

Estudo de técnicas de aprendizado de máquina para sistema de apoio à decisão do tratamento de câncer de próstata

Diogo Pereira Almeida¹, João Victor O. Moreira¹, João Victor R. Tagliarini¹

¹Engenharia da Computação – Faculdade de Engenharia de Sorocaba (FACENS)
Sorocaba, SP – Brazil

{diogo.meida, joaovictorosa.tagliarini}@gmail.com,
joao.oliver2012@hotmail.com

Abstract. Prostate cancer is the second most common type in Brazil, with forecasts indicating growth in the coming years. This project aims to develop a machine learning model as a decision support tool for healthcare professionals, suggesting treatments based on clinical data. The public database from the Cancer Hospital Registry of the Fundação Oncocentro de São Paulo was used, containing information about patients and tumors. After data processing and analysis, the algorithms K-Nearest Neighbors, Logistic Regression, Decision Tree, Random Forest, Naive Bayes and XGBoost were trained and evaluated. Accuracy, precision, recall, and F1-score metrics were used to compare the models, with XGBoost achieving the best performance, reaching an accuracy of 60%.

Resumo. O câncer de próstata é o segundo mais frequente no Brasil, com previsão de crescimento nos próximos anos. Este projeto visa desenvolver um modelo de aprendizado de máquina como ferramenta de apoio à decisão médica, sugerindo tratamentos a partir de dados clínicos. Utilizou-se o banco de dados público do Registro Hospitalar de Câncer da Fundação Oncocentro de São Paulo, com informações sobre os pacientes e tumores. Após o processamento e análise dos dados, foram treinados e avaliados os algoritmos K-Vizinhos mais próximos, Regressão Logística, Árvore de Decisão, Naive Bayes, Floresta Aleatória e XGBoost. As métricas de acurácia, precision, recall e F1-score foram utilizadas na comparação dos modelos, com o melhor sendo o XGBoost, com acurácia de 60%.

1. Introdução

No Brasil, o câncer de próstata não só é o segundo mais frequente entre os homens, superado apenas pelo câncer de pele não melanoma, mas também o segundo mais comum avaliando os dois sexos, conforme o Ministério da Saúde [Brasil 2020]. Entre os sintomas, estão o sangue na urina, aumento do tempo para iniciar ou terminar de urinar, perda de força no jato e aumento da frequência urinária.

Dessa maneira, diante de suspeitas, o médico especialista pode solicitar exames como o de toque retal e o de concentração do antígeno específico da próstata (PSA) no sangue. Se houver alterações, uma biópsia confirmará o diagnóstico. Confirmado o tumor, há uma gama de tratamentos, por exemplo a cirurgia, radioterapia, quimioterapia e hormonioterapia, que podem ser combinadas ou aplicadas individualmente [Brasil 2020].

Nesse sentido, segundo James (2024), as ocorrências de câncer de próstata vão aumentar de 1,4 milhão por ano em 2020, para 2,9 milhões até 2040, por conta da elevação

da expectativa de vida da população global. Além disso, nos 20 anos, os óbitos passarão de 375 mil por ano para 700 mil, um acréscimo de mais de 80%, principalmente nas nações de média e baixa renda. Já no país, as estimativas do triênio 2023-2025 são de 71.730 registros por ano, como apontou a publicação de Brasil (2022). Com dados consolidados, registrou-se 16.292 mortes pela doença em 2022, com média de 44 óbitos por dia.

Nesse panorama, a tecnologia vem sendo estudada para auxiliar no diagnóstico e tratamento dos quadros de câncer de próstata, principalmente com inteligência artificial. Por exemplo, o estudo da Universidade de Oxford e da Universidade de Manchester que utilizou Redes Neurais para analisar dados genéticos de milhares de amostras de tumores, em nove países. A pesquisa descobriu a existência de dois tipos diferentes de evoluções do câncer, o que pode mudar o entendimento das classificações de diagnóstico e planos de tratamento [Woodcock 2024].

Além disso, na área de aprendizado de máquina, há artigos acadêmicos focados em analisar os dados do Instituto Nacional de Câncer dos Estados Unidos, para indicar a intervenção cirúrgica ou não, dado as características clínicas do 3 paciente [Tang et al. 2024; Pan et al. 2024]. No âmbito nacional, Araújo (2024) também estudou a possibilidade de utilização de modelos na predição de tumores malignos, no Hospital Universitário da Universidade Federal do Maranhão.

2. Objetivo geral

Em vista do cenário crescente de casos, o objetivo geral é construir um modelo de aprendizado de máquina, baseado em dados reais, para sugerir o tratamento para um determinado diagnóstico, como uma ferramenta de apoio à decisão do profissional da saúde. Sendo assim, o estudo justifica-se pela possibilidade de contribuir com a melhoria da qualidade de vida dos homens acometidos pela doença, principalmente no panorama do tratamento excessivo, o qual pode gerar estresse e ansiedade desnecessários, incontinência urinária e impotência sexual como consequência dos resultados dos procedimentos [Brasil 2020].

3. Objetivos específicos e contribuições

Como objetivos específicos, listam-se os seguintes:

- Analisar as informações do Banco de Dados RHC (Registro Hospitalar de Câncer) público da Fundação Oncocentro de São Paulo;
- Pré-processar as colunas, mantendo as mais relevantes e lidando com valores nulos e inconsistentes;
- Treinar os modelos de aprendizado de máquina K-Vizinhos mais próximos, Regressão Logística, Árvore de Decisão, Floresta Aleatória, *Naive Bayes* e *XGBoost*;
- Obter as métricas de acurácia, *precision*, *recall* e *F1-score* e realizar uma análise crítica comparativa.

Como contribuições, espera-se com o projeto auxiliar os profissionais no processo de decisão do tratamento, analisando uma grande quantidade de informações complexas que podem trazer novos *insights* sobre o tema.

4. Fundamentação teórica

O presente tópico aborda os conceitos médicos necessários ao entendimento do trabalho, apresentando o que é o câncer de maneira geral e mais especificamente sobre o câncer de próstata, sintomas, diagnósticos e tratamentos. Além disso, traça um panorama da inteligência artificial e aprendizado de máquina na área da saúde, bem como explicará os modelos e métricas utilizadas no trabalho.

4.1. Câncer

Uma célula pode ser descrita como a menor unidade viva do corpo humano. Essas estruturas fazem parte do princípio do mecanismo biológico, sendo responsáveis por funções como gerar energia para o corpo e manter o funcionamento de suas atividades. Para isso, apresentam especializações destinadas a cada tipo de tecido do organismo: epitelial, conjuntivo, nervoso e muscular [Linhares 2023].

Para que o ciclo se mantenha de forma sustentável, a maior parte das células com capacidade de renovação precisa passar pela etapa de divisão celular, constituindo, assim, fases de início, meio e fim do tempo de vida. Nesse sentido, existem dois tipos de divisão celular: a meiose e a mitose. A meiose é responsável pela divisão das células sexuais do corpo humano, enquanto a mitose é o processo pelo qual o organismo multiplica o outro tipo de células e faz a renovação dos tecidos. Um câncer, por definição, se origina de células somáticas, cuja divisão ocorre por mitose, já que a causa de um tumor é a reprodução descontrolada de uma célula que permanece em constante multiplicação no organismo [American Cancer Society 2025].

Dessa forma, entende-se que o câncer é um agente que surge quando alguma parte desse processo de divisão celular sofre uma mutação, fazendo com que células defeituosas, que deveriam ter seu ciclo encerrado ou se autodestruir, continuem se multiplicando indiscriminadamente. Com isso em vista, a mitose pode ser separada em um estágio inicial e quatro fases de desenvolvimento: interfase, prófase, metáfase, anáfase e telófase [Openstack College 2016].

A atenção deve ser voltada para a interfase, pois é nesse período que se encontram os principais causadores de mutações capazes de gerar um câncer. A interfase é dividida em três etapas: G1, S e G2. G1 é o momento em que ocorre a síntese de proteínas, fazendo com que a célula aumente de tamanho; S é a fase em que o DNA é duplicado; e G2 é a etapa de verificação, que checa se o DNA foi copiado corretamente. Na maioria dos casos, o câncer pode surgir devido a falhas nos mecanismos de controle do ciclo celular durante essas três fases [Openstack College 2016].

Considerando o processo de divisão celular, é importante saber que existem dois grandes grupos de câncer: o que ocorre em células sanguíneas (hematológicas) e o que ocorre em qualquer outro tecido do organismo, sendo nomeado como câncer de tumor sólido [American Cancer Society 2025].

Um nódulo é um agregado de células e tecidos que se formam em um determinado local. Por definição, todo tumor será um nódulo, mas um nódulo nem sempre será um tumor. Entende-se que um tumor pode ser classificado como benigno ou maligno, a depender de suas características. Em um tumor benigno, as células não invadem outros tecidos além daquele no qual surgiram, e o acúmulo de massa naquela região não afeta o bom funcionamento das demais células. Já em um tumor maligno, as células cancerígenas

podem se espalhar para diferentes tecidos e causar disfunções no organismo [Araújo 2024].

Embora, o câncer surja em determinado momento da divisão celular, os motivos pelos quais essa mutação ocorre podem advir de vários fatores - sejam eles genéticos, relacionados a hábitos, estilo de vida, ou outros. Em geral, é difícil apontar com precisão qual desses fatores foi o principal responsável pelo acontecimento.

4.2. Câncer de próstata

A próstata, órgão existente apenas em pessoas do sexo masculino, tem a responsabilidade de produzir o líquido que nutre e protege os espermatozoides no sêmen. Ela pode ser encontrada logo abaixo da bexiga e na frente do reto (parte final do intestino) e envolve a parte inicial da uretra, que é o canal por onde a urina e o sêmen passam para sair do corpo, como mostra a Figura 1 abaixo. Importante ressaltar que a próstata em um homem jovem possui um formato ovalado medindo aproximadamente três centímetros na maior dimensão, entretanto, esse valor tenderá a crescer conforme o envelhecimento [Oncoguia 2024].

