

Dupla predição baseada em Aprendizado de Máquina para priorização de pacientes com câncer de mama no Sistema Único de Saúde

Samantha Mendonça¹, Bianca Costa¹, Álef Monteiro¹, Abraão Paixão¹, Italo P. Caliari¹, Hendrio Bragança¹, Wington L. Vital¹, Antonio P. Fontenele¹

¹Núcleo de Capacitação em Inteligência Artificial (NCIA), Fundação Paulo Feitoza - FPFtech, Av. Danilo de Matos Areosa, 1170, Distrito Industrial, Manaus, Amazonas, 69075-351, Brazil.

{samantha.252329, bianca.252310, alef.220364, abraao.252332,
italo.caliari, hendrio.braganca, wington.vital,
antonio.fontenele}@fpf.br

Abstract. *Breast cancer is the most frequent neoplasm among Brazilian women, facing bottlenecks in compliance with Law No. 12,732/2012. This study proposes a dual-prediction framework as a proof of concept, separately evaluating: (i) the biological risk of malignancy (Wisconsin Diagnostic dataset) and (ii) the healthcare risk of treatment delay (DATASUS data). Integration via Priority Score is proposed for future application in real care workflows. In malignancy prediction, the SVM model achieved 98.2% accuracy. Regarding the healthcare axis, a Voting Classifier obtained an AUC-ROC of 0.6439 and a lift of 1.29×, identifying groups with a delay rate of 85.63%. Exploratory analyses (by State/UF and CNES) and model stratification indicate that delay risk is strongly linked to institutional and geographical contexts. It is concluded that the Priority Score can support proactive queue regulation in the SUS, prioritizing cases with higher operational vulnerability.*

Resumo. *O câncer de mama é a neoplasia mais comum entre mulheres no Brasil, enfrentando gargalos institucionais no cumprimento da Lei nº 12.732/2012. Este estudo propõe um framework de dupla predição como prova de conceito, analisando dois eixos em bases públicas distintas: (i) risco biológico de malignidade (base Breast Cancer Wisconsin Diagnostic) e (ii) risco assistencial de atraso no tratamento (Painel Oncologia/DATASUS). Propõe-se a integração via Priority Score para cenários futuros de vinculação de dados clínicos e administrativos. Na predição de malignidade, o modelo SVM alcançou 98,2% de acurácia. No eixo assistencial, um ensemble Voting Classifier obteve AUC-ROC de 0,6439 e lift de 1,29×, identificando grupos com taxa de atraso de 85,63%. As análises por UF e CNES indicam que o risco de atraso é fortemente associado ao contexto geográfico e institucional. Assim, o Priority Score proposto pode auxiliar a regulação de filas no SUS, priorizando casos com maior vulnerabilidade operacional.*

1. Introdução

O câncer de mama origina-se de diversos tecidos mamários e possui alto potencial metastático, representando um grave desafio de saúde pública [Verras et al. 2023]. No Brasil, é a neoplasia feminina mais incidente, com estimativa de 73.610 novos casos anuais e forte heterogeneidade regional; desfechos favoráveis dependem tanto do diagnóstico preciso quanto da celeridade no tratamento [Instituto Nacional de Câncer 2022] [Freihat et al. 2025] [Soliman et al. 2024].

Embora a Lei nº 12.732/2012 estabeleça o limite de 60 dias para o início do tratamento no SUS, o Painel Oncologia revela dificuldades sistêmicas, com atrasos superiores a 60% devido a gargalos operacionais e desigualdades regionais [de Medeiros et al. 2020] [Gomes 2013] [Instituto Nacional de Câncer 2020].

A literatura de IA na área foca majoritariamente na perspectiva "micro", como a classificação de malignidade via características morfométricas [Uchikov et al. 2024] [Ghantasala et al. 2023] [S et al. 2022], negligenciando o risco assistencial de atraso e barreiras estruturais pós-diagnóstico [Soliman et al. 2024].

Este trabalho propõe um *framework* modular de dupla predição: os riscos biológico e assistencial são avaliados em bases distintas. A contribuição principal reside na integração em nível de decisão (*Priority Score*), demonstrando o potencial de priorização operacional via dados administrativos. O artigo segue organizado em: Trabalhos Relacionados (Seção 2), Referencial Teórico (Seção 3), Metodologia (Seção 4), Resultados e Discussão (Seção 5) e Conclusão (Seção 6).

