

Comparação entre métodos estatísticos e de aprendizado de máquina para predição de risco em síndromes coronarianas agudas

Alleh Nogueira

Introdução

Revisão de Literatura

Objetivo

O objetivo primário deste estudo é comparar o desempenho de modelos, desenvolvidos através de métodos de regressão estatística e aprendizado de máquina, para prever eventos cardiovasculares adversos maiores. Os objetivos secundários deste estudo são: (1) desenvolver e comparar o desempenho entre modelos estatísticos e baseados em aprendizado de máquina para prever sangramento maior; (2) estimar o *trade-off* teórico entre os riscos cardiovascular e hemorrágico, para cada paciente; (3) avaliar teoricamente as consequências clínicas dos modelos preditivos desenvolvidos.

Métodos

Delineamento e pacientes do estudo

Estabelecido em um hospital soteropolitano terciário, nosso Registro de Síndromes Coronarianas Agudas é uma coorte prospectiva delineada para avaliar consecutivamente os desfechos de pacientes que apresentaram-se agudamente com angina ou equivalentes angiosos, independentemente dos achados eletrocardiográficos ou bioquímicos à admissão. Entre setembro de 2011 e julho de 2019, incluiu-se todos os pacientes com quadro clínico consistente com síndrome coronariana aguda e que preencheram pelo menos um dos seguintes critérios objetivos: (1) injúria miocárdica aguda, definida por alteração dinâmica dos níveis séricos de troponina acima do percentil 99; (2) alterações isquêmicas agudas ao eletrocardiograma, incluindo-se desnivelamentos do segmento ST típicos, inversões simétricas de onda T, e/ou bloqueio de ramo esquerdo presumivelmente novo; (3) presença de obstrução luminal coronariana de pelo menos 70% em um paciente com história de angina crônica, ou de ondas Q patológicas em um paciente com história sugestiva de infarto miocárdico prévio. Para incrementar a representatividade da coorte, não houve critérios de exclusão nem interferência dos pesquisadores para condução do estudo. Todos os participantes proveram, por escrito, seu consentimento livre e esclarecido para sua inclusão. Em concordância com legislações nacionais e internacionais, o protocolo do estudo fora previamente aprovado pelo comitê de ética em pesquisa local sob o certificado de apresentação de apreciação ética ...

Desfechos do estudo

Um modelo de regressão estatística foi comparado a dois modelos de aprendizado de máquina. Na análise primária, os modelos foram desenvolvidos para prever a ocorrência de um desfecho intra-hospitalar composto de morte por todas as causas e reinfarto miocárdico não fatal. Na análise secundária, os modelos foram desenvolvidos para prever a ocorrência de sangramento maior durante o internamento, definido como sangramento tipo 3 ou 5 conforme as definições do *Bleeding Academic Research Consortium* (BARC).

Seleção dos potenciais preditores

Os dados foram colhidos à admissão pelos cardiologistas assistentes e armazenados em uma base de dados virtual. Com base em conhecimentos prévios, 23 variáveis foram selecionadas como candidatos a potenciais preditores. Características demográficas e relativas à história clínica retratam-se em 15 variáveis: idade, sexo, índice de massa corporal, tabagismo atual, hipertensão arterial sistêmica, diabetes mellitus, dislipidemia, doença arterial periférica e/ou carotídea, doença arterial coronariana prévia, infarto agudo do miocárdio prévio, história familiar de doença arterial coronariana prematura (definida como acometimento de qualquer parente de 1º grau com idade inferior a 55 e 65 anos para o sexo masculino e feminino, respectivamente), acidente vascular encefálico prévio, doença renal crônica, e uso prévio de aspirina. As demais 8 variáveis referem-se ao exame físico e investigações diagnósticas complementares: frequência cardíaca, pressão arterial sistólica, classe de Killip de II a IV, alterações isquêmicas ao eletrocardiograma, injúria miocárdica, NT-pró-BNP sérico, creatinina sérica, e hemoglobina sérica. Para a regressão logística, o número máximo de potenciais preditores foi determinado pela regra geral de 10 eventos por parâmetro preditor, e incluiu-se como potenciais preditores as variáveis com menor valor de p significativo ($< 0,05$) na análise univariada de associação com o desfecho. Como não há limitações para o número máximo de variáveis predictoras em modelos de aprendizado de máquina, todas as 23 variáveis disponíveis foram utilizadas.

Derivação e desempenho dos modelos preditivos

A coorte foi aleatoriamente dividida, em uma razão de 4:1, em dois subgrupos: uma coorte de derivação, contendo 80% dos pacientes, foi utilizada para derivar os modelos de regressão logística, treinar os modelos de aprendizado de máquina e ajustar seus parâmetros; uma coorte de validação, contendo 20% dos pacientes, foi utilizada para testar o desempenho dos modelos desenvolvidos. Para cada desfecho, empregou-se um método estatístico de regressão logística e dois classificadores de aprendizado de máquina – *support vector machine* e *extreme gradient boosting* – para gerar três modelos preditivos. Mensurou-se o desempenho de cada modelo de três modos: (1) discriminação, avaliada pela estatística c de Harrell, numericamente igual à área sobre a curva *Receiver Operator Characteristic* (ROC); (2) calibração,

Trade-off entre os riscos cardiovascular e hemorrágico

Para cada desfecho – evento cardiovascular adverso maior ou sangramento maior – selecionou-se o modelo preditivo com melhor discriminação. Os pacientes foram divididos em decis conforme seus riscos estimados. Calculou-se o risco observado, para cada decil. Para comparar as classificações entre ambos os desfechos, combinou-se suas categorias de riscos preditos. O *trade-off* teórico entre os riscos cardiovascular e hemorrágico foi avaliado pela diferença absoluta dos riscos observados de eventos cardiovasculares e hemorrágicos, em cada decil de risco.

Análise estatística

Variáveis categóricas são descritas por frequências absolutas e relativas e variáveis numéricas por média e desvio padrão ou mediana e intervalo interquartil para distribuições normais e não normais, respectivamente. A normalidade da distribuição foi avaliada por uma combinação de inspeção visual, medidas de assimetria e curtose, e testes formais de normalidade. Utilizou-se o teste t para avaliar diferenças entre variáveis contínuas paramétricas, o teste U de Mann-Whitney para variáveis não paramétricas, o teste χ^2 para variáveis categóricas, e o teste exato de Fisher para tabelas 2x2. Utilizou-se a correção de Bonferroni para duas múltiplas comparações entre as discriminações do modelo estatístico e de cada modelo de aprendizado de máquina. Para tratar as variáveis incompletas, 40 conjuntos de dados multiplamente imputados mediante 100 iterações foram criados e analisados utilizando-se o pacote “mice” [4]. A significância estatística foi definida por um valor de p bicaudal inferior a 0,05. Todas as análises foram conduzidas no R, versão 4.1.0 (*R Foundation for Statistical Computing*, Viena, Austria).

Cálculo do tamanho amostral