

Predicting Asthma Exacerbations in Digital Health Ecosystems Using Machine Learning Frameworks

Herbert Jung Sin, Martony Demes da Silva

Curso Superior de Tecnologia de Gestão de Dados
Centro de Educação Aberta e a Distância (CEAD)
Universidade Federal do Piauí (UFPI)
Teresina – PI – Brasil

herbert.sin@ufpi.edu.br, martony.silva@ufpi.edu.br

Abstract. *Predicting acute asthma exacerbations is critical for digital health and Patient-Generated Health Data integration. This study evaluates an intelligent screening model for the Afilaxy app to classify risk profiles using GINA clinical criteria. Four machine learning algorithms were evaluated on synthetic data: Logistic Regression, Random Forest, Gradient Boosting, and XGBoost. The XGBoost model achieved 91.3% AUC and 89.2% Recall at the optimal threshold. Despite strong classification performance, the 10.8% false-negative rate warrants clinical caution. This proof-of-concept establishes a foundation for future on-device deployment.*

Resumo. *A triagem do risco de exacerbações asmáticas é crítica para a saúde digital e dados gerados por pacientes (PGHD). Este trabalho avalia um modelo de triagem inteligente para o aplicativo Afilaxy, classificando perfis de risco com base nos critérios da GINA. Testou-se Regressão Logística, Random Forest, Gradient Boosting e XGBoost em dados sintéticos. O XGBoost obteve AUC de 91,3% e Sensibilidade de 89,2% no limiar otimizado. Apesar do forte desempenho, a taxa de 10,8% de falsos negativos exige cautela clínica. O estudo estabelece uma prova de conceito para futura implementação em dispositivos móveis.*

1. Introdução

A asma brônquica afeta mais de 260 milhões de pessoas no mundo, sendo responsável por 450 mil mortes anuais evitáveis [Global Initiative for Asthma 2024]. No Brasil, figura entre as principais causas de internação no SUS, gerando alto impacto socioeconômico por custos diretos e absenteísmo [Carvalho-Pinto et al. 2022]. Embora o manejo eficaz dependa do reconhecimento precoce de sinais inflamatórios pré-crise, o monitoramento ambulatorial exibe um “vazio assistencial” [Patel et al. 2021]: um hiato temporal entre consultas esporádicas que força decisões baseadas em relatos retrospectivos e imprecisos dos pacientes.

Como consequência desse modelo reativo, os dados coletados tardiamente são frequentemente afetados por vieses de memória e subestimação de sintomas leves. Preditores fundamentais, como variações bruscas no índice de qualidade do ar (*Air Quality Index* - AQI) e flutuações microclimáticas de umidade e temperatura, passam despercebidos pela triagem convencional até culminarem em crises agudas e urgências hospitalares.

Para parametrizar a gravidade dessas ocorrências, as diretrizes da *Global Initiative for Asthma* (GINA) categorizam o controle clínico do paciente com base nos sintomas das últimas quatro semanas (Tabela 1).

Tabela 1. Critérios GINA para Avaliação do Controle da Asma

Sintomas avaliados (últimas 4 semanas)	Classificação do Controle (Itens Ativos)
1. Sintomas diurnos ($> 2\times$ /semana)	Controlada: Nenhum item presente
2. Despertares noturnos por asma*	Parcialmente Controlada: 1 a 2 itens
3. Uso de medicação de alívio ($> 2\times$ /semana)	Não Controlada: 3 a 4 itens
4. Qualquer limitação de atividades físicas	

Fonte: Adaptado de GINA (2024). *Nota: O campo de sintomas noturnos integra o cálculo do score e o treinamento, enquanto a variável de tratamento preventivo (*onControllerMedication*), introduzida na v2.3.1 do aplicativo, permanece suspensa por ausência desse vetor no dataset proxy utilizado.

A expansão de dispositivos móveis viabilizou a coleta de Dados Gerados por Pacientes (*Patient-Generated Health Data* - PGHD). Contudo, traduzir esse fluxo heterogêneo em alertas clínicos acionáveis é um desafio, visto que scores heurísticos falham em mapear correlações não lineares. Para mitigar isso, este trabalho introduz a modelagem preditiva no ecossistema do aplicativo *Afilaxy* [Afilaxy 2025] (ativo em Android e em beta no iOS). Visando substituir sua lógica atual de pesos fixos por modelos inteligentes locais (*on-device*) sob escassez inicial de dados históricos nativos, adota-se um *dataset proxy* consolidado, enriquecido algoritmicamente para emular os módulos de *check-in* e sensores do aplicativo.

O objetivo geral deste artigo é avaliar a viabilidade teórica e estatística de algoritmos de aprendizado de máquina supervisionado para a classificação do risco de exacerbação asmática (definido pelo rótulo *high_risk* segundo as diretrizes GINA). Em termos pragmáticos, a "viabilidade computacional" investigada neste estudo refere-se especificamente à capacidade dos algoritmos de obter alto desempenho preditivo (priorizando a Sensibilidade) mantendo uma arquitetura de baixa complexidade paramétrica compatível com restrições de inferência móvel. A principal contribuição científica deste trabalho reside na formulação de uma abordagem de triagem integrada ao ecossistema *Afilaxy*, combinando a adaptação das diretrizes clínicas GINA para rotulamento automático e o estudo comparativo de modelos com foco em segurança clínica, estabelecendo os requisitos teóricos para a futura implantação em borda.

