

Arquitetura Multiagente Explicável para Predição de Desfechos Clínicos: Um Framework de IA Confiável Aplicado ao Monitoramento da Tuberculose

Ronilson W. S. Pereira¹, Igor Falcão², Saul Carneiro³,
Marcos Seruffo², Karla Figueiredo¹

¹Instituto de Matemática e Estatística – Universidade do Estado do Rio de Janeiro
Rio de Janeiro – RJ – Brasil

²Instituto de Tecnologia – Universidade Federal do Pará
Belém – PA – Brasil

³Hospital Universitário João de Barros Barreto – Universidade Federal do Pará
Belém – PA – Brasil

ronilsonengenharia@gmail.com, igorufpa2013.4@gmail.com

saul@ufpa.br, seruffo@ufpa.br, karlafigueiredo@ime.uerj.br

Abstract. *This work proposes an Explainable Multi-Agent Framework for decision support in public health, specifically focused on tuberculosis monitoring. The architecture employs specialized agents to coordinate data processing, risk prediction, and explainability. By integrating the XGBoost classifier with the SHAP method and the SMOTENC balancing technique, the framework identifies high-risk patients with 92.56% estimated probability in critical cases, doubling the recall for unfavorable outcomes compared to traditional models. The results demonstrate that the proposed modularity ensures predictive performance combined with high transparency, highlighting the framework's portability for monitoring other chronic conditions and providing qualified support for health managers.*

Resumo. *Este trabalho propõe um framework multiagente de IA Explicável para apoio à decisão em saúde pública, com foco no monitoramento da tuberculose. A arquitetura utiliza agentes especializados para coordenar o processamento de dados, a predição de risco e a explicabilidade. Ao integrar o classificador XGBoost ao método SHAP e à técnica de balanceamento SMOTENC, o framework identifica pacientes de alto risco com 92,56% de probabilidade estimada em casos críticos, dobrando o recall para desfechos desfavoráveis em relação a modelos tradicionais. Os resultados demonstram que a modularidade proposta garante desempenho preditivo aliado a elevados níveis de transparência, evidenciando a portabilidade da arquitetura para o monitoramento de outras condições crônicas e o suporte qualificado a gestores de saúde.*

1. Introdução

O monitoramento inteligente de desfechos clínicos em cenários de saúde pública exige arquiteturas que não apenas processem grandes volumes de dados, mas que operem

de forma transparente e adaptativa. No ecossistema de *e-Health*, a integração de sistemas de apoio à decisão requer que as previsões envolvam elevados níveis de confiabilidade e interpretabilidade [Azevedo and Dusek 2025]. Entretanto, o avanço recente da Inteligência Artificial (IA) tem sido marcado pelo uso de modelos de Aprendizado de Máquina (*Machine Learning* - ML), que frequentemente operam como "caixas-pretas", dificultando a auditoria dos resultados por gestores e profissionais de saúde [Hameed et al. 2025, Ribeiro et al. 2025].

Nesse cenário, abordagens baseadas em sistemas multiagentes (SMA) emergem como uma alternativa robusta para a construção de entidades autônomas e proativas, alinhadas à tendência de *AI Agents* que integram percepção e planejamento [Hughes et al. 2025]. Essa visão arquitetural favorece a construção de sistemas modulares onde a tecnologia e os atores humanos interagem de forma integrada [Born et al. 2025]. A utilização de agentes especializados permite que a lógica de aquisição de dados, a previsão de risco e a explicabilidade (*eXplainable AI* - XAI) operem de maneira coordenada mas independente, permitindo que o sistema tome decisões granulares sobre a necessidade de re-calibração ou geração de alertas específicos, facilitando a manutenção e a escalabilidade para diferentes domínios clínicos.

Para validar a eficácia desta arquitetura, este trabalho utiliza como estudo de caso o monitoramento da Tuberculose (TB). A TB permanece como uma emergência global cuja complexidade exige ferramentas capazes de oferecer inteligência contextualizada para a decisão [Batista et al. 2025, World Health Organization 2025]. Surge, então, uma necessidade crítica: não basta que um sistema aponte o risco de um desfecho clínico desfavorável (como o abandono do tratamento); é fundamental que a arquitetura forneça as justificativas e as incertezas por trás dessa previsão, transformando a saída algorítmica em um suporte qualificado para a intervenção humana.

