



UNIVERSIDADE D
COIMBRA

Daniela Filipa Neves Guia

MODELOS BASEADOS EM
INTELIGÊNCIA ARTIFICIAL PARA
DETECÇÃO DE OSTEOPOROSE:
PROJETO OSSEUS

Dissertação no âmbito do Mestrado em Engenharia Biomédica no
ramo de Instrumentação Biomédica orientada pelo Professor
Doutor César Alexandre Domingues Teixeira e Agnaldo Cruz e
apresentada ao departamento de Física da Faculdade de Ciências e
Tecnologia da Universidade de Coimbra

Setembro de 2023



FACULDADE DE
CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE DE
COIMBRA

Daniela Filipa Neves Guia

MODELOS BASEADOS EM
INTELIGÊNCIA ARTIFICIAL PARA
DETECÇÃO DE OSTEOPOROSE:
PROJETO OSSEUS

Faculdade de Ciências e Tecnologia

Departamento de Física

Dissertação apresentada à Universidade de Coimbra para cumprimento dos requisitos
necessários à obtenção do grau de Mestre em Instrumentação Biomédica

Orientador(s):

Prof. Dr. César Alexandre Domingues Teixeira (DEI-FCTUC)

Prof. Dr. Agnaldo Cruz (LAIS-UFRN)

Setembro 2023

Esta cópia da tese é fornecida na condição de que, quem a consulta, reconhece que os direitos de autor são pertença do autor da tese e que nenhuma citação ou informação obtida a partir dela pode ser publicada, sem referência apropriada.

This copy of this thesis has been supplied on condition that anyone who consults it is understood to recognize that is copyright rests with its author and that no quotation from the thesis and no information derived from it may be published without proper acknowledge.

AGRADECIMENTOS

Foram muitos os desafios que atravessei durante os 10 meses em que estive a realizar esta dissertação de mestrado. Felizmente, tive o apoio da família e dos amigos para concretizar este objetivo e finalizar este trabalho que tão importante foi para mim.

Todos os que estiveram perto de mim e me ajudaram a chegar onde estou, sabem o quanto me custou e agradeço-lhes toda a força e motivação que contribuíram para que conseguisse concluir o último projeto do curso.

Por isso, gostaria de agradecer a todos os que sempre me apoiaram desde o início, especialmente à minha família, aos meus pais, à minha irmã, marido e também aos meus grandes colegas e amigos Maria Melo, Margarida Realinho, Daniela Lima, Beatriz Bastos, Patrícia Pereira, Margarida Barreto, Helena Rebelo, Alexandra Alves, Bernardo Marques, Ana Fernandes, Alfonso Brandão, Rui Magalhães, Beatriz Melo, Joana Forte Teresa Nascimento, Ana Ferreira, Micaela Ponte, Sara Carvalho, Francisco Paiva e Sara Rodrigues.

Gostaria também de expressar a minha gratidão pelo apoio e disponibilidade ao professor César Teixeira que, mesmo com os constrangimentos pessoais que pudessem surgir, me deu acompanhamento ao longo do processo e tornou o meu trabalho mais significativo.

Gostaria de também agradecer ao professor Agnaldo que esteve disposto a envolver-me num dos seus projetos e partilhou os dados para o desenvolvimento de todo o trabalho.

*” Se todo em cada coisa.
Põe quanto és no mínimo que fazes.”
— Fernando Pessoa*

RESUMO

A osteoporose, uma das doenças ósseas mais prevalentes mundialmente, constitui um desafio global crescente. Crônica, sistêmica e silenciosa, esta doença é caracterizada por fragilidade óssea, tendo como principal consequência as fraturas. Apresenta impactos profundos na qualidade de vida e aumento da mortalidade, principalmente em idosos.

Apesar dos fatores de risco já conhecidos, o diagnóstico permanece complexo, dispendioso e muitas vezes inacessível, o que enfatiza a necessidade de soluções eficazes de triagem e prevenção.

Nesse contexto, este estudo visa explorar uma solução baseada nos dados do projeto *OSSEUS* com recurso a correlações entre fatores de risco e Inteligência Artificial. Em particular foram usados algoritmos de *Machine Learning* como Regressão Logística e *Support Vector Machine* para diagnóstico precoce da osteoporose.

O processo envolve treino e aprendizagem dos algoritmos, seguido da avaliação do desempenho com base em métricas como sensibilidade, especificidade e F-measure. Embora os resultados tenham demonstrado valores de sensibilidade e F-measure em linha com as expectativas, é crucial melhorar a especificidade para minimizar falsos positivos e aprimorar a precisão do modelo preditivo.

Concluindo, de acordo com os modelos desenvolvidos, esta análise oferece insights valiosos na identificação da osteoporose, proporcionando e impulsionando o desenvolvimento de soluções preventivas mais eficazes, reduzindo tanto o impacto desta condição na qualidade de vida dos doentes como o impacto económico associado.

Keywords: osteoporose, prevenção, fatores de risco e diagnóstico, algoritmos de inteligência artificial, modelos preditivos

ABSTRACT

Osteoporosis, one of the most prevalent bone diseases worldwide, is an increasingly global problem. Chronic, systemic, and silent, this disease is characterized by bone fragility, with fractures as its primary consequence. It exerts profound impacts on quality of life and increased mortality, particularly among the elderly.

Despite the known of risk factors, diagnosis remains complex, costly, and often inaccessible, underscoring the need for effective screening and preventive solutions.

In this context, this study aims to explore a solution based on data from the OSSEUS project, using correlations between risk factors and Artificial Intelligence, employing Machine Learning algorithms such as Logistic Regression and Support Vector Machine for early diagnosis of osteoporosis.

The process involves algorithm training and learning, followed by performance assessment based on metrics like sensitivity, specificity, and F-measure. While the results have demonstrated sensitivity and F-measure values in line with expectations, improving specificity is crucial to minimize false positives and enhance predictive model accuracy.

In conclusion, according to the developed models, this analysis provides valuable insights into osteoporosis identification, thereby fostering the development of more effective preventive solutions. This not only reduces the impact of this condition on patients' quality of life but also mitigates the associated economic burden.

Keywords: osteoporosis, prevention, risk assessment and diagnosis, artificial intelligence algorithms, predictive models.

ÍNDICE

Lista de Abreviaturas.....	10
Lista de Tabelas	12
Lista de Figuras	13
1. Introdução.....	14
1.1 Contextualização.....	14
1.2 Motivação.....	16
1.3 Objetivos.....	18
2. Conceitos base.....	20
2.1. A Osteoporose.....	20
2.1.1. Fisiologia óssea.....	20
2.1.2. Consequências.....	20
2.1.3. Classificação.....	21
2.1.4. Fatores de risco.....	22
2.1.4.1. Idade.....	22
2.1.4.2. Menopausa/género	23
2.1.4.3. Fatores genéticos	24
2.1.4.4. Corticoides.....	25
2.1.4.5. Doenças.....	26
2.1.4.6. Hábitos /Estilos de vida.....	26
2.1.5. Diagnóstico/Avaliação	27
2.1.5.1. Absorciometria de raios X de Dupla energia.....	28
2.1.5.2. Tomografia Computadorizada Quantitativa.....	28
2.2. Inteligência Artificial.....	29
2.3. Projeto <i>OSSEUS</i>	31
2.2.1. Contribuição da Engenharia Biomédica no <i>OSSEUS</i>	32

2.2.2. Contribuição da Engenharia Eletrônica no <i>OSSEUS</i>	34
2.2.3. Contribuição da Engenharia de software e computacional no <i>OSSEUS</i>	35
3. Estado da arte.....	35
4. Materiais e Métodos.....	41
4.1. Dados.....	41
4.2. Pré-Processamento dos dados	46
4.2.1. Seleção de <i>features</i> e avaliação de covariáveis.....	46
4.2.1.1 Teste <i>Kruskal–Wallis</i>	46
4.2.1.2 Teste de Correlação de <i>Pearson</i>	47
4.3. Preparação de dados para a Classificação.....	47
4.4. Algoritmos/Classificadores de <i>Machine learning</i>	48
4.4.1. Regressão Logística.....	48
4.4.2. <i>Support Vector Machine</i>	50
4.5. Avaliação	52
4.5.1. Matriz confusão.....	52
4.5.2. Sensibilidade, Especificidade e <i>F-measure</i>	54
5. Resultados e discussão.....	56
5.1. Descrição estatística dos dados.....	56
5.2. Análise da relevância e redundância das <i>features</i>	59
5.2.1. Relevância.....	59
5.2.2. Redundância.....	60
5.3. Classificação.....	61

6. Conclusão e Trabalho Futuro.....	67
Referências.....	69
Anexos.....	78

LISTA DE ABREVIATURAS

IA Inteligência Artificial

ML *Machine Learning*

DMO Densidade Mineral Óssea

OMS Organização Mundial de Saúde

SD *Standart Desviation*

DXA/DEXA *Dual-energy X-ray Absorptiometry* -Absorciometria de raios-X de dupla energia

EUA Estados Unidos da América

DPOC Doença Pulmonar Obstrutiva Crônica

AR Artrite Reumatóide

PTH Hormona Paratiroídeia

QTC *Quantitative Computed Tomography* -Tomografia Computadorizada Quantitativa

UFRN Universidade Federal do Rio Grande do Norte

LAIS Laboratório de Inovação Tecnológica em Saúde

MIT *Massachusetts Institute of Technology*

EM Eletromagnética

IMC Índice de Massa Corporal

ANN Artificial Neural Network

MLP Multi Layer Perceptron

AUROC Area Under the Receiver Operating characteristic Curve

SVM *Support Vector Machine*

RL Regressão logística

RLM Regressão logística Múltipla

KW Kruskal-Wallis

TP Verdadeiros Positivos

TN Verdadeiros Negativos

FP Falsos Positivos

FN Falsos Negativos

Sp Especificidade

Sn Sensibilidade

P Precisão

LISTA DE TABELAS

Tabela 1. Variáveis da base de dados do projeto <i>OSSEUS</i>	45
Tabela 2. Features ordenadas segundo Ranking dos valores H/X2 do resultado do teste KW.....	59
Tabela 3. Médias gerais dos classificadores, sensibilidade, especificidade e <i>F-measure</i> e precisão.....	61
Tabela 4. Resultados do algoritmo RLM.....	63
Tabela 5. Resultados do algoritmo SVM.....	65

LISTA DE FIGURAS

Figura 1. Regressão logística gráfico.....	49
Figura 2. <i>Support Vector Machine</i> gráfico.....	51
Figure 3. Exemplo de uma matriz confusão binária.....	53
Figura 4. Esquema representativo da metodologia do estudo.....	55
Figure 5. Estatística descritiva geral da população em estudo.....	57
Figure 6. Matriz da correlação de <i>Pearson</i>	60
Figura 7. Gráfico classificador RLM.....	62
Figura 8. Gráfico classificador SVM.....	64

1. Introdução

Este projeto foi desenvolvido no âmbito da unidade curricular Projeto do Mestrado em Instrumentação Biomédica da Faculdade de Ciências e Tecnologias da Universidade de Coimbra, no ano letivo de 2022/2023. O documento descreve o processo de elaboração de modelos computacionais baseados em inteligência artificial (IA) que estudam a predição de Osteoporose na população recorrendo aos dados do Projeto *OSSEUS*.

O mesmo documento é submetido a avaliação para a obtenção do grau de Mestre em Engenharia Biomédica, mais propriamente no ramo de Instrumentação Biomédica.

1.1. Contextualização

A osteoporose é uma doença esquelética sistêmica caracterizada por baixa massa óssea e deterioração da microarquitetura do tecido ósseo, com consequente aumento da fragilidade óssea e suscetibilidade a fraturas [1,2].

Operacionalmente, é definida com base na avaliação da densidade mineral óssea (DMO). A DMO é um indicador fundamental da saúde óssea e é medida por meio de técnicas como a absorciometria de raios-X de dupla energia (DXA), a técnica mais amplamente utilizada. Os resultados são frequentemente expressos em *scores* T, que comparam a DMO do paciente com a DMO média de uma população jovem e saudável [3,4].

De acordo com os critérios estabelecidos pela Organização Mundial da Saúde (OMS), a osteoporose é diagnosticada quando o score T < -2,5 SD [3].

Um score T de -1 a -2,5 indica osteopenia (baixa densidade mineral óssea, mas não tão grave como a osteoporose).

A osteoporose leva, portanto, à perda óssea, deixando os ossos tão fracos que uma simples queda ou colisão causa uma fratura. No entanto, no desenvolvimento da osteoporose, muitas vezes há um longo período de latência antes do aparecimento da principal manifestação clínica, as fraturas patológicas. Muitas vezes é uma doença silenciosa até que ocorram fraturas, o que causa importantes problemas secundários de saúde e até a morte [1,5].

Atualmente, estima-se que mais de 200 milhões de pessoas tenham osteoporose, e a taxa de incidência aumenta com a idade [1,3,4].

Embora seja observada em toda a população é em geral mais comum no sexo feminino, raça caucasiana e faixas etárias mais elevadas. Com uma população envelhecida e uma expectativa de vida mais longa, a osteoporose está a tornar-se cada vez mais uma epidemia global, pois é bastante prevalente nos idosos. Mais de 70% das pessoas com mais de 80 anos são afetadas [1,6].

De acordo com estatísticas da *International Osteoporosis Foundation*, uma em cada três mulheres e um em cada cinco homens com mais de 50 anos terão uma fratura osteoporótica durante a vida. Em todo o mundo, ocorrem aproximadamente 9 milhões de fraturas por ano como resultado da osteoporose. Cada fratura é um sinal de outra iminente [1,4].

Especificamente no Brasil, onde os nossos dados foram adquiridos, em 10 milhões de pessoas, 1 em 17 tem osteoporose e estima-se que num período de 20 anos houve um aumento de 50% no número de pessoas com maior probabilidade de desenvolver osteoporose. Além disso, o aumento para as faixas etárias acima de 65 anos terá aumento de 700% [5,7].

A situação em Portugal, segundo estatísticas, cerca de 12,2% da população portuguesa sofre de osteoporose, sendo cerca de 17% na população feminina e 2,6% na população masculina [8].

Segundo uma revisão de literatura realizada as estimativas de custos são baseadas em aspetos específicos de cada país (demografia, epidemiologia, economia, sistema de saúde e serviço de saúde), o que sugere cuidados de saúde amplamente diversos nos padrões de tratamento.

Para exemplificar, no Reino Unido, o tempo médio de internamento após uma fratura é de 30 dias, ao contrário da Suécia que é de 15 dias. Em Portugal o tratamento das fraturas da anca é feito de forma conservadora, enquanto noutros países da Europa o tratamento é maioritariamente cirúrgico. Neste contexto de dificuldades em medir com precisão o custo desta doença, conclui-se que os custos são substanciais.

No Brasil, num período de três anos, foram realizados 3.252.756 procedimentos relacionados com o tratamento da osteoporose na população idosa brasileira, o que gerou um gasto de R\$ 288.986.336,15 [6,7].

Em Portugal, os custos médios individuais relacionados à fratura foram estimados em € 13.434 [12.290; 14.576] para o primeiro ano e € 5.985 [4.982; 7045] pelo segundo ano após a fratura [8].

O aumento destes custos é diretamente afetado pela idade, chegando a dobrar quando comparados entre idosos e jovens. Numa projeção mundial, os custos atuais da fratura da anca são de US\$ 3,6 bilhões para homens e US\$ 19 bilhões para mulheres, até 2050 a previsão é de US\$ 14 bilhões para homens e US\$ 73 bilhões para mulheres. Estas estimativas de custo direto das fraturas osteoporóticas não levam em consideração os chamados custos indiretos para a economia, que envolvem incapacidade e perda de produtividade [4,9].

1.2. Motivação

Como já referido anteriormente a osteoporose é uma doença que afeta uma grande percentagem da população mundial e para além disso uma condição que envolve imensos gastos financeiros no seguimento da sua maior consequência que são as fraturas.

Paralelamente a este facto, o diagnóstico existente apresenta grandes constrangimentos no que diz respeito à acessibilidade e custos.

Assim, a prevenção da osteoporose e a redução de consequências e custos associados a esta patologia é uma das principais preocupações de saúde pública atualmente [1,3,4, 10].

A avaliação da massa óssea existente, determinando o risco de fratura com base nessa informação clínica é o objetivo crucial e final para avaliar pacientes com osteoporose, de forma a tomar decisões sobre a intervenção terapêutica mais apropriada [3,6].

A osteoporose geralmente é subtratada e subestimada, em parte porque é uma doença clinicamente silenciosa até que se manifeste na forma de fratura. Fraturas essas que causam importante morbidade e podem causar mortalidade.

Contudo, para além da diminuição da qualidade de vida, a osteoporose resulta num aumento da expectativa de vida ajustada por incapacidade e grande carga financeira para os sistemas de seguro saúde dos países responsáveis pelo atendimento desses pacientes [4,10,11].

Com o diagnóstico precoce esta doença pode ser prevenida, avaliando a DMO e intervindo antes da ocorrência de fraturas. Portanto, aumentar a conscientização desta realidade entre os médicos, o que, por sua vez, facilita o aumento da conscientização da população geral, será eficaz na prevenção dessa epidemia [1].

No entanto, acontece que o diagnóstico de osteoporose é muito complexo, caro e de difícil acesso na comunidade, globalmente, o exame de DMO ainda não é incluído como uma ferramenta de rastreio generalizado durante o exame de saúde regular [12].

Uma vez que apenas 1 em cada 3 pacientes com osteoporose é diagnosticado, os médicos precisam de melhorar a sua capacidade de identificar pacientes candidatos à triagem de DMO. Embora existam dados limitados sobre a correlação direta entre triagem eficaz e morbidade e mortalidade por fraturas, foi provado que o aumento de fraturas está associado a aumentos de morbidade e mortalidade. Portanto, identificar pacientes em risco, fazer um diagnóstico oportuno, implementar medidas de prevenção e iniciar terapia farmacológica para pacientes apropriados podem ajudar a minimizar o risco de fratura [13].