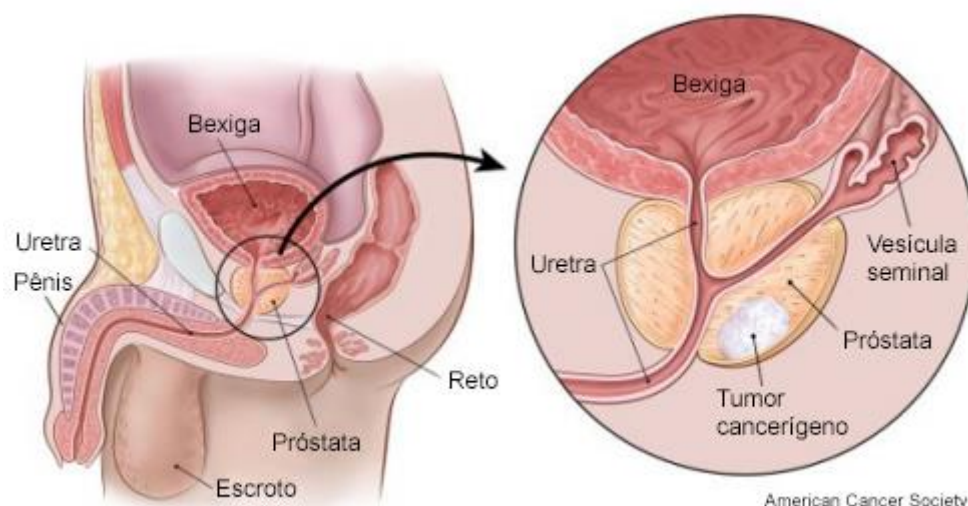


Figura 1. Localização da próstata

Como mencionado, o câncer de próstata é originado de células que começam a se dividir sem controle. O principal tipo de câncer que pode surgir dessa disfunção é o adenocarcinoma, embora outros tipos também possam ocorrer, como: tumores neuroendócrinos, carcinomas de células transicionais, carcinoma de pequenas células e sarcomas [Araújo 2024].

Um ponto a se destacar sobre a próstata é que muitos homens, principalmente com o avanço da idade, podem desenvolver alguma forma de câncer prostático. A questão está em compreender se isso os afetará clinicamente ao longo da vida, pois pode-se viver sem nunca descobrir que possui alguma anomalia na próstata. Isso ocorre porque o desenvolvimento foi tão lento que levaria muitos anos até se tornar preocupante. Casos assim são frequentemente identificados em autópsias de indivíduos que faleceram sem jamais saber da doença [American Cancer Society 2023].

4.3. Sintomas

Portanto, é muito importante que os homens reconheçam os sintomas, os quais devem ser observados ao longo da vida para identificar possíveis problemas no funcionamento do corpo. Sinais que podem indicar complicações tumorais incluem: dificuldades para urinar, presença de sangue na urina ou no sêmen, e disfunção erétil. Quanto às dores, estas podem ocorrer nas costas, quadris, tórax ou nos ossos, resultado da multiplicação do câncer pelo organismo. Uma grande dificuldade perceptível é que muitos desses sintomas também são comuns a diversas outras condições médicas, o que dificulta para o paciente reconhecer que esses sinais possam estar relacionados a um quadro clínico mais sério [American Cancer Society 2025].

4.4. Exame de toque retal

Conforme o Oncoguia (2024), o exame de toque retal é um procedimento clínico simples e de baixo custo, utilizado como método auxiliar no rastreamento e na investigação de possíveis alterações na próstata. Ele consiste na introdução do dedo indicador, protegido por luva e lubrificado, no reto do paciente, permitindo ao médico avaliar o tamanho, a consistência e a presença de irregularidades na superfície da glândula prostática. Apesar de não ser capaz de detectar todos os casos de câncer de próstata, o exame possibilita identificar nódulos ou áreas endurecidas que podem sugerir a presença de um tumor, principalmente quando a doença já se encontra em estágios mais avançados.

Embora a sensibilidade seja limitada, especialmente em fases iniciais, o toque retal continua sendo uma ferramenta relevante, pois permite a detecção de alterações que podem não ser percebidas por exames laboratoriais, como a dosagem de Antígeno Prostático Específico (PSA). Além de ser rápido e seguro, com a associação com outros métodos diagnósticos, como o PSA e a biópsia, aumenta-se consideravelmente a precisão da identificação de casos suspeitos. Contudo, vale ressaltar que fatores como preconceito, constrangimento e falta de informação ainda fazem com que muitos homens evitem a realização do exame, o que dificulta a detecção precoce da doença [Oncoguia 2024].

4.5. Concentração de PSA

Em continuidade, existem recursos que auxiliam os médicos a determinarem o momento adequado para indicar uma biópsia. Isso é feito por meio de exames que detectam alterações no corpo, possivelmente causadas por complicações prostáticas. O exame mais amplamente conhecido é a dosagem do PSA (Antígeno Prostático Específico).

O PSA é uma proteína produzida pelas células epiteliais da próstata, cuja função principal é ajudar na liquefação do sêmen. Sua importância para o exame reside no fato de que, em condições normais, apenas uma pequena quantidade dessa proteína atinge a corrente sanguínea. Assim, a concentração de PSA no sangue de homens saudáveis tende a ser baixa. No entanto, a inflamação e a proliferação de células prostáticas aumentam a liberação dessa proteína no sangue, elevando a concentração, segundo o *American Cancer Society* (2025).

A quantidade considerada normal de PSA no sangue não possui um valor fixo, pois tende a aumentar com o avanço da idade. Por esse motivo, a literatura médica considera 4ng/mL como um valor de atenção, recomendando-se que homens que ultrapassem esse índice realizem exames complementares. Contudo, resultados acima ou abaixo desse limiar não garantem completamente o diagnóstico, já que abaixo de dele, há 15% de ter

câncer, entre 4ng/mL e 10ng/mL, 25%, e acima de 10ng/mL, 50%. [Araújo 2024].

Embora o PSA seja um marcador amplamente utilizado para rastreamento e monitoramento do câncer de próstata, suas limitações diagnósticas devem ser consideradas, uma vez que diversos fatores podem modular suas concentrações, resultando em falsos positivos ou falsos negativos. Variáveis interferentes incluem: atividade sexual recente (particularmente ejaculação nas 48 horas que antecedem o exame), prática de atividades que promovam compressão perineal (como ciclismo ou equitação), realização prévia de exame de toque retal, biópsia prostática recente, infecções do trato urinário, prostatites e uso de determinados fármacos. Dessa forma, recomenda-se um preparo adequado e anamnese detalhada antes da coleta do PSA, para mitigar interferências e aumentar a acurácia diagnóstica [Hampl 2024].

Além disso, diagnósticos falso-positivos podem ser extremamente desconfortáveis para o paciente, pois implicam a realização de exames extensos, alguns invasivos, gerando ansiedade e outros problemas, principalmente de ordem psicológica. De forma mais grave, pode haver falso-negativo, fazendo com que o paciente não busque tratamento, permitindo que o câncer avance para estágios mais difíceis de reverter.

Por fim, além dos casos de falso-positivo e falso-negativo, há uma terceira situação preocupante: a de homens diagnosticados com câncer de próstata em estágio tão brando que jamais afetariam suas vidas (como mencionado anteriormente). Nesses casos, o paciente é exposto a tratamentos desgastantes e, mesmo sem necessidade clínica, pode viver com ansiedade pela possibilidade de progressão da doença. Por isso, muitos profissionais recomendam que o paciente esteja atento aos sintomas e busque ajuda médica ao perceber sinais de desconforto.

4.6. Biópsia da próstata

Caso seja detectado um PSA fora do comum, um exame de toque retal ou sintomas apontem para um possível caso clínico, é bem provável que o paciente precise passar por um exame de biópsia para tentar atestar com maior precisão se ele possui a doença ou não.

O exame de biópsia se inicia com o médico removendo algumas amostras da região da próstata. Após isso, é realizada a análise dessas amostras, por meio de processos como ressonância magnética ou exames de imagem. Contudo, sabe-se que, mesmo em casos como esses, é possível que não seja confirmado um câncer de próstata, ainda que o paciente esteja com a doença em desenvolvimento. Caso os sintomas e os outros exames ainda causem suspeitas ao médico, ele pode indicar novos procedimentos ao paciente para confirmar a existência ou não do tumor [American Cancer Society 2025].

Se realmente for confirmada a presença de células cancerígenas na região da próstata, o próximo passo será classificar o estágio em que aquele câncer se encontra, para que, com isso, seja possível selecionar um tratamento adequado ao paciente.

Para fazer a classificação do estadiamento de um câncer, é preciso avaliar questões como o tamanho, a probabilidade de se multiplicar para outras regiões, a velocidade com que o faz e entre outros fatores. Dessa forma, o sistema Gleason é amplamente aceito para realizar essa classificação, por meio de agrupamentos por notas, e, com isso, definir a atenção que deve ser dada ao tumor.

4.7. Sistema Gleason

Conforme o *American Cancer Society* (2025), o sistema Gleason, amplamente utilizado há muitos anos, atribui notas ao câncer de próstata com base na similaridade das células cancerosas ao tecido normal da próstata. Este sistema avalia o grau de anormalidade celular, fornecendo uma pontuação que ajuda a prever o comportamento do câncer. Cânceres com notas mais altas apresentam células que se desviam mais do aspecto do tecido prostático normal, indicando uma maior agressividade e uma maior probabilidade de crescimento e disseminação rápida.

