi: 10.1109/RAICS.2018.8635080.
- [91] A. S. Shirkorshidi, S. Aghabozorgi, and T. Ying Wah, "A Comparison study on similarity and dissimilarity measures in clustering continuous data," *PLoS One*, vol. 10, no. 12, pp. 1–20, 2015, doi: 10.1371/journal.pone.0144059.
- [92] F. Cao, J. Liang, D. Li, L. Bai, and C. Dang, "A dissimilarity measure for the k-Modes clustering algorithm," *Knowledge-Based Syst.*, vol. 26, pp. 120–127, 2012, doi: 10.1016/j.knosys.2011.07.011.
- [93] A. Kassambara, "Multivariate Analysis I: Practical Guide To Cluster Analysis in R. Unsupervised Machine Learning," *STHDA*, p. 188, 2017.
- [94] A. Jansana *et al.*, "Multimorbidity clusters among long-term breast cancer survivors in Spain: Results of the SURBCAN study," *Int. J. Cancer*, vol. 149, no. 10, pp. 1755–1767, 2021, doi: 10.1002/ijc.33736.
- [95] C. E. Coombes, X. Liu, Z. B. Abrams, K. R. Coombes, and G. Brock, "Simulation-derived best practices for clustering clinical data," *J. Biomed. Inform.*, vol. 118, 2021, doi: 10.1016/j.jbi.2021.103788.
- [96] P. J. Rousseeuw, "Silhouettes: A graphical aid to the interpretation and validation of cluster analysis," *J. Comput. Appl. Math.*, vol. 20, 1987, doi: 10.1016/0377-0427(87)90125-7.
- [97] T. Ullmann, C. Hennig, and A. L. Boulesteix, "Validation of cluster analysis results on validation data: A systematic framework," *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, vol. 12, no. 3, pp. 1–19, 2022, doi: 10.1002/widm.1444.
- [98] S. Ray, "Introduction to Machine Learning and Different types of Machine Learning Algorithms," *Proc. Int. Conf. Mach. Learn. Big Data, Cloud Parallel Comput. Trends, Perspectives Prospect. Com.*, pp. 35–39, 2019.
- [99] N. Lopes and L. Cavique, "Segmentação de clientes e análise dos atributos mais relevantes dos clusters," *Rev. Ciências da Comput.*, pp. 29–48, 2022.
- [100] D. Rizzuto, R. J. F. Melis, S. Angleman, C. Qiu, and A. Marengoni, "Effect of Chronic Diseases and Multimorbidity on Survival and Functioning in Elderly Adults," *J. Am. Geriatr. Soc.*, vol. 65, no. 5, pp. 1056–1060, 2017, doi: 10.1111/jgs.14868.
- [101] L. Bayer-Oglesby, A. Zumbrunn, and N. Bachmann, "Social inequalities, length of hospital stay for chronic conditions and the mediating role of comorbidity and discharge destination: A multilevel analysis of hospital administrative data linked to the population census in Switzerland," *PLoS One*, vol. 17, 2022, doi: 10.1371/journal.pone.0272265.
- [102] R. Benzo *et al.*, "Comorbidity burden and health-related quality of life in men with advanced prostate cancer," *Res. Sq.*, 2023, [Online]. Available:

- <https://pubmed.ncbi.nlm.nih.gov/36824747/%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC9949262>
- [103] M. Adjakly *et al.*, “Prostate cancer: The main risk and protective factors-Epigenetic modifications,” *Ann. Endocrinol. (Paris)*, vol. 76, no. 1, pp. 25–41, 2015, doi: 10.1016/j.ando.2014.09.001.
- [104] S. D. Brookman-May *et al.*, “Latest Evidence on the Impact of Smoking, Sports, and Sexual Activity as Modifiable Lifestyle Risk Factors for Prostate Cancer Incidence, Recurrence, and Progression: A Systematic Review of the Literature by the European Association of Urology Section of ,” *Eur. Urol. Focus*, vol. 5, no. 5, pp. 756–787, 2019, doi: 10.1016/j.euf.2018.02.007.
- [105] Registo Oncológico Nacional, *De todos os tumores na população residente em Portugal*. 2018.
- [106] J. Souza, J. V. Santos, V. B. Canedo, A. Betanzos, D. Alves, and A. Freitas, “Importance of coding co-morbidities for APR-DRG assignment: Focus on cardiovascular and respiratory diseases,” *Heal. Inf. Manag. J.*, vol. 49, no. 1, pp. 47–57, 2020, doi: 10.1177/1833358319840575.

