SISTEMA DE SUPORTE À DECISÃO CLÍNICA BASEADO EM ÁRVORE DE DECISÃO PARA CLASSIFICAÇÃO DE RISCO DE CÂNCER DE PRÓSTATA

Bruno Olímpio Garcia Lorca, Davi Oliveira da Cunha, Kaio Silva Castilho, Wagner dos Santos Clementino de Jesus.

Colégio Técnico Antônio Teixeira Fernandes – Colégio Univap Centro, Rua Paraibuna, 75, Jardim São Dimas - 12245-020 - São José dos Campos-SP, Brasil, bruno.olimpio.lorca@gmail.com, davioliveiradacunha26@gmail.com, kaiocastlho900@gmail.com, wsantoscj@gmail.com.

Resumo

O câncer de próstata é o segundo tipo de tumor mais frequente entre os homens no Brasil, somando dezenas de milhares de novos casos a cada ano, sendo o Antígeno Prostático Específico (PSA) o principal marcador de rastreamento, ainda que órgão-específico e não doença-específico, com acurácia reduzida na zona cinzenta (4,0 a 10,0 ng/mL), faixa em que apenas 25% a 33% das biópsias indicadas confirmam malignidade. Diante disso, desenvolveu-se um sistema de suporte à decisão clínica baseado em Árvore de Decisão para classificar o risco de câncer de próstata (Benigno ou Suspeito) a partir do PSA Total, do PSA Livre, da Densidade do PSA e da idade, treinado sobre dados sintéticos e integrado a uma interface em C#/WPF. Os resultados mostraram acurácia de 97,53% e *recall* de 99,26% para a classe Suspeito, superando os 87,17% relatados na literatura com dados reais. O modelo foi otimizado para o *recall*, dado o maior custo de falsos negativos em triagem oncológica. Conclui-se que o modelo, avaliado apenas com dados sintéticos, é interpretável e promissor como apoio à triagem urológica na zona cinzenta do PSA, devendo ser testado futuramente com dados reais.

Palavras-chave: Antígeno Prostático Específico. Árvore de Decisão. Aprendizado de Máquina. Sistema de Suporte à Decisão Clínica. Zona Cinzenta do PSA.

Área do Conhecimento: Técnico.

Introdução

O câncer de próstata é um dos problemas de saúde pública mais relevantes entre a população masculina, respondendo por cerca de 72 mil novos casos por ano no Brasil (INCA, 2022) e figurando como o segundo câncer sólido mais diagnosticado em homens globalmente (Rawla, 2019). O Antígeno Prostático Específico (PSA, do inglês *Prostate-Specific Antigen*) consolidou-se como o principal exame de rastreamento, mas constitui um marcador órgão-específico, e não doença-específico: níveis elevados podem decorrer tanto do câncer quanto de condições benignas, como a Hiperplasia Prostática Benigna (Corrêa *et al.*, 2003). Essa sobreposição é particularmente crítica na zona cinzenta (PSA Total entre 4,0 e 10,0 ng/mL), em que apenas 25% a 33% das biópsias indicadas confirmam malignidade (Corrêa *et al.*, 2003; Santos *et al.*, 2022), gerando biópsias desnecessárias e riscos evitáveis de dor, sangramento e infecção.

Diante dessa limitação, a literatura tem investigado o emprego de técnicas de aprendizado de máquina como suporte à decisão clínica, com modelos que superam o uso isolado do PSA na predição de malignidade prostática (Chen *et al.*, 2022). Entre esses algoritmos, a Árvore de Decisão destaca-se por sua interpretabilidade: o processo classificatório é expresso por regras explícitas de “se-então”, compreensíveis por profissionais de saúde sem formação em computação (Rezende, 2003). A aplicação do algoritmo J48 de Árvore de Decisão a dados reais de 530 pacientes do Hospital das Clínicas da Universidade Estadual de Campinas (UNICAMP) permitiu obter taxa de classificação correta de 87,17% ao subdividir o PSA em quatro intervalos séricos (Reis, 2023). Em outro estudo, obteve-se acurácia de 80% com redes neurais artificiais aplicadas a variáveis clínicas prostáticas (Araujo *et al.*, 2024) — resultado inferior ao obtido com dados reais por meio de Árvore de Decisão (Reis, 2023), apesar da maior capacidade representacional das redes neurais, reforçando o potencial de modelos arbóreos interpretáveis nesse contexto clínico.