As principais contribuições deste trabalho residem na proposição de uma arquitetura multiagente explicável que organiza o fluxo de previsão de desfechos clínicos em entidades especializadas e com autonomia de orquestração. O *framework* integra o algoritmo *XGBoost* [Chen et al. 2015] a explicações SHAP locais e globais, mitigando a opacidade em processos decisórios críticos e garantindo a transparência necessária para sistemas de saúde. Além disso, a pesquisa apresenta uma avaliação experimental de técnicas de balanceamento de dados e otimização de limiares, demonstrando como a arquitetura proposta consegue equilibrar a *performance* técnica com a sensibilidade clínica exigida no monitoramento de agravos de saúde.

Além da introdução, este artigo está estruturado da seguinte maneira: na Seção 2 são apresentados os trabalhos correlatos à pesquisa; na Seção 3 são descritos os materiais e métodos empregados; a Seção 4 expõe os resultados alcançados; e, por fim, a Seção 5 aborda as conclusões do estudo, juntamente com sugestões para investigações futuras.

2. Trabalhos Correlatos

Esta seção apresenta os trabalhos que fundamentam o desenvolvimento desta arquitetura, analisando abordagens de previsão em saúde pública, sistemas baseados em agentes e o uso de IA Explicável.

Em relação à previsão de desfechos em tuberculose no cenário brasileiro, o trabalho de [Rodrigues et al. 2024] utilizou dados do SINAN para estimar o risco de abandono

utilizando algoritmos como *Random Forest* e *LightGBM*, alcançando AUC entre 0,71 e 0,72. Por sua vez, [Yasin et al. 2024] propuseram um modelo para prever o tempo de internamento prolongado, integrando técnicas de XAI para identificar fatores de risco em uma aplicação web. Observou-se, de maneira geral, que as soluções existentes focam na performance algorítmica isolada e na identificação de variáveis, mas frequentemente negligenciam o tratamento de desbalanceamento severo em bases de dados nacionais e a integração das explicações em um fluxo de trabalho coordenado.

No contexto de arquiteturas para sistemas de saúde inteligente, [Liu et al. 2025] apresentaram uma estrutura de agentes autônomos que integram planejamento e percepção para interação adaptativa em ambientes médicos. Complementarmente, [Tanasã et al. 2026] descreveram o *design* de assistentes baseados em orquestração de agentes, enfatizando a importância da rastreabilidade do raciocínio em assistentes virtuais. Entretanto, estas soluções são predominantemente voltadas para dados não estruturados ou modelos de linguagem, deixando uma lacuna na aplicação de sistemas multiagentes para o processamento de dados tabulares epidemiológicos que exigem alta interpretabilidade e suporte à decisão clínica baseada em evidências estruturadas.

Embora o suporte à decisão em saúde pública seja um tópico de alta relevância, as contribuições que consideram a explicabilidade como requisito arquitetural em sistemas pervasivos ainda são escassas. A carência de *frameworks* que unifiquem predição de alta performance (como o XGBoost) com justificativas locais e globais (SHAP) em uma estrutura de agentes coordenados é uma das principais justificativas para esta pesquisa. As lacunas observadas, como a opacidade dos modelos de "caixa-preta" e a falta de modularidade para lidar com bases de dados reais e desbalanceadas, motivaram a realização do presente trabalho.

3. Materiais e Métodos

Esta seção detalha a arquitetura do *framework* proposto, projetado para a predição de desfechos clínicos e suporte à decisão através de agentes inteligentes. A metodologia prioriza uma organização baseada em entidades autônomas, garantindo que o processo seja auditável e escalável. A Figura 1 ilustra a arquitetura geral do sistema, destacando o fluxo de dados e a orquestração proativa entre os agentes.

3.1. Estudo de Caso e Base de Dados

Para validar a arquitetura proposta, utilizou-se um estudo de caso focado no monitoramento da Tuberculose (TB). Os dados são provenientes de um repositório público¹, totalizando 103.846 prontuários coletados via sistema TBWEB entre 2006 e 2016 no estado de São Paulo. O conjunto de dados abrange variáveis demográficas e clínicas, onde a classe alvo define o desfecho do tratamento (favorável vs. desfavorável).