APÊNDICE A

Tabela A 1 -Revisão da literatura

Autores	Ano	Objetivo	Tipo de dados	Nº indivíduos	Idade	Método	Número de condições crónicas consideradas	Codificação	Conclusões
Milagros Ruiz et al. [64]	2015	Identificar empiricamente o paciente idoso complexo com base no grau de multimorbilidade	Dados administrativos, de 2012/2013	2788900	Entre os 65 e os 95 anos	Análise de correspondência múltipla	20 definidas nos índices de Charlson e Elixhauser	ICD-10	3 grupos distintos: - Grupo 1 (multimorbilidade ≤ 2), associado a cancro e/ou metástases; - Grupo 2 (multimorbilidade de 3, 4 ou 5), associado a doença pulmonar obstrutiva crónica, doença pulmonar, reumatismo e osteoporose; - Grupo 3 (multimorbilidade 6) associado a insuficiência cardíaca, acidente vascular cerebral, diabetes, hipertensão e enfarte do miocárdio; Propuseram que os grupos 2 e 3 fossem considerados de idosos complexos.

Rachel L. Harrington et al. [70]	2020	Identificar e caracterizar subgrupos de multimorbilidade em sobreviventes de cancro pediátrico	Reivindicações de seguros, de 2009 a 2014, nos EUA	3687	Entre os 2 e os 18 anos	Análise de classes latentes	<i>Agency for Healthcare Research and Quality Clinical Classification Software (CCS) and Chronic Condition Indicator (CCI) algorithms</i>	ICD-9-CM	5 grupos distintos: 1- Morbilidade pediátrica geral 2 - Sistema nervoso central (SNC) 3 - Condições de saúde mental 4 - Sistema endócrino 5 - SNC e endócrino; Demonstraram que o grupo 4 apresenta uma maior probabilidade de hospitalização e o grupo 2 uma maior eventualidade de recorrer às urgências.
Jonás Carmona-Pérez et al. [71]	2021	Caracterizar a existência de clusters de multimorbilidade em pacientes com doenças respiratórias crónicas (DPOC, asma, rinite e apneia obstrutiva do sono) e o seu impacto na mortalidade.	Registos de saúde eletrónicos de pacientes em Espanha	96,059	Entre os 15 e os 65 anos	Análise de Clusters	114 <i>expanded diagnostic</i> clusters (EDCs) definidos por Salisbury et al [0]	ICD-9-CM	23 clusters; Identificaram associações de doenças cardiometabólicas com problemas de saúde mental, doenças dermatológicas com distúrbios músculo-esqueléticos, e doenças neurológicas com transtornos por uso de substâncias e neoplasias, entre outras; Doenças cardiovascular e

									doenças metabólicas foram das condições mais comuns em vários clusters;
									O cluster com maior probabilidade de mortalidade foi identificado em homens entre os 45 e 64 anos, que incluía associações entre distúrbios de uso de substâncias, condições neurológicas e cancro.
Nathanael Fillmore et al. [65]	2021	Definir padrões de multimorbilidade e o seu impacto em idosos veteranos, dos Estados Unidos, com mieloma múltiplo.	Registos clínicos e registos de saúde eletrónicos, de 2004 a 2017	5076	>65 anos	Análise de classes latentes	66 condições crónicas da <i>CMS Chronic Conditions Data Warehouse</i>	ICD-9 e ICD-10	5 padrões de multimorbilidade: 1 - Pacientes com baixa comorbilidade (hipertensão, artrite e cataratas) 2 - Doenças cardiovasculares e metabólicas 3 - Doenças psiquiátricas e uso de substâncias 4 - Doenças respiratórias crónicas e tabagismo 5 - Disfunção multissistémica e deficiências sensoriais; Todos os padrões apresentavam uma prevalência elevada de anemia;