Este trabalho apresenta um sistema de suporte à decisão clínica fundamentado em um modelo de Árvore de Decisão, com o objetivo geral de classificar o risco de câncer de próstata — como Benigno ou Suspeito — a partir de parâmetros analíticos obtidos via quimioluminescência (PSA Total, PSA Livre, Densidade do PSA e idade), buscando aumentar a precisão interpretativa em relação aos métodos de triagem convencionais, com ênfase no desempenho do modelo especificamente na zona cinzenta, faixa em que a interpretação clínica isolada do PSA é mais complexa.

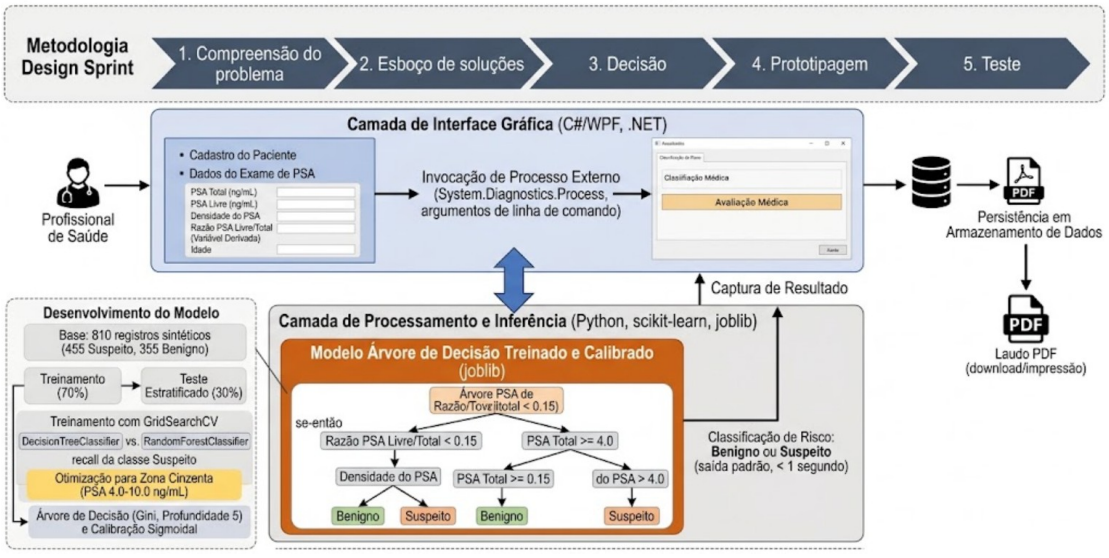
De modo específico, o trabalho busca: estruturar um banco de dados com as variáveis críticas de ensaios de quimioluminescência, com tratamento de ruídos e normalização; implementar a Árvore de Decisão via bibliotecas de aprendizado de máquina em Python, como o scikit-learn; avaliar o desempenho do modelo na zona cinzenta do PSA, validando a eficácia da razão PSA Livre/Total no processo decisório; e mensurar a performance por métricas de classificação — acurácia, precisão, *recall* (sensibilidade) e *F1-score* —, etapa preliminar a uma futura comparação com diagnósticos padrão-ouro e com as diretrizes da Sociedade Brasileira de Urologia, a ser conduzida com dados clínicos reais.

Metodologia

O desenvolvimento do sistema seguiu o método ágil *Design Sprint* (Knapp; Zeratsky; Kowitz, 2017), que comprime o ciclo de criação de uma solução em cinco fases sequenciais — compreensão do problema, esboço de soluções, decisão, prototipagem e teste. A arquitetura resultante opera em duas camadas complementares: uma camada de processamento e inferência, desenvolvida em Python (Python Software Foundation, 2024), responsável pelo aprendizado de máquina; e uma camada de interface gráfica, desenvolvida em C# (Microsoft, 2024a) com o framework Windows Presentation Foundation (WPF) (Microsof't, 2024b), responsável por apresentar os resultados ao profissional de saúde. A comunicação entre as duas camadas é realizada por invocação de processo externo, pelo pacote *System.Diagnostics* do ecossistema .NET — mais especificamente de sua classe *Process*: a interface C# inicia o interpretador Python como subprocesso, transmite os parâmetros clínico-laboratoriais do paciente como argumentos de linha de comando e captura, pela saída padrão, a classificação de risco retornada, com tempo total inferior a um segundo em todos os cenários testados.