Assim sendo, as notas vão de 1 (células mais semelhantes ao tecido normal) a 5 (células anormais com risco de infiltração em outras partes). Como o câncer de próstata frequentemente apresenta áreas com mais graus de diferenciação, a pontuação de Gleason é determinada pela soma dos dois padrões predominantes encontrados na amostra. Ou seja, o grau mais comum e o segundo mais comum. Então, dessa forma, se uma amostra obteve a pontuação “3 + 4 = 7” indica que a maior parte do tumor é grau 3, seguido por uma parte menor de grau 4 [Conte 2025].

Com base nessa pontuação, os tumores costumam ser classificados em três grupos, segundo Takano (2020):

- Gleason ≤ 6 : bem diferenciado (baixo grau);
- Gleason = 7: moderadamente diferenciado (grau intermediário);
- Gleason de 8 a 10: pouco diferenciado (alto grau).

Contudo, percebeu-se que tumores com a mesma pontuação total podem ter comportamentos diferentes. Por exemplo, tumores com “3 + 4 = 7” tendem a apresentar um prognóstico melhor do que aqueles com “4 + 3 = 7”. Da mesma forma, tumores com pontuação 8 são, na maioria das vezes, são menos agressivos do que os de pontuação 9 ou 10 [Araújo 2024].

Diante disso, foi desenvolvida uma nova forma de classificação chamada grupos de grau, que oferece uma maneira mais clara de comunicar a gravidade do câncer, especialmente para os pacientes. Essa escala vai de 1 a 5, segundo o *American Cancer Society* (2025):

- Grupo 1: Gleason ≤ 6 ;
- Grupo 2: Gleason 3 + 4 = 7;
- Grupo 3: Gleason 4 + 3 = 7;
- Grupo 4: Gleason 8;
- Grupo 5: Gleason 9 a 10.

4.8. Classificação de Tumores Malignos

Além do Gleason, outra classificação utilizada é a TNM (Classificação de Tumores Malignos). Essa sigla refere-se ao estadiamento, descrição da extensão anatômica do tumor, criado pela União Internacional contra o Câncer (UICC), possibilitando identificar em que estágio a doença se encontra [Fundação Oncocentro de São Paulo 2022]. Dessa forma, a letra T descreve as características do tumor primário, com as possibilidades indicadas na Tabela 1 a seguir.

Classificação	Descrição
TX	Tumor primário não pode ser avaliado
T0	Nenhuma evidência de tumor primário
T1	Tumor clinicamente não aparente, não palpável nem visível por imagem
T1a	Tumor encontrado incidentalmente em exame histológico em 5% ou menos do tecido ressecado
T1b	Tumor encontrado incidentalmente em exame histológico em mais de 5% do tecido ressecado
T1c	Tumor identificado por biópsia com agulha, por exemplo, devido a antígeno prostático específico (PSA) elevado
T2	Tumor confinado à próstata
T2a	Tumor envolve até metade de um lobo ou menos
T2b	Tumor envolve mais da metade de um lobo, mas não ambos
T2c	Tumor envolve ambos os lobos
T3	Tumor se estende além da cápsula prostática
T3a	Extensão extracapsular (unilateral ou bilateral), incluindo envolvimento microscópico do colo da bexiga
T3b	Tumor invade vesículas seminais
T4	Tumor é fixo ou invade estruturas adjacentes além das vesículas seminais: esfíncter externo, reto, músculos levantadores do ânus e/ou parede pélvica

Tabela 1. Classificação T – TNM

Já a sigla N refere-se aos linfonodos, estruturas essenciais para o sistema imunológico, responsáveis por filtrar micro-organismos e células anormais e produzir linfócitos (células de defesa), como demonstra a Tabela 2 abaixo.

Classificação	Descrição
NX	Linfonodos regionais não podem ser avaliados
N0	Sem metástase em linfonodos regionais
N1	Metástase em 1 a 2 linfonodos regionais
N2	Metástase em 3 a 6 linfonodos regionais
N3	Metástase em 7 ou mais linfonodos regionais

Tabela 2. Classificação N – TNM

Por fim, o índice M indica a presença de metástase, ou seja, o processo pelo qual as células malignas se desprendem do tumor primário e se instalam em outras regiões do corpo. Seus valores podem ser M0 (sem metástase) ou M1 (com metástase em outros tecidos) [Sobel, Gospodarowicz e Wittekind 2009].

Essas nomenclaturas são importantes para definir o estágio de um câncer, mas é essencial compreender que, devido à variedade de tipos existentes, não se pode generalizar todos com uma única classificação. Por isso, as letras são combinadas para poder descrever situações distintas, dependendo do tumor, o que dá origem a grupos de estágio clínico, como exibe a Tabela 3 inferior.

Estádio	T	N	M
I	T1, T2a	N0	M0
II	T2b, T2c	N0	M0
III	T3	N0	M0
IV	T4	N0	M0
IV	Qualquer T	N1	M0
IV	Qualquer T	Qualquer N	M1

Tabela 3. Grupo de estágio clínico

Por fim, há outro tipo de agrupamento, o qual leva em consideração o TNM, PSA e Gleason apresentados anteriormente, como mostra a Tabela 4. Contudo, caso essas duas últimas informações não estejam disponíveis, não é possível utilizar essa categorização.

Grupo	T	N	M	PSA (ng/mL)	Gleason
I	T1a–c	N0	M0	< 10	≤ 6
I	T2a	N0	M0	< 10	≤ 6
IIA	T1a–c	N0	M0	< 20	7
IIA	T1a–c	N0	M0	≥ 10 e < 20	≤ 6
IIA	T2a, T2b	N0	M0	< 20	≤ 7
IIB	T2c	N0	M0	Qualquer	Qualquer
IIB	T1–2	N0	M0	≥ 20	Qualquer
IIB	T1–2	N0	M0	Qualquer	≥ 8
III	T3a, T3b	N0	M0	Qualquer	Qualquer
IV	T4	N0	M0	Qualquer	Qualquer
IV	Qualquer	N1	M0	Qualquer	Qualquer
IV	Qualquer	Qualquer	M1	Qualquer	Qualquer

Tabela 4. Grupos de prognóstico

4.9. Monitoramento ativo

Cânceres de próstata no estado inicial podem apresentar uma evolução lenta e não demonstrar sintomas significativos. Em casos como esses, opta-se pelo monitoramento ativo do paciente, realizando acompanhamento regular com exames (como PSA, biópsias periódicas e exames de imagem) para avaliar a progressão do câncer, sem intervenção imediata. Isso evita ações mais invasivas, contudo, deve ser realizada com cautela, para não permitir um avanço silencioso, o qual pode trazer mais complicações futuras [American Cancer Society 2023].

4.10. Cirurgia

A prostatectomia radical consiste na remoção total da próstata e dos tecidos ao redor do órgão, incluindo os linfonodos regionais, vesículas seminais e vasos entre os testículos e a uretra. Essa é uma opção comum para homens com câncer localizado que ainda não se espalhou para outras partes do corpo. A cirurgia pode ser realizada de forma aberta ou minimamente invasiva (usando robôs ou técnicas laparoscópicas). Como consequência, o procedimento causa infertilidade e pode levar a incontinência urinária e disfunção erétil, porém, é um dos métodos de maior eficácia [American Cancer Society 2023].

4.11. Radioterapia

A radioterapia é um dos principais tratamentos para o câncer de próstata, sendo indicada conforme o estágio da doença. Pode ser utilizada como tratamento inicial em casos localizados, como alternativa à cirurgia, ou associada a outros métodos em situações de maior risco. Ela também pode ser recomendada quando o câncer retorna após outro procedimento ou se espalha para outras áreas do corpo do paciente. Em estágios avançados, além de ajudar no controle da doença, a radioterapia contribui para aliviar sintomas, como dores ósseas [American Cancer Society 2025].

Ainda, segundo a *American Cancer Society* (2025). Entre as modalidades de radioterapia, a de feixe externo de raio-x costuma ser a alternativa mais comum e pode acontecer com aplicações convencionais ou em sessões reduzidas com doses mais altas. Alguns dos efeitos colaterais que essa técnica pode causar, estão alterações intestinais e urinárias, além de disfunção erétil a longo prazo e fadiga.

4.12. Terapia Hormonal

A terapia hormonal é recomendada para reduzir o nível de andrógenos ou bloquear a entrada deles nas células cancerígenas. Esses hormônios são produzidos nos testículos, nas glândulas adrenais e até mesmo pelas células cancerígenas. Por essa razão, o tratamento hormonal visa retardar a progressão do câncer, embora dificilmente, por si só, seja capaz de curá-lo [American Cancer Society 2023].

Por esse motivo, comumente o tratamento hormonal é combinado com a radioterapia, como forma de evitar a progressão do câncer. Ainda assim, de acordo com a *American Cancer Society* (2023), caso o câncer persista após cirurgias ou radioterapia, a terapia hormonal pode voltar a ser empregada, sendo viável em situações em que o câncer se espalha para outros locais ou quando há dificuldade para determinar a localização exata do tumor.

Embora represente uma boa opção, seu maior problema está no fato de que, caso o câncer continue progredindo, há chances de ele desenvolver resistência a esse tipo de

medicação, tornando-se uma opção descartável ao longo do tratamento [*American Cancer Society* 2023].

4.13. Quimioterapia

Tratamento menos comum para o câncer de próstata, a quimioterapia é utilizada principalmente em situações em que o câncer se espalha para além da próstata. Outro fator é que esse tratamento pode ser aplicado simultaneamente à terapia hormonal, sem conflitos, sendo também recomendável para casos em que a terapia hormonal deixa de surtir efeito [*American Cancer Society* 2023].

