André Luís Franco Cotia

APLICAÇÃO DE MINERAÇÃO DE DADOS PARA ANÁLISE DE RISCO DE INFECÇÃO HOSPITALAR

Tese apresentada à Faculdade Israelita de Ciências da Saúde Albert Einstein para obtenção do título de Doutor em Ciências da Saúde.

São Paulo
2026

André Luís Franco Cotia

APLICAÇÃO DE MINERAÇÃO DE DADOS PARA ANÁLISE DE RISCO DE INFECÇÃO HOSPITALAR

Tese apresentada à Faculdade Israelita de Ciências da Saúde Albert Einstein para obtenção do título de Doutor em Ciências da Saúde.

Orientador: Prof. Dr. Alexandre Rodrigues Marra

São Paulo
2026

Cotia, André Luís Franco
Aplicação de mineração de dados para análise de risco de infecção hospitalar / André Luís Franco Cotia. -- São Paulo, 2026.
x, 36 f.

Tese (Doutorado) – Faculdade Israelita de Ciências da Saúde Albert Einstein. Instituto Israelita de Ensino e Pesquisa Albert Einstein. Programa de Pós-Graduação em Ciências da Saúde.

Título em inglês: *Application of data mining for healthcare-associated infection risk analysis.*

1. Higiene das mãos. 2. Registros eletrônicos de saúde. 3. Infecção hospitalar. 4. Aprendizado de máquina.

Elaborada pelo Sistema Einstein Integrado de Bibliotecas

FACULDADE ISRAELITA DE CIÊNCIAS DA SAÚDE ALBERT EINSTEIN

Coordenador do curso de pós-graduação: Prof. Dr. Luciano Cesar Pontes de Azevedo

André Luís Franco Cotia

APLICAÇÃO DE MINERAÇÃO DE DADOS PARA ANÁLISE DE RISCO DE INFECÇÃO HOSPITALAR

Presidente da banca: Prof. Dr. Alexandre Rodrigues Marra

BANCA EXAMINADORA

Membros titulares:
Prof. Dr. Marcelino de Souza Durão Junior
Dr. Thiago Zinsly Sampaio Camargo
Prof. Dr. João Paulo Marochi Telles

Membros suplentes:
Prof. Dr. Oscar Fernando Pavão dos Santos
Dra. Silvana Maria de Almeida
Prof. Dr. Lauro Vieira Perdigão Neto

Aprovada em: 09/04/2026.

Dedicatória

À minha esposa, Amanda. Por me incentivar desde o início, e por permanecer ao meu lado durante todo o percurso. Pela resiliência diante dos sacrifícios que vieram com esse compromisso, especialmente após o nascimento de nossa filha, Ada. Mesmo diante das intensas demandas na criação de nossa filha você ofereceu seu apoio inabalável. Sem você, não teria sido possível conciliar os desafios dos primeiros meses de vida da Ada com os compromissos do doutorado. Sou eternamente grato.

Aos meus pais, Hélio e Adriana, meu mais sincero reconhecimento. Desde muito cedo, vocês me ensinaram o valor do estudo, do esforço e da dedicação. Priorizando a minha formação, garantindo que tivesse acesso às melhores oportunidades. Foi graças a vocês que o estudo se tornou um elemento essencial em minha trajetória, e por isso serei sempre grato.

Agradecimentos

Agradeço, em primeiro lugar, ao meu orientador, Dr. Alexandre Rodrigues Marra, por ter me guiado de maneira excepcional ao longo desse caminho. Sua presença constante em todas as etapas do processo, sua atenção minuciosa aos detalhes e sua preocupação genuína foram essenciais. Sua orientação e a generosidade em compartilhar experiências de vida e de profissão enriqueceram imensamente meu crescimento acadêmico e pessoal.

Aos amigos e colegas de profissão que me acompanharam e me inspiraram ao longo desse caminho: Lucas Z. Pocebon, Victor R. Procaci, Aline C. Queiroz de Leão, Natanael S. Adiwardana, Ruan de A. Fernandes, Eduardo T. Pandini e João Paulo M. Telles. Suas palavras de incentivo, seu compartilhamento de conhecimento e sua camaradagem foram um grande suporte.

Meus sinceros agradecimentos, ao Sérgio B. Wey e ao Thiago Zinsly S. Camargo, por abrirem portas, mostrarem o caminho e pelos conselhos valiosos que tanto influenciaram minha jornada profissional.

Agradeço à equipe de Tecnologia da Informação do Hospital Israelita Albert Einstein, representada pelo José Edgar Vieira de Barros, pelo apoio indispensável na coleta e organização dos dados.

Aos estatísticos Anderson Paulo Scorsato e Elivane da Silva Victor, do Serviço de Estatística do Einstein Hospital Israelita, pela contribuição fundamental na análise dos dados, que foi essencial para o desenvolvimento deste trabalho.

Aos demais coautores do artigo científico, cuja colaboração foi de suma importância na síntese dos resultados e na revisão criteriosa para a publicação.

Sumário

Dedicatória v
Agradecimentos vi
Lista de figuras viii
Lista de abreviaturas ix
Resumo x
1 INTRODUÇÃO 1
1.1 Objetivos 5
2 MÉTODOS 6
2.1 Aspectos éticos 9
2.2 Análise estatística 9
3 RESULTADOS 12
4 DISCUSSÃO 17
5 CONCLUSÃO 23
6 APRESENTAÇÃO DO ARTIGO 24
7 MÉTRICAS 25
7.1 Métricas do periódico 25
7.1.2 Fator de impacto 25
7.2 Métricas do artigo 26
7.2.1 Citações 26
7.2.2 Altimetria 26
7.3 Métricas do autor 27
7.3.1 Redes científicas 27
7.3.2 Trabalhos publicados 27
7.3.3 Mediana do fator de impacto dos trabalhos 28
7.3.4 Citações recebidas 29
7.3.5 H-index e m-quociente 29
8 REFERÊNCIAS 32
Abstract

Lista de figuras

Figura 1. Detecção do crachá do profissional da saúde pelo sensor dentro do dispensador de álcool e pelo sensor junto ao leito 7
Figura 2. Coleta e armazenamento de dados 8
Figura 3. Resumo dos algoritmos de *machine learning* e diversos conjuntos de dados realizados 10
Figura 4. Resumo das métricas de qualidade do modelo 15
Figura 5. Razões de chances das variáveis preditoras derivadas de modelo linear generalizado com validação cruzada 15
Figura 6. Importância das variáveis dos modelos *random forest* sem validação cruzada em 5 partes 16
Figura 7. Importância das variáveis dos modelos XGBoost 16

Lista de abreviaturas

AUC	Área sob a curva (*Area under the curve*)
CDC	*Centers for Disease Control and Prevention*
CID	Classificação Internacional de Doenças
GAMLSS	*Generalized additive models for location, scale and shape*
GLM	*Generalized linear model*
HICPAC	*Healthcare Infection Control Practices Advisory Committee*
IRAS	Infecções relacionadas à assistência à saúde
NHSN	*National Healthcare Safety Network*
ROSE	*Random Over Sampling Examples*
TI	Tecnologia da informação
UTI	Unidade de terapia intensiva
XGBoost	*Extreme gradient boosting*

Resumo

Introdução: As infecções relacionadas à assistência à saúde aumentam a morbidade, a mortalidade e os custos com saúde. A higiene das mãos eficaz é crucial para a prevenção, mas alcançar a alta adesão continua sendo um desafio. **Objetivo:** Explorar o uso de *machine learning* para integrar um sistema eletrônico de auditoria de higiene das mãos com registros eletrônicos de saúde para prever infecções relacionadas à assistência à saúde. **Métodos:** Um estudo de coorte retrospectivo foi realizado em um hospital brasileiro durante 2017-2020. A adesão à higiene das mãos foi registrada eletronicamente e os dados dos pacientes foram coletados dos prontuários eletrônicos. Os desfechos primários foram as infecções relacionadas à assistência à saúde conforme as definições de vigilância do *Centers for Disease Control and Prevention/National Healthcare Safety Network*. Os algoritmos de *machine learning*, balanceados com *Random Over Sampling Examples*, foram utilizados para modelagem preditiva, incluindo *generalized linear model*; *generalized additive models for location, scale, and shape*; *random forest*; *support vector machine*; e *extreme gradient boosting*. **Resultados:** Cento e vinte e cinco dos 6.253 pacientes (2%) desenvolveram infecções relacionadas à assistência à saúde e 920.489 oportunidades de higiene das mãos (49,3% de adesão) foram analisadas. Observou-se uma correlação direta entre a adesão à higiene das mãos e as infecções relacionadas à assistência à saúde. O algoritmo *generalized linear model* com balanceamento pelo *Random Over Sampling Examples* demonstrou desempenho superior, com 84,2% de sensibilidade, 82,9% de especificidade e uma área sob a curva de 93%. **Conclusão:** Integrar sistemas eletrônicos de auditoria de higiene das mãos com registros eletrônicos de saúde e utilizar modelos de *machine learning* pode aprimorar a vigilância do controle de infecções e prever os desfechos dos pacientes. Mais pesquisas são necessárias para validar esses achados e integrá-los à prática clínica.

Descritores: Higiene das mãos; Registros eletrônicos de saúde; Infecção hospitalar; Aprendizado de máquina

1 INTRODUÇÃO

As infecções relacionadas à assistência à saúde (IRAS) têm um impacto negativo significativo na população, levando ao aumento dos custos de saúde e maior morbidade.[1] Nos Estados Unidos, por exemplo, com dados ajustados para 2007, o custo anual das IRAS pode chegar a US$ 45 bilhões.[2] O *Centers for Disease Control and Prevention* (CDC) estima que 98.000 mortes ocorrem devido a infecções adquiridas em hospitais por ano.[3]

A identificação dos fatores de risco para IRAS é essencial para a implementação de medidas preventivas eficazes. Existem variáveis bem estabelecidas que aumentam o risco de se desenvolver IRAS, fatores como idade, comorbidades, tempo de internação, procedimentos invasivos, terapias imunossupressoras, e tempo de uso de dispositivos invasivos, entre outros. A adoção de pacotes de medidas de prevenção, higienização adequada das mãos, treinamento das equipes em protocolos de segurança, a manipulação adequada de dispositivos, as técnicas de desinfecção e assepsia, além da remoção oportuna de dispositivos, interferem diretamente na diminuição do risco de IRAS.[4]

A utilização de variáveis individuais dos pacientes em escores de risco para diversas doenças e infecções tem sido empregada há décadas na medicina. Nos últimos anos, o desenvolvimento de novos modelos preditivos e algoritmos de *machine learning* tem se tornado cada vez mais frequente, visando aprimorar a acurácia desses escores e auxiliar na intervenção precoce em pacientes de maior risco. Por exemplo, um estudo de coorte retrospectivo demonstrou uma redução significativa na deterioração precoce do estado clínico e na necessidade de cuidados mais intensivos associada a uma intervenção baseada em um modelo de deterioração de *machine learning* em pacientes hospitalizados, evidenciando a eficácia dessa abordagem.[5] Esses avanços contribuem para a otimização do tempo, aumentando a eficácia e o impacto das ações preventivas, além de reduzir o tempo para o início do tratamento precoce e, consequentemente, os custos associados à assistência em saúde.

