UNIVERSIDADE DE SÃO PAULO
FACULDADE DE SAÚDE PÚBLICA
PROGRAMA DE PÓS-GRADUAÇÃO EM EPIDEMIOLOGIA

MARIANE FURTADO BORBA

Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde

São Paulo
2023

MARIANE FURTADO BORBA

Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde

Versão original

Tese apresentada à Faculdade de Saúde Pública da Universidade de São Paulo para obtenção do título de Doutora em Ciências pelo Programa de Pós-graduação em Epidemiologia.

Área de concentração: Epidemiologia

Versão corrigida. Contendo as alterações solicitadas pela comissão julgadora em 19 de maio de 2023.

Orientador: Prof. Dr. Alexandre Dias Porto Chiavegatto Filho

Coorientador: Prof. Dr. André Filipe de Moraes Batista

São Paulo

2023

Catalogação da Publicação
Ficha elaborada pelo Sistema de Geração Automática a partir de dados fornecidos pelo(a) autor(a)
Bibliotecária da FSP/USP: Maria do Carmo Alvarez - CRB-8/4359

Furtado Borba, Mariane
Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde / Mariane Furtado Borba; orientador Alexandre Dias Porto Chiavegatto Filho; coorientador André Filipe de Moraes Batista. -- São Paulo, 2023.
99 p.

Tese (Doutorado) -- Faculdade de Saúde Pública da Universidade de São Paulo, 2023.

1. Generalização. 2. Machine learning. 3. Modelos Preditivos. 4. Decisões em Saúde. I. Dias Porto Chiavegatto Filho, Alexandre, orient. II. de Moraes Batista, André Filipe, coorient. III. Título.

Tese de autoria de Mariane Furtado Borba, sob o título **"Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde"**, apresentada à Faculdade de Saúde Pública da Universidade de São Paulo, para obtenção do título de Doutora em Ciências pelo Programa de Pós-graduação em Epidemiologia, na área de concentração Métodos e Técnicas de Análise em Epidemiologia, aprovada em 19 de Maio de 2023 pela comissão julgadora constituída pelos doutores:

Prof. Dr. Alexandre Dias Porto Chiavegatto Filho
Universidade de São Paulo
Presidente

Prof. Dr. José Leopoldo Ferreira Antunes
Universidade de São Paulo

Prof. Dr. Bruno Pereira Nunes
Universidade Federal de Pelotas

Profa. Dra. Hellen Geremias dos Santos
Instituto Carlos Chagas da Fundação Oswaldo Cruz

Dedico esta tese a todos os pesquisadores e professores que foram minha inspiração e guia em minha trajetória acadêmica, e que me ensinaram que o conhecimento é uma jornada contínua e empolgante. Agradeço também à minha família e amigos que estiveram ao meu lado em cada passo dessa jornada, me apoiando e incentivando, e que sempre acreditaram em mim, mesmo quando eu mesma não acreditava.

Agradecimentos

Ao Prof. Dr. Alexandre Dias Porto Chiavegatto filho, orientador, por todos os ensinamentos, oportunidades e encorajamento constante. Por me ajudar a desenvolver habilidades críticas e analíticas, e além disso acreditar em minha capacidade mesmo quando eu mesma duvidei. Agradeço por ele ter criado um laboratório diverso e harmonioso cujo apoio e a troca fizeram uma enorme diferença na minha trajetória durante o doutorado.

Ao Prof. Dr. André Filipe de Moraes Batista, por ter desempenhado um papel fundamental em meu aprendizado. Sua dedicação em compartilhar conhecimentos foram fundamentais para o meu desenvolvimento como pesquisadora. Agradeço por ter guiado o caminho das pedras, mostrando-me como superar obstáculos e alcançar meus objetivos.

Aos professores da Faculdade de Saúde Pública da Universidade de São Paulo por terem desempenhado um papel essencial na minha formação acadêmica e pessoal. Com experiência, conhecimento e competência, eles ampliaram minha visão de mundo e me inspiraram a buscar a excelência em tudo o que faço. Sou grata pela receptividade à interdisciplinaridade resultante da minha trajetória acadêmica.

Às secretárias do programa de pós-graduação em epidemiologia, por todo o auxílio e atenção prestada durante todos esses anos.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) pela concessão da bolsa de estudos – Processo 8888.492630/2020-00 – que possibilitou a realização do meu curso de doutorado.

À Camila Prado e ao Leonardo Ciciarelli pelo suporte psicológico.

Aos pesquisadores membros do Laboratório de Big Data e Análise Preditiva em Saúde (LABDAPS) por todo apoio e trocas diárias.

À minha família e amigos por suportarem minhas ausências em momentos importantes.

Ao meu marido, Gabriel Weber Costa, por ser meu maior incentivador desde o início dessa caminhada.

"A natureza se manifesta melhor onde não tem obstáculo."

(Autor desconhecido)

Resumo

FURTADO, Mariane. **Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde**. 2023. 101 f. Tese (Doutorado em Ciências) – Faculdade de Saúde Pública, Universidade de São Paulo, São Paulo, 2023.

A utilização de algoritmos de inteligência artificial tem crescido rapidamente nos últimos anos, aumentando o seu potencial de aplicação em saúde pública. Algoritmos de machine learning (ML) são capazes de auxiliar na predição de desfechos complexos e na tomada de decisões por parte dos profissionais da área da saúde. Esta tese tem como objetivo analisar a capacidade de generalização dos algoritmos na área da saúde e aplicar modelos de ML para predições utilizando dados tabulares frequentemente coletados nos sistemas de saúde. A tese será defendida sob a forma de três artigos científicos. O primeiro artigo realizou uma revisão sistemática da literatura sobre a capacidade de generalização de modelos de ML em saúde. Os resultados indicaram que, apesar de ainda limitada, a literatura sobre generalização em saúde está crescendo nos últimos anos em parte como uma demanda das próprias revistas científicas. O segundo artigo desenvolveu e avaliou a performance da validação externa de um algoritmo de ML no contexto da predição de risco de mortalidade neonatal. O modelo foi desenvolvido utilizando Extreme Gradient Boosting (XGB) em dados de São Paulo de 2012 a 2015, incluindo 807.932 nascidos vivos e 5.518 óbitos neonatais. Foi realizada a validação externa do algoritmo em 1.161 municípios brasileiros, incluindo todas as capitais de estado para o ano de 2016, totalizando 2.848.052 nascidos vivos e 23.948 óbitos neonatais. Os resultados mostraram que os municípios que ofertam estruturas de maior complexidade obtiveram uma performance similar ou mesmo superior ao modelo base desenvolvido com dados do município de São Paulo. No terceiro e último artigo desta tese, foi realizada uma análise da aplicação da técnica de generalização conhecida como transfer learning nos dados da Rede IACOV-BR para predizer óbito entre pacientes internados por Covid-19 usando dados de prontuário de 16.236 pacientes de 18 hospitais brasileiros coletados no primeiro trimestre de 2020 durante o início da pandemia de Covid-19 no Brasil. A abordagem desse artigo propôs uma comparação entre uma nova solução capaz de predizer o progresso clínico dos pacientes com Covid-19 versus a abordagem já aplicada para predições tabulares em saúde. Os resultados indicam que apesar de promissora, a técnica de transfer learning convencional não se mostrou superior aos resultados de performance obtidos localmente com os algoritmos de boosting utilizados para dados tabulares. Os resultados desta tese apontam para a importância da generalização dos algoritmos de ML em saúde, ao mesmo tempo que os desafios técnicos ainda persistem em relação à manutenção da performance preditiva nas diferentes localidades.

Palavras-chaves: Modelos Preditivos. machine learning. Generalização. Decisões em Saúde.

Abstract

FURTADO, Mariane. **Generalization analysis of machine learning algorithms and their applications in optimizing health decisions**. 2023. 101 p. Thesis (Ph.D. in Science) – School of Public Health, University of Sao Paulo, Sao Paulo, 2023.

The use of artificial intelligence algorithms has significantly increased in recent years, increasing their potential for application in public health. ML algorithms (ML) can assist in the prediction of complex outcomes and in decision-making by healthcare professionals. This thesis aims to analyze the algorithmic generalization capability in healthcare and apply ML models for the prediction of health outcomes from tabular data frequently collected in healthcare systems. The thesis will be defended as three scientific articles. The first article conducted a systematic literature review on the generalization capability of ML models in healthcare. The results indicated that, although still limited, the literature on generalization in healthcare has been growing in recent years, in part as demand from journals themselves. The second article evaluated the performance of external validation of an ML algorithm in the context of predicting neonatal mortality risk. The model was developed using Extreme Gradient Boosting (XGB) on São Paulo data from 2012 to 2015, including 807,932 live births and 5,518 neonatal deaths. External validation of the algorithm was performed in 1,161 Brazilian municipalities, including all state capitals in 2016, totaling 2,848,052 live births and 23,948 neonatal deaths. The results showed that municipalities offering more complex structures obtained similar or even superior performance to the base model developed with data from the municipality of São Paulo. In the third and final article of this thesis, an analysis of the application of the generalization technique known as transfer learning was performed on IACOV-BR Network data to predict death from Covid-19 using medical record data from 16,236 patients from 18 Brazilian hospitals collected in the first quarter of 2020 during the early Covid-19 pandemic in Brazil. The results indicate that, although promising, the conventional transfer learning technique did not prove superior to locally obtained performance results with traditional boosting algorithms. The approach of this article proposed a comparison between a new solution for predicting the clinical progress of Covid-19 patients versus the approach already applied for tabular predictions in healthcare. The results of this thesis point to the importance of the generalization of ML algorithms in healthcare, while technical challenges persist regarding the maintenance of predictive performance in different locations.

Keywords: Predictive Models. machine learning. Generalization. Health Decisions.

Preâmbulo

A presente tese de doutorado foi realizada no Programa de Pós-Graduação em Epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo, na linha de pesquisa métodos e técnicas de análise em epidemiologia, sob orientação do professor Dr. Alexandre Dias Porto Chiavegatto Filho, contando com o suporte financeiro, por meio de bolsa de doutorado, cota institucional de demanda social, processo 88887.492630/2020-00, da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – CAPES.

A tese foi desenvolvida junto ao Laboratório de Big Data e Análise Preditiva em Saúde – LABDAPS, sediado na Faculdade de Saúde Pública da USP, utilizando principalmente a ferramenta LABDAPS MLFlow, um framework que utiliza a linguagem Python em uma versão customizada da biblioteca PyCaret, que possui licença do MIT. Além dessa, bibliotecas como Numpy, Pandas, Scikit Learn, Tensorflow, Matplotlib, Seaborn, também foram utilizadas.

A tese divide-se em: introdução, objetivos, descrição da relação entre os artigos da tese; seção referente aos resultados com o primeiro artigo de generalização e seguindo dos artigos de aplicações; e por fim considerações finais e conclusão.

Os códigos em Python utilizados para a construção dessa tese estão disponíveis no repositório GitHub (https://github.com/MarianeFurtado).

Lista de figuras

Lista de tabelas

Lista de abreviaturas e siglas

AUC	Area Under the Receiver Operating Characteristic Curve
AUPRC	Area Under Precision-Recall Curve
CEP	Código de Endereçamento Postal
CBO	Classificação Brasileira de Ocupações
CNES	Cadastro Nacional de Estabelecimentos de Saúde
COREN	Conselho Regional de Enfermagem
CPF	Cadastro de Pessoa Física
CRM	Conselho Regional de Medicina
DATASUS	Departamento de Informática do Sistema Único de Saúde
DNV	Declaração de Nascido Vivo
DO	Declaração de Óbito
DUM	Data da Última Menstruação
ESF	Estratégia Saúde da Família
F1	F1 Score
FSP	Faculdade de Saúde Pública
IA	Inteligência Artificial
IACOV-BR	Inteligência Artificial para Covid-19 no Brasil
IBGE	Instituto Brasileiro de Geografia e Estatística
IPEA	Instituto de Pesquisa em Economia Aplicada
LABDAPS	Laboratório de Big Data e Análise Preditiva em Saúde
LGPD	Lei Geral de Proteção de Dados Pessoais
LightGBM	Light Gradient Boosting Machine

ML	machine learning
NPV	Negative Predictive Value
PPV	Positive Predictive Value
RG	Registro Geral
ROC	Receiver Operating Characteristic Curve
SD	Standard Deviation
SDG	Sustainable Development Goals
SHAP	SHAPley Additive exPlanations
SINASC	Sistema de Informação sobre Nascidos Vivos
SIM	Sistema de Informação sobre Mortalidade
TL	Transfer Learning
TRIPOD	Transparent Report of a multivariate prediction model for Prognosis or Individual Diagnosis
USP	Universidade de São Paulo
UF	Unidade da Federação
XGBoost	eXtreme Gradient Boosting

Sumário

1 Introdução

Inteligência artificial (IA) refere-se à capacidade de máquinas realizarem ações compatíveis com a inteligência humana (1). Como principal subcampo de IA, a área de machine learning (ML) utiliza dados para o aprendizado de regras complexas, frequentemente relacionadas à tomada de decisão (2). ML tem sido cada vez mais utilizada no campo da saúde, introduzindo muitos desafios e possibilidades (3). O crescimento desse conjunto de técnicas tem ocorrido por meio do aprimoramento dos algoritmos, da disponibilidade de grandes bancos de dados em plataformas online com acesso simplificado e do avanço da computação de baixo custo (4, 5). No contexto da medicina personalizada, que leva em consideração enfoques individuais, a aplicação de técnicas de ML pode dar suporte ao diagnóstico e prognóstico de doenças, auxiliando profissionais de saúde na tomada de decisão (6, 7, 8). Além disso, o aumento contínuo do custo dos cuidados com saúde traz o potencial do uso de ML para melhorar a eficiência da gestão e dos investimentos em saúde pública (9, 10).

Segundo 11 para melhorar a eficiência da saúde é importante reduzir os erros humanos, tanto em diagnósticos, prognósticos, tratamentos e prevenção, quanto no fluxo e gestão do trabalho. Em decisões sobre diagnóstico e prognóstico, o uso de ML pode auxiliar na identificação precoce da doença e otimizar a escolha dos exames levando em consideração eficiência e custo (12). Quanto à prescrição do tratamento, pode trazer ganhos na redução de erros de posologia, aprimorando a escolha do tratamento e da forma de administração (13), além de otimizar o tempo necessário para o retorno do paciente ao consultório médico.

Atualmente na saúde, ML tem sido mais usada para aumentar a eficiência e gestão dos serviços. Embora o desenvolvimento de tecnologias em torno de sistemas de saúde esteja em um processo de ascensão em países desenvolvidos, ainda existem muitas lacunas do uso dessas tecnologias em países de média e baixa renda (14). No Brasil, ML tem grande potencial de uso no sistema público de saúde, dada a falta de profissionais especialistas em diversas regiões brasileiras, principalmente nas mais remotas do país. Além disso, descobertas realizadas em um país tão diverso e desigual quanto o Brasil pode também contribuir para o desenvolvimento e implementação da tecnologia em países com realidades similares para tornar os seus processos em saúde mais rápidos e eficientes.