									Pacientes com mieloma múltiplo e patologias presentes nos clusters 2 a 5 demonstraram elevadas taxas de mortalidade e recurso a serviços de urgência.
Mona Mahmoud et al. [66]	2021	Identificar as condições crónicas e os clusters de comorbilidade associados à mortalidade intra-hospitalar em idosos com COVID-19	Registos médicos de pacientes admitidos no hospital Martino, em Itália, entre janeiro e abril de 2020	219	≥ 70 anos	<i>Bernoulli Latent Block Model</i>	-	-	3 clusters: 1 – Metabólico-renal-cancro 2 - Neurocognitivo 3 – Não especificado; Os clusters Metabólico-renal-cancro e Neurocognitivo revelaram fortes associações com a mortalidade hospitalar por COVID-19.
María Matesanz - Fernández et al. [72]	2021	Identificar padrões em clusters de doenças em pacientes de um hospital geral e descrever as características e evolução de cada grupo	Registos clínicos de admissão ao internamento e registos de enfermagem, de 2000 a 2015	170,978 internações correspondentes a 74,220 pacientes	Média 76 (15–108) anos	Análise de correspondência múltipla	32 condições crónicas utilizando uma versão modificada do <i>German Multicare</i> [20]	ICD-9-CM	5 clusters; Esses clusters são constituídos por uma série de doenças de alta incidência, relacionadas com a idade e o sexo, apresentando as suas próprias características clínicas e evolução ao longo do tempo.

Francisco T. T. Lai et al. [73]	2021	Obter e comparar perfis de multimorbilidade de duas populações diferentes (Hong Kong e Zurique)	Registos clínicos de alta hospitalar, de todos os hospitais públicos de Hong Kong (2010-2013) e do Hospital Universitário de Zurique (2009-2017)	20,000 (10,000 pacientes selecionados aleatoriamente de cada local)	≥ 45 anos	Clustering hierárquico - aglomerativo	30 patologias baseadas nos algoritmos de Tonelli [21]	ICD-9 e ICD-10	9 clusters; Apenas 2 dos 9 clusters (cluster de AVC e cluster doença renal crónica) eram comuns nos dois locais; Identificaram padrões de multimorbilidade específicos de cada local; Em ambos os locais, especialmente em HK, as díades de doenças enquadravam-se principalmente nas mesmas especialidades médicas.
Magda Bucholtz et al. [67]	2022	Caracterizar padrões de multimorbilidade em indivíduos hospitalizados com COVID-19	Registos de admissões hospitalares	6036	Média 72 (57–82) anos	<i>Clustering – k-modes</i>	17 definidas pela medida de comorbilidade de Elixhauser	ICD-10	6 clusters: 1 - Doenças cardiovasculares e renais 2 - Distúrbios mentais e neurológicos e doenças metabólicas e endócrinas 3 - Doenças respiratórias e cardiovasculares 4 – Cancro, doenças genitourinárias, metabólicas e endócrinas 5 - Doenças metabólicas e endócrinas e doenças cardiovasculares 6 - Distúrbios mentais e neurológicos e doenças cardiovasculares

Os clusters apresentaram diferenças significativas na taxa de mortalidade, variando de 15,3% a 29,4%.									
O risco de mortalidade hospitalar por COVID-19 é maior em pessoas com várias condições crónicas subjacentes e varia consoante o número e a combinação das doenças apresentadas.									
G. Cezard et al. [74]	2022	Compreender a sequência de doenças crónicas que conduzem à multimorbilidade, através de um método longitudinal	Censos da Escócia de 2001 associados a registos clínicos de admissão ao internamento; Os participantes foram seguidos durante um período de 10 anos (2001-2011)	6300	Entre os 40 e os 74 anos	Clustering hierárquico	3 condições crónicas (diabetes, DCV e cancro), mais prevalentes na população e possíveis de identificar com precisão em registos de hospitalização	ICD-10	7 clusters: 1 - Transição rápida para multimorbilidade, numa fase tardia; 2 - DCV com transição lenta para multimorbilidade; 3 - Cancro com transição lenta para multimorbilidade; 4 - Diabetes com transição lenta para multimorbilidade; 5 - Transição rápida para diabetes e DCV; 6- Transição rápida para multimorbilidade e morte; 7- Transição rápida para cancro e DCV;

