**Delirium Prediction using Machine Learning Models on Preoperative Electronic Health Records Data**

**(Davoudi et al., 2017)**

Davoudi, A., Ebadi, A., Rashidi, P., Ozrazgat-Baslanti, T., Bihorac, A., & Bursian, A. C. (2017). Delirium Prediction using Machine Learning Models on Predictive Electronic Health Records Data. *Proceedings - IEEE 17th International Symposium on Bioinformatics and Bioengineering, BIBE 2017*, 568–573. https://doi.org/10.1109/BIBE.2017.00014

Os Registos de Saúde Electrónicos (EHR) são concebidos principalmente para registar informações relevantes sobre os pacientes durante a sua estadia no hospital para fins administrativos. Além disso, proporcionam uma fonte de dados eficiente e acessível para a investigação médica, tal como a previsão dos resultados dos pacientes. Neste estudo, utilizámos os Registos de Saúde Electrónicos pré-operatórios para prever o delírio pós-operatório. Comparámos o desempenho de sete modelos de aprendizagem de máquinas na previsão de delírios: modelos lineares, modelos de aditivos generalizados, florestas aleatórias, máquina vectorial de suporte, redes neuronais, e aumento do gradiente extremo. Entre os modelos avaliados neste estudo, as florestas aleatórias e o modelo aditivo generalizado superaram os outros modelos em termos da métrica de desempenho global para a previsão do delírio, particularmente no que diz respeito à sensibilidade. Verificámos que a idade, o abuso de álcool ou drogas, o estatuto socioeconómico, a questão médica subjacente, a gravidade do problema médico, e o cirurgião assistente podem afetar o risco de delírio.

**Introdução:**

Delirium é um neuropsiquiátrico transitório comum desordem exposta abruptamente com flutuações em consciência e estado mental (1). O delírio está ligado a múltiplos acontecimentos adversos, incluindo o aumento da morbilidade e mortalidade, estadias prolongadas na unidade de cuidados intensivos (UCI), e ventilação mecânica prolongada (2). A prevalência do delírio atinge os 73% entre pacientes de UCI cirúrgica, custando US entre $38 e $150 mil milhões por ano (3, 4). Estudos mostram que um terço de os casos de delírio podem beneficiar de uma prevenção multifactorial medidas e tratamentos (5-9). A elevada taxa de prevalência e o potencial para uma intervenção bem sucedida exige métodos de previsão precisos para identificar os pacientes com maior risco de desenvolver delírios. Estudos anteriores mostram que o delírio é atualmente subreconhecido (10), pelo que a previsão para o delírio pós-operatório está limitada a modelos de avaliação subjectiva do risco do médico que frequentemente dependem de uma elaborada extração de dados (11). Para remediar esta situação problema, os modelos de aprendizagem automática podem ser utilizados para prever risco de delírio. A patogénese do pós-operatório O delírio ainda não está completamente determinado. No entanto, o delírio, há muitos fatores que se demonstra contribuírem para maior risco de desenvolvimento de delírios. Estes fatores incluem, mas não estão limitados a: idade, tipo de admissão, procedimento cirúrgico primário. Estes fatores são registados em registos de saúde electrónicos e têm sido utilizados no literatura para previsão de delírios (6, 12, 13).

Neste estudo, utilizaremos o sistema electrónico pré-operatório Dados de registo de saúde (EHR) para previsão de delírios. EHR Os dados são recolhidos rotineiramente para todos os pacientes na admissão, e muitos estudos têm-nas utilizado para a previsão da saúde resultados durante e após a estadia no hospital (14). Anterior obras que utilizaram dados EHR para a previsão de delírios, têm usado principalmente modelos de regressão multivariada devido a a sua facilidade de interpretação e análise (12, 13). Usando análise de regressão logística, vários outros estudos têm identificou características importantes no conjunto de dados pré-operatórios contribuindo para o desenvolvimento do delírio, com até 87% área sob as características operacionais do receptor (ROC) curva (AUC) (12, 13, 15). Embora promissoras, elas características adicionais mais utilizadas, tais como o EstadoMini-Mental Pontuação do exame, acuidade visual e geriátrico Escala de Depressão, que não são avaliações de rotina em os hospitais. Outros têm sido aplicados a populações, tais como doentes idosos, doentes idosos com cirurgia da anca, ou pacientes na UCI (12, 13, 16).

Nos últimos anos, as técnicas de aprendizagem de máquinas têm tem sido cada vez mais utilizada na investigação médica para análise dados médicos complexos (17-21). Neste artigo, utilizámos os dados pré-operatórios recolhidos rotineiramente na admissão para desenvolvimento de modelos de aprendizagem de máquinas para prever o delírio.