3.1.1. Aspectos Éticos

Em conformidade com a Resolução n.º 466/12, a pesquisa utilizou exclusivamente dados anonimizados de domínio público, dispensando a necessidade de revisão por Comitê

¹https://figshare.com/articles/dataset/tuberculosis-data-06-16_csv/8066663?file=15032345

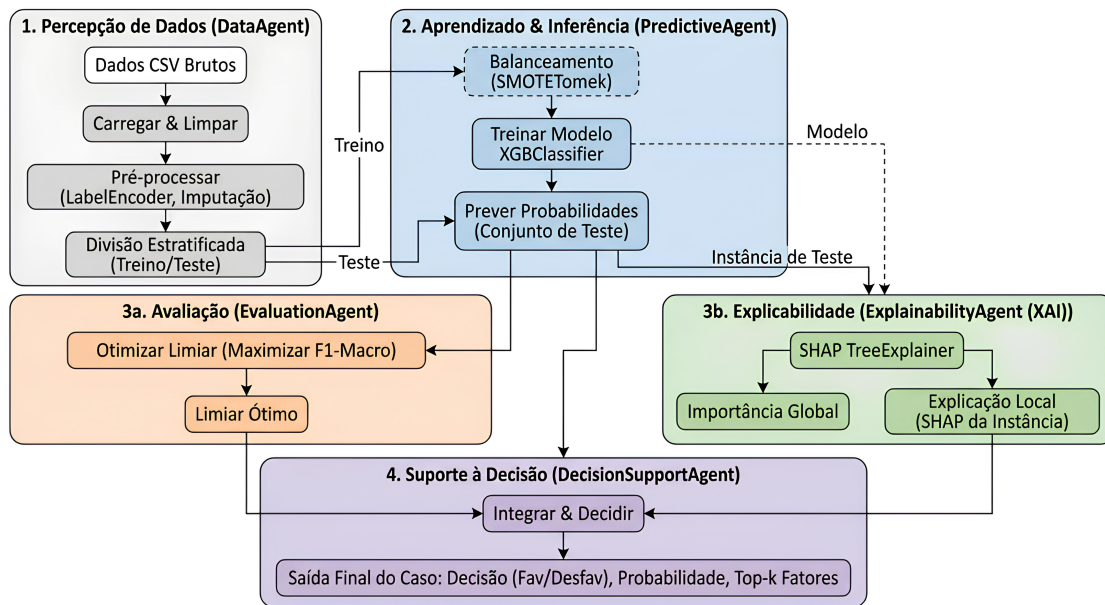


Figura 1. Arquitetura Multiagente do Framework de Apoio à Decisão Explicável.

de Ética em Pesquisa com Seres Humanos, conforme as normas vigentes para dados secundários não identificáveis.

3.2. Arquitetura Multiagente do Framework

O sistema adota o paradigma de sistemas multiagentes (SMA), decompondo funcionalidades complexas em entidades autônomas com responsabilidades específicas. A modularidade permite que cada componente seja aprimorado ou substituído sem comprometer a integridade do fluxo, que é estruturado em quatro estágios principais conforme a Figura 1:

- 1. Percepção de Dados (DataAgent):** Responsável pela ingestão, limpeza e pré-processamento (Label Encoding e imputação). Este agente realiza a divisão estratificada (treino/teste) e executa a seleção de atributos (*Feature Selection*) do tipo *Embedded* para priorizar variáveis de maior impacto clínico [Nassreddine et al. 2025].
- 2. Aprendizado e Inferência (PredictiveAgent):** Encapsula o motor de aprendizado de máquina. Utiliza a técnica *SMOTENC* [Dewi et al. 2025] para balanceamento no conjunto de treino e o algoritmo *XGBoost* [Chen et al. 2015] para gerar as probabilidades de desfecho no conjunto de teste.
- 3. Avaliação e Explicabilidade:** Estágio composto por dois agentes que operam de forma coordenada para garantir a confiabilidade do suporte à decisão:
 - **EvaluationAgent (3a):** Realiza a calibração clínica buscando o limiar (*threshold*) ótimo que maximize o $F1$ -score macro, garantindo equidade entre as classes.
 - **ExplainabilityAgent (3b):** Provê transparência via método SHAP [Ma et al. 2023], gerando a importância global das variáveis e a justificativa local para cada instância.

4. **Suporte à Decisão (DecisionSupportAgent):** Atua como a interface final de integração. Este agente consolida a predição, a probabilidade e os fatores de influência em um relatório interpretável, transformando a saída técnica em inteligência acionável.