São então necessários métodos apropriados para triagem, diagnóstico e monitoramento destes pacientes. O desenvolvimento de novas ferramentas para o diagnóstico desta condição de saúde pode ser feito antes da fratura por fragilidade, permitindo a prevenção, bem como o tratamento.

Os critérios para diagnóstico de osteoporose são revistos e um plano para a avaliação de causas secundárias de osteoporose tem vindo a ser discutido. Fasihi, L. et al defende que muitos investigadores também têm vindo a desenvolver modelos preditivos utilizando fatores de risco para o rastreio da osteoporose [10,14].

Existem vários fatores de risco cientificamente comprovados que influenciam a presença/desenvolvimento da osteoporose. Fatores esses que podem entrar na base do

desenvolvimento de algoritmos de risco ou diagnóstico assim como ao nível de programas de rastreio.

Tendo em conta que a era da inteligência artificial (IA) chegou e todas as esferas da vida estão a envolver-se nela, a medicina não é exceção. A medicina assistida por IA experimentará um salto qualitativo, e a compreensão das pessoas sobre a IA médica aumentará também.

Foi desta realidade e necessidade que surgiu o projeto *OSSEUS* e continuam a surgir muitas outras soluções e ferramentas de auxílio no diagnóstico da Osteoporose, tendo por base o papel inovador e facilitador da IA.

1.3. Objetivos

O estudo visa explorar uma abordagem para a deteção precoce da osteoporose, utilizando dados do projeto *OSSEUS* em conjunto com técnicas de IA. Ao estabelecer correlações entre fatores de risco e padrões de dados, especialmente por meio de modelos preditivos de *Machine Learning* (ML), como Regressão Logística e *Support Vector Machine*, pretende-se encontrar uma possível ferramenta que auxilie o diagnóstico desta doença, tornando-o mais acessível e eficaz, como uma solução preventiva em saúde pública.

O projeto *OSSEUS* originou um equipamento que permite a realização de exames preventivos para diagnóstico de osteoporose. É um dispositivo portátil, de fácil utilização e de baixo custo, proporcionando e potencializando a medição de uma forma não invasiva da DMO através de radiação eletromagnética.

Assim, é importante perceber quais os fatores que podem prever risco de desenvolver esta condição da osteoporose, assim como identificar os que têm mais peso neste sentido, de forma a desenvolver planos estratégicos de prevenção e tratamento na população.

Uma das aplicações mais importantes dos algoritmos de IA e ML na área da saúde é sua capacidade de descobrir e detetar doenças e sintomas, para que possam ser diagnosticados mais precocemente. Foram identificadas até ao momento várias áreas críticas na saúde em que a tecnologia de ML pode fazer uma grande diferença [16].

Em geral, o objetivo é aprimorar e auxiliar o trabalho dos profissionais de saúde, melhorando a sua precisão no que diz respeito à osteoporose.

2. Conceitos base

2.1. A osteoporose

2.1.1. Fisiologia Óssea

Dos cerca de 2 kg de cálcio presentes no corpo humano, 98% são depositados nos ossos. Além disso, toda a produção de células sanguíneas (glóbulos brancos, vermelhos e plaquetas) é realizada na medula óssea que preenche as cavidades internas de muitos ossos. Os ossos também atuam como proteção para os órgãos mais vulneráveis do corpo, como coração, pulmões, cérebro e medula espinhal [16].

Diferentes tipos de células estão contidos no osso: osteoprogenitores, osteoblastos, osteócitos e osteoclastos.

Os osteócitos monitoram e mantêm o conteúdo de proteínas e minerais na matriz circundante, estes são constantemente reciclados.

Os osteócitos também direcionam a liberação de cálcio do osso para o sangue e a deposição de sais de cálcio na matriz circundante.

Os osteoblastos encontrados nas superfícies interna e externa do osso são responsáveis pela produção de osso novo por meio de um processo denominado osteogênese. Supõe-se que os osteoblastos respondam a estímulos hormonais ou mecânicos para iniciar a osteogênese.

Existem também os osteoclastos, que são células responsáveis pela remoção dos sais de cálcio. Deve haver um equilíbrio entre osteoclastos e osteoblastos, porque quando os osteoclastos removem os sais de cálcio mais rapidamente, os ossos ficam fracos [16,17].

2.1.2. Consequências

A osteoporose é um distúrbio esquelético caracterizado pela diminuição da resistência óssea, dependente da densidade e qualidade óssea, o que predispõe a fraturas. A gravidade do trauma não é a causa fundamental da fratura, mas a fragilidade do osso e as fraturas ocorrem, por ordem de frequência, na coluna vertebral (muitas vezes passam despercebidas e não são relatadas), na região distal do antebraço (fratura de *Colles*) e fêmur proximal (pacientes com mais de 75 anos) [18].

No caso das fraturas de punho e antebraço, por exemplo, capacidades de escrita, refeições ou tarefas de autocuidados, acabam por ser afetadas. Assim como nas fraturas de coluna

ou anca, a mobilidade fica comprometida, dificultando atividades diárias como caminhar, dobrar, levantar, etc.

Esta consequência leva a complicações comuns como, cifose, dor crônica, comprometimento respiratório e alto risco de desenvolver pneumonia. A maioria dos pacientes perde a capacidade de viver de forma independente, pois a funcionalidade fica limitada [3].

Situações como estas levam a que as fraturas osteoporóticas sejam consideradas a principal consequência da doença, podendo levar a uma perda de independência e bem-estar, aumento da morbidade, mortalidade e fazem com que a qualidade de vida das pessoas com osteoporose seja gravemente comprometida [1,3,5].

2.1.3. Classificação

A osteoporose pode ser classificada em duas categorias, primária e secundária. Primária, quando relacionada com o processo normal de envelhecimento, ou secundária, quando ocorre na presença de uma doença subjacente ou medicação. [1,4,6,14,19]

A osteoporose primária é dividida em três tipos: osteoporose pós-menopáusicas (Tipo I), osteoporose relacionada à idade (Tipo II) e osteoporose idiopática. A osteoporose pós-menopausa (Tipo I) e a osteoporose relacionada à idade (Tipo II) são as formas primárias mais comuns de perda óssea observadas na prática clínica. O primeiro (Tipo I) desenvolve-se em mulheres com deficiência de estrogênio, enquanto o segundo (Tipo II) ocorre em homens e mulheres, pois a densidade óssea diminui com o envelhecimento.[21] Tanto a osteoporose Tipo I como a Tipo II estão associadas à baixa absorção intestinal de cálcio e baixa massa óssea. No entanto, a causa da absorção reduzida de cálcio é bem diferente. No Tipo I, os mecanismos que resultam na perda óssea levam indiretamente ao comprometimento da absorção de cálcio. No Tipo II, no entanto, postula-se que a redução da massa óssea ocorre como resultado da má absorção de cálcio e, portanto, pode-se esperar que o tratamento com cálcio e/ou vitamina D tenha mais efeito [1].

A osteoporose secundária refere-se aos pacientes nos quais um fator causador ou processo de doença é identificável. Diversas doenças crônicas podem causar osteoporose

secundária, incluindo hipercortisolismo, hipertireoidismo, hiperparatireoidismo, doença pulmonar obstrutiva crônica, abuso de álcool e até imobilização [6, 10, 13].

2.1.4. Fatores de risco

Os fatores de risco para osteoporose são divididos em duas categorias: modificáveis e não modificáveis. [4,6,19]

Peso, tabagismo, consumo de álcool, sedentarismo, deficiência de cálcio na dieta e uso prolongado de glicocorticoides estão entre os fatores de risco para o grupo de osteoporose modificável [19].

Ao contrário, sexo, idade, raça, características genéticas e doenças estão entre os fatores de risco para o grupo de osteoporose não modificável [21].

De acordo com vários estudos, as causas mais comuns de osteoporose idiopática são a idade avançada e a menopausa. Esta última afeta 35% das mulheres com mais de 50 anos e 50% das mulheres com mais de 75 anos.

O tratamento crônico com corticoides é a principal causa de osteoporose secundária a medicamentos e é conhecida internacionalmente como osteoporose induzida por glicocorticoides [18].

2.1.4.1. Idade

Os pacientes mais velhos são muito mais suscetíveis a fraturas em qualquer DMO do que os pacientes mais jovens devido a vários fatores, incluindo a qualidade do osso envelhecido, que envolve mais do que a DMO [22,23].

A perda óssea ocorre ao longo da vida em mulheres e homens devido à deficiência de cálcio, deficiência hormonal e alterações na formação óssea [24].

O osso mais antigo também pode perder a sua capacidade de remodelar e adicionar osso em locais críticos microscópicos de suporte de peso ou stress [25].

Fatores nutricionais contribuem para a perda óssea, incluindo baixa ingestão de cálcio e vitamina D, desnutrição, diminuição da absorção de cálcio pelo trato gastrointestinal e comprometimento da conversão renal de vitamina D. Tudo isso ocorre com o aumento da idade [23].

Os idosos também são mais propensos a ter diminuição do equilíbrio, visão e massa muscular em relação aos pacientes mais jovens, e muitas vezes tomam medicamentos que

podem afetar o equilíbrio e a cognição. Todos esses fatores aumentam ainda mais o risco de fratura de pacientes idosos em comparação com pacientes mais jovens com a mesma DMO [26].

Desta forma, idosos têm um risco particularmente alto de fraturas osteoporóticas não apenas por causa de anormalidades, mas também por fatores que afetam a incidência de quedas. Os adultos mais velhos são mais suscetíveis a fraturas do que os mais jovens com a mesma DMO [27].

A deficiência de vitamina D é mais comum em idosos, e a suplementação de mulheres deficientes parece retardar a perda óssea, pelo menos durante o inverno. Cálcio mais vitamina D pode reduzir substancialmente o risco de fratura da anca em idosos frágeis. Um estudo randomizado mostrou que o estrogênio continua eficaz na prevenção de fraturas vertebrais em mulheres idosas. Não há razão para acreditar que a eficácia de outros agentes diminua com a idade [28].

Evidências associadas a fatores nutricionais, particularmente cálcio e vitamina D, são revistas, assim como a associação do risco de quedas com fratura. A avaliação da fragilidade nos idosos, a sarcopenia associada e a multimorbidade são já consideradas na avaliação do risco de queda e fratura e no controle da osteoporose [25].

Nos humanos, a massa óssea aumenta desde o nascimento até os 30-35 anos de idade. Em seguida, geralmente começa um declínio lento, embora no caso das mulheres um período de perda óssea mais rápida tenda a se sobrepor por cerca de 10 anos após a menopausa. Pode-se ver que as três estratégias óbvias para alcançar uma boa massa óssea seriam: 1. Construir ossos melhores em primeiro lugar. 2 Para evitar a perda de osso antes que ela ocorra ou pelo menos para diminuir a taxa de perda. 3 Encontrar meios de aumentar a massa óssea mesmo quando perdida [29].

2.1.4.2. Menopausa/ gênero

De acordo com os dados do Censo dos EUA de 2010 para a população total (não institucionalizada e institucionalizada), estima-se que 8,2 milhões de mulheres com 50 anos ou mais foram diagnosticadas com osteoporose (em comparação com 2 milhões de homens) e um adicional de 27,3 milhões de mulheres tinham baixa massa óssea. Aproximadamente 71% das fraturas osteoporóticas em pessoas com 50 anos ou mais ocorrem em mulheres. O objetivo desta Diretriz de Prática Clínica é fornecer

recomendações atualizadas e baseadas em evidências para a prevenção, triagem e diagnóstico da osteoporose pós-menopausa [1].

A massa óssea é geralmente estável em indivíduos saudáveis na pré-menopausa. À medida que os níveis de estrogênio diminuem na menopausa, a reabsorção óssea pelos osteoclastos aumenta e excede a capacidade de formar osso novo pelos osteoblastos. Isso leva à perda óssea e perda da microarquitetura do osso trabecular e cortical, o que aumenta o risco de fratura. A massa óssea pode começar a diminuir antes da menopausa, com uma fase acelerada de perda óssea durante a transição menopáusicas [28].

Esses fatores podem, portanto, ser mais difundidos em relação ao gênero, pois nas mulheres, a menopausa prematura e a perda da função ovariana antes da menopausa são outros fatores de risco para a osteoporose [4].

As mulheres típicas da pré-menopausa geralmente começam a perder massa óssea quando os níveis de estrogênio diminuem e as taxas de renovação da remodelação óssea e a atividade dos osteoclastos aumentam devido à falta dos efeitos supressores do estrogênio nos receptores ósseos. Isso pode ocorrer na menopausa natural ou numa idade mais precoce após a ooforectomia completa [30].

Os homens também perdem massa óssea com o avanço da idade devido à perda de hormonas. Neste cenário, os osteoblastos podem ser restaurados e o equilíbrio normal da renovação óssea e reparo/substituição é perdido.

2.1.4.3.Fatores genéticos

Os nossos genes podem determinar a variação na DMO e o gene do recetor de vitamina D modula a transcrição de genes alvo envolvidos na captação de cálcio ou formação óssea, incluindo proteínas de ligação ao cálcio 61-63 e osteocalcina.

A osteoporose também pode ser causada por mutações no gene do colágeno I alfa 1 (COL1A1) que tem sido consistentemente associado ao risco de fratura [31].

De acordo com alguma literatura, a raça é um fator determinante da DMO e influencia a incidência de fraturas. Os resultados de estudos entre diferentes grupos raciais e étnicos demonstram que os valores médios mais altos de DMO e as taxas mais baixas de fratura da anca foram relatados para mulheres negras e tiveram uma taxa de mortalidade mais

alta durante a hospitalização por fratura do que mulheres brancas segundo dados da Health Care Financing Administration de 1980 a 1982 [6].

Além disso, o tamanho do corpo é outro fator que afeta o risco de fratura. Um estudo em mulheres brancas não hispânicas mais velhas mostrou que mulheres mais velhas com corpo menor correm maior risco de fratura da anca devido aos valores mais baixos de DMO da anca. Embora todas as medidas do tamanho corporal (incluindo peso corporal total, variação percentual do peso desde os 25 anos, massa magra, massa gorda, percentual de gordura corporal, circunferência da anca, índice de massa corporal e índice de massa corporal modificado) tenham sido associadas ao risco de fratura da anca, a medição do peso corporal total por si só foi considerado suficiente para determinar o risco de fratura e não foi melhorado por medições de outros atributos de tamanho e composição corporal.

Mulheres com histórico materno de fratura da anca têm aproximadamente duas vezes mais chances de sofrer fraturas da anca do que mulheres sem esse histórico familiar [6].

2.1.4.4.Corticoides

Os corticosteroides são a principal causa de osteoporose secundária devido ao tratamento medicamentoso. Doses superiores a 5 mg por dia e períodos de tratamento superiores a 3 meses aumentam o risco de osteoporose e fraturas por fragilidade. Portanto, é essencial estar ciente das medidas para reduzir o risco de osteoporose nestes pacientes [18].

Os glicocorticoides são usados no tratamento de várias doenças, incluindo doenças autoimunes, inflamatórias, dermatológicas, respiratórias, malignidades e transplantes de órgãos sólidos. Estudos mostram que cerca de 30 a 50% dos pacientes que recebem glicocorticoides desenvolvem fraturas [18, 32].

Estes glicocorticoides têm uma ampla variedade de efeitos diretos e indiretos no osso, que foram recentemente revistos em detalhes por Henneickle e colegas [33].

Nas células ósseas, incluindo osteócitos, osteoblastos e osteoclastos, há múltiplos efeitos diretos numa fase inicial. Em seguida, a estimulação glicocorticoide de osteoclastos induz sobrevivência prolongada, permitindo reabsorção óssea excessiva. Eles também podem induzir a apoptose dos osteócitos, contribuindo para o risco de fratura precoce que ocorre antes que a DMO seja reduzida e, no final, os glicocorticoides reduzem o recrutamento

de precursores osteoblásticos, levando à diminuição da diferenciação e função osteoblástica, resultando na diminuição da formação óssea.

Além disso, diminuições na reabsorção de cálcio, supressão da hormona do crescimento, alteração nas hormonas sexuais e alterações na pulsatilidade da paratireoide são efeitos indiretos que contribuem para a perda óssea induzida por glicocorticoides [32].

2.1.4.5. Doenças

Existem diversas patologias relacionadas com a Osteoporose.

A doença pulmonar obstrutiva crônica (DPOC) é uma das causas secundárias da osteoporose.

A prevalência global combinada de osteoporose entre indivíduos com DPOC é de 38%, sendo uma das comorbidades comuns que afetam o custo da DPOC tratada em hospitais. As fraturas osteoporóticas estão associadas à diminuição da qualidade de vida e aumento da mortalidade em pacientes com DPOC [34].

A artrite reumatóide (AR) está associada à osteoporose devido à inflamação sistêmica ativa, imobilização e uso de glicocorticóides. Existem duas formas de osteoporose ocorrer na AR: perda óssea generalizada com distribuição axial e perda óssea periarticular ou localizada nas proximidades das articulações inflamadas [31].

A literatura relata que a taxa de fraturas de coluna ou quadril é maior em pacientes com AR em comparação com pacientes com osteoporose primária [31].

2.1.4.6. Hábitos/Estilos de vida

O consumo de álcool e a relação com a DMO não apresenta peso significativo na literatura. Há um estudo que sugere que o consumo baixo a moderado de álcool pode até exercer efeitos protetores sobre a saúde óssea em mulheres idosas [31].