2. Trabalhos Relacionados

2.1. Predição de malignidade e diagnóstico citológico

Em bases do tipo Breast Cancer Wisconsin, estudos clássicos com aprendizado supervisionado reportam alto desempenho na classificação benigno/maligno. Em [Hugar and V 2024], a Regressão Logística atingiu 97,3% de acurácia, superando Random Forest e Linear SVC. De forma similar, [Maheshwar and Kumar 2019] comparou Árvore de Decisão, Naïve Bayes, KNN e SVM, com melhor resultado para SVM (97,7%). Em nossa experimentação na WDBC, a SVM manteve desempenho competitivo e, após otimização, alcançou 98,2% (Tabela 1), reforçando a perspectiva “micro” baseada em atributos morfométricos.

2.2. Monitoramento assistencial e a lei dos 60 dias no SUS

No eixo assistencial, [Gomes 2013] discute a Lei nº 12.732/2012 (prazo de 60 dias) e destaca que sua efetividade depende da capacidade estrutural da rede, do diagnóstico precoce e do financiamento, reforçando a necessidade de monitoramento contínuo do cumprimento do marco legal.

2.3. Proposta de integração: framework de dupla predição

Trabalhos em oncologia reforçam que estruturas de ensemble e fusão podem melhorar desempenho em tarefas complexas e que a operacionalização em fluxos reais ainda enfrenta desafios de interpretabilidade e escala [Kumar 2025]. Abordagens híbridas (extração manual + profundo) também apontam ganhos de confiabilidade

Tabela 1. Síntese comparativa: abordagem proposta vs. artigos relacionados

Autores	Base de dados	Técnicas principais	Desempenho
Hugar & V (2024)	Wisconsin BreastCancer Dataset (400X)	Logistic Regression, Linear SVC, SGD Classifier, Random Forest	Logistic Regression: 97,3% Linear SVC: 80,4% SGD Classifier: 94,7% Random Forest: 94,7%
Maheshwar & Kumar (2019)	BreastCancer Wisconsin Dataset	Decision Tree, Naïve Bayes, KNN, SVM	SVM: 97,7% KNN: 96,5% Naive Bayes: 95,4% Decision Tree: 93,3%
Abordagem proposta (2025)	BreastCancer Wisconsin Dataset (Diagnostic)	Logistic Regression, KNN, SVM	SVM Otimizado: 98,2% SVM: 97,4% Logistic Regression: 96,5% KNN: 95,6%

[Thepade et al. 2025], e a literatura reconhece que uma única fonte de dados pode ser insuficiente para decisões clínicas robustas [Kalamkar and Mary A. 2020]. Neste contexto, nossa proposta utiliza integração em nível de decisão para conectar risco biológico estimado (módulo clínico) a risco operacional/assistencial, visando apoiar priorização em cenários futuros com dados vinculáveis.

3. Referencial Teórico

3.1. O problema da classificação de células mamárias

A classificação histopatológica de tumores mamários é crucial para decisões terapêuticas e prognóstico [Cserni 2020]. Parâmetros morfométricos nucleares, como área e perímetro, possuem alto poder discriminativo entre lesões benignas e malignas [Cui et al. 2006]. Métodos de IA, incluindo aprendizado de máquina e profundo, identificam padrões morfológicos sutis, atingindo alto desempenho na classificação, embora enfrentem desafios na captura de nuances finas e necessidade de grandes volumes de dados anotados.

3.2. Modelagem do atraso assistencial e uso de dados administrativos

O atraso no tratamento oncológico agrava desfechos clínicos e a complexidade terapêutica [Ang et al. 2025]. Em câncer de mama, o intervalo diagnóstico-tratamento é um indicador crítico de qualidade [Silva et al. 2024]. Este atraso é predominantemente sistêmico, vinculado a fatores organizacionais, territoriais e à capacidade institucional em sistemas públicos [Ferreira et al. 2017]. Dados administrativos processados por aprendizado de máquina permitem capturar interações não lineares entre variáveis clínicas e demográficas para monitorar o desempenho populacional [Frosch et al. 2023].

3.3. Aprendizado supervisionado para classificação

No aprendizado supervisionado, a Regressão Logística favorece a interpretabilidade em análises de risco, enquanto as SVM são eficazes na separação não linear em espaços de alta dimensão [Géron 2019]. O KNN classifica por proximidade, sendo sensível à dimensionalidade. Métodos de *ensemble* melhoram o desempenho combinando preditores: o *Voting Classifier* agrega decisões, enquanto o *Gradient Boosting* (e variantes como XGBoost, CatBoost e Hist-GB) constrói modelos sequenciais para otimizar a predição em dados tabulares.

4. Metodologia

O framework propõe uma abordagem de dupla predição com dois módulos independentes: (i) risco biológico de malignidade tumoral e (ii) risco assistencial de atraso no início do tratamento. Nesta prova de conceito, cada módulo é treinado e avaliado em uma base pública distinta. As saídas probabilísticas P_{bio} e P_{sus} são posteriormente combinadas em um escore de priorização em nível de decisão (*Priority Score*), destinado a cenários futuros em que dados clínicos e administrativos possam ser integrados no fluxo assistencial.

