UNIVERSIDADE FEDERAL DE UBERLÂNDIA
FACULDADE DE MEDICINA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIAS DA SAÚDE

READMISSÃO HOSPITALAR POTENCIALMENTE EVITÁVEL EM POPULAÇÃO PEDIÁTRICA: MODELOS DE PREDIÇÃO

NAYARA CRISTINA DA SILVA

UBERLÂNDIA
2023

NAYARA CRISTINA DA SILVA

READMISSÃO HOSPITALAR POTENCIALMENTE EVITÁVEL EM POPULAÇÃO PEDIÁTRICA: MODELOS DE PREDIÇÃO

Tese apresentada ao Programa de Pós-graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Doutora em Ciências da Saúde.

Área de concentração: Ciências da Saúde

Orientadora: Geórgia das Graças Pena

Coorientador: André Ricardo Backes

Coorientador: Marcelo Keese Albertini

UBERLÂNDIA
2023

Dados Internacionais de Catalogação na Publicação (CIP)
Sistema de Bibliotecas da UFU, MG, Brasil.

S586r
2023

Silva, Nayara Cristina da, 1990-
Readmissão hospitalar potencialmente evitável em população pediátrica [recurso eletrônico] : modelos de predição / Nayara Cristina da Silva. - 2023.

Orientadora: Geórgia das Graças Pena.
Coorientador: André Ricardo Backes.
Coorientador: Marcelo Keese Albertini.
Tese (Doutorado) - Universidade Federal de Uberlândia, Programa de Pós-Graduação em Ciências da Saúde.
Modo de acesso: Internet.
Disponível em: http://doi.org/10.14393/ufu.te.2023.8026
Inclui bibliografia.

1. Ciências médicas. I. Pena, Geórgia das Graças, 1980-, (Orient.). II. Backes, André Ricardo, 1981-, (Coorient.). III. Albertini, Marcelo Keese, 1984-, (Coorient.). IV. Universidade Federal de Uberlândia. Programa de Pós-Graduação em Ciências da Saúde. V. Título.

CDU: 61

André Carlos Francisco
Bibliotecário - CRB-6/3408

UNIVERSIDADE FEDERAL DE UBERLÂNDIA
Coordenação do Programa de Pós-Graduação em Ciências da Saúde
Av. Pará, 1720, Bloco 2H, Sala 11 - Bairro Umuarama, Uberlândia-MG, CEP 38400-902
Telefone: (34) 3225-8628 - www.ppcsa.famed.ufu.br - ppcsa@famed.ufu.br

ATA DE DEFESA - PÓS-GRADUAÇÃO

Programa de Pós-Graduação em:	Ciências da Saúde				
Defesa de:	Tese de Doutorado Nº 01/PPCSA				
Data:	24.03.2023	Hora de início:	08:00h	Hora de encerramento:	12:00h
Matrícula do Discente:	11913CSD016				
Nome do Discente:	Nayara Cristina da Silva				
Título do Trabalho:	Readmissão hospitalar potencialmente evitável em população pediátrica: modelos de predição.				
Área de concentração:	Ciências da Saúde				
Linha de pesquisa:	2: DIAGNÓSTICO, TRATAMENTO E PROGNÓSTICO DAS DOENÇAS E AGRAVOS À SAÚDE				
Projeto de Pesquisa de vinculação:	NUTRIÇÃO CLÍNICA				

Reuniu-se em web conferência pela plataforma Mconf-RNP, em conformidade com a PORTARIA Nº 36, DE 19 DE MARÇO DE 2020 da COORDENAÇÃO DE APERFEIÇOAMENTO DE PESSOAL DE NÍVEL SUPERIOR - CAPES, pela Universidade Federal de Uberlândia, a Banca Examinadora, designada pelo Colegiado do Programa de Pós-graduação em Ciências da Saúde, assim composta: Profs. Drs. Ricardo Araújo Rios (UFBA), Pedro Luiz Lima Bertarini (UFU), Anita Maria da Rocha Fernandes (UNIVALI), Barbara Perez Vogt (UFU) e Geórgia das Graças Pena (UFU), orientadora da candidata.

Iniciando os trabalhos, a presidente da mesa, Profa. Dra. Geórgia das Graças Pena, apresentou a Comissão Examinadora e a candidata, agradeceu a presença do público, e concedeu a Discente a palavra para a exposição do seu trabalho. A duração da apresentação da Discente e o tempo de arguição e resposta foram conforme as normas do Programa.

A seguir o senhor(a) presidente concedeu a palavra, pela ordem sucessivamente, aos(às) examinadores(as), que passaram a arguir o(a) candidato(a). Ultimada a arguição, que se desenvolveu dentro dos termos regimentais, a Banca, em sessão secreta, atribuiu o resultado final, considerando o(a) candidato(a):

Aprovada.

Esta defesa faz parte dos requisitos necessários à obtenção do título de Doutor.

O competente diploma será expedido após cumprimento dos demais requisitos, conforme as normas do Programa, a legislação pertinente e a regulamentação interna da UFU.

Nada mais havendo a tratar foram encerrados os trabalhos. Foi lavrada a presente ata que após lida e achada conforme foi assinada pela Banca Examinadora.

Documento assinado eletronicamente por **Anita Maria da Rocha Fernandes**, **Usuário Externo**, em 24/03/2023, às 10:27, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Pedro Luiz Lima Bertarini**, **Professor(a) do Magistério Superior**, em 24/03/2023, às 11:58, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Barbara Perez Vogt**, **Professor(a) do Magistério Superior**, em 24/03/2023, às 12:00, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Ricardo Araújo Rios**, **Usuário Externo**, em 24/03/2023, às 12:03, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

Documento assinado eletronicamente por **Geórgia das Graças Pena**, **Professor(a) do Magistério Superior**, em 24/03/2023, às 12:10, conforme horário oficial de Brasília, com fundamento no art. 6º, § 1º, do Decreto nº 8.539, de 8 de outubro de 2015.

A autenticidade deste documento pode ser conferida no site https://www.sei.ufu.br/sei/controlador_externo.php?acao=documento_conferir&id_orgao_acesso_externo=0, informando o código verificador **4357020** e o código CRC **834DFA07**.

Referência: Processo nº 23117.020590/2023-80 SEI nº 4357020

FOLHA DE APROVAÇÃO

Nayara Cristina da Silva

READMISSÃO HOSPITALAR POTENCIALMENTE EVITÁVEL EM POPULAÇÃO PEDIÁTRICA: MODELOS DE PREDIÇÃO

<u>**Presidente da banca (orientador): Profa. Drª. Geórgia das Graças Pena**</u>

Tese apresentada ao Programa de Pós- graduação em Ciências da Saúde da Faculdade de Medicina da Universidade Federal de Uberlândia, como requisito parcial para obtenção do título de Doutora em Ciências da Saúde.

Área de concentração: Ciências da Saúde

Banca Examinadora

Titular: Prof. Dr. Ricardo Araújo Rios
Instituição: Universidade Federal da Bahia

Titular: Prof. Drª. Anita Maria da Rocha Fernandes
Instituição: Universidade do Vale do Itajaí

Titular: Prof. Dr. Pedro Luiz Lima Bertarini
Instituição: Universidade Federal de Uberlândia

Titular: Profª. Drª. Barbara Perez Vogt
Instituição: Universidade Federal de Uberlândia

Suplente: Prof. Dr. Dalcimar Casanova
Instituição: Universidade Tecnológica Federal do Paraná

Suplente: Profª. Dr. Matheus de Souza Gomes
Instituição: Universidade Federal de Uberlândia

DEDICATÓRIA

À Deus, a minha família, amigos e a todos aqueles que fazem parte da minha vida e que compartilharam comigo cada passo desta caminhada.

AGRADECIMENTOS

A Deus, pela presença constante em minha vida, direcionando sempre minhas escolhas e por me impulsionar na direção dos meus sonhos.

Á minha família, minha mãe Magali e minhas mães do coração vó Ilda e tia Suely, que são minha maior inspiração e exemplo. Obrigada pelos ensinamentos, apoio, orações e amor incondicional a cada novo dia e por compreenderem minhas ausências em alguns momentos. Vocês são a minha base. Ao meu avô e pai Jesus Francisco "Vô Dola" (*in memoriam*), minha maior saudade, pelo exemplo de vida imagino sua vibração se estivesse aqui por mais essa vitória em minha vida. À minha irmã Naessa, minha incentivadora nos momentos de desânimo, me motivando a perseguir meus sonhos.

Ao meu esposo Rodrigo, obrigada pelo amor, paciência, companheirismo, compreensão e pelas palavras de conforto e incentivo nessa fase importante da minha vida.

À minha orientadora, professora Drª Geórgia das Graças Pena, pela receptividade, por confiar e acreditar no meu potencial, pelos conhecimentos e experiências de vida transmitidos ao longo destes anos. Agradeço por me ensinar a aprender, refletir, pesquisar e interrogar. Tenho muito orgulho de tê-la como orientadora e como uma das principais responsáveis pela minha formação profissional. Que eu possa sempre contar com o privilégio da sua amizade.

Aos meus coorientadores, professor Dr. Marcelo Keese Albertini e Dr. André Ricardo Backes, por terem me acolhido e ensinado tanto. Obrigada pela paciência, disponibilidade e cuidado ao longo destes anos.

Ao professor Dr. Laurence Rodrigues do Amaral, por disponibilizar seu tempo e seus conhecimentos de forma tão gentil, cuja contribuição foi fundamental para o bom desenvolvimento deste trabalho.

Aos professores Dr. Ricardo Araújo Rios e Dr. Laurence Amaral, que muito contribuíram no exame de qualificação, resultando num trabalho muito melhor.

Aos professores Dr. Ricardo Araújo Rios, Dr. Pedro Luiz Lima Bertarini, Drª. Anita Maria da Rocha Fernandes e Drª. Barbara Perez Vogt por terem aceitado participar como banca examinadora da defesa desta tese. Suas contribuições serão muito válidas.

Aos funcionários do Setor de Tecnologia da Informação da Universidade Federal de Uberlândia, em especial aos analistas Wander e Sara, por toda paciência e ajuda na extração dos dados.

Ao Programa de Pós-Graduação em Ciências da Saúde, em especial às técnicas administrativas Gisele de Melo Rodrigues e Viviane Gonçalves, por toda colaboração e suporte.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) agradeço o apoio financeiro por meio do edital APQ-01331-21 o qual viabilizou o desenvolvimento da pesquisa resultante nesta tese e nos produtos da mesma.

Às demais pessoas que contribuíram direta ou indiretamente na elaboração deste trabalho ou participaram da minha vida, e que, por ventura, eu tenha me esquecido de agradecer.

"Sonhos determinam o que você quer.
Ação determina o que você conquista."
Aldo Novak

"Sê forte e corajoso. Não temas, nem te espantes;
porque o Senhor teu Deus está contigo,
por onde quer que andares"

Bíblia Sagrada (Josué 1:9)

RESUMO

Introdução: Readmissões hospitalares potencialmente evitáveis (RHPE) são eventos complexos com impacto negativo tanto para o paciente como para o sistema de saúde. Na população pediátrica um episódio de readmissão pode ser ainda pior, podendo impactar no desenvolvimento da criança ou adolescente, afetando negativamente o desenvolvimento motor, cognitivo, emocional e psicossocial a curto, médio e principalmente em longo prazo. Assim, o desenvolvimento de modelos de predição, especialmente utilizando o aprendizado de máquina, tem sido promissor para minimizar este desfecho. Entretanto, ainda são escassos os estudos, especialmente os interpretáveis com potencial aplicação prática. **Objetivos:** Desenvolver modelos de predição de RHPE de 30 dias, para crianças e adolescentes, internados em hospital de nível terciário. **Métodos:** Foi realizada coorte retrospectiva com dados de todos pacientes pediátricos (0 a 18 anos) internados em hospital terciário entre janeiro 2014 a dezembro 2018 (n=9.080). Internações que resultaram em óbito, altas hospitalares contra orientação médica ou readmissões planejadas foram excluídas. Foram coletados dados demográficos, clínicos, nutricionais e exames bioquímicos. Para o primeiro manuscrito, estimou-se um modelo de predição baseado em um sistema de pontuação (escore HOSPITAL) e posteriormente, classificou-se os pacientes em grupos de baixo, intermediário e alto risco. No segundo manuscrito, utilizou-se o algoritmo J48 e aplicou validação cruzada *leave-one-out* para desenvolver árvores de decisão interpretáveis. Para o terceiro manuscrito, construiu-se modelos baseados em *machine learning* utilizando vários algoritmos: *classification and regression tree - CART*, *random forest – RF*, *gradient boosting machine – GBM*, *extreme gradient boosting - XGBoost, decision tree - DT* e *logistic regression - LR*. Para comparar o desempenho dos modelos, calculamos a áreas sob a curva (AUC, *area under the receiver operating curve*). Outras medidas de desempenho também foram consideradas, como sensibilidade, especificidade, índice de *Youden* e acurácia, quando pertinentes. **Resultados:** A frequência de RHPE em 30 dias variou de 9,5 a 11,70%. O escore HOSPITAL apresentou boa capacidade discriminatória (AUC 0,80 IC 95% 0,77-0,83) para RHPE em população pediátrica (Manuscrito 1). Para melhorar as estimativas utilizamos técnicas de *machine learning*, visando construir um modelo de predição, interpretável e de fácil aplicação na pratica clínica. A árvore de decisão (algoritmo J48, aplicada a 63,6% dos casos novos) demonstrou que alteração na proteína C reativa, de hemoglobina e de sódio e o não acompanhamento nutricional foram os atributos que contribuíram para a RHPE (Manuscrito 2). Por fim, o algoritmo XGBoost apresentou o melhor índice de *Youden*. Os preditores foram: diagnóstico de câncer, idade, níveis de hemácias, leucócitos, amplitude de distribuição dos glóbulos vermelhos e sódio, admissão eletiva e multimorbidade (Mansucrito 3). **Conclusão:** O escore HOSPITAL pode ser utilizado para população pediátrica, o XGBoost mostrou boa discriminação e, a árvore de decisão embora com menor potencial discriminatório para RHPE, identificou atributos importantes para a prática clínica, especialmente o não acompanhamento nutricional. Além disso, os três modelos de predição desenvolvidos oferecem vantagens uma vez que foram encontrados atributos clínicos que fazem parte da rotina de assistência, de fácil obtenção e baixo custo, permitindo a obtenção do resultado de probabilidade de readmissão em tempo real se implementada no sistema de internação hospitalar. Por fim, destacamos a importância de melhorar a qualidade e a especificidade dos registros de dados hospitalares, como por exemplo, os dados nutricionais para construir modelos de predição de readmissão hospitalar.

Palavras-chave: Aprendizado de Máquina, Readmissão Hospitalar, Regras de Predição Clínica, Pediatria, Fatores de Risco

ABSTRACT

Background: Potentially avoidable hospital readmissions (PAHR) are complex events with a negative impact on both the patient and the health system. In the pediatric population, a readmission episode can be even worse, and may impact the development of the child or adolescent, negatively affecting motor, cognitive, emotional and psychosocial development in the short, medium and especially long term. Thus, the development of prediction models, especially using machine learning, has been promising to minimize this outcome. However, studies are still scarce, especially those that can be interpreted with potencial practical application. **Objective:** To develop 30-day PAHR prediction models for children and adolescents admitted to a tertiary hospital. **Methods**: A retrospective cohort was performed with data from all pediatric patients (0 to 18 years old) admitted to a tertiary hospital between January 2014 and December 2018 (n: 9,080). Admissions that resulted in death, hospital discharges against medical advice or planned visits/readmissions were excluded. Demographic, clinical, nutritional and biochemical data were collected. For the first manuscript, a prediction model based on a scoring system (HOSPITAL score) was estimated and, subsequently, patients were classified into low, intermediate and high risk groups. In the second manuscript, we used the J48 algorithm and applied leave-one-out cross-validation to develop interpretable decision trees. For the third manuscript, models based on machine learning were built, several algorithms were investigated: classification and regression tree - CART, random forest – RF, gradient boosting machine – GBM, extreme gradient boosting - XGBoost, decision tree - DC e logistic regression - LR. To compare the performance of the models, we computed the area under the receiver operating curve (AUC). Other performance measures were also calculated, such as sensitivity, specificity, Youden's J index and accuracy, when relevant. **Results**: The frequency of PAHR in 30 days ranged from 9.5 to 11.70%. The HOSPITAL score showed good discriminatory ability (AUC of 0.80 95% CI 0.77-0.83) for PAHR in a pediatric population (Manuscript 1). To improve the estimates, we used machine learning techniques, aiming to build a predictive model, interpretable and easy to apply in clinical practice. The decision tree (J48 algorithm, applied to 63.6% of new cases) showed that changes in C-reactive protein, hemoglobin and sodium levels and lack of nutritional monitoring were the attributes that contributed to PAHR (Manuscript 2). Finally, the XGBoost algorithm presented the best Youden's J index. The predictors (XGBoost) were: cancer diagnosis, age, levels of red blood cells, leukocytes, red cell distribution width and sodium, elective admission and multimorbidity (Manuscript 3). **Conclusion:** The HOSPITAL score can be used for the pediatric population, the XGBoost showed good discrimination and the decision tree, although with less discriminatory potential for PAHR, identified important attributes for clinical practice, especially the lack of nutritional monitoring. Besides, the three prediction models developed offer advantages, since clinical attributes were found that are part of the care routine, easy to get and low cost, allowing to obtain the readmission probability result in real time if implemented in the admission system hospital. Finally, we highlight the importance of improving the quality and specificity of hospital data records, such as nutritional data, to build hospital readmission prediction models.

Keywords: Machine Learning, Hospital Readmission, Clinical Decision Rules, Pediatrics; Risk factors

LISTA DE ILUSTRAÇÕES

LISTA DE TABELAS

Manuscrito 2

Manuscrito 3

LISTA DE ABREVIATURAS E SÍMBOLOS

AUC	*Area under the receiver operating curve*
IA	Inteligência Artificial
CART	*Classification and regression tree*
CEP	Comitê de ética em Pesquisa
DT	*Decision tree*
GBM	*Gradient boosting machine*
LR	Logistic regression
LOOCV	Validação cruzada *leave-one-out*
ML	*Machine learning*
PAHR	*Potentially avoidable hospital readmissions*
RHPE	Readmissões hospitalares potencialmente evitáveis
RF	*Random forest*
SIH	Sistema de Internação Hospitalar
SUS	Sistema Único de Saúde
TIH	Tempo de Internação Hospitalar
XGBoost	*Extreme gradient boosting*

SUMÁRIO

APRESENTAÇÃO

Trata-se de uma Tese de Doutorado, estruturada no formato alternativo aprovado pelo Colegiado do Programa de Pós-graduação em Ciências da Saúde (PPCSA) da Universidade Federal de Uberlândia (UFU). Este formato autoriza a apresentação dos artigos científicos resultantes do projeto de pesquisa. Portanto, a tese está organizada nas seguintes seções: (1) elementos pré-textuais (capa, folha de rosto, ficha catalográfica, folha de aprovação, dedicatória, agradecimentos, resumo na língua portuguesa, *Abstract*, Lista de Quadros e Tabelas, Lista de Abreviatura e Siglas, Sumário), (2) Introdução que destaca a importância do estudo; (2) Fundamentação teórica que norteia as hipóteses do presente trabalho; (3) Objetivos onde são expostos os propósitos do estudo; (4) Resultados que contemplam os artigos elaborados; (5) Considerações finais que discorre sobre a síntese dos principais resultados do estudo (6) Perspectivas futuras destacando os principais pontos que podem ser abordados nos estudos futuros.

O primeiro manuscrito é um estudo original intitulado "**Validation of the HOSPITAL score as predictor of 30-day potentially avoidable readmissions in pediatric hospitalized population: retrospective cohort study**" que teve como objetivo validar o uso de um escore conhecido na população adulta (HOSPITAL *score*) para readmissões potencialmente evitáveis em 30 dias em população pediátrica. Este manuscrito está publicado na revista *European Journal of Pediatrics* (fator de impacto na *Journal Citaton Reports/Clarivate*: 3.860).

O segundo manuscrito é um estudo original intitulado "**Decision tree model predicts avoidable hospital 30 days readmissions in pediatric population**" que teve como objetivo construir um modelo preditivo interpretável usando um algoritmo de árvore de decisão considerando os fatores clínicos e o acompanhamento nutricional a fim de identificar com risco de readmissão hospitalar potencialmente evitável em 30 dias. Este manuscrito encontra-se em revisão na revista *Nutrition* (fator de impacto na *Journal Citaton Reports/Clarivate*: 4.893).

