FATORES PREDITORES DA MORTALIDADE DE IDOSOS COM INFECÇÕES RELACIONADAS À ASSISTÊNCIA À SAÚDE, INTERNADOS EM UNIDADE DE TERAPIA INTENSIVA DE UM HOSPITAL UNIVERSITÁRIO NO SUL DO BRASIL, UTILIZANDO ALGORITMOS DE *MACHINE LEARNING*

Eduardo Falco Knaut (PIBIC/CNPq/FA/Uem), Luciano de Andrade (Orientador), e-mail: eduardofknaut@gmail.com.

Universidade Estadual de Maringá / Centro de Ciências da Saúde/Maringá, PR.

Área e subárea: medicina/clínica médica.

Palavras-chave: *machine learning*, infecções relacionadas à assistência à saúde (IRAS), idosos.

Resumo:

Este projeto busca traçar um panorama entre as infecções relacionadas à assistência à saúde em idosos em UTI com os fatores preditivos prognósticos. As doenças infecciosas são uma das principais causas de morte em ambiente hospitalar, principalmente na população mais velha. Com dados de um hospital universitário do sul do Brasil, coletou-se variáveis demográficas e clínicas para a previsão utilizando o *machine learning*. No geral, totalizou-se 250 infecções com 48 variáveis diferentes. Sete algoritmos foram criados para os cálculos e a análise feita com 6 métricas, além de outros métodos propostos. O modelo *Random Forest* obteve a melhor atuação, com acurácia de 0,824, kappa de 0,598, sensibilidade de 80%, especificidade de 83,3%, F1 de 0,667 e AUROC igual a 0,881. Outros modelos não trabalharam muito bem. Ao todo, com uma amostra limitada e estratégias matemáticas para avaliar a importância dos medidores clínicos, o tempo de estadia se destacou. Tal fato, o qual é amplamente disponível, não deve ser subestimado pelos profissionais da saúde.

Introdução

IRAS “é aquela adquirida após a admissão do paciente e que se manifeste durante a internação ou após a alta, quando puder ser relacionada com a internação ou procedimentos hospitalares”[1]. Caracterizada como uma das maiores ameaças à segurança dos enfermos, em ambientes voltados ao tratamento de doenças, elas representam uma grande tormenta aos cuidadores.

Além da ameaça à vida, as infecções nosocomiais são custosas para um país. Ao se tratar desse assunto, a prevenção é fundamental. Sendo os mais comuns organismos bactérias e fungos, o conhecimento dos meios de contaminação, a vigilância por órgãos especializados e a criação de

protocolos de controle da propagação são as principais medidas para garantir a segurança de cada unidade de saúde. O grande número de informações disponíveis no panorama das UTIs sobrecarrega e dificulta um julgamento preciso por médicos. A avaliação prematura a cada caso confirmado com infecção nosocomial em Unidade de Terapia Intensiva (UTI) pode ser um instrumento crucial para o cuidado individualizado.

Há alguns anos, o *machine learning* (ML) vem sendo usado na área da saúde com diversos objetivos. Neste estudo, nós visamos criar modelos preditivos a fim de averiguar a mortalidade de idosos em UTI com IRAS, com base em dados de um hospital universitário no sul do Brasil.

Materiais e métodos

1. Fonte de dados

A coleta de dados foi realizada em duas etapas: a primeira se baseou nos prontuários físicos arquivados do Serviço de Controle de Infecção Hospitalar (SCIH), a CCIH do HUM. A segunda fase foi realizada junto ao Sistema de Gestão da Assistência à Saúde do SUS (GSUS), focado apenas nos pacientes previamente selecionados.

2. Critérios de coleta e amostra

Somente pacientes idosos (60 anos ou mais) internados na UTI do Hospital Universitário de Maringá (HUM) entre 2014 a 2019 e diagnosticados com IRAS no setor foram considerados. Ao todo, 60 variáveis, foram contabilizadas. Dentre elas, avaliou-se dados demográficos, tempo entre movimentações intra-hospitalares, características da IRAS, tratamentos, testes laboratoriais e o desfecho.

3. Análise dos dados e técnica do ML

Onze variáveis não foram incluídas nos algoritmos de ML, apenas para quantificação e descrição da amostra. Elas incluem datas, antibióticos utilizados e microrganismos identificados. Antes do início do treinamento, um teste envolvendo toda a amostra com *Random Forest* (RF) foi feito para comparar a importância das variáveis.

Para a criação dos modelos, 7 algoritmos de classificação binária do desfecho, morte ou alta, com 48 preditores, foram implementados. O manejo dos hiperparâmetros foi realizado automaticamente por *repeated cross-validation*, com 10 *folds* e 10 *repeats*. O treino incluiu 70% das infecções. O restante foi utilizado no teste com avaliação por 6 métricas: acurácia, kappa, área embaixo da curva *Receiving Operating Characteristic* (AUROC), sensibilidade, especificidade e escore F1. Além disso, a soma das probabilidades de cada modelo foi dividida entre a razão do resultado

dicotômico, produzindo um número pelo qual o log de 2 foi retirado. Este valor serviu para criar uma comparação entre as previsões e o real.