4.1. Predição de risco clínico

No eixo clínico, utilizou-se o *Breast Cancer Wisconsin Diagnostic Dataset* (569 amostras), derivado de biópsias por punção aspirativa por agulha fina (PAAF), disponibilizado pela UCI Machine Learning Repository. Conforme [Street et al. 1993], as variáveis são obtidas a partir da segmentação de núcleos celulares via contornos ativos (*Snakes*), resultando em atributos morfológicos sumarizados por estatísticas (média, erro padrão e valor extremo), totalizando 30 preditores; o alvo é binário (maligno/benigno). Os atributos foram padronizados (*StandardScaler*) e os dados divididos em treino e teste (80/20).

4.2. Predição de risco assistencial

No eixo assistencial, utilizou-se uma amostra de 20% do Painel Oncologia (DATASUS), com 995 mil registros (2013–2024), integrando SISCAN, SIA-SUS e SIH-SUS, obtida via Kaggle. Após remoção de duplicidades e correção de inconsistências temporais, definiu-se a variável alvo binária *atraso_60_dias*, positiva quando *TEMPO_TRAT* excede 60 dias. Atributos foram organizados em dimensões (i) territorial/institucional (UF e CNES), (ii) temporal (mês do diagnóstico) e (iii) demográfica (sexo e idade), conforme sumarizado na Tabela 2.

Tabela 2. Descrição das variáveis do dataset Painel Oncologia (DATASUS)

Nome da Coluna	Descrição
year_diag	Ano em que foi realizado o diagnóstico (AAAA).
month_diag	Mês em que foi realizado o diagnóstico (extraído de AAAAMM).
IDADE	Idade do paciente na data do diagnóstico.
SEXO	Sexo biológico do paciente.
UF_RESID	Estado (Unidade Federativa) de residência do paciente (Código IBGE).
MUN_RESID	Município de residência do paciente (Código IBGE).
UF_DIAGN	Estado onde o diagnóstico foi realizado.
MUN_DIAG	Município onde o diagnóstico foi realizado.
CNES_DIAG	Código do estabelecimento de saúde (CNES) onde ocorreu o diagnóstico.
atraso_60_dias	Indicador de início de tratamento após 60 dias do diagnóstico

4.3. Validação e modelagem

Para refletir o uso prospectivo e evitar vazamento temporal, adotou-se validação por divisão cronológica, com exclusão de janelas incompletas (viés de maturação), garantindo que apenas ciclos assistenciais finalizados compusessem treino e avaliação. Foram comparados modelos de Regressão Logística (baseline, com ponderação de classes e ajuste por busca em grade com validação temporal) e métodos baseados em árvores impulsionadas por gradiente (HistGradientBoosting, XGBoost e CatBoost), escolhidos por capturarem não linearidades e interações entre variáveis administrativas. Para maior estabilidade, as melhores abordagens foram combinadas em um ensemble do tipo *Voting Classifier* com votação suave, avaliado em conjunto de teste independente.

4.4. Índice de Priorização

Propõe-se um mecanismo de priorização em nível de decisão que combina as probabilidades dos dois módulos (P_{bio} e P_{sus}). Nesta prova de conceito, o escore é discutido como componente para cenários futuros em que variáveis clínicas e administrativas possam ser vinculadas; para fins ilustrativos, a demonstração numérica utiliza $P_{bio} = 0,95$ como valor fixo (simulação), sem vinculação indivíduo-a-indivíduo entre bases:

$$\text{PriorityScore} = \sqrt{P(\text{maligno} \mid \text{células}) \times P(\text{atraso} > 60 \mid \text{contexto SUS})}. \quad (1)$$

A efetividade empírica reportada neste trabalho concentra-se no módulo assistencial, avaliado por análise de *lift* ao comparar a taxa média de atraso na população com a taxa no subgrupo de maior risco.

5. Resultados e Discussão

5.1. Resultados no dataset Wisconsin

No conjunto Breast Cancer Wisconsin Diagnostic, os modelos supervisionados apresentaram elevado desempenho na classificação de tumores mamários. O SVM otimizado destacou-se, alcançando acurácia de 98,2% no conjunto de teste, com eliminação completa de falsos positivos e apenas dois falsos negativos (Figura 1). Os resultados obtidos superam os achados de [Maheshwar and Kumar 2019], que relataram 97,7% de acurácia para SVM no mesmo conjunto de dados, e [Hugar and V 2024], que obtiveram 97,3% com Regressão Logística. Essa diferença, embora percentualmente pequena, pode ser clinicamente relevante, sobretudo por estar associada à redução de falsos negativos, sugerindo que a etapa de otimização de hiperparâmetros via *GridSearchCV* foi determinante para refinar a decisão do modelo. A Tabela 3 resume o comparativo das métricas globais entre os modelos.