Pandemias como a de Covid-19 servem de alerta para a sociedade sobre a necessidade na agilidade em processos de decisão, tanto na identificação de doenças como na sinalização da existência de novos patógenos circulantes e soluções baratas e viáveis em cenários de poucos recursos financeiros (15). Muitos trabalhos com o uso de ML foram desenvolvidos durante os momentos mais críticos dessa pandemia, e com isso surgiram também novas questões que demandam discussões éticas. Uma delas é a necessidade de transparência em relação a como os modelos foram criados e em qual cenário foram testados, sendo que neste último ponto encontra-se o desafio deste trabalho, referente a quão generalizáveis podem ser os modelos e quais os seus riscos de vieses.

Este estudo é uma pesquisa quantitativa de predição com aplicações de aprendizado supervisionado para dados estruturados com algoritmos de ML, composto por artigos científicos que discutem e aplicam estratégias de generalização, com o objetivo de esclarecer os aspectos ainda não consolidados na literatura de ML. O primeiro artigo é referente a uma revisão sistemática sobre a generalização de modelos de ML em saúde. O segundo artigo é um estudo aplicado de generalização de modelos de ML, em um projeto de predição de risco de óbito neonatal, e o terceiro traz uma aplicação de transferência de aprendizado (transfer learning), um método promissor para as aplicações práticas em saúde pública.

1.1 Machine learning e decisões em saúde

A análise preditiva consiste na predição de resultados por meio da análise de dados passados. No contexto de ML, são utilizados algoritmos para predizer a ocorrência de um evento pelo treinamento de conjuntos de dados que identificam padrões para a predição de um evento futuro (16). Para que isso seja possível, três elementos devem ser considerados: a qualidade dos dados de treinamento, os algoritmos para a construção do modelo e a qualidade do conjunto de dados de teste. A qualidade dos dados refere-se não apenas à qualidade na coleta, como também à quantidade e o tipo de dado avaliado em termos de sua força como preditor para o desfecho a ser predito. Um exemplo é o fato de dados laboratoriais terem frequentemente mais informações preditivas do que apenas a anamnese clínica (8). Os algoritmos são responsáveis pela modelagem dos dados recebidos, e o teste de qualidade avaliará performance das predições em novos dados. Os modelos de ML podem ser classificados pelo grau de supervisão humana recebida (16). A forma mais

comum divide os modelos de aprendizado entre supervisionados, não supervisionados, semi-supervisionados e por reforço.

No caso dos modelos de aprendizado supervisionado, existem rótulos, ou seja, informação direta sobre o resultado esperado e para todos os desfechos durante o processo de treinamento. Simplificadamente, os dados são em geral separados em treino e teste por métodos como holdout e validação cruzada, e a avaliação de performance do modelo é analisada com os resultados obtidos no conjunto de teste. Geralmente, a proporção é de 70% para dados para treinamento e 30% para teste, variando de acordo com o tamanho do conjunto de dados e do problema em questão. A validação cruzada (ou cross-validation) também é uma técnica utilizada para avaliar o desempenho de um modelo e reduzir o risco de overfitting, que ocorre quando o modelo se ajusta muito bem aos dados de treinamento, mas não generaliza bem para novos dados. Mas ao contrário do método holdout, que divide o conjunto de dados em dois subconjuntos, a validação cruzada divide o conjunto de dados em várias partições (ou folds). O processo consiste em dividir o conjunto de dados em k partições. Logo, para cada fold, o modelo é treinado nos k -1 folds restantes e avaliado no fold atual. O processo é repetido k vezes, de forma que cada fold seja utilizado exatamente uma vez como conjunto de teste. Ao final, é calculada a média dos resultados obtidos em todas as iterações para avaliar o desempenho do modelo. Comumente, a validação cruzada é mais utilizada para selecionar hiperparâmetros ou para amostras pequenas.

O aprendizado supervisionado divide-se em dois principais tipos de modelos: os de classificação e os de regressão. Nos modelos de classificação, os resultados a serem preditos são classes ou categorias, sendo possível trabalhar com dados para classificação binária ou multinomial. Os desfechos preditos nesse caso podem ser um resultado discreto, indicando a classe, ou um valor contínuo, indicando uma probabilidade de pertencimento a uma classe (17). Por outro lado, quando o objetivo está em predizer um resultado contínuo aplica-se um modelo de regressão.

No caso do aprendizado não supervisionado, os dados utilizados não possuem rótulo, ou seja, não existe uma resposta correta a ser predita. O objetivo está em encontrar padrões e relações entre os fatores (18). Alguns exemplos são os algoritmos de agrupamento, de detecção de anomalias e de redução de dimensão. Já o aprendizado semi-supervisionado, consiste em uma combinação entre o aprendizado supervisionado e o não supervisionado, onde existem apenas parte dos dados rotulados e é comumente utilizado no reconhecimento de imagens. Diferentemente dos anteriores, o aprendizado por reforço não possui treino e

teste ou possibilidade de rótulo, mas envolve um agente que recebe retornos sobre suas ações na forma de recompensas e penalidades (19).

1.2 Tipos de Dados

1.2.1 Dados Estruturados vs Não Estruturados

Com a crescente quantidade de dados gerados todos os dias, a análise de dados tornou-se cada vez mais importante para empresas e organizações em todos os setores, incluindo a área da saúde. No entanto, nem todos os dados são iguais. Os dados podem ser classificados em duas categorias principais: dados estruturados e não estruturados.

No contexto da saúde, muitos trabalhos com ML têm realizado predições por meio de dados não estruturados (7, 20, 21, 22). Dados não estruturados referem-se a informações que não estão organizadas em uma estrutura de dados predefinida, como tabelas ou bancos de dados relacionais. Esses dados geralmente são encontrados em formatos como texto, áudio, vídeo, imagens, e outros tipos de dados não padronizados. Com a quantidade crescente desses dados disponíveis, o uso de IA, e especialmente ML, tem se tornando cada vez mais importantes nesse contexto.

Por outro lado, grande parte dos dados de fato coletados nos diversos sistemas de informação em saúde consistem em dados estruturados ou tabulares, isso é, dados facilmente organizados em linhas e colunas, como dados de sintomas, histórico de saúde do paciente, sinais vitais, variáveis sociodemográficas, resultados de alguns testes de diagnóstico, entre outras informações relevantes para a predição. Alguns desses trabalhos tratam de classificação para o atendimento (23), no-show (24), readmissão hospitalar (25), diagnósticos (26), prognósticos (27, 28, 29), óbito (30, 31, 32, 33). O presente estudo utilizou dados estruturados para desenvolver modelos preditivos de aprendizado supervisionado com desfechos de classificação.

1.3 Generalização de modelos

A generalização algorítmica está relacionada à capacidade de um algoritmo aprender padrões e aplicá-los a novos grupos que não fizeram parte do treinamento (34). Esses grupos podem diferir do treinamento sobre tempo, espaço, ou serem resultados de processos de

aleatorização como no método Holdout (divisão dos dados entre um conjunto de treino e um conjunto de teste). A habilidade do modelo de obter boas medidas de desempenho, a variar de acordo com o domínio a ser analisado, é o que caracteriza um modelo generalizável. Nesse contexto, algo importante de ser notado é que, em modelos de ML, validação interna refere-se ao treinamento e validação do modelo, enquanto a validação externa refere-se ao teste. Outro ponto a ser abordado é o uso de transfer learning (TL) na capacidade de generalização. O TL permite o uso de modelos pré-treinados em tarefas relacionadas, acelerando o treinamento e melhorando a generalização, principalmente quando há limitações nos conjuntos de dados em relação ao tamanho da amostra. Ao utilizar conhecimentos anteriores, o TL consolida as representações aprendidas, podendo resultar em melhor desempenho e aplicabilidade em novos problemas.

Os principais desafios para a capacidade de generalização de modelos de ML incluem situações tanto de overfitting como underfitting (modelo que não se ajusta bem aos dados de treinamento); amostra de treinamento de tamanho insuficiente ou fortemente desbalanceada, ou seja, quando uma das classes é muito maior em termos de observações do que a outra; distribuição dos dados de teste diferente dos dados nos quais o modelo foi treinado; seleção inadequada ou ausência de seleção de variáveis, incluindo assim muito ruído no modelo; existência de outliers (dados discrepantes); seleção inadequada dos hiperparâmetros ou dos algoritmos testados, entre outros (35). Além disso, mesmo na presença de uma generalização adequada do modelo, é importante monitorar continuamente seu desempenho com novos dados, a fim de manter a performance preditiva.

A utilização de ML em saúde pública tem o potencial de auxiliar os profissionais e gestores de saúde em diversas áreas. Para isso, existe a necessidade da compreensão das capacidades e limitações da generalização algorítmica para que se possa garantir a confiabilidade dos modelos construídos, em termos da não disseminação de vieses humanos e do fornecimento de resultados úteis para a sociedade.

As generalizações em ML são desejáveis, mas podem não ser sempre possíveis, e isso não invalida modelos locais que passaram por análises de validação interna e que possibilitem a minimização de erros humanos. Países em desenvolvimento, como o caso do Brasil, podem se beneficiar de modelos de ML, pois possuem uma realidade de um sistema de saúde com um elevado número de pacientes e um número insuficiente de profissionais especialistas em diversas regiões e áreas rurais (36).

Generalizações podem beneficiar doenças menos comuns, tornando possível o acesso a dados com o padrão da doença em maior escala. Além disso, modelo generalizáveis permitem um maior monitoramento dos resultados preditivos, devido à existência de mais pessoas usando o mesmo modelo. Apesar dos benefícios, a estrutura de pesquisa sobre modelos generalizáveis precisa preocupar-se com o fato de que algumas empresas estejam dispostas a usar o mesmo modelo, por fatores como o barateamento do serviço prestado, para múltiplos compradores sem garantia de qualidade final.

Estudos científicos sobre a capacidade de generalização dos modelos de ML podem proporcionar uma expansão do planejamento e utilização de ML para a realização de análises por pesquisadores e profissionais de saúde, aumentando a clareza em torno do tópico de generalização, permitindo também maior confiabilidade de quem terá acesso à tecnologia. Este trabalho desenvolveu análises gerais e aplicações práticas para o avanço de lacunas na literatura sobre a capacidade de generalização de modelos preditivos em saúde.

1.3.1 Transfer Learning

Em estudos com o uso de ML, o objetivo é predizer o que ocorrerá no futuro com base em dados passados. Isso pode levar a grandes desafios na presença de um padrão de dados no treinamento diferente do modelo inicial. Nesse cenário, TL é uma técnica de ML que permite que um modelo treinado em uma tarefa específica seja, em parte, reutilizado para resolver outras tarefas relacionadas. Por exemplo, um modelo construído com dados de registros eletrônicos de uma rede hospitalar para predizer a necessidade de internação em unidade de terapia intensiva (UTI) pode ser reutilizado para a predição de diagnostico de uma doença especifica, bem como pode também ser utilizado como base de conhecimento para a predição de internações em UTI em um hospital com baixa capacidade hospitalar que resulte em um histórico menor do desfecho para aprendizagem no treinamento. TL dispensa a necessidade de que os dados de treinamento sejam independentes e identicamente distribuídos dos dados de teste (37) possibilitando que seja possível obter resultados de performance quando as distribuições de dados, espaço e/ou tempo mudam (38).

Em saúde, o uso do TL tem se mostrado uma abordagem promissora para aprimorar a precisão dos modelos preditivos. TL permite que o conhecimento adquirido em uma tarefa seja aplicado em outra tarefa para otimização dos resultados preditivos (39, 40, 41, 42). Em

outras palavras, é a prática de utilizar um modelo pré-treinado em uma grande quantidade de dados como ponto de partida para um novo modelo, que será treinado em um conjunto de dados menor e possivelmente diferente. Essa técnica é particularmente útil em situações em que não há dados suficientes para treinar um modelo do zero ou quando o custo de coletar esses dados é muito alto. Além disso, tem sido comumente utilizada para o reconhecimento de imagens, porém possui grande potencial também para aplicação em dados estruturados (43).

Na prática, TL costuma ser aplicado com o uso de redes neurais, envolvendo a reutilização de modelos de redes neurais pré-treinados em tarefas relacionadas como ponto de partida para treinar um novo modelo de rede neural em uma tarefa específica. Esse processo permite que o novo modelo de rede neural aprenda com menos dados, reduzindo o tempo e o custo de treinamento, além de melhorar a performance preditiva. O processo de TL começa com o pré-treinamento do modelo em um grande conjunto de dados, geralmente de grandes proporções. Em seguida, o modelo é ajustado para a tarefa específica por meio de um processo chamado de fine-tuning, que ajusta os pesos das camadas do modelo pré-treinado em um conjunto de dados menor e relacionado à tarefa. Durante o fine-tuning, o modelo pré-treinado é refinado com dados específicos da tarefa, permitindo que ele aprenda representações mais precisas e adaptadas ao novo problema.

Dessa forma, apesar dos conhecidos desafios na aplicação de TL no conceito não estruturado devido à heterogeneidade dos dados (44), esse estudo pretende analisar as potencialidades da transferência de aprendizado por meio de uma aplicação prática com dados hospitalares estruturados.

1.4 Objetivos

- **Objetivo geral**: Analisar a capacidade de generalização de modelos preditivos de ML em tarefas de classificação para dados tabulares em saúde.
- **Objetivos específicos**:

I. Revisar a bibliografia sobre a capacidade de generalização de modelos preditivos de ML em saúde;

II. Aprimorar e validar a capacidade de generalização de um modelo preditivo para predição de mortalidade neonatal no Brasil;

III. Desenvolver um modelo de transferência de aprendizado, para a reutilização do aprendizado adquirido para a predição de óbito para pacientes internados por Covid-19.

2 Dados

2.1 Regras para a revisão bibliográfica sobre generalização de modelos de ML em saúde

Para o levantamento da literatura acerca da capacidade de generalização de modelos de ML em saúde foram utilizadas buscas nas plataformas PubMed: (ML OR artificial intelligence) AND (structured data OR tabular data) AND (health OR healthcare) AND (generalize OR generalization OR generalisability); Embase: (machine AND learning OR artificial) AND intelligence AND (structured AND data OR tabular) AND data AND (health OR healthcare) AND (generalize OR generalization OR generalisability); e Web of Science: (ML OR artificial intelligence) AND (structured data OR tabular data) AND (health OR healthcare) AND (generalize OR generalization OR generalizability), publicados entre janeiro de 2018 e dezembro de 2022. As palavras-chave utilizadas para a busca podem ser vistas no Quadro 1:

Quadro 1. Estratégia de busca adaptada a cada base de dados.

Database	**Rules**
Embase	(ML or artificial intelligence) and (structured data OR tabular data) and (health or healthcare) AND (generalize OR generalization OR generalisability)
PubMed	(machine AND learning OR artificial) AND intelligence AND (structured AND data OR tabular) AND data AND (health OR healthcare) AND (generalize OR generalization OR generalisability)
Web of Science	((ML or artificial intelligence) and (structured data OR tabular data) and (health or healthcare) AND (generalize OR generalization OR generalisability))

Os dados foram transferidos para o software Rayyan para a avaliação de elegibilidade para a revisão sistemática.

2.2 Sistema de Informações de Mortalidade

O Sistema de Informação sobre Mortalidade (SIM) foi estabelecido em 1975 e informatizado em 1979 pelo Ministério da Saúde, com o objetivo de coletar dados de mortalidade e das doenças que levaram a óbito em todo o Brasil.