O terceiro manuscrito é um estudo original intitulado " **Machine learning for hospital readmission prediction in pediatric population**" que teve como objetivo desenvolver modelos de predição, para identificar população pediátrica com alto risco de readmissão potencialmente evitável em 30 dias usando aprendizado de máquina. Pretende-se a publicação na revista *Pediatrics* (fator de impacto na *Journal Citaton Reports/Clarivate*: 7.124).

1. INTRODUÇÃO

A readmissão hospitalar é um desfecho de importante impacto em saúde, tanto para o indivíduo quanto para o sistema de saúde. A taxa de readmissão hospitalar potencialmente evitáveis (RHPE) em 30 dias entre crianças e adolescentes ainda é elevada, variando de 4,4% a 29,5% (EHWERHEMUEPHA et al., 2018a, 2020b; TAYLOR; ALTARES SARIK; SALYAKINA, 2020; TOOMEY et al., 2016). Entre os principais fatores de risco, comumente associados à readmissão hospitalar pediátrica, estão o tempo de internação prolongado (EHWERHEMUEPHA et al., 2018a; FEUDTNER et al., 2009; ZHOU et al., 2019a), a idade (FEUDTNER et al., 2009; ZHOU et al., 2019a) e a multimorbidade (ZHOU et al., 2019b). Estudos revelam ainda que a readmissão hospitalar pode influenciar negativamente na qualidade de vida dos pacientes (COYNE, 2006; SILVA et al., 2018) com consequências em curto e longo prazos (CAHAYAG, 2020; DELVECCHIO et al., 2019; PUFAL et al., 2018; RASHIKJ CANEVSKA, 2018; SILVA et al., 2018), além de contribuir substancialmente para o aumento dos custos em saúde (GAY et al., 2015; KANE et al., 2021; MARKHAM et al., 2018).

O primeiro passo para diminuir as taxas de readmissão hospitalar pediátrica e minimizar seus efeitos deletérios é a identificação dos pacientes com maior risco de readmitir. Essa ação permite o direcionamento de intervenções preventivas para pacientes com maior necessidade, otimizando o atendimento e a alocação de recursos financeiros. A avaliação do risco de readmissão hospitalar foi abordada sob diferentes perspectivas e ferramentas preditivas foram desenvolvidas e validadas anteriormente, na tentativa de identificar precocemente indivíduos com alto risco de readmissão (DONZE et al., 2013; EHWERHEMUEPHA et al., 2020b; VAN WALRAVEN et al., 2010; WOLFF et al., 2019). A regressão logística tem sido uma das técnicas mais utilizadas nos estudos observacionais para construir modelos preditivos, incluindo os escores HOSPITAL (DONZE et al., 2013), LACE (HAN; FLUCK; FRY, 2021; VAN WALRAVEN et al., 2010) e PREDAM (SHADMI et al., 2015). Embora essas abordagens convencionais pareçam úteis, elas se baseiam principalmente em um conjunto fechado de variáveis clínicas e são direcionadas a populações específicas com escopo limitado e desempenho inconsistente (WEINREICH et al., 2016). Além disso, essa abordagem não é capaz de retratar a complexidade do processo e a dinâmica não estacionária das readmissões hospitalares, que podem mudar ao longo do tempo dependendo de diferentes tratamentos e condições durante as internações (TURGEMAN; MAY, 2016). Portanto, muitos fatores podem ser subestimados ou realmente não percebidos durante a evolução do paciente, uma vez que, podem escapar das percepções humanas da equipe de saúde.

Diante disso, recentemente, estudos têm sugerido o emprego de técnicas de *machine learning* (ML), para desenvolver ferramentas preditivas de readmissões hospitalares com potencial melhoria da capacidade preditiva, em comparação com a abordagens estatísticas tradicionais (ASHFAQ et al., 2019; BAIG et al., 2019; FUTOMA; MORRIS; LUCAS, 2015; MORGAN et al., 2019). Todavia, até o momento poucos estudos empregaram técnicas de ML no cenário clínico hospitalar como ferramenta de predição de readmissão hospitalar especialmente na população pediátrica (EHWERHEMUEPHA et al., 2020b; SYMUM; ZAYAS-CASTRO, 2021; TAYLOR; ALTARES SARIK; SALYAKINA, 2020; WOLFF et al., 2019; ZHOU et al., 2021). Quando as empregam, geram modelos de difícil interpretação por não deixarem claro quais são os preditores e a direção de sua associação, o que dificulta a aplicação desses modelos na prática clínica. Portanto, ainda são necessários estudos que empreguem técnicas de ML para desenvolver modelos de predição de RHPE, que sejam facilmente interpretáveis e implementados na prática clínica.

Como é do conhecimento geral da sociedade, os recursos em saúde são limitados e desfechos hospitalares como as RHPE, além de terem impacto negativo para o paciente também oneram o sistema de saúde. Assim, a busca por ferramentas capazes de identificar os pacientes com maior probabilidade de readmissão tem motivado estudos e seguem sendo um desafio para as pesquisas científicas. A prioridade é desenvolver ferramentas de fácil implementação e compreensão, capazes de predizer a readmissão com boa capacidade de predição. Para os serviços de saúde é imprescindível garantir a interpretabilidade dos modelos gerados, de forma que os profissionais de saúde compreendam de forma racional as relações entre as variáveis dependentes e independentes. Além disso, ferramentas que incluam preditores relevantes para a população pediátrica, bem como acrescente um olhar para os dados nutricionais, que podem contribuir com diversos desfechos em saúde e que são comumente negligenciados por muitos estudos. Portanto, o desenvolvimento de ferramentas de predição para RHPE, que sirvam de apoio às decisões clínicas e otimizem a ordenação do fluxo de trabalho tornando os processos mais ágeis e eficientes, ainda são necessárias.

2. REFERÊNCIAL TEÓRICO

2.1 Readmissão hospitalar pediátrica: impactos negativos para o paciente, para a família e para o serviço de saúde

A readmissão hospitalar é definida como uma internação dentro de um curto período de tempo pré-especificado após a alta hospitalar. Um período padrão internacionalmente utilizado para considerar como retorno precoce de um paciente é de 30 dias (KANSAGARA et al., 2011). As readmissões podem ou não estarem clinicamente relacionadas com a internação hospitalar inicial. As readmissões relacionadas com uma condição prévia são as mais relevantes, uma vez que podem refletir a qualidade do serviço e ocasionar aumento de custos hospitalares. No presente estudo, exploramos a readmissão hospitalar potencialmente evitável, definida como uma nova internação em período igual ou inferior a 30 dias, relacionada à condição médica previamente codificada ou decorrente de complicação do tratamento. Para tanto, excluímos todas os pacientes internados em enfermarias com internações planejadas (obstetrícia, ginecologia, transplante e cirurgia eletiva) e pacientes que faleceram antes da alta.

A readmissão hospitalar pediátrica é um desfecho complexo e multifatorial, que ainda permanece elevado, com taxas que variam de 4,0% a 29,5% (EHWERHEMUEPHA et al., 2018a, 2020b; TAYLOR; ALTARES SARIK; SALYAKINA, 2020; TOOMEY et al., 2016). Essa variação ocorre devido a diferentes tipos de planejamento do estudo, métodos de cálculo, variabilidade nos tipos de serviços e sistemas de saúde específicos. Além disso, a readmissão está associada a impactos negativos para a saúde do paciente, tais como prejuízos no desenvolvimento da criança (CAHAYAG, 2020; DELVECCHIO et al., 2019; PUFAL et al., 2018; RASHIKJ CANEVSKA, 2018; SILVA et al., 2018), aumento substancial dos custos (GAY et al., 2015; KANE et al., 2021; MARKHAM et al., 2018), do tempo internação (FEUDTNER et al., 2009; MARKHAM et al., 2018) e da mortalidade hospitalar (HARTMAN et al., 2017) como também gerando impactos em sua família (MUMFORD et al., 2018).

A hospitalização é um evento estressante e traumático na vida de um indivíduo, implicando no rompimento das relações sociais e familiares. Para crianças e adolescentes, a hospitalização pode ter um impacto mais substancial por ter mecanismos emocionais ainda limitados para lidar com essa experiência (COYNE, 2006; SILVA et al., 2018). Nesse sentido, além da internação ser um evento negativo em potencial, as readmissões hospitalares nessa fase da vida podem ser piores. A internação pediátrica pode impactar no desenvolvimento da criança, afetando negativamente o desenvolvimento motor, cognitivo, emocional e psicossocial em curto, médio e principalmente em longo prazo (CAHAYAG, 2020; DELVECCHIO et al., 2019; PUFAL et al., 2018; RASHIKJ CANEVSKA, 2018; SILVA et al., 2018). Cabe ressaltar que a internação pediátrica traz efeitos negativos não apenas para os pacientes, mas também para a rede familiar envolvida, que muitas vezes precisa conciliar os cuidados de saúde com a rotina laboral e familiar (MUMFORD et al., 2018), fortalecendo ainda mais a importância de desenvolvimento de estudos nesse contexto.

Além disso, as readmissões hospitalares pediátricas podem sobrecarregar os serviços de saúde, estando associadas com aumento substancial dos custos (GAY et al., 2015; KANE et al., 2021; MARKHAM et al., 2018) e do tempo de internação (FEUDTNER et al., 2009; MARKHAM et al., 2018). Um estudo avaliando readmissões hospitalares de 30 dias revelou que crianças que readmitiram passaram o dobro de dias no hospital (considerando a admissão anterior e atual) em comparação com crianças que não readmitiram e tiveram custos hospitalares aproximadamente 2,3 vezes maiores (MARKHAM et al., 2018).

2.2 Fatores de risco associados à readmissão hospitalar

Por se tratar de um desfecho complexo e multifatorial, muitos fatores podem estar associados à readmissão hospitalar pediátrica. Os principais fatores de risco apontados pelos estudos são a multimorbidade (EHWERHEMUEPHA et al., 2018a; NIEHAUS et al., 2022; ZHOU et al., 2019b, 2021), o maior número de internações anteriores (EHWERHEMUEPHA et al., 2018a; ZHOU et al., 2019b, 2021), o tempo de internação prolongado (EHWERHEMUEPHA et al., 2018a; FEUDTNER et al., 2009; NIEHAUS et al., 2022; ZHOU et al., 2019b), a idade (EHWERHEMUEPHA et al., 2018a; FEUDTNER et al., 2009; ZHOU et al., 2019b), polifarmácia (EHWERHEMUEPHA et al., 2020a, 2020b) e a presença de diagnósticos ou condições como anemia, desnutrição e câncer (EHWERHEMUEPHA et al., 2018b; KUMAR et al., 2019).

O tempo de internação hospitalar (TIH), por sua vez, é definido como o número de dias que um paciente permanece no hospital. É comumente utilizado como um indicador de gravidade da doença, bem como, da eficiência hospitalar, impactando no consumo de recursos hospitalares. Além disso, o TIH tem sido apontado como um indicador da qualidade do cuidado prestado (AWAD; BADER–EL–DEN; MCNICHOLAS, 2017; HUANG et al., 2013). O TIH é influenciado por inúmeros fatores ambientais, tais como diagnóstico médico, gravidade da doença, desnutrição, idade, entre outros (CORREIA; WAITZBERG, 2003; PIRLICH et al., 2006; REZENDE et al., 2004). Além disso, um maior TIH, acarreta em redução da rotatividade dos leitos hospitalares, com consequente elevação dos custos, onerando o Sistema Único de Saúde (SUS), que trabalha muitas vezes próximo ao limite dos seus recursos financeiros (WAITZBERG; CAIAFFA; CORREIA, 2001).

Estudos têm revelado que as taxas de readmissão pediátrica variam de acordo com a idade (EHWERHEMUEPHA et al., 2018a; FEUDTNER et al., 2009; ZHOU et al., 2019b). Um estudo que avaliou os fatores de risco associados à readmissão hospitalar não planejada de 30 dias por todas as causas em um hospital infantil terciário revelou que pacientes com mais de 13 anos tinham quase 1,5 vez mais chance de serem readmitidos em comparação com pacientes mais jovens. Segundo os autores, isso pode estar relacionado aos adolescentes e adultos jovens com condições de saúde crônicas subjacentes (ZHOU et al., 2020). Em outras palavras, à medida que a idade avança, fatores de risco cumulativos podem aumentar a frequência de doenças crônicas ou agravos à saúde, como drogadição, a hipertensão, obesidade, entre outros, que contribuiriam mais intensamente para complicações clínicas e necessidade de internações hospitalares. Uma revisão sistemática, que visou identificar os fatores de risco associados às readmissões hospitalares pediátricas não planejadas, apontou que a idade inferior a 12 meses ou entre 13 e 18 anos é um dos fatores de risco para readmissão hospitalar (ZHOU et al., 2016). Outro estudo que avaliou readmissões pediátricas não planejadas nos Estados Unidos mostrou que as taxas de readmissão entre bebês foi 2,1%, triplicando para 6,7% após a primeira infância. Segundo os autores, a prevalência de crianças com condições crônicas está aumentando devido à melhora da sobrevida no período neonatal e ao avanço no cuidado, levando a necessidades médicas adicionais e internações no futuro (AMRITPHALE et al., 2021).

A multimorbidade, mais frequentemente definida como a "coocorrência de duas ou mais doenças crônicas ou agudas e condições médicas em uma pessoa" (VAN DEN AKKER; BUNTINX; KNOTTNERUS, 1996), é frequente nos ambientes hospitalares. Especula-se que a multimorbidade pode ser *proxy* de gravidade do paciente estando relacionada com desfechos hospitalares negativos e aumento dos custos hospitalares. Diversos estudos avaliando readmissão hospitalar pediátrica mostraram que a multimorbidade é um dos fatores de risco para a readmissão (EHWERHEMUEPHA et al., 2018a; NIEHAUS et al., 2022; ZHOU et al., 2019b, 2021). Uma coorte retrospectiva multinacional com pacientes hospitalizados, encontrou associação forte e linear de readmissões potencialmente evitáveis em 30 dias com a multimorbidade. Os autores relatam que ter pelo menos quatro sistemas corporais envolvidos ou nove doenças crônicas dobra o risco de readmissões potencialmente evitáveis em 30 dias (AUBERT et al., 2019).

A presença de diagnóstico de câncer, por exemplo, tem sido apontado por alguns estudos, como fator de risco para readmissões hospitalares (EHWERHEMUEPHA et al., 2018a, 2020a, 2020b; HOENK et al., 2021). O câncer é uma doença complexa, de etiologia multifatorial, cujo tratamento pode acarretar em múltiplas internações, decorrente de complicações infecciosas, anormalidades hidroeletrolíticas e diminuição do estado funcional (DONZE et al., 2013; HOENK et al., 2021). A anemia, definida pelos níveis de hemoglobina diminuídos, também se associa negativamente com a readmissão hospitalar na primeira infância (KUMAR et al., 2019), bem como ocasiona prejuízos na função cognitiva, no crescimento e no desenvolvimento psicomotor (AISHVARYA et al., 2021). A anemia é um problema de saúde pública, comum em crianças hospitalizadas, de etiologia multifatorial, que envolve uma interação complexa entre nutrição e outros fatores clínicos como: perda de sangue, redução da expectativa de vida dos eritrócitos, inflamação e eritropoiese e/ou hemodiluição insuficientes (JUTRAS et al., 2020). A presença de anemia está significativamente associada à readmissão hospitalar pediátrica (KUMAR et al., 2019).

Por fim, outro importante fator de risco, muitas vezes negligenciado pelos estudos, é o estado nutricional. A perda de componentes corporais anteriores à admissão ou o declínio do estado nutricional durante a hospitalização afetam adversamente os resultados clínicos (CORREIA; PERMAN; WAITZBERG, 2017). Como bem reconhecido na literatura, as crianças possuem demandas nutricionais de energia e proteína aumentadas para o seu crescimento e desenvolvimento considerando a relação ponderal quando comparadas aos adultos (KYLE; PICHARD, 2006; MCCARTHY et al., 2019), o que as tornam mais vulneráveis aos efeitos deletérios de um estado nutricional prejudicado. Além disso, a desnutrição infantil pode, em longo prazo, impactar o crescimento e a trajetória de desenvolvimento cognitivo (MCCARTHY et al., 2019). A desnutrição correlaciona-se com inúmeros fatores, anteriores e posteriores à hospitalização, além de estar associada à doença subjacente ou ao seu tratamento. Um estudo que avaliou o efeito da desnutrição no risco de readmissão não planejada de sete dias em pediatria demonstrou que pacientes mais jovens e desnutridos possuíam maiores chances de readmissão do que seus pares com mais idade, revelando uma interação entre idade e estado nutricional (EHWERHEMUEPHA et al., 2018b)

Apesar de alguns fatores terem sido apontados por estudos anteriores, muitos fatores de risco podem ser subestimados ou até mesmo não serem reconhecidos pela equipe de saúde durante a internação. Por isso, continua sendo importante os esforços na identificação dos fatores que interferem na readmissão hospitalar pediátrica.

2.3 Técnicas para predição de readmissão hospitalar pediátrica

Identificar precocemente os indivíduos com maior probabilidade de readmissão é uma das estratégias para reduzir suas taxas em nível hospitalar, permitindo direcionar o cuidado para os pacientes com maior risco, otimizando o atendimento e consequentemente, reduzindo os custos hospitalares. A readmissão hospitalar pediátrica é um processo complexo e não estacionário, podendo ser afetado inclusive por fatores anteriores que vão para além da internação.

Nos últimos anos foram desenvolvidas ferramentas capazes de identificar indivíduos com alto risco de readmissão utilizando estatísticas convencionais (DONZE et al., 2013; HAN; FLUCK; FRY, 2021; SHADMI et al., 2015; VAN WALRAVEN et al., 2010) e mais recentemente empregando técnicas de ML (EHWERHEMUEPHA et al., 2020b; SYMUM; ZAYAS-CASTRO, 2021; WOLFF et al., 2019; ZHOU et al., 2021). Entretanto, os modelos de predição encontrados na literatura (Quadro 1), seja utilizando abordagens estatísticas tradicionais ou técnicas de ML, possuem algumas limitações, o que dificulta sua aplicação na prática clínica. Alguns modelos apresentam uma discriminação baixa ou moderada (WOLFF et al., 2019; ZHOU et al., 2021), excluem subgrupos pediátricos como os neonatos (EHWERHEMUEPHA et al., 2020b), utilizam preditores derivados para a população adulta, tais como o *Charlson Comorbidity Index* (EHWERHEMUEPHA et al., 2018a, 2020a, 2020b; HAN; FLUCK; FRY, 2021), utilizam métricas diferentes para avaliar o tempo (7, 10, 28 e 30 dias) de readmissão hospitalar (EHWERHEMUEPHA et al., 2018b, 2018a, 2020a; ZHOU et al., 2016, 2020). Além disso, modelos de predição empregando técnicas de ML são na maioria das vezes de difícil compressão, não sendo interpretáveis (WOLFF et al., 2019).

Portanto, ainda são necessários estudos que busquem desenvolver modelos de predição de readmissão hospitalar pediátrica que resultem em preditores de fácil obtenção, com plausibilidade biológica e que propicie a interpretação de forma racional das relações entre as variáveis independentes e a dependente. No melhor cenário em saúde, os modelos devem ser capazes de auxiliar os profissionais de saúde a identificar os pacientes com elevado risco de readmitir possibilitando implementar intervenções em tempo hábil, visando reduzir a ocorrência desse desfecho e consequentemente os diminuir os custos hospitalares e outros impactos negativos mencionados nos tópicos anteriores.

Quadro 1 - Resumo das características dos estudos que avaliaram readmissões hospitalares pediátricas em 30 dias.