Resultados e Discussão

Este estudo incluiu 180 pacientes diagnosticados com 250 infecções na UTI do HUM. Houve 104 óbitos (57,8%), 45 mulheres e 59 homens, e 76 altas (42,2%), 37 mulheres e 39 homens. Na divisão por grupos, 56 ficaram entre 60-69 anos, 70 entre 70-79 anos e 54 com 80 anos ou mais. Entre os óbitos, 25 estavam no primeiro grupo, 40 no segundo e 39 no terceiro. Dentre as infecções registradas, 160 corresponderam à primeira da internação, 49 já eram à segunda e 24 à terceira. O restante variou entre 4 até 7 diferentes infecções. A vasta maioria dos sítios (163) foram pneumonias. O segundo lugar foi das infecções relacionadas a acessos vasculares, totalizando mais de 40. Por fim, 35 pacientes tiveram IRAS por via urinária. Outros fatores predisponentes contabilizaram mínimos diferentes sítios. Dos 250 casos, 124 contém um agente etiológico identificado e 43 dois, com predominância dos gêneros *Acinetobacter*, *Candida*, *Enterococcus*, *Staphylococcus*, *Klebsiella*, *Pseudomonas* e *Escherichia*.

No modelo inicial com RF, os tempos entre as estadias nos setores obtiveram as maiores importâncias. A tentativa de previsão dos 34 casos restantes a partir do método de soma das probabilidades resultou em 21 corretos (62%), 6 incertos (17,5%) e 7 incorretos (20,5%).

A alta serventia do RF é exposta por sua melhor atuação na tabela 1. Com o maior número de acertos e com um valor kappa razoável, os resultados são promissores. Os outros modelos tiveram resultados menos relevantes, muitos dos quais a acurácia observada pouco superou a esperada, e alguns ainda obtiveram valores kappa negativos.

Tabela 1. Resultado final das métricas utilizadas para avaliação dos modelos.

Modelos	Acurácia	Kappa	Sensibilidade	Especificidade	Escore F1	AUROC
RF	0,824	0,598	0,800	0,833	0,667	0,881
GLMNET	0,647	0,227	0,500	0,727	0,500	0,727
KKNN	0,647	0,227	0,500	0,727	0,500	0,614
PDA	0,706	0,422	0,556	0,875	0,667	0,697
PAM	0,647	NA	NA	0,647	NA	0,551
C5.0TREE	0,500	-0,116	0,273	0,609	0,261	0,502
PLS	0,647	NA	NA	0,647	NA	0,553

NA: não é possível o cálculo; GLMNET: *elastic-net regularized generalized linear*; KKNN: *K-nearest neighbors; PDA: penalized discriminant analysis;* PAM: *nearest shrunken centroids;* C5.0TREE: *single C5.0 tree; PLS: partial least squares.*

Conclusões

Apesar de um estudo monocêntrico e de pequena amostra, fazendo uso da ferramenta do ML, muitas técnicas foram propostas para contornar as adversidades e associações de grande interesse foram elucidadas. Embora de baixa validade externa prospectiva, pelo uso de dados em diferentes estágios temporais, os quais não estariam disponíveis em um mesmo período, a proposta pode destacar um dos grandes pontos esquecidos pelos profissionais, o tempo de permanência no hospital. Os modelos tinham acesso exatamente a quando cada paciente iria morrer ou ser dispensado da UTI. Tal valor se provou muito útil para avaliar a mortalidade.

Agradecimentos

Ao CNPq, à Universidade Estadual de Maringá e ao meu orientador, por todo apoio e estímulo.

Referências

AGÊNCIA NACIONAL DE VIGILÂNCIA SANITÁRIA (ANVISA). **Portaria n° 2.616, de 12 de Maio de 1998**. Disponível em: <http://bvsms.saude.gov.br/bvs/saudelegis/gm/1998/prt2616_12_05_1998.html>. Acesso em 30/03/2020.

CENTERS FOR DISEASE CONTROL AND PREVENTION (CDC). **CDC/NHSN surveillance definitions for specific types of infections**. Publicado em 2021. Disponível em <https://www.cdc.gov/nhsn/pdfs/pscmanual/17pscnosinfdef_current.pdf>. Acesso em 15/06/2021.

SOUSA, A. F. L, et al. **Deaths among the elderly with ICU infections**. Rev. Bras. Enferm., Brasília, v. 70, n. 4, p. 733-739, Agosto 2017. Disponível em: <http://www.scielo.br/scielo.php?script=sci_arttext&pid=S0034-71672017000400733&lng=en&nrm=iso>. Acesso em 13/04/2020.

LIU, X. et al. **Interpretable machine learning model for early prediction of mortality in elderly patients with multiple organ dysfunction syndrome (MODS): a multicenter retrospective study and cross validation**. Disponível em: <https://arxiv.org/abs/2001.10977>. Acesso em 22/05/2020.