3.3. Interpretabilidade e Fundamentação do SHAP

Para interpretar as decisões, o **ExplainabilityAgent** utiliza o método SHAP (*SHapley Additive exPlanations*), um arcabouço unificado que calcula a contribuição de cada recurso na predição final. Baseado na teoria dos jogos cooperativos, o valor de *Shapley* quantifica a contribuição de cada característica para diferentes subconjuntos de recursos.

O valor de *Shapley* do recurso i é obtido por uma soma ponderada sobre todas as possíveis combinações de inclusão, conforme expresso na Equação (1) [Paul et al. 2024]:

$$\Phi_i(x_i) = \sum_{S \subseteq M \setminus \{i\}} \frac{|S|!(|M| - |S| - 1)!}{|M|!} [f(S \cup \{i\}) - f(S)], \quad (1)$$

onde: $\Phi_i(x_i)$ representa a contribuição média do recurso i para a previsão do modelo; M é o conjunto total de recursos de entrada; S representa um subconjunto de recursos que não contém o recurso i ; $|S|$ é a cardinalidade (número de recursos) do subconjunto S ; e o símbolo ! denota a operação de fatorial. Esta fundamentação matemática permite que o agente decomponha a probabilidade final em fatores clínicos auditáveis, garantindo que a importância atribuída a variáveis como `nDosesSeg` seja matematicamente consistente e livre de ambiguidades interpretativas.

3.4. Seleção de Atributos e Variáveis de Estudo

A partir de um experimento de sensibilidade conduzido pelo **DataAgent** (variando K entre 5 e 30), identificou-se que a utilização das 10 variáveis com maior ganho de informação otimiza o desempenho do sistema ($F1$ -score macro $\approx 0,70$). A Tabela 1 detalha os preditores selecionados para compor o modelo final.

Tabela 1. Variáveis Selecionadas pelo DataAgent (Top 10).

Ranking	Variável	Descrição Técnica/Clínica	Tipo
1	DROGADICAO	Histórico de dependência de drogas.	String
2	nDosesSeg	Doses do 3° ao 6° mês de tratamento.	Numérico
3	faixaEtaria	Grupo etário do paciente.	String
4	hiv	Status de infecção por HIV.	String
5	tipoCaso	Tipo de entrada (ex: novo, recidiva).	String
6	instTrat	Tipo de instituição (ex: prisão, hospital).	String
7	COMUNICEXA	Contatos examinados no entorno do paciente.	Numérico
8	ALCOOLISMO	Histórico de consumo de álcool.	String
9	classif	Classificação clínica (ex: pulmonar).	String
10	TIPOCUP	Tipo de ocupação principal.	String

4. Resultados e Discussão

A avaliação do *framework* multiagente considerou a eficácia da seleção de atributos e o desempenho preditivo sob diferentes configurações de balanceamento e calibração. O rigor aplicado pelo **DataAgent** permitiu reduzir o espaço de busca para as 10 variáveis de maior ganho de informação, resultando em um modelo robusto. A Tabela 2 sintetiza a comparação entre a configuração com dados originais (Sem Balanceamento) e a abordagem proposta, evidenciando a mais elevada eficiência anotada nos resultados da arquitetura coordenada.

Tabela 2. Comparação de Desempenho: Dados Originais vs. Framework Proposto (SMOTENC Otimizado).

Configuração	Acurácia	Recall (Classe 0)	F1-Score Macro	Threshold
Sem Balanceamento	0,90	0,25	0,65	0,50
Proposto (Otimizado)	0,85	0,50	0,68	0,47

A Figura 2 apresenta as curvas ROC e *Precision-Recall*. O modelo alcançou uma AUC-ROC de 0,875 (sem balanceamento) e 0,843 (com SMOTENC). Contudo, a análise detalhada revela que o modelo tradicional, apesar da acurácia elevada, negligencia a classe de risco com um *recall* de apenas 0,25. Esse fenômeno ocorre porque o classificador convencional prioriza a classe majoritária, o que compromete severamente a detecção clínica do abandono de tratamento.