A adequada higiene das mãos é considerada uma ação crucial para prevenir a disseminação de IRAS, incluindo aquelas causadas por organismos multirresistentes e outros agentes bacterianos e virais entre pacientes hospitalizados e profissionais de saúde.[6]

O reconhecimento dos benefícios da higiene das mãos remonta ao início do século XIX, quando um farmacêutico francês demonstrou a eficácia de soluções à base de cloretos como desinfetantes, recomendando seu uso para prevenir a transmissão de doenças contagiosas. Na década de 1840, Ignaz Semmelweis evidenciou que a limpeza das mãos com soluções de cloro entre atendimentos reduzia significativamente a mortalidade materna em uma clínica de Viena, atribuindo a febre puerperal à transmissão de partículas cadavéricas pelas mãos dos profissionais de saúde. Nos Estados Unidos, as primeiras diretrizes formais sobre a higiene das mãos em ambientes hospitalares foram publicadas pelo CDC em 1975, com recomendações que evoluíram nas décadas seguintes. A década de 1980 foi um marco na evolução dos conceitos de higiene das mãos, com a publicação das primeiras diretrizes nacionais em vários países, seguidas por outras em anos subsequentes. Em 1995 e 1996, o CDC e o *Healthcare Infection Control Practices Advisory Committee* (HICPAC) recomendaram o uso de sabão antimicrobiano ou agentes antissépticos sem água para a higienização das mãos ao sair dos quartos de pacientes com patógenos multirresistentes. Apesar dessas iniciativas, a adesão dos profissionais de saúde às práticas recomendadas ainda representa um desafio significativo.[7,8]

Os motivos mais comuns apontados pelos profissionais de saúde para a não adesão à higiene das mãos incluem falta de tempo, sobrecarga de trabalho, conhecimento inadequado, ceticismo em relação à eficácia da higiene das mãos como método de prevenção, localização inconveniente de pias e dispensadores de sabão, além da ausência de incentivos para a adesão à prática. Estudos demonstraram que a conformidade com a higiene das mãos pode ser influenciada pela acessibilidade aos produtos e pelas formulações desses produtos à base de álcool (líquidos, géis e espumas). Além disso, sabemos que o uso de antissépticos à base de álcool exige menos tempo do que a lavagem com sabão e água, sendo igualmente eficaz para a maioria dos patógenos.[7,9]

Um marco importante nas campanhas globais para promover a higiene das mãos no cenário da assistência à saúde foi o desenvolvimento do conceito dos "Cinco Momentos para a Higiene das Mãos". Esse conceito define cinco situações-chave em que os profissionais de saúde devem higienizar as mãos para prevenir infecções: antes de tocar o paciente, antes de realizar procedimentos assépticos, após risco de exposição a fluidos corporais, após tocar o paciente e após contato com superfícies próximas a ele. Criado no contexto da Campanha Nacional de Higiene das

Mãos na Suíça e integrado à Estratégia Multimodal de Melhoria da Higiene das Mãos da OMS, o conceito buscava facilitar a adesão às práticas de higiene ao vincular ações específicas a resultados concretos. Entre 2006 e 2008, foi amplamente adotado em mais de 400 hospitais, consolidando-se como uma ferramenta essencial na prevenção de IRAS.[8]

A questão da baixa adesão à higiene das mãos é ainda mais complicada pelos desafios enfrentados na medição da conformidade. As auditorias convencionais de higiene das mãos, realizadas por um observador humano, demandam muito tempo de profissionais treinados e têm capacidade limitada, cobrindo apenas uma pequena porcentagem do total de oportunidades. Os observadores podem ser profissionais designados exclusivamente para essa função e que circulam pelo hospital, profissionais de saúde inseridos nas suas próprias unidades ou trabalhadores de outras unidades. Também foi demonstrado que as taxas de adesão observadas diretamente podem não ser precisas, pois não se correlacionam com outras métricas de higiene das mãos, como contadores eletrônicos de lavagem das mãos ou volumes de produto utilizados. Tipicamente, as auditorias ocorrem principalmente antes da entrada e após a saída dos quartos dos pacientes.[10]

Os sistemas eletrônicos de auditoria de higiene das mãos vêm mudando gradualmente esse cenário, com um aumento considerável no número de oportunidades auditadas, possibilitando auditoria em tempo real, 24 horas por dia em todos os pacientes internados. Os métodos de monitoramento de higiene das mãos podem variar significativamente. Alguns utilizam sensores para detectar vapores de álcool nas mãos dos profissionais de saúde ou identificadores por radiofrequência para registrar quando a higienização foi realizada. Outros sistemas incluem o uso de câmeras direcionadas para dispensadores de álcool em gel para monitorar a adesão, enquanto algumas tecnologias empregam transmissores de ultrassom para medir o cumprimento das práticas. Independentemente do método utilizado, esses sistemas têm a capacidade de transmitir os dados de adesão de forma *wireless* para uma central de monitoramento, permitindo exibir as taxas individuais de conformidade dos profissionais em tempo real no local ou para os líderes das unidades.[11–14] O armazenamento desses dados e a integração com o prontuário eletrônico pode possibilitar a inclusão da taxa de adesão de higiene das mãos como uma variável de risco nos escores de risco para IRAS. O impacto no aumento das taxas de adesão à higiene das mãos pelos sistemas eletrônicos de auditoria ainda é variável, dependendo do sistema utilizado e de outros fatores, como o

treinamento dos profissionais de saúde, os processos e a estrutura de cada unidade hospitalar.[10,15–17]

O aumento das taxas de adesão à higiene das mãos pode ser alcançado principalmente por meio de sistemas eletrônicos de auditoria com tecnologia de feedback em tempo real, utilizando luzes ou sinais sonoros.[18] Vale notar que esses sistemas eletrônicos de auditoria de higiene das mãos frequentemente não são integrados com prontuários eletrônicos, o que pode representar desafios para uma análise de dados abrangente e para a coordenação dos esforços de controle de infecções. O uso desses sistemas eletrônicos ainda não é amplamente adotado, em parte devido ao custo e à necessidade de instalação da infraestrutura adequada. Mesmo as instituições que têm capacidade para implementar sistemas eletrônicos de higiene das mãos precisam trabalhar de perto com o departamento de engenharia para verificar se pode haver interferência com outros equipamentos.[10,16,17]

O impacto das taxas de higiene das mãos na incidência de IRAS ainda é uma inferência indireta, sendo comum a correlação entre a redução das IRAS e o aumento das taxas de higiene das mãos em uma área específica. No entanto, relações diretas nem sempre são observadas. Por exemplo, pode ocorrer um aumento no número de casos de IRAS, mesmo com o aumento das taxas de higiene das mãos no mesmo período. Em tais casos, é necessário considerar outros fatores de risco nessas análises.[19]

Para auxiliar na análise do impacto direto da higiene das mãos sobre a incidência de IRAS, programas e algoritmos sofisticados de *machine learning* podem ser empregados. Estes utilizam uma variedade de técnicas estatísticas e computacionais, possibilitando o desenvolvimento de análises de conjuntos de dados complexos, mesmo em larga escala. Diferente dos métodos convencionais de análise de fatores de risco, que muitas vezes se baseiam em algoritmos simples ou árvores de decisão, os modelos de *machine learning* podem otimizar iterativamente e automaticamente modelos matemáticos aos dados disponíveis.[20]

Nesse contexto, abordagens baseadas em *data mining* e *machine learning* têm ganhado destaque na área da saúde. *Data mining* refere-se ao processo de extração sistemática de padrões, associações e tendências relevantes a partir de grandes bases de dados, utilizando métodos estatísticos, matemáticos e computacionais. Já o *machine learning*, um subcampo da inteligência artificial, compreende um conjunto de algoritmos capazes de aprender automaticamente a partir

dos dados, identificando relações complexas e não lineares entre variáveis, sem a necessidade de regras explicitamente programadas. Diferentemente dos métodos estatísticos tradicionais, essas técnicas permitem lidar com grandes volumes de dados heterogêneos, integrar múltiplas fontes de informação e atualizar continuamente os modelos à medida que novos dados são incorporados, tornando-se particularmente adequadas para a análise de sistemas complexos, como aqueles relacionados ao risco de infecções hospitalares e à adesão a medidas de prevenção.[21–23]

O conceito de *machine learning* surgiu a partir da convergência entre estatística, ciência da computação e teoria da informação, com o objetivo de desenvolver métodos capazes de reconhecer padrões em dados e generalizar esse conhecimento para novas observações. Desde as primeiras formulações, esses métodos têm sido utilizados para identificar regularidades, dependências e estruturas latentes em conjuntos de dados complexos. Os algoritmos de *machine learning* operam por meio do ajuste iterativo de modelos matemáticos aos dados disponíveis, buscando minimizar erros de predição ou maximizar medidas de desempenho previamente definidas. Esses algoritmos podem ser organizados em diferentes classes, como aprendizado supervisionado, não supervisionado e semi-supervisionado, a depender da disponibilidade de desfechos conhecidos. Em aplicações clínicas, os modelos supervisionados são amplamente empregados para estimar riscos e prever desfechos, ao passo que técnicas não supervisionadas são utilizadas para explorar padrões ocultos, agrupar indivíduos com características semelhantes e identificar comportamentos atípicos. Essa capacidade de aprender padrões complexos a partir de dados observacionais torna o *machine learning* particularmente relevante para a análise de processos multifatoriais e dinâmicos na área da saúde.[22,24,25]

1.1 Objetivos

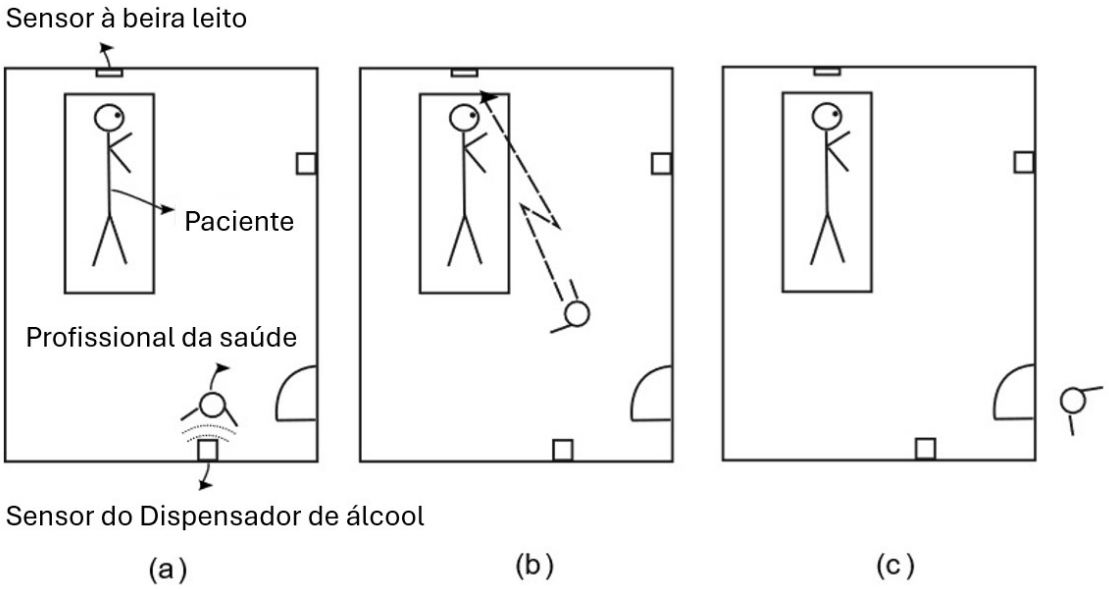
1. Empregar técnicas de *machine learning* para analisar a integração dos dados de um sistema eletrônico de auditoria de higiene das mãos com variáveis de riscos para IRAS coletadas dos prontuários eletrônicos dos pacientes;
2. Avaliar o impacto da variável taxa de adesão à higiene das mãos na predição de IRAS através de algoritmos de machine learning.

2 MÉTODOS

Foi realizado um estudo de coorte retrospectivo, consolidando dados dos prontuários eletrônicos de pacientes que foram hospitalizados em 80 leitos de semi-intensiva e 4 leitos de unidades de terapia intensiva (UTI), com dados de higiene das mãos capturados e armazenados eletronicamente para cada leito, de janeiro de 2017 a dezembro de 2020, no Hospital Israelita Albert Einstein, um hospital privado terciário de alta complexidade, localizado em São Paulo, Brasil, que possui um total de 745 leitos.

Os episódios de higiene das mãos foram registrados por dispensadores eletrônicos de álcool gel (PURELL Hand Instant Sanitizer – 62% álcool etílico + 4% álcool isopropílico, bolsa de 1 litro). O dispensador de álcool gel (modelo NXT de 1 litro) registra apenas um episódio em cada período de 2 segundos, mesmo que várias doses de álcool sejam dispensadas. Dispensadores de clorexidina (2% clorexidina) também estavam disponíveis para uso, mas esses dispensadores não tinham contadores eletrônicos. Ambos os dispensadores liberavam o mesmo volume de produto por uso (aproximadamente 1,3mL) e estavam localizados dentro dos quartos dos pacientes. Todos os dispensadores eram do mesmo tipo (1 litro). O bico dos dispensadores não ficava obstruído porque a cada 48 horas um enfermeiro (não plantonista) verificava o número dentro dos contadores de sinalização e verificava o bico ou trocava o dispensador para trocar os dispensadores vazios.

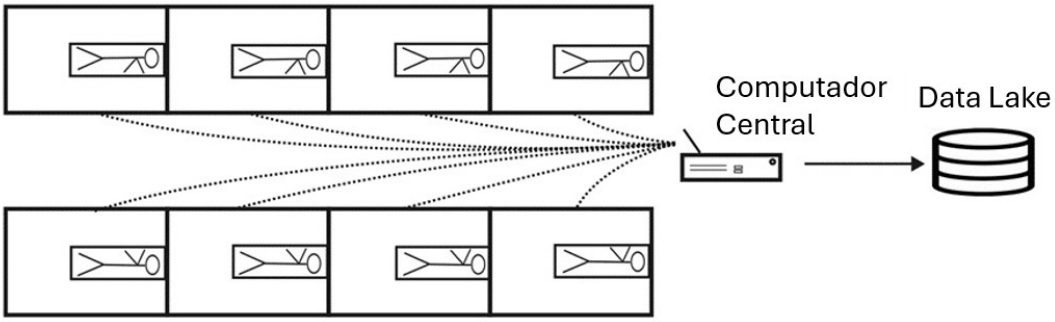
Os profissionais de saúde utilizavam crachás com microchip de radiofrequência, identificados por categoria profissional. O sensor dentro do dispensador de álcool detectava o evento de higiene das mãos e o crachá de identificação do profissional de saúde enviava essa informação ao sensor da cama. Quando o profissional de saúde se aproximava da cama do paciente, o sensor da cama classificava esse profissional como cumprindo as regras de higiene das mãos. Por outro lado, se o profissional de saúde não realizava a higiene das mãos, ele era classificado como não cumpridor das regras de higiene das mãos.[26,27] (Figura 1).