No entanto, os pacientes osteoporóticos devem ser regularmente aconselhados sobre a cessação tabágica e ingestão de álcool. Uma meta-análise identificou uma associação não linear entre o consumo de álcool e o risco de fratura de quadril. O consumo moderado de álcool foi inversamente associado ao risco de fratura da anca, enquanto o consumo excessivo de álcool foi associado a um risco elevado de fratura da anca. O consumo de

álcool (baixo e moderado/alto) pode ter um impacto prejudicial na saúde óssea nos compartimentos cortical e trabecular no rádio distal em homens, e resultados semelhantes foram encontrados nos compartimentos trabecular e tibial distal de mulheres com consumo mínimo de álcool e baixo consumo de álcool, respectivamente, sugerindo que evitar o álcool pode ser benéfico para a saúde óssea [35].

O tabagismo continua a ser uma variável controversa também, segundo alguns artigos é considerado fator de risco para osteoporose e está relacionado à perda de massa óssea e aumento do risco de fraturas osteoporóticas. No entanto, outros investigadores não encontraram nenhuma ligação entre tabagismo e risco de fratura.

O tabaco é aparentemente responsável, pelo menos em parte, pela eficiência reduzida da absorção intestinal de cálcio [6].

A hormona paratiroídea (PTH) e os metabólitos da vitamina D desempenham um papel vital na regulação da homeostase do cálcio e do metabolismo ósseo e, num estudo em jovens fumadores do sexo masculino, o PTH sérico mostrou um nível crescente em fumadores pesados [31].

No que diz respeito aos estilos de vida sedentários, Nancy. E, et al concluiu que um baixo nível de atividade física está positivamente correlacionado com o risco de fratura em alguns estudos [30].

O exercício pode ser benéfico na atenuação da perda óssea. Em vários estudos de revisão, o exercício demonstrou reduzir principalmente a perda de DMO e aumentar a rigidez óssea, validando a sua capacidade de atuar como uma barreira ao desenvolvimento da osteoporose [36]. Especificamente, exercício em carga com impacto mostrou uma correlação positiva com a DMO. O treino de resistência progressiva de alta intensidade pode melhorar a altura vertebral, bem como a DMO da coluna lombar e do colo do fêmur. Além disso, caminhar também pode ajudar a reduzir a reabsorção óssea [37].

2.1.5. Diagnóstico/ Avaliação

As avaliações radiológicas e laboratoriais da DMO são a ferramenta padrão atualmente utilizadas para diagnosticar a osteoporose. Vários métodos de imagem foram desenvolvidos para medir a DMO, sendo os mais comuns a Absorciometria de raios X de

Dupla energia (DEXA ou DXA - *Dual-energy X-ray Absorptiometry*) e a Tomografia Computadorizada Quantitativa (QCT - *Quantitative Computed Tomography*) [3].

Atualmente, a Fundação Nacional de Osteoporose, a Sociedade Internacional de Densitometria Clínica e as diretrizes da OMS consideram o DXA central da anca e/ou coluna como a medida preferencial para o diagnóstico de osteoporose. Um varrimento de absorciometria de raios X dupla pode ser concluído em cinco minutos com exposição mínima à radiação [3].

Os resultados de um teste DXA são relatados como um score T. A OMS define a osteoporose como um score T de DMO menor ou igual a 2,5 desvios padrão [3, 38].

2.1.5.1. Absorciometria de raios X de Dupla energia

A DXA é considerada a medida *gold standard* dos métodos utilizados para o diagnóstico da osteoporose. Este teste é capaz de medir o conteúdo mineral ósseo em qualquer local do corpo, mas geralmente é usado em locais centrais (coluna lombar e fêmur proximal) e locais periféricos, incluindo o antebraço distal.

Isso é feito passando dois feixes de diferentes energias através do osso no local a ser medido, expondo o paciente a níveis de radiação aproximadamente 90% menores do que uma radiografia padrão do tórax.

A unidade de medida da densidade óssea com o uso de DXA é a densidade de área (g/cm²); no entanto, a DMO é relatada como um T-score com base nessa medição. As técnicas periféricas de DXA analisam a DMO no rádio distal e no calcâneo com alta precisão e baixa exposição à radiação [3].

2.1.5.2. Tomografia Computadorizada Quantitativa

QCT também pode ser usado para medir DMO da coluna lombar ou locais periféricos. Embora a DMO possa ser medida por QCT em qualquer sistema de tomografia computadorizada, o equipamento deve ser calibrado usando um fantoma de calibração que contém elementos com densidades padronizadas de hidroxiapatita de cálcio. A precisão da QCT da coluna vertebral na previsão de fratura da coluna vertebral é

comparável à do DXA, mas tem a vantagem de medir a DMO volumétrica verdadeira ou tridimensional, em contraste com a DMO areal obtida do DXA.

A QCT pode distinguir entre osso cortical e trabecular e, portanto, é mais sensível a alterações na DMO causadas pela maior taxa de renovação óssea do osso trabecular. Também é preciso o suficiente para detetar alterações esqueléticas ao longo do tempo e pode ser usado para acompanhar o estado da doença ou para monitorar os resultados da terapia da osteoporose.

Embora essas técnicas de imagem sejam úteis para determinar a DMO ao diagnosticar a osteoporose, o papel da medição da DMO na avaliação da eficácia do tratamento, particularmente na redução do risco de fratura, permanece obscuro [3].

Pacientes com diagnóstico de osteoporose também devem ter uma avaliação laboratorial de sua função renal e tireoidiana, 25-hidroxivitamina D e nível de cálcio. Os marcadores de formação óssea são libertados pelos osteoblastos e geralmente são medidos no soro ou na urina [3].

2.2. Inteligência Artificial

No que diz respeito à IA, o termo teve origem nos Estados Unidos, quando um grupo de cientistas se reuniu numa conferência no Dartmouth College (1956), época em que a sua essência se referia a algoritmos estabelecidos pela análise de dados existentes e, às vezes, por auto-aprendizagem [38,39,40].

Após décadas de desenvolvimento, os avanços no campo da IA demonstraram sucesso numa variedade de tarefas clínicas e aplicação de *big data*, supercomputação, redes de sensores, neurociência e em outras áreas [40].

Atualmente, a IA constitui um sistema que combina poder computacional com conjuntos de dados (idealmente *big data*) para permitir a resolução de problemas. Um sub-tipo da IA é o ML, que utiliza vários algoritmos para aprender com os dados. O ML pode ser baseado em diferentes abordagens de aprendizagem, sendo as mais importantes a aprendizagem supervisionada e a aprendizagem não supervisionada. Na última década, o ML supervisionado revelou-se o mais eficiente e, portanto, tornou-se o principal pilar dos em aplicações na área da saúde com suporte de IA. Para treinar um sistema baseado em

aprendizagem supervisionada, a máquina tem de ser alimentada com dados já disponíveis e robustos, pois a qualidade desses dados de entrada determinará a qualidade da saída [41].

Em geral, o desenvolvimento de um algoritmo de ML específico requer um conjunto de dados de treino e um conjunto de dados de teste. Em alguns casos, um terceiro conjunto de dados é reservado apenas para fins de validação [14].

A análise preditiva consiste na aplicação de algoritmos para entender a estrutura dos dados existentes e gerar regras de previsão. Esses algoritmos podem ser utilizados num cenário não supervisionado, onde apenas preditores (co variáveis) estão disponíveis no conjunto de dados, ou em problemas supervisionados, quando, além dos preditores, também está disponível uma resposta de interesse, responsável por orientar a análise [37,42].

Na área da saúde, modelos preditivos podem ser utilizados para estimar o risco de determinada ocorrência, dado um conjunto de características socioeconômicas e demográficas, relacionadas ao estilo de vida e condições de saúde, entre outras. Os seus resultados, quando combinados com medidas de saúde pública aplicadas ao nível populacional, podem trazer resultados positivos na redução de custos e na ocorrência de intervenções, como tratamentos e ações preventivas.

De facto, foram feitos progressos substanciais em muitas áreas da medicina humana e a IA foi gradualmente integrada à prática médica diária e fez progressos consideráveis nessa área. No geral, a IA na medicina pode ser dividida em 2 subtipos, virtual e física, com a primeira incluindo, por exemplo, soluções de imagem e ferramentas de suporte à decisão de tratamento, e a segunda, por exemplo, no desenvolvimento de prótese e cirurgia assistida por robô. No que diz respeito ao risco/diagnóstico da osteoporose, a IA desempenha atualmente o papel principal, com soluções disponíveis (ou em desenvolvimento) para facilitar o diagnóstico, avaliação do risco de fratura, deteção de fratura, avaliação da qualidade óssea e decisão do tratamento [42].

2.3. Projeto *OSSEUS*

O Projeto *OSSEUS* foi realizado pela Universidade Federal do Rio Grande do Norte (UFRN) em parceria com o Laboratório de Inovação Tecnológica em Saúde (LAIS), e com colaboração de investigadores do *Massachusetts Institute of Technology* (MIT), nos Estados Unidos.

O projeto teve como foco principal a criação de um dispositivo que realiza exames triagem de casos de osteoporose na população de forma preventiva, evitando exames de alta complexidade, permitindo medir de uma forma não invasiva a DMO através de radiação eletromagnética. Isto é, permite uma triagem inicial e seleção dos pacientes que precisam de uma avaliação adicional com densitometria.

O dispositivo *OSSEUS* utiliza um microcontrolador e outros dispositivos periféricos para medir a atenuação eletromagnética (EM) no tecido ósseo. Começa por recolher informações do paciente por meio de um questionário digital. Em seguida, mede a atenuação do sinal EM entre duas antenas, selecionando o ponto de maior potência como referência. O teste é realizado no dedo do paciente (falange média do dedo médio) e os resultados são exibidos no equipamento e enviados para armazenamento em nuvem. O dispositivo emite uma onda EM, (que opera a uma frequência de 2,45 GHz através do tecido ósseo) que é atenuada pelo dedo do paciente, e um circuito converte o sinal em informações digitais. Com base nas características do paciente e na atenuação do sinal, o dispositivo pode indicar a necessidade de um exame DXA [47].

Assim, a radiação ionizante necessária para os testes de DXA é substituída por radiação eletromagnética de micro-ondas não ionizantes.

Os vários estudos existentes até ao momento, no âmbito deste projeto, demonstram que o *OSSEUS* é eficaz na medição da atenuação. O equipamento é capaz de diferenciar estruturas ósseas com densidades diferentes, caracterizando a sua microarquitetura.

Este dispositivo é acessível, de fácil utilização e transporte, não invasivo, de baixo custo e permite a deteção precoce da osteoporose, reduz os custos e otimiza encaminhamentos para testes de alta complexidade [47].

Este exame realizado pelo *OSSEUS*, para além de ser não invasivo, pode ser conduzido em intervalos de tempo mais curtos, uma vez que não apresenta quaisquer efeitos

colaterais para os pacientes, ao contrário de outros métodos tradicionais. Esta ferramenta resulta da combinação de várias áreas da engenharia, nomeadamente a engenharia biomédica, a engenharia eletrónica, a informática e o software.

2.3.1. Contribuição da Engenharia Biomédica no OSSEUS

O conhecimento das propriedades dielétricas dos tecidos corporais envolvidos neste estudo é essencial, visto que as ondas eletromagnéticas apresentam comportamentos distintos consoante o tecido ao qual são aplicadas. Para evitar que estes comportamentos inesperados comprometam os resultados, algoritmos avançados foram implementados para isolar a área específica destinada ao diagnóstico auxiliar. A falange foi selecionada para este diagnóstico de osteoporose, uma vez que, após vários testes, estudos têm indicado que medidas ósseas em locais periféricos, como as falanges, têm revelado potencial no diagnóstico de osteoporose, permitindo identificar pacientes com elevado risco de fraturas, com sensibilidade igual à do *gold standard* (DXA da coluna) [42, 43].

2.3.2. Contribuição da Engenharia Eletrónica para o OSSEUS

Através desta área de conhecimento, compreendemos que as ondas eletromagnéticas desempenham um papel crucial na determinação da permissividade óssea e, consequentemente, da DMO. Antenas desempenham um papel importante neste processo. Uma antena é um transdutor que converte corrente elétrica alternada contida num condutor numa onda eletromagnética de igual frequência, irradiando-a para o espaço livre. O tipo mais comum de antena Yagi é utilizado em antenas de TV e é facilmente encontrado em residências que recebem sinais VHF e UHF. O maior elemento é denominado refletor e os demais são elementos diretores menores do que o elemento energizado. Vários parâmetros críticos afetam o desempenho de uma antena Yagi, incluindo potência irradiada, diretividade e eficiência [44].

A proteção eletromagnética é relevante em circuitos que são sensíveis à interferência de radiação eletromagnética, sendo essencial bloquear essa interferência. Esta técnica é denominada blindagem e pode ser implementada através de uma caixa metálica em torno

do circuito ou de um material metálico em forma de grade. A eficácia da blindagem varia consoante diversos fatores como a frequência, geometria, tipo de campo e direção [45].

Absorvedores de ondas eletromagnéticas são dispositivos compostos que absorvem radiação incidente conforme a transmissibilidade de frequência, refletividade e dispersão. Eles podem ser ressonantes ou de banda larga, dependendo do comportamento da frequência. O dispositivo *Ecosorb* é um exemplo de absorvedor de micro-ondas que atenua interferência eletromagnética convertendo-a em calor. Pode ser aplicado em espaços livres e cavidades, reduzindo taxas de reflexão e ressonâncias [46].

Com a expertise em engenharia eletrônica, foi possível ainda projetar e desenvolver componentes eletrônicos integrados para a emissão e recepção de sinais eletromagnéticos, seguido pelo seu processamento através de microprocessadores [47].

2.3.3. Contribuição da Engenharia de Software e Computacional para o OSSEUS

Estas áreas foram essenciais para integrar todo o hardware desenvolvido, armazenar e processar os dados gerados pelo sistema. Estes dados fornecerão informações cruciais na identificação de grupos de risco, direcionando recursos de diagnóstico, prevenção e tratamento adequado de doenças osteometabólicas, bem como monitorar a evolução dessas doenças na população ao longo do tempo.

Os dispositivos atuais são miniaturizados e capazes de converter operações analógico-digitais, fornecer resultados através de interfaces gráficas online, analisar dados e processar redes neurais através do conceito de Internet das Coisas (IoT). Estes dispositivos, de tamanho reduzido e elevada capacidade de processamento, estão em constante evolução, oferecendo retrocompatibilidade e velocidade de processamento melhorada [47].

O resultado final da combinação de todas estas áreas é uma ferramenta funcional, adequada para validação clínica, permitindo a realização de testes auxiliares para determinar a DMO e diagnosticar doenças osteometabólicas como a osteoporose. A patente para este registo já foi gerada.

No Brasil, mais do que proporcionar avanços no diagnóstico da osteoporose e facilitar a prevenção de consequências mais graves e custos, o projeto *OSSEUS* tem vindo a ter uma importância grande na condução de políticas públicas, com a incorporação da tecnologia nos cuidados primários.

Um dos objetivos que se espera com estudos futuros com base neste projeto é conseguir identificar os fatores de risco que mais influenciam na progressão da doença e, assim, conseguir desenvolver modelos para deteção precoce [47].

3. Estado da arte

Assim como o *OSSEUS* muitos outros estudos são constantemente realizados e muitas outras soluções têm sido construídas neste sentido, tendo a IA vindo a desempenhar um papel muito importante.

Diante deste cenário, esforços crescentes foram realizados nos últimos anos por cientistas de IA para apoiar e melhorar o acompanhamento desta doença, nomeadamente no que diz respeito à identificação do risco/diagnóstico [37].

Foi na década de 90 que foi publicado um primeiro estudo relatando o uso de fatores de risco integrados com auxílio de redes neurais artificiais como *Multi Layer Perceptron* (MLP) para identificar um subconjunto de alto risco de fraturas osteoporóticas. Este estudo comprovou que esta técnica é um elemento importante como ferramenta de triagem para aqueles que realmente precisam de avaliação da DMO e para auxiliar em estratégias cada vez mais eficazes de prevenção de fraturas [48].

Já no final da década surgiu outra integração com MLP, porém, havia uma grande limitação para o processo utilizado, pois eram necessárias para a implementação da técnica amostras vivas - cortes histológicos do osso trabecular e posterior digitalização do mesmo [49].

Num estudo de 2002 foi desenvolvido um método chamado *Osteoporosis Risk Assessment Tool for Asians* (OSTA), para direcionar a população feminina chinesa com alto risco de fraturas para medição de DMO, utilizando um cálculo que envolve duas variáveis para baixo e alto risco de osteoporose. A combinação dos exames de ultrassom com a ferramenta proporcionou uma sensibilidade de 91% na identificação de indivíduos com maior risco de osteoporose [50].

Rizzi et al., em 2004, propuseram uma estratégia de prevenção de baixo custo para estimar a ocorrência de osteoporose, tendo como parâmetros um conjunto de características individuais correlacionadas com a DMO. Para este estudo, um modelo neuronal baseado em mistura de funções gaussianas (*Gaussian Mixture Model*), treinado pelo procedimento de maximização da expectativa hierárquica de divisão (SHEM) foi proposto com base na abordagem de máxima verossimilhança. Constatou-se a partir das

características de entrada, as mais importantes são idade e entrada na menopausa. Em relação ao risco de fratura, o IMC deve ser considerado para fornecer informações úteis para a estimativa dos valores de DMO [51].