A principal fonte dos dados do SIM é a declaração de óbito (DO), sendo padronizada nacionalmente e distribuída em três vias. O documento é preenchido pelo médico que atendeu o paciente, ou, na sua ausência, por duas pessoas que tenham presenciado ou verificado o óbito. As declarações de óbito são coletadas pela Secretaria de Saúde do

Município ou do Estado no estabelecimento de saúde e os seus dados são digitalizados e inseridos no SIM, sendo posteriormente disponibilizados no site do DATASUS.

Os dados do SIM são amplamente utilizados para a vigilância epidemiológica, além de servirem de base para a elaboração de indicadores sociodemográficos. Os seus resultados proporcionam a produção de estatísticas de mortalidade brasileira e a construção de alguns dos principais indicadores de saúde.

A DO sintetiza os dados de:

- Identificação: tipo de óbito (fetal, não-fetal), data do óbito, hora do óbito, naturalidade, data de nascimento, idade (em anos) e para menos de 1 ano (meses, dias, horas, minutos, ignorado), sexo (masculino, feminino, ignorado), raça/cor (branca, preta, amarela, parda, indigena), situação conjugal (solteiro, casado, viúvo, separado judicialmente, união estável, ignorada), escolaridade (sem escolaridade, fundamental I, fundamental II, médio, superior incompleto, superior completo, ignorado e série; ocupação habitual segundo a Classificação Brasileira de Ocupações (CBO);
- Residência: município de residência, código do município de residência;
- Ocorrência: local de ocorrência(hospital, outros estabelecimentos de saúde, domicílio, via pública, outros, ignorado), estabelecimento, código CNES do estabelecimento, município de ocorrência, código do município e unidade da federação;
- Fetal ou menor de um ano (de preenchimento exclusivo para óbitos fetais): idade da mãe, escolaridade da mãe, ocupação habitual, número de filhos vivos, número de perdas fetais ou abortos, número de semanas de gestação, tipo de gravidez (única, dupla, tripla e mais, ignorada), tipo de parto (vaginal, cesáreo, ignorado), morte em relação ao parto (antes, durante, depois, ignorado), peso da criança ao nascer em gramas e número da declaração de nascido vivo;
- Condições e causa do óbito (mulher em idade fértil): a morte ocorreu (na gravidez, no parto, no aborto, até 42 dias após o parto, de 43 dias a 1 ano após o parto, não ocorreu nesses períodos, ignorado), recebeu assistência médica durante a doença que ocasionou a morte (sim, não, ignorado) diagnóstico confirmado por necrópsia (sim, não, ignorado), causa direta da morte (diagnóstico, tempo entre o início da doença e o óbito e CID), outras condições que contribuíram para a morte;
- Médico: Óbito atestado por médico (assistente, substituto, IML, SVO, outro), município e unidade da federação do SVO ou IML;

- Causas externas: tipo de óbito (acidente, suicídio, homicídio, outros ou ignorado), acidente de trabalho (sim, não, ignorado), fonte da informação (boletim de ocorrência, hospital, família, outro, ignorado), descrição do evento incluindo tipo de local de ocorrência e endereço caso tenha ocorrido em via pública;
- Cartório: nome do cartório, código do cartório, número do registro, data e município;
- Localidade sem médico: nome do declarante e de duas testemunhas.

2.3 Sistema de Informações sobre Nascidos Vivos

O Sistema de Informação sobre Nascidos Vivos (SINASC) foi implantado a partir de 1990 pelo Ministério da Saúde com o objetivo de coletar dados sobre nascimentos em todo o território brasileiro.

A principal fonte dos dados do SINASC é a declaração de nascido vivo (DNV), sendo padronizada nacionalmente e distribuída em três vias. Assim como no caso do SIM, no SINASC o documento também é preenchido por um profissional de saúde que atendeu o paciente. As declarações de nascidos vivos são coletadas pela Secretaria de Saúde do Município, e são agregados aos níveis estadual e federal, sendo posteriormente disponibilizados no site do DATASUS.

Os dados do SINASC são amplamente utilizados para recolhimento de informações sobre nascimentos e utilizados para a construção de indicadores demográficos e sociais.

As variáveis coletas na DNV são relacionadas a dados de:

- Identificação do recém-nascido: data do nascimento, hora do nascimento, sexo (masculino, feminino, ignorado), peso da criança ao nascer em gramas, índice Apgar no 1º e no 5º minuto, detecção de alguma anomalia congênita (sim, não, ignorado);
- Local de ocorrência: local (hospital, Outro estabelecimento de saúde, domicílio, outros, ignorado), estabelecimento de saúde, código CNES, município de ocorrência, código do município e unidade da federação;
- Mãe: escolaridade (sem escolaridade, fundamental I, fundamental II, médio, superior incompleto, superior completo, ignorado) e série; ocupação habitual segundo a CBO, data de nascimento da mãe, idade da mãe, naturalidade, situação conjugal (solteira, casada, viúva, separada judicialmente, união estável, ignorada), raça/cor (branca,

preta, amarela, parda, indígena), município de residência, código do município de residência; unidade da federação;

- Pai: idade do pai;
- Gestação e parto: número de gestações anteriores, número de partos vaginais, número de cesáreas, número de nascidos vivos, número de perdas fetais/abortos, data da última menstruação DUM, número de semanas de gestação (se DUM ignorada), método utilizado para estimar (exame físico, outro método, ignorado), número de consultas pré-natal (ignorado), mês de gestação em que iniciou o pré-natal (ignorado), tipo de gravidez (única, dupla, tripla ou mais, ignorado), apresentação (cefálica, pélvica ou podálica, transversa, ignorado), o trabalho de parto foi induzido (sim, não, ignorado), tipo de parto (vaginal, cesáreo, ignorado), cesárea ocorreu antes do trabalho de parto iniciar (sim, não, não se aplica, ignorado), nascimento assistido por (médico, enfermeira/obstetriz, parteira, outros, ignorado);
- Anomalia congênita: descrição das anomalias observadas;
- Preenchimento: data do preenchimento, função (médico, enfermeiro, parteira, funcionário do cartório, outros);
- Cartório: nome do cartório, número do cartório, registro, data, município, unidade da federação.

Para esta tese, dois bancos de dados, obtidos do Linkage do SINASC e do SIM, foram avaliados, um disponibilizado pela Secretaria municipal de Saúde de São Paulo, referente a nascidos vivos no município de São Paulo entre 2012 e 2017, e outro pelo Ministério da Saúde, referentes a nascidos vivos em municípios de todo o País entre 2014 e 2017.

Para a identificação dos óbitos neonatais de nascidos vivos no município de São Paulo, o procedimento de Linkage foi realizado em duas etapas. Na primeira, foi filtrados pela Secretaria Municipal da Saúde de São Paulo capital dados entre os anos de 2012 e 2017, possuindo um total de registros igual a 1.202.843, sendo em sequência excluídos os nascimentos anteriores a 2012 ou posteriores a 2017, com idade gestacional inferior a 15 semanas ou posteriores a 45 semanas e nascimentos que não ocorreram no município de São Paulo. Após as exclusões, os dados do SINASC resultaram em 1.163.153 registros de nascidos vivos. Em relação ao Sistema de Informação sobre Mortalidade (SIM), foram recebidos dados entre os anos de 2012 e 2018, totalizando 313.756 registros, sendo em

sequência excluídos os óbitos infantis acima de 5 anos de idade ou cuja a idade no momento do óbito encontra-se desconhecida, ignorada ou nula. Após as exclusões, os dados do SIM resultaram em 16.532 registros de óbitos infantis. O linkage dos dados do SINASC e SIM ocorreu em duas etapas, sendo a primeira um linkage determinístico com as informações do número de identificação do nascimento quando preenchido também na declaração de óbito.

Na segunda etapa, foi realizado um linkage probabilístico utilizando o nome da mãe, data de nascimento e sexo da criança. O linkage probabilístico foi utilizado apenas quando a similaridade encontrada superou 85%. O banco de dados resultante do linkage chegou a 12.986 registros, ou seja, cerca de 81% dos registros originais constam no banco de dados total. Neste trabalho, foram utilizados apenas os dados de nascidos vivos entre 2012 e 2015 (N=807.932) com óbitos até o vigésimo oitavo dia do nascimento (N=5.518). Todos os dados foram recebidos já anonimizados.

O segundo linkage recebido de dados do SIM e SINASC foi disponibilizado pelo Ministério da Saúde com dados de nascimentos em 2014 (N=2.979.260), 2015 (N=3.017.670) e 2016 (N=2.857.800) e óbitos menores de um ano entre 2014 e 2017 (N=103.842). Para os nascidos vivos de diferentes municípios brasileiros, os óbitos neonatais foram identificados por meio de Linkage realizado em três etapas: a primeira, determinística, pelo número da DVN, sendo responsável por 80% da vinculação; a segunda, também determinística, por uma chave formada pelo nome da mãe, data de nascimento da criança e sexo, com incremento de 9% na vinculação; o terceiro, probabilístico, utilizou o nome da mãe, data de nascimento e sexo para pareamento com o Link Plus 12 (CDC-Atlanta), adicionando um incremento de 4% na vinculação. O resultado foi 93% de vinculação. Apenas 7.697 registros, ou cerca de 7% dos óbitos infantis, não foram localizados no SINASC. Foram utilizados apenas os registros de nascimentos de 2016 com óbitos entre 2016 e 2017 de forma a homogeneizar os resultados de teste. Para esse trabalho, foram selecionados os dados de municípios que apresentavam pelo menos 2 óbitos.

Assim como o linkage produzido pelo município de São Paulo, os dados do Ministério também foram recebidos anonimizados e livres de variáveis sensíveis. Os dados podem ser acessados em Diniz et al. (2022).

2.4 IACOV-BR

A rede de Inteligência Artificial para Covid-19 no Brasil (IACOV-BR) foi criada entre março e junho de 2020, pelo Laboratório de Big Data e Análise Preditiva em Saúde (LABDAPS) da Faculdade de Saúde Pública (FSP) da Universidade de São Paulo (USP), incluindo informações de 16.236 indivíduos como dados clínicos, laboratoriais e demográficos de pacientes com suspeita ou confirmação de covid-19 de 18 hospitais (públicos e/ou privados) das cinco regiões brasileiras. O projeto teve o financiamento da Microsoft ("AI for Health COVID-19 Grant"), da Fundação de Apoio à Pesquisa do Estado da Paraíba (Fapesq) e do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) do Ministério da Ciência, Tecnologia e Inovações. A rede coletou informações durante a triagem dos pacientes para o desenvolvimento de modelos de ML para a predição do prognóstico de covid-19, como a necessidade de ventilação mecânica, de internação em UTI ou risco de óbito. Os dados foram recebidos dos hospitais já anonimizados. As variáveis coletadas foram:

- Sociodemográficas: idade (em anos), gênero (masculino, feminino);
- Clínicas: resultado RT-PCR para Covid-19 (positivo, negativo), data do pedido do RT-PCR, data de internação em leito comum, data de inicio do uso de ventilação mecânica, data de internação em UTI, data de óbito, data de alta hospitalar, pressão arterial sistólica (mmHg), pressão arterial diastólica (mmHg), pressão arterial média (mmHg), temperatura corporal, frequência cardíaca (número de batimentos cardíacos por minutos), frequência respiratória (número de movimentos respiratórios por minuto);
- Laboratoriais: hemoglobina (g/dL), plaquetas (quantidade/mm^3), hematócrito (%), hemácias (quantidade x $10^6/mm^3$), concentração de hemoglobina corpuscular média (g/dL), hemoglobina corpuscular média (pg), amplitude de distribuição dos glóbulos vermelhos (%), volume corpuscular médio (fL), leucócitos (quantidade/mm^3), neutrófilos (quantidade/mm^3), linfócitos (quantidade/mm^3), basófilos (quantidade/mm^3), eosinófilos (quantidade/mm^3), monócitos (quantidade/mm^3).

3 Métodos

3.1 Glossário de Machine learning

3.1.1 Machine learning

ML é uma subárea da inteligência artificial que permite que modelos captem padrões a partir de dados sem serem explicitamente programados. Ele fornece uma maneira eficaz de extrair insights e padrões úteis a partir de conjuntos de dados, tornando-se uma tecnologia cada vez mais relevante, especialmente nas áreas da saúde. Nesta seção, serão explorados os conceitos básicos de ML.

- **Algoritmo**: conjunto de instruções ou regras que um modelo segue para aprender a partir dos dados. Um algoritmo de ML é responsável por determinar como um modelo aprende a partir dos dados, incluindo como ele processa, transforma e ajusta os dados para fazer predições.
- **Baseline**: modelo utilizado como base de comparação para a performance dos resultados, sendo responsável pela performance mínima esperada.
- **Boosting**: conjunto de classificadores fracos ou simples (árvores de decisão), ligeiramente correlacionados com a classificação verdadeira, para criar um classificador forte, melhor do que um classificador aleatório. Esses algoritmos aprendem de forma iterativa, ajustando os erros dos modelos anteriores.
- **Calibração**: capacidade do modelo de apontar as probabilidades preditas corretas.
- **Camada**: conjunto de neurônios artificiais em uma rede neural.
- **Camada oculta**: camadas de uma rede neural entre a camada de entrada e de saída.
- **Convergência**: ocorre quando no modelo não pode melhorar mais com o incremento de iterações, ou seja, os valores de função de custo (perda) mudam pouco ou nada a cada iteração adicional.
- **Classe majoritária**: classe do desfecho com maior número de exemplos.
- **Classe minoritária**: classe do desfecho com menor número de exemplos.
- **Cluster**: agrupamento de acordo com o padrão dos dados.
- **Classificação binária**: modelo de classificação que prediz classes mutuamente exclusivas.

- **Dados categóricos**: composto de variáveis binárias (dummies) e/ou com múltiplas categorias, que costumam representar uma categoria ou classe para descrever características qualitativas.
- **Dados contínuos**: composto de variáveis quantitativas ou numéricas, que podem assumir um número infinito de valores em um intervalo contínuo. Eles são medidos em uma escala contínua e podem ser fracionados em quantidades menores, como números decimais ou frações.
- **Dados desbalanceados**: quando a tarefa de classificação possui uma classe predominante. Geralmente os modelos de saúde são desbalanceados em relação ao desfecho positivo (existência de doença).
- **Desfecho**: resultado a ser predito, o Y da equação ou variável dependente.
- **Deep learning**: modelo de rede neural com mais do que uma camada oculta.
- **Engenharia de features**: modificação de variáveis a serem incluídas no modelo.
- **Ensemble**: modelos treinados onde as predições resultantes são agregadas, ou calculadas em torno da média.
- **Época**: total de observações ou exemplos sobre o tamanho do lote, ou seja, é a passagem completa de todo o conjunto de dados de treinamento durante o processo de treinamento de um modelo. No decorrer de cada época os parâmetros do modelo são atualizados.
- **Exemplo**: o mesmo que instância, observação, ou linha dos dados tabulares.
- **Função objetivo**: métrica a ser otimizada.
- **Generalização**: capacidade do modelo performar corretamente em novos dados.
- **Hiperparâmetros**: são configuração externas ao modelo, controlando a qualidade do modelo, não sendo diretamente estabelecidos pelos dados, por exemplo a taxa de aprendizado de uma rede neural.
- **Houdout**: estratégia que oculta dados do treinamento.
- **Importância de variáveis**: explicação de como as variáveis foram importantes para o algoritmo na construção do modelo.
- **Iteração**: número de vezes que um modelo de ML é treinado em um conjunto de dados durante o processo de ajuste de pesos e parâmetros.
- **Lote**: conjunto de observações utilizados em uma iteração durante o treinamento.
- **Modelo**: representação matemática treinada em um conjunto de dados para aprender padrões que levem aos desfechos.