Tabela 3. Desempenho dos modelos — Métricas de Classificação

Modelo	Acurácia	Precisão	Recall	F1 Score	FP	FN
SVM Otimizado	98.2%	98.3%	98.2%	98.2%	0	2
SVM	97.4%	97.5%	97.4%	97.3%	0	3
Regressão Logística	96.5%	96.5%	96.5%	96.5%	1	3
KNN	95.6%	95.7%	95.6%	95.6%	1	4

Nota: F1-Score calculado a partir de limiar de decisão ajustado para maximizar o equilíbrio entre precisão e sensibilidade, conforme procedimento descrito na metodologia.

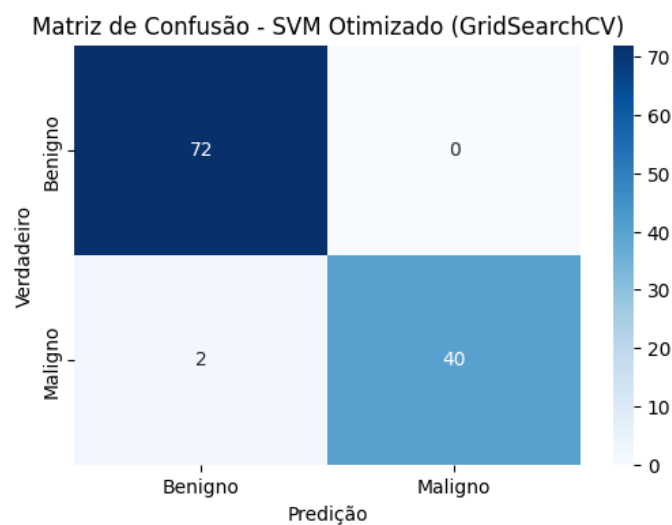


Figura 1. Matriz de confusão do modelo SVM Otimizado no conjunto de teste. Destaca-se a ausência de Falsos Positivos e a ocorrência mínima de Falsos Negativos.

Fonte: Autoria própria.

A análise de importância das características indicou predominância de atributos relacionados a valores extremos de textura, simetria e área nuclear (Figura 2), combinados com critérios histopatológicos clássicos, como o sistema de Nottingham. Esses achados reforçam a aceitabilidade biológica do modelo, sugerindo que ele captura padrões morfológicos reais associados à malignidade.

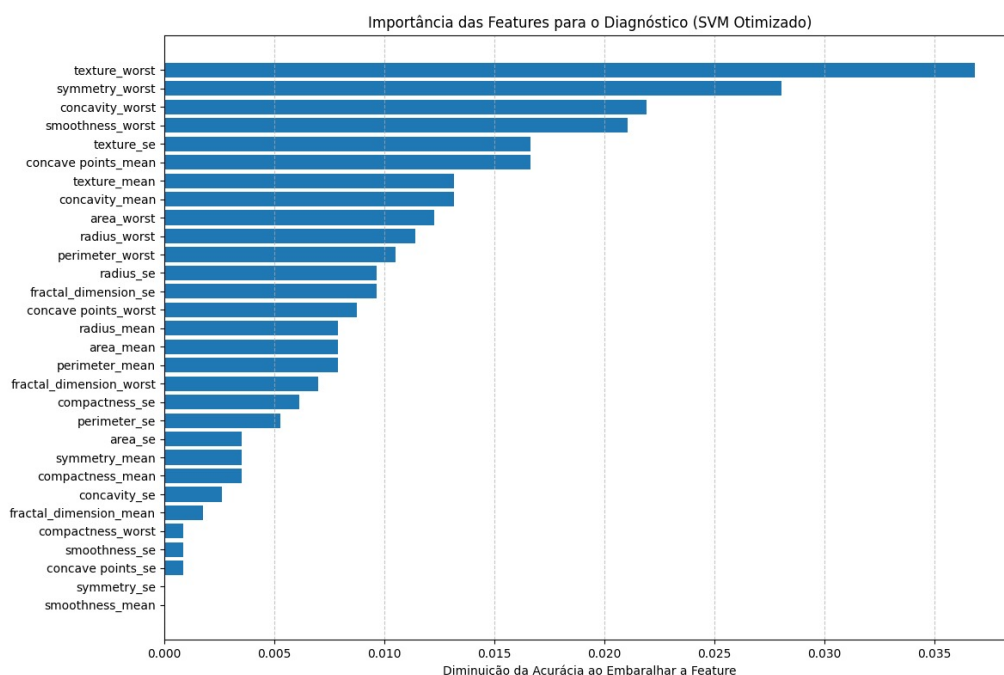


Figura 2. Importância das características para o diagnóstico no modelo SVM Otimizado. As variáveis relacionadas à textura, simetria e suavidade (especialmente os valores extremos/"piores") demonstraram maior poder preditivo.