Em 2005, Wang et al. [52] desenvolveram um sistema inteligente de apoio ao diagnóstico com integração de modelos heterogêneos, compostos por MLP e árvores de decisão. Observou-se que os estudos desenvolvidos com sistemas homogêneos não produziram um ganho significativo de desempenho. Por outro lado, os conjuntos híbridos superaram os homogêneos gerando resultados com acerto de 85,79%. Esses estudos também propõem a construção de mais conjuntos de tipos e tamanhos variados, utilizando fatores de risco selecionados por técnicas de estimação. Nesse mesmo ano, Sadatsafavi et al. [53] desenvolveram alguns modelos de rede neural a partir de um conjunto de dados de 2.158 mulheres iranianas na pós-menopausa em que os valores de DMO foram medidos por DEXA. Eles tiveram os resultados mais eficazes no seu trabalho usando entre 3 e 5 fatores somáticos (idade, peso, anos de pós-menopausa, esteroides e estrogênio) atingindo 86,3% de sensibilidade e 72,1% de especificidade.

Um ano mais tarde, Chiu et al. propuseram uma rede neural para identificar a osteoporose na população idosa (idade média de 63 anos, variando de 50 a 91 anos, 157 homens e 1246 mulheres). Com seleção aleatória de três grupos, sendo 703 para o grupo de treino, 350 participantes para o grupo de seleção e os outros 350 no conjunto de teste. Os fatores de risco mais relevantes incluíram características demográficas, medidas antropométricas e dados clínicos (sexo, idade, peso, altura, índice de massa corporal, estado pós-menopausa e consumo de café) [54].

Lemineur et al., em 2007, fez uso de redes neurais com o algoritmo de treino de *Levenberg-Marquardt* combinado com três parâmetros de DMO (DMO do colo do fêmur, DMO do corpo total e DMO da coluna L2 L4) juntamente com a idade dos 304 pacientes. Este estudo teve como objetivo identificar pacientes com osteoporose antes da ocorrência de fratura. 81,66% dos resultados alcançados foram estimados corretamente [55].

Em 2009, o trabalho de Leslie et al. [56] realizado no Canadá com 4.015 mulheres com 50 anos ou mais, foi implementado um programa de vigilância da osteoporose como auxílio para informar, orientar o rastreamento, a prevenção e o tratamento da doença. O uso de

algoritmos e métodos de classificação foi aplicado a diferentes conjuntos de diagnósticos hospitalares, diagnósticos de referências médicas e prescrições de medicamentos para tratamento de doenças. Os resultados obtidos alcançaram uma sensibilidade total de 93,3%, especificidade variando de 50,8 a 91,4% e 81,2 a 99,1% para discriminação osteoporótica de DMO normal. Os algoritmos que incluíram prescrições de medicamentos como variáveis obtiveram maior sensibilidade em relação aos baseados apenas em exames diagnósticos.

Mantzaris [57] realizou também um estudo na Grécia (2020), onde a maioria dos sujeitos era do sexo feminino e apenas uma baixa percentagem dos casos do sexo masculino. O objetivo deste estudo foi estimar o risco de osteoporose por meio da aplicação de Redes Neurais Probabilísticas (PNN) e Quantificação de Vetores de Aprendizagem (LVQ), sendo o NNP o de melhor desempenho, com eficácia de 96,58%. Os fatores idade, sexo, altura e peso foram identificados como os mais determinantes, obtendo-se o resultado da área sob a curva (AUC) para cada fator somático. O sexo, entretanto, não teve muita relevância, pois neste estudo o número de homens foi ínfimo em relação ao feminino, o que podemos identificar como uma limitação.

Considerando os fatores somáticos, Cos Juez et al. [58] desenvolveram um método matemático para identificar mulheres na pós-menopausa com osteoporose, considerando apenas fatores nutricionais e estilo de vida como fatores influentes no diagnóstico. As variáveis obtidas por meio de questionário, juntamente com a DMO dos pacientes calculada por densitometria, foram processadas por meio de algoritmos genéticos para identificar os fatores mais relevantes na identificação. Posteriormente, MLPs foram usadas para construir um modelo matemático que determinava a relação entre as variáveis de entrada e a DMO.

Em 2012, Jennane et al. [59] argumentaram através dos seus estudos que a medição do osso cortical é mais apropriada do que a medição do osso trabecular para triagem de osteoporose usando SVM com radiografias dentárias panorâmicas de 69 mulheres na pós-menopausa. Como resultados, a acurácia das medidas corticais com média de largura e variância foi de 87% e das medidas trabeculares com média de comprimento e ângulo de 65%.

Um estudo mais abrangente foi realizado em 2013 num grupo de 1.674 mulheres coreanas na pós-menopausa, cujo objetivo era avaliar o risco de osteoporose. Foi feita uma comparação com vários modelos baseados em algoritmos populares de ML, como SVM, RF, ANN e LR, com quatro instrumentos convencionais de decisão clínica: OST, ORAI, SCORE e OSIRIS. Foi quantificado que 1.674 mulheres apresentavam osteoporose em qualquer uma das seguintes localizações: quadril, colo do fêmur ou coluna lombar. O SVM previu o risco de osteoporose com uma AUC de 0,827, precisão de 76,7%, sensibilidade de 77,8% e especificidade de 76,0% na anca total, colo do fêmur ou coluna lombar. Os fatores somáticos selecionados pelo SVM foram idade, altura, peso, índice de massa corporal, duração da menopausa, duração da amamentação, terapia com estrogênio, hiperlipidemia, hipertensão, osteoartrite e diabetes mellitus [60].

Outro estudo realizado por Taфраouti et al. [61] abordou a integração de SVM com extração de características baseada no modelo de movimento Browniano Fracionário para identificação de osteoporose. Para isso, eles obtiveram como base para o estudo imagens de raios X de ossos de 77 pacientes. Nesse processo, a imagem original é subdividida em sub-imagens, e em cada sub-imagem é aplicada uma rotação em um ângulo θ , obtendo-se um sinal que é modelado pelo movimento browniano fracionário. As características de textura de cada imagem são obtidas pela concatenação das características de todas as suas sub-imagens. Estes são inseridos como parâmetros em quatro funções *Kernel* (Polynomial, Quadrática, RBF, Linear), com o melhor resultado em termos de precisão sendo a função kernel linear e polinomial 95 e 93%, respectivamente.

Vários estudos foram realizados em 2015, Iliou et al. [62], utilizou um conjunto de 589 dados extraídos da população grega, que realizou exames de densitometria óssea e laboratorial, aplicando um classificador MLP com o método de validação cruzada de dez vezes (*k-fold cross-validation*). Neste estudo, consideramos três e cinco fatores diagnósticos para a predição do risco de osteoporose, classificando-os em três categorias: normal, osteopenia e osteoporose. Liu et al. [63] utilizaram nos seus estudos uma amostra de 725 casos controle de ambos os sexos, dos quais 228 casos eram pacientes internados no Hospital Nacional da Universidade de Taiwan. O modelo preditivo em questão envolveu apenas idosos acima de 60 anos e foca na prevenção do risco de fratura de quadril. Todos os pacientes foram entrevistados com o mesmo questionário, os dados

coletados foram inseridos num banco de dados, para identificar as variáveis mais significativas foi aplicada a análise de sensibilidade e abordagem de pesos de conexão. Com este método, as variáveis de entrada foram classificadas de acordo com a importância, identificando as 10 variáveis principais (DMO total, queda em casa, altura, IMC, hipertensão, incontinência fecal, escolaridade), que mostram uma grande proporção de contribuição para prever fratura da anca. Neste estudo, modelos de RNA de retropropagação de três camadas foram aplicados a pacientes masculinos e femininos separadamente. O modelo masculino teve um desempenho melhor que o feminino por ter um menor grau de complexidade.

Yu et al., em 2016 [64] utilizaram como amostras os casos de 119 pacientes internados, sendo 55 pacientes com osteoporose, 64 pacientes sem osteoporose, idade média de 65 anos, 49 homens e 79 mulheres. Do total de 17 parâmetros, seis são características de imagem, seis são clínicas extraídas por ortopedistas e radiologistas e cinco queixas principais. Como resultados as sensibilidades foram 94,5 e 63,6%, especificidade 96,9 e 87,5%.

Mais recentemente em 2022, segundo um estudo de Hans, et al., o desempenho de muitos desses algoritmos de IA recém-desenvolvidos demonstrou ser pelo menos comparável ao de especialistas médicos, ou mesmo superior. No entanto, mesmo que os resultados do estudo pareçam promissores à primeira vista, eles sempre devem ser interpretados com cautela. O uso de padrões de referência inadequados ou a seleção de variáveis de pouco ou nenhum valor na prática clínica são limitações encontradas com frequência. Consequentemente, há uma clara necessidade de pesquisa clínica de alta qualidade neste campo da IA. Isso poderia, por exemplo, ser alcançado através do estabelecimento de uma "estrutura de melhores práticas" acordada internacionalmente que considere todas as partes interessadas relevantes [37].

Num estudo desenvolvido por Fasihi, et al., neste ano 2023, os modelos foram desenvolvidos e comparados usando oito algoritmos de IA (árvore de decisão, floresta aleatória (RF), k-vizinhos mais próximos, SVM, *Gradient Boost* (GB), árvores extras, *Ada Boost* (AB) e MLP para prever a osteoporose e recomendar programas desportivos [14].

Os resultados mostraram que para prever indivíduos saudáveis, osteopenia e osteoporose, o algoritmo FR com uma área sob a curva ROC AUROC de 0,91 teve o melhor desempenho em homens, enquanto o algoritmo GB com AUROC de 0,95 teve o melhor desempenho em mulheres em comparação com outros algoritmos de classificação. A previsão do algoritmo RF em mulheres e homens, com AUROC de 0,96 e 0,99, respectivamente, mostrou o melhor desempenho no diagnóstico do tipo de exercício para indivíduos saudáveis e aqueles com osteopenia e osteoporose.

Ainda neste estudo de Fasihi, et al conclui-se também que um dos principais objetivos atualmente é determinar os fatores de risco para osteoporose em dados clínicos que incluem características físicas, história pessoal e médica e exames laboratoriais em homens e mulheres [14].

Contudo o estudo demonstrou que a aplicação dos algoritmos de IA num ambiente clínico pode ajudar os prestadores de cuidados primários a classificar pacientes com osteoporose e melhorar o tratamento, recomendando programas de exercícios apropriados.

Ao usar algoritmos de classificação na prática clínica como uma ferramenta de rastreio/triagem, tanto os médicos quanto os pacientes estariam mais cientes dos fatores de risco da osteoporose e tomariam mais medidas preventivas nos estados iniciais da doença para evitar resultados adversos [14, 37].

Mas, o uso destes algoritmos na saúde pública ainda é incipiente. Consequentemente, há uma clara necessidade de pesquisa clínica de alta qualidade neste campo da IA. O desenvolvimento da IA exige que especialistas em ética participem na supervisão de todo o processo para resolver questões éticas em relação a dados, alocação de recursos e prática. Métodos apropriados para triagem, diagnóstico e monitoramento desses pacientes desempenharão um papel importantíssimo, por isso vários investigadores visaram continuar a desenvolver modelos preditivos utilizando fatores de risco para a triagem da osteoporose [14, 37].

Portanto, as investigações de IA não devem limitar-se à precisão e sensibilidade do relatório, mas deve concentrar-se na natureza das doenças, na sua etiologia e patogênese, e deve enriquecer a nossa compreensão e conhecimento. Algoritmos interpretáveis serão reconhecidos por mais pessoas e profissionais de saúde e trarão tratamento médico baseado em IA para a vida das pessoas [39].

4. Materiais e Métodos

4.1. Dados

Um total de 505 indivíduos foram incluídos no estudo.

Os dados utilizados para estudo pertencem à base de dados do Projeto *OSSEUS*, dados estes que contêm informações recolhidas em indivíduos da população brasileira com recurso ao equipamento.

No que diz respeito às componentes do equipamento *OSSEUS*, existe um teste específico não invasivo através de um dispositivo e durante a aquisição desses dados são também recolhidos dados clínicos e pessoais do paciente através de questionários.

Este dispositivo *OSSEUS* utiliza um microcontrolador e outros dispositivos periféricos para medir a atenuação EM no osso. Relativamente ao procedimento do teste realizado é o seguinte:

1. Recolha de informações do paciente: O operador preenche as características do paciente num questionário eletrónico, fornecendo informações relevantes, que estão mencionadas na tabela 1, como idade, sexo, dimensões dos dedos, histórico de fraturas, etc. (que são também objeto de estudo neste trabalho)
2. Medição da atenuação do sinal EM: O *OSSEUS* realiza a medição de atenuação do sinal EM entre duas antenas. O ponto de maior potência recebida durante a medição é selecionado como referência ou calibração.
3. Teste de atenuação: Após a calibração, o teste de atenuação é realizado no dedo do paciente. O dispositivo emite uma onda eletromagnética numa faixa de frequência pré-definida (2,5GHz) em direção ao dedo do paciente, mais especificamente, à falange média do dedo médio.
4. Conversão de Sinal: Uma antena do dispositivo recebe a onda EM após esta ter sido atenuada pelo obstáculo do dedo do paciente. Um circuito acoplado à antena identifica a intensidade do sinal analógico e converte-o num sinal digital.
5. Cálculo da Diferença: Com base nas características do paciente e na atenuação do sinal EM, o microcontrolador processa as informações. A variável "diferença" é calculada subtraindo o valor de tensão medido para cada amostra do valor de referência (sinal máximo medido sem obstáculos), que é de 1700 mV.

A variável "diferença" não representa um significado biofísico específico. Trata-se da diferença entre o sinal recebido numa antena nas situações com a amostra original e a amostra alterada servindo como obstáculo para o sinal transmitido, isto é, com o tecido ósseo do dedo do paciente. Esta é relevante pois indica o grau de atenuação do sinal, que está relacionado à densidade do tecido ósseo. Tecidos ósseos mais porosos causarão menos atenuação no sinal EM transmitido em relação aos valores de referência que representam indivíduos saudáveis ou patológicos. Portanto, essa diferença permite classificar o tecido ósseo do paciente como normal ou com baixa densidade [47].

Com base neste procedimento que foi descrito, quer as variáveis recolhidas no teste específico quer as variáveis recolhida através do questionário vão entrar então para o algoritmo/classificação do *OSSEUS*. Estes dados/variáveis recolhidos no Projeto *OSSEUS* e que foram utilizadas para o desenvolvimento deste presente estudo estão elencados na tabela abaixo:

Nome	Descrição	Tipo
ID	Número de identificação do paciente	Inteiro
Sexo	Género do paciente	Texto Opções: Masculino ou Feminino
Idade	Idade do paciente	Inteiro
Peso	Peso do paciente em Kg	Inteiro
Altura	Altura do paciente em cm	Inteiro
IMC	Índice de massa corporal	Inteiro
Raça	Raça do paciente	Texto Opções: “Branco”, “preto”, “indígena”, “pardo”, “mulato”, “caboclo” ou “cafuzo”
Álcool	Paciente consome álcool	Texto Opções: sim ou não

Tabaco	Paciente consome tabaco ou já consumiu	Texto Opções: sim ou não
Atividade física	Paciente pratica atividade física	Texto Opções: sim ou não
Leite	Paciente consome leite ou derivados	Texto Opções: sim ou não
Cálcio	Paciente toma suplementos de cálcio	Texto Opções: sim ou não
Vitamina D	Paciente faz exposição ao sol ou toma suplementação de vitamina D	Texto Opções: sim ou não
Quedas	Paciente sofreu fratura devido a uma queda	Texto Opções: sim ou não
Pais com osteoporose	Algum dos pais do paciente foi diagnosticado com osteoporose ou fraturou um osso após uma queda pequena	Texto Opções: sim ou não
Pais com hipercefose	Algum dos pais do paciente apresenta hipercefose	Texto Opções: sim ou não
Corticoesteroides	Paciente tomou corticóides (cortisona, prednisona etc.) por mais de 3 meses	Texto Opções: sim ou não
Artrite	Paciente foi diagnosticado com artrite reumatóide	Texto Opções: sim ou não
Doenças	Paciente apresenta outros diagnósticos como hipertireoidismo, glândulas paratireoides hiperativas, diabetes ou distúrbios gastrointestinais como doença de Crohn ou doença celíaca,etc...)	Texto Opções: sim ou não

Testosterona	Paciente sofreu impotência, falta de libido ou outros sintomas relacionados com baixos níveis de testosterona	Texto Opções: sim ou não
Menopausa	Paciente está na menopausa	Texto Opções: sim ou não
Score	Resultado com base no DEXA, que é considerado como padrão ouro no diagnóstico da Osteoporose	Texto Opções: normal, baixa, osteoporose
MedialComprimento	Comprimento da falange medial em mm	Inteiro
MedialLargura	Largura da falange medial em mm	Inteiro
MedialAltura	Altura da falange medial em mm	Inteiro
Desvio	Desvio padrão da leitura da densidade óssea na coluna apontado pelo DEXA	Número racional
Desviof	Desvio padrão da leitura da densidade óssea no fêmur apontado pelo DEXA (caso solicitado)	Número racional
Desviob	Desvio padrão da leitura da densidade óssea no antebraço apontado pelo DEXA (caso solicitado)	Número racional
Desviot	Desvio padrão da leitura da densidade óssea do corpo inteiro relatado pelo DEXA (se solicitado)	Número racional
MedialComprimento	Comprimento da falange medial em mm	Inteiro
MedialLargura	Largura da falange medial em mm	Inteiro
MedialAltura	Altura da falange medial em mm	Inteiro

Calibração	Intensidade do sinal do <i>OSSEUS</i> sem obstáculo entre as antenas (sem o dedo do paciente)	Número racional
Leitura	Intensidade do sinal do equipamento <i>OSSEUS</i> com obstáculo entre as antenas (com o dedo do paciente)	Número racional
Diferença	Diferença entre a medição com dedo do paciente no equipamento <i>OSSEUS</i> e a medição no vazio (calibração)	Número racional

Tabela 1. Variáveis da base de dados do projeto *OSSEUS*

4.2. Pré-Processamento dos dados

Para realizar a análise e desenvolver o modelo com estes dados do projeto *OSSEUS*, foi utilizado o *software Matlab* versão R2021B, que trabalha com os dados no formato Excel original. Inicialmente, identificámos os fatores clínicos mais significativos da osteoporose e os dados foram organizados. Foi ainda criada uma nova variável, chamada "diagnóstico", com base na variável "score", que indica se o paciente tem ou não tem osteoporose.