Muito semelhante à terapia hormonal, seu uso dificilmente levará à cura, mas tende a atrasar a proliferação do câncer. Para tumores de próstata, sua aplicação pode ocorrer por via oral, com doses administradas em ciclos, cuja frequência pode variar conforme as respostas apresentadas pelo paciente [*American Cancer Society* 2023].

Mesmo que esse procedimento apresente certa eficiência em retardar o avanço do câncer, seus efeitos colaterais podem ser mais desagradáveis que os da terapia hormonal, incluindo sintomas como queda de cabelo, feridas na boca, vômitos, diarreia, cansaço, além de maior risco de infecções e sangramentos. Em casos mais graves, pode ainda provocar reações alérgicas e danos nos nervos periféricos, sendo imprescindível realizar ajustes ou suspender as doses do medicamento, conforme a necessidade [*American Cancer Society* 2023].

4.14. Imunoterapia

Nessa opção, o câncer é tratado com uma vacina conhecida como *Sipuleucel-T*, que funciona de forma diferente das vacinas tradicionais. Em vez de prevenir uma contaminação futura, ela estimula o próprio sistema imunológico a combater as células de câncer de próstata. Essa abordagem é indicada para estágios avançados da doença, quando o paciente já não responde mais à terapia hormonal [*American Cancer Society* 2023].

A vacina é produzida individualmente: retiram-se os leucócitos (glóbulos brancos) do paciente, que são então tratados com uma célula de proteína presente no câncer. Após esse processo, essas células são injetadas novamente no paciente, por meio de infusão intravenosa, segundo o *American Cancer Society* (2023).

Todavia, esse tratamento não impede o avanço do câncer; sua função é prolongar a vida do paciente pelo maior tempo possível enquanto for eficaz. Seus efeitos colaterais costumam incluir febre, calafrios, fadiga e dores de cabeça. Em situações em que o câncer está mais avançado e são necessárias doses mais agressivas, sintomas como pressão alta, dificuldade para respirar e danos a órgãos saudáveis também podem ocorrer [*American Cancer Society* 2023].

4.15. Inteligência artificial na saúde

Nesse sentido, com o avanço atual dos estudos práticos, das capacidades computacionais das máquinas e aumento dos conjuntos de dados, a área de inteligência artificial (IA) vem ganhando relevância no meio da saúde. Segundo o Conselho Federal de Enfermagem [Brasil 2024], a Organização Mundial da Saúde (OMS) coloca a tecnologia com alto potencial de melhorar a execução dos serviços, com rapidez e assertividade no diagnóstico e tratamento, ajudar no atendimento, pesquisas, confecção de fármacos,

vigilância e rastreamento de doenças.

Ainda conforme o Conselho [Brasil 2024], os investimentos privados em IA na saúde atingiram 28,9 bilhões de dólares globalmente, somados os números de 2022 com os cinco anos anteriores. Com tal marca, o setor ficou na primeira posição no ranking de capital aportado, superando o mercado financeiro e de varejo. No Brasil, a perspectiva é de 42 bilhões de reais em tecnologias inovadoras na saúde, até 2026, previsto na estratégia nacional [Victor 2024].

Dessa maneira, de forma conceitual, Norvig e Russel (2013) definem IA como “o estudo de agentes que recebem percepções do ambiente e executam ações”. Dentre alguns tópicos de especialização, encontram-se o processamento de linguagem natural, visão computacional, robótica e aprendizado de máquina. Sobre este último, deu-se início na década de 1960, com foco em aprender padrões e tendências a partir de dados e utilizá-los em cenários semelhantes [Araújo 2024].

4.16. Aprendizado de máquina na saúde

Em continuidade, o aprendizado de máquina possui algumas abordagens, como a aprendizagem supervisionada, não supervisionada e por reforço. Dessa forma, a primeira consiste em observar elementos que possuem atributos e uma saída alvo bem definida e mapear as entradas para essa saída. Ela se subdivide em classificação, saída categórica e de conjunto finito, e regressão, numérica e contínua. Já a segunda, é quando não se tem uma saída explícita, então se detectam agrupamentos e associações lógicas das entradas. Por fim, na terceira, o agente aprende por meio de um sistema de recompensa ou punições [Norvig e Russel 2013].

Por exemplo de supervisionada, um modelo de classificação poderia receber uma tabela de informações, cujas linhas correspondam a dados das células de um tumor e uma das colunas indique de qual câncer se trata. Sendo assim, a tarefa seria aprender a classificar o tipo cancerígeno da formação, com base nos atributos de uma nova entrada.

Um outro modelo de regressão seria capaz de analisar uma tabela com dados clínicos do paciente, idade, sexo, tipo e estágio de câncer, tratamento aplicado, com uma coluna exibindo o tempo de sobrevida em meses. Portanto, o intuito seria prever este tempo para um novo caso a ser verificado.

No contexto da não supervisionada, uma possibilidade seria analisar uma tabela de um único tipo de câncer, com as entradas representando as expressões dos genes dos pacientes. A partir disso, compreender as similaridades e diferenças para agrupar em conjuntos e encontrar potencialmente novos subgrupos desse câncer em questão.

Já na aprendizagem por reforço, um modelo poderia ser treinado em um ambiente simulado de aplicação de radioterapia em células cancerígenas. Com isso, o objetivo seria analisar a posição, ângulo e intensidade do feixe, reforçando positivamente quando se maximizasse o dano ao tumor e aplicando punição quando não fosse suficiente ou atingisse células saudáveis.

4.17. Controvérsias da inteligência artificial na saúde

Por outro lado, a inteligência artificial pode ter um comportamento tendencioso ou até preconceituoso, conforme os dados de treinamentos forem obtidos. Ainda, o ponto de segurança e privacidade das informações, bem como de quem é a responsabilidade legal

caso algum problema ocorra, são questões amplamente discutidas atualmente, sem consenso geral. Além disso, no aspecto técnico, há uma preocupação de aumento da discrepância da assistência médica entre países em desenvolvimento e desenvolvidos, devido aos custos de desenvolvimento e implantação das ferramentas [Brasil 2023].

Em consonância, a Sociedade Brasileira de Medicina Tropical [Brasil 2023] também publicou algumas diretrizes recomendadas pela OMS no uso dessa tecnologia. Do lado médico, a tomada de decisão final que deve ser realizada por humanos e os profissionais da saúde devem receber treinamento adequado para utilização. Na parte dos desenvolvedores, eles devem não só serem responsáveis por monitorar, avaliar a performance e garantir o correto funcionamento, bem como serem transparentes ao prestar esclarecimento sobre as etapas e se atentarem à segurança e diversidade dos dados, para evitar vieses.

4.18. K-Vizinhos mais próximos

O K-Vizinhos Mais Próximos (KNN) é um modelo de aprendizado de máquina supervisionado baseado em distância, que parte da premissa de que dados similares tendem a estar concentrados em uma mesma região do espaço de entrada. Dessa forma, na etapa de treinamento, o algoritmo apenas armazena as amostras. Já na fase de teste, calcula a distância entre a entidade a ser predita e todos os demais objetos. Em seguida, seleciona as K menores distâncias, de acordo com o parâmetro definido pelo usuário. Por fim, a classe mais frequente (moda) entre os vizinhos mais próximos é atribuída ao novo dado [IBM 2025].

Entre os métodos de distância, o mais comum é a euclidiana, descrita na equação 1 a seguir. Sendo assim, “i” e “j” denotam diferentes pontos, “a” um atributo e “f” a quantidade de *features*.

$$d(X_i, X_j) = \sqrt{\sum_{a=1}^f (X_i^a - X_j^a)^2}$$

(1)

Dentre as vantagens estão a fácil implementação, interpretação e treinamento. Porém, como desvantagens, tem-se a perda de performance com o crescimento da base de dados, sensibilidade à valores redundantes, irrelevantes ou exceções, bem como sofre com a dispersão das informações em cenários de alta dimensionalidade, que dificulta o trabalho em um ambiente rarefeito [IBM 2025].

4.19. Regressão logística

Os modelos lineares são aqueles que tentam encontrar uma equação (equação 2) que combina as variáveis de entrada (x_i) com os respectivos coeficientes (w_i) e um viés (a), ajustando os pesos a cada iteração, para diminuir a saída da função de custo, a qual mede o erro das previsões [Scikit-learn 2025].

$$y = a + \sum_{i=1}^n x_i w_i$$

(2)

A regressão logística é um modelo utilizado em classificação, para avaliar a probabilidade de um elemento ser de determinado rótulo, com base nas variáveis independentes. Tal algoritmo combina ao final uma função logística (sigmoide) em formato de ‘S’, por exemplo da equação 3 e Figura 2, que restringe as previsões num intervalo de 0 a 1, permitindo em problemas binários configurar um limiar, por exemplo

0.5, que valores acima deste são da classe A e abaixo disto da classe B [AWS 2025].

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

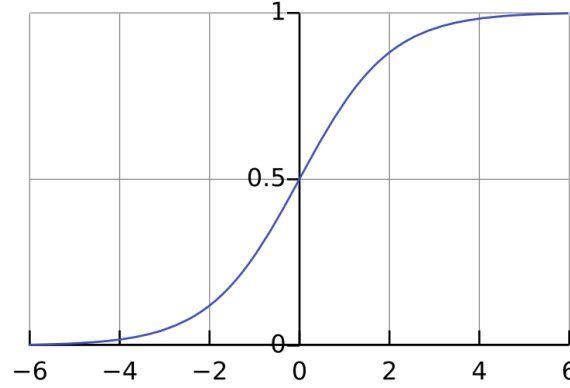


Figura 2. Exemplo de função logística

Quando o problema envolve mais de dois rótulos, aplica-se a regressão logística multinomial, uma extensão da anterior. O modelo utiliza a função *softmax* (equação 4) para generalizar a função sigmoide, atribuindo uma probabilidade a cada classe, de forma que a soma das delas seja igual a 1. Sobre os argumentos, “K” representa a quantidade de classes e “z” o resultado da função linear. Por fim, o modelo decidirá a classificação predita por aquela de maior chance [Universidade de Stanford 2025].

$$softmax(z_i) = \frac{e^{z_i}}{\sum_{j=1}^K e^{z_j}} \quad (4)$$

4.20. Árvore de decisão

Árvores de classificação são modelos preditivos que organizam decisões em forma de estrutura hierárquica, semelhante a uma árvore. Cada nó interno representa uma condição de decisão com base em um atributo, e cada folha representa uma classe final (rótulo de saída). O processo de criação da árvore envolve dividir os dados em subconjuntos com base em critérios que maximizam a "pureza" desses subconjuntos, ou seja, que os tornem o mais homogêneos possível em relação à variável alvo [IBM 2025].