- **Modelo de classificação**: predição resultante é uma classe, podendo ser binária ou múltipla.
- **Neurônio**: unidade dentro de uma camada oculta que recebe uma entrada e aplica uma função para gerar um valor de saída.
- **Normalização**: conversão em um intervalo de valores padrão.
- **One-hot encoding**: técnica de representação de dados categóricos que consiste em transformar cada categoria em um vetor binário de tamanho igual ao número de categorias possíveis.
- **Otimização de hiperparâmetros**: seleção da melhor configuração de hiperparâmetros.
- **Parâmetros**: são configurações internas das variáveis em um modelo de ML, usado para o ajuste do modelo aos dados, como por exemplo os pesos de uma rede neural ou os coeficientes de uma regressão.
- **Pré-processamento**: conjunto de técnicas e etapas que são aplicadas aos dados brutos antes de serem utilizados como entrada para um modelo de ML.
- **Regularização dropout**: regularização utilizada para treinar redes neurais e reduzir overfitting.
- **ReLU**: tipo de função de ativação em uma rede neural.
- **Redes neurais**: modelo com uma camada oculta.
- **Ruído**: outlier ou dados que não seguem um padrão ou não possuem informações relevantes para a tarefa de aprendizado.
- **Variância**: sensibilidade das predições na presença de alguma alteração nos dados de entrada.
- **Variáveis preditoras**: também conhecidas como features, atributos ou características, correspondendo as variáveis independentes da equação.
- **Validação cruzada**: técnica utilizada em ML para avaliar a capacidade de generalização de um modelo. Consiste em dividir o conjunto de dados disponível em partições de mesmo tamanho, com treinamento repetido em k-1 partições e validado na partição restante.
- **Viés**: tendência de um modelo de ML a aprender ou predizer certos resultados com base na forma como os dados de treinamento foram coletados ou rotulados, podendo surgir do conjunto de dados de treinamento, do algoritmo utilizado e das

suposições do modelo. O viés pode levar a problemas de discriminação, injustiça e/ou desigualdade.

- **Taxa de aprendizado**: hiperparâmetro que determina a magnitude da velocidade dos ajustes feitos aos pesos.
- **Threshold**: valor de corte para determinar a classe resultante de um modelo de ML. Quando usado no conjunto de teste, o modelo produzirá uma pontuação de probabilidade para cada classe. O threshold de classificação é usado para converter a pontuação de probabilidade em uma classe para o desfecho.

Métricas de Desempenho

Para a avaliação do desempenho dos modelos preditivos de classificação foi utilizada a Área abaixo da curva Receiver Operating Characteristic (ROC). A curva ROC é uma curva sobre a linha de discriminação que possibilita a comparação dos classificadores intervalo [0,1], com valores que se aproximam de 1 indicando classificações melhores. Essa métrica se mostra adequada quando se dispõe classes desproporcionais em problemas de classificação binária (45).

- **Precisão**: capacidade do modelo identificar corretamente os exemplos positivos, ou a proporção de instâncias positivas que foram classificadas como positivas. Pode ser calculada pela divisão de números verdadeiros positivos (TP) pela soma dos verdadeiros positivos e dos falsos positivos (FP).

$$Precisão = \frac{TP}{(TP + FP)}$$

- **Sensibilidade/Recall**: a proporção de instâncias classificadas corretamente como positivas em relação ao total de instâncias de fato positivas. Pode ser calculada pela divisão de número de verdadeiros positivos pela soma dos verdadeiros positivos e falsos negativos.

$$Sensibilidade = \frac{TP}{(TP + FN)}$$

- **F1 score**: média harmônica entre precisão e sensibilidade. Pode ser calculada por:

$$F1 = \frac{2 * (Precisão * Sensibilidade}{(Precisão + Sensibilidade)}$$

- **AUCPR**: a área abaixo da curva Receiver Operating Characteristic (ROC) de precision-recall, avalia a habilidade do modelo em predizer corretamente as instâncias positivas, em relação às instâncias negativas. A AUCPR dá mais peso ao desempenho do modelo na identificação correta das instâncias positivas.
- **Especificidade**: a proporção de instâncias negativas classificadas corretamente como negativas em relação ao total de instâncias negativas. Pode ser calculada pela divisão de número de verdadeiros negativos (TN) pela soma dos verdadeiros negativos e falsos positivos (FP).

$$Especificidade = \frac{TN}{(TN + FP)}$$

- **Valor predito negativo (VPN)**: capacidade do modelo identificar corretamente os exemplos negativos, ou a proporção de instâncias negativas classificadas como negativas. Pode ser calculada pela divisão de números verdadeiros negativos pela soma dos verdadeiros negativos e dos falsos negativos (FN).

$$VPN = \frac{TN}{(TN + FN)}$$

- **Brier score**: medida de avaliação que permite conferir a calibração do modelo, ou seja, serve para verificar a qualidade do modelo. Pode ser calculada por:

$$Brier = \frac{1}{n} * \sum (y - p)^2$$

Sendo n o número total de instâncias; y a classe real da instância e p a probabilidade predita.

4 Resultados

4.1 Artigo 1. Generalization of machine learning models: A scope review of the health applications

Keywords: Machine learning; healthcare; generalization

Abstract

Machine learning applications have been a growing area of scientific research over the last few years. Consequently, the need for increasingly accurate and robust models has intensified in several areas, especially in healthcare. Epidemiological studies commonly seek to develop methodologies that have external validation, i.e. that prove to be effective when applied to other populations. These challenges are increasingly important for studies on machine learning in healthcare, and the debate if this should be a requirement for predictive algorithms is still incipient. The present study performed a scope analysis to identify the most prevalent outcomes and directions in the field of machine learning generalizations in healthcare. We found that despite the challenges of generalization being frequently discussed in the literature, few studies have actually tested the generalization ability of machine learning algorithms with real patient data.

4.1.1 Introduction

Machine learning (ML) can improve health outcomes by facilitating the analysis of complex and multidimensional data, with the potential of supporting decision-making processes and enabling personalized treatment strategies (6, 46, 8). However, the effectiveness of ML models depends on their ability to generalize well to new data. The generalizability of a model refers to its ability to correctly predict the occurrence of events when exposed to a different population from its training set (47). In machine learning studies, generalization usually refers to the replicability of the predictive metrics to other data, with temporal or spatial differentiation, when compared to the training and validation data, and can be an indicator of the reliability of the model (48).

In machine learning for health applications, algorithms are frequently applied to predict the occurrence of an event, requiring a training set that will guide the model to learn patterns for predicting future events (16). The use of machine learning in healthcare can help promote the prioritization of care by considering individual factors (49). However, due to the heterogeneity of the input data, there are usually no guarantees of model extrapolation, given the current limitations of the technology. In the healthcare sector, where data is often limited and expensive to obtain, the ability to generalize supervised ML models is particularly important to ensure that models can be deployed in real-world environments.

Most epidemiological studies use tabular data, especially, clinical and laboratory data, that can be organized into tables. Although tabular data is vast in health, it is still less studied in machine learning applications than unstructured data. On unstructured data the algorithms need extremely large databases to get reasonable results. Tabular data, on the other hand, may need leaner samples if they have strong predictors.

Health databases used for predictive analyzes with ML often have a low number of observations with low prevalence of the outcome, and are based on data from a hospital or hospital network, which can restrict their ability to generalize.

We performed a scope review on the generalization of supervised ML models in healthcare, with the objective of analyzing the main research outcomes regarding the generalization of predictive algorithms in healthcare in order to identify its existing applications, in addition to addressing its limitations, gaps, and main trends.

4.1.2 Material and Methods

This scope review was developed to answer the following question: What are the main generalization issues regarding machine learning algorithms in healthcare? Registered in Prospero International prospective register of systematic reviews with ID CRD42023409789.

Inclusion criteria

In this scope review, we selected as eligible articles that met the following criteria: (a) ML articles using supervised learning with a predictive goal; (b) articles with information collected at the individual level; and (c) articles focusing on tabular/structured features such as socioeconomic and demographic characteristics, health conditions, and clinical and laboratory results.

Exclusion criteria

The following were excluded: (a) articles that used synthetic data (random data division resulting from the hold out method); (b) articles that used only genetic or pharmacological data, due to the focus on common data entered in hospital records; (c) articles using unstructured data; and (d) review articles.

Outcome

The outcomes of interest comprise any tabular health outcome, such as sepsis, hospital readmission, and death.

An electronic search was performed on the following databases: Embase, US National Library of Medicine, National Institutes of Health – PubMed, and Web of Science (WoS), using keywords such as "artificial intelligence", "machine learning", "structured data", "tabular data", "health", "healthcare", "generalization", and "generalizability". The search strategy per each database is shown in Frame 1. Searches were restricted to articles

published between January 2018 and December 2022, and no language restrictions were applied.

Frame 1. Adapted search strategy for Embase, PubMed, and Web of Science databases.

Database	**Rules**
Embase	(ML or artificial intelligence) and (structured data OR tabular data) and (health or healthcare) AND (generalize OR generalization OR generalisability)
PubMed	(machine AND learning OR artificial) AND intelligence AND (structured AND data OR tabular) AND data AND (health OR healthcare) AND (generalize OR generalization OR generalisability)
Web of Science	((ML or artificial intelligence) and (structured data OR tabular data) and (health or healthcare) AND (generalize OR generalization OR generalisability))

Two authors (MF and KAM) independently reviewed all individual manuscripts. In order to collaboratively select, organize, and manage decisions about the articles identified in this systematic review, the reviewers used the Rayyan software. All articles identified from all research databases were exported to a spreadsheet and entered Rayyan, in which duplicate articles were removed. Subsequently, the title and abstract of each study were independently assessed by the two reviewers, to evaluate its eligibility for review, considering the pre-established inclusion and exclusion criteria. In this first screening step, papers that clearly did not fulfill the inclusion and exclusion criteria were removed. Those that met the criteria or those that the reviewers were not able to decide, based on titles and abstracts only, had their full texts evaluated independently by the two reviewers. Finally, for all papers included in the review, data extraction was performed independently by the same two reviewers.

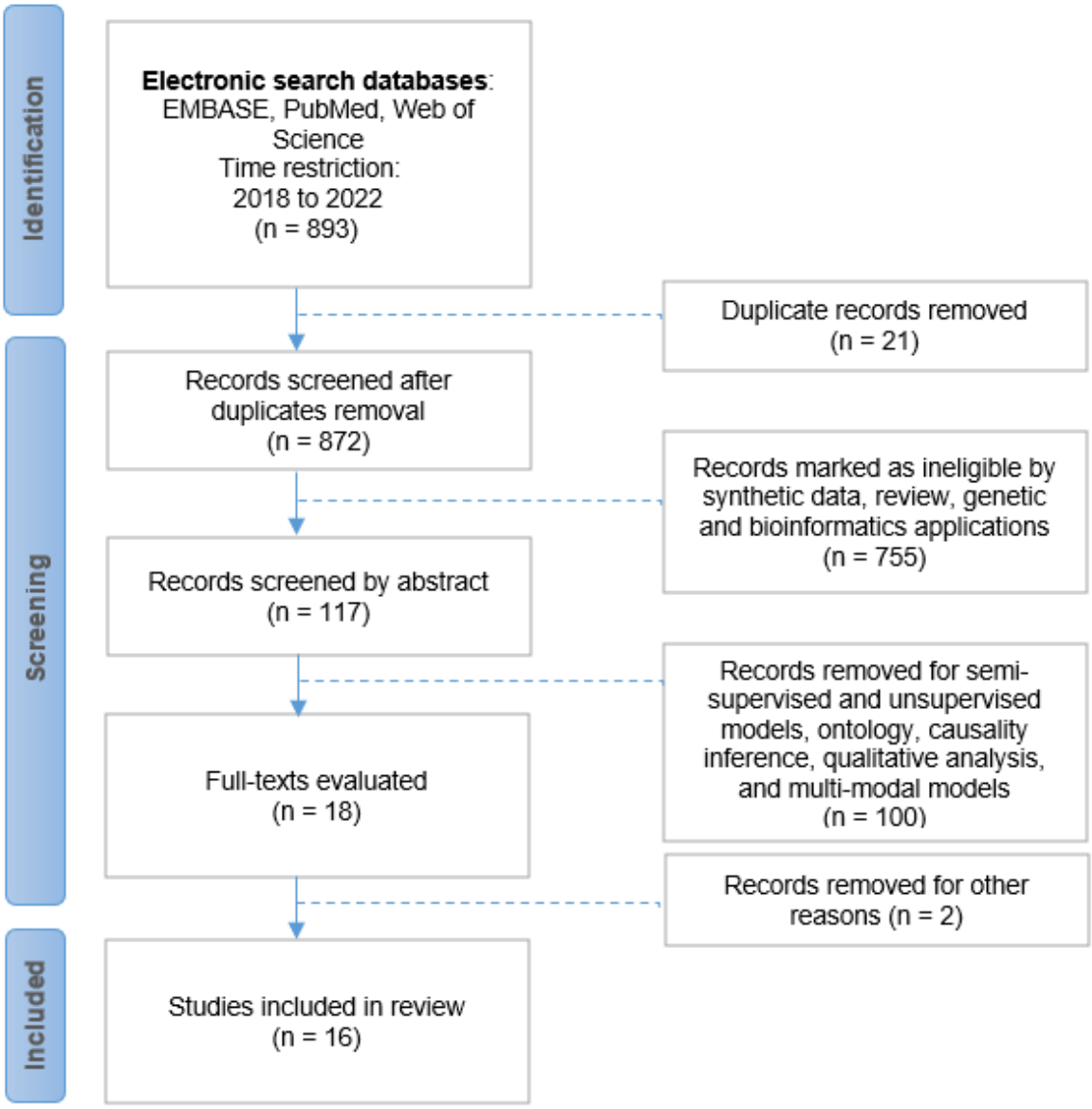
The information collected from each manuscript included identification details, information about the data, information about the models and its generalization ability, and main findings. Regarding identification, it was collected the title of the paper, authors, and year of publication. On the data, we collect details such as the country, data settings, sample size, and years of coverage of the data. We also extracted information about the type of prediction problem (classification or regression), the method used for dividing training and testing data, the algorithms used, and the metrics reported, such as Area Under Receiver Operating Curve (AUC), accuracy, and F-score. Regarding generalization, we evaluated whether the models were tested in diverse scenarios, including different

temporal, spatial, and ethnic-racial contexts. Finally, the two reviewers identified the outcome of each study and collected its main findings.

4.1.3 Results

The combined search from all three databases yielded 893 studies. After the removal of duplicates, screening of titles and abstracts, and full-text evaluation, 16 studies were included in this systematic review. The process of identifying and selecting relevant scientific articles is shown in Figure 1.