Seguidamente, com recurso à *Statistic Toolbox* do *Matlab* foram realizadas análises estatísticas descritivas das variáveis, o que nos permitiu entender e caracterizar melhor a população do estudo.

As variáveis foram uniformizadas e renomeadas para o idioma português. Também foi verificada a validade dos dados, garantindo que não havia valores ausentes (NaNs) ou outros dados discrepantes (*outliers*).

Além disso, tratámos as variáveis ordinais e nominais, convertendo-as em números para facilitar o processamento. Isto é, as respostas "Sim" e "Não" por 0 e 1,

Uma etapa de standardização foi realizada para as *features* numéricas (idade, IMC e diferença) do conjunto de dados. Para cada *feature* (F), calculámos a média (m) e o desvio padrão (σ) e, em seguida, aplicou-se a fórmula de standardização z-score:

$$Fstd = \frac{F-m}{\sigma} \quad (1)$$

O objetivo desta standardização foi tornar as variáveis significativamente comparáveis para análise, entre *features* e modelos, por forma a ficarem todas numa escala comum.

4.2.1. Seleção de *features* e avaliação de covariáveis

Um conceito muito importante presente na fase de pré-processamento de dados na criação de um modelo de ML é a seleção de *features*.

Em problemas de classificação de ML, existem muitas *features* com base nos quais a classificação final pode ser feita.

Neste trabalho existem diferentes *features* que afetam o resultado da variável “diagnóstico” na base de dados. Primeiramente, foi necessário compreender quais delas eram as mais relevantes para o estudo e quais contribuíam mais para a deteção do risco de osteoporose. Em segundo lugar, devemos minimizar a redundância, por exemplo, se houver duas *features* com alta correlação, podemos ignorar uma delas e usar apenas uma, pois manter as duas é basicamente irrelevante para o modelo, já que influenciam o resultado de maneira semelhante.

Portanto, para conhecer a relevância das *features* e estudar o peso de cada uma no *outcome*, variável “diagnóstico”, utilizámos o teste de *Kruskal Wallis* e para verificar a redundância, o teste de correlação de *Pearson*.

4.2.1.1. Teste Kruskal–Wallis

Foi considerado que as *features* são independentes, e aplicou-se o teste não paramétrico de *Kruskal-Wallis* (KW), um caso estendido do teste U de *Mann-Whiney*.

É usado para os modelos de classificação binários, para comparar mais de dois grupos independentes através do cálculo da estatística H. A estatísticas H (equivalente ao valor de Chi-sq) para cada *feature* é dada por:

$$H = 1 + \frac{12}{n \times (n+1)} \sum ni \times (Ri - R)^2 \quad (2)$$

- R_i : Média de classificações para as amostras pertencentes à classe i
- R : Média de todas as classificações para todas as classes
- n_i : Número de padrões pertencentes à classe i
- n : Número total de padrões

As *features* foram então selecionadas com base na sua classificação de acordo com as estatísticas H , visto que as *features* mais discriminantes estão relacionadas com maior valor de H . Posto isto, após ordenar as *features* com base no teste de KW, as *features* com $H=0$ foram excluídas, e as demais foram submetidas à correlação de *Pearson* para analisar redundância [65].

4.2.1.2. Teste de Correlação de *Pearson*

Posteriormente ao teste KW, o subconjunto de *features* selecionados foi inspecionado quanto à redundância, calculando o coeficiente de correlação de *Pearson* entre elas.

Em seguida, foram retiradas as *features* que apresentaram valor de correlação absoluta superior a 0,90, que foram consideradas como fortemente correlacionadas, restando apenas uma *feature* delas para o subconjunto final. Aqui, as variáveis foram convertidas para o tipo *double* para ser possível executar a função *corrplot*. (Zero (0) e um (1) foram usados para questões com respostas sim e não (0 =não e 1 =sim) e quanto ao gênero feminino=0 e masculino=1.

Após este teste, foram, portanto, eleitas as *features* relevantes e não redundantes para então serem submetidas a classificadores [66].

4.3. Preparação de dados para a Classificação

Para determinar qual o conjunto de *features* que melhor prediz a variável “diagnóstico”, começamos por organizar os dados numa tabela Excel, ordenadas de acordo com o ranking do teste KW e após a remoção das *features* muito correlacionadas. Esta tabela constitui a nossa base de dados para os classificadores.

As *features* que foram consideradas para a análise perfazem um total de dezassete e incluem: idade, menopausa, IMC, corticoides, níveis de vitamina D, níveis de cálcio, diferença, histórico de quedas, gênero, atividade física, consumo de tabaco, consumo de

álcool, presença de artrite, raça, níveis de testosterona, consumo de leite e se os pais têm hipercifose e se pais têm osteoporose.

Em seguida, desenvolvemos um programa que de forma cumulativa vai adicionando *features*. Começamos com uma combinação composta por 2 *features* e aumentamos gradualmente a complexidade, testando combinações até dezessete *features* juntas. A intenção é tentar perceber como a adição de *features* afetam a avaliação e os resultados dos nossos modelos.

Além disso, para melhorar a avaliação dos modelos, gerou-se de forma aleatória vários conjuntos de dados para treino e teste. Para cada conjunto implementou-se o procedimento descrito acima de adicionar cumulativamente as *features*. No presente estudo, 70% dos dados foram utilizados para treinar os algoritmos ($n = 505 \times 0,70 = 354$) e 30% para testar o desempenho preditivo dos modelos ajustados ($n = 505 \times 0,30 = 151$).

4.4. Algoritmos/ Classificadores de *Machine Learning*

O principal objetivo da aplicação de algoritmos/classificadores de ML é ajustar um modelo que relacione a resposta, Y, aos preditores/features, X, a fim de prever esse evento em observações futuras. No desenvolvimento de classificadores supervisionados é requerida a existência de um conjunto de dados já anotados que é utilizado para construir um modelo capaz de prever a classificação de outros dados futuros e, por isso, também são conhecidos como modelos preditivos.

No caso deste estudo quer-se prever a ausência (classe 0) ou presença (classe 1) de osteoporose.

Para o presente estudo, dois algoritmos foram selecionados para comparação de classificação: Regressão Logística Múltipla (RLM) e *Support Vector Machine* (SVM).

4.4.1. Regressão logística

A classificação por regressão logística (RL) é amplamente usada em ML para prever a probabilidade de ocorrência de um evento ou resultado específico.

A RL é um modelo estatístico utilizado para modelar a relação entre uma variável dependente binária (ou seja, uma variável que possui apenas dois valores possíveis, como

“sim” ou “não”, “1” ou “0”) e uma ou mais variáveis independentes. Esta emprega a função logística, uma curva sigmoide que mapeia a variável “y” como uma função de “x”.

Em contraste com a simples definição de uma linha reta ou um hiperplano, a RL utiliza a função logística para expressar a saída como um valor entre dois possíveis. [11,67]

A forma da função logística é dada por:

$$f(x) = \frac{1}{1+e^{-x}} \quad (3)$$

Visualmente, essa equação gera uma curva em forma de "S", como ilustrado na Figura 1 abaixo.

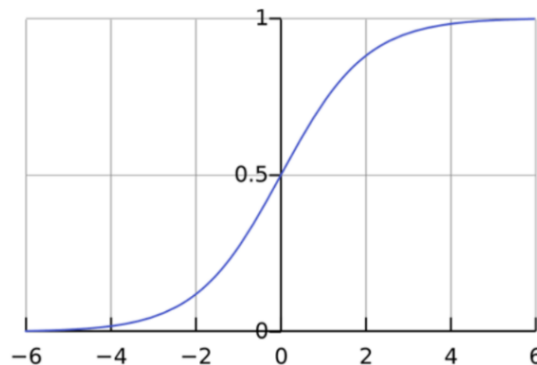


Figura 1. Regressão logística gráfico. Imagem extraída do artigo *Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after PKP for osteoporotic vertebral compression fractures, 2023* [68].

Neste estudo, o método de regressão logística tem como objetivo modelar relações entre múltiplas variáveis independentes e uma variável dependente - uma abordagem conhecida como Regressão Logística Múltipla (RLM).

A RLM difere da regressão linear múltipla, uma vez que a variável de resultado (a variável dependente) é dicotômica, representando a presença ou ausência de osteoporose. O seu propósito é desenvolver o modelo mais apropriado, eficiente e biologicamente válido para descrever a relação entre o resultado e um conjunto de preditores. As *features*

independentes são referidas como covariáveis. Na RLM, as variáveis preditoras podem ser de diferentes tipos de dados, incluindo categóricos, ordinais e numéricos. Uma análise cuidadosa de várias variáveis preditoras é essencial para identificar aquelas que melhor preveem o resultado desejado [67, 68].

Dessa forma, lidamos com múltiplas variáveis explicativas que influenciam o valor da variável dependente. Para modelar esses conjuntos de dados de entrada, as fórmulas de regressão logística assumem uma relação linear entre as diferentes variáveis independentes. É possível ajustar a função sigmoide e calcular o valor de saída final como:

$$y = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n)}} \quad (4)$$

Aqui, os símbolos β_i representam os coeficientes de regressão, um valor que indica a contribuição de cada variável preditora para o modelo.

É aconselhado que os dados de treino devem ser standardizados, usando a normalização z-score, tal como foi descrito na secção 4.2 Em resumo, a RL é uma técnica de classificação aplicada a diferentes conjuntos de dados para prever se um ponto de dados pertence a uma classe específica. A hipótese sigmoide é utilizada para calcular a probabilidade de "y" pertencer a uma classe determinada. Os dados de treino são normalizados através o *score* Z para garantir uma análise precisa e justa [11, 67].

4.4.2. Support Vector Machine

A classificação *Support Vector Machine* (SVM) foi realizada usando as mesmas variáveis/ features selecionadas e em relação à mesma variável de saída/*outcome*-“diagnóstico”.

O treino SVM vai encontrar um hiperplano num espaço N-dimensional (N - número de *features*) que divide os dados de treino em duas classes, maximizando a margem de separação entre as classes.

Existem muitos hiperplanos possíveis que podem ser escolhidos para separar as duas classes de pontos de dados. O objetivo do treino SVM é encontrar um hiperplano que melhor separe as diferentes classes de dados, maximizando a margem entre os pontos de dados mais próximos de cada classe. Esses pontos são conhecidos como vetores de suporte, o termo que confere nome ao algoritmo. A margem, nesse contexto, representa a distância entre o hiperplano e esses vetores de suporte [68,69,70].

Apesar do fato de que o hiperplano divide os dados linearmente, os SVMs podem ser usados para resolver problemas não lineares traduzindo os dados em espaços dimensionais superiores onde eles podem ser separados linearmente—esse mapeamento é realizado usando funções de *kernel*.

Isto é, o SVM assume diferentes formas, como o SVM linear e o SVM não linear. O primeiro é aplicável a cenários em que os dados podem ser claramente separados por um hiperplano linear. Por outro lado, o SVM não linear utiliza funções de *kernels*, para transformar o espaço de entrada num espaço multidimensional, possibilitando a separação quando uma separação linear não é suficiente [69].

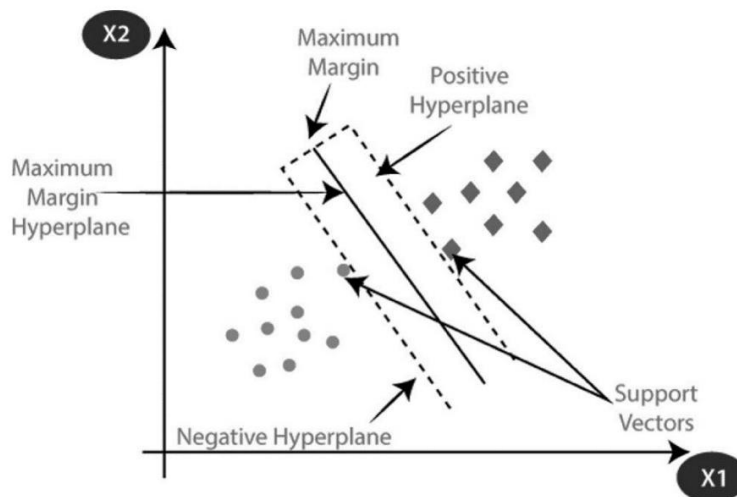


Figura 2. Support Vector Machine gráfico. Imagem extraída do artigo *Support Vector Machines (SVMS) Based Advanced Healthcare System Using Machine Learning Techniques*, 2022 [69].

Este algoritmo adota uma abordagem de classificação de amostras, porém não fornece diretamente estimativas das probabilidades de associação de classe presentes no conjunto de dados, o que representa uma diferença fundamental em relação à RLM.

4.5. Avaliação

Após a aprendizagem é necessário avaliar. Para a avaliação foi construída a matriz de confusão do modelo, de forma a visualizar quanto o modelo está a acertar. Com base na matriz de confusão, foram encontrados alguns parâmetros de avaliação como Sensibilidade, Especificidade, Precisão e *F-measure*.

4.5.1. Matriz confusão

A matriz de confusão é uma das métricas mais simples e fáceis de utilizar ao resolver problemas de classificação para encontrar a eficácia e precisão relativas ao modelo de classificação. Posto isto, no nosso projeto essa matriz de confusão vai identificar verdadeiros positivos (TP), verdadeiros negativos (TN), falsos positivos (FP) e falsos negativos (FN).

Esta medida foi utilizada para visualizar e resumir as avaliações de desempenho dos métodos utilizados após o algoritmo de classificação, no caso do nosso estudo esta é uma classificação binária, sendo zero (0): Indivíduo não tem osteoporose e um (1): Indivíduo tem osteoporose.

A Figura 3 mostra de forma genérica a matriz de confusão para o problema de classificação binária. Cada coluna da matriz representa instâncias numa classe prevista, enquanto cada linha representa a instância numa classe real ou vice-versa.

		Valor predito	
		0	1
Valor real	0	TN	FP
	1	FN	TP

Figure 3. Exemplo de uma matriz confusão binária

No estudo, as situações podem ser descritas da seguinte forma:

- Verdadeiro Positivo (TP): Quando tanto o valor previsto pelo modelo quanto o valor real são iguais a um (1). Isso indica que o modelo corretamente identificou uma pessoa com osteoporose.
- Verdadeiro Negativo (TN): Quando o modelo prevê um valor de zero (0) e o valor real também é zero (0). Nesse caso, o modelo acertadamente reconheceu a ausência de osteoporose na pessoa.
- Falso Positivo (FP): Ocorre quando o valor previsto pelo modelo é um (1), mas o valor real é zero (0). Isso significa que o modelo erroneamente previu que uma pessoa tem osteoporose, quando na verdade não tem.
- Falso Negativo (FN): Se o valor previsto for zero (0) enquanto o valor real é um (1), isso equivale a classificar erroneamente uma pessoa com osteoporose como não tendo a condição.

Estas definições ajudam a compreender as diferentes combinações entre as previsões do modelo e os valores reais, permitindo avaliar o desempenho do modelo de forma mais precisa.

4.5.2. Especificidade, Sensibilidade, Precisão e *F-measure*

Com base nos campos presentes na matriz confusão podemos obter os valores de especificidade (*Sp*), sensibilidade (*Sn*), precisão (*P*) e *F-measure*, as métricas mais frequentes para classificação e avaliação.

A sensibilidade e a especificidade, representam respetivamente a capacidade de o classificador detetar resultados positivos ou a identificação de resultados negativos, respetivamente.

A precisão explica quantos doentes previstos como tendo osteoporose realmente tinham.

Para a condição em que dois modelos têm baixa precisão e alta sensibilidade ou vice-versa, torna-se difícil comparar esses modelos, portanto, o *F-Measure* fornece uma forma de combinar *Sn* e *P* numa única medida que capta ambas as propriedades.

O cálculo dessas métricas de desempenho é feito com recurso às equações:

$$Sn = \frac{TP}{TP+FN} \quad (5)$$

$$Sp = \frac{TN}{TN+FP} \quad (6)$$

$$P = \frac{TP}{TP+FP} \quad (7)$$

$$Fmeasure = \frac{2 \times Sn \times P}{Sn+P} \quad (8)$$

- *Sn*: Sensibilidade
- *Sp*: Especificidade
- *P*: Precisão
- *TP*: Verdadeiros Positivos
- *TN*: Verdadeiros Negativos
- *FP*: Falsos Positivos

A visão geral da metodologia com as etapas e métodos executados está representada no esquema presente na figura 4.