									Os grupos que transitaram rapidamente para multimorbilidade e morte eram constituídos por pacientes mais vulneráveis, mais velhos, com nível de educação inferior e mais propensos a viver em áreas mais carenciadas.
Lynn Robertso n et al. [68]	2022	Identificar e caracterizar clusters de multimorbilidade em pacientes hospitalizados	Dados administrativos de hospitalizações, na Escócia, em 2014	11389	Entre os 51 e os 79 anos	<i>Partitioning around Medoids</i> (PAM)	30 patologias baseadas nos algoritmos de Tonelli	ICD-10	10 clusters: hipertensão; asma; abuso de substâncias; doença renal crónica e diabetes; doença renal crónica; dor crónica; cancro; insuficiência cardíaca; diabetes; hipotireoidismo; As mulheres apresentam-se em maior percentagem em clusters de dor e hipotireoidismo.
Yan Zhang et al. [69]	2022	Analisar e comparar associações entre múltiplas doenças crónicas, em adultos e idosos chineses, consoante a idade e o sexo.	Registos clínicos de admissão ao internamento da Plataforma Nacional de Informações de Saúde de Shenzhen, de 2017 a 2018.	306,264	≥50 anos	<i>Combinação entre Association Rule Mining (ARM) e Árvores de decisão.</i>	40 condições crónicas seleccionadas de acordo com as doenças mais mencionadas, em estudos prévios, como tendo um impacto significativo em indivíduos chineses	ICD-10	Entre os pacientes com ≥ 65 anos as doenças vasculares periféricas encontram-se fortemente associadas a doenças cardíacas, cerebrovasculares e dislipidemia;

									Associações mais acentuadas foram observadas entre catarata senil e glaucoma, tanto em homens como mulheres;
									Verificaram associações únicas consoante o género, como: - Osteoporose e cancro em homens - Anemia e doença renal crónica em mulheres idosas.
Anna Péfoyo Koné et al. [63]	2023	Identificar combinações de doenças entre pacientes com cancro e outras condições e avaliar o respetivo impacto	Dados de base de dados administrativas de pacientes com diagnóstico de cancro entre 2003-2013 e <i>follow-up</i> até 2018	549,248	≥18 anos	Clustering – k-means	17 condições crónicas selecionadas pelos autores	ICD-9 e ICD-10	Verificaram que as combinações de doenças variaram de acordo com o tipo de cancro e a idade, no entanto eram semelhantes entre os géneros; O maior cluster observado foi o cancro e a ansiedade, com 25% dos indivíduos a apresentar igualmente artrose; O cancro também tende a ocorrer juntamente com hipertensão ou artrose.

Valerie Kuan et al. [76]	2023	Estimar a frequência de combinações de doenças e identificar associações de doenças para todas as idades numa população multiétnica	Registos de saúde eletrónico de indivíduos incluídos no <i>Clinical Practice Research Datalink</i> , em Inglaterra, de 1 de abril de 2010 a 31 de março de 2015	3,872,451	>1 ano	<i>Network analysis</i>	308 condições de saúde física e mental, com base no uso intensivo de recursos médicos, prevalência e importância clínica	Algoritmo <i>phenotyping</i>	A depressão comumente ocorreu agrupada à ansiedade, enxaqueca, obesidade, condições atópicas, surdez, distúrbios dos tecidos moles e distúrbios gastrointestinais em todos os subgrupos; Insuficiência cardíaca ocorreu, frequentemente, relacionada com hipertensão, artrose, angina estável, enfarte agudo do miocárdio, doença renal crónica, diabetes tipo 2 e DPOC; Fraturas da coluna vertebral revelaram-se fortemente associadas a malignidade em indivíduos negros, mas com osteoporose em indivíduos brancos; A hipertensão demonstram-se associada a distúrbios renais em indivíduos com idade entre 20 e 29 anos, mas com dislipidemia, obesidade e diabetes tipo 2 em indivíduos com ≥ 40 anos;
--------------------------	------	---	---	-----------	--------	-------------------------	--	------------------------------	--

									O cancro da mama foi associado a diferentes comorbilidades em indivíduos de diferentes grupos étnicos.
Philip McLoone et al. [75]	2023	Classificar indivíduos com artrite reumatoide e ≥ 2 condições crónicas e descrever a associação entre as diferentes classes de condições, número e resultados de saúde adversos	Registos clínicos de pacientes com alta hospitalar e certificados de óbito, entre 2006 a 2010	2,566	Entre os 37 e os 72 anos	Análise de classes latentes	42 condições crónicas definidas na literatura publicada, anteriormente, sobre a multimorbilidade	ICD-10	5 grupos distintos: 1 - Distúrbios da tiroide; 2 - Dor crónica; 3 - Asma; 4 – Hipertensão; 5 – Cancro; Hipertensão e dor crónica denotaram altamente prevalentes em todas as classes; No grupo 5, a combinação mais comum foi cancro e hipertensão.