Fonte: Autoria própria.

5.2. Caracterização do atraso assistencial no SUS

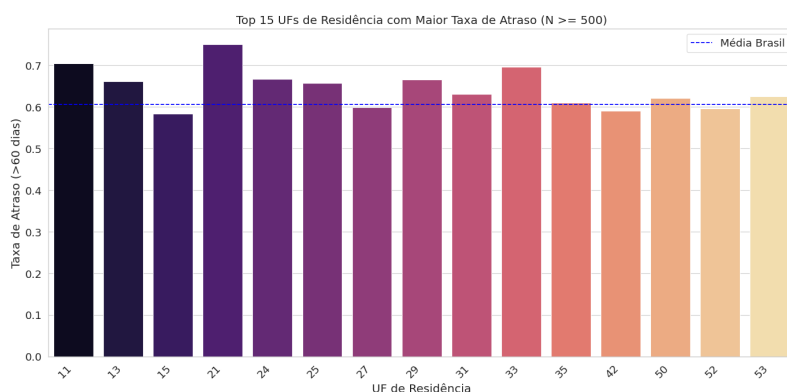
A análise do Painel Oncologia confirma o atraso no tratamento do câncer de mama como um problema estrutural no SUS: cerca de 60,2% das pacientes excedem o prazo legal de 60 dias, corroborando achados de [Instituto Nacional de Câncer 2020] e [de Medeiros et al. 2020].

Os dados revelam heterogeneidade territorial, com estados como Maranhão, Rondônia e Rio de Janeiro superando 70% de atraso (**Figura 3a**). Embora a literatura associe o atraso ao nível socioeconômico [Ferreira et al. 2017] e à distância dos centros de saúde [Silva et al. 2024], esta análise não identificou significância estatística quanto à distância.

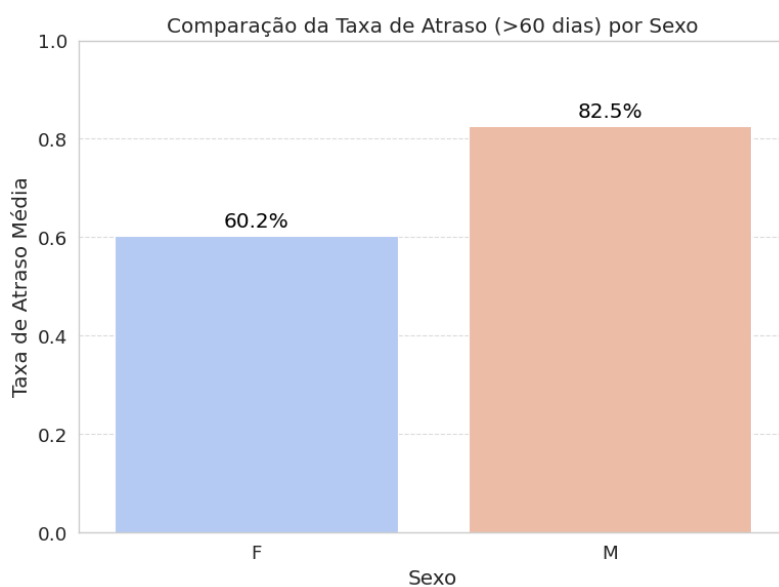
A análise por unidade de saúde (CNES) expõe gargalos severos, com instituições atingindo 100% de atraso, o que demanda investigação de fatores operacionais. Identificou-se ainda que homens possuem taxa de atraso superior (82,5%) à das mulheres (60,2%) (**Figura 3b**). Por fim, observou-se sazonalidade nos meses de novembro e dezembro e uma tendência de agravamento temporal do problema (**Figura 3c**).