A Figura 1 sintetiza esse fluxo de ponta a ponta, das cinco fases do *Design Sprint* à arquitetura de operação e ao *pipeline* de treinamento: o profissional de saúde insere o cadastro do paciente e os dados do exame de PSA na interface, que aciona a classe *Process* para invocar o módulo Python; a Árvore de Decisão treinada e calibrada classifica o risco como Benigno ou Suspeito; o resultado retorna para avaliação médica; os registros são persistidos em armazenamento de dados; e o laudo é gerado em arquivo no formato Portable Document Format (PDF) para *download* e impressão.

Figura 1 – Fluxo de concepção, arquitetura e treinamento do sistema, das fases do *Design Sprint* à geração do laudo em PDF.



Fonte: Autores, 2026.

Os valores de PSA Total e PSA Livre têm origem na quimioluminescência, técnica de imunoensaio que emprega reações antígeno-anticorpo marcadas com substâncias quimioluminescentes, gerando emissão de luz proporcional à concentração do analito na amostra sérica, com especificidade analítica superior a 99% (Pereira; Bertollo; Zarife, 2010). A dosagem simultânea do PSA Total e do PSA Livre permite calcular a razão PSA Livre/Total — variável derivada de especial relevância na zona cinzenta, uma vez que valores inferiores a 15% associam-se clinicamente a maior probabilidade de malignidade (Oliveira *et al.*, 2014). O modelo foi treinado sobre 810 registros clínico-laboratoriais (455 Suspeito, 355 Benigno), com as variáveis PSA Total, PSA Livre, razão PSA Livre/Total, Densidade do PSA e idade, submetidas a pré-processamento e normalização.

A definição do algoritmo de classificação decorreu de um processo empírico e comparativo: dois modelos candidatos — o *DecisionTreeClassifier* e o *RandomForestClassifier*, ambos do scikit-learn (Pedregosa *et al.*, 2011) — foram avaliados sob condições estritamente equivalentes por meio do *GridSearchCV*, com busca de hiperparâmetros de profundidade, critério de divisão e, para a Random Forest, número de estimadores, sob validação cruzada estratificada de cinco partições. Ambos os classificadores foram configurados com *class_weight='balanced'*, mitigando o desbalanceamento moderado entre classes, e otimizados pelo *recall* da classe Suspeito — e não pela acurácia geral —, uma vez que, em triagem oncológica, o custo de um falso negativo supera substancialmente o de um falso positivo (Géron, 2021). A Árvore de Decisão e a Random Forest atingiram exatamente o mesmo *recall* médio sob validação cruzada (98,12%), configurando um empate técnico; como o *pipeline* elege o primeiro modelo a atingir o melhor escore, a Árvore de Decisão — com critério Gini e profundidade máxima de cinco níveis — foi selecionada como modelo-base, preservando a interpretabilidade defendida por Rezende (2003) sem custo de desempenho preditivo frente ao algoritmo de *ensemble* de Breiman (2001).

O modelo vencedor foi então submetido à calibração probabilística pelo módulo *CalibratedClassifierCV*, com o método sigmoidal (escalonamento de Platt) sob validação cruzada de cinco partições, ajustando as probabilidades emitidas para refletir mais fielmente o risco real de cada desfecho. O classificador calibrado é persistido em disco por meio da biblioteca joblib (Joblib Developers, 2025), sendo carregado pelo módulo de inferência a cada nova predição sem necessidade de novo treinamento.

A avaliação quantitativa do modelo foi conduzida sobre uma partição de teste estratificada, correspondente a 30% da base de dados (243 de 810 registros), inteiramente reservada e não utilizada em nenhuma etapa do treinamento ou da otimização de hiperparâmetros. Complementarmente, a avaliação funcional baseou-se em cenários de teste sintéticos, porém clinicamente plausíveis, cobrindo PSA normal, PSA elevado e zona cinzenta (favorável e desfavorável). Cabe destacar que, em razão das restrições éticas e regulatórias associadas ao acesso a prontuários hospitalares reais, todos os dados utilizados neste trabalho — tanto no treinamento do modelo quanto na avaliação funcional — são sintéticos e não envolvem prontuários ou informações de pacientes reais; o presente estudo não se enquadra, portanto, como pesquisa envolvendo seres humanos para fins de submissão a Comitê de Ética em Pesquisa (CEP), ficando a validação estatística plena com dados clínicos reais e anotados por especialistas registrada como trabalho futuro, conforme discutido a seguir.