(a) Detecção do crachá do profissional da saúde pelo sensor dentro do dispensador de álcool, após acionamento do dispensador; (b) Detecção do crachá do profissional da saúde pelo sensor junto ao leito, após aproximação do leito; (c) Saída do profissional do quarto do paciente.
Figura 1. Detecção do crachá do profissional da saúde pelo sensor dentro do dispensador de álcool e pelo sensor junto ao leito

Vários profissionais de saúde podem estar no quarto ao mesmo tempo e, para cada um, o sensor da cama os classificará como cumpridores ou não cumpridores das regras de higiene das mãos. Quando o profissional da saúde sai do quarto, o sensor da cama continua monitorando o crachá até que o intervalo de tempo sem detectar o crachá atinja três minutos. Se o sensor detectar o crachá durante esse intervalo (por exemplo, se o profissional de saúde retornar ao quarto), o intervalo de tempo é redefinido.

O cálculo da adesão à higiene das mãos foi realizado considerando apenas os profissionais de saúde que usavam o crachá com microchip detector por radiofrequência, identificados por categoria profissional, pois não é possível monitorar a entrada e saída de profissionais que não utilizavam o crachá. A adesão à higiene das mãos foi calculada dividindo a soma dos eventos de higiene das mãos pela soma de todos os eventos (eventos cumpridos e não cumpridos). Um computador central na enfermaria coletava todos os dados das camas utilizando uma Rede Zigbee (Figura 2).



O computador central da unidade coleta todos os dados junto ao leito utilizando uma Rede Zigbee.

Figura 2. Coleta e armazenamento de dados

Esses dados eram armazenados localmente em um banco de dados MySQL (https://www.mysql.com/) onde a adesão à higiene das mãos era calculada por um programa em Python (https://www.python.org/) usando os dados brutos enviados pelos dispositivos nas camas. Todos os cálculos foram processados em tempo real. Um banco de dados com a adesão à higiene das mãos calculada foi armazenado em um *data lake* acessível por outros programas, por exemplo, o software de prontuário eletrônico do hospital.

Variáveis individuais de interesse foram coletadas dos prontuários eletrônicos dos pacientes para análise dos fatores de risco de IRAS. As informações referentes às comorbidades foram extraídas a partir dos códigos da Classificação Internacional de Doenças (CID) registrados nos prontuários eletrônicos dos pacientes. Foram considerados os diagnósticos documentados durante a hospitalização ou previamente registrados no prontuário eletrônico, conforme disponibilidade dos dados, para compor as variáveis clínicas utilizadas nas análises de risco. Essas variáveis foram disponibilizadas em uma plataforma de pesquisa baseada em um *data lake* em nuvem, com acesso exclusivo concedido ao profissional de tecnologia da informação (TI) envolvido no projeto e à equipe estatística para condução das análises do projeto. Ambos receberam senhas individuais para acessar os dados.

As variáveis coletadas incluíram: tempo de permanência na unidade de semi-intensiva ou na UTI, sexo, idade, comorbidades (*diabetes mellitus*, hipertensão arterial sistêmica, insuficiência renal, neoplasias, doenças hepáticas, doenças pulmonares, anemia, infecção pelo HIV, neutropenia, transplante de órgão sólido, transplante de medula óssea, uso de drogas injetáveis), dispositivos invasivos (cateter venoso central, cateter vesical de demora, ventilação mecânica), ventilação não invasiva

e procedimentos e tratamentos realizados (uso de corticosteroides, uso de drogas vasoativas, quimioterapia, transfusão de componentes sanguíneos, cirurgia, hemodiálise, nutrição parenteral exclusiva). Os desfechos primários considerados foram IRAS, seguindo as definições baseadas nos critérios estabelecidos pela definição de vigilância do CDC/*National Healthcare Safety Network* (NHSN) de infecções associadas aos cuidados de saúde e critérios para tipos específicos de infecções no ambiente de cuidados agudos.[28] Consideramos os dados anteriores ao diagnóstico de uma infecção para avaliar se essas variáveis realmente contribuem para o risco ou eram apenas desfechos da infecção.

2.1 Aspectos éticos

Este estudo foi aprovado pelo Comitê de Ética do Hospital Israelita Albert Einstein, protocolo número 5147-22, CAAE 57509922.4.0000.0071. Os dados coletados foram processados pela equipe de TI, anonimizados e os números dos prontuários dos pacientes foram desidentificados para preservar a confidencialidade, seguindo a Lei Geral de Proteção de Dados.

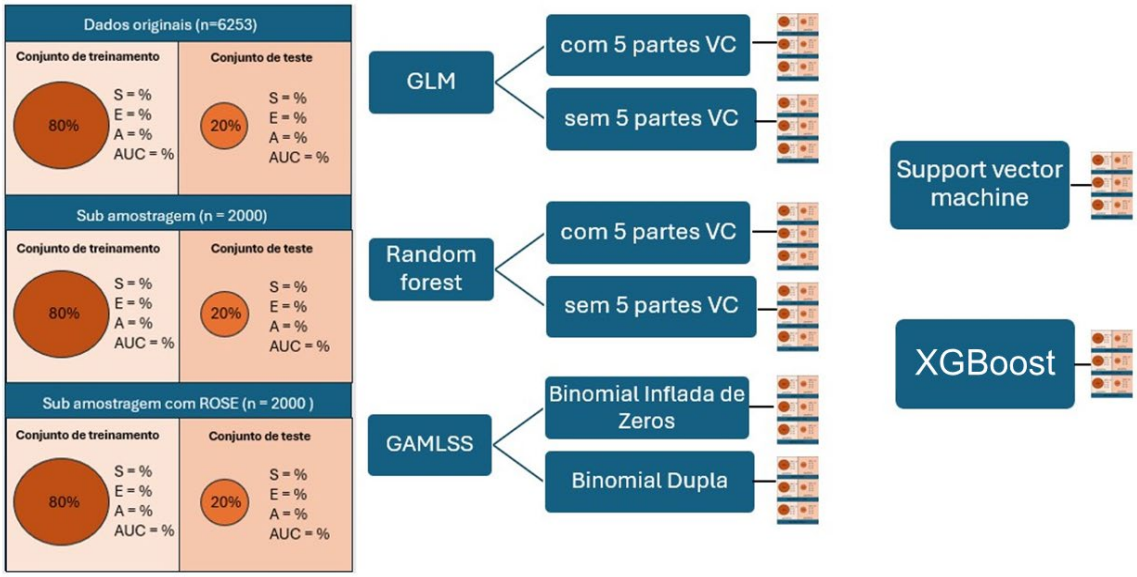
2.2 Análise estatística

O teste de Anderson-Darling foi utilizado para verificar a normalidade das variáveis quantitativas. O teste de Mann-Whitney comparou a variável qualitativa de IRAS com os preditores quantitativos. As associações entre variáveis qualitativas foram avaliadas por meio dos testes qui-quadrado ou exato de Fisher, conforme apropriado.

Adicionalmente, foi realizada análise de regressão logística para estimar a associação entre os preditores e o desfecho IRAS, sendo conduzidas análises univariadas e multivariadas. Na análise univariada, cada variável foi avaliada individualmente, com estimativa das razões de chances (odds ratios, OR) e respectivos intervalos de confiança de 95%. Para a construção do modelo multivariado, foram consideradas as variáveis com relevância clínica e/ou aquelas que apresentaram associação na análise bivariada ($p < 0,20$), permitindo o ajuste para potenciais fatores

de confusão. Os resultados foram apresentados em tabela contendo as estimativas brutas e ajustadas, com seus respectivos intervalos de confiança e valores de p.

O teste de associação entre a variável desfecho IRAS e as variáveis qualitativas foi reavaliado utilizando uma subamostra balanceada de 150 episódios, com distribuição semelhante entre os grupos com e sem infecção (aproximadamente 1:1). Essa estratégia foi adotada para reduzir o desequilíbrio entre os grupos observado na amostra original, minimizando possíveis vieses e permitindo uma comparação mais adequada entre as variáveis qualitativas.[29] As variáveis “sexo”, “cateter venoso central” e “doenças pulmonares” não apresentaram valor de $p < 0,05$, não sendo possível rejeitar a hipótese de independência em relação ao desfecho. As demais variáveis apresentaram valor de $p < 0,05$, indicando associação estatisticamente significativa com o desfecho, sendo, portanto, selecionadas para as etapas subsequentes da análise do modelo preditivo, que foi conduzida utilizando os seguintes algoritmos de *machine learning*: *generalized linear model* (GLM), regressão logística; *generalized additive models for location, scale, and shape* (GAMLSS); *random forest*; *support vector machine*; e *extreme gradient boosting* (XGBoost) (Figura 3). A validação do modelo foi realizada por meio da comparação entre as predições e os casos reais de IRAS, utilizando o software R (versão 4.1.1) e banco de dados MySQL.



A: acurácia; AUC: área sob a curva; E: especificidade; GAMLSS: *generalized additive models for location, scale, and shape*; GLM: *generalized linear model*; S: sensibilidade; VC: validação cruzada; XGBoost: *extreme gradient boosting*.

Figura 3. Resumo dos algoritmos de *machine learning* e diversos conjuntos de dados realizados

Para aprimorar a estimativa do modelo e reduzir o erro de previsão, foi aplicada a validação cruzada de 5 vezes nos algoritmos GLM e *random forest*. Modelos GAMLSS foram explorados por sua flexibilidade, utilizando distribuições "Zero-Inflated Binomial" e "Double Binomial" para lidar com o excesso de zeros[30] (Figura 3).

Cada algoritmo foi testado com 3 amostras, que incluem: o conjunto de dados original, um conjunto de amostras de 2.000 pacientes e uma amostra gerada pela técnica *Random Over Sampling Examples* (ROSE), que criou dados sintéticos para lidar com o desequilíbrio de classes, utilizando o pacote ROSE[31] (Figura 3). Os conjuntos de dados foram divididos em 80% para treinamento e 20% para teste utilizando o pacote caTools[32] (Figura 3). Os resultados do GLM foram apresentados como razão de chances (= *odds ratio*, OR), e as pontuações de importância de Gini foram analisadas para *random forest* e XGBoost. O desempenho preditivo foi avaliado utilizando as métricas de precisão, sensibilidade, especificidade, acurácia e área sob a curva (AUC – *area under the curve)*.

3 RESULTADOS

Nosso estudo analisou dados de 6.253 pacientes ao longo de um período de 4 anos, durante o qual foram relatadas 125 IRAS, envolvendo 2% do número total de pacientes. Destas, houve 33 casos de infecção por *Clostridioides difficile*, 10 casos de infecção respiratória inferior associada à ventilação mecânica, 9 casos de infecção da corrente sanguínea relacionada ao cateter venoso central, 5 casos de infecção do trato urinário associada à cateter vesical de demora, e 68 casos de IRAS não associadas a dispositivos invasivos, incluindo 21 identificadas como infecção de sítio cirúrgico.

Um total de 920.489 oportunidades de higiene das mãos foi registrado entre várias categorias profissionais, com uma taxa de conformidade para higiene das mãos de 49,3% (Tabela 1). O tempo médio de permanência dos pacientes nos leitos auditados foi de 12,6 dias (desvio padrão de 23,5), com uma mediana de 6 dias (intervalo interquartil de 3 a 13).

Dos pacientes, 55,3% eram do sexo masculino. A média de idade foi de 72,5 anos (desvio-padrão de 18,6) enquanto a mediana foi de 76 anos (intervalo interquartil de 60 a 87). A prevalência de comorbidades foi a seguinte: diabetes mellitus (0,2%), hipertensão arterial sistêmica (0,5%), insuficiência renal (0,4%), neoplasias (1,5%), doenças hepáticas (0,05%), doenças pulmonares (9,5%) e anemia (0,3%). Considerando a baixa prevalência dessas comorbidades nos dados coletados, incluímos apenas a variável de doença pulmonar nas análises de regressão logística simples e no teste de associação com subamostra de 150 pacientes.

Em relação aos dispositivos invasivos, os percentuais de uso foram os seguintes: cateter venoso central (7,6%), cateter vesical de demora (28,4%) e ventilação mecânica (8%). A proporção de pacientes submetidos a procedimentos e terapias foi a seguinte: corticosteroides (54,5%), drogas vasoativas (26,8%), quimioterapia (2,2%), transfusão de componentes sanguíneos (25,1%), procedimentos cirúrgicos (26,2%), hemodiálise (3,3%) e ventilação mecânica não invasiva (25,9%).