Figure 1 – Integrative representation of data extraction steps including identification, screening and inclusion steps reported according to PRISMA guidelines



Characteristics of the studies

The studies in this review were mostly concentrated in the United States (75%, 12 out of the 16 papers), with the rest divided between Brazil, Mexico, Canada, Denmark, Singapore, and China. Among the collected variables, there were socioeconomic and demographic characteristics, clinical data, laboratory data, criminal data, medications, and medical history. Topics included risk of death (18.8%, 3 articles (50, 51, 52), hospital admission or readmission (3 articles (53, 54, 55), risk of suicide (12.5%, 2 articles (56, 57), postoperative complications (2 articles (58, 59), and one paper of each of the following subjects: survived (60), diabetes (61), sepsis (62), Fabry disease (63), child health diagnosis (64), and risk of crime after psychiatric admission (65).

Table 1 – Studies characteristics

Study	Country	Outcome	Algorithms	Sample size	Training and testing strategy	Metrics								Generalization
						AUC	AUPRC	Precision	Recall	Accuracy	F-score	Specificity	NPV	
Barak-Corren et al.	US	risk of suicide	NB	3,714,105	50% for validation	yes	no	yes	no	no	no	yes	no	In space
Bonde et al.	US	postoperative complications	NN	5,881,881	79 train, 19 validation and 2 test	yes	no	no	no	no	no	no	no	Synthetic
Chi et al.	US	risk of death	LSTM	35,521	80 train, 10 validation, 10 test	yes	no	yes	yes	yes	yes	no	yes	For race
Edgcomb et al.	US	hospital readmission	CART (DT)	15,644	10- folds CV	yes	yes	yes	yes	yes	yes	no	yes	Synthetic
Hill et al.	US	risk of death	RF, EN, XGBoost, LR	52,894	per year	yes	no	yes	yes	yes	no	yes	no	In time
Jefferies et al.	US	Fabry disease	Not informed	1,004,978	75 train, 25 test	yes	no	no	no	no	no	no	no	Synthetic
Jorge et al.	US	hospital admission	DT, RF, NB, LR	1996	10-folds CV	yes	no	yes	no	no	yes	no	no	Synthetic
Kumar et al.	Singapore	Diabetes	LR, SVM, Catboost, NN	561	5-folds CV	yes	no	no	no	no	no	no	no	Synthetic
Peterson et al.	US	hospital admission	Lasso, Ridge, EN, RF, GBT, Perceptron, SVM, KNN, Ensemble	8,439	80 train, 20 test	yes	no	yes	yes	yes	yes	no	no	Synthetic
Sanz; Reverter; Valim	US, Canada, Mexico	survived	SVM, KC, Cox	50 and 300	5- folds Nested CV	yes	no	no	yes	yes	yes	yes	no	Synthetic
Tomasev et al.	US	postoperative complications	LSTM, LR	703,782	80 train, 5 validation, 5 calibration, 10 test	yes	no	yes	yes	no	yes	yes	yes	In time and space
Trinhammer et al.	Denmark	risk of crime after psychiatric admission	LR, RF, XGBoost, LightGBM	45,720	80 train, 20 test	yes	no	yes	yes	yes	yes	no	yes	Synthetic
Wang; Li; NG	China	child health diagnosis	RF, LR	Not informed	train and test	yes	no	yes	yes	yes	yes	no	no	Synthetic
Wilimitis et al.	US	risk of suicide	Lasso, RF	120,398	5-folds CV	yes	yes	yes	yes	no	no	yes	yes	Synthetic
Xie et al.	US	risk of death	LR, SR, Lasso, RF	58,976	70 train, 10 validation, 20 test	yes	no	no	yes	no	no	yes	no	Synthetic
Yuan et al.	China	sepsis	XGBoost, LR, SVM, NN	434	80 train, 20 test	yes	no	yes	yes	yes	yes	yes	no	Synthetic

Note: US = United States, NB = Naive Bayes, LSTM = Long short-term memory, RF = Random Forest, EN = Elastic Net XGBoost = Extreme Gradient Boosting Trees, LR = Logistic Regression, DT = Decision Tree, SVM = Support Vector Machine, NN = Neural Network, GBT = Gradient Boosting trees, KNN = K-nearest neighbors, KC = Kernel Cox, LightGBM = Light Gradient Boosting Machine, SR = Stepwise Regression, CV = Cross Validation, Synthetic: result of shuffling and randomizing data for training and testing division.

Performance measures

The metrics most reported by the manuscripts were AUC (100%), Area Under Precision-Recall Curve (AUPRC) (12.5%), precision (68.8%), sensibility or recall (68.8%), specificity (43.8%), negative predictive value (NPV) (31.3%), accuracy (50%), and f-score (56.3%).

Overall findings

Every article performed a supervised classification learning analysis. Generalization issues were analyzed by time/period, and synthetic generalization was tested through random division and by shuffling the database into training and testing, both in 70-30 and 80-20 training and testing division, and also using cross-validation with 5 and 10 folds. Around 88% of the papers used synthetic generalization. The two papers that tested generalization by analyzing its results in another dataset or in prospective data were 56 and 59, which performed generalization analysis in two hospitals and in different years, respectively.

4.1.4 Discussion

We performed a scope review of studies that analyzed the generalization of machine learning algorithms in healthcare. Our survey identified 16 articles from platforms such as PubMed, Embase, and Web of Science. We found that despite the issue of generalization being frequently discussed, few studies have actually tested the generalization of machine learning algorithms with real patient data.

The studies included in this review showed that the generalization of machine learning models applied to health outcomes is important, but the methodology to perform this generalization is still incipient. Most of the literature has showed the feasibility of synthetic generalizing the models, with only two papers (56, 59) testing the generalizability of time or space. Synthetic generality was frequently defined as the unseen data (test) resulting from the randomization of the total data, thus being an indication of adaptation

of the model to new data but without the characteristic of being in a different time and/or space.

There are many benefits, especially economic ones, regarding the generalization of health models. For companies and model developers, it allows the creation of a single model and being able to sell it to multiple buyers. For health services, ensuring the generalization of models applied in the system improves local development and equity. However, these same reasons can incentivize the enforcement of algorithms through unreliable predictions.

Algorithmic generalizations for vulnerable subgroups (ethnic-racial, and groups for different socioeconomic backgrounds) are especially important and necessary, but even these may not always be technically possible. Despite not being a consensus, when identifying the limitations of the model created, it may be more interesting to have a model built specifically for this minority group, tailoring the model and therefore having more than one model for the same outcome. This study has a few limitations. The first is regarding the choice of keywords for the search engines, which can change frequently in a relatively new area such as machine learning for healthcare. Another limitation is the different ways of understanding the generalization of models by the studies, with the absence of standardization of paradigms to assess the quality of the generalization of algorithms. Despite its challenges, this type of bibliographical survey helps in the search for clarifications about commonly used terms and the correct way to apply them.

This study points to the many challenges of applied generalization studies in machine learning for healthcare data. There is a growing need within the literature for a better understanding of generalization issues and how to measure them. As the field matures, new challenges will emerge regarding how to scale locally-validated models. While generalizable models are desirable, they should not prevent accurate local models from being used.

Contributors

Author 1 (MF): Conceptualization, data curation, methodology, writing – review & editing.

Author 2 (KAM): Data curation, writing – review & editing.

Author 3 (HSS): Writing – review & editing.

Author * (ACF): Conceptualization, supervision, writing – review & editing.

Declaration of Interests

None.

Funding

CAPES.

Data Sharing Statement

Appendix. Supplementary materials

Ethics approval

None.

4.2 Artigo 2. Population-based validation of machine learning models for neonatal mortality prediction

Keywords: Neonatal mortality, Birth records, Population Health, Health prediction, Machine learning, Artificial Intelligence

Abstract

Despite recent improvements, neonatal mortality rates remain high worldwide, especially in developing countries. Predictive machine learning (ML) algorithms can improve targeted public policies and increase preventive measures in neonatal health. In this study, we evaluated the generalization ability of machine learning algorithms to predict neonatal deaths, testing their performance across different geographic regions of Brazil, a highly unequal country. We analyzed linkage data from the two main Brazilian national health sources, i.e. the Information System on Live Births (SINASC) and the Mortality Information System (SIM). The algorithm was trained with 2012-2015 data from São Paulo, the most populous city in Brazil, and included 807,932 newborns and 5,518 neonatal deaths. The test set comprised data from all Brazilian cities with at least two occurrences of neonatal deaths among children born in 2016, which included 2,848,052 live births and 23,948 neonatal deaths. We applied a popular machine learning algorithm with ten predictor variables routinely collected for newborns in Brazil: place of birth, mother's age, mother's education, mode of delivery, Apgar scores at 1st and 5th minutes, birth weight, presence of congenital anomaly, and gestational age. In the city of São Paulo, where the algorithm was originally trained, we found an AUROC of 0.97, which decreased only slightly for other state capitals (with a mean of 0.96 and a standard deviation of 0.01), but more for other Brazilian cities (mean of 0.91, and a standard deviation of 0.11). The results bring new insights to the generalization ability of artificial intelligence models in highly unequal areas. We identified that although there was a reduction in the predictive performance of the algorithm, the metrics remained acceptable even for smaller cities.

4.2.1 Introduction

Neonatal mortality is defined as the death of newborns within their first 28 days of life (66), and is considered an important overall public health indicator (67). Within the last three decades, there was a decrease of approximately 52% in the worldwide neonatal mortality rate, but this decline was lower than the observed for the mortality of children up to 5 years old during the same period (68). Also, the magnitude of the reduction in neonatal mortality in Latin America and the Caribbean is decreasing over time, and is below the global average (69).

In 2019, around 2.4 million neonatal deaths occurred worldwide (69). In developing countries, the neonatal mortality rate is still much higher than in developed countries (68, 70). Recent studies have found that in low-middle and upper-middle-income countries, neonatal deaths account for approximately 57% and 55% of deaths of children under 1 year of age, respectively (71).

In Brazil, although there has been a consistent reduction in infant mortality during the last 30 years, the result is still considered insufficient for international standards (72). In 2020, there were 16,716 preventable neonatal deaths recorded in the country (73), with the highest rates concentrated in the North and Northeast regions of the country, reflecting a well-known scenario of socioeconomic inequalities.

The third Sustainable Development Goals (SDG) is to eradicate preventable deaths of newborns and children under 5 years of age in all countries (74, 75). Over the last few decades, the implementation of several public policies has contributed to the reduction of the neonatal mortality rate in Brazil (76, 77, 78, 79). Even considering these advances, the country still needs to eradicate its high neonatal mortality from preventable causes (80, 81) in order to achieve the SDG target by 2030.

Artificial intelligence algorithms have the potential to support stakeholders and healthcare professionals in improving decision-making (82). Machine learning algorithms have already been used to predict high-risk pregnancy (83, 84), premature mortality (85, 86), neonatal mortality (87, 88), fetal growth and weight (89, 90, 85, 91), conditions associated with maternal health (92, 93), among others, but the generalization ability of these algorithms in diverse and unequal settings is still unknown.

In addition to predicting neonatal mortality with high performance, being able to apply an algorithm trained on one city to a different sociodemographic context is an important asset to health systems. This study aims to evaluate the generalization capacity of machine learning models by evaluating the performance in all Brazilian cities of an algorithm developed in the most populous Brazilian city. Considering the continental dimension of Brazil and its heterogeneity of income and health infrastructure, these results can assist clinical decision-making by improving the use of machine learning models for all regions of the country.

4.2.2 Methods

Design and settings

We used linkage data from two official Brazilian national registries: the Information System on Live Births (SINASC, acronym from the Brazilian name Sistema de Informações sobre Nascidos Vivos) and the Mortality Information System (SIM, acronym from the Brazilian name Sistema de Informação sobre Mortalidade). Both information systems are part of the official Health Surveillance systems from the Brazilian Ministry of Health.

Data linkage procedures were carried out by the Brazilian Ministry of Health. It was performed in three steps, two of which were deterministic and responsible for 89% of the total linkage, and one that was probabilistic, adding 4% new data. In the first step, the linkage was performed by the ID of the live birth registry from both systems. The second step resulted from matching a key composed of the mother's name, date of birth, and sex in both systems. Finally, the previous key was also used for the probabilistic linkage of the data.

A predictive machine learning algorithm was developed using only data from de city of São Paulo, which has the highest concentration of births in Brazil. The training model included data from the years 2012 to 2015 and comprised 807,932 newborns and 5,518 neonatal deaths. This database was provided by the municipal health department of Sao Paulo and its probabilistic linkage was conducted using the mother's name, date of birth, and the name of the deceased child (94).

In order to evaluate the performance of the model on new, unseen data, we used data from the year 2016 from the 3,842 cities within all 27 Brazilian states with at least

two occurrences of neonatal deaths among children born in 2016, which represent 69% of all Brazilian cities.

We used the following variables for the development of the predictive model: place of birth (hospital, other healthcare facilities, at home, others), mother's age in years, mother's education (categorized as formal education, incompleted primary education, completed primary education, completed high school, some college, and completed college or more), mode of delivery (vaginal or cesarean), 1st and 5th minutes Apgar scores, birth weight in grams, presence of congenital anomaly, and gestation age in weeks. The age of the children at death was used as a proxy for death and recategorized as a binary outcome.

Data analysis

All analyses were conducted in Python. We used a popular machine learning algorithm (XGBoost) with the following previously tuned hyperparameters, selected using Hyperopt on the training data: maximum tree depth for base learners, 5; subsample ratio of the training instance, 0.95; subsample ratio of columns when constructing each tree, 0.7; the number of gradients boosted trees, 150; balancing of positive and negative weights, 3; the minimum sum of instance weight needed in a child, 6; minimum loss reduction required to make a further partition on a leaf node of the tree, 0.0555; and L1 regularization term on weights, 0.5. The hyperparameters optimization was performed with scoring F1 and 100 maximum evaluations.

We performed the analyses using the minimum data set of perinatal indicators from the World Health Organization (95), i.e., maternal age, place of delivery, mode of delivery, birth weight, and gestational age. Additionally, we tested the model further including 3 important variables: 1st minute Apgar score, 5th minute Apgar score, and congenital anomaly. We also performed sub-group analyses, with variables not added in the general model, for vulnerable populations to analyze the fairness of the model, by evaluating the performance of the algorithm among nonwhite mothers, mothers with low education, and teenage mothers (¡ 18 years).

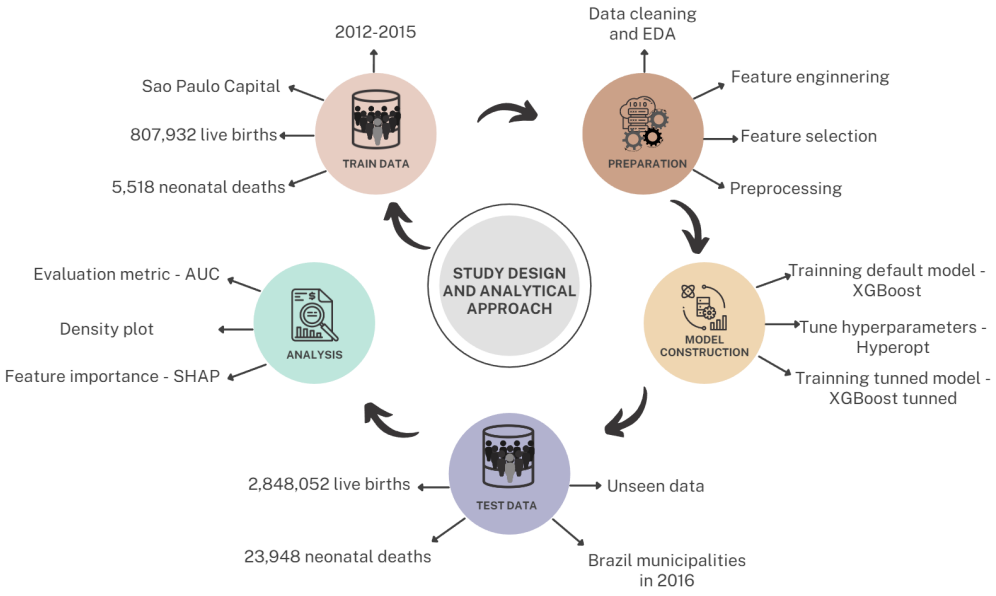
The performance of the models was evaluated based on the following metrics: area under the receiver operating characteristic curve (AUC), sensitivity (recall), specificity, positive predictive value (PPV) or precision, negative predictive value (NPV), the area

under the prediction-recall curve (AUPRC), and the F1-score. We also estimated Brier Scores, to analyze the calibration of the predictive models.