Autor, ano	População	Local de estudo	Tipo de estudo	Desfecho	Readmissão em 30 dias (%)	Método utilizado	Desempenho dos modelos AUC (*Area Under The Curve*)
Estudos com abordagem tradicional							
Toomey, et. al., 2016	<18 anos N = 305	Hospital pediátrico	Transversal	Readmissões potencialmente evitáveis em 30 dias	29,5	Regressão logística (RL) multivariada	-
Ehwerhemuepha et.al., 2018a	> 28 dias e < 18 anos N = 38.143	Hospital pediátrico terciário	Retrospectivo	Readmissões não planejadas em 30 dias	10,4	Regressão logística (RL) multivariada Validação via sistema de pontuação LACE	0,79 – RL 0,68 – LACE
Han et.al., 2021	≤ 18 anos N = 12.421	Hospital do "*National Health Service*"	Retrospectivo	Readmissão por todas as causas em 28 dias	-	Validação via sistema de pontuação LACE	0,80
Zhou et.al., 2020	≤ 16 anos N = 73.132	Hospital pediátrico terciário	Coorte retrospectivo	Readmissão hospitalar não planejada em 30 dias	4,5	Regressão logística multivariada	0,64
Estudos com abordagem de *Machine learning*							
Ehwerhemuepha et.al., 2020c	≤ 18 anos	Hospitais do "*Cerner Health Facts Database*"	Retrospectivo	Readmissão pediátrica	12,6	Validação da plataforma *HealtheDataLab* Modelo 1 (M1) - Readmissão não planejada - centro único Modelo 2 (M2) - Readmissões planejadas e não planejadas - centro único	0,82 – M1 0,87 – M2

Zhou et.al., 2021	≤ 16 anos N = 940	Hospital pediátrico terciário	Caso-controle pareado retrospectivo	Readmissão hospitalar não planejada de 30 dias	4,5	Random Forest (RF) Regressão Logística (RL) Regressão Logística - *Stepwise* *Elastic Net – Lasso e Ridge* *Gradient Boosted Tree* (GB)	0,60 – RL 0,61 – RL *Stepwise* 0,63 – *Elastic net* 0,64 – RF 0,65 – GB
Taylor et al., 2020	≤ 18 anos N = 1.111.323	Hospitais pediátricos	Coorte retrospectiva	Readmissões não planejadas de 3, 7 e 30 dias	4,4	Modelos Generalizados Mistos (GLMM) *Extreme Gradient Boosting* (XGBoost)	0,81 – XGBoost 0,78 – GLMM
Symum et. al., 2021	< 18 anos	Hospitais pediátricos	Retrospectivo	Readmissão hospitalares não planejadas de 30 dias	8,3	Regressão logística - *backward stepwise* *Decision tree* (DT) - C4.5 *Support Vector machines* (SVM) - *Polynomial Kernel* *Gradient Boosting* (GB)	0,73 – SVM 0,69 – RL 0,67 – GB 0,64 – DT
Wolf et.al., 2019	< 18 anos N = 56.558	Hospital pediátrico	Retrospectivo	Readmissão por todas as causas em 30 dias	3,7	*Support Vector machines* (SVM) *Perceptron multicamadas* – AutoMLP (MLP1) *Perceptron multicamadas* – *artificial neural network* (MLP 2) *Naïve Bayes* (NB)	0,65 – NB 0,64 – MLP 1 0,53 – MLP 2 0,59 – SVM

Fonte: Próprio autor

2.4 O uso das técnicas de *machine learning:* potenciais aplicações nos serviços de saúde

Nos últimos anos houve um rápido crescimento na geração, armazenamento e disponibilização de dados, devido ao estabelecimento e universalização de tecnologias inteligentes como *Internet of Things* (*smart TV*, *alexa*, *wearables*: *smartwatches* e fones de ouvido e sensores para medir glicemia, batimentos) (HARRISON; SIDEY-GIBBONS, 2021; SARKER, 2021; SIDEY-GIBBONS; SIDEY-GIBBONS, 2019). Na assistência à saúde não foi diferente, observou-se ampla adoção de sistemas de prontuário eletrônico e utilização da telemedicina (HARRISON; SIDEY-GIBBONS, 2021; SARKER, 2021; SIDEY-GIBBONS; SIDEY-GIBBONS, 2019). Esse desenvolvimento da tecnologia possibilitou o fornecimento de grandes quantidades de dados históricos e clínicos documentados, ou seja, houve um ganho substancial de conhecimento longitudinal relacionado à saúde. Avanços tecnológicos baseados em inteligência artificial (IA) têm potencial para aprimorar a pesquisa médica e os cuidados clínicos e podem apoiar a equipe de saúde no diagnóstico e predição de doenças e personalização do cuidado (HARRISON; SIDEY-GIBBONS, 2021; SARKER, 2021; SIDEY-GIBBONS; SIDEY-GIBBONS, 2019).

O aprendizado de máquina - *machine learning* (ML) é uma subdisciplina da IA responsável por ajudar a impulsionar os avanços tecnológicos. O ML emprega diversos algoritmos, capazes de avaliar processos de alta complexidade com rapidez e precisão, detectando e extraindo padrões ou tendência internas, em grandes conjuntos de dados com uma diversidade de variáveis, para os quais a percepção humana pode ser falha (DAVENPORT; KALAKOTA, 2019; SARKER, 2021). Os algoritmos de aprendizado de máquina são divididos principalmente em quatro categorias: supervisionado, não supervisionado, semi-supervisionado e aprendizado por reforço. Algoritmos supervisionados visam fazer predições sobre um desfecho (variável dependente) com base em um conjunto de atributos (variáveis independentes). Estes algoritmos utilizam um conjunto de dados rotulados (desfecho informado, ou seja, a saída de dados é previamente conhecida) para aprender como os atributos e os desfechos estão relacionados, com o intuito de prever resultados de novos conjuntos de dados. O analista ajuda o sistema a construir o modelo transferindo o conhecimento do pesquisador para a máquina. Em contrapartida, os algoritmos não supervisionados visam encontrar padrões previamente indefinidos em conjuntos de dados, por exemplo, agrupando observações semelhantes em *clusters*. Eles usam dados que não foram “rotulados” por um supervisor humano (ou seja, observações que não foram categorizadas *a priori*). Por outro lado, os semi-supervisionados consistem numa hibridização dos métodos supervisionados e não supervisionados, usam dados rotulados e não rotulados. Já o aprendizado por reforço é baseado em recompensa ou penalidade (HARRISON; SIDEY-GIBBONS, 2021; SARKER, 2021; SIDEY-GIBBONS; SIDEY-GIBBONS, 2019).

Um importante campo de aplicação do ML é a tomada de decisão inteligente por meio da análise preditiva orientada por dados. A base da análise preditiva é capturar e explorar as relações entre variáveis explicativas e previstas de eventos anteriores para prever um resultado (SARKER, 2021). As técnicas de ML podem auxiliar na tomada de decisões clínicas inteligentes no domínio da saúde. O aprendizado de máquina pode contribuir na resolução de problemas de diagnóstico e prognóstico clínicos por meio da predição de doenças identificando os pacientes com maior risco, possibilitando o gerenciamento de pacientes permitindo alocação eficiente de recursos clínicos, processamento de imagens de diagnósticos, entre outras.

Nos últimos anos, em meio à crescente disponibilidade de dados de saúde, houve aumento do interesse na aplicação de diferentes métodos de ML para prever desfechos hospitalares, entre eles, a readmissão hospitalar (EHWERHEMUEPHA et al., 2020b; SYMUM; ZAYAS-CASTRO, 2021; TOOMEY et al., 2016; WOLFF et al., 2019; ZHOU et al., 2021). Na literatura é possível encontrar estudos sugerindo que as técnicas de aprendizado de máquina podem superar abordagens estatísticas tradicionais, como a regressão logística, tendo o potencial de melhorar a previsão de resultados negativos em saúde (ASHFAQ et al., 2019; BAIG et al., 2019; FUTOMA; MORRIS; LUCAS, 2015; MORGAN et al., 2019). A perspectiva é que o conhecimento adquirido por meio da aplicação de modelos de ML contribua substancialmente com os serviços, possibilitando ordenar o fluxo de atendimento, tornando os processos mais ágeis e eficientes para gerar resultados clínicos efetivos. De forma a priorizar os pacientes que mais necessitem, auxiliando os profissionais de saúde a tomar decisões, a planejar tratamentos e consequentemente, a melhorar a qualidade da assistência e reduzir os custos em saúde. Os pacientes também podem se beneficiar à medida que a previsão possibilitará o diagnóstico oportuno, o que permite o tratamento personalizado e eficaz, potencialmente prevenindo complicações decorrente do tratamento incorreto ou tardio, bem como redução do tempo de internação e consequentemente de possíveis complicações psicológicas (MICHAILIDIS et al., 2022).

Contudo, apesar de implicações que remontam à década de 1970, muitos profissionais de saúde ainda não estão familiarizados com o ML e não sabem como ele pode ser aplicado na prática hospitalar (HANDELMAN et al., 2018). Além disso, muitos modelos baseados em ML encontrados na literatura, apesar de obter boa predição não podem ser facilmente interpretáveis (WOLFF et al., 2019; ZHOU et al., 2021). Para aplicações em saúde, o fácil entendimento do modelo é, sem dúvida, uma propriedade importante tendo em vista à complexidade subjacente dos desfechos clínicos e o impacto potencial de decisões erradas (JOVANOVIC et al., 2016). Sendo assim, um dos grandes desafios dessa inovação tecnológica é a interpretabilidade dos modelos gerados, que se preocupem em explicar as relações entre as variáveis independentes e dependentes de forma racional e prática. Diante disso, é necessário ter uma compreensão fundamental dos modelos preditivos que forem implementados como suporte à decisão clínica. Modelos que além de serem capazes de prever o desfecho com boa precisão, sejam aplicáveis na prática clínica de forma racional.

3. OBJETIVOS

3.1 Objetivo Geral

Desenvolver modelos de predição de readmissão hospitalar potencialmente evitável em 30 dias, em crianças e adolescentes internados em um hospital universitário de nível terciário.

3.2 Objetivos Específicos

- Validar escore conhecido na população adulta (HOSPITAL *score*) para prever readmissões potencialmente evitáveis em 30 dias em população pediátrica.
- Aprimorar os métodos convencionais de predição de readmissão hospitalar potencialmente evitável por meio da utilização de técnicas de *machine learning,* identificando os atributos associados à readmissão hospitalar potencialmente evitável em 30 dias.
- Explorar a utilidade de técnicas de interpretabilidade aplicadas em classificadores caixa-preta de alto desempenho.

4 RESULTADOS

4.1 Primeiro manuscrito

European Journal of Pediatrics
https://doi.org/10.1007/s00431-022-04795-z

RESEARCH

Validation of the HOSPITAL score as predictor of 30-day potentially avoidable readmissions in pediatric hospitalized population: retrospective cohort study

Nayara Cristina da Silva[1] · Marcelo Keese Albertini[2] · André Ricardo Backes[3] · Geórgia das Graças Pena[1]

Received: 19 May 2022 / Revised: 26 December 2022 / Accepted: 28 December 2022

O artigo não será disponibilizado na íntegra por questões legais relacionadas ao sigilo exigido pela revista. Acesse o artigo através do link: https://link.springer.com/article/10.1007/s00431-022-04795-z.

Reforça-se que este artigo contou com o apoio financeiro da Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) por meio do edital APQ-01331-21, conforme pode ser confirmado na página 6 do artigo original: "*This study received financial support from "the Foundation for Research Support of the State of Minas Gerais" (FAPEMIG – Fundação de Amparo à Pesquisa do Estado de Minas Gerais) in a public call process number APQ-01331–21.The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript*".

4.2 Segundo manuscrito

Decision tree model predicts avoidable hospital 30 days readmissions in pediatric population

Authors: Nayara Cristina da Silva, Laurence Rodrigues do Amaral, Marcelo Keese Albertini, André Ricardo Backes, Geórgia das Graças Pena*

Affiliations

[a] Graduate Program in Health Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil, Address: Pará Av, 1720, Campus Umuarama, Uberlândia, Minas Gerais, Brazil. Zip Code 38400-902. nayara.cristina.silva@outlook.com ORCID:0000-0003-3551-475X, georgia@ufu.br ORCID: 0000-0002-0360-223X

[b] Laboratory of Bioinformatics and Molecular Analysis (LBAM), Federal University of Uberlandia, Patos de Minas, Minas Gerais, Brazil. ZIP Code 38700-002. laurence@ufu.br ORCID: 0000-0003-4681-5451

[c] School of Computer Science, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil. Zip Code 38408-100. albertini@ufu.br ORCID: 0000-0002-1846-946X

[d] Department of Computing, Federal University of Sao Carlos, Sao Carlos, São Paulo, Brazil. ZIP Code 13565-905. arbackes@yahoo.com.br ORCID: 0000-0002-7486-4253

***Corresponding author:** Geórgia das Graças Pena, Graduate Program in Health Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil. Address: Pará Av, 1720, Campus Umuarama +55 34 3225-8584, Zip Code 38.405-320 Uberlândia-MG, E-mail: georgia@ufu.br, georgiapena@gmail.com

Word Count: 2315
Number of tables: 1
Number of figures: 2

Abstract

Background and Objective: Identifying patients at high risk of avoidable readmission remains a challenge for healthcare professionals. Despite the recent interest in Machine Learning in this topic, studies are scarce and commonly using only black box algorithms. The aim of our study was to build an interpretable predictive model using a decision tree inference to identify pediatric patients at the risk of 30-day potentially avoidable readmissions. **Methods**: Retrospective cohort study conducted with all patients under 18 years old admitted to a tertiary university hospital (n=528). Demographic, clinical and nutritional data were collected from electronic databases. The outcome was the potentially avoidable 30-days readmissions. The J48 algorithm was used to develop the best-fit trees capable of classify the outcome efficiently. Leave-one-out cross-validation was applied and we computed the area under the receiver operating curve (AUC). **Results**: The most important attributes of the model were C Reactive Protein, hemoglobin and sodium levels, besides nutritional monitoring. We obtained an AUC of 0.65 and accuracy of 63.3% the full training and leave-one-out cross-validation. **Conclusion:** Our model allows the identification of 30-day potentially avoidable readmissions through practical indicators facilitating timely interventions by the medical team, and it could contribute to reduce this outcome.

Keywords: Hospital Readmission; Pediatrics; Decision tree; Algorithms; Supervised Machine Learning

Introduction

Pediatric hospital readmissions have received attention in recent decades. The 30-day readmission rate for hospitalized children is still high, ranging from 4.40 to 29.50% [1–3]. Studies show that hospital readmission can negatively influence patients' quality of life [4,5] with short and long-term consequences [4,6–9]. Besides, they can contribute substantially to the increase in healthcare costs [10–12]. A study of pediatric patients revealed that the hospital cost for all admissions and readmissions is US$ 17.3 billion, of which 21.5% (US$ 3.71 billion) was spent during a readmission hospital stay [12]. Another study found that of the $11.6 billion spent annually for all hospitalizations, being US$ 2.0 billion (16.9% of total hospitalization costs) related to all-cause readmissions within 30 days [11].

Despite that, identifying patients at high risk of readmission and implementing timely interventions remains a challenge for healthcare professionals. Recently, predictive modeling has been pointed out as an efficient method to stratify the risk of readmission, allowing the targeting of preventive interventions to patients at risk, thus optimizing the allocation of clinical resources [13]. Tools capable of early identification of patients at risk of readmission, have been proposed in order to helping to minimize the incidence of hospital readmissions [2,14–17]. However, there is still a lack of practical and easily understood predictive models to support clinical decisions. The reported models are often poorly designed, being mainly based on black-box algorithms [2,14–17] which makes it impossible to know how clinical factors led to forecasts.

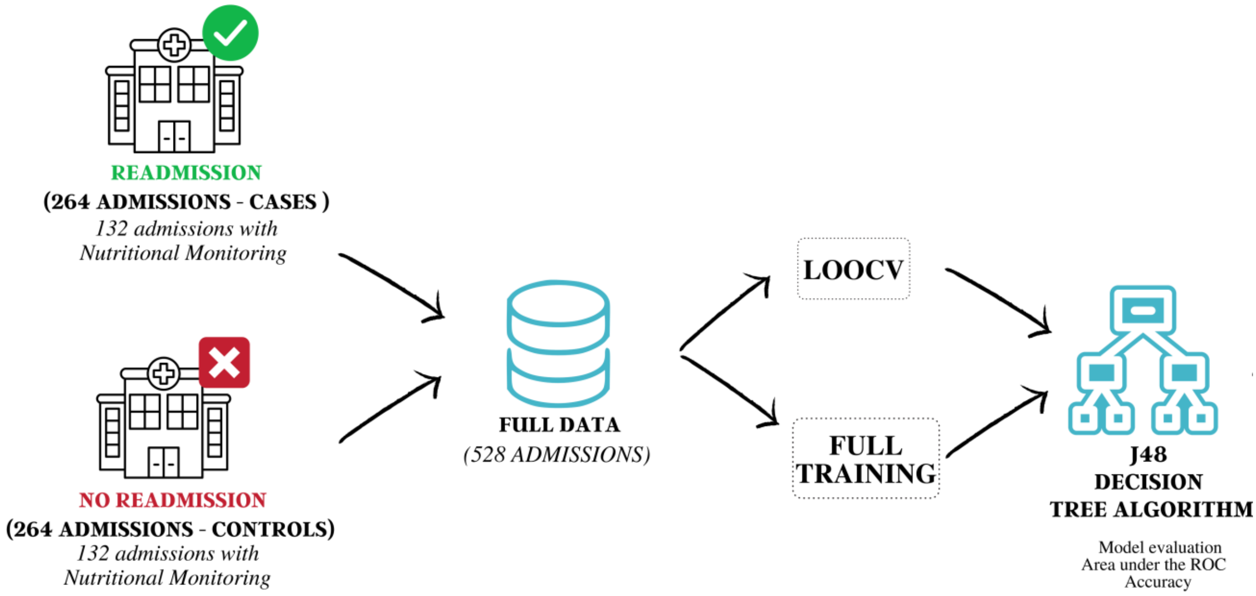
For healthcare applications, the model's interpretability is as important as its performance. So, when it is possible to observe the attributes and the decision paths rationally, the predictive clinical model became easier their application by health team. Given this, decision trees, based on supervised machine learning approach, can be an excellent option. Since this method relates the nodes to each other hierarchically [18], resulting in an easy model to interpret. Therefore, the aim of our study is to build an interpretable predictive model using a decision tree algorithm to identify patients at the risk of 30-day potentially avoidable readmissions.

Methods

Study design and Settings

A retrospective cohort study was conducted at a tertiary university hospital from January 1st, 2014 to December 31st, 2018. Children and adolescents between 0 and 18 years old were included, who had all data retrieved from electronic databases. We excluded hospitalizations that resulted in hospital death (not at risk for readmission outcome), discharges against medical advice (not at the opportunity to implement care plan and discharge instruction) and patients with incomplete data in the electronic databases.

For this study, a 1:1 nested case–control design was performed, as classifiers do not work well with unbalanced data. To overcome class imbalance, we included patients who readmitted and had complete data (cases) and randomly selected patients with complete data who did not readmit (controls) (**Figure 1**). The University's Ethics Committee approved this study (CAAE 51706221.3.0000.5152, protocol number: 5.003.236).



LOOCV: leave-one-out cross-validation

Figure 1 Study flow diagram

Predictors selection

Demographic data (age and sex) and clinical data (wards, admission type, diagnoses and length of hospital stay - LOS), presence or absence of any nutritional monitoring during hospitalization and biochemical exams (blood count, leukogram, C-reactive protein - CRP and sodium) were obtained from electronic databases.

We classified the age of patients in six groups, according to childhood and adolescence periods of growth and development: <1 years, ≥ 1 to <5 years, ≥ 5 to <9 years, ≥9 to < 13 years, ≥13 to < 16 years and ≥16 years. Length of hospital stay was categorized into quartiles: < 8, ≥ 8 to < 17, ≥ 17 and < 38 and ≥ 38 days. All blood tests were performed at a single laboratory, CRP was measured by methods immunoturbidimetry using Cobas® 6000 analyzer, sodium measured by potentiometric methods and hematological parameters were analyzed using an automated Sysmex XN-3000™ hematology analyzer. We categorized biochemical exams as altered and normal, considering age and sex (Supplementary Material). Nutritional data were not filled in a standardized way in the electronic databases, so it was not possible to classify the nutritional status of patients. Therefore, we could only observe whether the child had any nutritional monitoring during hospitalization. Therefore, we created a variable showing patients who had nutritional monitoring during hospital stay (with or without nutritional monitoring during hospitalization).

Outcome

The outcome was the 30-day potentially avoidable readmissions, considered as a new admission within this short period after the immediately previous hospital discharge. Thus, all unavoidable readmissions, all patients admitted to wards with planned hospitalizations (obstetrics, gynecology and transplant) or with predictable admissions, such as labour/delivery, and chemotherapy or radiotherapy treatments (ambulatory care) were exclude.

Statistical Analysis

Descriptive data were summarized using proportions or means (± standard deviation, SD). For the statistical analysis, we use the R Project (version 4.0.3), the RStudio (version 4.0.2), and considered the 95% confidence intervals (95% CI). The Machine Learning based decision tree algorithm J48, present in Weka suite, was used to develop best-fit trees in order to select the minimum set of characteristics capable of classify patients at risk of 30-day potentially avoidable readmissions efficiently. The leave-one-out cross-validation (LOOCV) applied to estimate the classification accuracy and test the generalizability of the model. We computed the area under the receiver operating curve (AUC). We performed the analyses using WEKA software (Waikato Environment for Knowledge Analysis, version 3.6.1).