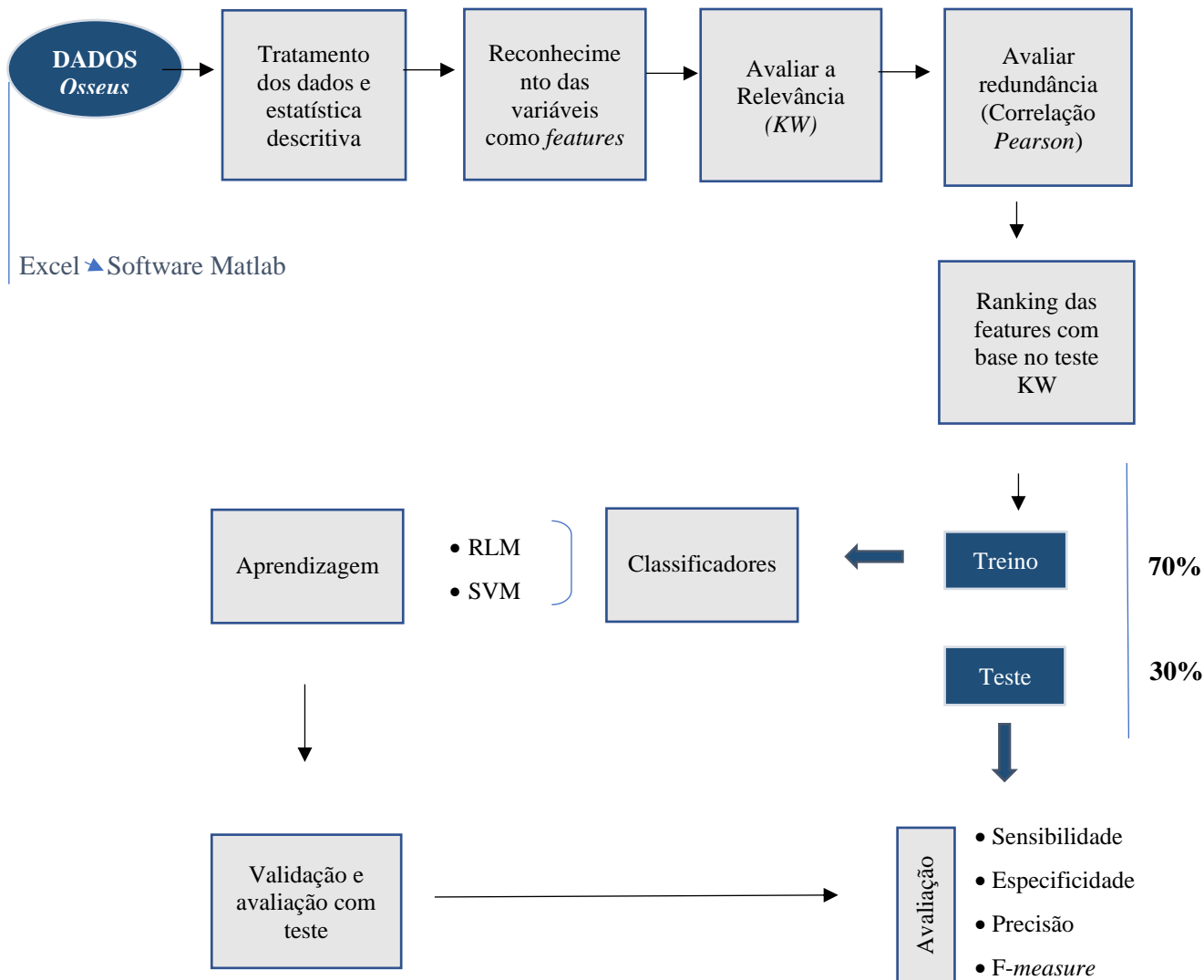


Figura 4. Esquema representativo da metodologia do estudo

5. Resultados e discussão

5.1. Descrição estatística dos dados

No total, foi incluído no estudo uma amostra de 505 indivíduos, n=505.

De acordo com a estatística descritiva inicial, a média de idade é de 62,23 anos e a média do IMC é de 33,44. Há uma alta percentagem de mulheres, 91%.

Em relação à nossa variável de saída/*Outcome*, 78,4% (396 indivíduos) dos sujeitos apresentam diagnóstico de osteoporose.

Em geral 82% da população em estudo está na menopausa, 58% nunca caiu, apenas 27% toma corticoides e uma percentagem insignificante tem artrite (8%). No que diz respeito à suplementação, 41% tomam suplementos de cálcio e 42% tomam suplementos de vitamina D.

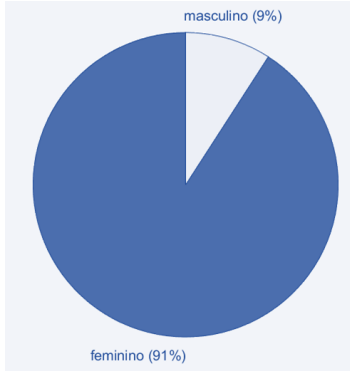
No que diz respeito a hábitos/estilos de vida, vários indivíduos apresentam comportamentos sedentários (76%) e 12% ,80%, 37% consomem álcool, leite e tabaco respetivamente.

Cerca de metade dos indivíduos (47%) possuem outras doenças.

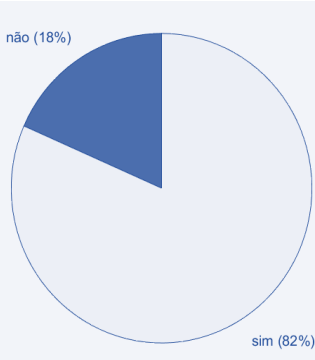
Relativamente à componente genética, quase 100% da população em estudo não tem pais com hipercifose (97%) e 34% possui pais com osteoporose.

A Figura 5 apresenta uma descrição geral das características da população presente no estudo.

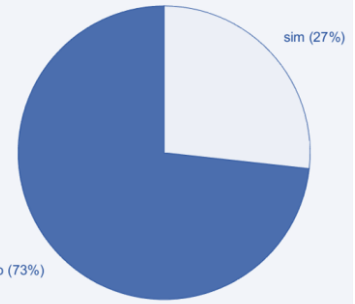
Genero



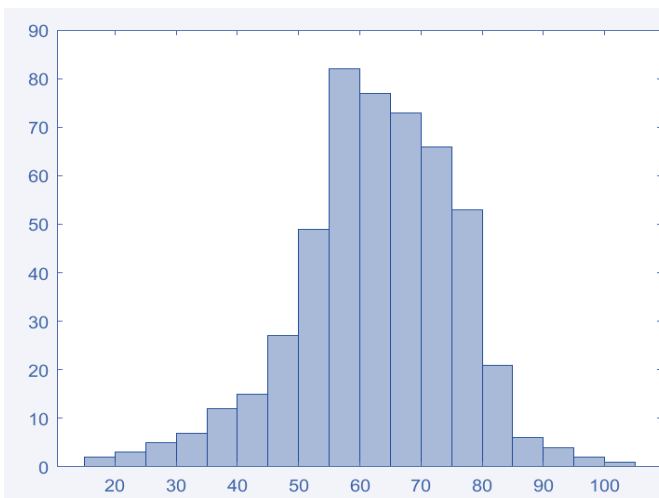
Menopausa



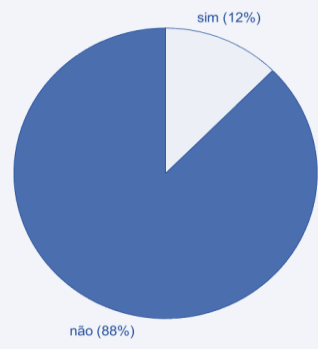
Corticoides



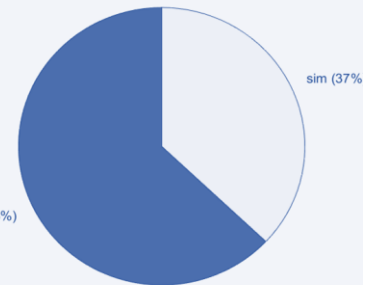
Idade



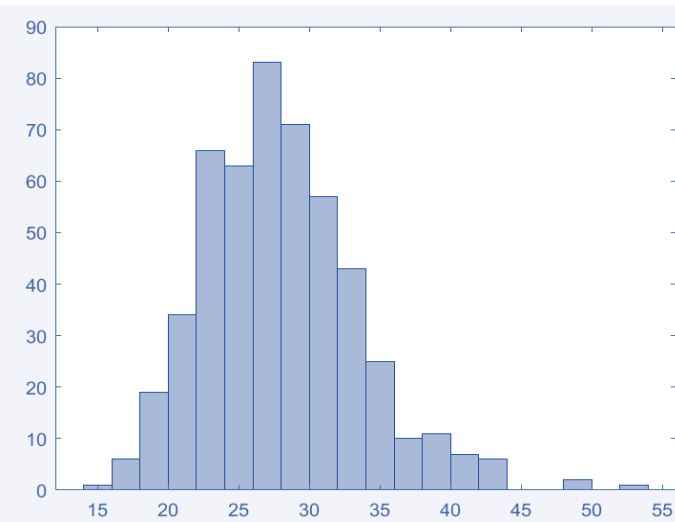
Alcool



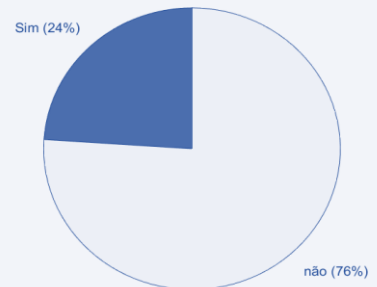
Tabaco



IMC



Atividade física



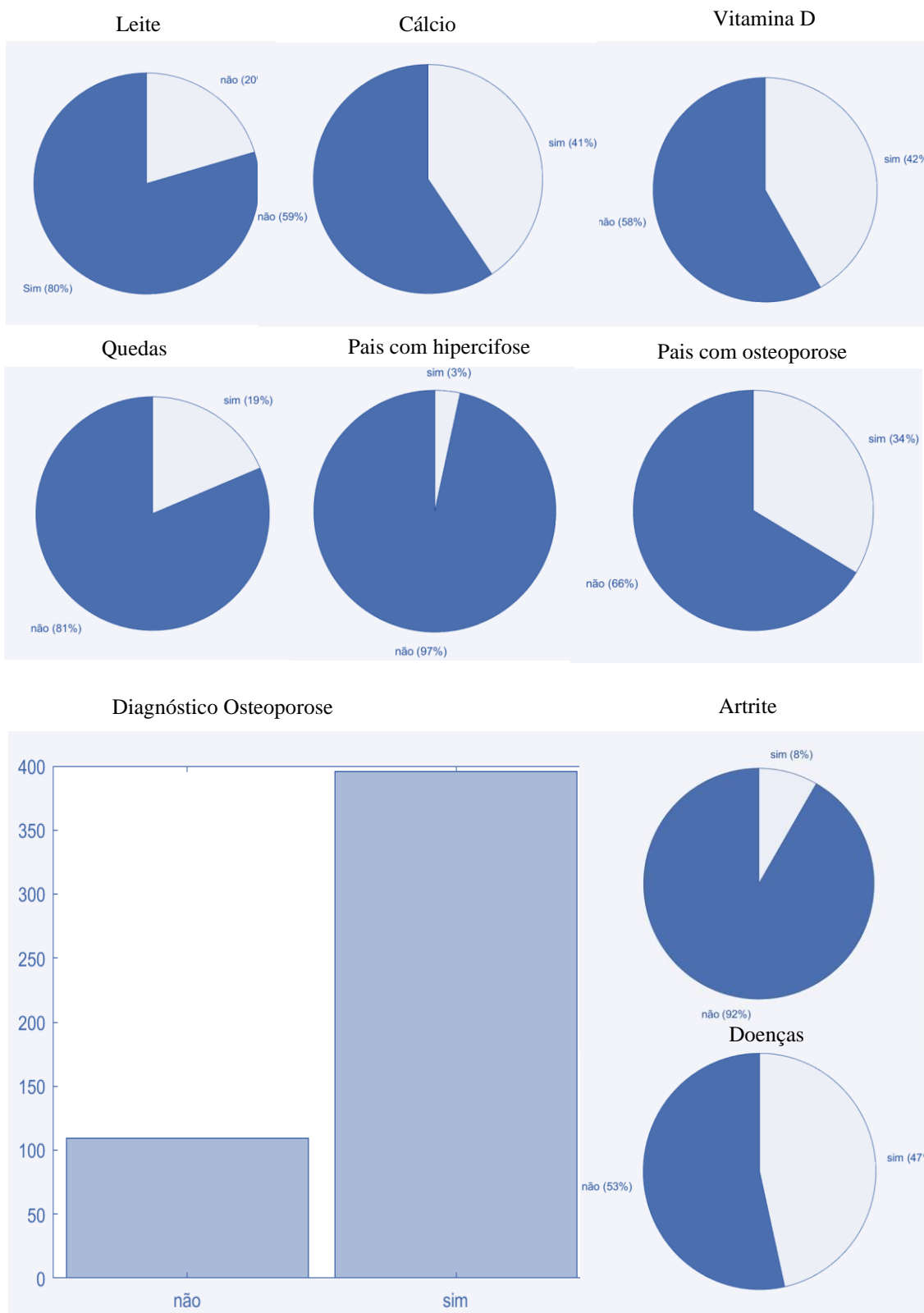


Figure 5. Estatística descritiva geral da população em estudo

5.2 Análise da relevância e redundância das *features*

5.2.1 Relevância

Os resultados do teste de *KW* estão presentes na tabela 2 abaixo, onde podemos observar os valores de *H*. Os testes realizados na integra encontra-se em anexo no código.

<i>FEATURES</i>	<i>H</i>
Idade	81.63
Menopausa	58.00
IMC	20.58
Corticoides	15.00
Vitamin D	13.14
Cálcio	12.78
Diferença	5.81
Quedas	4.10
Genero	3.63
Atividade física	3.50
Tabaco	3.50
Álcool	3.12
Artrite	1.32
Raça	1.42
Testosterona	0.96
Leite	0.75
Pais com hipercifose	0.04
Pais com osteoporose	0
Doenças	0

Tabela 2. *Features* ordenadas segundo *Ranking* dos valores *H* do resultado do teste *KW*

A tabela demonstra que a *feature* com maior relevância para detetar a osteoporose é a idade. Seguidamente vem a menopausa, IMC, corticóides, vitamina D e cálcio.

Features como a diferença, quedas, género, atividade física, tabaco, álcool, artrite, testosterona, leite, pais com hipercifose, apresentam um valor de H menor o que sugere que a correlação destas com o *outcome* osteoporose é menor.

Por último as *features* pais com osteoporose e doenças apresentam um valor de H igual a zero. Zero indica que os dados de amostra correspondem exatamente ao que se esperaria se a hipótese nula estivesse correta. Assim, um zero (0) significa que as duas variáveis são completamente independentes, nenhuma relação existe entre as duas variáveis.

5.2.2 Redundância

O gráfico resultante do teste de Correlação de *Person* está presente abaixo na figura 6, apresentando numa matriz os valores de correlação entre todas as *features*.

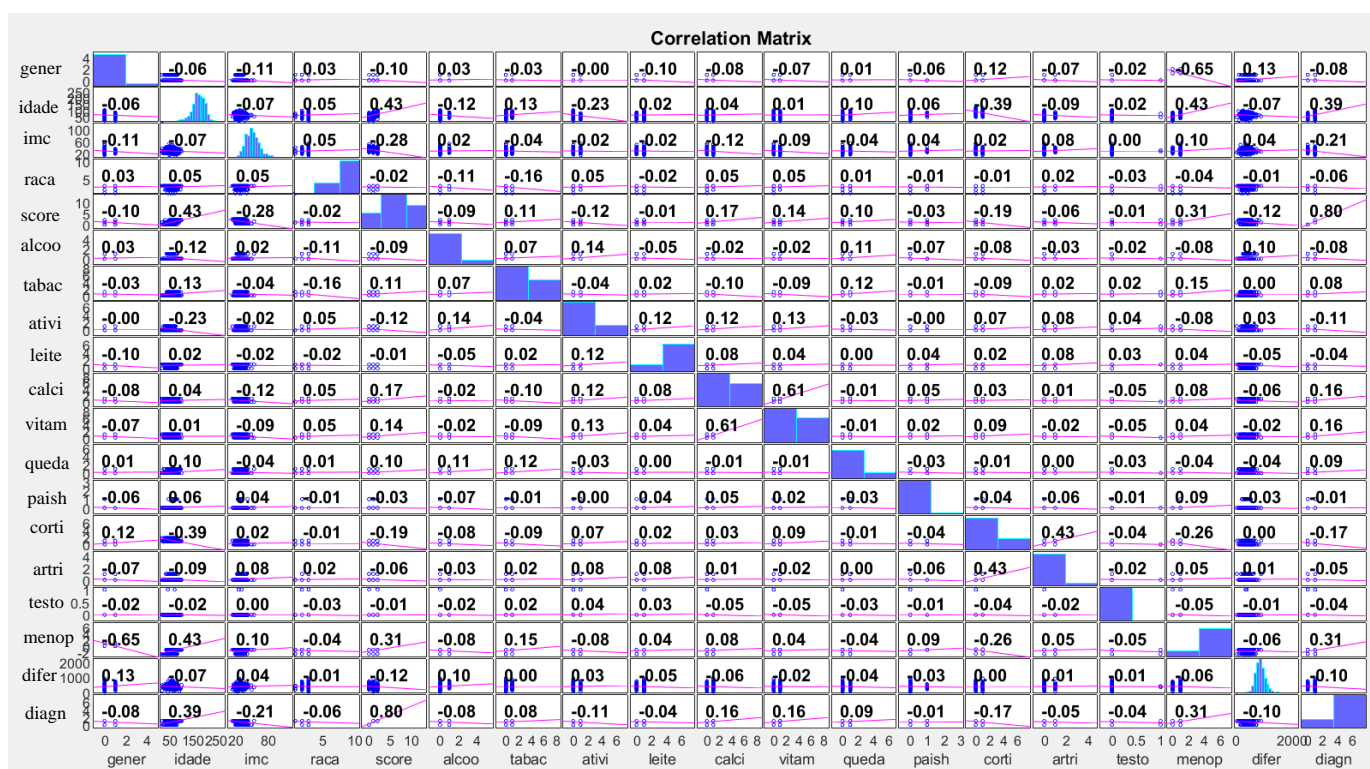


Figure 6. Matriz da correlação de *Pearson*.

A matriz de correlação é simétrica, ou seja, a correlação entre X e Y é a mesma que a correlação entre Y e X.