Ainda segundo a IBM, entre as métricas para guiar a divisão, tem-se a entropia (equação 5), que mede a impureza ou incerteza do conjunto (S), ou seja, quanto mais misturadas estiverem as classes, maior a entropia. Nesse sentido, a entropia varia de 0, quando todos os valores são de uma única classe, a 1, quando as classes estão distribuídas igualmente nas amostras. Conforme a fórmula abaixo, “n” representa a quantidade de classes e “p” proporção de exemplos da classe “i”.

$$Entropia(S) = - \sum_{i=1}^n p_i \log_2(p_i) \quad (5)$$

Após calculada, é obtido o quanto a entropia diminui após se dividir o conjunto com base em um atributo, o ganho de informação. O atributo com o maior ganho de informação é escolhido para a divisão naquele ponto da árvore. Como mostra a equação 6, Entropia(S) é a entropia antes da divisão, S_v é o subconjunto de S no qual o atributo A tem valor v [IBM 2025].

$$Ganho(S, A) = Entropia(S) - \sum_{v \in \text{valores}(A)} \frac{|S_v|}{|S|} Entropia(S_v) \quad (6)$$

4.21. Floresta Aleatória

A Floresta Aleatória é um algoritmo baseado em múltiplas árvores de decisão, no qual cada uma é treinada com um conjunto de dados diferente e um subconjunto de atributos diferentes. Para cada árvore, o algoritmo escolhe aleatoriamente um subconjunto dos dados de treino, com reposição, ou seja, alguns exemplos podem aparecer mais de uma vez, e outros podem ficar de fora. Depois que todas estão treinadas, cada uma prediz a classe alvo para determinada amostra, sendo o resultado final definido por maioria simples [IBM 2025].

Como benefícios dessa técnica, ela pode reduzir o risco de *overfitting* das árvores de decisão, generalizar melhor e lidar com dados de maiores dimensões e complexidade. Por outro lado, pode consumir mais memória, ser mais lento e menos interpretável [IBM 2025].

4.22. Naive Bayes

O algoritmo *Naive Bayes* é um tipo de classificador baseado em probabilidade, construído a partir do Teorema de Bayes. Esse teorema possibilita a atualização da probabilidade de um evento com base em uma nova evidência observada, ou seja, permite inverter a direção da probabilidade condicional. Conforme a equação 7 a seguir, $P(Y|X)$ é a probabilidade a posteriori (probabilidade de Y dado que X ocorreu), $P(X|Y)$ é a probabilidade condicional (probabilidade de observar X se Y for verdadeiro), $P(Y)$ é a probabilidade a priori (a chance de Y ocorrer antes de qualquer evidência) e $P(X)$ é a probabilidade marginal (probabilidade de X ocorrer, independentemente de Y) [KAVLAKOGLU e RUSSI 2025].

$$P(Y|X) = \frac{P(X|Y) * P(Y)}{P(X)} \quad (7)$$

O *Naive Bayes* é chamado de "ingênuo" porque assume que todas as variáveis do modelo são independentes entre si, ou seja, o valor de uma não influencia o valor de outra. Ainda assim, essa simplificação facilita os cálculos e reduz a complexidade computacional, tornando o modelo rápido e eficiente, mesmo com poucos dados e que a suposição não reflita a realidade com precisão. Além disso, por $P(X)$ ser o mesmo para todas as classes e não afetar a comparação entre elas, pode ser ignorado [KAVLAKOGLU e RUSSI 2025].

Portanto, para prever um novo rótulo, o modelo calcula as probabilidades condicionais de cada classe das Y possíveis, escolhendo aquele mais provável, conforme a equação 8, multiplicando a probabilidade a priori pelas probabilidades a posteriori de todas as n features da nova amostra [KAVLAKOGLU e RUSSI 2025].

$$Classe = \arg \max_{y \in Y} [P(y) * \prod_{i=1}^n P(X_i|y)] \quad (8)$$

4.23. XGBoost

O *XGBoost* (*Extreme Gradient Boosting*) é uma implementação eficiente e otimizada da técnica de *boosting* por gradiente, amplamente utilizada em tarefas de aprendizado supervisionado. Seu diferencial está na forma como combina diversas árvores de decisão rasas para formar um modelo forte, mais preciso e robusto [IBM 2025].

Diferentemente da Floresta Aleatória, que treina modelos em paralelo e combina os resultados por votação, o *boosting* adota uma abordagem sequencial: cada novo modelo

é treinado para corrigir os erros dos anteriores, dando mais peso às instâncias com maior erro [IBM 2025].

O *XGBoost* utiliza a descida do gradiente para minimizar a função de perda. A cada iteração, calcula-se o gradiente dessa função em relação às previsões atuais, o que orienta os ajustes necessários no modelo, pela derivada da função de custo. Esse processo continua até atingir um mínimo da função de perda, preferencialmente 0, indicando um local de mínimo local ou global, ou até que um critério de parada seja alcançado, como número máximo de iterações ou ausência de melhoria no desempenho [IBM 2025].

4.24. Cluster Centroids

Cluster Centroids é uma técnica de tratativa de dados desbalanceados por *undersampling* da classe majoritária. Ele é baseado no algoritmo K-médias e mantém apenas um ponto mais próximo para cada *n* centroides de *cluster*, definido pelo usuário. Geralmente, a quantidade de *clusters* é definida pela quantidade de elementos da classe minoritária [Imbalanced Learn 2025].

Sobre o K-médias, ele inicia plotando os dados em um espaço das *features* analisadas e definindo pontos aleatórios, que serão os centros dos agrupamentos. A próxima etapa é calcular a distância, geralmente euclidiana, para cada amostra, determinando assim que os pontos mais próximos de um centroide de um *cluster* são semelhantes entre si. Após isso, a posição do centro é recalculada, para ocupar a média das distâncias dos elementos do *cluster*. Esses dois últimos processos são repetidos iterativamente, até um momento de convergência ou critério de máximo de tentativas configuradas [KAVLAKOGLU e WINLAND 2025].

4.25. Métricas para modelos de classificação

A acurácia representa a proporção de previsões corretas feitas pelo modelo em relação ao total de previsões. Em outras palavras, ela indica quantas vezes o modelo acertou, considerando tanto as classes positivas quanto as negativas. Embora seja uma métrica intuitiva e amplamente utilizada, ela pode ser enganosa em cenários com classes desbalanceadas, como mostra a equação 9.

$$\text{Acurácia} = \frac{(\text{Verdadeiros Positivos} + \text{Verdadeiros Negativos})}{\text{Total de Amostras}} \quad (9)$$

A precisão mede a proporção de exemplos classificados como positivos que realmente pertencem à classe positiva. Essa métrica é útil quando o objetivo é minimizar falsos positivos, ou seja, evitar que o modelo classifique como positivo algo que não é, conforme equação 10 [Microsoft 2024].

$$\text{Precisão} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Positivos})} \quad (10)$$

O *recall* avalia a capacidade do modelo em identificar corretamente todos os exemplos que pertencem à classe positiva. É especialmente importante em situações onde perder um caso positivo pode ter consequências significativas, como em diagnósticos médicos, segundo equação 11 [Microsoft 2024].

$$\text{Recall} = \frac{\text{Verdadeiros Positivos}}{(\text{Verdadeiros Positivos} + \text{Falsos Negativos})} \quad (11)$$

O *F1-score* é a média harmônica entre precisão e recall, sendo uma métrica que busca

equilibrar essas duas medidas. Ele é particularmente útil quando há uma necessidade de balancear os erros de falsos positivos e falsos negativos, e quando os dados estão desbalanceados, equação 12 abaixo [Microsoft 2024].

$$F1Score = 2 * \frac{(Precisão * Recall)}{(Precisão + Recall)} \quad (12)$$

5. Trabalhos relacionados

No presente tópico são explicados quatro trabalhos correlatos, os quais ressaltam a importância do tema, ao trazer diferentes visões no contexto do câncer de próstata com inteligência artificial. Os dois primeiros de Pan et al. (2024) e Tang et al. (2024) falam sobre a escolha de tratamento no cenário dos Estados Unidos. Já Maeda (2022), no Brasil, discute um contexto pós tratamento, analisando a recidiva do câncer, bem como Araújo (2024), que debate o momento inicial, de diagnóstico da doença.

Nesse sentido, para o uso de aprendizado de máquina nessa área da saúde, destaca-se o artigo de Pan et al. (2024), no qual a pesquisa propôs um modelo de aprendizado profundo com a personalização do tratamento, minimizando incertezas que pudessem ocasionar tratamentos desnecessários, como casos de prostatectomia radical, que é a remoção total do órgão. Para isso, foi desenvolvido um modelo, capaz de estimar o efeito individual do tratamento em dados de sobrevivência.