Results

Of the 528 patients aged between 0 and 18 years, 60.2% (318) were male, 33.5% (177) had under one year of age. The frequency of 30-day potentially avoidable readmissions was 50.0% (264). Of these 31.10% (82) had a length of hospital stay less than 8 days, 70.8% (187) and 85.6% (226) had hemoglobin and CRP levels altered, respectively (**Table 1**).

Table 1 Demographic, clinical and biochemical variables by potentially avoidable 30- day readmission

Predictors	**All %(n)**	**30-day readmissions %(n)**	
		No 50 (264)	**Yes** 50 (264)
Sex, male %(n)	60.2 (318)	53.8 (142)	66.7 (176)
Age group, years			
< 1	33.5 (177)	43.9 (116)	23.1 (61)
≥ 1 and < 5	26.9 (142)	25.8 (68)	28.0 (74)
≥ 5 and < 9	13.8 (73)	12.9 (34)	14.8 (39)
≥ 9 and < 13	11.7 (62)	5.3 (14)	18.2 (48)
≥ 13 and < 16	7.8 (41)	6.8 (18)	8.7 (23)
≥ 16	6.3 (33)	5.3 (14)	7.2 (19)
Wards			
Clinical	76.1 (402)	70.1 (185)	82.2 (217)
Urgent/Emergency/Surgical	10.6 (56)	9.1 (24)	12.1 (32)
Newborns	13.3 (70)	20.8 (55)	5.7 (15)
Admission type			
Elective	9.1 (48)	11.0 (29)	7.2 (19)
Urgent or Emergency	90.9 (480)	89.0 (235)	92.8 (245)
Length of hospitalization, days			
< 8	23.5 (124)	15.9 (42)	31.1 (82)
≥ 8 and < 17	25.2 (133)	31.4 (83)	18.9 (50)
≥ 17 and < 38	26.1 (138)	23.9 (63)	28.4 (75)
≥ 38	25.2 (133)	28.8 (76)	21.6 (57)
Hemoglobin, g/dL			
Normal	38.3 (202)	47.3 (125)	29.2 (77)
Altered	61.7 (326)	52.7 (139)	70.8 (187)

Red Cell Distribution Width, %			
Normal	2.3 (12)	2.7 (7)	1.9 (5)
Altered	97.7 (516)	97.3 (257)	50.2 (259)
C-reactive protein, mg/dL			
Normal (< 0.5)	22.0 (116)	29.5 (78)	14.4 (38)
Altered (≥ 0.5)	78.0 (412)	70.5 (186)	85.6 (226)
Sodium, mEq/L			
Normal (≥ 134)	94.7 (500)	92.8 (245)	96.6 (255)
Altered (< 134)	5.3 (28)	7.2 (19)	3.4 (9)
Nutritional monitoring			
Yes	50.0 (264)	50.0 (132)	50.0 (132)
No	50.0 (264)	50.0 (132)	50.0 (132)

Considering all available predictors, a decision tree inferred by from the J48 method was constructed to classify patients with a risk of 30-day potentially avoidable readmissions (**Figure 2**). The decision tree algorithm to classify readmission vs non-readmission proposed the use of CRP and hemoglobin, sodium levels and nutritional data, obtaining an AUC of 0.65 and accuracy of 63.3% the full training (FULL) and leave-one-out cross-validation (LOOCV) (**Figure 2**).

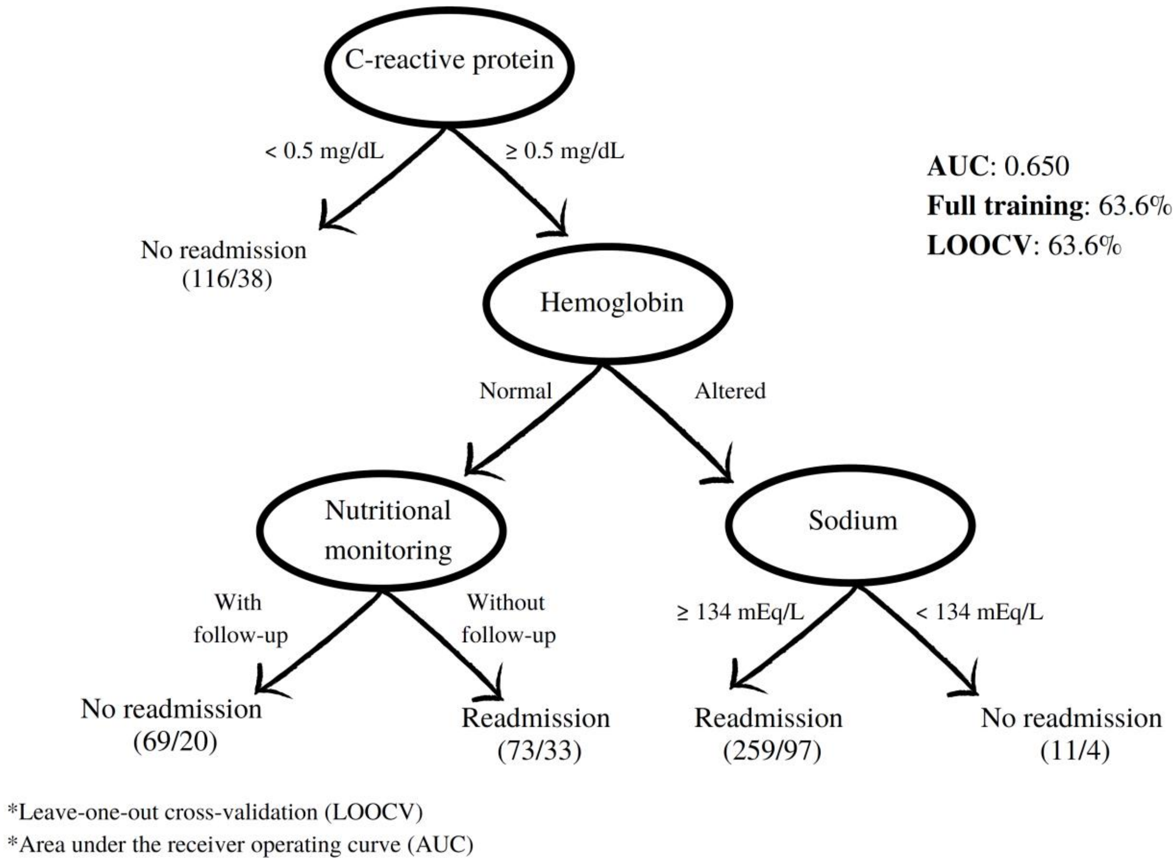


Figure 2 Decision tree algorithm proposed to differentiate patients with 30-day potentially avoidable readmissions. The total number of classified admissions (correct and incorrect) for each class is shown in parentheses for each terminal node. Incorrectly classified admissions appear after slash "/". The area under the receiver operating curve (AUC), full training (FULL) and leave-one-out cross-validation (LOOCV) accuracies is shown in the figure.

Discussion

In this study, we used the J48 algorithm to build a classification model for 30-day potentially avoidable readmissions. The most important attributes for the model were CRP, hemoglobin and sodium levels, besides nutritional monitoring. Our findings were confirmed using the leave-one-out cross-validation. To the best of our knowledge, our study is the first to build a prediction model based on a decision tree with only three levels and confirmation by leave-one-out cross-validation. Being a model of easy understanding and application in clinical practice, making clear the contribution and direction of each association, as well as using attributes routinely found in hospital services. Furthermore, the rule found by our model applies to 63.6% of new cases.

Previous studies had reported many risk factors involved with increased risk of hospital readmission: age [19–21], multimorbidity [13,15,20,22], prolonged duration of the last hospital stay [19,22,23], polypharmacy [17,24] and presence of diagnostics/conditions like anemia, malnutrition, cancer and global developmental delay [1,21]. However, hospital readmission is still a recurring problem and difficult clinical management, involved with short and long-term deleterious effects [4,6–9], besides contributing substantially to hospital costs [10–12]. The early identification of patients at greater risk of readmission provides opportunities for targeting interventions and allocating clinical and financial resources. In this sense, predictive models have been proposed in the literature, with variable performances such as AUC of 0.65 using Naive Bayes for all-cause 30-day readmission [14], AUC of 0.65 with Gradient Boosted for 30-day unplanned hospital readmissions [15], AUC of 0.73 using Support Vector Machines with Polynomial Kernel for at-discharge models [16], and even AUC of 0.81 with XGBoost for unplanned readmissions within 30 days [2] all for 30-day hospital readmission. However, there are few practical and interpretable models that easy to understand and apply, capable of supporting clinical decisions.

In this sense, we use a Machine Learning decision tree-based algorithm in order to build an interpretable model capable of identifying patients at risk of hospital readmission. Despite presenting a modest performance, AUC=0.65, our results are relevant and capable of identifying new patients with a risk of readmission in 63.6% of cases (LOOCV= 63.6%). Therefore, our model can effectively contribute to clinical practice, since it is a model that is easy to understand and apply in hospital routine, besides employing only relevant and easily got attributes in medical services.

In the model built, using the J48 algorithm, we identified that the most relevant attribute was the CRP levels, with more information in each iteration, being placed as the root of our decision tree, in which their high levels contribute to a risk of readmission. CRP is an acute phase protein, considered a sensitive and rapid response marker of inflammation. Studies have suggested that high CRP concentrations are correlated with the presence of ongoing organ dysfunction [25,26]. So, the elevated CRP may be an indirect marker of disease severity, and therefore may be related to a higher risk of hospital readmission [25,26]. A study found that high CRP levels were associated with a higher risk of readmission at 7 days [25], and also with a higher risk of adverse outcome after discharge from the intensive care unit [26].

Our decision tree also used hemoglobin levels in order to identify patients at risk of readmission. For patients with normal hemoglobin levels, it is necessary to assess the presence or absence of nutritional monitoring during hospitalization. Those patients without nutritional support during hospitalization have a risk of being readmitted when compared to patients who are followed up by a nutritionist.

Nutritional data have already been explored in previous observational studies, however, in most cases, they only assess the association of malnutrition with hospital readmission [1,21]. However, they cannot address the relevance of nutritional monitoring during hospitalization. At least 80% of patients admitted to a hospital are expected to undergo nutritional screening within the first 24 hours of admission [27]. Nutritional screening is the initial step allowing the identification of patients at nutritional risk and early intervention when necessary, minimizing deleterious effects related to nutritional status [28]. However, regarding the nutritional approach in pediatric patients at the hospital level, there is still no consensus and the tools are scarce and little used [29]. A study carried out in Brazil revealed that 43.3% of medical records did not contain any records of the children's nutritional status [30]. According to the authors, this reiterates the under-reporting of this important data by the entire health team that assists hospitalized children [30]. Because of this deficiency, many hospitalized patients may not receive any type of nutritional monitoring, which would make it difficult to identify nutritional losses with negative effects on their health. Therefore, it is relevant, besides malnutrition, to assess the effectiveness of nutritional monitoring and its contribution to hospital readmission.

On the other hand, if the hemoglobin levels are altered, it is necessary to assess the sodium levels. One study found that the hemoglobin level is inversely correlated with 30-day hospital readmission rates [31]. Low hemoglobin levels may be related to anemia, a condition often diagnosed in hospitalized children [32–34] that can be both a symptom and a complication of many diseases. Studies suggest that anemia is a negative prognostic factor and may contribute to the worsening of clinical outcomes, besides negatively affecting the child's health, with long-term deleterious effects [32,35]. Moreover, sodium levels may contribute to identifying patients at risk of hospital readmission. Abnormal sodium levels are one of the most common electrolyte disturbances in hospitalized patients and have been associated with worse clinical outcomes. Studies have revealed that hyponatremia is associated with hospital readmission [36–39]. However, in our study, we found no association for sodium levels below 134 mEq/L. One hypothesis, for the absence of association, may be because of the low frequency of hyponatremia (5.3%) in the evaluated patients. Nevertheless, sodium levels above 134 mEq/L were associated with hospital readmission, and this rule applies to almost 50% of the patients evaluated. Studies suggest that sodium levels may be a marker of the severity of the underlying disease, being related to an increase in negative health outcomes such as mortality [36,40,41], increased length of stay [36,38,41], or yet hospital readmission [36–39].

Hospital readmission is a challenging outcome, it contributes substantially to increased costs and is often associated with adverse health outcomes. Therefore, models capable of predicting the risk of readmission are of interest, these tools can help identify and reduce readmission, improve overall patient care and reduce healthcare costs. In our study we built a classification model for 30-day potentially avoidable readmissions using the J48 algorithm. We showed that one of the relevant predictors was nutritional monitoring, often neglected by predictive models. Future studies should be carried out exploring nutritional data, aiming to deepen knowledge and make health professionals aware of the importance of nutritional screening.

Some limitations for this study need to be pointed out. First of all, the extrapolation of the data must be careful, since this study was carried out with pediatric patients from a tertiary university hospital. Secondly, the small sample size, since the absence of complete data made it impossible to have a larger database. Algorithms are more effective when used in large databases. However, the present study has strengths; we evaluated all available data during the study period, applied LOOCV cross-validation, used predictors relevant to the pediatric population and were easily accessible, and finally, we sought to build a model that was easy to interpret and apply in practice.

Conclusion

The decision tree model found making showed that the CRP, hemoglobin, sodium levels and nutrition monitoring are the most important to classify 30-day potentially avoidable readmissions. Our model allows the identification of individuals at risk of readmission, in an easy and practical way, facilitating the targeting of interventions by the medical team, and contributing to minimize this outcome.

Acknowledgments: The technology and information center (CTI) team of the Clinical Hospital of the Federal University of Uberlandia for helping to databases extraction. Nayara. C. Silva received a doctoral fellowship from "The Foundation for Research Support of the State of Minas Gerais" (FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais). André R. Backes gratefully acknowledges the financial support of CNPq (Grant #307100/2021-9). Marcelo K. Albertini gratefully acknowledges the financial support of CNPq (Grant #406418/2021-7).

Conflicts of Interest: The authors declare no conflict of interest.

Funding Sources: This study received financial support from "The Foundation for Research Support of the State of Minas Gerais" (FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais) in a public call process number APQ-01331-21. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abbreviations:
AUC: area under the curve
CRP: C-reactive protein
FULL: full training
LOOCV: leave-one-out cross-validation
LOS: length of hospital stay
ROC: receiver operating characteristic
SD: standard deviation

Statement of Authorship

MSc Silva conceptualized and investigation the study, designed the methodology, performed the formal analyzes and data curation the study, drafted the initial manuscript and reviewed the manuscript.
Dr Laurence Amaral designed the methodology, performed the formal analyzes and data curation the study, reviewed the manuscript.
Drs Marcelo Albertini and André Backes designed the methodology, reviewed the manuscript.
Dr.ª Pena conceptualized and investigation the study, reviewed the manuscript.
All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

References

[1] Ehwerhemuepha L, Bendig D, Steele C, Rakovski C, Feaster W. The Effect of Malnutrition on the Risk of Unplanned 7-Day Readmission in Pediatrics. Hosp Pediatr 2018;8:207–13. https://doi.org/10.1542/hpeds.2017-0195.

[2] Taylor T, Altares Sarik D, Salyakina D. Development and Validation of a Web-Based Pediatric Readmission Risk Assessment Tool. Hosp Pediatr 2020;10:246–56. https://doi.org/10.1542/hpeds.2019-0241.

[3] Toomey SL, Peltz A, Loren S, Tracy M, Williams K, Pengeroth L, et al. Potentially Preventable 30-Day Hospital Readmissions at a Children's Hospital. Pediatrics 2016;138:e20154182. https://doi.org/10.1542/peds.2015-4182.

[4] Silva VLS da, França GVA de, Munhoz TN, Santos IS, Barros AJD, Barros FC, et al. Hospitalization in the first years of life and development of psychiatric disorders at age 6 and 11: a birth cohort study in Brazil. Cad Saude Publica 2018;34:1–13. https://doi.org/10.1590/0102-311x00064517.

[5] Coyne I. Children's Experiences of Hospitalization. J Child Heal Care 2006;10:326–36. https://doi.org/10.1177/1367493506067884.

[6] Delvecchio E, Salcuni S, Lis A, Germani A, Di Riso D. Hospitalized Children: Anxiety, Coping Strategies, and Pretend Play. Front Public Heal 2019;7:1–8. https://doi.org/10.3389/fpubh.2019.00250.

[7] Cahayag V. Hospitalization and Child Development: Effects on Sleep, Developmental Stages, and Separation Anxiety. Nurs | Sr Theses 2020;17. https://doi.org/doi.org/10.33015/dominican.edu/2020.NURS.ST.09.

[8] Pufal EC, Müller AB, Bandeira PFR, Valentini NC. Motor development in the hospitalized infant and its biological and environmental characteristics. Clin Biomed Res 2018;38:66–73. https://doi.org/10.4322/2357-9730.75638.

[9] Rashikj Canevska O. The Impact of Hospitalization on Psychophysical Development and Everyday Activities in Children. Annu Fac Philos Skopje 2018;71:465–70. https://doi.org/10.37510/godzbo1871465rc.

[10] Markham JL, Hall M, Gay JC, Bettenhausen JL, Berry JG. Length of Stay and Cost of Pediatric Readmissions. Pediatrics 2018;141:e20172934. https://doi.org/10.1542/peds.2017-2934.

[11] Gay JC, Agrawal R, Auger KA, Del Beccaro MA, Eghtesady P, Fieldston ES, et al. Rates and Impact of Potentially Preventable Readmissions at Children's Hospitals. J Pediatr 2015;166:613-619.e5. https://doi.org/10.1016/j.jpeds.2014.10.052.

[12] Kane JM, Hall M, Cecil C, Montgomery VL, Rakes LC, Rogerson C, et al. Resources and Costs Associated With Repeated Admissions to PICUs. Crit Care Explor 2021;3:e0347. https://doi.org/10.1097/CCE.0000000000000347.

[13] Niehaus IM, Kansy N, Stock S, Dötsch J, Müller D. Applicability of predictive models for 30-day unplanned hospital readmission risk in paediatrics: a systematic review. BMJ Open 2022;12:e055956. https://doi.org/10.1136/bmjopen-2021-055956.

[14] Wolff P, Graña M, Ríos SA, Yarza MB. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. Biomed Res Int 2019;2019:1–9. https://doi.org/10.1155/2019/8532892.

[15] Zhou H, Albrecht MA, Roberts PA, Porter P, Della PR, Della PR. Using machine learning to predict paediatric 30-day unplanned hospital readmissions: A case-control retrospective analysis of medical records, including written discharge documentation. Aust Heal Rev 2021;45:328–37. https://doi.org/10.1071/AH20062.

[16] Symum H, Zayas-Castro J. Identifying Children at Readmission Risk: At-Admission versus Traditional At-Discharge Readmission Prediction Model. Healthcare 2021;9:1334. https://doi.org/10.3390/healthcare9101334.

[17] Ehwerhemuepha L, Gasperino G, Bischoff N, Taraman S, Chang A, Feaster W. HealtheDataLab – a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. BMC Med Inform Decis Mak 2020;20:115. https://doi.org/10.1186/s12911-020-01153-7.

[18] Hajjej F, Alohali MA, Badr M, Rahman MA. A Comparison of Decision Tree Algorithms in the Assessment of Biomedical Data. Biomed Res Int 2022;2022:1–9. https://doi.org/10.1155/2022/9449497.

[19] Feudtner C, Levin JE, Srivastava R, Goodman DM, Slonim AD, Sharma V, et al. How Well Can Hospital Readmission Be Predicted in a Cohort of Hospitalized Children? A Retrospective, Multicenter Study. Pediatrics 2009;123:286–93. https://doi.org/10.1542/peds.2007-3395.

[20] Zhou H, Roberts PA, Dhaliwal SS, Della PR. Risk factors associated with paediatric unplanned hospital readmissions: a systematic review. BMJ Open 2019;9:e020554. https://doi.org/10.1136/bmjopen-2017-020554.

[21] Kumar D, Swarnim S, Sikka G, Aggarwal S, Singh A, Jaiswal P, et al. Factors Associated with Readmission of Pediatric Patients in a Developing Nation. Indian J Pediatr 2019;86:267–75. https://doi.org/10.1007/s12098-018-2767-0.

[22] Ehwerhemuepha L, Finn S, Rothman M, Rakovski C, Feaster W. A Novel Model for Enhanced Prediction and Understanding of Unplanned 30-Day Pediatric Readmission. Hosp Pediatr 2018;8:578–87. https://doi.org/10.1542/hpeds.2017-0220.