5.3. Modelagem preditiva do atraso assistencial

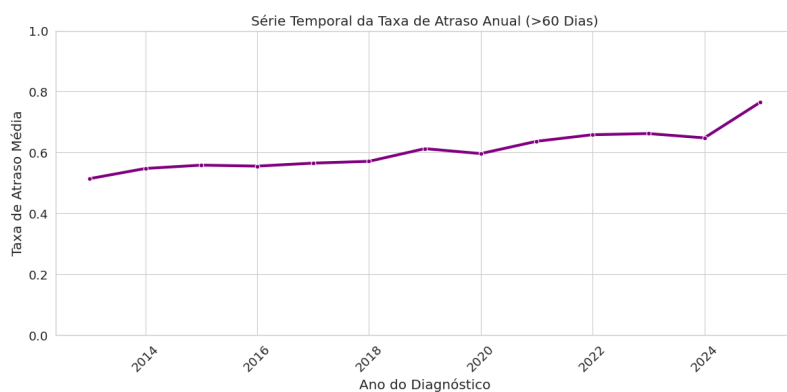
Modelou-se a probabilidade de atraso assistencial superior a 60 dias a partir de variáveis institucionais, geográficas e demográficas. Foram avaliados um baseline linear (Regressão Logística, com ponderação de classes e otimização por busca em grade) e métodos não lineares baseados em árvores impulsionadas por gradiente (HistGradientBoosting, XGBoost e CatBoost), por capturarem interações entre atributos administrativos.



(a) Taxa de atraso por Unidade Federativa (UF).



(b) Taxa de atraso por sexo (Masculino vs. Feminino).



(c) Evolução temporal do atraso assistencial.

Figura 3. Análise multidimensional do atraso assistencial no SUS: (a) geografia, (b) sexo e (c) evolução temporal.

Para aumentar robustez, combinou-se os melhores modelos em um ensemble do tipo *Voting Classifier* com votação suave, no qual as probabilidades são agregadas por média ponderada:

$$\hat{P}_{\text{ensemble}}(y = 1 | x) = \frac{\sum_{m=1}^M w_m P_m(y = 1 | x)}{\sum_{m=1}^M w_m}. \quad (2)$$

No projeto, $M = 3$ e foram utilizados pesos $[1, 1, 2]$ para HistGB, XGB e CatBoost, respectivamente.

Para integração em nível de decisão, define-se o *Priority Score* como média geométrica entre risco biológico (P_{bio}) e risco assistencial (P_{sus}), enfatizando casos em que ambos são elevados:

$$\text{PriorityScore}(x) = \sqrt{P_{bio} \cdot P_{sus}}. \quad (3)$$

Em termos empíricos, a Regressão Logística (baseline) obteve AUC-ROC de 0,6517 e F1-Score de 0,7317 na validação, enquanto os modelos GBDT apresentaram ganhos principalmente em F1-Score. O ensemble aumentou a estabilidade das previsões e alcançou AUC-ROC de 0,6439 no teste (Tabela 4). Apesar de modesto em termos absolutos, o desempenho é compatível com a natureza administrativa/ruidosa dos dados e é útil para estratificação operacional; em linha com [Frosch et al. 2023], variáveis geográficas e institucionais (como UF e CNES) são determinantes para capturar disparidades sistêmicas ignoradas por modelos puramente clínicos.

Tabela 4. Comparação de desempenho dos modelos (F1-Score e AUC-ROC).

Modelo	Conjunto	F1-Score	AUC-ROC
Logistic Regression (Baseline)	Validação	0.7317	0.6517
Logistic Regression (Otimizado)	Validação	0.7428	0.6544
HistGradientBoosting	Validação	0.7714	0.6359
XGBoost	Validação	0.7721	0.6604
CatBoost	Validação	0.7999	0.6401
Logistic Regression (Otimizado)	Teste	0.7454	0.6369
Ensemble	Teste	0.7921	0.6439

5.4. Prova de valor e priorização assistencial

A utilidade prática do modelo foi avaliada por meio de análise de *lift*. Enquanto a taxa média de atraso no conjunto de teste foi de 66,19%, o grupo correspondente aos 10% de maior risco identificado pelo modelo apresentou taxa de atraso de 85,63%, resultando em um *lift* de $1,29\times$ em relação a uma seleção aleatória (Figura 4).

A integração proposta entre risco biológico e risco assistencial define um *Priority Score* concebido para identificar, prospectivamente, pacientes em maior vulnerabilidade operacional. Nesta prova de conceito, a utilidade prática demonstrada empiricamente concentra-se no módulo assistencial (predição de atraso e análise de *lift*); já o *Priority Score* é apresentado como mecanismo de decisão para uso quando informações clínicas e administrativas puderem ser acopladas no nível do indivíduo. Scores elevados (por exemplo, 0,96, no exemplo ilustrativo com $P_{bio} = 0,95$) representariam casos de alta agressividade tumoral associados a contextos institucionais com elevada probabilidade de falha operacional, configurando um ponto cego do sistema assistencial.