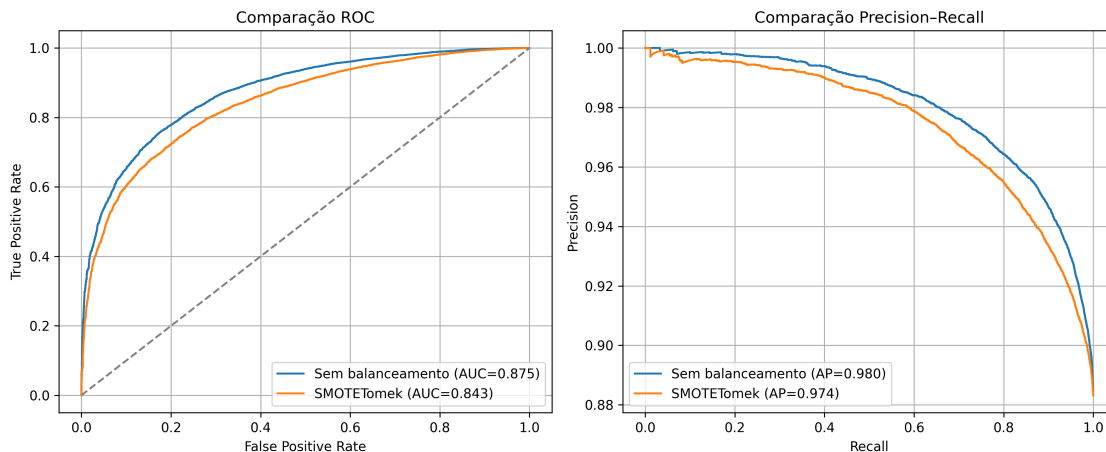


Figura 2. Comparação das métricas de avaliação entre os modelos Sem Balanceamento e SMOTENC.

Em contraste, a configuração proposta, orquestrada pelo **PredictiveAgent** e refinada pelo **EvaluationAgent**, elevou o *recall* da classe minoritária para 0,50 e o *F1-score macro* para 0,68. Embora um *recall* de 0,50 indique que o sistema ainda não detecta todos os casos, ele representa um **aumento de 100% na sensibilidade clínica** em relação à abordagem tradicional. No contexto da tuberculose, onde o custo epidemiológico de um falso negativo supera o custo operacional de investigar falsos positivos, dobrar a capacidade de detecção justifica plenamente o *trade-off* técnico e torna a ferramenta viável para o suporte à vigilância em saúde pública.

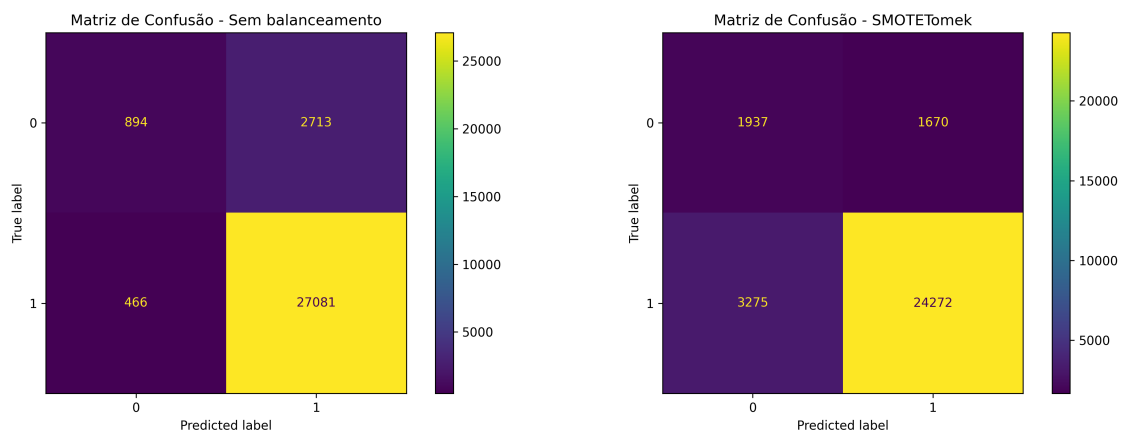


Figura 3. Matrizes de Confusão - Configuração Original vs. Framework Proposto.

Além do desempenho quantitativo, a integração da explicabilidade via **ExplainabilityAgent** é o diferencial central da arquitetura (Figura 4). A variável `nDosesSeg` apresenta o maior impacto médio na decisão, corroborando a literatura que aponta a continuidade do tratamento como o principal determinante do sucesso terapêutico. Adicionalmente, a relevância de `COMUNICEXA` (exame de contatos) e `hiv` demonstra que o modelo captou a complexidade da doença sob padrões epidemiológicos reais, priorizando marcadores de vulnerabilidade social e clínica.