Em continuidade, Pan et al. (2024) trabalhou com uma base de mais de 35 mil pacientes do programa de Vigilância, Epidemiologia e Resultados Finais dos Estados Unidos. Os resultados demonstraram que aqueles cujo tratamento foi condizente com as recomendações do modelo tiveram melhores taxas de sobrevida. Esse desempenho foi avaliado, entre outras métricas, pelo HR (razão de risco), que mede o risco de um evento ocorrer em um grupo em relação a outro ao longo do tempo, sendo que um valor de HR inferior a 1 indica menor risco.

No caso do modelo feito por Pan et al. (2024), foi observado um valor ajustado de 0,76, representando uma redução de 24% no risco de morte para os pacientes que seguiram as recomendações. Além disso, a diferença média no tempo restrito de sobrevivência foi de até 3,81 meses, sendo que o modelo também identificou como variáveis influentes os fatores clínicos, como níveis de PSA, idade, escala de Gleason, estadiamento e presença de metástase.

Também no âmbito de pesquisas voltadas à personalização do tratamento do câncer de próstata, tem-se a publicação de Tang et al. (2024) baseado em 140.294 casos extraídos da mesma base de dados que o anterior. O estudo aplicou dez algoritmos de aprendizado de máquina para diferenciar pacientes candidatos a tratamentos cirúrgicos e não cirúrgicos, sendo alguns deles Máquina de Vetores de Suporte (SVM), KNN, Árvore de Decisão, Regressão Logística, Floresta Aleatória, Rede Neural e *AdaBoost*, *XGBoost*, *LightGBM* e *CatBoost*.

As performances de Tang et al. (2024) foram avaliadas pela área sob a curva (AUC), acurácia, *recall* e *precision*, enquanto o método SHAP (*Shapley Additive Explanations*) elucidou os principais fatores que influenciaram nas predições. Para comparar as classes estudadas, também se utilizaram da análise de sobrevida, revelando que pacientes submetidos a cirurgia apresentaram aumento de 20,36% na taxa de sobrevivência em 10 anos em relação aos não cirúrgicos, com a prostatectomia radical alcançando 89,2%.

Dentre os modelos avaliados por Tang et al. (2024), o *CatBoost* destacou-se como o que apresentou o melhor desempenho, com AUC de 96,8%, acurácia, *precision* e *recall* próximas de 88%, evidenciando alta capacidade preditiva e robustez para generalização em diferentes conjuntos de dados. A interpretação por meio do método SHAP revelou que as variáveis clínicas tradicionais, como o estágio T da Classificação de tumores malignos (TNM), a idade, a porcentagem de biópsias positivas, o nível de PSA e a escala de Gleason, foram os fatores mais relevantes para a recomendação do tratamento mais adequado.

Em continuidade, no cenário nacional, o trabalho de Maeda (2022) aplica algoritmos supervisionados, como *Naive Bayes* e SVM, para prever a recidiva de câncer a partir da base de dados da Fundação Oncocentro de São Paulo (FOSP), a mesma utilizada neste trabalho.

Conforme Maeda (2022), os melhores resultados puderam ser observados na validação cruzada, com um acurácia igual a 99%, *precision* de 99% e o *F1-score* equivalente a 99%. Os valores obtidos com o treinamento se mostraram bastante elevados por conta do desbalanceamento entre as classes, destacando a importância do tratamento dos dados. Após aplicar a técnica de amostragem sintética da classe minoritária (SMOTE), as métricas mudaram para 67% de acurácia, *precision* de 95% e *F1-score* de 78%.

Por fim, a tese escrita por Araújo (2024) propõe uma abordagem baseada em algoritmos de aprendizado de máquina aplicados anteriormente à triagem de biópsias do tumor, utilizando exclusivamente variáveis clínicas. Esse projeto visa contribuir para a redução de exames invasivos desnecessários e minimizar os gastos envolvidos nesses exames.

Ainda sobre a tese de Araújo (2024), os algoritmos avaliados incluíram, entre outros, *Naive Bayes*, KNN, Árvore de Decisão e Redes Neurais. Entre esses, o classificador *Naive Bayes* apresentou o melhor desempenho, alcançando uma acurácia de 89,09%, *recall* de 92%, *precision* de 86,67% e AUC de 91,87%. Esses resultados foram obtidos com um treino com 219 pacientes e validação com 55, do Hospital Universitário da Universidade Federal do Maranhão.

Em síntese, é possível perceber a relevância acadêmica do estudo do problema, com diversos trabalhos abordando mais de um escopo da área com aprendizado supervisionado de máquina. Essa revisão da literatura auxilia no entendimento das métricas e modelos mais utilizados, bem como as características médicas que mais influenciam e descrevem o quadro da doença do paciente.

6. Materiais e métodos

Para o projeto, foi utilizada a linguagem *Python* no ambiente do *Google Colab*, junto a bibliotecas de manipulação e visualização de dados como *Pandas*, *Numpy*, *Matplotlib* e *Seaborn*, e outras focadas em aprendizado de máquina, como *Scikit-learn*, *Imbalanced-learn*, e *Xgboost*.

Nesse sentido, inicialmente se obteve a base de dados da Fundação Oncocentro de São Paulo (FOSP), em um arquivo .dbf de 3Gb. O conjunto continha informações de casos de todos os cânceres dos hospitais e clínicas do estado, entre 2000 e 2024, portanto, foi necessário filtrar apenas os de câncer de próstata e transformou-se o resultado em CSV

para facilitar a manipulação.

Em sequência, o CSV resultante continha 104 colunas e 131.437 linhas. Após a revisão do dicionário fornecido pela FOSP, selecionou-se 26 colunas, retirando aquelas não relacionadas ao câncer de próstata, pós tratamento e informações de identificação do paciente ou clínica.

Dessa forma, analisou-se mais a fundo as colunas selecionadas, sendo elas IDADE, DIAGPREV (se o paciente teve diagnóstico prévio), BASEDIAG (base do diagnóstico), EC (Estádio clínico), ECGRUP (Grupo de estadiamento), T (TNM-T), N (TNM-N), M (TNM-M), PSA, GLEASON, NAOTRAT (o motivo de não se ter aplicado tratamento), NENHUM (nenhum tratamento aplicado), CIRURGIA (cirurgia aplicada), RADIO (radioterapia aplicada), QUIMIO (quimioterapia aplicada), HORMONIO (hormonioterapia aplicada), TMO (transplante de medula óssea aplicada), IMUNO (imunoterapia aplicada), OUTROS (outros tratamentos aplicados), ULTINFO (última informação do paciente), CONSDIAG (diferença em dias entre consulta e diagnóstico), TRATCONS (diferença em dias entre consulta e tratamento), DIAGTRAT (diferença em dias entre diagnóstico e tratamento), PERDASEG (se houve perda de segmento no tratamento, RECENENHUM (se teve recidiva) e CIDO (Classificação internacional de doenças oncológicas).

Portanto, iniciou-se a tratativa de valores nulos, substituindo pela moda o único valor faltante de CIDO e por 0 em CONSDIAG, TRATCONS e DIAGTRAT, já que estavam vazios quando nenhum tratamento foi aplicado. Posteriormente, filtrou-se apenas os casos bem sucedidos e completos, ou seja, sem perda de segmento, sem recidiva e sem óbito pelo câncer, removendo as colunas PERDASEG, RECENENHUM, ULTINFO e NAOTRAT. Também, foram excluídas linhas relacionadas a informações de T, N e M pós tratamento, definidas como “Y”, e quando T era 0, por conta de não existir tumor. Sendo assim, o conjunto final ficou com 84.954 linhas e 22 colunas.

Para a classe alvo, as colunas NENHUM, CIRURGIA, RADIO, QUIMIO, OUTROS, HORMONIO, TMO e IMUNO eram booleanas, indicando a aplicação do respectivo tratamento. Dessa maneira, fez-se uma revisão da literatura e combinou-se elas em uma só colunas coluna categórica chamada TRATAMENTO, como mostra a Figura 3 abaixo.

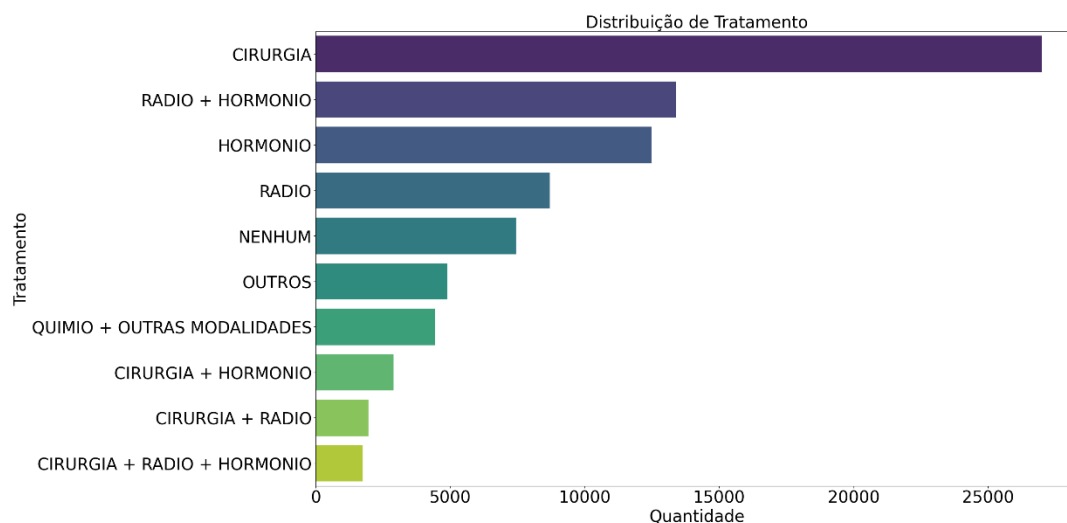


Figura 3. Tratamentos

Para a IDADE, o intervalo ia de 0 a 101 anos, com uma concentração maior entre 60 e 75 anos. Para lidar com os outliers, optou-se por remover as linhas abaixo de 39 e acima de 95, por não conterem uma dezena de informações. Sobre DIAGPREV, ela continha apenas duas classes, 2 – Diagnóstico anterior e 1 – Sem diagnóstico anterior, por isso foi transformada em booleana. Também, removeu-se ECGRUP por ter uma correlação exata para com EC, representando uma redundância das informações.