Tabela 1. Dados sobre higiene das mãos, número de pacientes, número de infecções relacionadas à assistência à saúde e mediana do tempo de internação

Dados	
Número de pacientes	6.253
Número de oportunidades de higiene das mãos	920.489
Número de higiene das mãos realizadas	453.997
Taxa de adesão a higiene das mãos (%)	49,3
Número de IRAS (% do total de pacientes) (%)	125 (2)
Número de infecções por *Clostridioides difficile* (% do total de IRAS)	33 (26,4)
Número de infecções respiratórias inferiores associadas à ventilação mecânica (% do total de IRAS)	10 (8)
Número de infecções da corrente sanguínea relacionadas à cateter venoso central (% do total de IRAS)	9 (7,2)
Número de infecções do trato urinário associadas à cateter urinário de demora (% do total de IRAS)	5 (4)
Mediana do tempo de internação (dias)	6 (IQR 3-13)

Análises bivariadas foram conduzidas em grupos com e sem IRAS, analisando as variáveis coletadas nos prontuários médicos. Nessa análise inicial, houve uma diferença nas taxas de adesão à higiene das mãos entre os grupos com e sem IRAS (p=0,05), com uma mediana de 0,52 (intervalo interquartil 0,33 – 0,72) no grupo sem IRAS e 0,45 (intervalo interquartil 0,33 – 0,63) no grupo com IRAS. A mediana de tempo de permanência na unidade entre os grupos foi de 6 e 39 dias (p<0,001), respectivamente nos grupos sem e com IRAS. A mediana de idade foi de 76 e 82 anos (p=0,003) nos grupos sem e com IRAS, respectivamente.

Na análise de variáveis qualitativas, não houve associação estatisticamente significativa (p<0,05) entre IRAS e as variáveis de sexo e doenças pulmonares. Todas as outras variáveis apresentaram valor-p<0,05. Após as análises inferenciais bivariadas, conduzimos modelos de regressão logística múltipla, bem como análises univariada e multivariada, com o conjunto de dados (Tabela 2). Na análise multivariada, a taxa de adesão à higiene das mãos apresentou uma razão de chances de 0,26 (IC 95% 0,10 – 0,65, p=0,004). Outras variáveis, como maior tempo de permanência, quimioterapia, transfusão de sangue, procedimento cirúrgico, hemodiálise, ventilação mecânica não invasiva e cateter vesical de demora, foram estatisticamente associadas a um risco maior de IRAS na análise multivariada (Tabela 2).

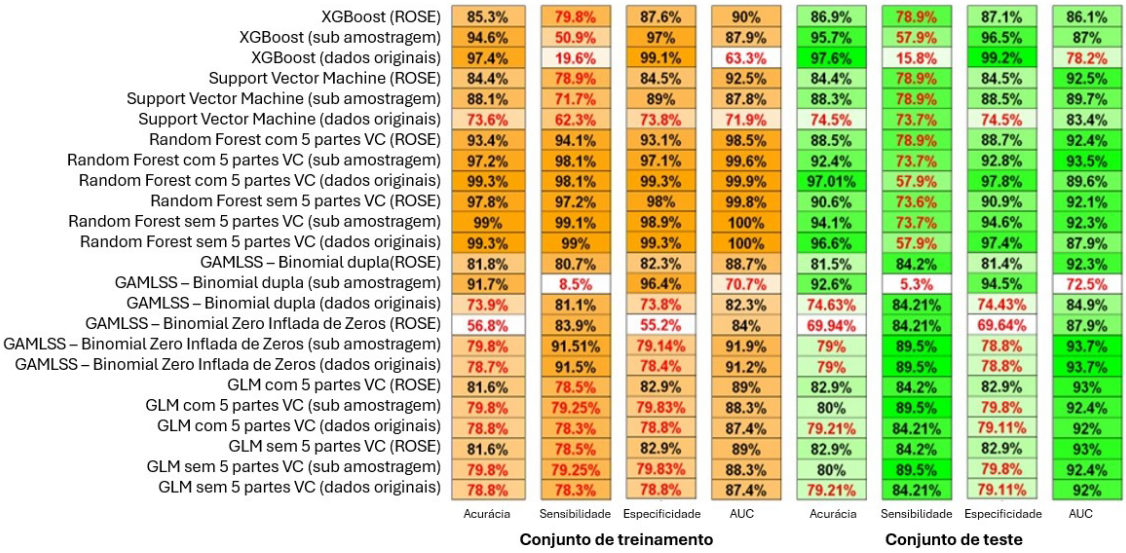
Tabela 2. Análises univariada e multivariada do modelo de regressão logística múltipla no conjunto de dados original

Variáveis	IRAS	Sem IRAS	Análise univariada			Análise multivariada		
			RC	IC 95% (LI – LS)	Valor-p	OR	IC 95% (LI – LS)	Valor-p
Gênero, masculino	73 (58,4)	3387 (55,27)	1,15	(0,78 – 1,72)	0,459	1,21	(0,79 – 1,88)	0,373
Idade (anos), média (DP)	77,17 (18,06)	72,43 (18,66)	1,01	(1,00 – 1,02)	0,021	1,00	(0,99 – 1,02)	0,230
Taxa adesão de higiene das mãos, média (DP)	0,48 (0,22)	0,52 (0,26)	0,53	(0,25 – 1,02)	0,106	0,26	(0,10 – 0,65)	0,004
Tempo de internação, mediana (dias) (ICQ)	39 [20 – 85]	6 [3 – 12]	1,02	(1,01 – 1,02)	<0,001	1,01	(1,00 – 1,01)	<0,001
Cateter venoso central, sim	32 (25,6)	447 (7,29)	3,97	(2,46 – 6,21)	<0,001	0,75	(0,41 – 1,34)	0,357
Corticosteroides, sim	107 (85,6)	3301 (53,87)	5,17	(3,08 – 9,32)	<0,001	1,38	(0,76 – 2,64)	0,300
Droga vasoativa, sim	79 (63,2)	1598 (26,08)	4,47	(3,02 – 6,69)	<0,001	1,08	(0,63 – 1,84)	0,776
Neoplasia, sim	5 (4)	90 (1,47)	3,13	(1,08 – 7,16)	0,015	1,84	(0,60 – 4,53)	0,224
Doença pulmonar, sim	17 (13,6)	579 (9,45)	1,80	(1,03 – 2,98)	0,028	1,76	(0,92 – 3,22)	0,074
Quimioterapia, sim	13 (10,4)	126 (2,06)	4,83	(2,30 – 9,08)	<0,001	3,49	(1,51 – 7,44)	0,002
Transfusão de componentes sanguíneos, sim	91 (72,8)	1484 (24,22)	7,55	(5,00 – 11,69)	<0,001	1,67	(0,94 – 3,01)	0,081
Cirurgia, sim	87 (69,6)	1550 (25,29)	6,58	(4,38 – 10,09)	<0,001	1,98	(1,09 – 3,64)	0,026
Hemodiálise, sim	33 (26,4)	176 (2,87)	11,01	(6,71 – 17,53)	<0,001	3,01	(1,62 – 5,45)	<0,001
Ventilação mecânica não-invasiva, sim	80 (64)	1541 (25,15)	4,87	(3,28 – 7,30)	<0,001	1,65	(1,00 – 2,72)	0,047
Ventilação mecânica invasiva, sim	47 (37,6)	458 (7,47)	6,62	(4,34 – 9,94)	<0,001	1,46	(0,85 – 2,47)	0,161
Cateter vesical de demora, sim	96 (76,8)	1681 (27,43)	8,47	(5,47 – 13,58)	<0,001	2,66	(1,55 – 4,67)	<0,001

IRAS: Infecções relacionadas à assistência à saúde; ICQ: intervalo interquartílico; DP: desvio padrão; LI: limite inferior; LS: limite superior; IC: intervalo de confiança; RC: razão de chances (= *odds ratio*, OR).

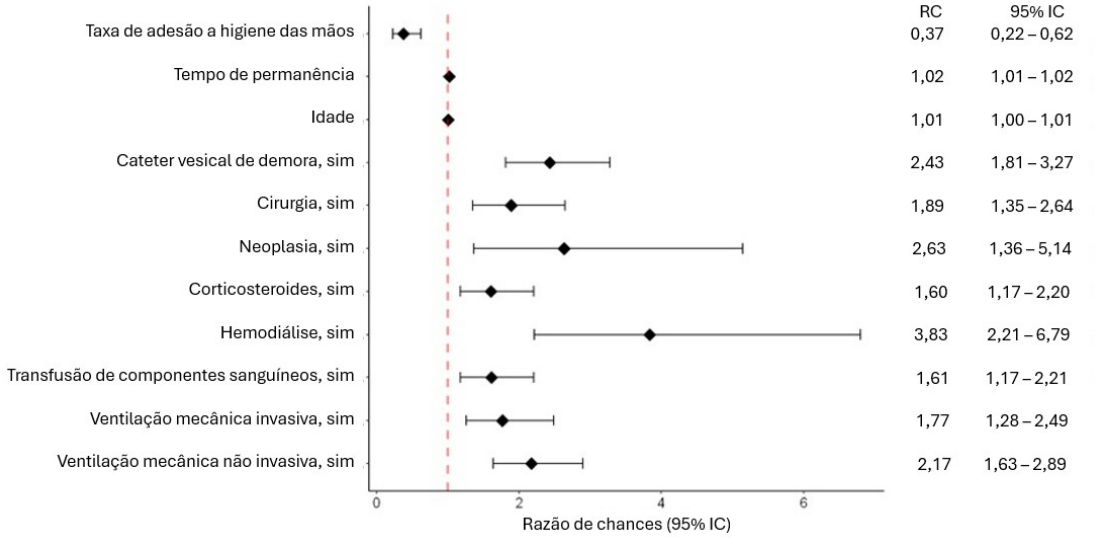
Subsequentemente, realizamos análises utilizando algoritmos de *machine learning.* O algoritmo com o melhor desempenho na previsão de IRAS, com base nas análises dos prontuários eletrônicos e da higiene das mãos, foi o GLM com e sem validação cruzada de 5 partes (Acurácia: 82,9; Sensibilidade: 84,2%; Especificidade: 82,9%; AUC: 93%). Com o uso do algoritmo de *random forest*, sem validação cruzada de 5 partes, a acurácia, sensibilidade, especificidade e AUC foram 90,6%, 73,6%, 90,9% e 92,1%, respectivamente. No entanto, ao aplicar o algoritmo *random forest* com validação cruzada de 5 partes, visando melhorar a estimativa do modelo, reduzir o erro de previsão e aprimorar as medidas de qualidade de ajuste, houve uma perda de acurácia, um sobre ajuste dos modelos. O modelo XGBoost alcançou uma acurácia de 86,9%, sensibilidade de 78,9%, especificidade de 87,1% e AUC de 86,1%. Com a técnica de *support vector machine*, os valores de acurácia, sensibilidade, especificidade e AUC foram 84,4%, 78,9%, 84,5% e 92,5%, respectivamente. Além disso, no algoritmo GAMLSS, sem validação cruzada de 5 partes, foram testadas as distribuições “Binomial Inflada de Zeros” e “Binomial Dupla”. O modelo com “Binomial Inflada de Zeros” teve acurácia de 69,9%, sensibilidade de 84,2%, especificidade de 69,6% e AUC de 87,9%. O modelo com a distribuição “Binomial Dupla” apresentou acurácia de 81,5%, sensibilidade de 84,2%, especificidade de 81,4% e AUC de 92,3%.

A variável taxa de adesão à higiene das mãos foi considerada importante para predição do risco de IRAS em todos os algoritmos testados. O GLM sem e com validação cruzada de 5 partes, o *random forest* sem validação cruzada de 5 partes e o XGBoost apresentaram o melhor desempenho quando modelados usando o conjunto de testes balanceado pelo método ROSE. Os resultados estão resumidos na figura 4. Na figura 5, apresentamos as razões de chances das variáveis no GLM com validação cruzada de 5 partes, que exibiu todas as métricas acima de 80%.



AUC: área sob a curva; GAMLSS: *generalized additive models for location, scale, and shape*; GLM*: generalized linear model*; ROSE: *Random Over Sampling Examples*; XGBoost: *extreme gradient boosting*. Dados originais, sub amostragem e ROSE, com conjunto de treinamento e conjunto de teste.