Ao contrário do trabalho anterior, não dependemos de avaliações no nosso modelo. Assim, o nosso modelo pode ser prontamente utilizados na prática, utilizando apenas dados EHR. Os nossos resultados também mostram um melhor desempenho preditivo em comparação com os modelos de classificação da literatura. Utilizámos vários métodos incluindo downsampling e SyntheticMinority Técnica de sobreamostragem (SMOTE) (22) para contrariar a desequilíbrio no resultado e melhorar o desempenho dos modelos. O resto do documento é o seguinte: nós explicar os métodos utilizados na secção II e apresentar o resultados na secção III, enquanto a secção IV inclui o discussão e conclusões.

**B. Análise**

O fluxo de trabalho analítico preditivo do estudo está delineado na Figura 1. Realizámos várias etapas de pré-processamento do conjunto de dados para melhorar a eficiência computacional e a robustez dos modelos de previsão. O pré-processamento de dados incluiu limpeza de dados com remoção de outliers, imputação de dados em falta, e optimização de variáveis categóricas e nominais (28). Para fazer face ao risco de sobreajustamento, dividimos os dados de forma aleatória. Em cada execução, 80% dos dados foram utilizados para o desenvolvimento de modelos e 20% para testes. A prevalência do delírio foi semelhante em cada partição, utilizando o desenho de amostragem. Dividimos ainda os dados de desenvolvimento em 80% para formação do modelo, e 20% para validação para afinar os parâmetros. Repetimos o processo 50 vezes para relatar medidas de desempenho e intervalos de confiança. Em cada execução, os dados foram reordenados antes da partição.

Treinámos o algoritmo sobre os coortes de desenvolvimento, enquanto os resultados reportados foram obtidos dos coortes de teste. Utilizámos 20% do coorte como coorte de teste (n=10,291) em cada uma das 50 execuções de validação cruzada de 5 vezes repetidas (resultou em 250 coortes diferentes). O tamanho total da amostra permite uma largura máxima de 95% de intervalo de confiança para a área abaixo da curva de 0,06 para cada modelo quando a prevalência da complicação prevista é de 3%.

Comparámos sete abordagens de modelação preditiva: Naïve Bayes (NB), modelo aditivo generalizado (GAM), regressão logística (LR), máquina de suporte vectorial (SVM), florestas aleatórias (RF), reforço do gradiente extremo (XGB), e redes neurais (NN). Naive Bayes é geralmente usado como um modelo generativo simples, uma categoria de modelos preditivos que aprende a distribuição dos dados de entrada usando a regra de Bayes. Por outro lado, os modelos discriminatórios aprendem um mapa direto dos dados de entrada para as etiquetas de resposta, tais como regressão logística e modelo aditivo generalizado (29). A regressão logística é amplamente utilizada na literatura médica, e tem sido usada anteriormente para modelos de previsão para delírios (30, 31); com o risco previsto a aumentar ou a diminuir monotonicamente. Os modelos aditivos generalizados são modelos de regressão aditiva que podem relaxar a hipótese de monotonicidade dos modelos logísticos e oferecer a vantagem de estimar funções de risco não-lineares para variáveis contínuas.

Máquina vectorial de suporte, florestas aleatórias, impulsionamento de gradientes extremos, e redes neurais estão entre as técnicas de aprendizagem de máquinas amplamente utilizadas, mas nunca foram utilizadas para a previsão de delírios antes. A SVM realiza a classificação encontrando uma fronteira de decisão separadora no espaço do recurso de entrada. As florestas aleatórias constroem muitas árvores de decisão e tipicamente classificam os dados de acordo com o modo das árvores de decisão. Estas árvores de decisão são treinadas através da divisão do conjunto de dados em subconjuntos num valor num nó, repetindo este processo em cada subconjunto de forma recursiva. As florestas aleatórias melhoram o seu desempenho fazendo uma média sobre múltiplas árvores de decisão treinadas em diferentes partes do conjunto de dados, e reduzindo assim o risco de sobreajustamento. O aumento do gradiente extremo funciona como um conjunto de modelos de previsão mais fracos (árvores de decisão aqui) de uma forma iterativa. A cada iteração, é construído um novo modelo que acrescenta um estimador para proporcionar uma melhor aproximação do que a iteração anterior, e cada modelo aprende a corrigir o modelo da fase anterior.

As redes neurais são modelos de aprendizagem mecânica inspirados em redes de neurónios biológicos. Contêm camadas de nós informáticos simples que funcionam como algoritmo de soma não linear, interligados por linhas de ligação ponderadas, com pesos a serem ajustados com novas amostras de treino (32, 33). Utilizámos as probabilidades de eventos calculadas pelos modelos preditivos para classificar os pacientes em categorias de eventos e não eventos. Aplicámos um corte definido como o valor onde a estatística J do Youden é maximizada (34).