Para CIDO, ela era uma coluna de alta cardinalidade, com 48 valores únicos, com apenas 2 deles se destacando muito, 85503 e 81403, somando mais de 97% da quantidade total. Nesse cenário, tomou-se a decisão de agrupar as menos frequentes em uma categoria chamada OUTROS. Para CONSDIAG, TRATCONS, DIAGTRAT, observou-se uma grande concentração entre 0 a 100 dias, mas com muitos valores altos que acabavam desvirtuando as médias. Em uma análise mais aprofundada, também se observou a baixa relação com o tratamento aplicado, portanto removeram-se essas colunas.

Ao final, restaram-se 84.840 linhas e 11 colunas IDADE, DIAGPREV, BASEDIAG, EC, T, N, M, PSA, GLEASON, CIDO e TRATAMENTO. Depois, as colunas categóricas foram passadas pelo método de *One-hot encoding*, resultando num total de 52 colunas. Posteriormente, para balanceamento da classe alvo, utilizou-se a técnica de *Cluster Centroids*, o que diminui a quantidade para 17.550 linhas, 1.755 para cada um dos 10 rótulos.

Para o treinamento e teste dos modelos, o conjunto de dados foi dividido em uma proporção de 70% para treinamento, 12.285 linhas, e 30% de teste, 5.265 amostras.

7. Resultados

Para o modelo de KNN, foram escolhidos 21 vizinhos, performando com acurácia de 55% e a Tabela 5 demonstrando a *precision*, *recall* e *F1-Score* de cada classe.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	0.90	0.92	0.91	526
CIRURGIA + HORMONIO	0.26	0.27	0.26	526
CIRURGIA + RADIO	0.28	0.53	0.36	527
CIRURGIA + RADIO + HORMONIO	0.26	0.39	0.31	527
HORMONIO	0.91	0.78	0.84	527
NENHUM	0.73	0.51	0.60	526
OUTROS	0.53	0.38	0.44	527
QUIMIO + OUTRAS MODALIDADES	0.51	0.42	0.46	526
RADIO	0.79	0.55	0.65	526
RADIO + HORMONIO	0.97	0.75	0.84	527

Tabela 5. Resultados KNN

Já para a Regressão Logística com 1.000 iterações, a acurácia foi de 58% com a Tabela 6 exibindo mais detalhes.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	0.98	0.94	0.96	526
CIRURGIA + HORMONIO	0.31	0.20	0.24	526
CIRURGIA + RADIO	0.35	0.61	0.44	527
CIRURGIA + RADIO + HORMONIO	0.32	0.38	0.35	527
HORMONIO	0.94	0.81	0.87	527
NENHUM	0.66	0.53	0.59	526
OUTROS	0.46	0.45	0.46	527
QUIMIO + OUTRAS MODALIDADES	0.50	0.55	0.52	526
RADIO	0.71	0.61	0.65	526
RADIO + HORMONIO	0.92	0.77	0.83	527

Tabela 6. Resultados Regressão Logística

Para a Árvore de Decisão com altura máxima de 10, a acurácia foi de 57% e a Tabela 7 mostra os resultados para cada classe.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	0.98	0.93	0.96	526
CIRURGIA + HORMONIO	0.32	0.21	0.25	526
CIRURGIA + RADIO	0.27	0.75	0.40	527
CIRURGIA + RADIO + HORMONIO	0.31	0.25	0.28	527
HORMONIO	0.90	0.82	0.86	527
NENHUM	0.78	0.51	0.62	526
OUTROS	0.62	0.32	0.43	527
QUIMIO + OUTRAS MODALIDADES	0.47	0.57	0.51	526
RADIO	0.80	0.58	0.67	526
RADIO + HORMONIO	0.95	0.75	0.84	527

Tabela 7. Resultados Árvore de Decisão

Para a Floresta Aleatória, foram definidas 100 árvores de no máximo 10 de altura, com acurácia de 59% e com a Tabela 8 a seguir.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	1.00	0.94	0.96	526
CIRURGIA + HORMONIO	0.37	0.17	0.24	526
CIRURGIA + RADIO	0.28	0.73	0.41	527

CIRURGIA + RADIO + HORMONIO	0.28	0.37	0.32	527
HORMONIO	0.99	0.81	0.89	527
NENHUM	0.87	0.52	0.65	526
OUTROS	0.64	0.37	0.47	527
QUIMIO + OUTRAS MODALIDADES	0.48	0.65	0.55	526
RADIO	0.97	0.56	0.71	526
RADIO + HORMONIO	0.99	0.76	0.86	527

Tabela 8. Resultados Floresta Aleatória

Para o *Naive Bayes*, a acurácia foi de 53% e está mais bem descrita na Tabela 9.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	0.68	0.93	0.79	526
CIRURGIA + HORMONIO	0.28	0.21	0.24	526
CIRURGIA + RADIO	0.30	0.65	0.41	527
CIRURGIA + RADIO + HORMONIO	0.29	0.35	0.32	527
HORMONIO	0.93	0.63	0.75	527
NENHUM	0.81	0.45	0.58	526
OUTROS	0.63	0.33	0.43	527
QUIMIO + OUTRAS MODALIDADES	0.44	0.56	0.50	526
RADIO	0.83	0.51	0.63	526
RADIO + HORMONIO	0.82	0.74	0.78	527

Tabela 9. Resultados *Naive Bayes*

Por fim, o *XGBoost* teve como parâmetros 100 árvores e 4 de profundidade máxima, com 60% de acurácia e a Tabela 10 trazendo as métricas mais específicas.

Tratamento	<i>Precision</i>	<i>Recall</i>	<i>F1-Score</i>	Quantidade
CIRURGIA	0.99	0.94	0.96	526
CIRURGIA + HORMONIO	0.40	0.20	0.27	526
CIRURGIA + RADIO	0.34	0.69	0.46	527
CIRURGIA + RADIO + HORMONIO	0.27	0.37	0.32	527
HORMONIO	0.95	0.82	0.88	527
NENHUM	0.74	0.56	0.63	526
OUTROS	0.51	0.44	0.47	527

QUIMIO + OUTRAS MODALIDADES	0.49	0.60	0.54	526
RADIO	0.84	0.59	0.69	526
RADIO + HORMONIO	0.98	0.76	0.86	527

Tabela 10. Resultados *XGBoost*

8. Conclusão

Em síntese, a performance dos modelos variou entre 53% e 60% de acurácia, apresentando melhor desempenho na identificação de classes de tratamento únicas ou mais representadas antes do balanceamento, como CIRURGIA, HORMONIO, RADIO e RADIO + HORMONIO. Por outro lado, os modelos enfrentaram maior dificuldade em prever rótulos que envolvem combinações de tratamentos menos frequentes, como CIRURGIA + HORMÔNIO, CIRURGIA + RADIOTERAPIA e CIRURGIA + RADIOTERAPIA + HORMÔNIO.

Em continuidade, modelos baseados em conjuntos de árvores apresentaram melhor desempenho, com destaque para o *XGBoost*, seguido pela Floresta Aleatória, por lidarem melhor com variáveis categóricas e múltiplas colunas. Em contrapartida, algoritmos como KNN e *Naive Bayes* obtiveram resultados inferiores, possivelmente por serem mais sensíveis ao aumento da dimensionalidade dos dados.

Apesar dos avanços, a performance dos modelos ainda é considerada insatisfatória para aplicação direta em contextos clínicos. Como próximos passos, sugere-se o estudo e a implementação de uma arquitetura de Rede Neural que possa capturar padrões mais complexos nos dados. Além disso, etapas anteriores de análise e pré-processamento devem ser reavaliadas, a fim de extrair novas percepções, assim como é recomendada a busca por bases de dados adicionais que complementem as classes com menor representatividade.

Concluiu-se, assim, o escopo previsto para o projeto, incluindo análise exploratória, pré-processamento dos dados, treinamento dos modelos e avaliação por métricas. Com as melhorias sugeridas, espera-se que o sistema possa futuramente auxiliar profissionais da saúde na sugestão de tratamentos, oferecendo apoio à decisão médica por meio da análise de grandes volumes de informações complexas, contribuindo para novos *insights*.

Dessa forma, almeja-se um tratamento mais assertivo, aumentando as chances de sucesso terapêutico, reduzindo efeitos colaterais decorrentes de intervenções excessivas ou ineficazes, e promovendo uma melhor qualidade de vida para os pacientes, familiares, amigos e demais envolvidos no processo.

Referências

- AMERICAN CANCER SOCIETY. What is cancer. 31 mar. 2025. Disponível em: <https://www.cancer.org/cancer/understanding-cancer/what-is-cancer.html>. Acesso em: 15 abr. 2025.
- AMERICAN CANCER SOCIETY. What is prostate cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/about/what-is-prostate-cancer.html>. Acesso em: 15 abr. 2025.