2. Referencial Teórico

O desenvolvimento de um modelo preditivo para o aplicativo *Afilaxy* exige a convergência entre a caracterização clínica das exacerbações, o processamento de fluxos contínuos de dados móveis, algoritmos de classificação e métricas voltadas à segurança em saúde.

2.1. Fisiopatologia da Asma, GINA e o Papel dos Dados de Saúde (PGHD)

A asma é uma doença inflamatória crônica das vias aéreas inferiores, caracterizada por hiperresponsividade brônquica e limitação variável ao fluxo aéreo, cujos sintomas (sibilância, dispneia e tosse) oscilam frente a gatilhos exógenos, como infecções e variações no Índice de Qualidade do Ar (*Air Quality Index* - AQI) [Guarnieri and Balmes 2014, Global Initiative for Asthma 2024]. As exacerbações representam o agravamento desse pro-

cesso. O manejo preconizado pela *Global Initiative for Asthma* (GINA) enfatiza o monitoramento ativo em curtas janelas temporais para estratificar o risco de crises [Global Initiative for Asthma 2024], justificando o uso de ferramentas digitais para automatizar essa detecção.

Nesse cenário, os Dados Gerados por Pacientes (*Patient-Generated Health Data* - PGHD) coletados via saúde móvel (*mHealth*) reconfiguraram o monitoramento clínico. Ao contrário dos registros transversais de Prontuários Eletrônicos (*Electronic Health Records* - EHR), o PGHD capta o cotidiano do indivíduo de forma longitudinal e contínua através de sintomas autorrelatados e telemetria climática passiva (temperatura, umidade e AQI por geolocalização) [Demiris et al. 2019, Patel et al. 2021]. Essa fusão de dados provê o arcabouço informacional necessário para treinar algoritmos preditivos supervisionados [Merchant et al. 2016].

2.2. Algoritmos de Classificação Supervisionada

Para mapear o risco de crises a partir do PGHD, quatro arquiteturas destacam-se pela robustez e eficiência computacional, cobrindo desde vieses lineares interpretáveis até abordagens de comitês (*ensembles*).

- **Regressão Logística:** Modelo probabilístico linear que estima a ocorrência de um evento aplicando a função sigmóide sobre uma combinação linear ponderada dos atributos de entrada [Hastie et al. 2009]. Embora limitada para capturar interações fortemente não lineares sem engenharia prévia de atributos, sua baixa complexidade computacional e alta interpretabilidade tornam-na um *baseline* indispensável no contexto médico.
- **Random Forest:** Método baseado em aprendizado em conjunto (*ensemble learning*) que constrói uma multiplicidade de árvores de decisão independentes durante o treinamento [Breiman 2001]. O modelo mitiga o sobreajuste (*overfitting*) injetando aleatoriedade na amostragem dos dados via *bagging* e na seleção restrita de atributos [James et al. 2021], consolidando a predição final por votação majoritária.
- **GBM e XGBoost:** Diferente do aprendizado paralelo, as arquiteturas de *boosting* operam sequencialmente. O GBM minimiza a função de perda treinando novas árvores sequenciais [Hastie et al. 2009], enquanto o *XGBoost* introduz penalidades de regularização estrutural (L_1 e L_2) em sua função de custo, prevenindo o sobreajuste. Sua arquitetura paralela em nível de *hardware* otimiza a execução local (*on-device*) no ecossistema *Aflaxy*.

2.3. Sensibilidade (Recall) como Métrica de Segurança Clínica

A validação de modelos aplicados à medicina diagnóstica difere de problemas genéricos onde a acurácia global é priorizada. Em ambientes de saúde digital, o custo clínico de um Falso Negativo (classificar um paciente em risco iminente como estável) sobrepuja drasticamente o impacto de um Falso Positivo, visto que a omissão de um alerta impede uma intervenção preventiva oportuna [Altman and Bland 2000]. Diante dessa assimetria, a Sensibilidade (ou *Recall*) assume o papel de métrica central de desempenho, pois quantifica a proporção exata de casos de risco real que o pipeline computacional detectou com sucesso [Altman and Bland 2000], garantindo o patamar de segurança mandatório para ferramentas de suporte à decisão.

3. Trabalhos Relacionados

O mapeamento de riscos respiratórios e a gestão proativa da asma por meio de ecossistemas digitais têm impulsionado soluções baseadas em Inteligência Artificial, concentrando-se na transição de sistemas reativos para o monitoramento contínuo suportado por Dados Gerados por Pacientes (PGHD) e telemetria ambiental. O cruzamento de históricos clínicos com variáveis exógenas é uma estratégia consolidada para prever crises agudas. Bousquet et al. 2017 estruturou a iniciativa europeia MASK-asthma, demonstrando que o uso de aplicativos móveis para o autorrelato diário de sintomas eleva a adesão do paciente e fornece um panorama longitudinal confiável da doença, embora restrito a heurísticas clínicas estáticas. Expandindo esse cenário, Merchant et al. 2016 integraram sensores acoplados a inaladores de resgate para capturar passivamente o uso da medicação junto a dados de geolocalização, correlacionando com sucesso as flutuações do índice de qualidade do ar (*Air Quality Index* - AQI) como preditores eficientes de exacerbações urbanas.