[23] Zhou H, Della P, Roberts P, Porter P, Dhaliwal S. A 5-year retrospective cohort study of unplanned readmissions in an Australian tertiary paediatric hospital. Aust Heal Rev 2019;43:662–71. https://doi.org/10.1071/AH18123.

[24] Ehwerhemuepha L, Pugh K, Grant A, Taraman S, Chang A, Rakovski C, et al. A Statistical-Learning Model for Unplanned 7-Day Readmission in Pediatrics. Hosp Pediatr 2020;10:43–51. https://doi.org/10.1542/hpeds.2019-0122.

[25] Ziv-Baran T, Wasserman A, Shteinvil R, Zeltser D, Shapira I, Shenhar-Tsarfaty S, et al. C-reactive protein and emergency department seven days revisit. Clin Chim Acta 2018;481:207–11. https://doi.org/10.1016/j.cca.2018.03.022.

[26] Gülcher SS, Bruins NA, Kingma WP, Boerma EC. Elevated C-reactive protein levels at ICU discharge as a predictor of ICU outcome: a retrospective cohort study. Ann Intensive Care 2016;6:5. https://doi.org/10.1186/s13613-016-0105-0.

[27] Indicadores de Qualidade em Terapia Nutricional Pediátrica. Int Life Sci Inst Do Bras 2017;3. https://doi.org/10.51723/ccs.v26i03/04.307.

[28] Raslan M, Gonzalez MC, Dias MCG, Paes-Barbosa FC, Cecconello I, Waitzberg DL. Aplicabilidade dos métodos de triagem nutricional no paciente hospitalizado. Rev Nutr 2008;21:553–61. https://doi.org/10.1590/S1415-52732008000500008.

[29] Joosten KFM, Hulst JM. Nutritional screening tools for hospitalized children: Methodological considerations. Clin Nutr 2014;33:1–5. https://doi.org/10.1016/j.clnu.2013.08.002.

[30] Sarni ROS, Carvalho M de FCC, Monte CMG do, Albuquerque ZP, Souza FIS. Anthropometric evaluation, risk factors for malnutrition, and nutritional therapy for children in teaching hospitals in Brazil. J Pediatr (Rio J) 2009;85:223–8. https://doi.org/10.2223/JPED.1890.

[31] Lin RJ, Evans AT, Chused AE, Unterbrink ME. Anemia in General Medical Inpatients Prolongs Length of Stay and Increases 30-Day Unplanned Readmission Rate. South Med J 2013;106:316–20. https://doi.org/10.1097/SMJ.0b013e318290f930.

[32] Melku M, Alene KA, Terefe B, Enawgaw B, Biadgo B, Abebe M, et al. Anemia severity among children aged 6-59 months in Gondar town, Ethiopia: A community-based cross-sectional study. Ital J Pediatr 2018;44:1–12. https://doi.org/10.1186/s13052-018-0547-0.

[33] Jutras C, Charlier J, François T, Du Pont-Thibodeau G. Anemia in Pediatric Critical Care. Int J Clin Transfus Med 2020;8:23–33.

[34] Sloniewsky D. Anemia and Transfusion in Critically Ill Pediatric Patients. A Review of Etiology, Management, and Outcomes. Crit Care Clin 2013;29:301–17. https://doi.org/10.1016/j.ccc.2012.11.005.

[35] Walter T, De Andraca I, Chadud P, Perales CG. Iron deficiency anemia: Adverse effects on infant psychomotor development. Pediatrics 1989;84:7–17.

[36] Lu H, Vollenweider P, Kissling S, Marques-vidal P. Prevalence and Description of Hyponatremia in a Swiss Tertiary Care Hospital : An Observational Retrospective Study. Front Med 2020;7:1–9. https://doi.org/10.3389/fmed.2020.00512.

[37] Donze J, Aujesky D, Williams D, Schnipper JL. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. JAMA Intern Med 2013;173:632–8. https://doi.org/10.1001/jamainternmed.2013.3023.

[38] Corona G, Giuliani C, Parenti G, Colombo GL, Sforza A, Maggi M, et al. The Economic Burden of Hyponatremia: Systematic Review and Meta-Analysis. Am J Med 2016;129:823-835.e4. https://doi.org/10.1016/j.amjmed.2016.03.007.

[39] Donzé JD, Beeler PE, Bates DW. Impact of Hyponatremia Correction on the Risk for 30-Day Readmission and Death in Patients with Congestive Heart Failure. Am J Med 2016;129:836–42. https://doi.org/10.1016/j.amjmed.2016.02.036.

[40] Girardeau Y, Jannot AS, Chatellier G, Saint-Jean O. Association between borderline dysnatremia and mortality insight into a new data mining approach. BMC Med Inform Decis Mak 2017;17:1–10. https://doi.org/10.1186/s12911-017-0549-7.

[41] Akirov A, Diker-Cohen T, Steinmetz T, Amitai O, Shimon I. Sodium levels on admission are associated with mortality risk in hospitalized patients. Eur J Intern Med 2017;46:25–9. https://doi.org/10.1016/j.ejim.2017.07.017.

4.3 Terceiro manuscrito

Machine learning for hospital readmission prediction in pediatric population

Authors: Nayara Cristina da Silva, MSc[a], Marcelo Keese Albertini, PhD[b], André Ricardo Backes, PhD[c], Geórgia das Graças Pena, PhD[a*]

Affiliations
[a] Graduate Program in Health Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil, Address: Pará Av, 1720, Campus Umuarama, Uberlândia, Minas Gerais, Brazil. Zip Code 38400-902. nayara.cristina.silva@outlook.com ORCID:0000-0003-3551-475X, georgia@ufu.br ORCID: 0000-0002-0360-223X
[b] School of Computer Science, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil. Zip Code 38408-100. albertini@ufu.br ORCID: 0000-0002-1846-946X
[c] Department of Computing, Federal University of Sao Carlos, Sao Carlos, São Paulo, Brazil. ZIP Code 13565-905. arbackes@yahoo.com.br ORCID: 0000-0002-7486-4253

***Corresponding author:** Geórgia das Graças Pena, Graduate Program in Health Sciences, Federal University of Uberlandia, Uberlandia, Minas Gerais, Brazil. Address: Pará Av, 1720, Campus Umuarama +55 34 3225-8584, Zip Code 38.405-320 Uberlândia-MG, E-mail: georgia@ufu.br, georgiapena@gmail.com

Short title: Machine learning for hospital readmission prediction

Conflicts of Interest: The authors declare no conflict of interest.

Funding Sources: This study received financial support from “The Foundation for Research Support of the State of Minas Gerais” (FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais) in a public call process number APQ-01331-21. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Word Count: 3051
Number of tables: 2
Number of figures: 4

Abbreviations
AUC: area under the receiver operating curve
CART: classification and regression tree
CRP: C-reactive protein
GBM: gradient boosting machine
ICD: International Statistical Classification of Diseases
IQR: interquartile ranges
LR: logistic regression
ML: machine learning
RF: random forest
RDW: red cell distribution width
SD: standard deviation

SHAP: SHApley Additive exPlanation
XGBoost: extreme gradient boosting

Article Summary: Pediatric readmissions are a burden on patients, families, and the healthcare system. Machine learning approaches, can predict potentially avoidable 30-day pediatric hospital readmission into tertiary assistance.

What's Known on This Subject: Readmission is a complex and multifactorial process. . Studies have shown that the use of artificial intelligence techniques can be promising to predict pediatric hospital readmission.

What This Study Adds: The algorithm XGBoost with bagging imputation obtained AUC of 0.814 (J-index of 0.484). Cancer diagnosis, age, red blood cells, leukocytes, red cell distribution width and sodium levels, elective admission, and multimorbidity were the most important characteristics to classified between readmission and non-readmission groups.

Statement of Authorship

Msc Silva conceptualized and investigation the study, designed the methodology, performed the formal analyses and data curation the study, drafted the initial manuscript and reviewed the manuscript.

Dr. Marcelo Albertini conceptualized and investigation the study, designed the methodology, performed the formal analyses and critically reviewed and revised the manuscript.

Dr.André Backes conceptualized and investigation the study, designed the methodology, critically reviewed and revised the manuscript.

Dr.ª Pena conceptualized and investigation the study, designed the methodology, critically reviewed and revised the manuscript.

All authors approved the final manuscript as submitted and agree to be accountable for all aspects of the work.

Abstract

Background and Objective: Pediatric readmissions are a burden on patients, families, and the healthcare system. In order to identify patients at higher readmission risk, more accurate techniques, as machine learning (ML), could be a good strategy to expand the knowledge in this area. The aim of this study was to develop predictive models capable of identifying children and adolescents at high risk of potentially avoidable 30-day readmission using ML. **Methods**: Retrospective cohort study was carried out with 9,080 patients under 18 years old admitted to a tertiary university hospital. Demographic, clinical, and biochemical data were collected from electronic databases. We randomly divided the dataset into training (75%) and testing (25%), applied downsampling, repeated cross-validation with five folds and ten repetitions, and the hyperparameter was optimized of each technique using a grid search via racing with ANOVA models. We applied six ML classification algorithms to build the predictive models, including classification and regression tree (CART), random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), decision tree and logistic regression (LR). The area under the receiver operating curve (AUC), sensitivity, specificity, Youden's J-index and accuracy were used to evaluate the performance of each model. **Results**: The 30-day hospital readmissions rate was 9.5%. Some algorithms presented similar AUC, both in the dataset training and in the dataset testing, such as XGBoost, RF, GBM and CART. Considering the Youden's J-index, the algorithm that presented the best index was XGBoost with bagging imputation, with AUC of 0.814 (J-index of 0.484). Cancer diagnosis, age, red blood cells, leukocytes, red cell distribution width and sodium levels, elective admission, and multimorbidity were the most important characteristics to classify between readmission and non-readmission groups. **Conclusion:** Machine learning approaches, especially XGBoost, can predict potentially avoidable 30-day pediatric hospital readmission into tertiary assistance. If implemented in the computer hospital system, our model can help in the early and more accurate identification of patients at risk, targeting health strategic interventions.

Keywords: Machine Learning; Hospital Readmission; Pediatrics; Risk factors; Clinical Decision Rules

Introduction

Pediatric hospital readmission is a complex, multifactorial and highly prevalent outcome, with negative impacts both for the individual 1–5 and for the health system 6–8. Identifying patients at higher risk of readmission is one of the first steps to reduce pediatric hospital readmission rates and minimize their deleterious effects. This identification is also required to allow the targeting of preventive interventions to patients with greater need, and also to optimize care and the allocation of financial resources.

Recent technological advances have enabled a greater information flow, improving the potential medical research and clinical care, in order to support the diagnosis and diseases prediction and personalization healthcare 9–11. Studies have shown that the use of artificial intelligence techniques can be promising to predict pediatric hospital readmission 12–15. Machine learning can assess highly complex outcomes quickly and accurately for which human perception may fail 11,16. However, this context is still incipient in the health care, and it is not a applicable reality in many hospitals. Many health professionals are still not familiar with the use of these techniques and do not know how to apply them in hospital practice 17. Moreover, few studies have been dedicated to predicting potentially avoidable pediatric hospital readmissions using machine learning 12–15 and the few models published previously had some limitations, such as the impossibility of interpretation 12,13, which make their application in clinical practice difficult. For health outcomes, the easy understanding of a model is essential, given the underlying complexity of clinical outcomes and the potential negative impact of health wrong decisions18.

In this sense, there is a lack of studies seeking the development of predictive models for pediatric hospital readmission, with a good performance, using predictors that are easy to obtain, and with biological plausibility. These models can facilitate the rational interpretation of clinical factors contribution and the path to decision. Therefore, the aim of our study was to develop predictive models capable to identify children and adolescents at high risk of potentially avoidable 30-day readmission.

Methods

Study design and Settings

We conducted a retrospective cohort study at a tertiary university hospital with data collected from January 1st, 2014 to December 31st, 2018, including all hospitalized children and adolescents aged 0 to 18 years old. All data were retrieved from the electronic databases. We excluded admissions that resulted in hospital death (not at risk for readmission outcome) and discharges against medical advice (not at the opportunity to implement care plan and discharge instruction) (Figure 1). The University's Ethics Committee approved this study (CAAE 51706221.3.0000.5152, protocol number: 5.003.236).

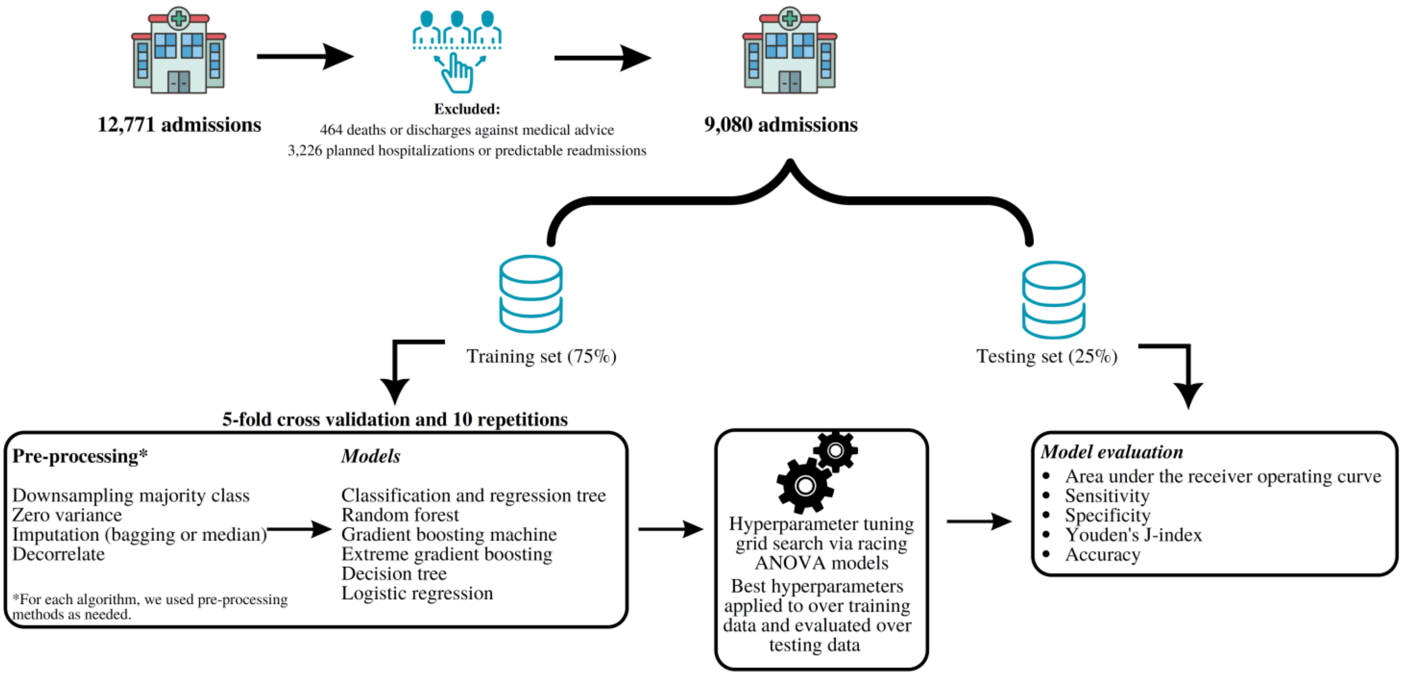


Figure 1 Study flow diagram

Predictors selection

Demographic and clinical data

We collected demographic (age, sex, ethnicity) and clinical data, such as diagnoses, multimorbidity, length of hospital stay (number of days), admission type (elective or urgent/emergency), and wards (clinical, urgent/emergency/surgical and newborns).

We categorized the diagnoses into new variables in order to show the presence or absence of a diagnosis of sepsis, diabetes, high blood pressure, neoplasia, heart, and kidney diseases. Multimobidity was defined as "the co-occurrence of multiple chronic or acute diseases and medical conditions within one person" 19. To quantify multimorbidity, we counted all the diagnoses of each patient according to the International Statistical Classification of Diseases (ICD) and Health-Related Problems.

Biochemical exams

We collected the biochemical tests: hematocrit, hemoglobin, red blood cell, leukocytes, red cell distribution width (RDW), C-reactive protein (CRP), and sodium. All blood tests were performed at a single laboratory and hematological parameters were analyzed using an automated Sysmex XN-3000™ hematology analyzer. CRP was measured by methods immunoturbidimetry using Cobas® 6000 analyzer, and sodium measured by potentiometric methods. We obtained last known laboratory values before discharge. However, since different physicians request different biochemical tests depending on the patient's clinical condition, the frequency of the exams could vary.

Outcome

The outcome of our study was potentially avoidable 30-day readmissions, considered as a new admission in this short period after the immediately previous hospital discharge. We excluded all unavoidable readmissions, i.e, all patients admitted to wards with planned hospitalizations (obstetrics, gynecology and transplant) or with predictable admissions, such as labour/delivery, and chemotherapy or radiotherapy treatments (ambulatory care), in order to achieve only avoidable readmissions.

Statistical Analysis

We randomly divided the dataset into training (75%) and testing (25%), using stratified sampling, to maintain proportionality between training and testing. The training dataset was used for the determination of hyperparameters and training processes using repeated cross-validation (RCV) process with five folds and ten repetitions, while the test dataset was only used in the final evaluation of the model's performance. To overcome the class imbalance, we used the downsampling method for each fold cross-validation, aiming to randomly reduce the size of the majority class, becoming equal the proportions of the classes. Each 5-cross-validation step consists in the following: (1) apply pre-processing techniques (including downsampling) to 4 data folds (alternatively) to be used for training; (2) fit model to pre-processed folds; (3) apply fitted model to hold-out fold and compute model evaluation metrics. The hyperparameters of each technique were optimized through a grid search via racing with ANOVA models, using 3500 combinations of different hyperparameter values. Subsequently, the best hyperparameter set found was used to train a model with the entire training set, computing the performance metrics (all dataset training) and, later, to the test set.

We selected several machine-learning techniques to build predictive models, including classification and regression tree (CART), random forest (RF), gradient boosting machine (GBM), extreme gradient boosting (XGBoost), decision tree, and logistic regression (LR). For each algorithm, we used pre-processing methods as needed. In the CART algorithm, we removal of zero variance (remove variables that are highly sparse and unbalanced). We used two imputation methods for RF and GBM, bagging (creates bagged tree models to impute missing data), and median (replaces missing values with the median of the training set). For XGBoost and LR we use two methods using decorrelate (remove variables that have large absolute correlations with other variables) and imputation using bagging and median. For the decision tree, we made three methods: removal of zero variance, decorrelate, and imputation via bagging and via median. In order to compare the performance of the models, we computed the area under the receiver operating curve (AUC). Other performance measures were also calculated, such as sensitivity and specificity, Youden's J-index, and accuracy (Figure 1).

Since the correct interpretation of a prediction model based on black-box machine learning models is a challenge, it is a crucial point for health professionals. We used methods to identify the contribution of each attribute in the prediction model and thus ensure the interpretability of the generated models. We used SHApley Additive exPlanation (SHAP) to provide consistent and accurate assignment values for each feature within each forecast model by building a graphical summary of the predictors[20]. The SHAP values evaluate the importance of the output resulting from adding feature A to all combinations of feature other than A[20].

We showed the descriptive data as proportions or means (± standard deviation, SD) or medians (with interquartile ranges - IQR). For the statistical analysis, we used the R Project (version 4.0.3), the RStudio (version 4.0.2), and considered the 95% confidence intervals (95% CI).

Results

Of the 9,080 patients included in the study, 60.3% (5,475) were male, with a mean age of 5.7 years and mean length of hospital stay of 15 days. The frequency of potentially avoidable 30-day readmission was 9.5% (862), of which 91.8% (791) had an Urgent or Emergency admission type and longer length of hospital stay (19.3 days) (Table 1).