Figura 4. Resumo das métricas de qualidade do modelo



IC: intervalo de confiança; RC: razão de chances (= *odds ratio*, OR). Resumo das razões de chances para cada variável pelo algoritmo de *generalized linear model* com validação cruzada em 5 partes, balanceada usando *Random Over Sampling Examples.*

Figura 5. Razões de chances das variáveis preditoras derivadas de modelo linear generalizado com validação cruzada

Nas figuras 6 e 7, as variáveis utilizadas na modelagem *random forest* sem validação cruzada de 5 partes e no XGBoost são mostradas em ordem de importância pela redução média no índice de Gini, destacando as baixas taxas de higiene das mãos como o segundo fator de risco direto mais importante para o desenvolvimento de IRAS, superado apenas pelo tempo de internação, com boa acurácia.

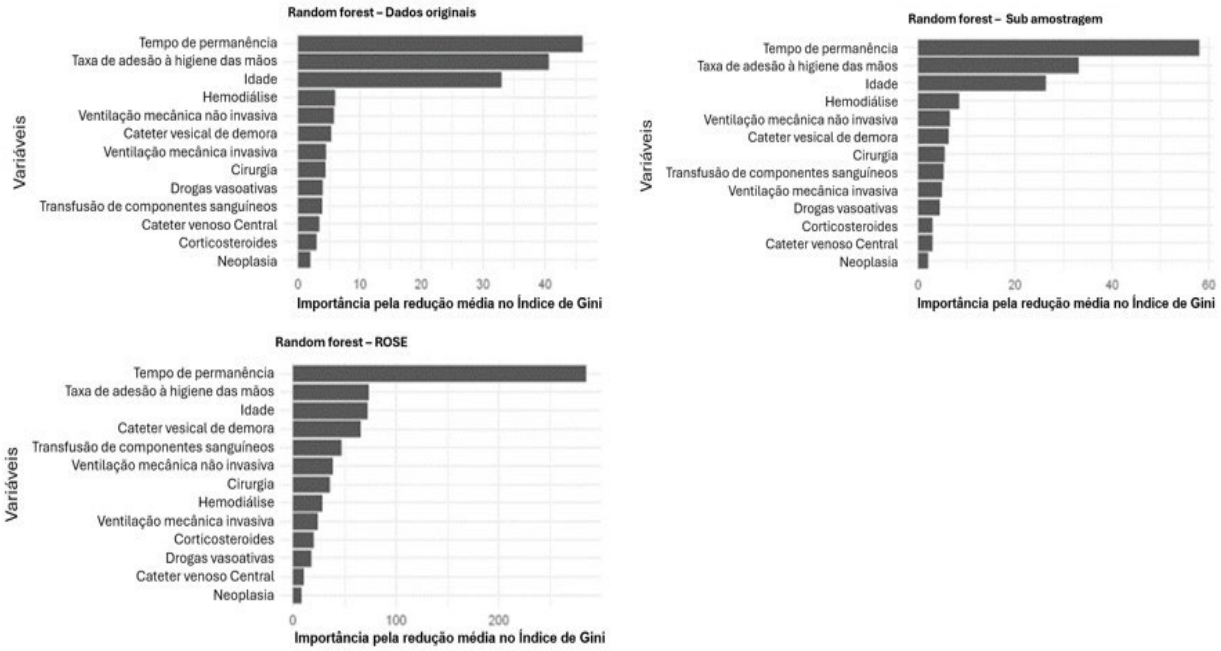


Figura 6. Importância das variáveis dos modelos *random forest* sem validação cruzada em 5 partes

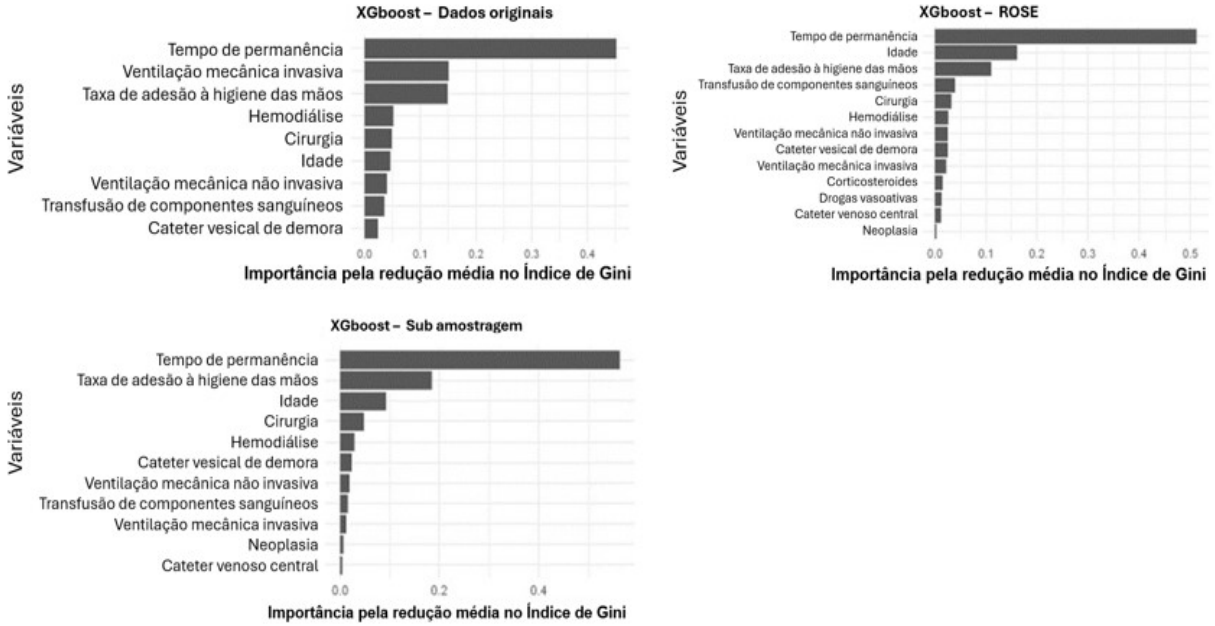


Figura 7. Importância das variáveis dos modelos XGBoost

4 DISCUSSÃO

Nosso estudo demonstrou a integração de técnicas de *machine learning* oferecendo uma nova abordagem para analisar a relação entre a adesão à higiene das mãos e IRAS. Este estudo demonstrou o potencial de empregar *data mining* e análises estatísticas para avaliar o impacto das práticas de higiene das mãos dos profissionais de saúde no risco de desenvolver IRAS, considerando outros fatores de risco individuais. Nossos achados permitiram confirmar que a adesão à higiene das mãos é um dos fatores mais relevantes para predizer IRAS.

Apesar de várias estratégias para melhorar a adesão à higiene das mãos em ambientes de saúde, o impacto na redução de IRAS permanece complexo. Embora os sistemas eletrônicos de auditoria de higiene das mãos ofereçam uma solução promissora ao aumentar o número de oportunidades auditadas, sua eficácia em melhorar as taxas de adesão varia amplamente. O sistema eletrônico de auditoria de higiene das mãos foi validado no hospital e comparado com o método de observação manual direta, mostrando uma correlação nas taxas de higiene das mãos. Dados sobre essa validação foram publicados em outros estudos realizados no Hospital Israelita Albert Einstein, São Paulo, Brasil desde 2013.[26,27]

Além disso, a correlação direta entre as taxas de higiene das mãos e a incidência de IRAS nem sempre é observada. Isso levanta questões sobre a adequação dos métodos de monitoramento atuais e a necessidade de abordagens mais sofisticadas para entender as dinâmicas subjacentes. Um dos fatores que pode contribuir para essa discrepância é o efeito Hawthorne, amplamente descrito em auditorias manuais de adesão à higiene das mãos. Esse efeito ocorre quando o comportamento dos profissionais de saúde é modificado pela percepção de estar sendo observado, levando a uma superestimação das taxas reais de adesão durante períodos de observação direta. Auditorias realizadas por observadores humanos tendem a capturar apenas uma fração limitada das oportunidades de higiene das mãos e, frequentemente, concentram-se em momentos previsíveis, como a entrada e saída dos quartos dos pacientes. Além disso, a presença física do observador pode induzir mudanças temporárias no comportamento, reduzindo a validade externa dos dados coletados e dificultando a extrapolação dos resultados para a prática cotidiana. Essas limitações reforçam a necessidade de métodos de monitoramento mais contínuos, objetivos e

integráveis, capazes de fornecer dados mais representativos do comportamento real dos profissionais de saúde ao longo do tempo.[33,34]

A integração de dados gerados por novas tecnologias de saúde no campo do controle de IRAS juntamente com dados do sistema eletrônico do paciente, por meio de modelos de *machine learning*, pode levar ao desenvolvimento de sistemas de alerta e pontuação de risco para IRAS no futuro. Esses sistemas possibilitam intervenções e a sensibilização das equipes de saúde para aumentar as práticas de higiene das mãos de forma individualizada para cada paciente.

O uso de técnicas de *data mining* e *machine learning* para analisar fatores de risco e criar escores para infecções hospitalares aumentou nas últimas duas décadas, destacando-se como uma abordagem promissora para suporte clínico em tempo real, com potencial para integração em sistemas de prontuário eletrônico. O uso de algoritmos de *machine learning*, como GLM e *random forest*, pode contribuir para maior precisão e exatidão em comparação com modelos preditivos tradicionais, como a regressão logística.[35] Um estudo recente utilizou o algoritmo *random forest* para desenvolver um modelo preditivo de infecções do trato urinário associadas à cateter vesical de demora. O modelo apresentou alta acurácia ao identificar a maioria dos casos positivos de infecções do trato urinário associadas à cateter vesical de demora e ao descartar corretamente a maioria dos indivíduos sem risco, alcançando sensibilidade de 97,63%, especificidade de 97,36% e valor preditivo positivo de 22,85%. Além disso, a ferramenta permitiu estimar o risco de infecções do trato urinário associadas à cateter vesical de demora até 48 horas após a remoção do cateter.[36] Outro estudo recente, conduzido em 2022, usou o algoritmo *random forest* em um grande conjunto de dados para avaliar a precisão no diagnóstico de pneumonia associada à ventilação mecânica nas primeiras 24 horas após a intubação, encontrando um bom desempenho (AUC: 84%, sensibilidade: 74%, e especificidade: 71%), com maior acurácia em comparação ao *Clinical Pulmonary Infection Score* previamente validado.[37]

Até o momento desta revisão não encontramos estudos na literatura que integrassem dados de sistemas eletrônicos de auditoria de higiene das mãos por paciente com dados de prontuários eletrônicos para a análise de fatores de risco e o impacto direto da higiene das mãos na ocorrência de IRAS. Alguns estudos analisaram indiretamente a causalidade entre o aumento das taxas de higiene das mãos e a redução de IRAS. Por exemplo, uma recente revisão sistemática da literatura encontrou 11 estudos que avaliavam dispositivos eletrônicos de auditoria de higiene das mãos para

aumentar as taxas de higiene das mãos.[38] Apenas dois estudos demonstraram uma redução estatisticamente significativa na densidade de IRAS relatadas após um aumento nas taxas de higiene das mãos usando sistemas eletrônicos de auditoria de higiene das mãos. No entanto, esses dois estudos não analisaram os fatores de risco em nível individual por paciente; foram conduzidos sem avaliações pré e pós-intervenção, e os dados não isolaram o impacto direto da higiene das mãos no aumento do risco de IRAS, considerando que os dados foram analisados globalmente e outras intervenções de prevenção ocorreram simultaneamente.[39,40]

O primeiro estudo foi conduzido ao longo de 14 meses antes da intervenção e 8 meses após, utilizando um sistema eletrônico de auditoria de higiene das mãos para avaliar a redução de IRAS. Nos meses finais do período pré-intervenção, foram realizadas atividades de treinamento focadas na prevenção de infecções do trato urinário associadas à cateter vesical de demora e infecções primárias da corrente sanguínea relacionadas a cateteres. Durante a intervenção, que envolveu o uso do sistema eletrônico, houve mudanças nos dispositivos invasivos empregados no hospital. Como os próprios autores destacaram, não foi possível isolar o impacto do sistema eletrônico de higiene das mãos na redução das IRAS. Não foram coletados dados sobre fatores de risco dos pacientes para IRAS, e as taxas de higiene das mãos, assim como as densidades de IRAS, foram calculadas de forma geral no hospital. Assim, a relação entre o aumento das taxas de higiene das mãos e a diminuição das IRAS foi indireta.[39]

O segundo estudo, de Akkoc et al., investigou-se o impacto de um sistema eletrônico de registro e lembrete de higiene das mãos na redução de IRAS, utilizando um desenho de pré e pós-intervenção. A pesquisa foi conduzida em um curto período, com 2 meses de auditoria convencional de higiene das mãos antes da intervenção e 2 meses após, com a introdução do sistema eletrônico. Durante o uso do sistema, houve um aumento nas taxas de higiene das mãos e uma redução significativa na densidade de IRAS em comparação com o período anterior à intervenção (31,89 *versus* 18,43%, *odds ratio* = 2,07; IC 95% = 1,16–3,69, p=0,013; RR = 0,58; IC 95% = 0,37–0,89, p=0,014). O estudo analisou o uso dos dispensadores de álcool em gel, mas não avaliou as oportunidades de higiene das mãos que não foram realizadas, além de os profissionais de saúde não utilizarem crachás durante a intervenção.[40]