No âmbito dos algoritmos classificadores, pesquisas buscam otimizar a assertividade frente a dados heterogêneos. A revisão sistemática de Alotaibi et al. 2020 evidenciou que modelos baseados em comitês de árvores, como *Random Forest* e *Gradient Boosting*, superam abordagens lineares tradicionais devido à capacidade de processar interações não lineares entre o microclima e os sintomas biológicos. De forma complementar, Patel et al. 2021 discutiram os gargalos de integração desses ecossistemas, apontando que a maior barreira para a adoção de IA na saúde móvel não reside na acurácia matemática isolada, mas sim na necessidade de os modelos operarem com alta sensibilidade para garantir a segurança do paciente, demandando baixa complexidade estrutural para viabilizar a execução local (*on-device AI*). A Tabela 2 sintetiza esse panorama metodológico.

Tabela 2. Comparativo metodológico dos trabalhos relacionados

Autor (Ano)	Técnicas Base	Origem Dados	Variáveis	Foco Principal
Bousquet (2017)	Heurísticas Clínicas	App Mobile Ativo	Sintomas diários	Adesão ao tratamento
Merchant (2016)	Correlação Espacial	Sensores + IoT	Resgate e AQI	Mapeamento de gatilhos
Alotaibi (2020)	RF / Boosting	Bases Transversais	Sinais e Clima	Benchmark algoritmos
Patel (2021)	Análise Estrutural	Prontuários PGHD	+ Clínicas móveis	e Integração e viabilidade

Fonte: Elaborada pelos autores (2025).

Apesar dos avanços na captação de PGHD e no benchmark de classificadores, observa-se na literatura uma lacuna de engenharia: a maioria dos modelos robustos funciona em silos na nuvem, dependendo de conexão constante à internet e ignorando restrições de infraestrutura do smartphone. O diferencial deste trabalho reside no desenvolvimento de uma abordagem embarcada integrada ao ecossistema real do aplicativo *Afilaxy*. A arquitetura proposta calibra o modelo XGBoost focado na Sensibilidade clínica (89,2% no limiar otimizado) para blindar o paciente contra falsos negativos, validando a viabilidade de conversão do artefato final para inferência leve na borda (*on-device AI*), eliminando a dependência de servidores e protegendo a autonomia do dispositivo.

4. Metodologia

A arquitetura metodológica deste trabalho fundamenta-se na Saúde Digital Preditiva e em *Machine Learning* para modelagem de risco respiratório. O desenho experimental estruturado garante o controle de sobreajuste (*overfitting*) e a robustez estatística frente a dados heterogêneos (Figura 1).¹