APÊNDICE B

Braquiterapia

Tabela B 1 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129	Sem Multimorbilidade n=4303
Arritmia, n(%)	0 (0)	93 (100)	13 (8.7)	20 (8.6)	126 (11)	-
Hipertensão, n(%)	517 (79)	57 (61)	90 (60)	180 (78)	844 (75)	-
Doença Pulmonar Obstrutiva Crónica, n(%)	75 (11)	4 (4.3)	14 (9.4)	20 (8.6)	113 (10)	-
Diabetes Mellitus, n(%)	0 (0)	0 (0)	31 (21)	232 (0)	263 (23)	-
Doença renal, n(%)	24 (3.7)	1 (1.1)	2 (1.3)	2 (0.9)	29 (2.6)	-
Úlcera péptica, n(%)	11 (1.7)	1 (1.1)	2 (1.3)	5 (2.2)	19 (1.7)	-
Tumor, n(%)	27 (4.1)	0 (0)	4 (2.7)	9 (3.9)	40 (3.5)	-
Obesidade, n(%)	57 (8.7)	13 (14)	25 (17)	41 (18)	136 (12)	-
Abuso de álcool, n(%)	0 (0)	0 (0)	149 (0)	0 (0)	149 (13)	-
Depressão, n(%)	42 (6.4)	2 (2.2)	9 (6)	6 (2.6)	59 (5.2)	-
Metástases, n(%)	36 (5.5)	8 (8.6)	7 (4.7)	10 (4.3)	61 (5.4)	136 (3.2)
Idade, mediana (IQR)	67 (62-71)	70 (64-73)	66 (61-70)	67 (63-71)	67 (62-71)	64 (59- 68)
Tempo de internamento, mediana (IQR)	2 (2-3)	2 (2-3)	2 (2-3)	2 (2-3)	2 (2-3)	5 (4, 7)
Peso Relativo de Internamento, média (s)	3.0824 ($\pm$ 1.1948)	3.5054 ($\pm$ 2.0303)	3.2215 ($\pm$ 1.3195)	3.2543 ($\pm$ 1.2792)	2.1754 ($\pm$ 0.9154)	5.35 ($\pm$ 1.56)
Morte, n(%)	11 (1.7)	2 (2.2)	2 (1.3)	4 (1.7)	19 (1.7)	6 (0.1)

Tabela B 2- Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
Programada, n(%)	622 (95)	87 (94)	143 (96)	222 (96)	1074 (95)

Urgente, n(%)	32 (4.9)	6 (6.5)	6 (4.0)	10 (4.3)	54 (4.8)
---------------	----------	---------	---------	----------	----------

Tabela B 3 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
Menor, n(%)	607 (93)	77 (83)	132 (89)	200 (86)	1016 (90)
Moderado, n(%)	31 (4.7)	9 (9.7)	14 (9.4)	25 (11)	79 (7.0)
Major, n(%)	17 (2.6)	7 (7.5)	3 (2.0)	7 (3.0)	34 (3.0)

Tabela B 4 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
Menor, n(%)	40 (6.1)	4 (4.3)	5 (3.4)	15 (6.5)	64 (5.7)
Moderado, n(%)	521 (80)	51 (55)	115 (77)	182 (78)	869 (77)
Major, n(%)	93 (14)	38 (41)	29 (19)	35 (15)	195 (17)
Extremo, n(%)	1 (0.2)	0 (0)	0 (0)	0 (0)	1 (<0.1)

Tabela B 5 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
Grupo A, n(%)	0 (0)	0 (0)	0 (0)	0 (0)	0 (0)
Grupo B, n(%)	0 (0)	0 (0)	1 (0.7)	0 (0)	1 (<0.1)
Grupo C, n(%)	11 (1.7)	1 (1.1)	3 (2.0)	3 (1.3)	18 (1.6)
Grupo D, n(%)	11 (1.7)	4 (4.3)	1 (0.7)	1 (0.4)	17 (1.5)
Grupo E, n(%)	57 (8.7)	10 (11)	4 (2.7)	23 (9.9)	94 (8.3)
Grupo F, n(%)	576 (88)	78 (84)	140 (94)	205 (88)	999 (88)

Tabela B 6 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
Norte, n(%)	238 (36)	42 (45)	31 (21)	81 (35)	392 (35)
Centro, n(%)	324 (49)	41 (44)	115 (77)	114 (49)	594 (53)
Lisboa, n(%)	89 (14)	10 (11)	2 (1.3)	36 (16)	137 (12)
Alentejo, n(%)	0 (0)	0 (0)	1 (0.7)	1 (0.4)	2 (0.2)
Algarve, n(%)	4 (0.6)	0 (0)	0 (0)	0 (0)	4 (0.4)

Tabela B 7 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser, para o conjunto de dados de braquiterapia.