Os sistemas eletrônicos de auditoria de higiene das mãos também podem ser usados em tempo real para melhorar a adesão e a qualidade da higiene das mãos. Não encontramos estudos com esse propósito que realizaram análises de

machine learning para calcular o impacto direto e o risco de IRAS. Um estudo recente utilizou um sistema de feedback visual em tempo real, com um sinalizador de semáforo, para avaliar o tempo adequado de secagem do desinfetante para as mãos antes da abertura de incubadoras em uma UTI neonatal e demonstrou um aumento significativo na adesão ao tempo de secagem do desinfetante antes de abrir a incubadora no grupo com sinalização por cores.[41] Outro estudo instalou câmeras de vídeo posicionadas para monitorar todas as pias e dispensadores de álcool em gel em uma UTI, registrando a prática de higiene das mãos pelos profissionais de saúde. Sensores nas portas detectavam entradas e saídas de indivíduos. Quando os auditores de vídeo remotos observavam um profissional realizando a higiene das mãos ao entrar ou sair, era atribuída uma aprovação; caso contrário, uma reprovação. O desempenho era exibido continuamente em painéis eletrônicos nos corredores, e relatórios resumidos eram enviados por e-mail aos supervisores. Os dados eram apresentados de forma agregada, incluindo informações gerais como data, hora da observação, número de oportunidades para higiene das mãos durante o período auditado e número de vezes em que a prática foi realizada, sem identificar individualmente os profissionais. O monitoramento remoto, aliado ao feedback em tempo real, foi associado a um aumento significativo na adesão à higiene das mãos.[42]

Essa tecnologia de *feedback* em tempo real sobre a higiene das mãos poderia ter seus dados capturados em um banco de dados ou incluídos no prontuário eletrônico para futuras análises de risco usando algoritmos de *machine learning,* como os utilizados em nosso estudo, GLM e *random forest*.

Há outros estudos que usaram técnicas de *machine learning* para avaliar fatores de risco para IRAS, mas sem o uso de dados de higiene das mãos. Uma revisão sistemática da literatura[20] de 2020 incluiu 27 estudos que usaram *machine learning* para esse propósito, com evidências moderadas de desempenho superior em comparação com métodos convencionais.[20] A maioria dos estudos identificados se concentrou na avaliação de fatores de risco para infecções específicas, como infecção por *Clostridioides difficile*,[43,44] infecção de sítio cirúrgico,[45–52] infecção primária de corrente sanguínea relacionada a cateter,[53,54] ventriculite, meningite,[55] pneumonia[56] e infecção do trato urinário associada à cateter vesical de demora.[57,58] Dois estudos utilizaram técnicas de *machine learning* para detecção de sepse,[59,60] enquanto outros oito analisaram fatores de risco relacionados às IRAS de forma geral.[61–68]

Nosso estudo teve limitações. É um estudo retrospectivo, com dados coletados automaticamente de prontuários eletrônicos, conduzido em um único centro, o que pode limitar a generalização de nossos achados para diferentes cenários de saúde. Validação externa em ambientes diversos é necessária para confirmar a aplicabilidade de nossos resultados em vários contextos. Apesar do considerável número de oportunidades de higiene das mãos auditadas pelo sistema eletrônico, é importante reconhecer que nem todos os eventos de higiene das mãos foram capturados. Isso ocorreu porque alguns profissionais de saúde não diretamente associados à unidade semi-intensiva ou à UTI, como o corpo clínico do hospital, não terem sido equipados com crachás contendo sensores. Importante ressaltar que o sistema eletrônico de auditoria avaliou predominantemente os momentos de higiene das mãos antes e após o contato com o paciente, não contemplando de forma direta todos os cinco momentos preconizados pela Organização Mundial da Saúde, como após risco de exposição a fluidos corporais, e após contato com superfícies próximas ao paciente. Consequentemente, nossos dados podem não representar totalmente as práticas reais de higiene das mãos de todos os profissionais de saúde no hospital. Além disso, a prevalência de algumas comorbidades em nossa população de estudo foi menor do que na população geral. A coleta de dados sobre variáveis como comorbidades a partir de informações do CID nos prontuários eletrônicos tem limitações, principalmente devido a erros de codificação ou falta de entrada das informações no sistema.[69] Essa discrepância pode limitar a generalização de nossos achados, especialmente ao avaliar a associação entre comorbidades e o risco de IRAS. Para minimizar esse viés, incluímos apenas a variável “doença pulmonar”, onde o percentual era semelhante ao encontrado na literatura, nas análises de regressão logística simples e nos algoritmos de *machine learning*. Levamos em consideração variáveis de risco para IRAS, como tempo de internação, uso de dispositivos invasivos e procedimentos/tratamentos, usando análise bivariada para comparar a adesão à higiene das mãos e outros fatores entre pacientes com e sem IRAS. Modelos de regressão logística multivariada foram usados para ajustar essas variáveis e isolar o efeito da higiene das mãos. Diferentes algoritmos de *machine learning*, incluindo GLM, *random forest* e XGBoost, preveem IRAS e ajustam o impacto de vários fatores, com técnicas como validação cruzada para melhorar a precisão do modelo. Tempo de internação e o uso de dispositivos invasivos foram especificamente analisados e incluídos nos modelos para controlar seus efeitos no risco de IRAS. Essas abordagens visam mitigar o impacto de vieses e fatores de confusão, fornecendo uma

compreensão mais clara do papel da higiene das mãos na prevenção de IRAS. Nossos resultados também destacam a importância da coleta precisa de dados e métodos analíticos robustos para identificar fatores de risco para IRAS. Além disso, os achados ressaltam o potencial dos modelos de *machine learning*, particularmente dos algoritmos GLM e *random forest*, na previsão de desfechos de pacientes e na orientação de intervenções direcionadas para mitigar o risco de IRAS.

5 CONCLUSÃO

Nosso estudo demonstrou o impacto da variável taxa de higiene das mãos no risco de desenvolvimento de IRAS por meio da validação interna de modelos de *machine learning.* Os modelos exibiram alta precisão, sendo capaz de avaliar o risco de desenvolvimento de IRAS, integrando dados das taxas de higiene das mãos obtidas de um sistema eletrônico de auditoria para o cuidado de cada paciente com variáveis individuais dos prontuários eletrônicos. O monitoramento contínuo e a análise detalhada das oportunidades de higiene das mãos podem ajudar a ajustar estratégias de intervenção e aumentar a adesão às práticas recomendadas. Mais pesquisas são necessárias para refinar esses modelos preditivos e validar sua eficácia em ambientes de saúde. Esforços devem ser feitos para integrar essas percepções na prática clínica para aprimorar as medidas de controle de infecções e melhorar os desfechos dos pacientes.

6 APRESENTAÇÃO DO ARTIGO

Cotia ALF, Scorsato AP, da Silva Victor E, Prado M, Gagliardi G, de Barros JEV, Generoso JR Jr, de Menezes FG, Hsieh MK, Lopes GOV, Edmond MB, Perencevich EN, Goto M, Wey SB, Marra AR. Integration of an electronic hand hygiene auditing system with electronic health records using machine learning to predict hospital-acquired infection in a health care setting. Am J Infect Control. 2025;53(1):58-64. doi: 10.1016/j.ajic.2024.09.012.

7 MÉTRICAS

7.1 Métricas do periódico

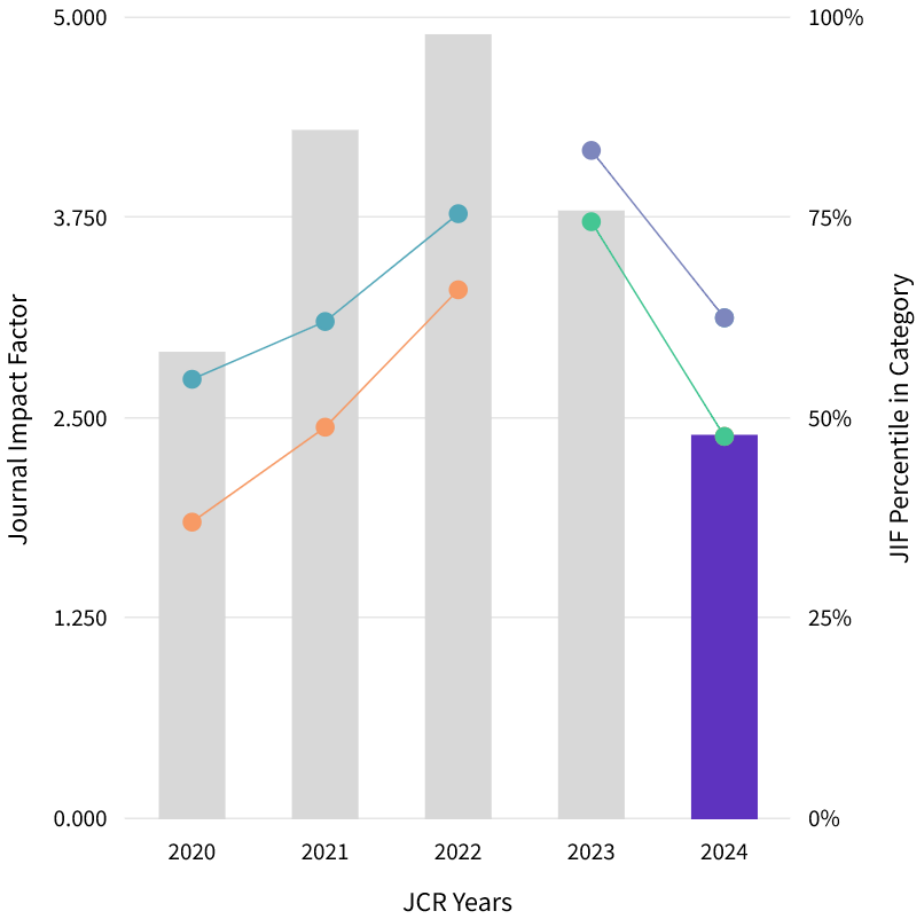
Métricas do periódico nos últimos 5 anos.

7.1.2 Fator de impacto

Periódico: American Journal of Infection Control

ISSN: 0196-6553.

Fator de Impacto: 2.4



Fonte: *Journal Citation Reports*

7.2 Métricas do artigo

7.2.1 Citações

Há 3 citações para o artigo nas bases de dados *Scopus*, *Web of Science* e *Google Scholar* e 0 em capítulos de livros.

Citações recebidas	Scopus/ Web of Science/ Google Scholar	Capítulo de livro	Total
Gastaldi S, Tartari E, Satta G, Allegranzi B. Advancing infection prevention and control through artificial intelligence: a scoping review of applications, barriers, and a decision-support checklist. Antimicrob Steward Healthc Epidemiol. 2025;5(1):e317.	1		1
Wey SB. Dr. SB Wey: a path of influence and discovery in hospital epidemiology. Antimicrob Steward Healthc Epidemiol. 2024;4(1):e179.	1		1
Vidhya B, Fallah M, Vallabhaneni R, Dhanamalar M. Prediction of Medical Care Costs using Random Forest Regressor for Personalized Healthcare. 3rd International Conference on Data Science and Information System, ICDSIS 2025, Conference paper.	1		1

Avaliação realizada em 26/01/2026.

7.2.2 Altimetria

Dados altimétricos coletados em 26/01/2026.

Escore	Variáveis	Quantidade
	Menções em rede	
	X	2
	Origem das menções	
	EUA	1
	Citações	
	Dimensions	
	Capítulo de livro	0
	Artigo	4
	Leituras	
	Mendeley	32

7.3 Métricas do autor

7.3.1 Redes científicas

Rede científica	Link para acesso ao perfil
ORCID	https://orcid.org/0000-0001-7141-5009
Lattes	http://lattes.cnpq.br/2512267784575870
ScopusID	https://www.scopus.com/authid/detail.uri?authorId=57223158718&origin=recordpage
Publons/ ResearchID	https://www.webofscience.com/wos/author/record/KIG-0670-2024
ResearchGate	https://www.researchgate.net/profile/Andre-Cotia-2
Google Scholar	https://scholar.google.com/citations?user=toodSRoAAAAJ&hl=pt-BR&oi=ao

7.3.2 Trabalhos publicados

Trabalhos publicados e indexados na base de dados referencial *Web of Science*.