O desempenho do classificador foi mensurado por meio de métricas de classificação calculadas com os módulos *accuracy_score*, *classification_report*, *confusion_matrix* e *recall_score* do scikit-learn (Pedregosa *et al.*, 2011), aplicados sobre as predições emitidas para a partição de teste. A acurácia corresponde à proporção de acertos totais — (VP + VN) / (VP + VN + FP + FN); a precisão de cada classe é dada por VP / (VP + FP), representando a confiabilidade de uma predição positiva; o *recall*, ou sensibilidade, corresponde a VP / (VP + FN), medindo a capacidade do modelo de identificar corretamente os casos efetivamente pertencentes à classe; e o *F1-score* sintetiza precisão e *recall* em uma única métrica, calculada pela média harmônica 2 × (precisão × *recall*) / (precisão + *recall*), especialmente útil na presença de desbalanceamento entre classes. Essas métricas foram apuradas tanto de forma agregada (acurácia geral) quanto por classe (precisão, *recall* e *F1-score* de Benigno e Suspeito), permitindo avaliação granular do comportamento do classificador diante dos dois tipos de erro possíveis.

Resultados

A execução do *pipeline* sobre a partição de teste produziu acurácia geral de 97,53% e *recall* de 99,26% para a classe Suspeito — métrica priorizada na seleção do modelo, conforme justificado na Metodologia. A Tabela 1 apresenta a matriz de confusão resultante, obtida sobre os 243 casos do conjunto de teste, na qual VN (Verdadeiro Negativo) e VP (Verdadeiro Positivo) representam os acertos do modelo para as classes Benigno e Suspeito, respectivamente, enquanto FP (Falso Positivo) e FN (Falso Negativo) representam os dois tipos de erro de classificação possíveis.

Tabela 1 – Matriz de confusão do classificador na partição de teste.

Classe real	**Previsto: Benigno**	**Previsto: Suspeito**	**Total**
Benigno	102 (VN)	5 (FP)	107
Suspeito	1 (FN)	135 (VP)	136
Total	103	140	243

Fonte: Autores, 2026.

A leitura da matriz evidencia uma assimetria proposital nos erros do modelo: dos 107 casos efetivamente Benignos, 102 foram corretamente classificados e apenas 5 foram erroneamente indicados como Suspeitos (falsos positivos); já entre os 136 casos efetivamente Suspeitos, 135 foram corretamente identificados e um único caso não foi detectado (falso negativo). Essa distribuição reflete diretamente a estratégia de otimização pelo *recall* descrita na Metodologia: como, em triagem oncológica, deixar de identificar um caso potencialmente maligno é clinicamente mais grave do que encaminhar desnecessariamente um paciente benigno para reavaliação, o modelo foi deliberadamente otimizado para errar, quando erra, a favor da cautela.

Esses mesmos valores permitem derivar a precisão, o *recall* e o *F1-score* por classe — métricas que, em conjunto, avaliam não apenas quantos casos o modelo acerta, mas também o tipo de erro que comete em cada classe. A Tabela 2 sintetiza esses indicadores.

Tabela 2 – Métricas de desempenho do classificador por classe.

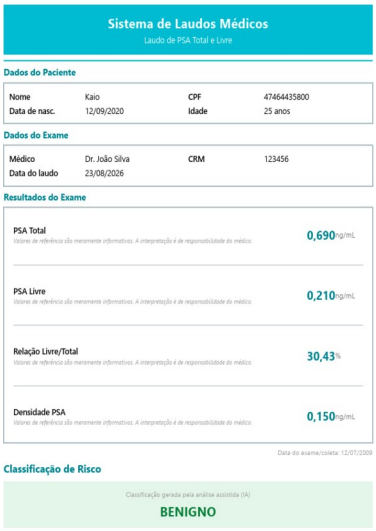
Classe	**Precisão**	***Recall* (Sensibilidade)**	***F1-Score***
Benigno	99,03%	95,33%	0,97
Suspeito	96,43%	99,26%	0,98

Fonte: Autores, 2026.