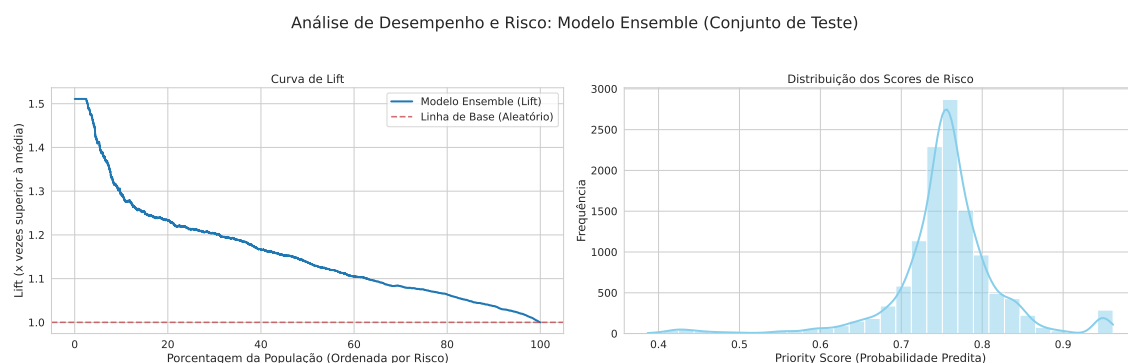


Figura 4. Análise de *lift* e distribuição de risco assistencial e biológico, destacando o subgrupo de maior prioridade segundo o Priority Score.

Fonte: Autoria própria.

6. Conclusão

Este estudo propôs um *framework* de dupla predição para apoiar a priorização oncológica no SUS, integrando o risco clínico de malignidade ao risco assistencial de atraso terapêutico. A abordagem supera a triagem tradicional ao articular diagnóstico e gestão de forma complementar.

No eixo clínico, modelos supervisionados baseados em características morfológicas ofereceram suporte confiável à classificação de tumores, priorizando a redução de falsos negativos. No eixo assistencial, o uso de aprendizado de máquina em dados do DATASUS revelou que o atraso no tratamento é fortemente condicionado por fatores territoriais e institucionais, refletindo desigualdades estruturais da rede oncológica.

A integração desses eixos via índice de priorização identificou grupos com elevado risco de atraso, apresentando ganho operacional superior a estratégias aleatórias. O *framework* demonstra potencial como ferramenta preventiva para auxiliar o cumprimento da Lei nº 12.732/2012.

Em suma, o trabalho evidencia o potencial do aprendizado de máquina para converter registros administrativos do SUS em inteligência preditiva. Como limitações, esta prova de conceito utilizou bases distintas para cada risco; trabalhos futuros preveem a integração de dados indivíduo-a-indivíduo, uso de modelos de *deep learning* e expansão para outros tipos de neoplasias.

7. Agradecimentos

Esta pesquisa foi realizada com o apoio da Fundação Amazônica de Amparo à Pesquisa e Desenvolvimento Tecnológico Desembargador Paulo dos Anjos Feitoza (FPF-tech) e Foxconn Brasil.

Referências

Ang, S. P., Lee, E., Chia, J. E., Iglesias, M., Di Vanna, M., Shambhavi, S., and Iglesias, J. (2025). Time-to-treatment initiation and its effect on all-cause mortality: Insights from the Surveillance, Epidemiology, and End Results database. *World Journal of Oncology*, 16(3):286–294. PMID: 40556964; PMCID: PMC12185122.