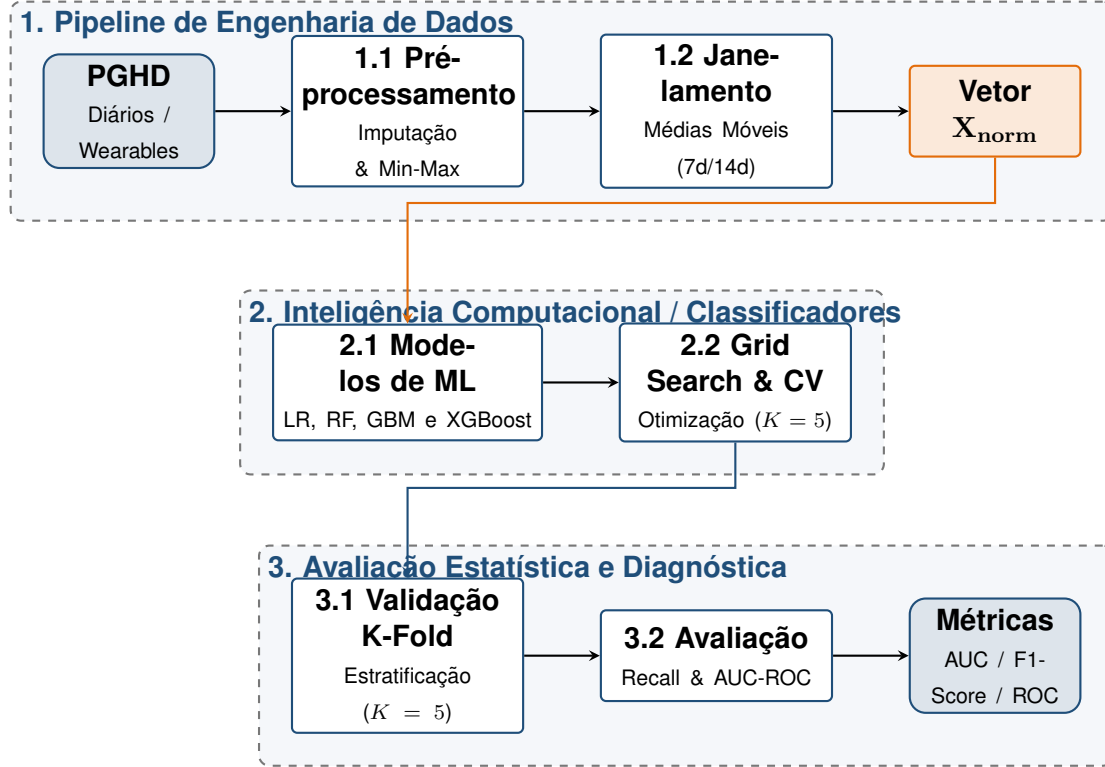


Figura 1. Arquitetura de dados ponta a ponta e engenharia de inteligência preditiva para monitoramento da asma crônica.

4.1. Definição do Target e Justificativa Clínica

A modelagem no *Afilaxy* foca na predição de exacerbações agudas no intervalo entre consultas, e não no diagnóstico de asma. O rótulo original da base (*Diagnosis*) foi descartado para mitigar *data leakage*. Em seu lugar, estruturou-se a variável binária *high_risk* ($y \in \{0, 1\}$) via engenharia de atributos sobre o histórico de sintomas.

A rotulagem baseou-se nas diretrizes da GINA, contabilizando cinco sintomas de *check-ins* diários: chiado (*Wheezing*), falta de ar (*ShortnessOfBreath*), aperto no peito (*ChestTightness*), tosse (*Coughing*) e sintomas noturnos (*NighttimeSymptoms*). O escore ($symptom_score \in [0, 5]$) mapeia o controle da asma segundo a Equação (1), definindo alto risco a partir de 3 indicadores:

$$high_risk = \begin{cases} 1, & \text{se } symptom_score \geq 3 \\ 0, & \text{se } symptom_score < 3 \end{cases} \quad (1)$$

¹O pipeline e os scripts estão disponíveis publicamente em: <https://github.com/herb-sin/afilaxy-ml>

Após a parametrização nos 2.392 registros da base *proxy*, a distribuição do *target* mostrou-se naturalmente balanceada, dispensando amostragem artificial: a classe *high_risk* = 1 representa **58,1%** (1.390 amostras) e *high_risk* = 0 representa **41,9%** (1.002 amostras). Essa proporção é consistente com a prevalência individual dos sintomas (50%–60%) e próxima ao modelo binomial esperado (57,7%), descartando viés no desenho do rótulo.

4.2. Descrição Detalhada das Etapas Experimentais

A execução experimental da metodologia proposta é estruturada em quatro etapas encaixadas, que englobam desde a estruturação e pré-processamento dos dados brutos até a otimização de hiperparâmetros e validação estatística dos classificadores, conforme detalhado a seguir.

4.2.1. Etapa 1: Ingestão e Tratamento de Dados

O ecossistema inicia-se com a coleta e agregação de dados de saúde gerados por pacientes (PGHD - *Patient-Generated Health Data*) em conformidade teórica com as perspectivas de Patel et al. (2021) e Demiris et al. (2019). O referencial metodológico para o pré-processamento segue as diretrizes de integridade de dados biológicos e de telemetria digital:

- **Tratamento de Omissões e Ruídos:** Registros nulos provenientes de falhas de sincronização de dispositivos vestíveis ou lapsos de preenchimento de diários digitais são tratados via imputação baseada na mediana móvel do paciente para variáveis contínuas, ou pela moda local para variáveis categóricas, prevenindo vieses de exclusão.
- **Normalização:** Para garantir a estabilidade computacional e a convergência dos algoritmos, aplica-se a transformação de escala linear Min-Max. Este procedimento realiza o mapeamento dos atributos numéricos originais uniformemente dentro do intervalo centrado entre zero e um, neutralizando disparidades de magnitude entre diferentes unidades de medida sem distorcer a dispersão dos dados.

4.2.2. Etapa 2: Engenharia de Atributos (*Feature Engineering*)

Para capacitar os classificadores a capturarem a dinâmica das exacerbações de asma, a base de dados bruta é transformada através da extração de três grupos de atributos preditivos:

1. **Janelas Temporais de Sintomas e Função Pulmonar:** Estruturação de variáveis agregadas (médias e desvios padrões móveis de 7 e 14 dias) a partir do score de sintomas diários relatados e dos parâmetros espirométricos dinâmicos de Volume Expiratório Forçado (FEV_1) e Capacidade Vital Forçada (FVC), permitindo mapear declínios funcionais graduais à luz das recomendações globais da GINA [Global Initiative for Asthma 2024].
2. **Variáveis Ambientais Exógenas:** Alinhamento georreferenciado de dados macroclimáticos (temperatura, umidade relativa do ar) e níveis de poluentes particu-

lados ($PM_{2.5}$ e PM_{10})², fundamentando-se na evidência clínica estabelecida por Guarneri e Balmes (2014) sobre o impacto da poluição na hiperresponsividade brônquica.