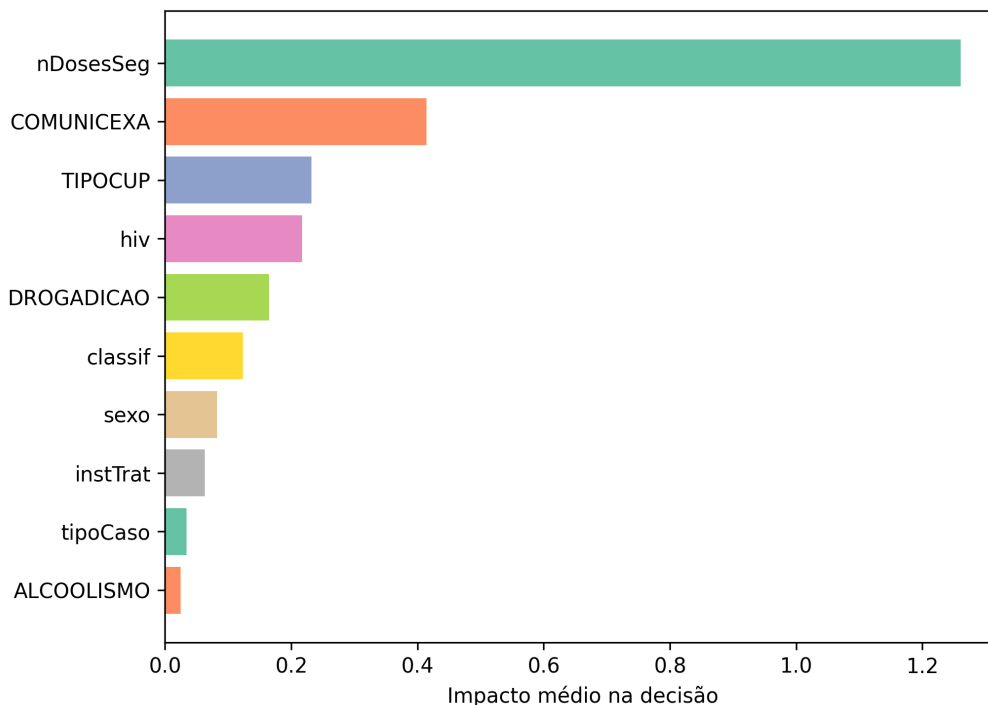


Figura 4. Principais Variáveis Explicativas (XAI Global) geradas pelo ExplainabilityAgent.

No nível individual, a Figura 5 demonstra como o agente identifica determinantes específicos para cada paciente, permitindo uma auditoria detalhada do risco. Observa-se

que, para o caso ilustrado, o baixo número de doses registradas (`nDosesSeg`) impulsionou o aumento do risco estimado. Em contrapartida, a realização de exames de contatos (`COMUNICEXA`) atuou como um fator de proteção, reduzindo a probabilidade de desfecho desfavorável.