- Cserni, G. (2020). Histological type and typing of breast carcinomas and the who classification changes over time. *PMC*, pages 25–41.
- Cui, Y., Koop, E. A., Van Diest, P. J., Kandel, R. A., and Rohan, T. E. (2006). Nuclear morphometric features in benign breast tissue and risk of subsequent breast cancer. *Breast Cancer Research and Treatment*, 97(2):103–107.
- de Medeiros, G. C., Teodózio, C. G. C., Fabro, E. A. N., Aguiar, S. S., Lopes, A. H. M., Conte, B. C., Silva, E. V., Coelho, L. L. P., Muniz, N. F., Schuab, S. I. P. C., Bergmann, A., and Thuler, L. C. S. (2020). Fatores associados ao atraso entre o diagnóstico e o início do tratamento de câncer de mama: um estudo de coorte com 204.130 casos no brasil. *Revista Brasileira de Cancerologia*, 66(3):e–09979.
- Ferreira, N. A. S., Carvalho, S. M. F., Valenti, V. E., Bezerra, I. M. P., Batista, H. M. T., de Abreu, L. C., Matos, L. L., and Adami, F. (2017). Treatment delays among women with breast cancer in a low socio-economic status region in Brazil. *BMC Women’s Health*, 17(1):13. PMID: 28222726; PMCID: PMC5320774.
- Freihat, O., Sipos, D., and Kovacs, A. (2025). Global burden and projections of breast cancer incidence and mortality to 2050: a comprehensive analysis of GLOBOCAN data. *Frontiers in Public Health*, 13:1622954. PMID: 41246094; PMCID: PMC12611818.
- Frosch, Z. A. K., Hasler, J., Handorf, E., et al. (2023). Development of a multilevel model to identify patients at risk for delays in cancer treatment initiation. *JAMA Network Open*, 6(8):e2328712. Título traduzido: Desenvolvimento de um modelo multinível para identificar pacientes com risco de atraso no início do tratamento do câncer.
- Ghantasala, G. S. P., Kunchala, A., R, S., B, V. N., Raparathi, Y., and Vidyullatha, P. (2023). Machine learning based ensemble classifier using Wisconsin dataset for breast cancer prediction. In *2023 International Conference on Integrated Intelligence and Communication Systems (ICIICS)*, pages 1–4, Kalaburagi, India.
- Gomes, F. d. B. C. (2013). Lei sobre tratamento de câncer no SUS: análise dos dispositivos e perspectivas. *Cadernos Ibero-Americanos de Direito Sanitário*, 2(1):81–87.
- Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, Sebastopol, CA, 2 edition.
- Hugar, S. and V, S. (2024). Breast cancer classification using logistic regression: An efficient and scalable approach. In *IEEE*.
- Instituto Nacional de Câncer (2020). *Relatório do intervalo entre diagnóstico e início do tratamento do câncer no SUS: dados do PAINEL-Oncologia*. Rio de Janeiro. Acesso em: 2 jan. 2026.
- Instituto Nacional de Câncer (2022). *Estimativa 2023: Incidência de Câncer no Brasil*. INCA, Rio de Janeiro. Acesso em: 2 jan. 2026.
- Kalamkar, S. and Mary A., G. (2020). Fusão de dados clínicos e técnicas de aprendizado de máquina para saúde inteligente. In *Conferência Internacional de Tecnologia da Indústria 4.0 (I4Tech) de 2020*, pages 211–216, Pune, India.

- Kumar, V. (2025). Avanços em inteligência artificial para oncologia: Uma revisão abrangente de técnicas, desafios e aplicações. In *Sétima Conferência Internacional sobre Inteligência Computacional e Tecnologias de Comunicação (CCICT)*, pages 40–47, Sonapat, India.
- Maheshwar and Kumar, G. (2019). Breast cancer detection using decision tree, naïve bayes, knn and svm classifiers: A comparative study pdf. In *IEEE*.
- S, N., J, S., M, V., S, S., and G, S. (2022). SVM-ANN optimized algorithm for the classification of breast cancer data as benign and malignant. In *2022 Smart Technologies, Communication and Robotics (STCR)*, pages 1–7. Acesso em: 2 out. 2025.
- Silva, F. d. P. C. d., Souza, M. C., and Bertoni, N. (2024). Factors associated with delayed initiation of breast cancer treatment at an oncology referral center in Juiz de Fora, Minas Gerais state, from 2010 to 2019: a cohort study. *Epidemiologia e Serviços de Saúde*, 33:e20231177.
- Soliman, A., Li, Z., and Parwani, A. V. (2024). Impact of artificial intelligence in breast cancer pathology: a review of the literature. *Diagnostic Pathology*, 19(38).
- Street, W. N., Wolberg, W. H., and Mangasarian, O. L. (1993). Nuclear feature extraction for breast tumor diagnosis. In Acharya, R. S. and Goldgof, D. B., editors, *Biomedical Image Processing and Biomedical Visualization*, volume 1905, pages 861 – 870. International Society for Optics and Photonics, SPIE.
- Thepade, S. D., Patil, R. Y., Fasate, A., Patil, A., Padhen, C., and Kadam, P. (2025). Aprimorando a detecção de câncer de mama usando aprendizado profundo híbrido e recursos artesanais. In *9ª Conferência Internacional sobre Computação, Comunicação, Controle e Automação (ICCUBEA)*, pages 01–07, Pune, India.
- Uchikov, P., Khalid, U., Dedaj-Salad, G. H., Ghale, D., Rajadurai, H., Kraeva, M., Kraev, K., Hristov, B., Doykov, M., Mitova, V., Bozhkova, M., Markov, S., and Stanchev, P. (2024). Artificial intelligence in breast cancer diagnosis and treatment: Advances in imaging, pathology, and personalized care. *Life*, 14(11):1451. PMID: 39598249; PMCID: PMC11595975.
- Verras, G.-I., Tchabashvili, L., Chlorogiannis, D.-D., Mulita, F., and Argentou, M.-I. (2023). Updated clinical evidence on the role of adipokines and breast cancer: A review. *Cancers*, 15(6):1572.