Interpretando os resultados podemos concluir então que as *features* apresentadas na matriz não são redundantes, pois para isso acontecer teria de existir correlações com um valor superior a 0.90 entre elas, condição que não se verifica perante a figura acima.

5.3 Classificação

Classificador	Média Sensibilidade	Média Especificidade	Média F-measure	Precisão
RLM	0,909± 0,019	0,288±0,074	0,866±0,018	0,823±0,017
SVM	0,922±0,026	0,199±0,075	0,883±0,022	0,836±0,019

Tabela 3. Médias gerais dos classificadores, sensibilidade, especificidade, *F-measure* e precisão

A Tabela 3 apresenta as métricas gerais de desempenho de dois classificadores de ML, o classificador de RLM e o classificador de SVM, com foco em três métricas específicas: Sensibilidade, Especificidade e *F-measure*.

Relativamente à sensibilidade, no contexto desta tabela, o Classificador LR obteve uma média de sensibilidade de aproximadamente 0,90, enquanto o Classificador SVM alcançou uma média de sensibilidade ligeiramente superior de cerca de 0,92. Isso indica que ambos os classificadores têm uma boa capacidade de identificar corretamente as instâncias positivas.

No que diz respeito à especificidade, o Classificador RLM obteve uma média de especificidade de cerca de 0,28, enquanto o Classificador SVM obteve uma média de especificidade de aproximadamente 0,19. Esses valores indicam que ambos os classificadores têm um desempenho mais limitado na identificação correta das instâncias negativas.

A *F-Measure* é uma métrica que combina precisão e sensibilidade, fornecendo uma medida geral do desempenho de um classificador. O Classificador RLM alcançou uma média de *F-Measure* de aproximadamente 0,86, enquanto o Classificador SVM atingiu uma média de *F-Measure* de cerca de 0,88. Isso mostra que ambos os classificadores têm um bom equilíbrio entre precisão e sensibilidade.

Em resumo, com base nas métricas apresentadas na tabela, o Classificador SVM apresenta um desempenho ligeiramente superior em termos de sensibilidade e *F-Measure* em comparação com o Classificador RLM. No entanto, o Classificador RML demonstra um melhor desempenho em termos de especificidade.

As figuras e as tabelas seguintes mostram a variação da performance dos classificadores com o número de *features*, primeiramente para o RLM e por último para SVM.

- RLM Algorithm

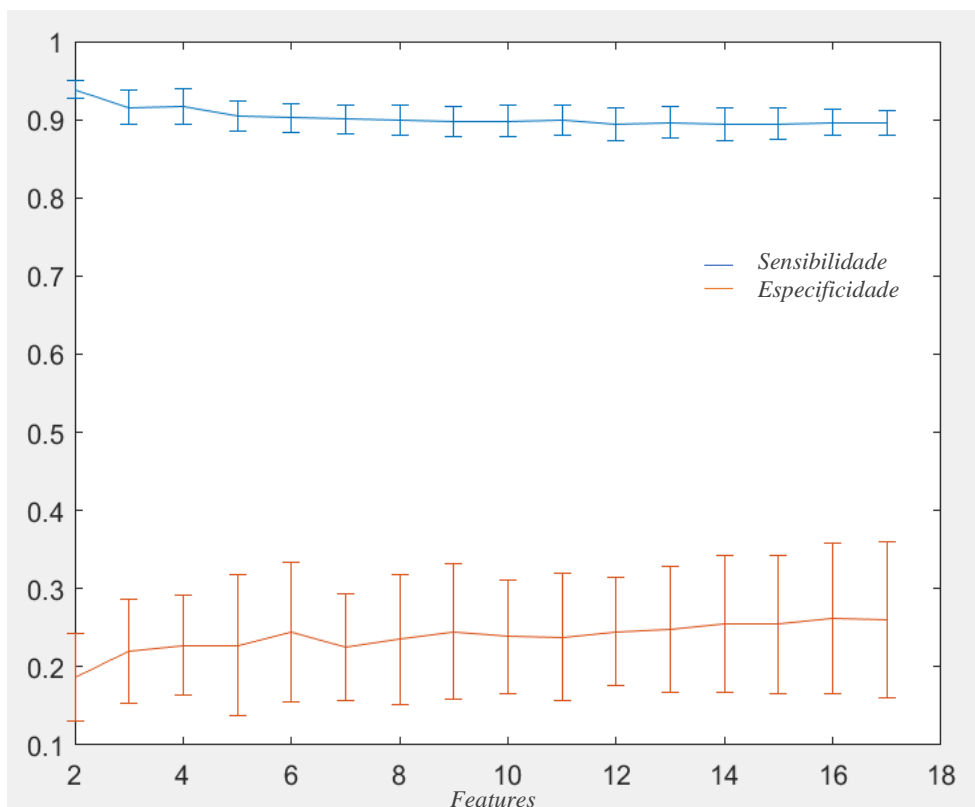


Figura 7. Gráfico classificador RLM

<i>Features</i>	Média Sensibilidade	Média Especificidade	Média F-measure
2	0,938 ±0,011	0,199 ±0,056	0,872 ±0,020
3	0,922 ±0,022	0,220 ±0,067	0,872 ±0,019
4	0,922 ±0,023	0,233 ±0,063	0,871 ±0,019
5	0,904 ±0,019	0,233 ±0,090	0,872 ±0,018
6	0,902 ±0,018	0,244 ±0,089	0,871 ±0,0188
7	0,900 ±0,018	0,233 ±0,069	0,872 ±0,019
8	0,909 ±0,019	0,235 ±0,082	0,872 ±0,019
9	0,909 ±0,019	0,245 ±0,087	0,871 ±0,019
10	0,909 ±0,020	0,244 ±0,073	0,872 ±0,019
11	0,909 ±0,020	0,244 ±0,082	0,872 ±0,019
12	0,894 ±0,021	0,245 ±0,069	0,872 ±0,019
13	0,909 ±0,020	0,255 ±0,080	0,871 ±0,018
14	0,894 ±0,021	0,255 ±0,088	0,871 ±0,018
15	0,895 ±0,020	0,254 ±0,089	0,871 ±0,018
16	0,909 ±0,016	0,262 ±0,096	0,871 ±0,019
17	0,909 ±0,016	0,261 ±0,100	0,871 ±0,018

Tabela 4. Resultados do algoritmo RLM

A tabela 4 e a figura 7 apresenta os resultados de avaliação do RLM para diferentes números de *features* consideradas no modelo. Os valores apresentados são as médias de sensibilidade, especificidade e *F-measure* obtidas em teste.

Como referido anteriormente, a sensibilidade representa a proporção de exemplos positivos corretamente classificados como positivos, enquanto a especificidade representa a proporção de exemplos negativos corretamente classificados como negativos. Ambas as métricas são importantes para avaliar o desempenho geral de um modelo de classificação.

Com base nos resultados deste classificador, a média de sensibilidade para a RLM varia de 0,894 a 0,938, enquanto a média de especificidade varia de 0,199 a 0,262.

Observando os valores da Tabela 4 o modelo com 2 *features* apresenta alta sensibilidade (0.938) e baixa especificidade (0.199).

Conforme mais *features* são adicionadas, a sensibilidade permanece relativamente alta, mas a especificidade também aumenta gradualmente.

É importante encontrar um equilíbrio entre sensibilidade e especificidade, dependendo do contexto em algumas aplicações, pode ser mais crítico minimizar os falsos positivos (aumentar a especificidade) para evitar diagnósticos errados, enquanto noutros casos, a ênfase pode estar em detetar o maior número possível de casos positivos (aumentar a sensibilidade).

Os resultados apresentam valores de *F-measure* a variar de 0,871 a 0,872, indicando um desempenho geral equilibrado em termos de precisão e sensibilidade. Isto significa que os valores de VP, VN, FP, FN aferidos não apresentam grandes distorções.

- SVM

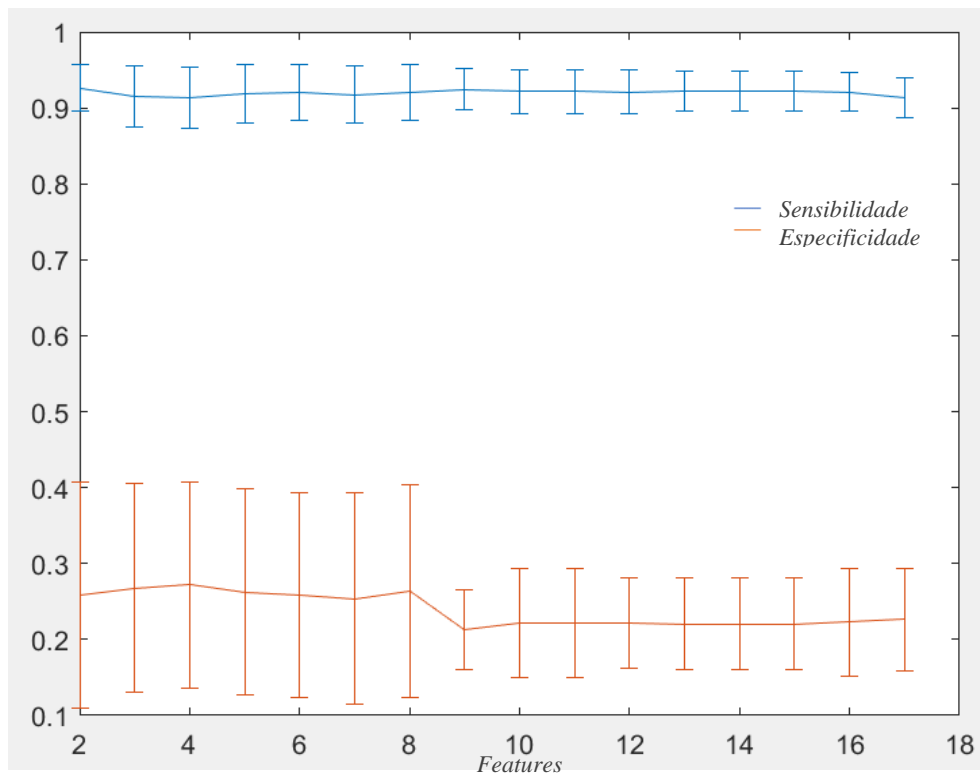


Figura 8. Gráfico classificador SVM

<i>Features</i>	Média Sensibilidade	Média Especificidade	Média F-measure
2	0,977 ±0,044	0,114 ±0,157	0,863 ±0,022
3	0,931 ±0,038	0,211 ±0,125	0,863 ±0,022
4	0,931 ±0,039	0,211 ±0,126	0,863 ±0,022
5	0,922 ±0,032	0,233 ±0,108	0,863 ±0,023
6	0,911 ±0,016	0,244 ±0,072	0,863 ±0,022
7	0,905±0,018	0,255 ±0,072	0,864 ±0,022
8	0,911 ±0,019	0,242 ±0,069	0,863 ±0,022
9	0,922 ±0,027	0,175 ±0,064	0,865 ±0,024
10	0,922 ±0,032	0,188 ±0,072	0,865 ±0,023
11	0,921 ±0,027	0,188 ±0,072	0,865 ±0,023
12	0,913 ±0,024	0,211 ±0,074	0,865 ±0,024
13	0,910 ±0,027	0,211 ±0,077	0,865 ±0,024
14	0,911 ±0,026	0,211 ±0,078	0,865 ±0,024
15	0,912 ±0,024	0,211 ±0,078	0,865 ±0,024
16	0,912 ±0,024	0,222 ±0,078	0,865 ±0,023
17	0,913 ±0,023	0,223 ±0,071	0,865 ±0,024

Tabela 5. Resultados do algoritmo SVM

A Tabela 5 e a figura 8 apresenta os resultados obtidos da avaliação de um modelo preditivo SVM para diferentes números de *features*.

O modelo apresenta valores de sensibilidade bastante elevados em todos os cenários (variando de 0,913 a 0,977), o que sugere que ele é eficaz em detetar casos de risco de osteoporose.

A especificidade varia de 0,114 a 0,255, indicando que o modelo tem uma variação considerável na capacidade de identificar corretamente casos de não risco de osteoporose. Valores mais próximos de 1.0 seriam desejáveis para uma alta especificidade.

Relativamente aos valores de *F-measure*, estes variam entre 0,863 e 0,865, o que sugere um desempenho geral consistente, mas não excepcional.

Em resumo, o modelo SVM parece ser eficaz na deteção de casos de risco de osteoporose, com uma sensibilidade alta, mas há espaço para melhorias na especificidade para evitar falsos positivos. A escolha do número certo de características pode afetar essas métricas, portanto, a seleção cuidadosa de características pode ser importante para otimizar o desempenho do modelo.

Ambos os modelos, RLM e SVM, têm desempenho bastante semelhante, com pequenas diferenças nas métricas de sensibilidade, especificidade e *F-measure*.

A RLM tende a ter uma especificidade ligeiramente mais alta (0,199 a 0,262), enquanto o SVM tem uma sensibilidade média um pouco maior (0,913 a 0,977).

O *F-measure*, que combina precisão e sensibilidade, é bastante consistente em ambos os modelos. A RLM parece ter valores de *F-measure* ligeiramente mais altos em média, indicando um melhor desempenho geral, balanceado em termos de precisão e sensibilidade.

6. Conclusão e Trabalho Futuro

O presente estudo conseguiu investigar e identificar a complexa relação entre a presença de osteoporose e diversos fatores de risco e características individuais. A relevância destas correlações encontradas no desenvolvimento de modelos preditivos da doença foi enfatizada, destacando-se a necessidade de compreender e identificar os elementos que podem influenciar essa condição debilitante.

Comparando os resultados obtidos nos dois classificadores, ambos apresentam um desempenho consistente, independentemente do número de características utilizadas. A escolha entre os dois dependerá das prioridades específicas a considerar. Se for mais importante maximizar a sensibilidade, o SVM pode ser preferível. Se a especificidade é uma prioridade maior, a RLM pode ser uma escolha melhor. Contudo, os dois modelos parecem ser eficazes na detecção de risco de osteoporose, com desempenho geral muito próximo um do outro.

Assim, os modelos construídos revelaram de um modo geral um panorama promissor, embora também tenham revelado desafios a ser abordados. A sensibilidade, que demonstra a capacidade de detetar corretamente a osteoporose, apresentou efetivamente valores encorajadores. No entanto, a especificidade, que denota a habilidade de identificar indivíduos saudáveis com precisão, não atingiu o nível desejado. Isto resulta na identificação errônea de indivíduos saudáveis como portadores da doença.

Esta lacuna sugere a necessidade de afinar os modelos para minimizar falsos positivos e otimizar a identificação precisa dos pacientes.

Verificou-se ainda notavelmente, a combinação de fatores específicos, como idade e menopausa, como indicadores valiosos na predição da osteoporose. Estas duas variáveis juntas demonstraram boa capacidade preditiva para detetar eficazmente a presença da doença.

Esta descoberta ressalta a importância de abordagens multidisciplinares que integram dados clínicos e biomarcadores para desenvolver modelos mais robustos e precisos.

Contudo, apesar de haver aspetos que podem ser aprimorados, é relevante destacar que os resultados obtidos estão em consonância com as conclusões da literatura científica, no que diz respeito à complexidade do número de *features* a serem consideradas num modelo (relação entre a quantidade de *features* incorporadas no modelo e o seu desempenho). Observou-se que adicionar mais *features* ao modelo resultou num desempenho inferior, logo um aumento excessivo no número de *features* não beneficia necessariamente os classificadores e pode comprometer a precisão do modelo, particularmente em contextos de modelação preditiva. Isto realça a importância da seleção criteriosa de *features* relevantes.

É importante sublinhar que este estudo enfrentou limitações, incluindo o tamanho reduzido da base de dados face à prevalência global da osteoporose e a desproporção de género na amostra. Essas limitações podem ter influenciado certos resultados, destacando a necessidade de realizar estudos mais amplos e representativos para validar e expandir as conclusões.

Portanto, no que concerne às perspectivas futuras, há várias direções que podem ser exploradas. Sugere-se uma expansão substancial da amostra, considerando uma distribuição equilibrada de géneros, a fim de obter resultados mais generalizáveis. Além disso, investigações adicionais devem explorar e integrar na abordagem diferentes métodos de extração de *features* e a aplicação de outros classificadores, visto que a variedade de opções é vasta.

Essa abordagem diversificada pode ser promissora para o diagnóstico e permitirá uma avaliação mais completa e aprimorada da predição da doença.

Concluindo, o aperfeiçoamento contínuo de ferramentas baseadas em IA, como aquela proposta neste estudo, é crucial para futuros estudos de forma a avançar no diagnóstico e prevenção da osteoporose. A evolução constante dessas soluções contribuirá de forma significativa para o combate eficaz desta condição, melhorando a qualidade de vida dos pacientes e abrindo caminho para uma abordagem mais personalizada e preventiva da saúde óssea e conseqüentemente dos custos associados à mesma.