4.2.3 Results

Figure 2 summarizes the machine learning analysis performed in this study with training data from the municipality of São Paulo, and test data for all Brazilian municipalities.

Figure 2 – Representation of the study design and analytical approach



The main characteristics of the training data, which comprised 807,932 live births and 5,518 neonatal deaths, are included in Supplement Table 1. Over half of the training population were children whose mothers completed high school, 54% identified as White, and more than 99% of the births occurred at a hospital facility. The prevalence of congenital anomalies was 2%.

The analysis that only included the WHO's minimum data set of perinatal indicators achieved an AUC of 0.911 (0.90-0.92 CI95%), 0.636 precision, 0.437 recall, 0.518 F1-score, 0.487 AUPRC, and 0.01 of Brier score. Despite the overall high predictive performance, when including variables such as 1st minute Apgar score, 5th minute Apgar score, and congenital anomaly, there is an important increase in the selected metrics. Specifically, an AUC of 0.965 (0.96-0.97 CI95%), 0.632 precision, 0.580 recall, 0.605 F1-score, 0.612

AUPRC, and 0.01 of Brier score. Therefore, we chose to keep all analysis results including these three variables as well.

Table 2 shows the death rate and the predictive performance of the model in each state capital. The neonatal mortality rate ranged from 4.14 in Vitoria, in the Southeast state of Espirito Santo, to 16.07 in Porto Velho, the capital of the Northern state of Rondônia. Sensitivity and negative predictive values were very high, i.e., > 99% for all models, which was expected given the rare nature of the output.

Table 2 – Predictive performance for neonatal mortality on the testing set with the XGBoost model for the capitals of Brazilian states, 2016.

Region	State	Capital	Death rate	AUC	Precision	Recall	F1	AUPRC	Brier
North	Acre	Rio Branco	7.92	0.930	0.585	0.425	0.492	0.445	0.007
	Amapá	Macapá	12.03	0.935	0.784	0.446	0.569	0.608	0.008
	Amazonas	Manaus	8.45	0.939	0.800	0.490	0.608	0.637	0.005
	Pará	Belém	14.83	0.958	0.750	0.456	0.567	0.614	0.010
	Rondônia	Porto Velho	16.07	0.962	0.815	0.469	0.595	0.699	0.010
	Roraima	Boa Vista	8.63	0.952	0.782	0.544	0.641	0.644	0.005
	Tocantins	Palmas	12.84	0.968	0.808	0.598	0.688	0.743	0.007
Northeast	Alagoas	Maceió	10.44	0.944	0.677	0.496	0.572	0.565	0.008
	Bahia	Salvador	14.10	0.978	0.707	0.646	0.675	0.727	0.009
	Ceará	Fortaleza	11.14	0.958	0.656	0.633	0.645	0.663	0.008
	Maranhão	São Luís	13.89	0.968	0.680	0.652	0.666	0.676	0.009
	Paraíba	João Pessoa	10.43	0.965	0.688	0.505	0.583	0.616	0.008
	Pernambuco	Recife	14.51	0.964	0.706	0.533	0.607	0.651	0.010
	Piauí	Teresina	15.84	0.965	0.802	0.519	0.630	0.696	0.010
	Rio G. do Norte	Natal	10.46	0.944	0.663	0.540	0.595	0.599	0.008
	Sergipe	Aracaju	15.63	0.950	0.750	0.513	0.610	0.651	0.010
Midwest	Distrito Federal	Brasília	8.07	0.940	0.517	0.555	0.577	0.600	0.008
	Goiás	Goiânia	14.40	0.957	0.677	0.555	0.610	0.631	0.010
	Mato Grosso	Cuiabá	10.67	0.970	0.712	0.542	0.615	0.658	0.007
	Mato G. do Sul	Campo Grande	8.86	0.950	0.605	0.526	0.563	0.610	0.007
Southeast	Espírito Santo	Vitória	4.14	0.937	0.714	0.444	0.548	0.534	0.003
	Minas Gerais	Belo Horizonte	8.61	0.957	0.589	0.608	0.599	0.620	0.007
	Rio de Janeiro	Rio de Janeiro	8.78	0.959	0.662	0.562	0.608	0.616	0.006
	São Paulo	São Paulo	7.47	0.965	0.632	0.580	0.605	0.611	0.006
South	Paraná	Curitiba	6.74	0.954	0.608	0.558	0.582	0.578	0.005
	Rio Grande do Sul	Porto Alegre	8.59	0.966	0.524	0.629	0.572	0.563	0.008
	Santa Catarina	Florianópolis	6.81	0.994	0.713	0.760	0.735	0.806	0.004

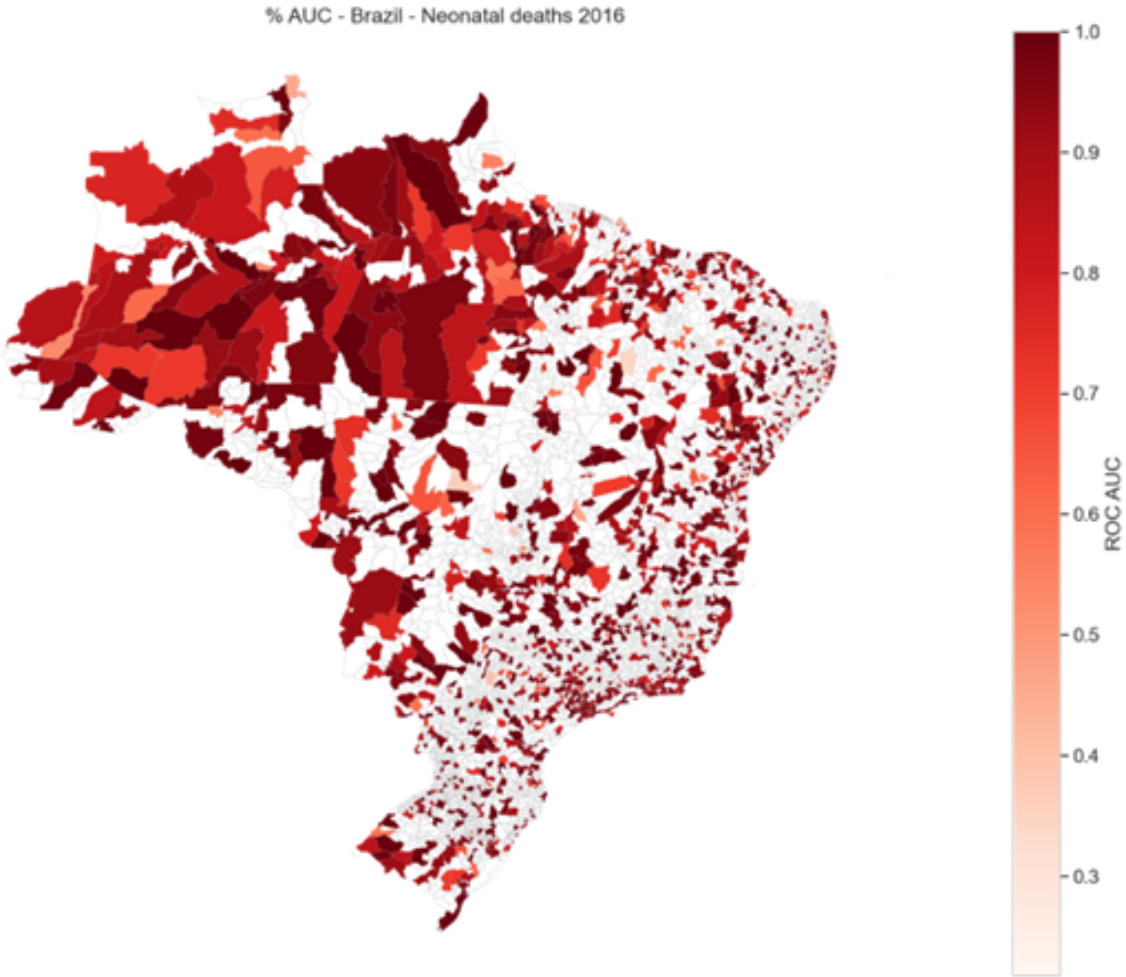
Note: AUC = area under the receiver operating characteristic curve; Recall = sensibility; Precision = positive predictive value; F1 = F1 score; AUPRC = area under prediction-recall curve; Brier = Brier Score.

The mean values of the AUC was 0.957 (with a standard deviation (SD) of 0.01), 0.689 for precision (0.08 SD), 0.548 for recall (0.08 SD), 0.606 for F1-score (0.05 SD), 0.632 for AUPRC (0.07 SD), and 0.01 for the Brier score (0.00 SD). Brier score values show how calibrated the models are in relation to the outcome and all state capitals presented models with a Brier score close to 0, indicating good calibration. All capital cities presented a very high AUC, with the lowest observed in Rio Branco (0.930). The northern region of Brazil had the lowest AUC values (approximately 0.949), while the

northest and south regions presented very similar, and high, performance metrics (0.960 and 0.971 respectively). The highest precision and sensitivity values were observed in the northeast region of the country.

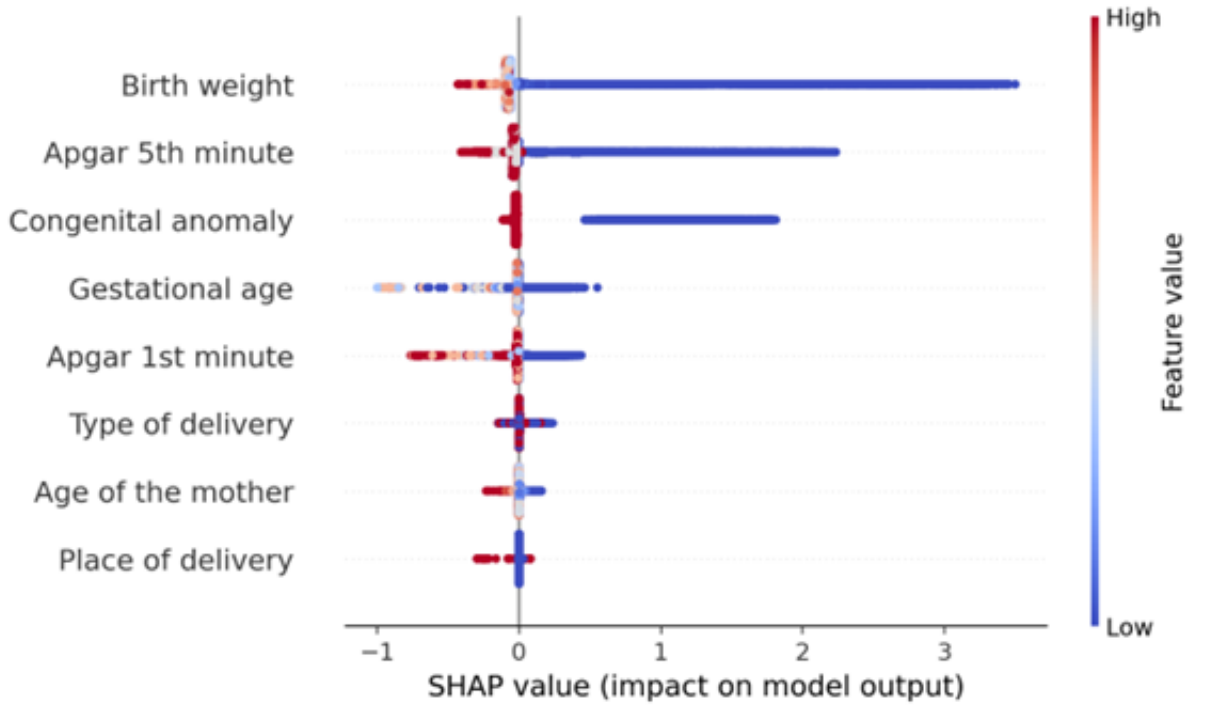
Figure 3 shows the area under the ROC curve for the municipalities with at least 2 deaths in 2016. A total of 1,161 cities were analyzed, distributed among all geographical areas of the country, and varying considerably by size. The mean AUC among all cities was 0.902, with a corresponding standard deviation of 0.12.

Figure 3 – Map of the AUC performance of Brazil municipalities to predict neonatal deaths in 2016.



The five variables with the highest predictive importance according to the Shapley Value are shown in Figure 4. In this graph, red shows that positive results increases the probability of neonatal mortality. Therefore, higher birth weight reduced the predicted increased risk of death, as well as not having congenital mortality and having values between 9 and 10 in the Apgar score in the first and fifth minutes.

Figure 4 – SHAP feature importance on the test set.



Supplemental Tables 5-7 report on findings from the sub-group analysis. All sub-group analyses were performed with vulnerable subgroups living in all 27 Brazilian capitals and the federal district. Overall, high model performance was observed for all subgroups: the mean AUC of neonatal mortality among children from teenage mothers and from low-education mothers was 0.947 for both groups, while for nonwhite mothers the AUC was 0.954.

4.2.4 Discussion

Our results found a high generalizing ability of the machine learning model for predicting the risk of neonatal death in a diverse and unequal country. Among the 27 Brazilian state capital cities, all models yielded an AUC higher than 0.93. Our models were developed using perinatal information from the mother and the baby that are routinely collected for live births in Brazil, which indicates its potential of being used for public policies at the national level.

Brazil is a diverse country with continental dimensions. The sociodemographic profile of the cities varies greatly within the country, as well as their healthcare resources. The prediction based on variables that are routinely collected allows the development

of models in other settings, and similar approaches should be tested for other countries. While our study showed the feasibility of generalizing machine learning models trained in one city to other locations with a different profile, there is also evidence of the feasibility of generating models to predict perinatal and neonatal mortality in high perinatal mortality rate areas (96) and in resource-limited settings (97). Neonatal mortality is an important public health indicator, and the implementation of data-driven tools based on a small number of easily collected predictors can contribute to reducing inequalities within and between countries.

The immediate postdelivery is a window of opportunity to identify the risk of neonatal death, and the predictors that contributed the most to our models, namely birth weight, congenital anomaly, and Apgar scores at the 1st and 5th minutes, are all collected at delivery. A recent systematic review on the topic identified that the most used attributes to predict neonatal mortality were birth weight, gestational age, child sex, Apgar score, maternal age, and the number of pregnancies (98). Regarding the time of data collection, a study with around 500,000 pregnancies developed several models for stillbirth and neonatal prediction and observed a robust increase in model performance when it included information from the delivery and the day after delivery (97).