Table 1 Demographic, clinical and biochemical variables by potentially avoidable 30-day readmission

Predictors	**All** (9,080)	**30-day readmissions, %(n)**	
		No 90.5 (8,218)	**Yes 9.5 (862)**
Sex, male [%(n)]	60.3 (5,475)	59.9 (4,926)	63.7 (549)
Age, years [means (SD)]	5.7 (6.0)	5.7 (6.1)	6.5 (5.3)
Admission type, Urgent or Emergency [%(n)]	83.6 (7,587)	82.7 (6,796)	91.8 (791)
Multimorbidity [means (SD)]	5.2 (4.9)	5.16 (4.8)	6.0 (5.9)
Length of hospitalization, days [means (SD)]	15.3(23.2)	14.8 (22.7)	19.3 (27.5)
Hematocrit, % [means (SD)]	36.7(8.8)	37.2 (8.8)	32.3(6.8)
Hemoglobin, g/dL [means (SD)]	12.6(3.2)	12.8 (3.2)	10.9 (2.3)
Red blood cell, million/mm^3 [means (SD)]	4.2(0.8)	4.26 (0.8)	3.8 (0.7)
Leukocytes, million/mm^3 [means (SD)]	10.9 (5.38)	11.2 (5.2)	8.9 (6.4)
Red Cell Distribution Width, % [means (SD)]	14.9(2.6)	14.8 (2.6)	15.1 (2.7)
C-reactive protein, mg/dL [means (SD)]	4.2(5.5)	4.1 (5.5)	4.6 (5.8)
Sodium, mEq/L [means (SD)]	138.6(3.6)	138.6 (3.6)	138.1 (3.2)
Diagnosis of cancer [%(n)]	6.7 (607)	3.0 (249)	41.5 (358)
Diagnosis of sepsis [%(n)]	5.0 (450)	4.8 (391)	6.8 (59)
Diagnosis of kidney diseases [%(n)]	3.8 (347)	3.5 (284)	7.3 (63)
Diagnosis of heart diseases [%(n)]	17.1 (1,549)	17.5 (1,438)	12.9 (111)
Diagnosis of hypertension [%(n)]	3.0 (269)	2.8 (228)	4.8 (41)
Diagnosis of diabetes [%(n)]	1.9 (169)	2.0 (161)	0.9 (8)

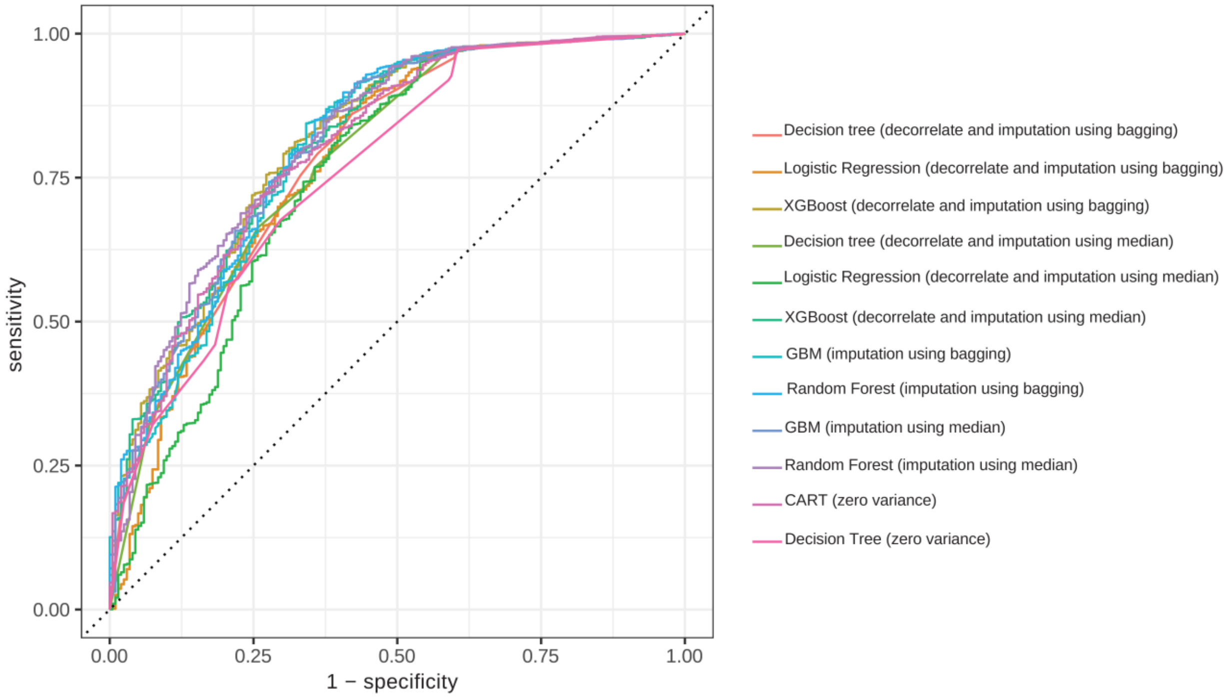
Table 2 shows the metrics of all models tested, both the average of the metrics found in the dataset training and the metrics found when applying the best model found in the dataset testing. Some algorithms presented similar AUC, both in the dataset training and in the dataset testing such as XGBoost, RF, GBM and CART. Considering the Youden's J-index, the algorithm that presented the best index was XGBoost with bagging imputation, AUC of 0.807 (J-index of 0.467) and AUC of 0.814 (J-index of 0.484) in the dataset training and testing, respectively. We can see that the imputation using median was a little different from the results using bagging imputation.

Table 2 Performance metrics for machine learning models

Methods	Dataset Training*					Dataset Testing				
	AUC	**Specificity**	**Sensitivity**	**Youden J Index**	**Accuracy**	**AUC**	**Specificity**	**Sensitivity**	**Youden J Index**	**Accuracy**
XGBoost (decorrelate and imputation using bagging)	0.807	0.638	0.829	0.467	0.811	0.814	0.668	0.816	0.484	0.803
RF (imputation using median)	0.809	0.664	0.795	0.459	0.782	0.815	0.663	0.801	0.464	0.789
XGBoost (decorrelate and imputation using median)	0.808	0.639	0.827	0.466	0.809	0.810	0.604	0.840	0.444	0.819
GBM (imputation using median)	0.801	0.611	0.843	0.454	0.821	0.810	0.619	0.853	0.472	0.833
RF (imputation using bagging)	0.810	0.671	0.790	0.461	0.779	0.809	0.688	0.782	0.471	0.774
CART (zero variance)	0.811	0.655	0.806	0.461	0.792	0.803	0.673	0.771	0.444	0.732
GBM (imputation using bagging)	0.803	0.632	0.818	0.449	0.800	0.801	0.658	0.813	0.472	0.800
DT (decorrelate and imputation using bagging)	0.783	0.631	0.792	0.423	0.777	0.785	0.604	0.823	0.427	0.804
DT (decorrelate and imputation using median)	0.784	0.653	0.777	0.430	0.765	0.781	0.653	0.740	0.393	0.732
LR (decorrelate and imputation using bagging)	0.772	0.538	0.881	0.419	0.848	0.778	0.599	0.853	0.452	0.830
DT (zero variance)	0.781	0.639	0.776	0.415	0.763	0.762	0.703	0.676	0.379	0.679
LR (decorrelate and imputation using median)	0.764	0.502	0.910	0.413	0.871	0.755	0.460	0.934	0.395	0.892

* Average results obtained from repeating cross-validation with the best hyperparameters found.
Legend: CART: classification and regression tree; RF: random forest; XGBoost: extreme gradient boosting; GBM: gradient boosting machine; DT: decision tree; LR: logistic regression; AUC: Area Under the Receiver Operating Characteristic Curve; RCV: repeated cross-validation

In Figure 2 reports the predictive performance for the experiments, the algorithm that presented the best performance was XGBoost with imputation via bagging with AUC of 0.814.



Legend: CART: classification and regression tree; XGBoost: extreme gradient boosting; GBM: gradient boosting machine

Figure 2 Comparison of area under the ROC for machine learning models to identify children and adolescents at high risk of potentially avoidable 30-day readmission

Figure 3 shows the SHAP feature importance, calculated as the average absolute SHAP value per feature, for the XGBoost model trained before for predicting potentially avoidable 30-day readmissions. Features with large absolute Shapley values were important. Cancer diagnosis, age, levels of red blood cells, leukocytes and RDW were the predictors that contributed most to the model.

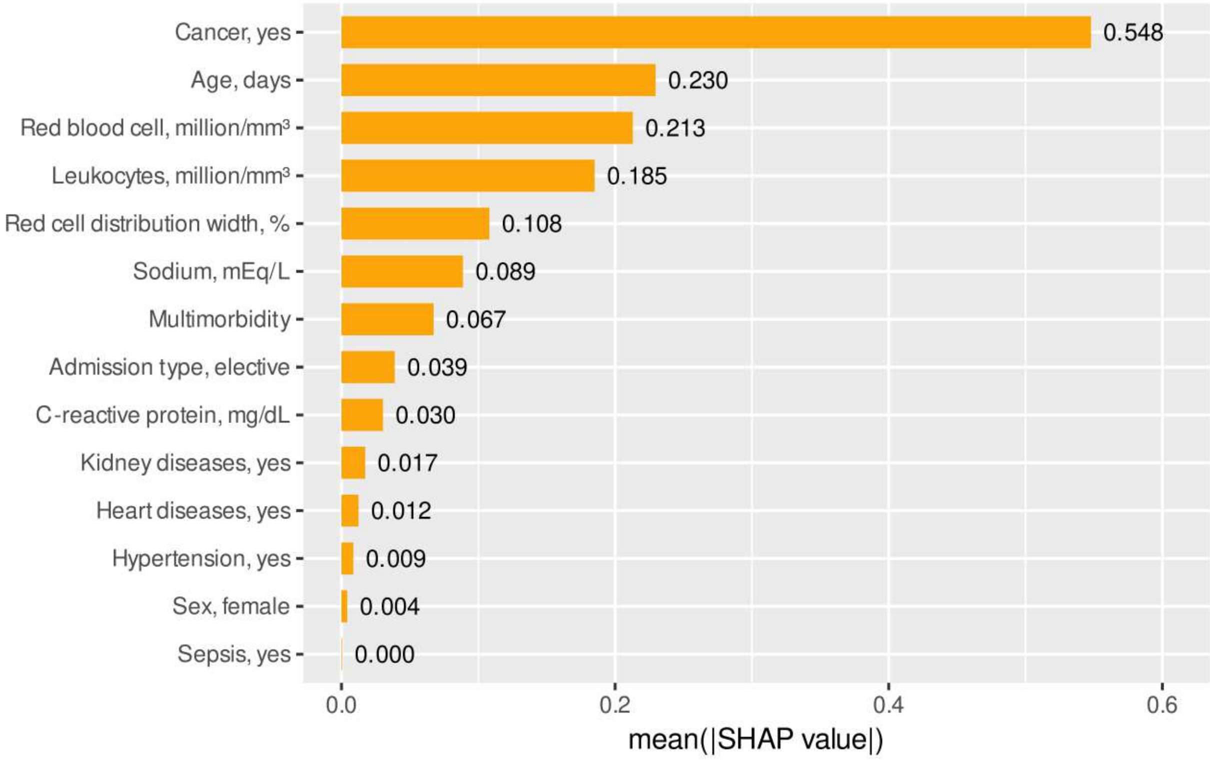


Figure 3 SHAP feature importance measured as the mean absolute Shapley values for the XGBoost model (decorrelate and imputation using bagging)

SHAP values were calculated and plotted (Figure 4). As shown, cancer diagnosis, age, levels of red blood cells, leukocytes, RDW and sodium, elective hospitalization, and multimorbidity were the most important features to classified between readmission and non-readmission groups. Cancer diagnosis has positive SHAP values, that is, shows a positive correlation. It means that patients diagnosed with cancer are more likely to be readmitted, as well as higher age patients, with lower levels of red blood cells, leukocyte, and sodium, higher levels of RDW and greater multimorbidity amount have a greater risk of readmission. On the other hand, patients with elective admission have a lower readmission risk.

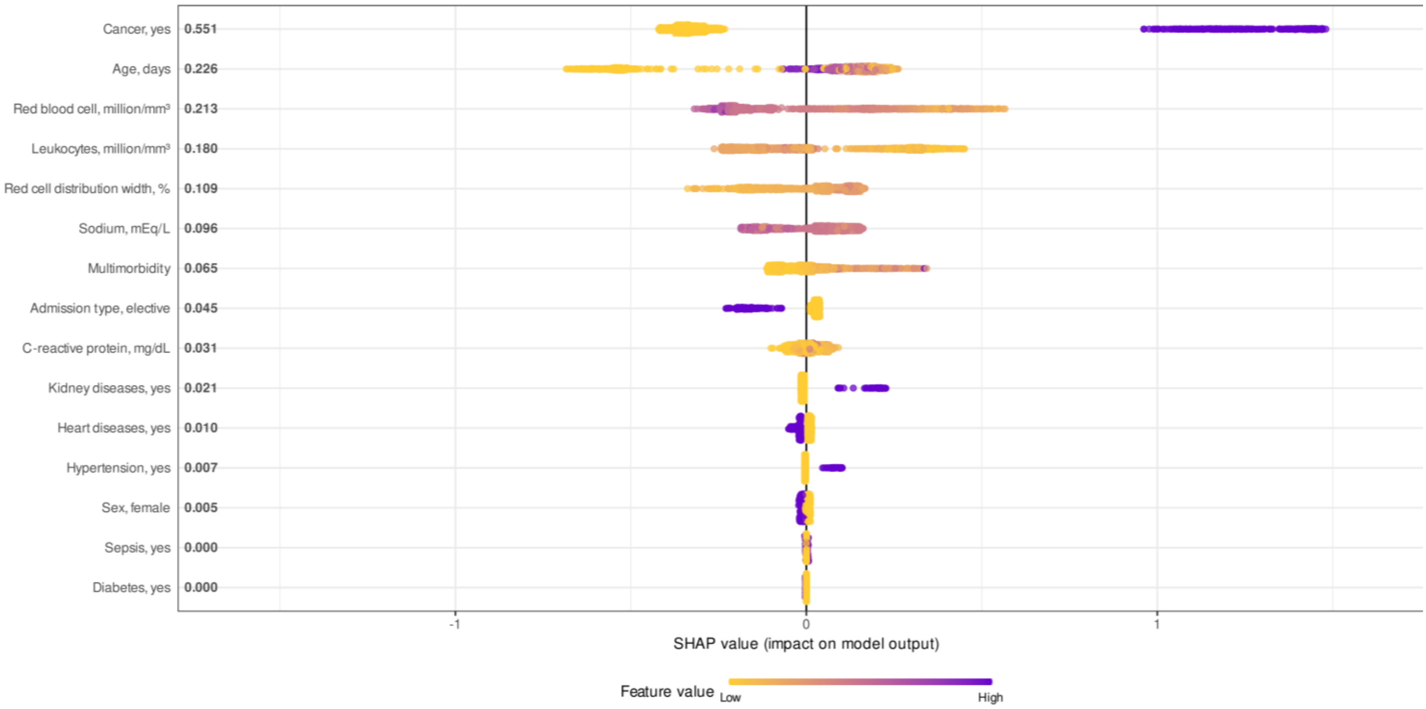
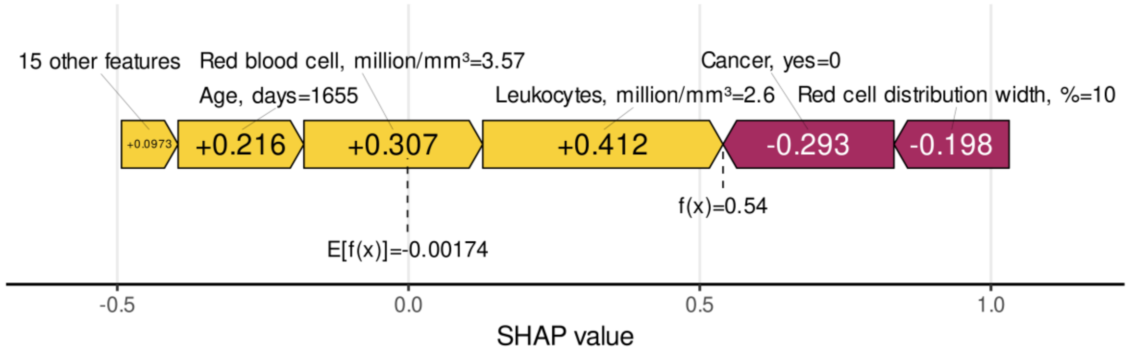


Figure 4 Impact that each feature had on the result of the full model estimated using SHapley Additive exPlanations (SHAP) values for XGBoost model (decorrelate and imputation using bagging) using the full training dataset. The plot rank feature by the sum of the magnitudes of the SHAP values across all samples. The color represents the feature value (purple high, yellow low). The x-axis measures the impact on the model output (right positive, left negative)

In order to explore the contribution of predictors and the clinical model application, we randomly selected two patients from the dataset (Figure 5). The results suggest that low levels of leukocytes and red blood cells (increase risk factors) and no cancer diagnosis (decrease risk factor) were the top three contributors to the readmission prediction.

A



B

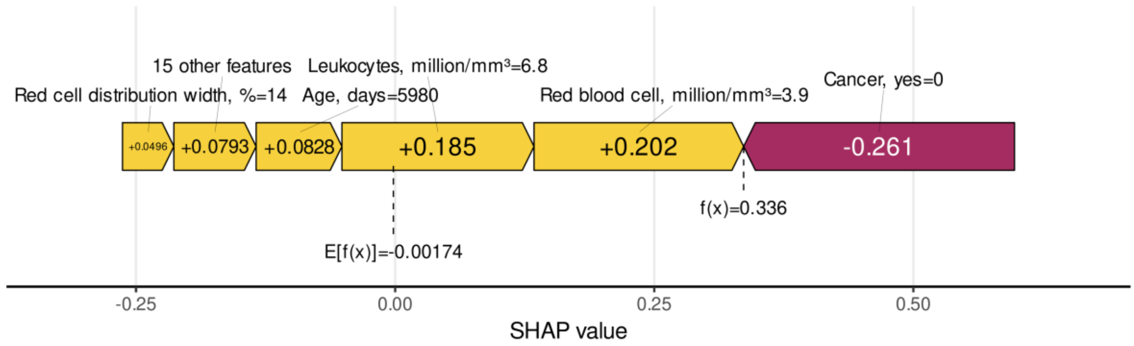


Figure 5 SHAP force plot for explaining of individual's prediction results using the XGBoost model (decorrelate and imputation using bagging). The yellow and red bars represent risk and protective factors, respectively, longer bars representing greater resource importance

Discussion

In this retrospective cohort, we developed and validated ML models, using demographic, clinical and biochemical variables to predict potentially avoidable 30-day pediatric hospital readmission. Some models had similar AUC, such as XGBoost, RF, GBM and CART, however, XGBoost obtained the best perfomance (AUC of 0.814, specificity of 66%, sensitivity of 81% and Youden's J-index of 0.484). Cancer diagnosis, age, levels of red blood cells, leukocytes, RDW, sodium, elective admission, and multimorbidity were the predictors with the greatest impact on potentially avoidable 30-day readmission. To the best of our knowledge, our study is among the first to apply machine learning techniques to predict potentially avoidable 30-day readmission on the pediatric population. We developed a prediction interpretable model with good predictive capacity, which allow to understand the contributing factors and the paths to the rational decision. Predictors with biological plausibility and easily available in most health services were used.

Pediatric hospital readmissions are harmful to patients and families[2,21] overload the health system[6–8] and are a recurrent outcome, ranging from 4 to 29.5%[15,22–24]. The frequency of 30-day pediatric hospital readmissions in our study was 9.50%, comparable to other studies[22,24–26]. Determining who is eligible to be discharged is a challenge that demands time and effort from the medical team. Tools capable of helping to identify patients at higher risk of readmission have been the subject of studies. Recently, studies using traditional statistics revealed that two tools developed and widely used to predict readmissions in adult patients (LACE index and HOSPITAL score) can predict hospital readmissions in the pediatric population[27,28]. However, the hospital readmission is a complex, multifactorial and non-stationary process, with factors that human perception may not fully understand, making an accurate prediction difficult. Studies have showed that machine learning techniques can favor the early identification of patients with a higher readmission risk, improve the prediction, and consequently, enable the personalization care. There are still few studies that apply these techniques to predict potentially avoidable 30-day readmission on the pediatric population[12–15]. The found models had limited usefulness and unsatisfactory performance. Some previous studies had limited predictive capacity[12,13], excluded pediatric subgroups, such as neonates[15], used predictors derived from the adult population, such as the Charlson Comorbidity Index[15], or yet, they were difficult to understand, not being interpretable[12,13].

The model generated by XGBoost showed a good accuracy (AUC of 0.814), specificity (66%), sensitivity (81%) and Youden's J-index (0.484) to identify the potentially avoidable 30-day readmission risk in a pediatric population. In the literature, the predictive capacity of models employing ML techniques to predict pediatric readmissions shows substantial variation. One study found an AUC of 0.65 with the Naive Bayes approach[12] and gradient boosted tree model[13], another found an AUC of 0.73 using Support Vector Machines with Polynomial Kernel for at-discharge models[14] and even

AUC of 0.81 with XGBoost for unplanned readmissions in 30 days[22]. The complexity of the hospital readmission process explains these differences found in the models' ability to discriminate, since a wide list of factors and interactions are involved in the causes for which a patient may be readmitted. Besides, the difficulty in finding good models is because of the metric used, the analysis methods and predictors used, as well as the variability of health services. Our study showed a model that performed better than most 30-day pediatric readmission models. We sought to consider the previous limitations of pediatric prediction algorithms, using only readily available predictors at no additional cost and relevant to the pediatric population.

