DEMONSTRAÇÃO COM DADOS FICTÍCIOS · nenhum paciente real · os números abaixo testam o software, não são evidência clínica

Cledon · demonstrador da rede federada

Piloto simulado: câncer de esôfago, resposta ao tratamento neoadjuvante

Três hospitais simulados guardam seus próprios registros. O coordenador envia uma análise aprovada, cada hospital executa localmente e devolve apenas resultados agregados. Gerado em 13/11/2026.

1. Rede participante

Hospitais fictícios e qualidade das bases. Nenhuma lista de pacientes sai dos hospitais: abaixo estão apenas contagens, com células menores que 5 suprimidas.

Hospital	Papel	Pacientes	Faltantes: albumina	RNL	Perda de peso	Chimarrão	Cachaça
Hospital A (Sul, simulado)	desenvolvimento	220	24	10	20	60	73
Hospital B (Minas Gerais, simulado)	desenvolvimento	160	10	<5	15	48	61
Hospital C (Bahia, simulado; teste externo)	teste externo	110	9	5	11	41	46

2. Protocolo do estudo

Item	Conteúdo
Pergunta	Em pacientes com carcinoma epidermoide de esôfago operados após quimiorradioterapia neoadjuvante, dados clínicos e laboratoriais disponíveis antes do tratamento preveem a resposta patológica completa, com desempenho mantido em outro hospital?
Objetivo secundário	Descrever, por hospital, as exposições documentadas em prontuário (tabagismo, álcool, chimarrão quente, cachaça), sem inferência causal.
Desfecho	resposta patológica completa (ypT0N0) no exame da peça cirúrgica
Data da previsão	início do tratamento neoadjuvante
População-alvo	Pacientes posteriormente operados após quimiorradioterapia neoadjuvante; o desempenho não se aplica automaticamente a todos que iniciam o tratamento.
Variáveis aprovadas	idade, sexo, ct, cn, localizacao, ecog, albumina_g_dl, hemoglobina_g_dl, rnl, perda_peso_pct, tabagismo_atual, alcool_pesado, chimarrao_quente, cachaca_regular, pcr
Modelo simples (comparador)	Tumor cT3–4, Linfonodo cN+
Modelo completo	Idade (+10 anos), Sexo feminino, Tumor cT3–4, Linfonodo cN+, Terço inferior, ECOG ≥ 1, Albumina (+1 g/dL), Hemoglobina (+1 g/dL), Razão neutrófilo/linfócito (log), Perda de peso (+1 ponto %)
Avaliação aprovada	Mesmos casos completos do modelo completo; semente e 1.000 reamostragens fixadas antes do teste externo.
Controles de saída	células < 5 suprimidas (com supressão complementar); modelo só com ≥ 50 casos completos; no máximo 200 rodadas de modelo por hospital; nenhuma linha individual
Versão e impressão digital	0.2-demonstracao · SHA-256 4b20710bcf887a97…

Cada hospital mantém a própria lista de protocolos aprovados, identificados pela impressão digital (SHA-256). Qualquer alteração no protocolo muda a impressão digital e exige nova aprovação.

3. Execução controlada

O coordenador fez 44 pedidos aos hospitais; 6 foram recusados. Cada hospital registrou todos os pedidos na sua auditoria local. As tentativas abaixo foram feitas de propósito e precisavam ser bloqueadas. O papel de cada hospital (desenvolvimento ou teste externo) é imposto pelo próprio hospital, não só pelo coordenador; antes do teste externo, cada modelo é congelado e registrado no hospital C, que aceita avaliá-lo uma única vez. O conjunto de casos e os parâmetros de avaliação também são fixados pelo protocolo.

Tentativa	Resposta do hospital
Exportar a lista de pacientes	BLOQUEADO: operação 'exportar_linhas' não está prevista no protocolo
Pedir o CEP dos pacientes (variável fora do protocolo)	BLOQUEADO: variáveis fora do protocolo: cep_ficticio
Usar uma versão do protocolo alterada sem nova aprovação	BLOQUEADO: protocolo não aprovado por este hospital (versão diferente da aprovada)
Treinar o modelo no hospital reservado para teste externo	BLOQUEADO: este hospital está reservado para teste externo e não participa do desenvolvimento

Tentativa	Resposta do hospital
Avaliar no teste externo um modelo que não foi congelado e registrado	BLOQUEADO: o teste externo só avalia um modelo congelado e registrado neste hospital
Repetir a avaliação externa do mesmo modelo congelado	BLOQUEADO: este modelo congelado já foi avaliado aqui; o teste externo é feito uma única vez

Objetivo secundário: exposições documentadas em prontuário, por hospital. Contagens com supressão de células pequenas e supressão complementar (para que a célula suprimida não possa ser recuperada pela diferença com o total).