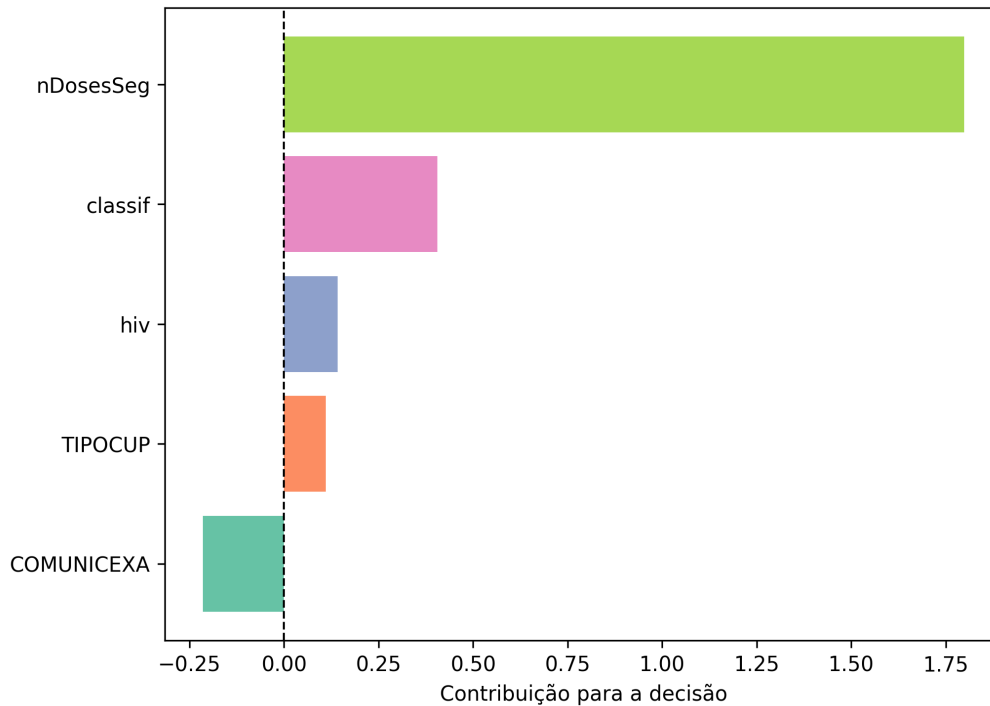


Figura 5. Explicação da Decisão Individual (XAI Local) para suporte clínico.

Epidemiologicamente, o registro do exame de contatos sinaliza uma vigilância em saúde ativa: o rastreo de pessoas próximas indica que o paciente está inserido em um fluxo de acompanhamento mais rigoroso, o que reduz as chances de falha terapêutica. Essa transparência transforma a predição matemática em um roteiro de intervenção: o profissional percebe que, embora a adesão medicamentosa esteja baixa, há um vínculo com a vigilância que pode ser explorado.

Finalmente, o **DecisionSupportAgent** atua na consolidação das informações, reforçando a natureza probabilística do suporte à decisão. Em cenários críticos, o agente é capaz de identificar predições com alto grau de confiança, como em casos de risco acen-tuado onde o sistema estima uma probabilidade de desfecho desfavorável de 92,56%. Ao apresentar esse valor de forma contínua e não apenas como um veredito binário, o *framework* convida o profissional de saúde ao exercício do julgamento crítico e à priorização imediata da intervenção. Esta abordagem consolida o sistema como uma ferramenta que prioriza a responsabilidade clínica e a transparência, transformando a saída algorítmica em um suporte qualificado e pervasivo para a gestão de agravos em saúde pública.

5. Considerações Finais

Este trabalho apresentou um *framework* multiagente explicável para a predição de desfechos clínicos, validado no monitoramento da tuberculose. A organização modular em

agentes especializados (**DataAgent**, **PredictiveAgent**, **EvaluationAgent**, **ExplainabilityAgent** e **DecisionSupportAgent**) permitiu converter dados brutos em suporte à decisão transparente e auditável, superando as limitações de opacidade comuns em modelos de aprendizado de máquina tradicionais.

Os resultados demonstraram que o uso da técnica *SMOTENC* e a otimização de limiares conduzida pelo **EvaluationAgent** foram cruciais para mitigar o viés da classe majoritária. Essa abordagem permitiu dobrar a capacidade de identificação de pacientes em risco (*Recall* de 0,50), garantindo uma sensibilidade clínica superior à abordagem *baseline*. A integração da explicabilidade via SHAP assegurou que predições críticas, como casos com probabilidade estimada de 92,56%, fossem acompanhadas por justificativas baseadas em fatores clínicos e sociais, promovendo a confiança necessária para a adoção do sistema em saúde pública.

Apesar dos avanços, o estudo apresenta limitações, como a dependência da completude dos registros administrativos e o foco inicial em dados do estado de São Paulo. Padrões epidemiológicos distintos podem exigir novas calibrações, o que reforça a necessidade de validações externas futuras. A principal contribuição deste trabalho reside, portanto, na proposição de uma estrutura modular e adaptável que facilita essa transição. Como investigações futuras, pretende-se explorar o aprendizado federado para preservar a privacidade dos dados entre diferentes unidades de saúde e a integração de modelos de linguagem (*LLMs*) para a geração de relatórios explicativos ainda mais acessíveis aos profissionais de saúde.