For health outcomes, the interpretability of the model is as important as its performance. In this sense, we applied techniques to “open the black-box” in order to make it possible to understand the contribution of each attribute to the prediction of hospital readmission. We showed patients diagnosed with cancer, higher age, lower red blood cells, leukocytes, sodium levels and a higher RDW level, as well as greater multimorbidity, presented a higher readmission risk. Furthermore, on individual exploration, we showed examples that the top-three contributors to prediction were lower red blood cells and leukocytes levels (increase risk factors) and no cancer diagnosis (decrease risk factor).

Cancer diagnosis was an important feature of our prediction model. Cancer is a complex disease, with multifactorial etiology, whose treatment can lead to multiple admissions because of infectious complications, hydroelectrolytic abnormalities and decreased functional capacity[29,30].

Age also had a substantial contribution. Studies have pointed out that pediatric readmission rates vary according to age[24,31,32]. A study that evaluated the risk factors associated with unplanned pediatric hospital readmission within 30 days revealed that patients older than 13 years were almost 1.5 times more likely to be readmitted compared to younger patients[33]. The prevalence of children with chronic conditions is increasing due to improved survival in the neonatal period and advances in medical care, leading to additional medical needs and future hospitalizations[34].

Other predictors capable of modifying the readmission risk were altered levels of red blood cells, leukocytes, RDW and sodium. Low levels of red blood cells may be a reflection of anemia, which can be negatively associated with pediatric hospital readmission[35], as well as cause impairments in cognitive function, growth and psychomotor development[36]. The change in RDW values (anisocytosis) beside to reflecting a profound dysregulation of erythrocyte homeostasis attributed to underlying metabolic disorders[37]. It may show of the underlying general health status, having been appointed as a marker of chronic inflammation and disease severity[37], associated with hospital mortality[38,39] and longer length of hospital stay[39,40]. Low levels of leukocytes (leukopenia) increases the susceptibility to develop infections, evidencing a fragility of the immune system, because of their inadequate response to fight infections, it is associated with hematological complications (sepsis and septic shock)[41], mortality[41,42] and readmission[42]. Our study also identified that sodium levels can change the risk of readmission. Other studies[29,43,44] have already identified as a predictor of hospital readmission hyponatremia. It is one of the most common electrolyte disturbances in hospitalized children, and excessive secretion of antidiuretic hormone when under stress, pain or nausea/vomiting, associated with water retention and increased sodium reabsorption are the principal causes of hyponatremia in children[45,46].

On the other hand, our study revealed that the greater readmission risk with the higher multimorbidity. Multimorbidity has been identified as a proxy for patient severity related to negative hospital outcomes and increased hospital costs. Other studies evaluating pediatric hospital readmission found out that multimorbidity is one of the risk factors for readmission[13,24,32].

While certain predictors of hospital readmission cannot be easily modified, our model allows for the early identification of at-risk patients. This enables improved coordination of discharge planning and targeted interventions, prioritizing those patients most in need of special care. By doing so, we can reduce hospital readmissions and, ultimately, lower health costs. To this end, we encourage the integration of predictive algorithms for readmission risk in electronic medical systems, providing additional tools to assist medical teams in identifying at-risk patients and planning interventions. In future studies, it would be beneficial to explore nutritional data in order to improve our understanding of its impact, and to raise awareness among health professionals about the importance of nutritional screening.

There are some limitations to this study. First, the extrapolation of the results should be done with caution, as our study was carried out with pediatric patients from a one hospital. Additionally, our population belongs to free medical care offered by the government with a specific dynamic of readmissions, different from a private health system, and may serve as a reference for other studies with a similar population. Second, our estimates are conservative, as we cannot analyze the number of readmissions that may have occurred by transfer to other institutions. Third, we use electronic medical record data, susceptible to random human errors in data entry. Fourth, even though our study evaluated a broad set of clinical variables, other conditions may be associated with readmission, such as social factors (income, education, living conditions) unavailable in our dataset.The study has also strength points. As mentioned, we collected wide range of variables and all available data for a large period. Even with some absence of data and with a more conservative strategy, our model was successful.

Conclusion

Machine learning approaches, especially XGBoost, can predict avoidable 30-day pediatric hospital readmission into tertiary assistance (AUC of 0.814, specificity of 66%, sensitivity of 81% and Youden's J-index of 0.484). Cancer diagnosis, age, levels of red blood cells, leukocytes, RDW, and sodium, elective admission and multimorbidity were the predictors with the greatest impact on potentially avoidable 30-day readmission. Our model can help in early and accurate identification of patients at risk, facilitating the strategic targeting of interventions, aiming to improve, optimize care and improve the experiences of patients and families, as well as reduce healthcare costs.

Acknowledgments
The technology and information center (CTI) team of the Clinical Hospital of the Federal University of Uberlandia for helping to databases extraction. Nayara. C. Silva received a doctoral fellowship from the "The Foundation for Research Support of the State of Minas Gerais" (FAPEMIG - Fundação de Amparo à Pesquisa do Estado de Minas Gerais). André R. Backes gratefully acknowledges the financial support of CNPq (Grant #307100/2021-9). Marcelo K. Albertini gratefully acknowledges the financial support of CNPq (Grant #406418/2021-7).

Funding sources
This study received financial support from "The Foundation for Research Support of the State of Minas Gerais" (FAPEMIG - Fundaç˜ ao de Amparo ` a Pesquisa do Estado de Minas Gerais) in a public call process number APQ-01331–21. The funder had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

References

1. Delvecchio E, Salcuni S, Lis A, Germani A, Di Riso D. Hospitalized Children: Anxiety, Coping Strategies, and Pretend Play. Front Public Heal. 2019;7(September):1-8. doi:10.3389/fpubh.2019.00250

2. Silva VLS da, França GVA de, Munhoz TN, et al. Hospitalization in the first years of life and development of psychiatric disorders at age 6 and 11: a birth cohort study in Brazil. Cad Saude Publica. 2018;34(5):1-13. doi:10.1590/0102-311x00064517

3. Cahayag V. Hospitalization and Child Development: Effects on Sleep, Developmental Stages, and Separation Anxiety. Nurs | Sr Theses. 2020;17. doi:doi.org/10.33015/dominican.edu/2020.NURS.ST.09

4. Pufal EC, Müller AB, Bandeira PFR, Valentini NC. Motor development in the hospitalized infant and its biological and environmental characteristics. Clin Biomed Res. 2018;38(1):66-73. doi:10.4322/2357-9730.75638

5. Rashikj Canevska O. The Impact of Hospitalization on Psychophysical Development and Everyday Activities in Children. Annu Fac Philos Skopje. 2018;71(January 2018):465-470. doi:10.37510/godzbo1871465rc

6. Markham JL, Hall M, Gay JC, Bettenhausen JL, Berry JG. Length of Stay and Cost of Pediatric Readmissions. Pediatrics. 2018;141(4):e20172934. doi:10.1542/peds.2017-2934

7. Gay JC, Agrawal R, Auger KA, et al. Rates and Impact of Potentially Preventable Readmissions at Children's Hospitals. J Pediatr. 2015;166(3):613-619.e5. doi:10.1016/j.jpeds.2014.10.052

8. Kane JM, Hall M, Cecil C, et al. Resources and Costs Associated With Repeated Admissions to PICUs. Crit Care Explor. 2021;3(2):e0347. doi:10.1097/CCE.0000000000000347

9. Harrison CJ, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction to natural language processing. BMC Med Res Methodol. 2021;21(1):158. doi:10.1186/s12874-021-01347-1

10. Sidey-Gibbons JAM, Sidey-Gibbons CJ. Machine learning in medicine: a practical introduction. BMC Med Res Methodol. 2019;19(1):64. doi:10.1186/s12874-019-0681-4

11. Sarker IH. Machine Learning: Algorithms, Real-World Applications and Research Directions. SN Comput Sci. 2021;2(3):160. doi:10.1007/s42979-021-00592-x

12. Wolff P, Graña M, Ríos SA, Yarza MB. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. Biomed Res Int. 2019;2019:1-9. doi:10.1155/2019/8532892

13. Zhou H, Albrecht MA, Roberts PA, Porter P, Della PR, Della PR. Using machine learning to predict paediatric 30-day unplanned hospital readmissions: A case-control retrospective analysis of medical records, including written discharge documentation. Aust Heal Rev. 2021;45(3):328-337. doi:10.1071/AH20062

14. Symum H, Zayas-Castro J. Identifying Children at Readmission Risk: At-Admission versus Traditional At-Discharge Readmission Prediction Model. Healthcare. 2021;9(10):1334. doi:10.3390/healthcare9101334

15. Ehwerhemuepha L, Gasperino G, Bischoff N, Taraman S, Chang A, Feaster W. HealtheDataLab – a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. BMC Med Inform Decis Mak. 2020;20(1):115. doi:10.1186/s12911-020-01153-7

16. Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. Futur Healthc J. 2019;6(2):94-98.

17. Handelman GS, Kok HK, Chandra R V, Razavi AH, Lee MJ, Asadi H. eDoctor: Machine learning and the future of medicine. J Intern Med. Published online 2018:1-17. doi:10.1111/joim.12822

18. Jovanovic M, Radovanovic S, Vukicevic M, Poucke S Van, Delibasic B. Artificial Intelligence in Medicine Building interpretable predictive models for pediatric hospital readmission using Tree-Lasso logistic regression. Artif Intell Med. 2016;72:12-21. doi:10.1016/j.artmed.2016.07.003

19. Van Den Akker M, Buntinx F, Knottnerus JA. Comorbidity or multimorbidity: What's in a name? A review of literature. Eur J Gen Pract. 1996;2(2):65-70. doi:10.3109/13814789609162146

20. Lundberg SM, Lee SI. A unified approach to interpreting model predictions. Adv Neural Inf Process Syst. 2017;2017-Decem(Section 2):4766-4775.

21. Coyne I. Children's Experiences of Hospitalization. J Child Heal Care. 2006;10(4):326-336. doi:10.1177/1367493506067884

22. Taylor T, Altares Sarik D, Salyakina D. Development and Validation of a Web-Based Pediatric Readmission Risk Assessment Tool. Hosp Pediatr. 2020;10(3):246-256. doi:10.1542/hpeds.2019-0241

23. Toomey SL, Peltz A, Loren S, et al. Potentially Preventable 30-Day Hospital Readmissions at a Children's Hospital. Pediatrics. 2016;138(2):e20154182. doi:10.1542/peds.2015-4182

24. Ehwerhemuepha L, Finn S, Rothman M, Rakovski C, Feaster W. A Novel Model for Enhanced Prediction and Understanding of Unplanned 30-Day Pediatric Readmission. Hosp Pediatr. 2018;8(9):578-587. doi:10.1542/hpeds.2017-0220

25. Zhou H, Della P, Roberts P, Porter P, Dhaliwal S. A 5-year retrospective cohort study of unplanned readmissions in an Australian tertiary paediatric hospital. Aust Heal Rev. 2019;43(6):662-671. doi:10.1071/AH18123

26. Berry JG, Toomey SL, Zaslavsky AM, et al. Pediatric readmission prevalence and variability across hospitals. JAMA - J Am Med Assoc. 2013;309(4):372-380. doi:10.1001/jama.2012.188351

27. Han TS, Fluck D, Fry CH. Validity of the LACE index for identifying frequent early readmissions after hospital discharge in children. Eur J Pediatr. 2021;180(5):1571-1579. doi:10.1007/s00431-021-03929-z

28. Silva NC da, Keese MA, Backes AR, Pena G das G. Validation of the HOSPITAL score as predictor of 30 - day potentially avoidable readmissions in pediatric hospitalized population : retrospective cohort study. Eur J Pediatr. Published online 2023. doi:10.1007/s00431-022-04795-z

29. Donze J, Aujesky D, Williams D, Schnipper JL. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. JAMA Intern Med. 2013;173(8):632-638. doi:10.1001/jamainternmed.2013.3023

30. Hoenk K, Torno L, Feaster W, et al. Multicenter study of risk factors of unplanned 30-day readmissions in pediatric oncology. Cancer reports (Hoboken, NJ). 2021;4(3):e1343. doi:10.1002/cnr2.1343

31. Feudtner C, Levin JE, Srivastava R, et al. How Well Can Hospital Readmission Be Predicted in a Cohort of Hospitalized Children? A Retrospective, Multicenter Study. Pediatrics. 2009;123(1):286-293. doi:10.1542/peds.2007-3395

32. Zhou H, Roberts PA, Dhaliwal SS, Della PR. Risk factors associated with paediatric unplanned hospital readmissions: a systematic review. BMJ Open. 2019;9(1):e020554. doi:10.1136/bmjopen-2017-020554

33. Zhou H, Della PR, Porter P, Roberts PA. Risk factors associated with 30-day all-cause unplanned hospital readmissions at a tertiary children's hospital in Western Australia. J Paediatr Child Health. 2020;56(1):68-75. doi:10.1111/jpc.14492

34. Amritphale N, Amritphale A, Vasireddy D, Batra M, Sehgal M, Gremse D. Age Based Trends for Unplanned Pediatric Rehospitalizations in the United States. Cureus. 2021;5(13 (12)):e20181. doi:10.7759/cureus.20181

35. Kumar D, Swarnim S, Sikka G, et al. Factors Associated with Readmission of Pediatric Patients in a Developing Nation. Indian J Pediatr. 2019;86(3):267-275. doi:10.1007/s12098-018-2767-0

36. Aishvarya V, Roshine P, Elayarani M, Mahendravarman P, Saravanan S. A study on impact of iron deficiency on cognition and anthropometry in pediatrics and clinical pharmacist's intercession to provide awareness about iron deficiency in tertiary care teaching hospital. Int J Contemp Pediatr. 2021;8(11):1848-1854. https://www.ijpediatrics.com/index.php/ijcp/article/viewFile/4510/2881

37. Salvagno GL, Sanchis-Gomar F, Picanza A, Lippi G. Red blood cell distribution width: A simple parameter with multiple clinical applications. Crit Rev Clin Lab Sci. 2015;52(2):86-105. doi:10.3109/10408363.2014.992064

38. Silva NC da, Prestes IV, Gontijo WA, Pena G das G. High red blood cell distribution width is associated with a risk of short-term mortality in hospitalized surgical , but not clinical patients. Clin Nutr ESPEN. 2020;39:150-156. doi:10.1016/j.clnesp.2020.06.023

39. Kim, Da Ha, Ha EJ, Park SJ, Jhang WK. Evaluation of the usefulness of red blood cell distribution width in critically ill pediatric patients. Medicine (Baltimore). 2020;36(July). doi:10.1097/MD.0000000000022075

40. Said AS, Spinella PC, Hartman ME, et al. Red blood cell distribution width: biomarker for red cell dysfunction and critical illness outcome? Pediatr Crit Care Med. 2018;18(2):134-142. doi:10.1097/PCC.0000000000001017.Red

41. Belok SH, Bosch NA, Klings ES, Walkey AJ. Evaluation of leukopenia during sepsis as a marker of sepsis-defining organ dysfunction. PLoS One. 2021;16(6):e0252206. doi:10.1371/journal.pone.0252206

42. Shah J, Balasubramaniam T, Yang J, Shah PS. Leukopenia and Neutropenia at Birth and Sepsis in Preterm Neonates of < 32 Weeks ' Gestation. Am J Perinatol. 2022;39(9):965-972. doi:10.1055/s-0040-1721133

43. Corona G, Giuliani C, Parenti G, et al. The Economic Burden of Hyponatremia: Systematic Review and Meta-Analysis. Am J Med. 2016;129(8):823-835.e4. doi:10.1016/j.amjmed.2016.03.007

44. Donzé JD, Beeler PE, Bates DW. Impact of Hyponatremia Correction on the Risk for 30-Day Readmission and Death in Patients with Congestive Heart Failure. Am J Med. 2016;129(8):836-842. doi:10.1016/j.amjmed.2016.02.036

45. Karageorgos S, Kratimenos P, Landicho A, et al. Hospital-Acquired Hyponatremia in Children Following Hypotonic versus Isotonic Intravenous Fluids Infusion. Children. 2018;5(10):139. doi:10.3390/children5100139

46. Zieg J. Pathophysiology of Hyponatremia in Children. Front Pediatr. 2017;5(October). doi:10.3389/fped.2017.00213

5 CONSIDERAÇÕES FINAIS

Nosso estudo desenvolveu modelos de predição para readmissões hospitalares potencialmente evitáveis de 30 dias para população pediátrica internada em um hospital de nível terciário. Utilizando abordagens tradicionais, validamos o escore HOSPITAL, inicialmente desenvolvido e validado para população adulta. O escore HOSPITAL mostrou boa capacidade para identificar risco de readmissão potencialmente evitável em 30 dias em uma população pediátrica em diferentes contextos clínicos e diagnósticos (AUC=0,80). Portanto, nosso estudo ampliou a utilidade do escore HOSPITAL estendendo sua utilização também para a população pediátrica.

Além disso, levando em consideração a complexidade do processo de readmissão hospitalar de natureza não estacionária, podendo ser afetado por fatores anteriores e também por fatores que ocorrem durante a internação, buscamos desenvolver uma ferramenta de predição empregando técnicas de aprendizado de máquina. Contudo, nossa prioridade era gerar modelos com boa capacidade de predição, facilmente compreendidos e implementados na prática clínica, utilizando preditores relevantes para a população pediátrica e incorporando dados nutricionais que muitas vezes são negligenciados.

Nosso estudo revelou que é possível utilizar técnicas de aprendizado de máquina para prever o risco de readmissão potencialmente evitável de 30 dias para população pediátrica de um hospital de nível terciário. Nossos modelos alcançaram boa discriminação e utilizaram atributos facilmente obtidos na prática clínica e prontamente disponíveis em muitos serviços de saúde. Além disso, mostramos a importância de valorizar os aspectos nutricionais dos pacientes, como fator de risco para readmissão hospitalar.

Portanto, podemos perceber que a capacidade de predição dos modelos gerados nesse estudo, seja utilizando técnicas tradicionais sejam empregando técnicas de aprendizado de máquina, foram semelhantes. Uma explicação para tal resultado pode ser a qualidade dos dados utilizados em nosso estudo, tínhamos muitos dados omissos, especialmente dados nutricionais, o que dificultou uma aplicação das técnicas de ML. No entanto, extraímos todos os dados disponíveis no sistema hospitalar e diversas tentativas foram realizadas para recuperar os dados ausentes.

6 PERSPECTIVAS FUTURAS

Apesar do progresso no desenvolvimento e na precisão dos modelos em prever e potencialmente reduzir a readmissão hospitalar pediátrica, esse desfecho continua sendo um processo complexo, que precisa ser melhor compreendido. Estudos futuros devem ser realizados com o intuito de validar os modelos desenvolvidos (validação interna prospectiva). Através do desenvolvimento e implementação de um programa de computador, por exemplo, que possibilite integrar esses algoritmos aos prontuários eletrônicos do Sistema de Internação Hospitalar usando dados em tempo real, gerando alertas para a equipe de saúde sobre o atual risco do paciente. Outra opção seria realizar estudos visando executar uma validação externa, verificando a generalização dos modelos desenvolvidos em outros serviços de saúde.

Além disso, novos estudos podem ser realizados visando a predição de outros desfechos em saúde que também geram impactos para os pacientes e para o serviço de saúde, tais como mortalidade hospitalar, diminuição do tempo de internação, diagnósticos oncológicos, cardíacos e renais, entre outros. Por fim, destacamos a importância de melhorar a qualidade dos registros de dados hospitalares, especialmente em relação aos dados nutricionais, muitas vezes negligenciados apesar de sua reconhecida importância clínica em diversos desfechos hospitalares negativos. Dados nutricionais de qualidade, poderiam possibilitar o desenvolvimento de estudos futuros empregando técnicas de *machine learning* visando identificar como os fatores nutricionais influenciam nos desfechos hospitalares. Modelos de predição utilizando técnicas de aprendizado de máquina são capazes de processar uma diversidade de dados de forma rápida e precisa, melhorar a disponibilidade e qualidade dos registros de dados de prontuários pode possibilitar uma expansão na construção de modelos de predição de readmissão hospitalar futuros bem como outros desfechos.