- AMERICAN CANCER SOCIETY. Signs and Symptoms of Prostate Cancer. 01 jan. 2025. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/detection-diagnosis-staging/signs-symptoms.html>. Acesso em: 18 abr. 2025.
- AMERICAN CANCER SOCIETY. Tests to Diagnose and Stage Prostate Cancer. 21 mar. 2025. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/detection-diagnosis-staging/how-diagnosed.html>. Acesso em: 18 abr. 2025.
- AMERICAN CANCER SOCIETY. Observation or Active Surveillance for Prostate Cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/watchful-waiting.html>. Acesso em: 21 abr. 2025.
- AMERICAN CANCER SOCIETY. Surgery for Prostate Cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/surgery.html>. Acesso em: 21 abr. 2025.
- AMERICAN CANCER SOCIETY. Hormone Therapy for Prostate Cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/hormone-therapy.html>. Acesso em: 21 abr. 2025.
- AMERICAN CANCER SOCIETY. Immunotherapy for Prostate Cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/vaccine-treatment.html>. Acesso em: 21 abr. 2025.
- AMERICAN CANCER SOCIETY. Chemotherapy for Prostate Cancer. 22 nov. 2023. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/chemotherapy.html>. Acesso em: 21 abr. 2025.
- AMERICAN CANCER SOCIETY. Radiation Therapy for Prostate Cancer. 28 mar. 2025. Disponível em: <https://www.cancer.org/cancer/types/prostate-cancer/treating/surgery.html>. Acesso em: 21 abr. 2025.
- ARAÚJO, Wesley Batista Dominices de. Método de auxílio ao diagnóstico de câncer de próstata utilizando aprendizado de máquina e dados clínicos. 2024. 141 f. Tese (Programa de Pós-graduação em Engenharia de Eletricidade/CCET) - Universidade Federal do Maranhão, São Luís, 2024.
- AWS. O que é regressão logística. 2025. Disponível em: <https://aws.amazon.com/pt/what-is/logistic-regression/>. Acesso em 31 maio 2025.
- BRASIL. Conselho Federal de Enfermagem. Inteligência artificial chega à saúde. 11 mar. 2024. Disponível em: <https://www.cofen.gov.br/a-inteligencia-artificial-chega-a-saude/>. Acesso em: 06 abr. 2025.
- BRASIL. Ministério da Saúde. Câncer de próstata. 2020. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/c/cancer-de-prostata>. Acesso em: 20 mar. 2025.
- BRASIL. Ministério da Saúde. 17/11 - Dia Nacional de Combate ao Câncer de Próstata. 2022. Disponível em: <https://bvsms.saude.gov.br/17-11-dia-nacional-de-combate-ao-cancer-de-prostata/#:~:text=De%20acordo%20com%20dados%20do,devido%20ao%20c%C3%A2ncer%20de%20pr%C3%B3stata>. Acesso em: 20 mar. 2025.
- BRASIL. Sociedade Brasileira de Medicina Tropical. Revolução da inteligência artificial: uso na saúde traz novas possibilidades. 10 maio 2023. Disponível em:

- <https://sbmt.org.br/revolucao-da-inteligencia-artificial-uso-na-saude-traz-novas-possibilidades/>. Acesso em: 06 abr. 2025.
- CONTE, Juliana. O que significa o escore de Gleason na biópsia da próstata?. UOL. 27 jan. 2025. Disponível em: <https://drauziovarella.uol.com.br/cancer/o-que-significa-o-escore-de-gleason-na-biopsia-da-prostata/>. Acesso em: 23 abr. 2025.
- FUNDAÇÃO ONCOCENTRO DE SÃO PAULO. TNM – Classificação de Tumores Malignos. 2022. Disponível em: <https://fosp.saude.sp.gov.br/fosp/diretoria-adjunta-de-informacao-e-epidemiologia/tnm/>. Acesso em: 22 abr. 2025.
- HAMPL, Daniel. O que pode alterar o PSA?. 2024. Disponível em: <https://danielhampl.com.br/glossario/o-que-pode-alterar-o-psa/>. Acesso em: 18 abr. 2025.
- IBM. O que é o algoritmo k-nearest neighbors (KNN)? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/knn>. Acesso em 31 maio 2025.
- IBM. O que é uma árvore de decisão? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/decision-trees>. Acesso em 31 maio 2025.
- IBM. O que é random forest? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/random-forest>. Acesso em 31 maio 2025.
- IBM. O que é o XGBoost? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/xgboost>. Acesso em 31 maio 2025.
- IMBALANCED LEARN. Cluster Centroids. 2025. Disponível em: https://imbalanced-learn.org/stable/references/generated/imblearn.under_sampling.ClusterCentroids.html. Acesso em 31 maio 2025.
- JAMES, Nicholas; TANNOCK, Ian; N'DOW, Jame. et al. The Lancet Commission on prostate cancer: planning for the surge in cases. The Lancet. 27 abr. 2024. Disponível em: [https://www.thelancet.com/journals/lancet/article/PIIS01406736\(24\)00651-2/fulltext](https://www.thelancet.com/journals/lancet/article/PIIS01406736(24)00651-2/fulltext). Acesso em: 25 mar. 2025.
- KAVLAKOGLU, Eda; RUSSI, Erica. IBM. O que são classificadores Naive Bayes? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/naive-bayes>. Acesso em 31 maio 2025.
- KAVLAKOGLU, Eda; WINLAND, Vanna. IBM. O que é agrupamento k-means? 2025. Disponível em: <https://www.ibm.com/br-pt/think/topics/k-means-clustering>. Acesso em 31 maio 2025.
- LINHARES, Rafaela. Tecidos do corpo humano. KENHUB. 30 out. 2023. Disponível em: <https://www.kenhub.com/pt/library/anatomia/tecidos-do-corpo-humano>. Acesso em: 15 abr. 2025.
- MAEDA, André Eidi; CROCCO, Pedro Ferreira; RUPPERT, Guilherme Cesar Soares. et al. Um estudo sobre a predição da recidiva de câncer usando técnicas de aprendizado de máquina. In: JORNADA DE INICIAÇÃO CIENTÍFICA DO CENTRO DE TECNOLOGIA DA INFORMAÇÃO RENATO ARCHER, 24, 2022, São Paulo. Anais [...]. Campinas: JICC, 2022. Disponível em: <https://www.gov.br/cti/pt-br/publicacoes/producao-cientifica/jicc/xxiv-jicc-2022/pdf/jicc-2022-paper-4.pdf>. Acesso em: 20 abr. 2025.

- MICROSOFT. Métricas da avaliação. 2024. Disponível em: <https://learn.microsoft.com/pt-br/azure/ai-services/language-service/custom-text-classification/concepts/evaluation-metrics>. Acesso em: 31 maio 2025.
- ONCOGUIA. A próstata. 16 abr. 2024. Disponível em: <https://www.oncoguia.org.br/conteudo/a-prostata/770/149/#:~:text=Em%20homens%20mais%20jovens%2C%20tem,%2C%20tornando%20do%20mais%20l%C3%ADquido>. Acesso em: 15 abr. 2025.
- ONCOGUIA. Exames de rastreamento para câncer de próstata. 16 mar. 2024. Disponível em: <https://www.oncoguia.org.br/conteudo/exames-de-rastreamento-para-cancer-de-prostata/5856/289/>. Acesso em 18 abr. 2025.
- OPENSTAX COLLEGE. The cell cycle. 27 maio 2016. Disponível em: <https://openstax.org/books/biology/pages/10-2-the-cell-cycle>. Acesso em 15 abr. 2025.
- PAN, Huiqing; WANG, Jiayi; SHI, Weizhong. et al. Quantified treatment effect at the individual level is more indicative for personalized radical prostatectomy recommendation: implications for prostate cancer treatment using deep learning. *Journal of Cancer Research and Clinical Oncology*. 04 jan. 2024. DOI 10.1007/s00432-023-05602-4. Disponível em: <https://doi.org/10.1007/s00432-023-05602-4>. Acesso em: 22 mar. 2025.
- RUSSEL, Stuart Jonathan; NORVIG, Peter. Inteligência artificial. 3. ed. Rio de Janeiro: Elsevier, 2013.
- SCIKIT-LEARN. Logistic Regression. 2025. Disponível em: https://scikit-learn.org/stable/modules/linear_model.html#logistic-regression. Acesso em: 31 maio 2025.
- SOBIN, L. H.; GOSPODAROWICZ, M. K.; WITTEKIND, Ch. TNM – Classification of Malignant Tumours. 7. ed. Reino Unido: Wiley-Blackwell, 2009.
- TAKANO, Luiz. Escala de Gleason: entenda seu significado no diagnóstico do câncer de próstata. 01 jul. 2020. Disponível em: <https://www.takanourologia.com.br/blog/escala-de-gleason-entenda-seu-significado-no-diagnostico-do-cancer-de-prostata/>. Acesso em: 23 abr. 2024.
- TANG, Shengsheng; ZHANG, Hongzheng; LIANG, Junhao. et al. Prostate cancer treatment recommendation study based on machine learning and SHAP interpreter. *Cancer Science*. 03 nov. 2024. Disponível em: <https://onlinelibrary.wiley.com/doi/full/10.1111/cas.16327>. Acesso em: 22 mar. 2025.
- UNIVERSIDADE DE STANFORD. Softmax Regression. 2025. Disponível em: <http://ufldl.stanford.edu/tutorial/supervised/SoftmaxRegression/>. Acesso em: 31 maio 2025.
- VICTOR, Natha. Ministério da Saúde. Grupo de Trabalho da Saúde analisa uso da inteligência artificial na assistência médica. 04 jun. 2024. Disponível em: <https://www.gov.br/saude/pt-br/assuntos/noticias/2024/junho/grupo-de-trabalho-da-saude-analisa-uso-da-inteligencia-artificial-na-assistencia-medica>. Acesso em: 06 abr. 2025.
- WOODCOCK, Dan J.; SAHLI, Atef; TESLO, Ruxandra. et al. Genomic evolution shapes

prostate cancer disease type. Cell Genomics. 13 mar. 2024. Disponível em: [https://www.cell.com/cell-genomics/fulltext/S2666-979X\(24\)00038-7](https://www.cell.com/cell-genomics/fulltext/S2666-979X(24)00038-7). Acesso em: 01 abr. 2025.