Os números confirmam o padrão observado na matriz de confusão: a classe Suspeito apresenta o maior *recall* (99,26%), refletindo a baixíssima taxa de falsos negativos, enquanto a classe Benigno apresenta a maior precisão (99,03%), refletindo o único caso Suspeito entre os 103 preditos como Benignos. Os F1-scores próximos de 1 (0,97 e 0,98) indicam que essa priorização do *recall* não comprometeu significativamente a precisão geral do modelo. O resultado é reforçado pelo fato de a Árvore de Decisão e a Random Forest terem atingido exatamente o mesmo *recall* médio sob validação cruzada (98,12%) — evidência de que o desempenho obtido não é uma peculiaridade de um único algoritmo, mas um padrão consistente entre dois modelos de naturezas estruturais distintas.

Na zona cinzenta, cenários com razão PSA Livre/Total favorável (acima de 15%) foram classificados como Benigno, e desfavoráveis (abaixo de 15%) como Suspeito, com PSA Total situado na própria zona cinzenta — comportamento que um corte estático de 4,0 ng/mL não reproduziria. A Figura 2 ilustra o laudo emitido automaticamente pela interface ao final do fluxo de atendimento, disponibilizado em PDF para *download* e impressão, com os valores laboratoriais inseridos, a razão PSA Livre/Total calculada e a classificação de risco produzida pelo modelo.

Figura 2 – Laudo de PSA gerado automaticamente pelo sistema, com classificação de risco emitida pelo modelo de Árvore de Decisão.



Fonte: Autores, 2026.

Discussão

O desempenho do sistema excede os trabalhos correlatos discutidos na Introdução: os 97,53% de acurácia superam os 87,17% relatados com dados reais da UNICAMP (Reis, 2023) e os 80% obtidos com redes neurais artificiais (Araujo *et al.*, 2024). Cabe cautela: esses resultados (Reis, 2023; Araujo *et al.*, 2024) advêm de dados reais, ao passo que os do presente trabalho decorrem de base sintética, ainda que clinicamente fundamentada. Ainda assim, reforça-se a adequação de um modelo arbóreo interpretável frente a abordagens de caixa-preta (Rezende, 2003), corroborada pelo empate técnico entre Árvore de Decisão e Random Forest (Breiman, 2001).

O presente estudo reconhece limitações. Em primeiro lugar, a avaliação apoiou-se exclusivamente em dados sintéticos, não em uma coorte real com diagnóstico histopatológico confirmado: os valores relatados atestam a correção do *pipeline*, mas não permitem inferir o desempenho em pacientes reais, o que exigirá avaliação futura com dados clínicos reais. Em segundo lugar, o modelo limita-se às variáveis PSA Total, PSA Livre, Densidade do PSA e idade, sem incorporar descritores anatomopatológicos de maior poder preditivo, como o Escore de Gleason (Reis, 2023).

Conclusão

Como prova de conceito computacional sobre a base de dados sintética utilizada, o sistema demonstrou que um modelo de Árvore de Decisão fundamentado no PSA Total, no PSA Livre, na Densidade do PSA e na idade do paciente é capaz de classificar o risco de câncer de próstata com acurácia de 97,53% e *recall* de 99,26% para a classe Suspeito, cumprindo o objetivo proposto de oferecer uma alternativa interpretável e auditável aos cortes binários estáticos tradicionalmente empregados na triagem urológica, com comportamento particularmente relevante na zona cinzenta do PSA.

Como trabalhos futuros, projetam-se: a avaliação com dados reais de pacientes, mediante aprovação prévia do CEP; a expansão das variáveis preditivas com o Escore de Gleason e o estadiamento clínico; a adoção de *hashing* criptográfico para credenciais; e a integração a prontuários eletrônicos via HL7/FHIR.

Referências

ARAUJO, W. B. D. de; SANTANA, E. E. C.; SILVA, N. P.; SOUSA JUNIOR, C. M.; MOURA, G. L.; SANTOS, J. A. L. M. dos; LOPES, P. L. A.; SILVA, W. do N.; GONÇALVES, J. P. P.; SILVA, F. C. B. R. **Rede neural artificial aplicada ao diagnóstico de câncer de próstata.** Journal of Health Informatics, v. 16, n. Especial, 2024.