Referências

- [1] Sözen, T., Özışık, L., Başaran, NÇ. (2017). An overview and management of osteoporosis. *Eur J Rheumatol.* 4(1):46-56. doi: 10.5152/eurjrheum.2016.048. Epub 2016 Dec 30. PMID: 28293453; PMCID: PMC5335887
- [2] Consensus development conference: Diagnosis, prophylaxis, and treatment of osteoporosis. *Am J Med.* 1993; 94:646–50.
- [3] Porter, J.L., Varacallo, M. (2022) Osteoporosis. 4. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2022 Jan–. PMID: 28722930.)
- [4] Salari, N., Ghasemi, H., Mohammadi, L., Behzadi, M.H., Rabieenia, E., Shohaimi, S., Mohammadi, M. (2021) The global prevalence of osteoporosis in the world: a comprehensive systematic review and meta-analysis. *J Orthop Surg Res.* 17;16(1):609. doi: 10.1186/s13018-021-02772-0. PMID: 34657598; PMCID: PMC8522202.)
- [5] Kanis, J. A., McCloskey, E. V., Johansson, H., Oden, A., Melton, L. J., 3rd, & Khaltsev, N. (2008). A reference standard for the description of osteoporosis. *Bone*, 42(3), 467–475. <https://doi.org/10.1016/j.bone.2007.11.001>
- [6] Lane, N.E. (2006). Epidemiology, etiology, and diagnosis of osteoporosis. *Am J Obstet Gynecol*, 194(2 Suppl): S3-11. doi: 10.1016/j.ajog.2005.08.047. PMID: 16448873.
- [7] Stolnicki, B., & Teixeira, B. C. (2020). The Impact of Hip Fractures in the Public Health System in Brazil (SUS) 2008 - 2017: The Orthopedist Task. *Revista brasileira de ortopedia*, 57(4), 552–559. <https://doi.org/10.1055/s-0040-1713762>
- [8] Marques, A., Lourenço, Ó., da Silva, J.A.P. et al. (2015). The burden of osteoporotic hip fractures in Portugal: costs, health related quality of life and mortality. *Osteoporos Int* 26, 2623–2630. <https://doi.org/10.1007/s00198-015-3171-5>

- [9] Tu, K. N., Lie, J. D., Wan, C. K. V., Cameron, M., Austel, A. G., Nguyen, J. K., Van, K., & Hyun, D. (2018). Osteoporosis: A Review of Treatment Options. *P & T: a peer-reviewed journal for formulary management*, 43(2), 92–104.
- [10] Bonura, F. (2009). Prevention, screening, and management of osteoporosis: an overview of the current strategies. *Postgrad Med. Jul*;121(4):5-17. doi: 10.3810/pgm.2009.07.2021. PMID: 19641263.
- [11] Shim, J.G., Kim, D.W., Ryu, K.H., Cho, E.A., Ahn, J.H., Kim, J.I., Lee, S.H. (2020). Application of machine learning approaches for osteoporosis risk prediction in postmenopausal women. *Arch Osteoporos.* 23;15(1):169. doi: 10.1007/s11657-020-00802-8. PMID: 33097976.
- [12] Ou Yang, W.Y., Lai, C.C., Tsou, M.T., Hwang, L.C. (2021). Development of Machine Learning Models for Prediction of Osteoporosis from Clinical Health Examination Data. *Int J Environ Res Public Health.* 18;18(14):7635. doi: 10.3390/ijerph18147635. PMID: 34300086; PMCID: PMC8305021.
- [13] Levine, J.P. (2022). Identification, diagnosis, and prevention of osteoporosis. *Am J Manag Care.* 17 Suppl 6: S170-6. PMID: 21761956.
- [14] Fasihi, L., Tartibian, B., Eslami, R. et al. (2022). Artificial intelligence used to diagnose osteoporosis from risk factors in clinical data and proposing sports protocols. *Sci Rep* 12, 18330. <https://doi.org/10.1038/s41598-022-23184-y>
- [15] Habebh, H., Gohel, S. (2021). Machine Learning in Healthcare. *Curr Genomics.* 16;22(4):291-300. doi: 10.2174/1389202922666210705124359. PMID: 35273459; PMCID: PMC8822225.
- [16] Martini, F., M. J. Timmmons & R. B. Tallitsch (2009), ‘Anatomia humana’.
- [17] Cruz, A. S., S. G. da Silva & B. H. de Castro (2013), Bone density measurement through electromagnetic waves, em ‘The 6th 2013 Biomedical Engineering International Conference’, pp. 1–5.

- [18] García-Bustínduy, M., Gantes, M.A. (2007) Corticoides e osteoporose [Corticosteróides e osteoporose]. *Actas Dermosifiliogr*, 98(8):526-30. Espanhol. doi: 10.1016/s0001-7310(07)70127-8. PMID: 17919426.
- [19] Pouresmaeili, F., Kamalidehghan, B., Kamarehei, M., Goh, Y.M. (2018). A comprehensive overview on osteoporosis and its risk factors. *Ther Clin Risk Manag*, 6; 14:2029-2049. doi: 10.2147/TCRM.S138000. PMID: 30464484; PMCID: PMC6225907.
- [20] Glaser, D. L., & Kaplan, F. S. (1997). Osteoporosis. Definition and clinical presentation. *Spine*, 22(24 Suppl), 12S–16S. <https://doi.org/10.1097/00007632-199712151-00003>
- [21] Pouresmaeili, F., Kamalidehghan, B., Kamarehei, M., Goh, Y.M. (2018). Uma visão abrangente sobre a osteoporose e seus fatores de risco. *Ther Clin Risk Manag*, 14:2029-2049. doi: 10.2147/TCRM.S138000. PMID: 30464484; PMCID: PMC6225907.
- [22] Ettinger, M.P. (2003). Aging bone and osteoporosis: strategies for preventing fractures in the elderly. *Arch Intern Med*, 13;163(18):2237-46. doi: 10.1001/archinte.163.18.2237. PMID: 14557222.
- [23] Seymour, D.G., & Stone, M. (1994). Perspectivas para o tratamento da osteoporose em idosos. *J R Soc Med*, 87 Supl 23(Suppl 23):26-30. PMID: 8064770; PMCID: PMC1294177.
- [24] Clarke, B.L., & Khosla, S. Fisiologia da perda óssea. *Radiol Clin North Am*. 2010 maio;48(3):483-95. doi: 10.1016/j.rcl.2010.02.014. PMID: 20609887; PMCID: PMC2901252.
- [25] Aspray, T.J., & Hill, T.R. Osteoporose e Envelhecimento do Esqueleto. *Bioquímica subcelular*. 2019; 91:453-476. doi: 10.1007/978-981-13-3681-2_16. PMID: 30888662
- [26] Fang, H., Deng, Z., Liu, J., Chen, S., Deng, Z., Li, W. (2022). O mecanismo de remodelação óssea após o envelhecimento ósseo. *Clin Interv Envelhecimento*,

17:405-415. doi: 10.2147/CIA.S349604. PMID: 35411139; PMCID: PMC8994557.

- [27] Coughlan, T., & Dockery, F. (2014). Osteoporose e risco de fratura em pessoas idosas. *Clin Med (Londres)*, 14(2):187-91. doi: 10.7861/clinmedicine.14-2-187. PMID: 24715132; PMCID: PMC4953292.
- [28] Hill, T.R., & Aspray, T.J. (2017). O papel da vitamina D na manutenção da saúde óssea em idosos. *Ther Adv Musculoskelet Dis*, 9(4):89-95. doi: 10.1177/1759720X17692502. Epub 2017 14 de fevereiro. PMID: 28382112; PMCID: PMC5367643
- [29] Ji, M.X., & Yu, Q. (2015). Osteoporose primária em mulheres na pós-menopausa. *Doença Crônica Transl Med*, 1(1):9-13. doi: 10.1016/j.cdtm.2015.02.006. PMID: 29062981; PMCID: PMC5643776.
- [30] Pista, N.E. (2006). Epidemiologia, etiologia e diagnóstico da osteoporose. *Am J Obstet Gynecol*, 194(2 Supl):S3-11. doi: 10.1016/j.ajog.2005.08.047. PMID: 16448873.
- [31] Panday, K., Gona, A., Humphrey, M.B. (2014). Medication-induced osteoporosis: screening and treatment strategies. *Ther Adv Musculoskelet Dis*, 6(5):185-202. doi: 10.1177/1759720X14546350. PMID: 25342997; PMCID: PMC4206646.)
- [32] Henneicke, H., Gasparini, S., Brennan-Speranza, T., Zhou, H., Seibel, M. (2014). Glucocorticoids and bone: local effects and systemic implications. *Trends Endocrinol Metab* 25: 197–211
- [33] Suheda, Ozcakil., Deniz, Sigirli., Ahmet, Ursavas., Esra, Uzaslan. (2020). COPD and Osteoporosis: Associated Factors in Patients Treated with Inhaled Corticosteroids, *International Journal of Chronic Obstructive Pulmonary Disease*, 15: 2441-2448, DOI: 10.2147/COPD.S274728)
- [34] Zhang, X., Yu, Z., Yu, M., Qu, X. (2015). Consumo de álcool e risco de fratura de quadril. *Osteoporos Int*, 26(2):531-42. doi: 10.1007/s00198-014-2879-y. Epub 2014 30 de setembro. PMID: 25266483.

- [35] Senderovich, H., & Kosmopoulos, A. (2018). Uma visão sobre o efeito dos exercícios na prevenção da osteoporose e fraturas associadas em indivíduos de alto risco. *Rambam Maimonides Med J*, 29;9(1): e0005. doi: 10.5041/RMMJ.10325. PMID: 29406844; PMCID: PMC5796736
- [36] Todd, J.A., & Robinson, R.J. (2003). Osteoporose e exercício. *Postgraduate Medical Journal*, 79:320-323
- [37] Dimai H. P. (2023). New Horizons: Artificial Intelligence Tools for Managing Osteoporosis. *The Journal of clinical endocrinology and metabolism*, 108(4), 775–783. <https://doi.org/10.1210/clinem/dgac702>
- [38] Jones, L.D., Golan, D., Hanna, S.A., Ramachandran, M. (2018). Artificial intelligence, machine learning and the evolution of healthcare: A bright future or cause for concern? *Bone Joint Res*, 5;7(3):223-225. doi: 10.1302/2046-3758.73.BJR-2017-0147.R1. PMID: 29922439; PMCID: PMC5987686
- [39] Jiang, L., Wu, Z., Xu, X., et al. (2021). Opportunities and challenges of artificial intelligence in the medical field: current application, emerging problems, and problem-solving strategies. *Journal of International Medical Research*, 49(3). doi:10.1177/03000605211000157
- [40] Benke, K., & Benke, G. (2018). Artificial Intelligence and Big Data in Public Health. *Int J Environ Res Public Health*, 10;15(12):2796. doi: 10.3390/ijerph15122796. PMID: 30544648; PMCID: PMC6313588.
- [41] Pepe, M.S. (2005). Evaluating technologies for classification and prediction in medicine. *Stat Med*, 30;24(24):3687-96. doi: 10.1002/sim.2431. Erratum in: *Stat Med*. 2006 Jul 15;25(13):2340. PMID: 16320261.
- [42] Silva, K. L. (2003). Avaliação tecnológica em saúde: densitometria óssea e terapêuticas, *Cad. Saúde Pública*, 19, 987–1003
- [43] Della Martina, M., Fabiani, G., Biasioli, A., Piccoli, V., Vascotto, L., Driul, L., Chittaro, M., & Marchesoni, D. (2008). Identification of patients with a high risk

of osteoporosis. Analysis of risk factors, ultrasonography and DEXA, in a female population in the North East of Italy. *Minerva ginecologica*, 60(2), 97–104

- [44] Balanis, C. (2009). *Teoria de Antenas - Análise e Síntese*, 3a edição, LTC, Rio de Janeiro, RJ
- [45] Hwang, P., A. Chen, C. Lou & J. Lin. (2014). Electromagnetic shielding effectiveness and functions of stainless steel/bamboo charcoal conductive fabrics, *Journal of Industrial Textiles*, 44(3), 477–494
- [46] Emerson, C. (2011). *ECCOSORB AN - Flexible foam sheet broadband microwave absorbers*, Randolph, MA.
- [47] Albuquerque, G., Cruz, A., Carvalho, D. et al. (2022). A method based on non-ionizing microwave radiation for ancillary diagnosis of osteoporosis: a pilot study. *BioMed Eng OnLine* 21, 70. <https://doi.org/10.1186/s12938-022-01038-y>
- [48] Ongphiphadhanakul, B., Rajatanavin, R., Chailurkit, L., Piaseu, N., Teerarungsikul, K., Sirisriro, R., Komindr, S., Pauvilai, G. (1997). Prediction of low bone mineral density in postmenopausal women by artificial neural network model compared to logistic regression mode. *J Med Assoc Thai*, 80:508–15.
- [49] Gregory, J.S., Junold, R.M., Undrill, P.E., Aspen, R.M. (1999). Analysis of trabecular bone structure using fourier transforms and neural networks. *IEEE Trans Inf Technol Biomed*, 3:289–94
- [50] Oh, L.K., Sedrine, W.B., Torralba, T.P., Kung, A., Fujiwara, S., Chan, S.P., Huang, Q.R., Rajatanavin, R., Tsai, K.S., Park, H.M., Reginster, J.Y. (2001). A simple tool to identify asian women at increased risk of osteoporosis. *Osteoporos Int*, 12:699–705
- [51] Rizzi, A., Panella, M., Paschero, M., Mascioli, F.F. (2004) Estimation of bone mineral density data using mog neural networks. In: *IEEE International Joint Conference on Neural Networks*, 4:3241–3246.
- [52] Wang, W., Richards, G., Rea, S. (2005). Hybrid data mining ensemble for predicting osteoporosis risk. *Annu Int Conf*, 1:886–9

- [53] Sadatsafavi, M., Moayyeri, A., Soltani, A., Larijani, B., Nouraei, M., Akhondzadeh, S. (2005). Artificial neural networks in prediction of bone density among post-menopausal women. *J Endocrinol Invest*, 28(7):425–31
- [54] Chiu, J.S., Li, Y.C., Yu, F.C., Wang, Y-F. (2006). Applying an artificial neural network to predict osteoporosis in the elderly. *Stud Health Technol Inform*, 124:609–14.
- [55] Lemineur, G., Harba, R., Kilic, N., Ucan, O.N., Osman, O., Benhamou, L. (2007). Efficient estimation of osteoporosis using artificial neural networks. In: *Industrial Electronics Society, IECON 2007. 33rd Annual Conference of the IEEE 2007. 2007; p. 3039–44.*
- [56] Leslie, W.D., Lix, L.M., Yogendran, M.S. (2011). Validation of a case definition for osteoporosis disease surveillance. *Osteoporos Int*, 22(1):37–46
- [57] Mantzaris, D., Anastassopoulos, G., Iliadis, L., Kazakos, K., Papadopoulos, H. (2010). Medical informatics and biomedical engineering. In: Papadopoulos H, Andreou AS, Brammer M, editors. *AIAI: IFIP international conference on artificial intelligence applications and innovations. Artificial intelligence applications and innovations. Proceedings of the 6th IFIP WG 12.5 international conference, AIAI 2010, Larnaca, Cyprus, 6–7 October 2010. Berlin: Springer; 2010. p. 120–127]*
- [58] Juez, F.J., Suárez, M.Á., Lasheras, F.S., & Murcia-Mazón, A. (2011). Application of neural networks to the study of the influence of diet and lifestyle on the value of bone mineral density in post-menopausal women. *Math. Comput. Model.*, 54, 1665-1670.
- [59] Harrar, K., Hamami, L., Akkoul, S., Lespessailles, E., Jennane, R. (2012). Osteoporosis assessment using multilayer perceptron neural networks. In: *2012 3rd International Conference on Image Processing Theory, Tools and Applications (IPTA)*. p. 217–21
- [60] Yoo, T.K., Kim, S.K., Kim, D.W., Choi, J.Y., Lee, W.H., Park, E.C. (2013). Osteoporosis risk prediction for bone mineral density assessment of postmenopausal women using machine learning. *Yonsei Med J*, 54:1321–30.

- [61] Tafraouti, A., Hassouni, M.E., Toumi, H., Lespessailles, E., Jennane, R. (2014). Osteoporosis diagnosis using fractal analysis and support vector machine. In: 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems. p. 73–7
- [62] Iliou, T., Anagnostopoulos, C.E., Stephanakis, I.M., & Anastassopoulos, G.C. (2017). A novel data preprocessing method for boosting neural network performance: A case study in osteoporosis prediction. *Inf. Sci.*, 380, 92-100.
- [63] Liu, Q., Cui, X., Chou, Y.C., Abbod, M.F., Lin, J., Shieh, J.S. (2015). Ensemble artificial neural networks applied to predict the key risk factors of hip bone fracture for elders. *Biomed Signal Process Control*, 21(4):146–56
- [64] Yu, X., Ye, C., Xiang, L. (2016). Application of artificial neural network in the diagnostic system of osteoporosis. *Neurocomputing*, 214(C):376–81.
- [65] Mann, H.B., & Whitney, D.R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18, 50-60.
- [66] Mukaka M. M. (2012). Statistics corner: A guide to appropriate use of correlation coefficient in medical research. *Malawi medical journal: the journal of Medical Association of Malawi*, 24(3), 69–71.
- [67] Tripepi, G., Jager, K. J., Dekker, F. W., & Zoccali, C. (2008). Linear and logistic regression analysis. *Kidney international*, 73(7), 806–810. <https://doi.org/10.1038/sj.ki.5002787>
- [68] Ma, Y., Lu, Q., Yuan, F., & Chen, H. (2023). Comparison of the effectiveness of different machine learning algorithms in predicting new fractures after PKP for osteoporotic vertebral compression fractures. *Journal of orthopaedic surgery and research*, 18(1), 62. <https://doi.org/10.1186/s13018-023-03551-9>
- [69] Payal, Manju., Ananth Kumar, T., Ajagbe, Sunday. (2022). Support Vector Machines (SVMS) Based Advanced Healthcare System Using Machine Learning Techniques. 10. 3007. 10.15680/IJIRCCE.2022.1005020|.

[70] Osisanwo, F.Y., Akinsola, J.E.T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., Akinjobi, J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison". International Journal of Computer Trends and Technology (IJCTT), V48(3):128-138. ISSN:2231-2803. www.ijcttjournal.org. Published by Seventh Sense Research Group

Anexos

BASE DE DADOS

Projeto *OSSEUS*

CÓDIGO

Link: https://github.com/DanielaGuia/IA_Osteoporose_Osseus.git

- Teste de KW
- Teste correlação de Pearson
- Classificador RL
- Classificador SVM