We used deaths by place of occurrence in order to create a model based on characteristics related to the moment of delivery and markers of severity at birth, which are influenced by factors related to the availability and access to the health service that performs the delivery or that has technology for neonatal care, so it is not possible to guarantee that this is an appropriate model to estimate risk of death during prenatal care.

The study has a few limitations. First, the most recent data used was from 2016, which is the year of the last national linkage available from the Ministry of Health. Second, we observed some variability between the vulnerable subgroups used in our secondary analysis, which requires closer monitoring of the model performance for these groups. Furthermore, the fact that the models generalized well in our study does not mean that the results will be necessarily replicated for a different outcome. Nevertheless, neonatal mortality is a relatively stable outcome, therefore only a small variation in the data distribution is expected, favoring the generalization of the models.

In conclusion, we tested the generalization of a machine learning algorithm to predict neonatal deaths between cities with a very diverse sociodemographic profile. The Brazilian Public Health System is undergoing a process of development and integration of

electronic health records, and it is expected that it will soon be possible to use data from newborns to readily inform the predicted risk of neonatal death. The development of tools to assess the risk of neonatal death can be an important asset to national health systems, and an important asset to achieving the SDG to end preventable deaths of newborns.

Contributors

Author 1 (MF): Conceptualization, data curation, methodology, writing – review & editing.

Author 2 (HSS): Writing – review & editing.

Author 3 (HGS): Methodology, review.

Author 4 (AFMB): Data curation, methodology.

Author 5 (CSGD): Conceptualization, review & editing.

Author * (ACF): Conceptualization, supervision, writing – review & editing.

Declaration of Interests

None.

Acknowledgments

Dácio Rabello, Maria de Fátima Marinho Souza, Carmen Simone Grilo Diniz, Margarida M. Tenório de Azevedo Lira, Marcel Reis Queiroz, Alexandre Chiavegatto Filho, Renata Wassermann, Eliana de Aquino Bonilha, André Filipe Batista, Jéssica Reis Queiroz, Maria Elizangela Ramos Junqueira, Celia Maria Castex Aly, Roberto Aparecido Moreira, Marina de Freitas. All participants in the research "Potential days of pregnancy lost (DPGP): an innovative measure of gestational age to assess maternal and child health interventions and outcomes"were funded in the so-called data science approaches to improve maternal and child health in Brazil by the CNPq and the Bill and Melinda Gates Foundation.

Funding

MF received a scholarship from Coordination for the Improvement of Higher Education Personnel (CAPES), and CSGD is responsible for the research Making "Too much, too soon" interventions in childbirth more visible to the information system, which received funding from Bill and Melinda Gates Foundation (Grant n° ID INV-027961) and National Council for Scientific and Technological Development (CNPq) (process number 445847/2020-4). https://www.gatesfoundation.org. https://www.gov.br/cnpq/pt-br. The funders didn't play any role in the study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Data Sharing Statement

Appendix. Supplementary materials

Ethics approval

Approved by the ethics committee of the School of Public Health of the University of Sao Paulo (CAEE: 98163018.2.0000.5421).

4.3 Artigo 3. Generalization of transfer learning algorithms for tabular healthcare

Keywords: Health tabular data, Prediction, Transfer Learning, Artificial Intelligence

Abstract

Background: Transfer learning is a promising new tool in healthcare and refers to the ability of artificial intelligence algorithms to learn predictive tasks and adapt the generated learning to other tasks. It has the potential to complement diagnostic and prognostic tasks in healthcare, especially in areas with a small number of patients for training the algorithms. Methods: We compared the predictive performances of popular machine learning algorithms to predict Covid-19 mortality in a sample of 18 Brazilian hospitals. We first used the algorithms to select the hospital with the highest potential for applying transfer learning and then used its patients to train neural networks with local data, using the learning of a selection of these layers for transferring the models to other hospitals. Findings: The results showed that despite being promising, the TL method for tabular data used conventionally in neural networks did not obtain superior results in boosting models with local training and testing, presenting limitations in relation to the size of the training sample. Interpretation: Transfer learning models for tabular data can be useful over the conventional technique approaches, but some technical challenges are still present.

4.3.1 Introduction

The use of Machine Learning (ML) techniques has become increasingly relevant in various fields, such as healthcare. By enabling the recognition of patterns through access to patient data, these techniques have the potential to assist professionals in better clinical decision-making (6, 8, 7). These tools aim to minimize errors and enable the humanization of care, directing the attention of healthcare professionals only to patients.

Transfer Learning (TL) is a technique that allows models trained on specific tasks to be applied to solve related tasks (39, 40, 41, 42), eliminating the need for the training data to be independently and identically distributed from the test data (37). Using these methods, it is possible to obtain acceptable performance results even when data distributions, space, and/or time change (38), which is especially relevant given the complexities of healthcare.

The quality of access to healthcare in Brazil has been a topic of scientific interest, particularly in light of the challenges faced during the Covid-19 pandemic, which highlighted the weaknesses of the country's healthcare system (99, 100, 101, 98, 102), as well as in other countries around the world (103, 104, 105). In this context, developing and adopting new technologies can significantly contribute to the improvement of healthcare services provided to the population.

Algorithmic generalization through transfer learning is a well-known but still poorly disseminated technique that can significantly contribute to solving clinical decisions (106). Health information systems have a wide variety of data types, with a considerable portion being structured or tabular data, i.e., information organized in rows and columns. These data often contain relevant information for health predictions, such as symptoms, medical history, vital signs, sociodemographic variables, and diagnostic test results. Several studies have used tabular data for applying artificial intelligence techniques in healthcare, demonstrating its potential (30, 23, 25, 27, 31, 32, 33, 26, 28, 29, 24).

This article presents an approach for artificial intelligence (AI) algorithmic generalization applied to healthcare, using TL for predicting death from Covid-19 tabular data in all five Brazilian geographic regions. Model generalization will be crucial for AI applications in healthcare, especially in countries with great geographic and population diversity, such as Brazil, and in situations where algorithms do not generalize well with

standard models. We propose a methodology for knowledge transfer from pre-trained models on health data from one region to another, with the aim of improving model generalization and consequently promoting their applicability in different contexts.

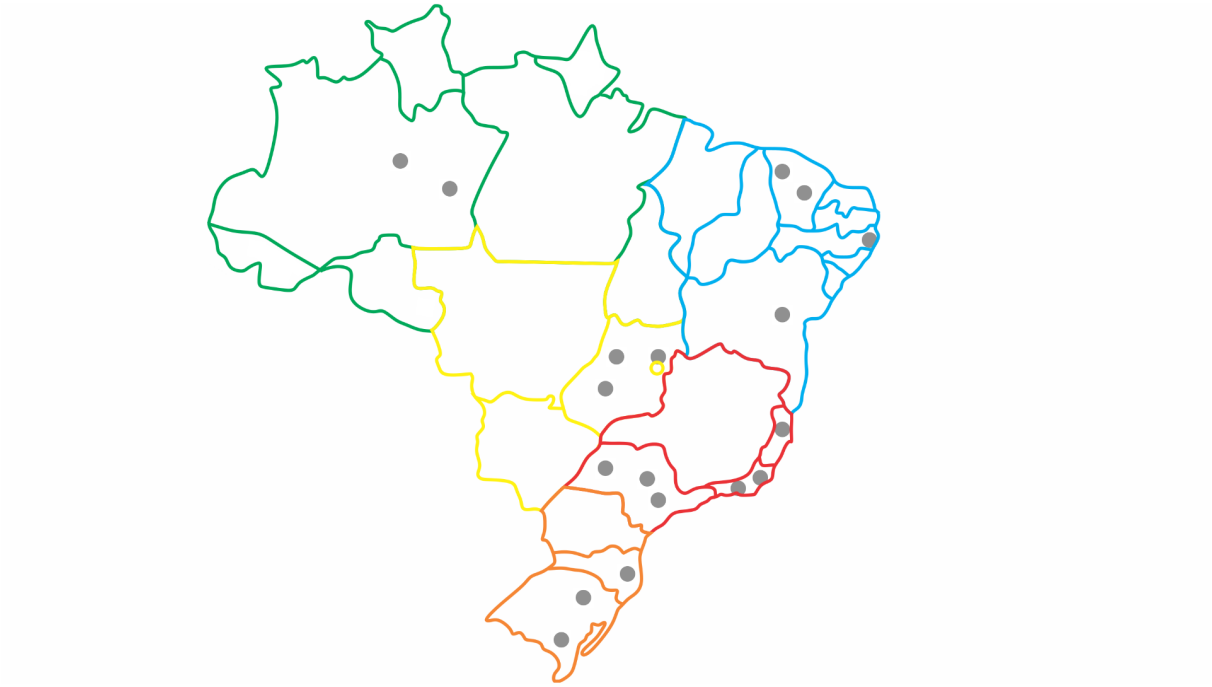
4.3.2 Methods

The Artificial Intelligence Network for Covid-19 in Brazil (IACOV-BR) is a retrospective observational cohort of Brazilian hospitalized patients with suspected or confirmed Covid-19 diagnoses from March to August 2020. It includes 16,236 individuals with clinical, laboratory, and demographic data from 18 hospitals across the five geographic regions of Brazil (Figure 5).

The first study of this cohort (107) aimed to identify the best data aggregation strategy to construct the training of models to predict invasive mechanical ventilation, admission to the intensive care unit (ICU), and death. The authors tested using only local data, and seven other strategies, including adding a different number of patients from other hospitals to the training data based on their geographic location. Predictive performance for some of the hospitals was very low due to the smaller volume of input data, which could potentially be solved with a transfer learning approach. In addition, among the compared strategies, the best performances were obtained with only the use of training data from the same hospital, emphasizing the difficulty of models in dealing with changes in data distribution.

For this analysis, a broad range of sociodemographic, clinical, and laboratory predictors were obtained from routinely collected electronic health records and used for the algorithms. These predictors were age (years), gender (male, female), systolic blood pressure (mmHg), diastolic blood pressure (mmHg), body temperature (continuous in Celsius degrees), heart rate (number of heart beats per minute), respiratory rate (number of respiratory movements per minute), hemoglobin count (g/dL), platelets count (quantity/mm^3), mean corpuscular hemoglobin concentration (g/dL), red cell distribution width (%), mean corpuscular hemoglobin (pg), leukocytes count (quantity/mm^3), neutrophils count (quantity/mm^3), lymphocytes count (quantity/mm^3), basophils count (quantity/mm^3), and monocytes count (quantity/mm^3).

Figure 5 – Map of Brazil with hospitals in the IACOV network according to geographic region.



This study was approved by the Research Ethics Committee (IRB) of the University of Sao Paulo (CAAE: 32872920.4.1001.5421), which included a waiver of informed consent. More details about the IACOV-BR methodology can be assessed in the published literature (107).

In order to apply TL with acceptable performance in new data, it is necessary to confirm that the model to be transferred obtained acceptable performance for its local test set. We therefore first analyzed the results of models built for each of the 18 hospitals locally (training and testing in the same hospital), using the holdout method that randomly divides the dataset into two subsets, with 70% allocated to the model training and 30% allocated to testing. For continuous variables were standardized with a mean of zero and a standard deviation of one using the Z-score method. Missing data for continuous variables were imputed by the median due to distribution of data. The mean of the variables in pre-processing was applied to the test set in order to avoid data leakage problems. The 10-fold cross-validation approach was applied to the training set for model hyperparameter optimization using the Hyperopt technique. Boosting algorithms commonly used in tabular data analysis (108) such as Extreme Gradient Boosting (XGBoost), Catboost, and Light Gradient Boosting Machine (LighGBM) were compared to a neural network model with Multilayer perceptron (MLP). This decision was made because boosting models are the

most popular and frequently show good performance (greater than 0.7) for this type of data, and TL applications (example in the Figure 6) are only possible through neural networks with more than one hidden layer.

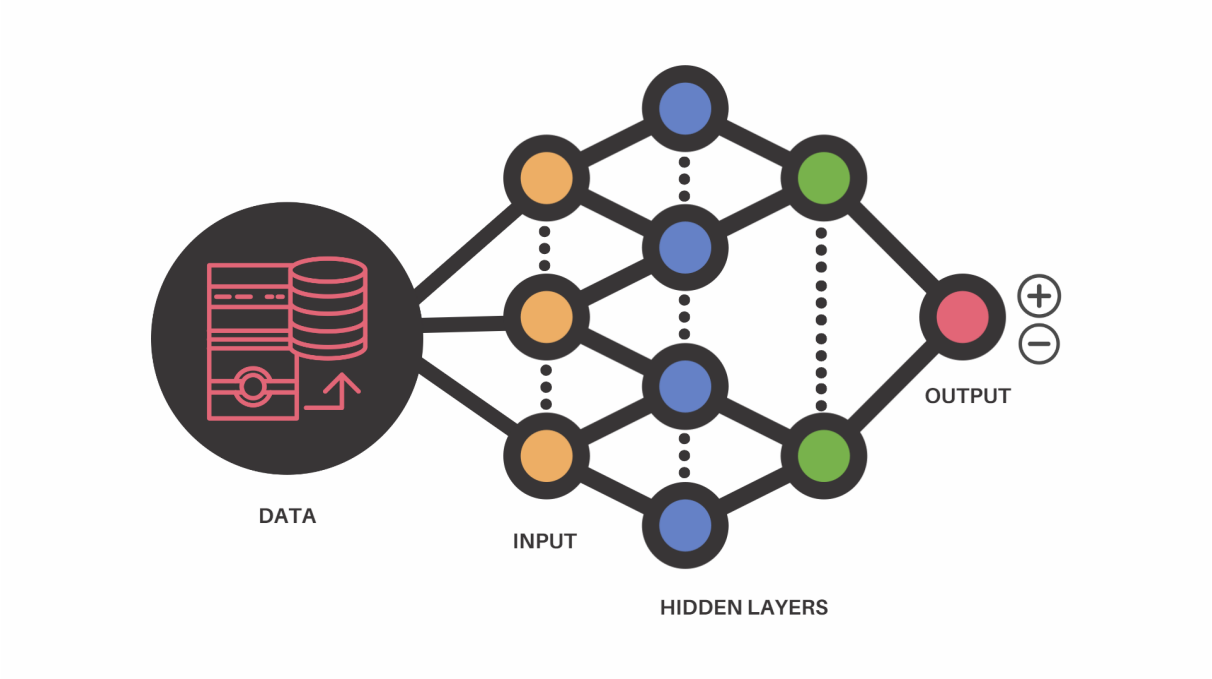
The boosting algorithms and the neural network served as a baseline for the results of the multi-layer neural network models. Through this initial analysis, we identified that the hospital with the best local performance in terms of outcome balancing was the Hospital de Clínicas da USP (HC-USP) located in the municipality of São Paulo. The incidence of death among HC-USP patients was 36%. For the MLP algorithm, the HC-USP local algorithm yielded an AUROC of 0.809 (95% Confidence Interval 0.77;0-85), with a recall of 0.528 and specificity of 0.879. Therefore, this hospital was selected as the reference for the transfer learning process.

In the following step, we used the Boruta algorithm, which selected a minimal model with 11 variables: age, gender, hemoglobin count, platelets count, mean corpuscular hemoglobin concentration, red blood cell distribution width, mean corpuscular hemoglobin, neutrophils count, leukocytes count, lymphocytes count, and monocytes count. All results from here are displayed with just the variables chosen by Boruta.