3. **Variabilidade de Variáveis Fisiológicas e do Ecossistema Mobile:** Mapeamento do desvio padrão móvel da frequência cardíaca, saturação de oxigênio (SpO_2) e padrões de eficiência do sono. Embora a arquitetura do aplicativo Afilaxy esteja preparada via estruturas (*scaffolds*) no HealthConnect e HealthKit para integrar tais módulos, eles permanecem inativos na versão v2.3.1 (retornando nulo no modelo de dados). Por essa razão, e pela ausência desses vetores na base proxy laboratorial, essas variáveis foram omitidas do treinamento do classificador final.

4.2.3. Etapa 3: Modelagem Preditiva e Otimização

O núcleo de inteligência computacional foi estruturado a partir de uma abordagem comparativa com quatro classes de algoritmos complementares: Regressão Logística (adotada como *baseline*), *Random Forest*, *Gradient Boosting Machine* (GBM) e o *XGBoost*.

Para extrair o teto de desempenho e garantir uma comparação equitativa, as arquiteturas foram submetidas a um processo de busca exaustiva em grade (*Grid Search*) integrado a um esquema de validação cruzada estratificada ($K = 5$) restrito ao subconjunto de treinamento. O espaço de busca multidimensional concentrou-se nos parâmetros estruturais mais críticos de cada modelo:

- **Regressão Logística:** Força da regularização inversa ($C \in \{0.01, 0.1, 1.0, 10.0\}$);
- **Modelos baseados em Árvores (RF, GBM e XGBoost):** Volume do comitê ($n_estimators \in \{50, 100, 200\}$) e profundidade máxima das ramificações ($max_depth \in \{3, 5, 7\}$);
- **Algoritmos de Boosting (GBM e XGBoost):** Fator de encolhimento ou taxa de aprendizado ($learning_rate \in \{0.01, 0.05, 0.1, 0.2\}$).

Os hiperparâmetros ótimos selecionados automaticamente em cada partição foram consolidados nos artefatos finais para a etapa subsequente de validação estatística e diagnóstica.

4.2.4. Etapa 4: Avaliação de Desempenho e Validação Estatística

Para mitigar o risco de estimativas excessivamente otimistas e garantir a generalização do modelo para novos indivíduos, adotou-se um protocolo de validação cruzada estratificada em 5 partições ($K = 5$) no conjunto de treino, seguido pela avaliação final em um conjunto de testes independente contendo 20% dos registros totais. Devido à reestruturação da variável resposta fundamentada nas diretrizes GINA, que resultou em uma distribuição equilibrada de classes (58,1% alto risco versus 41,9% baixo risco), o pipeline dispensou a aplicação de algoritmos de amostragem artificial como o SMOTE, preservando a distribuição epidemiológica real dos dados.

²Na versão v2.3.1 do app, apenas o índice AQI composto é armazenado no banco de dados. Embora as concentrações individuais de $PM_{2.5}$ e PM_{10} sejam consultadas via API da WAQI, elas não são persistidas no histórico do paciente, exigindo um ajuste de modelo de dados antes do retreinamento.

O desempenho preditivo da arquitetura foi quantificado analiticamente a partir da matriz de confusão, fundamentando-se nos conceitos de validação de testes diagnósticos descritos por Altman e Bland (2000). O comportamento do classificador final foi verificado por meio da extração da Sensibilidade (*Recall*), Especificidade, Valor Preditivo Positivo (VPP), Valor Preditivo Negativo (VPN) e a Área Sob a Curva ROC (AUC-ROC), assegurando o equilíbrio ideal entre a segurança clínica do usuário e a especificidade operacional do aplicativo.

4.3. Formulação da Tarefa Preditiva e Limitações Metodológicas

É fundamental pontuar que a formulação experimental adotada neste trabalho configura uma *classificação de estado/triagem de risco corrente* baseada na presença de sintomas ativos e variáveis exógenas simultâneas, e não uma *predição temporal antecipada* com janela de horizonte futuro delimitada (ex.: prever uma crise com N dias de antecedência).

Adicionalmente, como o rótulo *high_risk* é derivado das diretrizes GINA aplicando regras sobre os sintomas autorrelatados (*Wheezing*, *ShortnessOfBreath*, etc.), esses indicadores foram **excluídos** da matriz de atributos X para prevenir vazamento direto de dados (*data leakage*). Em seu lugar, a variável sintética *riskScore* — calculada a partir de exposição ambiental, função pulmonar e histórico de sintomas — introduz uma sobreposição indireta entre preditores e *target*. Esse fator contribui parcialmente para os valores de AUC observados, devendo o modelo ser interpretado como um emulador da lógica de triagem GINA adaptada ao ecossistema *Afilaxy*, e não como uma prova de aprendizado causal autônomo.