Machine Learning é um campo com grande potencial para aplicações em saúde, possibilitando tornar os processos médicos cada vez mais ágeis e eficientes. Os avanços tecnológicos, permitem integração dos sistemas com a análise de dados visando obter resultados clínicos efetivos, com otimização dos fluxos de trabalho e alocação eficiente dos recursos financeiros. Nesse sentido, estudos que visem desenvolver ferramentas capazes de contribuir para uma gestão organizada, possibilitando redirecionar os esforços das equipes de saúde e os recursos financeiros são de suma importância.

7 REFERÊNCIAS

AISHVARYA, V. et al. A study on impact of iron deficiency on cognition and anthropometry in pediatrics and clinical pharmacist ' s intercession to provide awareness about iron deficiency in tertiary care teaching hospital. **International Journal of Contemporary Pediatrics**, v. 8, n. 11, p. 1848–1854, 2021. https://doi.org/10.18203/2349-3291.ijcp20214158

AMRITPHALE, N. et al. Age Based Trends for Unplanned Pediatric Rehospitalizations in the United States. **Cureus**, v. 5, n. 13 (12), p. e20181, 2021. https://doi.org/10.7759/cureus.20181

ASHFAQ, A. et al. Readmission prediction using deep learning on electronic health records. **Journal of Biomedical Informatics**, v. 97, n. July, p. 103256, set. 2019. https://doi.org/10.1016/j.jbi.2019.103256

AUBERT, C. E. et al. Patterns of multimorbidity associated with 30-day readmission : a multinational study. **BMC Public Health**, v. 19, p. 1–8, 2019. https://doi.org/10.1186/s12889-019-7066-9

AWAD, A.; BADER–EL–DEN, M.; MCNICHOLAS, J. Patient length of stay and mortality prediction: A survey. **Health Services Management Research**, v. 30, n. 2, p. 105–120, 2017. https://doi.org/10.1177/0951484817696212

BAIG, M. M. et al. Machine Learning-based Risk of Hospital Readmissions : Predicting Acute Readmissions within 30 Days of Discharge. **2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)**, p. 2178–2181, 2019. https://doi.org/10.1109/EMBC.2019.8856646

CAHAYAG, V. Hospitalization and Child Development: Effects on Sleep, Developmental Stages, and Separation Anxiety. **Nursing | Senior Theses**, v. 17, 2020. https://doi.org/10.33015/dominican.edu/2020.NURS.ST.09

CORREIA, M. I. T. D.; PERMAN, M. I.; WAITZBERG, D. L. Hospital malnutrition in Latin America: A systematic review. **Clinical Nutrition**, v. 36, p. 958–967, 2017.

https://doi.org/10.1016/j.clnu.2016.06.025

CORREIA, M. I. T. D.; WAITZBERG, D. L. The impact of malnutrition on morbidity, mortality, length of hospital stay and costs evaluated through a multivariate model analysis. **Clinical Nutrition**, v. 22, n. 3, p. 235–239, 2003. https://doi.org/10.1016/S0261-5614(02)00215-7

COYNE, I. Children's Experiences of Hospitalization. **Journal of Child Health Care**, v. 10, n. 4, p. 326–336, 25 dez. 2006. https://doi.org/10.1177/1367493506067884

DAVENPORT, T.; KALAKOTA, R. The potential for artificial intelligence in healthcare. **Future Healthcare Journal**, v. 6, n. 2, p. 94–98, 2019. https://doi.org/10.7861/futurehosp.6-2-94

DELVECCHIO, E. et al. Hospitalized Children: Anxiety, Coping Strategies, and Pretend Play. **Frontiers in Public Health**, v. 7, n. September, p. 1–8, 2019. https://doi.org/10.3389/fpubh.2019.00250

DONZE, J. et al. Potentially Avoidable 30-Day Hospital Readmissions in Medical Patients. **JAMA Intern Med.**, v. 173, n. 8, p. 632–638, 2013. https://doi.org/10.1001/jamainternmed.2013.3023

EHWERHEMUEPHA, L. et al. A Novel Model for Enhanced Prediction and Understanding of Unplanned 30-Day Pediatric Readmission. **Hospital pediatrics**, v. 8, n. 9, p. 578–587, 2018a. https://doi.org/10.1542/hpeds.2017-0220

EHWERHEMUEPHA, L. et al. The Effect of Malnutrition on the Risk of Unplanned 7-Day Readmission in Pediatrics. **Hospital Pediatrics**, v. 8, n. 4, p. 207–213, 2018b. https://doi.org/10.1542/hpeds.2017-0195

EHWERHEMUEPHA, L. et al. A Statistical-Learning Model for Unplanned 7-Day Readmission in Pediatrics. **Hospital Pediatrics**, v. 10, n. 1, p. 43–51, 2020a. https://doi.org/10.1542/hpeds.2019-0122

EHWERHEMUEPHA, L. et al. HealtheDataLab – a cloud computing solution for data science and advanced analytics in healthcare with application to predicting multi-center pediatric readmissions. **BMC Medical Informatics and Decision Making**, v. 20, n. 1,

p. 115, 19 dez. 2020b. https://doi.org/10.1186/s12911-020-01153-7

FEUDTNER, C. et al. How Well Can Hospital Readmission Be Predicted in a Cohort of Hospitalized Children? A Retrospective, Multicenter Study. **Pediatrics**, v. 123, n. 1, p. 286–293, 2009. https://doi.org/10.1542/peds.2007-3395

FUTOMA, J.; MORRIS, J.; LUCAS, J. A comparison of models for predicting early hospital readmissions. **Journal of Biomedical Informatics**, v. 56, p. 229–238, ago. 2015. https://doi.org/10.1016/j.jbi.2015.05.016

GAY, J. C. et al. Rates and Impact of Potentially Preventable Readmissions at Children's Hospitals. **The Journal of Pediatrics**, v. 166, n. 3, p. 613- 619.e5, mar. 2015. https://doi.org/10.1016/j.jpeds.2014.10.052

HAN, T. S.; FLUCK, D.; FRY, C. H. Validity of the LACE index for identifying frequent early readmissions after hospital discharge in children. **European Journal of Pediatrics**, v. 180, n. 5, p. 1571–1579, 2021. https://doi.org/10.1007/s00431-021-03929-z

HANDELMAN, G. S. et al. eDoctor: Machine learning and the future of medicine. **Journal of Internal Medicine**, p. 1–17, 2018. https://doi.org/10. 1111/joim.12822

HARRISON, C. J.; SIDEY-GIBBONS, C. J. Machine learning in medicine: a practical introduction to natural language processing. **BMC Medical Research Methodology**, v. 21, n. 1, p. 158, 31 dez. 2021. https://doi.org/10.1186/s12874-021-01347-1

HARTMAN, M. E. et al. Readmission and Late Mortality After Critical Illness in Childhood*. **Pediatric Critical Care Medicine**, v. 18, n. 3, p. e112–e121, mar. 2017. https://doi.org/10.1097/PCC.0000000000001062

HOENK, K. et al. Multicenter study of risk factors of unplanned 30-day readmissions in pediatric oncology. **Cancer reports (Hoboken, N.J.)**, v. 4, n. 3, p. e1343, 2021. https://doi.org/10.1002/cnr2.1343

HUANG, Z. et al. Length of stay prediction for clinical treatment process using temporal similarity. **Expert Systems with Applications**, v. 40, n. 16, p. 6330–6339, 2013. https://doi.org/10.1016/j.eswa.2013.05.066

JOVANOVIC, M. et al. Artificial Intelligence in Medicine Building interpretable predictive models for pediatric hospital readmission using Tree-Lasso logistic regression. **Artificial Intelligence In Medicine**, v. 72, p. 12–21, 2016. https://doi.org/10.1016/j.artmed.2016.07.003

JUTRAS, C. et al. Anemia in Pediatric Critical Care. **International Journal of Clinical Transfusion Medicine**, v. 8, p. 23–33, 2020. https://doi.org/10.2147/IJCTM.S229764

KANE, J. M. et al. Resources and Costs Associated With Repeated Admissions to PICUs. **Critical Care Explorations**, v. 3, n. 2, p. e0347, 17 fev. 2021. https://doi.org/10.1097/CCE.0000000000000347

KANSAGARA, D. et al. Risk Prediction Models for Hospital Readmission: A Systematic Review. **JAMA**, v. 306, n. 15, p. 1688–1698, 2011. https://doi.org/10.1001/jama.2011.1515

KUMAR, D. et al. Factors Associated with Readmission of Pediatric Patients in a Developing Nation. **Indian Journal of Pediatrics**, v. 86, n. 3, p. 267–275, 2019. https://doi.org/10.1007/s12098-018-2767-0

KYLE, U. G.; PICHARD, C. The Dutch Famine of 1944-1945: A pathophysiological model of long-term consequences of wasting disease. **Current Opinion in Clinical Nutrition and Metabolic Care**, v. 9, n. 4, p. 388–394, 2006. https://doi.org/10.1097/01.mco.0000232898.74415.42

LIN, R. J. et al. Anemia in General Medical Inpatients Prolongs Length of Stay and Increases 30-Day Unplanned Readmission Rate. **Southern Medical Journal**, v. 106, n. 5, p. 316–320, maio 2013. https://doi.org/10.1097/SMJ.0b013e318290f930

MARKHAM, J. L. et al. Length of Stay and Cost of Pediatric Readmissions. **Pediatrics**, v. 141, n. 4, p. e20172934, abr. 2018. https://doi.org/10.1542/peds.2017-2934

MCCARTHY, A. et al. Prevalence of malnutrition in pediatric hospitals in developed and in-transition countries: The impact of hospital practices. **Nutrients**, v. 11, n. 2, p. 1–18, 2019. https://doi.org/10.3390/nu11020236

MICHAILIDIS, P. et al. Forecasting Hospital Readmissions with Machine Learning. **Healthcare**, v. 10, n. 6, p. 1–14, 2022. https://doi.org/10.3390/healthcare10060981

MORGAN, D. J. et al. Assessment of Machine Learning vs Standard Prediction Rules for Predicting Hospital Readmissions. **JAMA Network Open**, v. 2, n. 3, p. e190348, 8 mar. 2019. https://doi.org/10.1001/jamanetworkopen.2019.0348

MUMFORD, V. et al. Measuring the financial and productivity burden of paediatric hospitalisation on the wider family network. **Journal of Paediatrics and Child Health**, v. 54, n. 9, p. 987–996, set. 2018. https://doi.org/10.1111/jpc.13923

NIEHAUS, I. M. et al. Applicability of predictive models for 30-day unplanned hospital readmission risk in paediatrics: a systematic review. **BMJ Open**, v. 12, n. 3, p. e055956, 30 mar. 2022. https://doi.org/10.1136/bmjopen-2021-055956

PIRLICH, M. et al. The German hospital malnutrition study. **Clinical Nutrition**, v. 25, n. 4, p. 563–572, 2006. https://doi.org/10.1016/j.clnu.2006.03.005

PUFAL, E. C. et al. Motor development in the hospitalized infant and its biological and environmental characteristics. **Clinical & Biomedical Research**, v. 38, n. 1, p. 66–73, 2018. https://doi.org/10.4322/2357-9730.75638

RASHIKJ CANEVSKA, O. The Impact of Hospitalization on Psychophysical Development and Everyday Activities in Children. **The Annual of the Faculty of Philosophy in Skopje**, v. 71, n. January 2018, p. 465–470, 2018. https://doi.org/10.37510/godzbo1871465rc

REZENDE, I. et al. Prevalência da desnutrição hospitalar em pacientes internados em um hospital filantrópico em Salvador (BA), Brasil. **R. Ci. méd. biol.**, v. 3, n. 71, p. 194–200, 2004.

SARKER, I. H. Machine Learning: Algorithms, Real-World Applications and Research Directions. **SN Computer Science**, v. 2, n. 3, p. 160, 22 maio 2021. https://doi.org/10.1007/s42979-021-00592-x

SHADMI, E. et al. Predicting 30-Day Readmissions With Preadmission Electronic Health Record Data. **Medical Care**, v. 53, n. 3, p. 283–289, mar. 2015.

https://doi.org/10.1097/MLR.0000000000000315

SIDEY-GIBBONS, J. A. M.; SIDEY-GIBBONS, C. J. Machine learning in medicine: a practical introduction. **BMC Medical Research Methodology**, v. 19, n. 1, p. 64, 19 dez. 2019. https://doi.org/10.1186/s12874-019-0681-4

SILVA, V. L. S. DA et al. Hospitalization in the first years of life and development of psychiatric disorders at age 6 and 11: a birth cohort study in Brazil. **Cadernos de Saúde Pública**, v. 34, n. 5, p. 1–13, 28 maio 2018. https://doi.org/10.1590/0102-311x00064517

SYMUM, H.; ZAYAS-CASTRO, J. Identifying Children at Readmission Risk: At-Admission versus Traditional At-Discharge Readmission Prediction Model. **Healthcare**, v. 9, n. 10, p. 1334, 7 out. 2021. https://doi.org/10.3390/healthcare9101334

TAYLOR, T.; ALTARES SARIK, D.; SALYAKINA, D. Development and Validation of a Web-Based Pediatric Readmission Risk Assessment Tool. **Hospital Pediatrics**, v. 10, n. 3, p. 246–256, 2020. https://doi.org/10.1542/hpeds.2019-0241

TOOMEY, S. L. et al. Potentially Preventable 30-Day Hospital Readmissions at a Children's Hospital. **Pediatrics**, v. 138, n. 2, p. e20154182, 1 ago. 2016. https://doi.org/10.1542/peds.2015-4182

TURGEMAN, L.; MAY, J. H. A mixed-ensemble model for hospital readmission. **Artificial Intelligence in Medicine**, v. 72, p. 72–82, set. 2016. https://doi.org/10.1016/j.artmed.2016.08.005

VAN DEN AKKER, M.; BUNTINX, F.; KNOTTNERUS, J. A. Comorbidity or multimorbidity: What's in a name? A review of literature. **European Journal of General Practice**, v. 2, n. 2, p. 65–70, 1996. https://doi.org/10.3109/13814789609162146

VAN WALRAVEN, C. et al. Derivation and validation of an index to predict early death or unplanned readmission after discharge from hospital to the community. **Canadian Medical Association Journal**, v. 182, n. 6, p. 551–557, 6 abr. 2010. https://doi.org/10.1503/cmaj.091117

WAITZBERG, D. L.; CAIAFFA, W. T.; CORREIA, M. I. T. D. Hospital malnutrition:

The Brazilian national survey (IBRANUTRI): A study of 4000 patients. **Nutrition**, v. 17, n. 7–8, p. 573–580, 2001. https://doi.org/10.1016/S0899-9007(01)00573-1

WEINREICH, M. et al. Predicting the Risk of Readmission in Pneumonia. A Systematic Review of Model Performance. **Annals of the American Thoracic Society**, v. 13, n. 9, p. 1607–1614, set. 2016. https://doi.org/10.1513/AnnalsATS.201602-135SR

WOLFF, P. et al. Machine Learning Readmission Risk Modeling: A Pediatric Case Study. **BioMed Research International**, v. 2019, p. 1–9, 15 abr. 2019. https://doi.org/10.1155/2019/8532892

ZHOU, H. et al. Utility of models to predict 28-day or 30-day unplanned hospital readmissions: an updated systematic review. **BMJ Open**, v. 6, n. 6, p. e011060, 27 jun. 2016. https://doi.org/10.1136/bmjopen-2016-011060

ZHOU, H. et al. A 5-year retrospective cohort study of unplanned readmissions in an Australian tertiary paediatric hospital. **Australian Health Review**, v. 43, n. 6, p. 662–671, 2019a. https://doi.org/10.1071/AH18123

ZHOU, H. et al. Risk factors associated with paediatric unplanned hospital readmissions: a systematic review. **BMJ Open**, v. 9, n. 1, p. e020554, jan. 2019b. https://doi.org/10.1136/bmjopen-2017-020554

ZHOU, H. et al. Risk factors associated with 30-day all-cause unplanned hospital readmissions at a tertiary children's hospital in Western Australia. **Journal of Paediatrics and Child Health**, v. 56, n. 1, p. 68–75, 2020. https://doi.org/10.1111/jpc.14492

ZHOU, H. et al. Using machine learning to predict paediatric 30-day unplanned hospital readmissions: A case-control retrospective analysis of medical records, including written discharge documentation. **Australian Health Review**, v. 45, n. 3, p. 328–337, 2021. https://doi.org/10.1071/AH20062

APÊNDICE A – PARECER CEP

UNIVERSIDADE FEDERAL DE UBERLÂNDIA/MG

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: ESTUDO DE TÉCNICAS DE INTELIGÊNCIA ARTIFICIAL APLICADAS À ÁREA DA SAÚDE PARA DESFECHOS HOSPITALARES NEGATIVOS EM POPULAÇÃO PEDIÁTRICA

Pesquisador: Geórgia das Graças Pena

Área Temática:

Versão: 1

CAAE: 51706221.3.0000.5152

Instituição Proponente: Faculdade de Medicina

Patrocinador Principal: Financiamento Próprio

DADOS DO PARECER

Número do Parecer: 5.003.236

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

UBERLANDIA, 28 de Setembro de 2021

Assinado por:
Karine Rezende de Oliveira
(Coordenador(a))

Endereço: Av. João Naves de Ávila 2121- Bloco "1A", sala 224 - Campus Sta. Mônica
Bairro: Santa Mônica **CEP:** 38.408-144
UF: MG **Município:** UBERLANDIA
Telefone: (34)3239-4131 **Fax:** (34)3239-4131 **E-mail:** cep@propp.ufu.br

APÊNDICE B – INSTRUMENTO DE COLETA DE DADOS

Identificação:

Código do paciente: ____________
Data de internação no hospital: ___/___/___
Data de alta hospitalar: ___/___/___
Tipo de alta hospitalar: ☐ alta médica ☐ contra orientação médica
Tipo de admissão: _______
Dados sobre Óbito: ☐ Não ☐ Sim ___/___/___
Causa do óbito:
____
Data da Alta Hospitalar: ___/___/___
Data de nascimento: ___/___/___ Idade: ____dias
Gênero: ☐ F ☐ M

Variáveis Clínicas:

Enfermaria:
____
Diagnósticos médicos:
____

Sinais Vitais:

Temperatura: ____ºC
Frequência Cardíaca: _____ Frequência Respiratória: ________
Pressão Sistólica:_____ Pressão Diastólica:__________

Variáveis Bioquímicas:

Hemácias: _______ milhões/mm3 Hematócrito: _____%
Hemoglobina: _______ g%
VCM: _______ fL HCM: ________fL CHCM: _______ g/dL
Leucócitos: _______ mil/mm³ Linfócitos: _______ %
Bastonetes: ______% Segmentados: ______% Eosinófilos: ______ mm³
Basófilos:______ mm³ Monócitos: ______% Reticulócitos:______
Plaquetas: _______ mil/mm³ MPV:_______ VHS:________ mm/h
TAP:___________% RDW: ______%
Procalcitonina:______ ng/mL Proteína C reativa: ____mg/L
Ácido úrico: ______ mg/dL Creatinina:_____ mg/dL Ureia:_________ mg/dL
Bilirrubina direta: ___ mg/dL Bil. indireta: ____ mg/dL Bil. total: _____ mg/dL
Fosfatase alcalina:____ U/L TGP: ___ U/L TGO: _______ U/L
GAMA GT: ________ U/L Desidrogenase lática: ______ U/L
Proteínas Totais: ________ g/dL Albumina: _____g/L
Glicemia: _____ mg/dL Hemoglobina glicada:____%

Insulina: _____ microU/mL
Triglicerídeos: ___ mg/dL
Colesterol total: _____ mg/dL
LDL: ______ mg/dL
HDL: _____ mg/dL
VLDL: _____ mg/dL
Transferrina: ____ mg/dL
Índice de saturação de transferrina: ____%
Ferritina: ___ ng/mL
Ferro: ___mcg/dL
Capacidade de ligação do ferro: ___ mcg/dL
Lactato: ____ mmol/L
Creatinofosfoquinase (CPK): ______ ng/mL
Vitamina B12: ___
Vitamina D: ____
Vitamina B9:____
T3: ______ ng/dL
T4 livre: ___ ng/dL
TSH: ___ uIU/mL

Avaliação Nutricional:

Data da avaliação: ___/___/___
Peso Atual: ______kg
Altura: ______cm
IMC: _________kg/m^2
Peso Habitual: ______kg
Perda de Peso: ______kg
% Perda de Peso: ______%
Classificação z escore: ________
Conclusão da Avaliação Nutricional: _______