Variáveis	Cluster 1 n=655	Cluster 2 n=93	Cluster 3 n=149	Cluster 4 n=232	Total n=1129
-4, n(%)	36 (5.5)	0 (0)	13 (8.7)	0 (0)	49 (4.3)
-3, n(%)	28 (4.3)	0 (0)	8 (5.4)	27 (12)	63 (5.6)
-1, n(%)	12 (1.8)	0 (0)	3 (2.0)	1 (0.4)	16 (1.4)
0, n(%)	408 (62)	0 (0)	67 (45)	7 (3.0)	482 (43)
3, n(%)	63 (9.6)	0 (0)	11 (7.4)	1 (0.4)	75 (6.6)
4, n(%)	20 (3.1)	2 (2.2)	2 (1.3)	10 (4.3)	34 (3.0)
5, n(%)	14 (2.1)	62 (67)	6 (4.0)	8 (3.4)	90 (8.0)
12, n(%)	6 (0.9)	1 (1.1)	0 (0)	0 (0)	7 (0.6)
média (s)	10.2550 ($\pm$ 4.4592)	15.0323 ($\pm$ 4.6285)	10.8725 ($\pm$ 5.2030)	10.9440 ($\pm$ 4.2305)	9.8663 ($\pm$ 4.6531)

APÊNDICE C

Prostatectomia Radical

Tabela C 1 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a prevalência de comorbilidades de Elixhauser selecionadas, presença de metástases, idade, tempo de internamento, tratamento, peso relativo de internamento e mortalidade, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Arritmia, n(%)	0 (0)	0 (0)	366 (0)	0 (0)	16 (7.4)	382 (7.7)
Hipertensão, n(%)	2377 (84)	688 (75)	242 (66)	445 (72)	0 (0)	3752 (76)
Doença Pulmonar Obstrutiva Crónica, n(%)	151 (5.3)	40 (4.3)	21 (5.7)	40 (6.5)	216 (0)	468 (9.4)
Diabetes Mellitus, n(%)	0 (0)	921 (100)	53 (14)	161 (26)	15 (6.9)	1150 (23)
Doença renal, n(%)	85 (3.0)	27 (2.9)	13 (3.6)	13 (2.1)	5 (2.3)	143 (2.9)
Úlcera péptica, n(%)	61 (2.1)	11 (1.2)	3 (0.8)	4 (0.6)	4 (1.9)	83 (1.7)
Tumor, n(%)	105 (3.7)	15 (1.6)	6 (1.6)	16 (2.6)	6 (2.8)	148 (3)
Obesidade, n(%)	0 (0)	0 (0)	45 (12)	620 (100)	25 (12)	690 (14)
Abuso de álcool, n(%)	141 (5.0)	29 (3.1)	13 (3.6)	38 (6.1)	14 (6.5)	235 (4.7)
Depressão, n(%)	169 (6.0)	20 (2.2)	11 (3.0)	22 (3.5)	7 (3.2)	229 (4.6)
Metástases, n(%)	78 (2.7)	29 (3.1)	9 (2.5)	17 (2.7)	6 (2.8)	139 (2.8)
Idade, mediana (IQR)	65 (61-69)	65 (62-69)	66 (62-70)	64 (60-68)	64 (60-68)	65 (61-69)
Tempo de internamento, mediana (IQR)	6 (4-7)	6 (4-7)	6 (4-7)	5 (4-7)	5 (4-7)	6 (4-7)
Peso Relativo de Internamento, média (s)	6.0750 ($\pm$ 0.6450)	6.0814 ($\pm$ 0.6796)	6.3825 ($\pm$ 1.3017)	6.0758 ($\pm$ 0.6862)	6.0741 ($\pm$ 0.7619)	4.0695 ($\pm$ 0.3854)
Morte, n(%)	1 (<0.1)	0 (0)	4 (1.1)	1 (0.2)	0 (0)	6 (0.1)