5. Resultados e Discussão

Antes do treinamento, realizou-se uma análise estatística descritiva na base *proxy*, composta por 15 variáveis e 2.392 registros. O espaço de atributos compreende: 5 dimensões clínicas e demográficas (idade de 5 a 80 anos, gênero, etnia, histórico de tabagismo e IMC de 15,0 a 40,0 kg/m²); 2 parâmetros espirométricos ($FEV_1 \approx 2,54$ L/s e $FVC \approx 3,00$ L/s); 4 variáveis ambientais e comportamentais (exposição à poluição, pólen, poeira e atividade física induzida); e 4 variáveis sintéticas mapeadas ao ecossistema *Afilaxy* (*riskScore*, ocorrência de crise anterior, disponibilidade do inalador de resgate e contagem de acionamentos de emergência). Os 5 indicadores de sintomas ativos foram excluídos da matriz de atributos X para evitar vazamento de dados, conforme detalhado na Seção 4.3. Essa granularidade multifatorial provê o suporte analítico necessário para modelar a estabilidade clínica do paciente e simular o ecossistema de dados heterogêneos do aplicativo.

5.1. Análise do Desempenho Global dos Classificadores

Os algoritmos foram avaliados em um conjunto de testes independente com 20% da base original. A reestruturação do *target* via diretrizes GINA resultou em uma distribuição equilibrada para o aprendizado supervisionado, refletindo a realidade epidemiológica sem a necessidade de amostragem artificial.

A otimização via *Grid Search* com validação cruzada no treino definiu os hiperparâmetros do modelo *XGBoost*: taxa de aprendizado (*learning_rate*) de 0,20, profundidade máxima (*max_depth*) igual a 3 e 100 estimadores ($n_estimators = 100$). A profundidade rasa atuou como regularizador estrutural, reduzindo o risco de sobreajuste nos dados transversais e favorecendo a capacidade de generalização do classificador.

A Tabela 3 consolida as métricas obtidas por cada arquitetura na fase de avaliação.

Tabela 3. Desempenho Comparativo dos Modelos no Conjunto de Teste

Modelo	Média (AUC)	CV	AUC (Teste)	Sensibilidade (Recall)	F1-Score
Regressão Logística	0,942		0,933	0,867	0,870
Random Forest	0,862		0,881	0,899	0,843
Gradient Boosting	0,921		0,910	0,860	0,854
XGBoost	0,922		0,913	0,867	0,859

Fonte: Dados extraídos dos experimentos computacionais (2025).

Os dados demonstram que os modelos de *boosting* (XGBoost e Gradient Boosting) e a Regressão Logística operaram em faixa competitiva, com AUC entre 0,910 e 0,933 no conjunto de teste. A Regressão Logística obteve a maior AUC individual (0,933) e o maior *F1-Score* (0,870), enquanto o Random Forest registrou o maior Recall (0,899). O XGBoost (AUC = 0,913; *F1* = 0,859) apresentou desempenho competitivo com os demais modelos de *boosting*. Na ausência de testes estatísticos de hipótese formais (como Wilcoxon ou teste *t* pareado nas dobras da validação cruzada), não se deve inferir superioridade estatística estrita de uma arquitetura sobre a outra. O XGBoost é adotado como modelo de referência para implantação (*deployment*) em razão de suas propriedades de regularização estrutural (L_1 e L_2) e compatibilidade com inferência em borda (*on-device AI*), aspectos críticos no ecossistema *Afilaxy*.

5.2. A Métrica de Sensibilidade como Critério de Segurança Clínica

Sob a ótica de segurança em saúde digital, onde a omissão de um alerta preventivo pode levar ao agravamento do quadro respiratório, o modelo *Random Forest* obteve o maior Recall no limiar padrão (**89,9%**), enquanto XGBoost e Regressão Logística alcançaram **86,7%**. Com ajuste do limiar de decisão (Seção 5.3), o Recall do XGBoost pode ser elevado para **89,2%**.

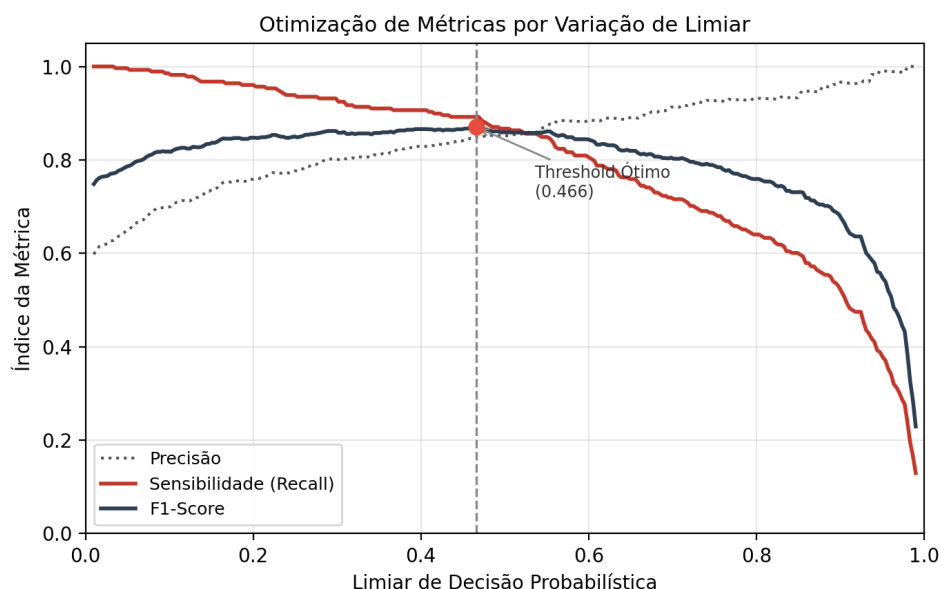
Embora essa taxa represente uma alta capacidade de detecção no conjunto testado, é indispensável reconhecer a relevância clínica da margem de erro remanescente: um *Recall* de 89,2% implica que **10,8% dos indivíduos em situação de alto risco não foram identificados pelo modelo** (falsos negativos). Em uma coorte em produção, essa taxa representa pacientes vulneráveis sem o devido alerta. Portanto, o resultado deve ser interpretado como um avanço na triagem automatizada, mas que não substitui o acompanhamento médico e nem garante a eliminação absoluta de crises.