Referências

- Azevedo, R. B. A. and Dusek, P. M. (2025). Informação em saúde como recurso estratégico para a governança pública. *Revista Científica FESA*, 3(27):112–123.
- Batista, J. F. C., Almeida-Santos, M. A., and Lima, S. O. (2025). Tendência dos indicadores epidemiológicos e fatores associados ao abandono de tratamento e óbito pela tuberculose em pessoas em situação de rua no brasil: estudo ecológico e transversal, 2014-2022. *Epidemiologia e Serviços de Saúde*, 34:e20240273.
- Born, M. B., Alves, B. C., Goulart, F. M., Caldas, L. B., and de Aguiar, M. S. (2025). Revisão sistemática da literatura integrando sistemas multiagente e large language models. In *Workshop-Escola de Sistemas de Agentes, seus Ambientes e Aplicações (WESAAC)*, pages 230–239. SBC.
- Chen, T., He, T., Benesty, M., Khotilovich, V., Tang, Y., Cho, H., Chen, K., Mitchell, R., Cano, I., Zhou, T., et al. (2015). Xgboost: extreme gradient boosting. *R package version 0.4-2*, 1(4):1–4.
- Dewi, R., Saleh, A., Tanjung, D. Y. H., Jinan, A., et al. (2025). Enhancing machine learning algorithm performance for pcs diagnosis using smotenc on imbalanced data. *JITK (Jurnal Ilmu Pengetahuan dan Teknologi Komputer)*, 11(1):55–63.
- Hameed, S., Nauman, M., Akhtar, N., Fayyaz, M. A., and Nawaz, R. (2025). Explainable ai-driven depression detection from social media using natural language processing and black box machine learning models. *Frontiers in Artificial Intelligence*, 8:1627078.

- Hughes, L., Dwivedi, Y. K., Malik, T., Shawosh, M., Albashrawi, M. A., Jeon, I., Dutot, V., Appanderanda, M., Crick, T., De', R., et al. (2025). Ai agents and agentic systems: A multi-expert analysis. *Journal of Computer Information Systems*, pages 1–29.
- Liu, F., Niu, Y., Zhang, Q., Wang, K., Dong, Z., Wong, I. N., Cheng, L., Li, T., Duan, L., Li, K., et al. (2025). A foundational architecture for ai agents in healthcare. *Cell Reports Medicine*, 6(10).
- Ma, F.-q., He, C., Yang, H.-r., Hu, Z.-w., Mao, H.-r., Fan, C.-y., Qi, Y., Zhang, J.-x., and Xu, B. (2023). Interpretable machine-learning model for predicting the convalescent covid-19 patients with pulmonary diffusing capacity impairment. *BMC Medical Informatics and Decision Making*, 23(1):169.
- Nassreddine, G., Nassereddine, M., and Al-Khatib, O. (2025). Ensemble learning for network intrusion detection based on correlation and embedded feature selection techniques. *Computers*, 14(3):82.
- Paul, S., Olymon, K., Martinez, G. S., Sarkar, S., Yella, V. R., and Kumar, A. (2024). Mldsp: Bacterial promoter prediction tool using dna structural properties with machine learning and explainable ai. *Journal of Chemical Information and Modeling*, 64(7):2705–2719.
- Ribeiro, R. D., Gomes, V., Gomes, D., Gomes, G., de Souza, E. V., Gomes, A., Sousa, R. d. C. V., and Gomes, L. H. (2025). Impacto da inteligência artificial na medicina: Revisão bibliográfica sobre diagnóstico, tratamento e sistemas de apoio à decisão. *LU-MEN ET VIRTUS*, 16(44):217–233.
- Rodrigues, M. M., Barreto-Duarte, B., Vinhaes, C. L., Araújo-Pereira, M., Fukutani, E. R., Bergamaschi, K. B., Kristki, A., Cordeiro-Santos, M., Rolla, V. C., Sterling, T. R., et al. (2024). Machine learning algorithms using national registry data to predict loss to follow-up during tuberculosis treatment. *BMC Public Health*, 24(1):1385.
- Tanasă, A.-M., Oprea, S.-V., and Bâra, A. (2026). Designing an architecture of a multi-agentic ai-powered virtual assistant using llms and rag for a medical clinic. *Electronics*, 15(2):334.
- World Health Organization (2025). *Global Tuberculosis Report 2025*. World Health Organization, Geneva. Licence: CC BY-NC-SA 3.0 IGO.
- Yasin, P., Yimit, Y., Cai, X., Aimaiti, A., Sheng, W., Mamat, M., and Nijati, M. (2024). Machine learning-enabled prediction of prolonged length of stay in hospital after surgery for tuberculosis spondylitis patients with unbalanced data: a novel approach using explainable artificial intelligence (xai). *European Journal of Medical Research*, 29(1):383.