Exposição documentada	Registro	Hospital A	Hospital B	Hospital C
Tabagismo atual	sim	113	68	40
Tabagismo atual	não	107	92	70
Tabagismo atual	não registrado	0	0	0
Álcool (consumo pesado)	sim	86	56	30
Álcool (consumo pesado)	não	134	104	80
Álcool (consumo pesado)	não registrado	0	0	0
Chimarrão quente	sim	104	9	0
Chimarrão quente	não	56	103	69
Chimarrão quente	não registrado	60	48	41
Cachaça regular	sim	9	29	10
Cachaça regular	não	138	70	54
Cachaça regular	não registrado	73	61	46

Faltantes e observados de cada variável são suprimidos em conjunto, em todas as operações: quando há menos de 5 faltantes, o número de observados também aparece como intervalo, para que uma consulta não desfaça a supressão de outra. Campos vazios ou com valor inválido contam como "não registrado" e não entram nos modelos.

Descrição de registros, não comparação de riscos: sem um grupo de comparação, estas contagens não dizem se uma exposição aumenta o risco. Em prontuários reais, "não registrado" costuma ser frequente e não significa "não exposto".

4. Resultado científico

Modelo desenvolvido de forma federada nos hospitais A e B (305 casos completos, 95 respostas completas; 6 rodadas). Em cada rodada, cada hospital devolveu apenas um vetor de 11 números (gradiente) e uma matriz 11×11 (hessiana). O resultado é idêntico ao de juntar as bases, sem juntá-las (verificado em `testes.py`).

Variável	Razão de chances estimada	IC 95%	Valor plantado nos dados fictícios
Idade (+10 anos)	0,85	0,64–1,14	0,95
Sexo feminino	1,58	0,86–2,90	1,16
Tumor cT3–4	0,57	0,33–0,98	0,55
Linfonodo cN+	0,67	0,39–1,13	0,61
Terço inferior	0,61	0,35–1,08	0,82
ECOG ≥ 1	0,41	0,24–0,72	0,74
Albumina (+1 g/dL)	1,79	0,98–3,27	1,57
Hemoglobina (+1 g/dL)	1,18	0,98–1,41	1,11
Razão neutrófilo/linfócito (log)	0,66	0,39–1,14	0,58
Perda de peso (+1 ponto %)	0,93	0,87–0,98	0,96

A coluna da direita mostra os efeitos inventados na geração dos dados. 9 dos 10 intervalos de 95% contêm o valor plantado (o esperado é cerca de 95% deles; com amostras pequenas, alguns ficam de fora por acaso). Isso mostra que o fluxo funciona; não diz nada sobre pacientes reais.

Comparação com o modelo simples e teste externo. O hospital C não participou do desenvolvimento. Os dois modelos são avaliados nos mesmos pacientes (casos completos do modelo completo). O desempenho em A e B é aparente, medido nos hospitais de desenvolvimento; não substitui validação interna com correção de otimismo.

Hospital	Modelo	Casos	Respostas completas	AUC (IC 95%)
A · desenvolvimento	simples (cT, cN)	171	53	0,59 (0,50–0,68)
A · desenvolvimento	completo	171	53	0,69 (0,60–0,78)
B · desenvolvimento	simples (cT, cN)	134	42	0,57 (0,47–0,67)
B · desenvolvimento	completo	134	42	0,76 (0,66–0,84)
C · teste externo	simples (cT, cN)	89	35	0,50 (0,39–0,61)
C · teste externo	completo	89	35	0,62 (0,49–0,74)

Calibração:

Hospital	Modelo	Brier	Observado/esperado	Inclinação de calibração
A · desenvolvimento	simples (cT, cN)	0,209	0,96	1,15
A · desenvolvimento	completo	0,194	0,99	0,84
B · desenvolvimento	simples (cT, cN)	0,212	0,97	0,89
B · desenvolvimento	completo	0,177	1,01	1,22
C · teste externo	simples (cT, cN)	0,246	1,24	0,32
C · teste externo	completo	0,239	1,33	0,58

AUC: capacidade de ordenar quem responde e quem não responde (0,5 = acaso). Observado/esperado e inclinação de calibração próximos de 1 indicam previsões bem calibradas. Nos dados fictícios, o hospital C foi gerado com taxa de resposta diferente, para mostrar como o teste externo revela problemas de calibração que o desenvolvimento não mostra.

Limites desta demonstração

- Dados fictícios demonstram o funcionamento do software, não eficácia clínica, distribuição real de pacientes ou segurança definitiva.
- Gradientes e hessianas podem vazar informação quando há poucos casos ou muitas rodadas; por isso há um mínimo de casos e um limite de rodadas, mas isso não é uma garantia formal de privacidade. Num piloto real, o modelo de ameaças e controles adicionais (agregação segura, privacidade diferencial) seriam avaliados com os hospitais.
- A supressão conjunta cobre faltantes e observados de cada variável; outras margens, como o número de casos completos de cada modelo, ainda podem ser combinadas entre consultas. Num piloto real, todas as saídas seriam registradas e checadas em conjunto antes da liberação.
- Os três "hospitais" rodam no mesmo computador, como processos separados. Num piloto real, cada nó ficaria na infraestrutura do hospital, com autenticação, criptografia e aprovação local de cada versão do código.

Arquivos: `protocolo/protocolo-piloto-esofago.json`, `no.py` (nó local), `coordenador.py`, `testes.py`, `saida/resultado.json` e as auditorias em `hospitais/*/auditoria_local.jsonl`.