5.3. Impacto da Otimização do Limiar de Decisão

A Figura 2 ilustra o comportamento dinâmico das métricas de validação frente à variação do limiar probabilístico do modelo XGBoost. O mapeamento visual evidencia o *trade-off* clássico em saúde: a elevação do limiar aumenta a Precisão, porém provoca queda acentuada na Sensibilidade.

Para o ecossistema *Afilaxy*, o ponto de operação padrão em 0,50 foi substituído pela calibração do limiar em **0,466** (sinalizado pela linha tracejada), elevando o *Recall* para **89,2%**. O limiar foi calibrado para maximizar o F1-Score, o que elevou o Recall sem degradar a precisão de forma proporcional — um compromisso alinhado ao princípio de

Figura 2. Comportamento das Métricas de Validação em Função do Limiar de Decisão



Fonte: Gráfico extraído do pipeline de calibração do experimento (2025).

precaução clínica, que prioriza a captura de exacerbações mesmo ao custo de um pequeno número adicional de falsos positivos.

5.4. Análise de Relevância dos Atributos

Para conferir interpretabilidade ao modelo de referência, extraiu-se a importância de atributos (*Feature Importance*) do algoritmo *XGBoost*, baseada no ganho relativo médio trazido para as divisões estruturais do *ensemble* (Figura 3).

Essa distribuição reflete a capacidade discriminativa no dataset, **sem implicar causalidade direta ou mecanismos fisiopatológicos**.

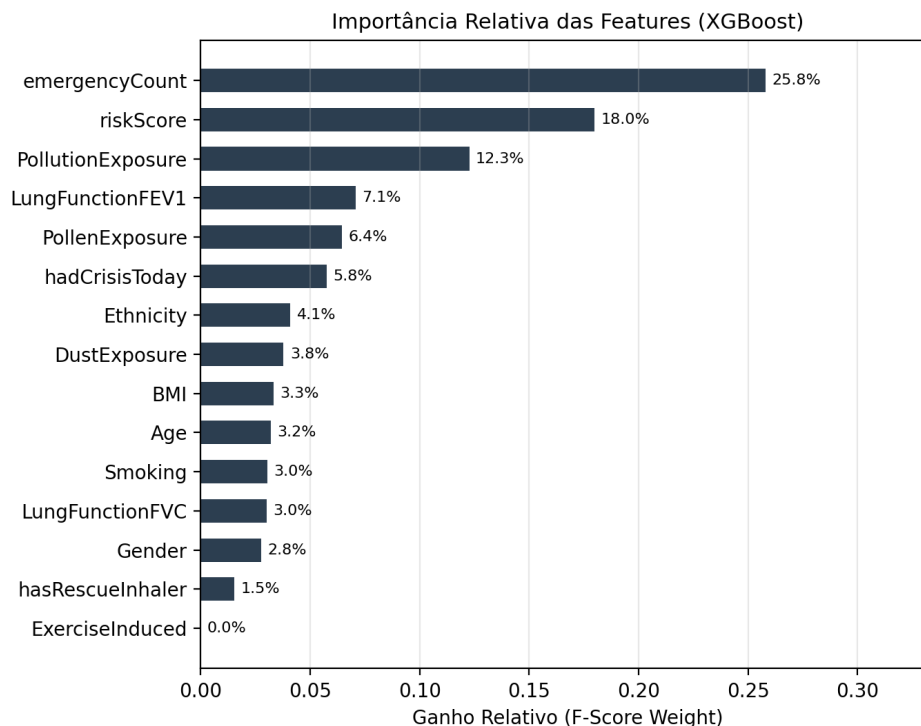
A variável *emergencyCount* (contagem de acionamentos do botão de emergência no app) emergiu como o preditor de maior peso (25,8%), seguida por *riskScore* (18,0%) — saída do motor heurístico do *Afilaxy* — e *PollutionExposure* (12,3%). Os indicadores espirométricos *FEV₁* (7,1%) e *PollenExposure* (6,4%) completam os cinco atributos mais relevantes, convergindo com o conhecimento clínico sobre gatilhos de exacerbação asmática. Notadamente, a inclusão do *riskScore* como *feature* cria dependência do modelo com a regra heurística estática, devendo ser removida em estudos de ablação futuros. Tais padrões mostram-se **compatíveis com o conhecimento clínico**, devendo ser validados prospectivamente com dados longitudinais nativos do aplicativo.