Tabela C 2 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a natureza ou modo de admissão hospitalar, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Programada, n(%)	2807 (99)	907 (98)	365 (100)	615 (99)	214 (99)	4908 (99)
Urgente, n(%)	32 (1.1)	14 (1.5)	1 (0.3)	5 (0.8)	2 (0.9)	54 (1.1)

Tabela C 3 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de severidade, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Menor, n(%)	2657 (94)	843 (92)	293 (80)	573 (92)	196 (91)	4562 (92)
Moderado, n(%)	164 (5.8)	71 (7.7)	60 (16)	43 (6.9)	18 (8,3)	356 (7.2)

Tabela C 4 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os quatro níveis de mortalidade, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Menor, n(%)	2323 (82)	738 (80)	163 (45)	96 (15)	89 (41)	3409 (69)
Moderado, n(%)	477 (17)	159 (17)	178 (49)	505 (81)	119 (55)	1438 (29)
Major, n(%)	34 (1.2)	23 (2.5)	22 (6.0)	19 (3.1)	8 (3.7)	106 (2.1)
Extremo, n(%)	5 (0.2)	1 (0.1)	3 (0.8)	0 (0)	0 (0)	9 (0.2)

Tabela C 5 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os grupos hospitalares, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Grupo A, n(%)	18 (0.6)	6 (0.7)	1 (0.3)	10 (1.6)	1 (0.5)	36 (0.7)
Grupo B, n(%)	81 (2.9)	32 (3.5)	14 (3.9)	23 (3.7)	7 (3.2)	157 (3.2)
Grupo C, n(%)	696 (25)	247 (27)	76 (21)	210 (34)	69 (32)	1298 (26)
Grupo D, n(%)	558 (20)	160 (17)	49 (14)	71 (12)	34 (16)	872 (18)
Grupo E, n(%)	1048 (37)	364 (40)	148 (41)	229 (37)	74 (34)	1863 (38)
Grupo F, n(%)	413 (15)	108 (12)	71 (20)	74 (12)	31 (14)	697 (14)
Outros	25	4	7	3	0	39

Tabela C 6 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo a região do hospital, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
Norte, n(%)	1227 (44)	371 (40)	146 (41)	228 (37)	100 (46)	2072 (42)
Centro, n(%)	670 (24)	228 (25)	95 (26)	168 (27)	41 (19)	1202 (24)
Lisboa, n(%)	824 (29)	287 (31)	114 (32)	211 (37)	67 (31)	1503 (31)
Alentejo, n(%)	51 (1.8)	20 (2.2)	4 (1.1)	8 (1.3)	7 (3.2)	90 (1.8)
Algarve, n(%)	42 (1.5)	11 (1.2)	1 (0.3)	2 (0.3)	1 (0.5)	57 (1.2)

Tabela C 7 - Clusters de multimorbilidade obtidos com o algoritmo PAM e descrições segundo os índices de Elixhauser, para o conjunto de dados de prostatectomia radical.

Variáveis	Cluster 1 n=2839	Cluster 2 n=921	Cluster 3 n=366	Cluster 4 n=620	Cluster 5 n=216	Total n=4962
-4, n(%)	1 (<0.1)	0 (0)	0 (0)	357 (58)	0 (0)	358 (7.2)
-3, n(%)	143 (5.0)	0 (0)	0 (0)	118 (19)	0 (0)	261 (5.3)
-1, n(%)	31 (1.1)	5 (0.5)	1 (0.3)	32 (5.2)	16 (7.4)	85 (1.7)
0, n(%)	2054 (72)	2 (0.2)	0 (0)	19 (3.1)	7 (3.2)	2082 (42)
3, n(%)	149 (5.2)	5 (0.5)	1 (0.3)	4 (0.6)	146 (68)	305 (6.1)
4, n(%)	83 (2.9)	42 (4.6)	14 (3.8)	2 (0.3)	8 (3.7)	149 (3.0)
5, n(%)	80 (2.8)	8 (0.9)	213 (58)	1 (0.2)	3 (1.4)	305 (6.1)
12, n(%)	59 (2.1)	8 (0.9)	14 (3.8)	4 (0.6)	0 (0)	85 (1.7)
média (s)	10.00951 (± 3.146905)	10.96091 (± 3.044702)	14.86612 (± 3.572513)	6.429032 (± 3.432214)	13.11574 (± 4.258019)	9.2308 (± 3.7933)