Publicações				
Ano	**FI**	**Categoria**	**Referência**	**Citações**
2021	1.9	Article	Paiva Y, Nagano DS, **Cotia ALF**, Guimarães T, Martins RCR, Perdigão Neto LV, Côrtes MF, Marchi AP, Corscadden L, Machado AS, Paula AI, Franco LAM, Neves PR, Levin AS, Costa SF. Colistin-resistant Escherichia coli belonging to different sequence types: genetic characterization of isolates responsible for colonization, community- and healthcare-acquired infections. Rev Inst Med Trop Sao Paulo. 2021;63:e38.	7
2022	4.7	Article	Chuffi S, Gomes-Gouvêa MS, Casadio LVB, Nastri ACSS, Gonzalez MP, **Cotia ALF**, Aranda AGD, Tenore SB, Ono SK, Malta FM, Madalosso G, Ferreira PRA, Carrilho FJ, Pinho JRR. The Molecular Characterization of Hepatitis A Virus Strains Circulating during Hepatitis A Outbreaks in São Paulo, Brazil, from September 2017 to May 2019. Viruses. 2021;14(1):73.	10
2023	4.3	Article	**Cotia A**, Oliveira Junior HA, Matuoka JY, Boszczowski Í. Clinical Equivalence between Generic Versus Branded Antibiotics: Systematic Review and Meta-Analysis. Antibiotics (Basel). 2023;12(5):935.	4

2024	3.0	Brief Communication	Fernandes RA, Carneiro M, Makdisse M, Adiwardana NS, Telles JP, Vidal CFL, **Cotia ALF**, Gray M; Brazilian Collaborative on Antimicrobial Stewardship. The Brazilian collaborative on antimicrobial stewardship: A value-based healthcare approach. Braz J Infect Dis. 2024;28(2):103738.	2
2024	3.8	Article	**Cotia ALF**, Scorsato AP, da Silva Victor E, Prado M, Gagliardi G, de Barros JEV, Generoso JR Jr, de Menezes FG, Hsieh MK, Lopes GOV, Edmond MB, Perencevich EN, Goto M, Wey SB, Marra AR. Integration of an electronic hand hygiene auditing system with electronic health records using machine learning to predict hospital-acquired infection in a health care setting. Am J Infect Control. 2024:S0196-6553(24)00720-X.	2
2025	3.8	Article	Malheiro DT, Parreira KCJ, Celeghini PD, Callado GY, **Cotia ALF**, Cendoroglo Neto M, Bragatte MAS, Negretto Schrarstzhaupt I, Sampaio V, Kobayashi T, Edmond MB, Marra AR. COVID-19 Reinfections in the City of São Paulo, Brazil: Prevalence and Socioeconomic Factors. Open Forum Infect Dis. 2025;12(4):ofaf181	1
2025	0	Article	Celeghini PD, **Cotia ALF**, Hsieh MK, et al. Systematic literature review and meta-analysis of urinary tract infections and sexually transmitted infections in symptomatic women: prevalence, patterns and clinical implications. Gynecology and Obstetrics Clinical Medicine. 2025;5:e000222	0
2025	3.3	Article	Chauvin AG, Pardo I, **Cotia ALF**, Rosmino IL, Marins TA, Dos Santos LM, Barduchi B, Toniolo AR, Dos Santos RG, Malheiro DT, Scorsato AP, Victor EDS, Edmond MB, de Almeida SM, Marra AR. Beyond the Virus: The Collateral Impact of COVID-19 on Antimicrobial Consumption, Microbial Resistance, and Pharmacoeconomics. Pathogens. 2025;14(11):1126.	0

7.3.3 Mediana do fator de impacto dos trabalhos

Fator de impacto das publicações	Mediana
1.9	3.6
4.7	
4.3	
3.0	
3.8	
3.8	
0	
3.3	

7.3.4 Citações recebidas

Publicações x citações		
Ano	**Publicações**	**Citações**
2021	1	7
2022	1	10
2023	1	4
2024	2	4
2025	2	0
Total	**8**	**25**

7.3.5 H-index e m-quociente

Cálculo do H-index	
Tópicos	Dados
Quantidade de publicações	8
Soma do número de citações	25
Artigos que receberam citações	6
Média de citações por item	3,1
H-index	3

Cálculo do m-quociente		
H-index	3	**0,6**
Número de anos de publicação	5	

Publicações – h -index				
Ano	**FI**	**Categoria**	**Referência**	**Citações**
2022	4.7	Article	Chuffi S, Gomes-Gouvêa MS, Casadio LVB, Nastri ACSS, Gonzalez MP, **Cotia ALF**, Aranda AGD, Tenore SB, Ono SK, Malta FM, Madalosso G, Ferreira PRA, Carrilho FJ, Pinho JRR. The Molecular Characterization of Hepatitis A Virus Strains Circulating during Hepatitis A Outbreaks in São Paulo, Brazil, from September 2017 to May 2019. Viruses. 2021;14(1):73.	10
2021	1.9	Article	Paiva Y, Nagano DS, **Cotia ALF**, Guimarães T, Martins RCR, Perdigão Neto LV, Côrtes MF, Marchi AP, Corscadden L, Machado AS, Paula AI, Franco LAM, Neves PR, Levin AS, Costa SF. Colistin-resistant Escherichia coli belonging to different sequence types: genetic characterization of isolates responsible for colonization, community- and healthcare-acquired infections. Rev Inst Med Trop Sao Paulo. 2021;63:e38.	7
2023	4.3	Article	**Cotia A**, Oliveira Junior HA, Matuoka JY, Boszczowski Í. Clinical Equivalence between Generic Versus Branded Antibiotics: Systematic Review and Meta-Analysis. Antibiotics (Basel). 2023;12(5):935.	4
2024	3.0	Brief Communication	Fernandes RA, Carneiro M, Makdisse M, Adiwardana NS, Telles JP, Vidal CFL, **Cotia ALF**, Gray M; Brazilian Collaborative on Antimicrobial Stewardship. The Brazilian collaborative on antimicrobial stewardship: A value-based healthcare approach. Braz J Infect Dis. 2024;28(2):103738.	2
2024	3.8	Article	**Cotia ALF**, Scorsato AP, da Silva Victor E, Prado M, Gagliardi G, de Barros JEV, Generoso JR Jr, de Menezes FG, Hsieh MK, Lopes GOV, Edmond MB, Perencevich EN, Goto M, Wey SB, Marra AR. Integration of an electronic hand hygiene auditing system with electronic health records using machine learning to predict hospital-acquired infection in a health care setting. Am J Infect Control. 2024:S0196-6553(24)00720-X.	2
2025	3.8	Article	Malheiro DT, Parreira KCJ, Celeghini PD, Callado GY, **Cotia ALF**, Cendoroglo Neto M, Bragatte MAS, Negretto Schrarstzhaupt I, Sampaio V, Kobayashi T, Edmond MB, Marra AR. COVID-19 Reinfections in the City of São Paulo, Brazil: Prevalence and Socioeconomic Factors. Open Forum Infect Dis. 2025;12(4):ofaf181	1
2025	0	Article	Celeghini PD, **Cotia ALF**, Hsieh MK, et al. Systematic literature review and meta-analysis of	0

			urinary tract infections and sexually transmitted infections in symptomatic women: prevalence, patterns and clinical implications. Gynecology and Obstetrics Clinical Medicine. 2025;5:e000222	
2025	3.3	Article	Chauvin AG, Pardo I, **Cotia ALF**, Rosmino IL, Marins TA, Dos Santos LM, Barduchi B, Toniolo AR, Dos Santos RG, Malheiro DT, Scorsato AP, Victor EDS, Edmond MB, de Almeida SM, Marra AR. Beyond the Virus: The Collateral Impact of COVID-19 on Antimicrobial Consumption, Microbial Resistance, and Pharmacoeconomics. Pathogens**.** 2025;14(11):1126.	0

8 REFERÊNCIAS

1. Liu X, Long Y, Greenhalgh C, Steeg S, Wilkinson J, Li H, et al. A systematic review and meta-analysis of risk factors associated with healthcare-associated infections among hospitalized patients in Chinese general hospitals from 2001 to2022. J Hosp Infect. 2023;135:37-49.

2. Scott, DR. The direct medical costs of healthcare-associated infections in US hospitals and the benefits of prevention. Division of Healthcare Quality Promotion National Center for Preparedness, Detection, and Control of Infectious Diseases, Coordinating Center for Infectious Diseases, Centers for Disease Control and Prevention; 2009 Mar [cited 2024 Oct 7]. 13 p. Available from: https://stacks.cdc.gov/view/cdc/11550.

3. Haque M, Sartelli M, McKimm J, Abu Bakar M. Health care-associated infections - an overview. Infect Drug Resist. 2018;11:2321–33.

4. Boev C, Kiss E. Hospital-acquired infections: current trends and prevention. Crit Care Nurs Clin North Am. 2017;29(1):51-65.

5. Gallo RJ, Shieh L, Smith M, Marafino BJ, Geldsetzer P, Asch SM, et al. Effectiveness of an artificial intelligence-enabled intervention for detecting clinical deterioration. JAMA Intern Med. 2024;184(5):557-62.

6. Lotfinejad N, Peters A, Tartari E, Fankhauser-Rodriguez C, Pires D, Pittet D. Hand hygiene in health care: 20 years of ongoing advances and perspectives. Lancet Infect Dis. 2021 Aug;21(8):e209-e221. Erratum in: Lancet Infect Dis. 2021;21(10):e302.

7. Boyce JM, Pittet D; Healthcare Infection Control Practices Advisory Committee; HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Guideline for hand hygiene in health-care settings. Recommendations of the Healthcare Infection Control Practices Advisory Committee and the HICPAC/SHEA/APIC/IDSA Hand Hygiene Task Force. Society for Healthcare Epidemiology of America/Association for Professionals in Infection Control/Infectious Diseases Society of America. MMWR Recomm Rep. 2002;51(RR-16):1-45, quiz CE1-4.

8. WHO Guidelines on Hand Hygiene in Health Care: First Global Patient Safety Challenge Clean Care Is Safer Care. Geneva: World Health Organization; 2009 [cited 2024 Nov 26]. (WHO Guidelines Approved by the Guidelines Review Committee). Available from: http://www.ncbi.nlm.nih.gov/books/NBK144013/.

9. Ibrahim MA, Chow C, Poh BF, Ang B, Chow A. Differences in psychosocial determinants of hand hygiene between health care professional groups: insights from a mixed-methods analysis. Am J Infect Control. 2018;46(3):253-60.

10. Chang NC, Reisinger HS, Jesson AR, Schweizer ML, Morgan DJ, Forrest GN, et al. Feasibility of monitoring compliance to the My 5 Moments and Entry/Exit hand hygiene methods in US hospitals. Am J Infect Control. 2016;44(8):938-40.

11. Edmond MB, Goodell A, Zuelzer W, Sanogo K, Elam K, Bearman G. Successful use of alcohol sensor technology to monitor and report hand hygiene compliance. J Hosp Infect. 2010;76(4):364-5.

12. Boyce JM. Measuring Healthcare Worker Hand Hygiene Activity: Current Practices and Emerging Technologies. Infect Control Hosp Epidemiol. 2011;32(10):1016-28.

13. Marra AR, Sampaio Camargo TZ, Magnus TP, Blaya RP, Dos Santos GB, Guastelli LR, et al. The use of real-time feedback via wireless technology to improve hand hygiene compliance. Am J Infect Control. 2014;42(6):608-11.

14. Boudjema S, Dufour JC, Aladro AS, Desquerres I, Brouqui P. MediHandTrace®: a tool for measuring and understanding hand hygiene adherence. Clin Microbiol Infect. 2014;20(1):22-8.

15. Generoso JR, Casaroto E, Neto AS, Prado M, Gagliardi GM, De Menezes FG, et al. Comparison of two electronic hand hygiene systems using real-time feedback via wireless technology to improve hand hygiene compliance in an intensive care unit. Antimicrob Steward Healthc Epidemiol. 2022;2(1):e127.

16. Lacey G, Zhou J, Li X, Craven C, Gush C. The impact of automatic video auditing with real-time feedback on the quality and quantity of handwash events in a hospital setting. Am J Infect Control. 2020;48(2):162-6.

17. Xu Q, Liu Y, Cepulis D, Jerde A, Sheppard RA, Reichle W, et al. Hand hygiene behaviours monitored by an electronic system in the intensive care unit - a prospective observational study. J Hosp Infect. 2022;123:126-34.

18. Tseng C, He X, Chen W, Tseng CL. Evaluating automatic hand hygiene monitoring systems: a scoping review. Public Health Pract (Oxf). 2022;4:100290.

19. Mouajou V, Adams K, DeLisle G, Quach C. Hand hygiene compliance in the prevention of hospital-acquired infections: a systematic review. J Hosp Infect. 2022;119:33-48.

20. Scardoni A, Balzarini F, Signorelli C, Cabitza F, Odone A. Artificial intelligence-based tools to control healthcare associated infections: a systematic review of the literature. J Infect Public Health. 2020;13(8):1061-77.