Utilizámos este ponto de corte do conjunto de dados de formação no teste conjunto de dados. Por causa do desequilíbrio entre os dois resultados classes, a precisão por si só não dá uma visão completa de o desempenho de discriminação dos modelos. Para o nosso estudo, comparámos os modelos utilizando precisão, AUC, sensibilidade como a proporção de verdadeiros positivos sobre o total de alvos positivos, e especificidade como a proporção de verdadeiros negativos sobre o total de alvos negativos.

Calculámos a importância da redução média do índice de impureza de Gini com base na sua importância. O índice de impureza de Gini pode ser calculado como na Equação (1). Cada vez que uma divisão é feita numa variável, os índices de Gini para os dois nós descendentes são menores do que o índice de Gini do nó pai. A diminuição do índice de Gini para cada variável é calculada pela soma das diminuições do índice de Gini para as variáveis sobre todas as árvores da floresta.

Discussão:

No nosso estudo, comparámos o desempenho dos modelos de previsão para o delírio utilizando dados de EHR. Estudámos sete modelos: regressão logística, modelos de aditivos generalizados, máquinas vectoriais de suporte, Bayes naïve, florestas aleatórias, aumento do gradiente extremo, e redes neurais. Entre estes modelos, florestas aleatórias e modelos de aditivos generalizados foram os modelos de previsão de delírios com melhor desempenho. Os dados que utilizámos para estes modelos estão disponíveis no ponto de acesso aos cuidados pré-operatórios e não requerem avaliações especializadas ou informação de auto-relato. Incluímos variáveis complexas, tais como códigos ZIP de residência e médico assistente. Os códigos ZIP podem atuar como um substituto das características socioeconómicas do bairro, que se tem demonstrado estar associado a múltiplos comportamentos de doença e saúde (36-38). O desempenho dos médicos assistentes também pode potencialmente ser um factor de resultados pós-operatórios (39, 40).

Para este estudo, aplicámos primeiro uma etapa de limpeza de dados (28) para reduzir os erros através da remoção dos valores aberrantes. A etapa de pré-processamento foi levada a cabo em consulta com clínicos que forneceram uma boa compreensão da natureza dos dados e das necessidades clínicas, e tornaram os algoritmos mais robustos e eficientes. O modelo aditivemodel generalizado e as florestas aleatórias foram os modelos preferidos neste estudo devido à sua precisão, eficiência relativa, e capacidade de é responsável pela não linearidade das variáveis. O modelo aditivo generalizado é um modelo orientado para os dados que tem a flexibilidade de captar a não-monotonicidade no resultado previsto. As florestas aleatórias, como uma abordagem baseada em árvores de decisão, é capaz de capturar a condicionalidade, as relações entre as características, e as relações não lineares entre as características e o resultado. Uma vez que a afinação dos parâmetros é crítica para as máquinas vectoriais de apoio, rede neural, florestas aleatórias, e aumento do gradiente extremo, uma afinação mais fina dos seus parâmetros pode potencialmente melhorar os resultados atuais.

Optimizámos os nossos parâmetros e modelo com base no índice máximo de Youden e não na precisão. Como resultado, o downsampling e SMOMTE não melhoraram significativamente o desempenho de florestas aleatórias e modelos de aditivos generalizados. Utilizámos a validação externa para reportar o verdadeiro desempenho dos modelos em dados não vistos. Utilizámos também modelos de florestas aleatórias para classificar as características utilizadas no modelo com base na sua redução média no índice de impurezas de Gini. Muitas das características escolhidas pelo modelo mostram as características socioeconómicas anteriormente ignoradas e a experiência do cirurgião para a previsão de delírios.

Conclusões

Estes modelos poderiam ser aplicados no ponto de acesso aos cuidados pré-operatórios. Não se baseiam em dados autodeclarados e testes especializados, e foram derivados de dados populacionais completos, recolhidos rotineiramente no período pré-operatório. Podem ajudar o pessoal de saúde na identificação de doentes com maior risco de tais complicações. Estes modelos de previsão também podem ajudar os pacientes a tomar decisões informadas sobre os seus procedimentos cirúrgicos e os riscos envolvidos. A validação prospectiva do modelo em diferentes populações pode melhorar o modelo para implementação em tempo real do fluxo de trabalho clínico para estratificação automatizada e simplificada do risco no período pré-operatório. O trabalho futuro inclui também a aplicação de outros métodos de trabalho com conjuntos de dados desequilibrados e a utilização de outros métodos de classificação. Pretendemos também estudar o desempenho dos modelos de previsão, acrescentando características intra-operatórias e pós-operatórias para captar qualquer complicação e a trajetória do paciente durante a sua estadia no hospital.