BREIMAN, L. **Random forests.** Machine Learning, v. 45, n. 1, p. 5-32, 2001. DOI: 10.1023/A:1010933404324.

CHEN, S.; JIAN, T.; CHI, C.; LIANG, Y.; LIANG, X.; YU, Y.; JIANG, F.; LU, J. **Machine learning-based models enhance the prediction of prostate cancer.** Frontiers in Oncology, v. 12, art. 941349, 2022. DOI: 10.3389/fonc.2022.941349.

CORRÊA, N. A. B.; COSTA, G. F. M.; MASSAMBANI, E. M.; MATUMOTO, F. H.; PAULA, M. M. M. **Diagnóstico precoce de carcinoma de próstata: antígeno prostático específico (PSA), um marcador quase ideal.** Revista Brasileira de Análises Clínicas, v. 35, n. 2, p. 63-64, 2003.

GÉRON, A. **Mãos à obra: aprendizado de máquina com Scikit-Learn, Keras & TensorFlow.** 2. ed. Rio de Janeiro: Alta Books, 2021.

INSTITUTO NACIONAL DE CÂNCER (INCA). **Estimativa 2023: incidência de câncer no Brasil.** Rio de Janeiro: INCA, 2022.

JOBLIB DEVELOPERS. **Joblib: running Python functions as pipeline jobs.** [S. l.: s. n.], 2025.

KNAPP, J.; ZERATSKY, J.; KOWITZ, B. **Sprint: o método usado no Google para testar e aplicar novas ideias em apenas cinco dias.** Tradução de André Anhê. Rio de Janeiro: Intrínseca, 2017.

MICROSOFT. **Guia do C#.** Redmond: Microsoft, 2024a.

MICROSOFT. **O que é o Windows Presentation Foundation (WPF).** Redmond: Microsoft, 2024b.

OLIVEIRA, G. R. C.; MACHADO, M. J. A.; COSTA, S. H. N. et al. **Avaliação dos níveis de PSA livre e total em pacientes com diferentes faixas etárias atendidos no laboratório clínico PUC-Goiás em outubro de 2013.** Estudos, Goiânia, v. 41, n. 3, p. 507-513, 2014.

PEDREGOSA, F.; VAROQUAUX, G.; GRAMFORT, A.; MICHEL, V.; THIRION, B.; GRISEL, O.; BLONDEL, M.; PRETTENHOFER, P.; WEISS, R.; DUBOURG, V.; VANDERPLAS, J.; PASSOS, A.; COURNAPEAU, D.; BRUCHER, M.; PERROT, M.; DUCHESNAY, E. **Scikit-learn: machine learning in Python.** Journal of Machine Learning Research, v. 12, p. 2825-2830, 2011.

PEREIRA, F. M.; BERTOLLO, L. A.; ZARIFE, M. A. S. **Comparação de dois testes automatizados por quimioluminescência para a detecção de anticorpos contra o vírus da hepatite C.** Revista Pan-Amazônica de Saúde, Ananindeua, v. 1, n. 4, p. 17-21, 2010.

PYTHON SOFTWARE FOUNDATION. **Python 3 documentation.** Wilmington: Python Software Foundation, 2024.

RAWLA, P. **Epidemiology of prostate cancer.** World Journal of Oncology, v. 10, n. 2, p. 63-89, 2019.

REIS, N. F. **Modelagem matemática para o crescimento de tumores e sua associação aos níveis séricos de marcadores tumorais: uma análise do Antígeno Prostático Específico em adenocarcinomas de próstata.** 2023. Dissertação (Mestrado em Matemática Aplicada) – Instituto de Matemática, Estatística e Computação Científica, Universidade Estadual de Campinas, Campinas, 2023.

REZENDE, S. O. **Sistemas inteligentes: fundamentos e aplicações.** Barueri: Manole, 2003.

SANTOS, R. O. M. dos; ABREU, M. M. de; MIGOWSKI, A.; ENGSTROM, E. M. **Ferramenta de apoio à decisão sobre o rastreamento do câncer de próstata no Brasil.** Revista de Saúde Pública, v. 56, p. 19, 2022.