21. Jordan MI, Mitchell TM. Machine learning: trends, perspectives, and prospects. Science. 2015;349(6245):255-60.

22. Badillo S, Banfai B, Birzele F, Davydov II, Hutchinson L, Kam-Thong T, et al. An introduction to machine learning. Clin Pharmacol Ther. 2020;107(4):871-85.

23. Saberi-Karimian M, Khorasanchi Z, Ghazizadeh H, Tayefi M, Saffar S, Ferns GA, et al. Potential value and impact of data mining and machine learning in clinical diagnostics. Crit Rev Clin Lab Sci. 2021;58(4):275-96.

24. An Q, Rahman S, Zhou J, Kang JJ. A comprehensive review on machine learning in healthcare industry: classification, restrictions, opportunities and challenges. Sensors (Basel). 2023;23(9):4178.

25. Habehh H, Gohel S. Machine learning in healthcare. Curr Genomics. 2021;22(4):291-300.

26. Oliveira Filho MA, Marra AR, Magnus TP, Rodrigues RD, Prado M, de Souza Santini TR, et al. Comparison of human and electronic observation for the measurement of compliance with hand hygiene. Am J Infect Control. 2014;42(11):1188-92.

27. Casaroto E, Generoso JR, Serpa Neto A, Prado M, Gagliardi GM, De Menezes FG, et al. Comparing human to electronic observers to monitor hand hygiene compliance in an intensive care unit. ASHE. 2022;2(1):e161.

28. Horan TC, Andrus M, Dudeck MA. CDC/NHSN surveillance definition of health care-associated infection and criteria for specific types of infections in the acute care setting. Am J Infect Control. 2008;36(5):309–32.

29. Kamran Siddiqui. Heuristics for sample size determination in multivariate statistical techniques. World Appl Sci J. 2013;27(2):285–7.

30. Stasinopoulos MD, Rigby RA, Heller GZ, Voudouris V, Bastiani FD. Flexible regression and smoothing: using GAMLSS in R. New York: Chapman and Hall/CRC; 2017. p. 571 [cited 2024 Jul 24]. Available from: https://www.taylorfrancis.com/books/9781351980388.

31. Lunardon N, Menardi G, Torelli N. ROSE: a package for binary imbalanced learning. The R Journal. 2014;6(1):79-89.

32. Jarek T, Dietze M. Tools: moving window statistics, GIF, Base64, ROC AUC, etc. 2022 [cited 2024 Jul 10]. Available from: https://cran.r-project.org/web/packages/caTools/caTools.pdf.

33. Hagel S, Reischke J, Kesselmeier M, Winning J, Gastmeier P, Brunkhorst FM, et al. Quantifying the hawthorne effect in hand hygiene compliance through comparing direct observation with automated hand hygiene monitoring. Infect Control Hosp Epidemiol. 2015;36(8):957–62.

34. Purssell E, Drey N, Chudleigh J, Creedon S, Gould DJ. The Hawthorne effect on adherence to hand hygiene in patient care. J Hosp Infect. 2020;106(2):311-317.

35. Sophonsri A, Lou M, Ny P, Minejima E, Nieberg P, Wong-Beringer A. Machine learning to identify risk factors associated with the development of ventilated hospital-acquired pneumonia and mortality: implications for antibiotic therapy selection. Front Med (Lausanne). 2023;10:1268488.

36. Sufriyana H, Chen C, Chiu HS, Sumazin P, Yang PY, Kang JH, et al. Estimating individual risk of catheter-associated urinary tract infections using explainable artificial intelligence on clinical data. Am J Infect Control. 2025;53(3):368-374.

37. Liang Y, Zhu C, Tian C, Lin Q, Li Z, Li Z, et al. Early prediction of ventilator-associated pneumonia in critical care patients: a machine learning model. BMC Pulm Med. 2022;22(1):250.

38. Piaggio D, Zarro M, Pagliara S, Andellini M, Almuhini A, Maccaro A, et al. The use of smart environments and robots for infection prevention control: a systematic literature review. Am J Infect Control. 2023;51(10):1175-1181.

39. McCalla S, Reilly M, Thomas R, McSpedon-Rai D, McMahon LA, Palumbo M. An automated hand hygiene compliance system is associated with decreased rates of health care-associated infections. Am J Infect Control. 2018;46(12):1381-86.

40. Akkoc G, Soysal A, Gul F, Kepenekli Kadayifci E, Arslantas MK, Yakut N, et al. Reduction of nosocomial infections in the intensive care unit using an electronic hand hygiene compliance monitoring system. J Infect Dev Ctries. 2021;15(12):1923–8.

41. van Gils RH, Kornelisse RF, Dankelman J, Helder OK. Validation of a hand hygiene visual feedback system to improve compliance with drying time of alcohol-based hand rub in a neonatal intensive care unit: the incubator traffic light system. J J Hosp Infect. 2024 Mar;145:210-17.

42. Armellino D, Hussain E, Schilling ME, Senicola W, Eichorn A, Dlugacz Y, et al. Using high-technology to enforce low-technology safety measures: the use of third-party remote video auditing and real-time feedback in healthcare. Clin Infect Dis. 2012;54(1):1-7.

43. Escobar GJ, Baker JM, Kipnis P, Greene JD, Mast TC, Gupta SB, et al. Prediction of Recurrent Clostridium Difficile Infection Using Comprehensive Electronic Medical Records in an Integrated Healthcare Delivery System. Infect Control Hosp Epidemiol. 2017;38(10):1196-203.

44. Oh J, Makar M, Fusco C, McCaffrey R, Rao K, Ryan EE, et al. A generalizable, data-driven approach to predict daily risk of Clostridium difficile infection at two large academic health centers. Infect Control Hosp Epidemiol. 2018;39(4):425-33.

45. Hu Z, Simon GJ, Arsoniadis EG, Wang Y, Kwaan MR, Melton GB. Automated detection of postoperative surgical site infections using supervised methods with electronic health record data. Stud Health Technol Inform. 2015;216:706–10.

46. Ke C, Jin Y, Evans H, Lober B, Qian X, Liu J, et al. Prognostics of surgical site infections using dynamic health data. J Biomed Inform. 2017;65:22-33.

47. Sohn S, Larson DW, Habermann EB, Naessens JM, Alabbad JY, Liu H. Detection of clinically important colorectal surgical site infection using Bayesian network. J Surg Res. 2017;209:168-173.

48. Weller GB, Lovely J, Larson DW, Earnshaw BA, Huebner M. Leveraging electronic health records for predictive modeling of post-surgical complications. Stat Methods Med Res. 2018;27(11):3271-85.

49. Campillo-Gimenez B, Garcelon N, Jarno P, Chapplain JM, Cuggia M. Full-text automated detection of surgical site infections secondary to neurosurgery in Rennes, France. Stud Health Technol Inform. 2013;192:572–5.

50. Kuo PJ, Wu SC, Chien PC, Chang SS, Rau CS, Tai HL, et al. Artificial neural network approach to predict surgical site infection after free-flap reconstruction in patients receiving surgery for head and neck cancer. Oncotarget. 2018;9(17):13768–82.

51. Soguero-Ruiz C, Fei WME, Jenssen R, Augestad KM, Álvarez JLR, Jiménez IM, et al. Data-driven temporal prediction of surgical site infection. AMIA Annu Symp Proc. 2015;2015:1164–73.

52. Sanger PC, Van Ramshorst GH, Mercan E, Huang S, Hartzler AL, Armstrong CAL, et al. A prognostic model of surgical site infection using daily clinical wound assessment. Am Coll Surg. 2016;223(2):259-270.e2.

53. Beeler C, Dbeibo L, Kelley K, Thatcher L, Webb D, Bah A, et al. Assessing patient risk of central line-associated bacteremia via machine learning. Am J Infect Control. 2018;46(9):986-91.

54. Parreco JP, Hidalgo AE, Badilla AD, Ilyas O, Rattan R. Predicting central line-associated bloodstream infections and mortality using supervised machine learning. J Crit Care. 2018;45:156-162.

55. Savin I, Ershova K, Kurdyumova N, Ershova O, Khomenko O, Danilov G, et al. Healthcare-associated ventriculitis and meningitis in a neuro-ICU: Incidence and risk factors selected by machine learning approach. J Crit Care. 2018;45:95-104.

56. Haas JP, Mendonça EA, Ross B, Friedman C, Larson E. Use of computerized surveillance to detect nosocomial pneumonia in neonatal intensive care unit patients. Am J Infect Control. 2005t;33(8):439-43.

57. Branch-Elliman W, Strymish J, Kudesia V, Rosen AK, Gupta K. Natural language processing for real-time catheter-associated urinary tract infection surveillance: results of a pilot implementation Trial. Infect Control Hosp Epidemiol. 2015;36(9):1004-10.

58. Taylor RA, Moore CL, Cheung KH, Brandt C. Predicting urinary tract infections in the emergency department with machine learning. PLoS ONE. 2018;13(3):e0194085.

59. Desautels T, Calvert J, Hoffman J, Jay M, Kerem Y, Shieh L, et al. Prediction of sepsis in the intensive care unit with minimal electronic health record data: a machine learning approach. JMIR Med Inform. 2016;4(3):e28.

60. Shimabukuro DW, Barton CW, Feldman MD, Mataraso SJ, Das R. Effect of a machine learning-based severe sepsis prediction algorithm on patient survival and hospital length of stay: a randomised clinical trial. BMJ Open Resp Res. 2017;4(1):e000234.

61. Cohen G, Hilario M, Sax H, Hugonnet S, Pellegrini C, Geissbuhler A. An application of one-class support vector machine to nosocomial infection detection. Stud Health Technol Inform. 2004;107(Pt 1):716–20.

62. Cohen G, Hilario M, Sax H, Hugonnet S, Geissbuhler A. Learning from imbalanced data in surveillance of nosocomial infection. Artif Intell Med. 2006;37(1):7-18.

63. Cohen G, Sax H, Geissbuhler A. Novelty detection using one-class Parzen density estimator. An application to surveillance of nosocomial infections. Stud Health Technol Inform. 2008;136:21–6.

64. Chang YJ, Yeh ML, Li YC, Hsu CY, Lin CC, Hsu MS, et al. Predicting Hospital-Acquired Infections by Scoring System with Simple Parameters. Gravenor MB, organizador. PLoS ONE. 24 de agosto de 2011;6(8):e23137.

65. Gerbier S, Yarovaya O, Gicquel Q, Millet AL, Smaldore V, Pagliaroli V, et al. Evaluation of natural language processing from emergency department computerized medical records for intra-hospital syndromic surveillance. BMC Med Inform Decis Mak. 2011;11(1):50.

66. Chen J, Pan QS, Hong WD, Pan J, Zhang WH, Xu G, et al. Use of an Artificial Neural Network to Predict Risk Factors of Nosocomial Infection in Lung Cancer Patients. Asian Pac J Cancer Prev. 2014;15(13):5349-53.

67. Gómez-Vallejo HJ, Uriel-Latorre B, Sande-Meijide M, Villamarín-Bello B, Pavón R, Fdez-Riverola F, et al. A case-based reasoning system for aiding detection and classification of nosocomial infections. Decis Support Syst. 2016;84:104–16.

68. Ehrentraut C, Ekholm M, Tanushi H, Tiedemann J, Dalianis H. Detecting hospital-acquired infections: a document classification approach using support vector machines and gradient tree boosting. Health Informatics J. 2018;24(1):24–42.

69. Surjan G. Questions on validity of International Classification of Diseases-coded diagnoses. Int J Med Inform. 1999;54(2):77–95.

Abstract

Introduction: Hospital-acquired infections increase morbidity, mortality, and healthcare costs. Effective hand hygiene is crucial for prevention, but achieving high compliance remains challenge. **Purpose:** Explore using machine learning to integrate an electronic hand hygiene auditing system with electronic health records to predict Hospital-acquired infections. **Methods:** A retrospective cohort study was conducted at a Brazilian hospital during 2017-2020. Hand hygiene compliance was recorded electronically, and patient data were collected from electronic health records. The primary outcomes were hospital-acquired infections per Centers for Disease Control and Prevention/National Healthcare Safety Network surveillance definitions. Machine learning algorithms, balanced with Random Over Sampling Examples, were utilized for predictive modeling, including generalized linear models; generalized additive models for location, scale, and shape; random forest; support vector machine; and extreme gradient boosting. **Results:** 125 of 6,253 patients (2%) developed HAIs and 920,489 hand hygiene opportunities (49.3% compliance) were analyzed. A direct correlation between hand hygiene compliance and hospital-acquired infections was observed. The generalized linear models algorithm with Random Over Sampling Examples demonstrated superior performance, with 84.2% sensitivity, 82.9% specificity, and a 93% area under the curve. **Conclusion:** Integrating electronic hand hygiene auditing systems with electronic health records and using machine learning models can enhance infection control surveillance and predict patient outcomes. Further research is needed to validate these findings and integrate them into clinical practice.

Keywords: Hand hygiene; Electronic health record; Cross infection; Machine learning