5.5. Discussão de Viabilidade Técnica e Implantação no App Real

A avaliação do modelo provê diretrizes para o desenvolvimento contínuo do aplicativo *Afilaxy*:

- **Superação do Motor Heurístico:** A transição de regras fixas para modelos baseados em árvores permite identificar padrões não lineares entre clima e sintomas, tornando a triagem de risco mais dinâmica.

Figura 3. Pesos de Importância Relativa dos Atributos no XGBoost. Nota: Os rótulos refletem as variáveis da base proxy (Asthma Disease Dataset); o mapeamento para os campos nativos do app requer validação prévia.



Fonte: Dados do experimento (2025).

- **Uso de Dados Proxy e Necessidade de Validação Longitudinal:** É indispensável reforçar que os resultados numéricos foram obtidos em uma base sintética e transversal (*Asthma Disease Dataset*). Em produção real, o sistema enfrentará séries temporais ruidosas, falhas de preenchimento por parte do usuário e ausência de atributos demográficos e clínicos basais (como idade, IMC e histórico de tabagismo, que atualmente não são coletados pelo perfil de usuário do app), exigindo novos testes em dados longitudinais nativos.
- **Hipótese de Inferência na Borda (On-Device AI):** A conversão do modelo para *.tflite* ou *.mlmodel* fundamenta a viabilidade de execução local, mas essa integração não está implementada na v2.3.1. Além disso, a computação das janelas de 7 e 14 dias exige um pipeline local (ou Cloud Function) para agregar o histórico do Firestore. Salienta-se que **não foram realizados testes empíricos de bateria, latência, RAM ou tamanho de arquivo**, sendo essa validação um trabalho futuro indispensável.
- **Dependência de Hardware para Parâmetros Espirométricos:** A utilização das variáveis de função pulmonar (FEV_1 e FVC) no modelo pressupõe a integração com espirômetros portáteis certificados via Bluetooth. Esta barreira de hardware deve ser considerada no roadmap de implantação do ecossistema.

6. Conclusão

Este trabalho apresentou uma prova de conceito para triagem inteligente do risco de exacerbação asmática no aplicativo *Afilaxy*. O modelo *XGBoost* demonstrou elevada ca-

pacidade discriminativa em dados sintéticos, alcançando AUC de 91,3% e Sensibilidade de 89,2% no limiar otimizado, superando abordagens puramente heurísticas no monitoramento contínuo.

Como limitações, destacam-se o viés do *dataset proxy* transversal, a sobreposição funcional entre atributos e o rótulo de risco, e a ausência de testes de desempenho em hardware móvel. Destaca-se que o *Afilaxy* possui arquitetura para integrar 7 variáveis adicionais (como frequência cardíaca e SpO_2 via HealthConnect/HealthKit), atualmente pendentes de ativação nativa. Trabalhos futuros devem focar no retreinamento dos modelos com dados longitudinais reais, testes de significância estatística, modelagem temporal preditiva e *benchmarking* empírico de latência e consumo energético da inferência *on-device*.

Referências

- Afilaxy (2025). Afilaxy: A rede social de quem tem asma. <https://afilaxy.com/>. Acesso em: 23 ago. 2026.
- Alotaibi, S. et al. (2020). Machine learning applications in asthma monitoring and management: A systematic review. *IEEE Access*, 8:109508–109521.
- Altman, D. G. and Bland, J. M. (2000). Diagnostic tests 2: Predictive values. *BMJ*, 321(7258):430.
- Bousquet, J. et al. (2017). Next-generation care pathways for asthma with mobile technology: Mask-asthma (mobile airways sentinel network). *Journal of Allergy and Clinical Immunology*, 140(5):1245–1249.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1):5–32.
- Carvalho-Pinto, R. P. et al. (2022). Recomendações para o manejo da asma grave da sociedade brasileira de pneumologia e fisiologia - 2021. *Jornal Brasileiro de Pneumologia*, 48(1):e20210273.
- Demiris, G. et al. (2019). Patient-generated health data as a resource for research and clinical care. *Methods of Information in Medicine*, 58(S 01):e1–e2.
- Global Initiative for Asthma (2024). *Global Strategy for Asthma Management and Prevention*. Global Initiative for Asthma (GINA). Disponível em: <https://www.ginasthma.org>. Acesso em: 17 jun. 2026.
- Guarnieri, M. and Balmes, J. R. (2014). Outdoor air pollution and asthma. *The Lancet*, 383(9918):724–734.
- Hastie, T., Tibshirani, R., and Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer Science & Business Media, New York, 2nd edition.
- James, G., Witten, D. C., Hastie, T., and Tibshirani, R. (2021). *An Introduction to Statistical Learning: with Applications in R*. Springer, New York, 2nd edition.
- Merchant, R. K. et al. (2016). Integrating digital health tools into the management of chronic respiratory diseases. *JMIR mHealth and uHealth*, 4(2):e40.
- Patel, S. et al. (2021). Bridging the gap: patient-generated health data and digital health ecosystems. *npj Digital Medicine*, 4(1):112.