We also analyzed the training of the MLP for the selected hospital for the final model (model to be transferred). In this case, the same steps of data division mentioned earlier were performed, but now with hyperparameter optimization performed through 10-fold stratified k-fold cross-validation using GridSearch, which exhaustively evaluates all possible combinations of hyperparameters (48). The best set of selected hyperparameters was applied for training the baseline model. The baseline MLP was then applied to the local test data to compare the performance achieved for other algorithms. Finally, the best threshold was selected for the final model for the comparison between the metric f1-score superior to 0.7 (previously established) and the error value in the final model. Model selection was based on the predictive performance of the area under the ROC curve (AUC). In addition to AUC, metrics such as sensitivity, specificity, precision, negative predictive value (NPV), F1 score, and Brier score were also reported.

Finally, TL was performed by means of testing the model trained for the reference hospital (HC- USP) for the remaining 17 hospitals, allowing the comparison with their local predictive results. Using the model with variables resulting from Boruta, for each hospital was used the Adam optimizer with a significantly low learning rate of 0.0001, and binary cross-entropy loss function. The Adam optimizer was selected because it is

Figure 6 – Architectural example of a transfer learning model



computationally more efficient. A lower learning rate is expected to make training slower, but it can help the model to converge to a more stable and better solution.

4.3.3 Results

We analyzed a total of 8,477 individuals with positive RT-PCR results for Covid-19 and applied a transfer learning framework from the reference hospital (HC-USP) to 17 other hospitals from different regions of Brazil to predict patient mortality. Supplementary materials include Table 1 with a description of demographic data from the combined datasets, Table 2 with the optimized hyperparameters for each model in the local analysis, Table 3 with the results of the local test models for all variables, and Table 4 with the results of the local test models for the Boruta-selected variables. For the transfer learning analysis, the best hyperparameters selected for the neural network through GridSearch were a dropout rate of 0.3, a learning rate of 0.001, and 32 neurons.

The results of the NN model, as a reference for the application of transfer learning, trained and tested at HC-USP, obtained an AUC curve of 0.809, with a precision of 0.708, sensitivity of 0.528, F1-score of 0.605, specificity of 0.879, and negative predictive value of 0.770, with the best threshold at approximately 0.4. The same model was tested with Boruta variables. These results can be accessed in Table 12 of the supplementary material.

The results of the NN model correspond to the local analysis, with training and testing of the algorithms for each hospital separately using two approaches: one with all variables from the database and another with variables selected by the Boruta variable selection algorithm. All metrics presented refer to test data. Only the HC-USP results are presented for two models, one trained with all variables (AUC curve of 0.809, with a precision of 0.708, sensitivity of 0.528, F1-score of 0.605, specificity of 0.879, and negative predictive value of 0.770, with the best threshold at approximately 0.4), and the other with variables selected by Boruta (Table 12). For other hospitals, consult Table. The results of the TL application are shown in Table 3.

However, in some cases when learnings from the reference model were applied to other hospitals, the models started to overestimate positive results with an excess of false positives, indicating that the negative class was largely ignored. For three hospitals (Hospital São Francisco, Hospital Escola da UFPel, e Hospital de Urgências de Trindade), the low sample size made it impossible to obtain a performance result for neural networks models. Most hospitals performed better in local models with boosting algorithms (LightGBM and XGBoost), indicating that TL with neural network algorithms did not improve predictive performance. For hospitals with a larger sample, the results of boosting models and Neural Networks were closer, showing the a higher ability of Neural Networks to deal with a larger number of observations.

4.3.4 Discussion

Our study developed a classification transfer learning model for structured data to predict Covid-19 mortality using data from the largest country in Latin America. Our findings demonstrated the feasibility of applying transfer learning techniques to structured health data, however, our results showed that the technique was not superior to models that trained and tested on local data.

Predictive models derived from generalization techniques for patient prognosis have the potential to support decisions in large and diverse areas. However, environments with low data collection can be a limiting factor for obtaining results that can improve the flow of care services on a large scale. As deep neural networks often require many observations, non-conventional ways of applying for learning transfer, using boosting algorithms, could be able to obtain better results. A certain level of generalization from ML models, especially

Table 3 – Results of transfer learning for neural network in the test data.

Place	**AUROC**	**Precision**	**Recall**	**F1**	**Specificity**	**NPV**
Hospital Santa Casa São Paulo - SP	0.709	0.574	0.670	0.618	0.647	0.734
Hospital de Clínicas da USP São Paulo - SP	0.790	0.613	0.640	0.626	0.776	0.795
Hospital Português da Bahia Salvador - BA	0.826	0.238	0.536	0.330	0.874	0.962
Hospital UNIMED Fortaleza - CE	0.709	0.321	0.273	0.295	0.914	0.894
Hospital Regional de Luiziania Luiziania - GO	0.606	0.674	0.586	0.627	0.556	0.461
Hospital Moinhos de Vento Porto Alegre - RS	0.918	0.302	0.929	0.456	0.756	0.989
Hospital UNIMED Rio de Janeiro - RJ	0.750	0.421	0.750	0.539	0.680	0.897
Hospital Universitário Clementino Fraga Filho Rio de Janeiro - RJ	0.699	0.528	0.594	0.559	0.702	0.755
Hospital Santa Lúcia Brasília - DF	0.614	0.303	0.588	0.400	0.662	0.865
Hospital Santa Júlia Manaus - AM	0.816	0.367	0.917	0.524	0.698	0.978
Hospital Santa Catarina Blumenau - SC	0.960	0.167	1.000	0.286	0.643	1.000
Hospital São Francisco Mogiguaçu - SP	-	-	-	-	-	-
Hospital de Clínicas da UFPE Recife - PE	0.720	0.563	0.750	0.643	0.682	0.833
Hospital Escola da UFPel Pelotas - RS	-	-	-	-	-	-
Hospital de Urgências de Trindade Trindade - GO	-	-	-	-	-	-
Hospital Universitário Walter Cantídio Fortaleza - CE	0.427	0.182	0.333	0.235	0.438	0.636
Hospital Evangélico de Vila Velha Vila Velha - ES	0.519	0.375	0.750	0.500	0.615	0.889
Hospital Universitário Getúlio Vargas Manaus - AM	0.519	0.375	0.750	0.500	0.615	0.889

in healthcare, is desired for most real-world applications. Although not all tasks can be generalized beyond their sample locations, performing external tests (off-site validation) such as the one proposed in this study will allow an understanding of the limitations and possibilities of these models, advancing knowledge and facilitating the implementation of these algorithms in the real world.

The quality of the collected data has a significant influence on the predictive ability of models, and completeness and correct filling are essential requirements (109). However, prediction strength is equally important, hence the selection of variables that have good predictive ability is critical for algorithm development. In health, this especially includes the use of laboratory test results as reliable predictor variables (110).

Careful and critical analysis of issues such as data bias is essential before implementing algorithmic models, to ensure that human biases are not reproduced and perpetuated against vulnerable groups. Errors and recording failures are expected, especially in multicenter studies and during critical periods, such as a pandemic. To minimize these factors, we applied different techniques to deal with scenarios of data missing and possible errors in the database. Thus, it became possible to address these limitations and advance with greater reliability in data analysis and decision-making based on predictive models.

The study has limitations, the main being the size of the samples collected per hospital. Because it was a critical moment of the pandemic, data were collected over a short period, thus reducing our sample size. Another issue is the presence of data heterogeneity among hospitals. On the one hand, this may have decreased the ability of transfer learning algorithms to generalize to unequal settings, but on the other hand, it mimics the actual real-world challenges that arise from applying machine learning models in clinical practice.

This study aimed to analyze the predictive performance of transfer learning technique in a large and diverse country. Our findings showed that the use of transfer learning, although a promising approach to improve the performance of machine learning models in tabular data classification problems with limited datasets, may not increase predictive performance when the databases are not large enough for accurate model training.

Contributors

MF: Conceptualization, data curation, methodology, writing – review editing

HSS: Writing – review editing

ACF: Conceptualization, supervision, writing – review editing

Declaration of Interests

None.

Acknowledgments

This work was supported by National Council for Scientific and Technological Development (CNPq) under Grant Number 402626/2020-6, and Microsoft (Microsoft AI for Health COVID-19 Grant). We would like to thank the IACOV-BR Network, in alphabetic order: Ana Claudia Martins Ciconelle (Institute of Mathematics and Statistics, University of São Paulo); Ana Maria Espírito Santo de Brito (Instituto de Medicina, Estudos e Desenvolvimento—IMED, São Paulo, São Paulo); Bruno Pereira Nunes (Universidade Federal de Pelotas—UFPel); Dárcia Lima e Silva (Hospital Santa Lúcia); Fernando Anschau (Setor de Pesquisa da Gerência de Ensino e Pesquisa do Grupo Hospitalar Conceição, RS – Brasil; Programa de Pós-Graduação em Neurociências da Universidade Federal do Rio Grande do Sul); Henrique de Castro Rodrigues (Serviço de Epidemiologia e Avaliação/Direção Geral do HUCFF/UFRJ); Hermano Alexandre Lima Rocha (Unimed Fortaleza. Fortaleza, Ceará, Brasil; Departamento de Saúde Comunitária. Universidade Federal do Ceará. Fortaleza, Ceará, Brasil); João Conrado Bueno dos Reis (Hospital São Francisco); Liane de Oliveira Cavalcante (Hospital Santa Julia de Manaus); Liszt Palmeira de Oliveira (Instituto Unimed-Rio; Universidade do Estado do Rio de Janeiro); Lorena Sofia dos Santos Andrade (Universidade de Pernambuco—UPE/UEPB); Luiz Antonio Nasi (Hospital Moinhos de Vento); Marcelo de Maria Felix (InRad—Institute of Radiology, School of Medicine, University of São Paulo); Marcelo Jenne Mimica (Departamento de Ciências Patológicas Faculdade de Ciências Médicas da Santa Casa de São Paulo); Maria Elizete de Almeida Araujo (Federal University of Amazonas, University Hospital Getulio Vargas, Manaus, AM, Brazil); Mariana Volpe Arnoni (Serviço de Controle de Infecção Hospitalar Santa Casa de São Paulo); Rebeca Baiocchi Vianna (Hospital Santa Lúcia); Renan Magalhães Montenegro Junior (Complexo Hospitalar da Universidade Federal do Ceará – EBSERH); Renata Vicente da Penha (Hospital Evangélico de Vila Velha); Rogério Nadin Vicente (Hospital Santa Catarina de Blumenau); Ruchelli França de Lima (Hospital

Moinhos de Vento); Sandro Rodrigues Batista (Faculdade de Medicina, Universidade Federal de Goiás, Goiânia, Goiás; Secretaria de Estado da Saúde de Goiás, Goiânia, Goiás); Silvia Ferreira Nunes (Fundação Santa Casa de Misericórdia do Pará—FSCMP; Mestrado Profissional em Gestão e Saúde na Amazônia); Tássia Teles Santana de Macedo (Escola Bahiana de Medicina e Saúde Pública); Valesca Lôbo e Sant'ana Nuno (Hospital Português da Bahia). We would also like to thank all those people who somehow contributed to the progress of this research, in alphabetical order: Adriana Weinfeld Massaia; Alexandre Amaral; Ana Maria Pereira Rangel; Antônia Célia de Castro Alcantara; Bruna Donida; Bruno Mendes Carmon; Carisi Polanczyk; Carolina Zenilda Nicolao; Claiton Marques de Jesus; Denise Corrêa Nunes; Diana Almeida; Eduardo Menezes Lopes; Elias Bezerra Leite; Elimar Ponzzo Dutra Leal; Fernanda Arns de Castro; Fernanda Colares de Borba Netto; Flávia Araújo; Flávio Lúcio Pontes Ibiapina; Gerência de Ensino e pesquisa do Complexo Hospitalar da Universidade Federal do Ceará – EBSERH; Hospital Português da Bahia; Humberto Bolognini Tridapalli; Iasmin Luiza Leite; Laura Freitas de Faveri; Lena Claudia Maia Alencar; Luciane Kopittke; Luciano Hammes; Luiz Alberto Mattos; Marly Suzielly Miranda Silva; Mayara Rocha de Oliveira; Mohamed Parrini; Pablo Viana Stolz; Paloma Farina de Lima; Paulo Pitrez; Pollyana Bueno Siqueira; Rafaella Côrti Pessigatti; Raul José de Abreu Sturari Junior; Rodrigo Smania Garrastazu Almeida; Rogério Farias Bitencourt; Rubens Vasconcelos Barreto; Tatiane Lima Aguiar; Thyago Gregório Mota Ribeiro.

Funding

Coordination for the Improvement of Higher Education (CAPES) process 88887.492630/2020-00, National Council for Scientific and Technological Development (CNPq) under Grant Number 402626/2020-6, and Microsoft AI for Health COVID-19 Grant.

Data sharing statement

Supplemental materials

5 Considerações Finais e conclusão

5.1 Implicações do uso de ML em Políticas Públicas de Saúde

Os modelos de predição resultantes das técnicas de generalização têm o potencial de auxiliar em atividades de monitoramento, identificando os pacientes que necessitam de maior atenção. Esses poderão ser utilizados em diferentes locais e tempo, permitindo um aprimoramento no fluxo de serviços de atendimento e do acesso à saúde pública de qualidade mesmo em locais mais remotos do país.

O uso de modelos de ML em saúde necessita de atenção e rigor sobre a avaliação dos dados utilizados, para que não sejam proliferados preconceitos que possam estar escondidos contidos nos dados (111, 17). Por isso, é importante compreender como os dados foram criados e como o algoritmo foi treinado para reduzir o risco de vieses (7). Alguns princípios éticos como transparência, justiça e equidade, não maleficência, responsabilidade e privacidade, possuem um grau de consentimento global em inteligência artificial (112). Assim, nenhum modelo deve ser implementado em saúde sem testes que verifiquem a existência de vieses humanos e erros sistemáticos, o que será apenas possível por meio do monitoramento e da avaliação constante dos resultados obtidos.

Apesar da expansão do número de estudos que aplicam ML com o objetivo de melhorar a saúde pública, ainda há um longo caminho a percorrer até que essas técnicas possam ser disponibilizadas à população em geral. Digitalização de dados, sistema que suporte a implementação de modelos de ML, profissionais de saúde e de tecnologia da informação capacitados para identificar vieses e necessidade de recalibrar os modelos, são alguns dos desafios da área. Além disso, formas de avaliar o uso desses modelos também precisam ser aprimoradas para que a experiência do usuário e do paciente sejam otimizadas. Modelos de ML são ferramentas para contornar muitos dos desafios vivenciados atualmente, mas não são solucionadoras de problemas sociais extremamente complexos.

5.2 Conclusão

A análise dos resultados dos três artigos permite concluir que a generalização de modelos de ML em saúde ainda é uma área em desenvolvimento. O primeiro artigo, uma revisão sistemática da literatura, destacou a falta de um consenso em relação aos

termos utilizados e à importância dada à generalização em trabalhos sobre o tema na área da saúde. Além disso, muitos artigos que utilizam o conceito de generalização não o definem de forma clara e precisa. Já o segundo artigo, que apresentou análises de modelos de ML generalizados para diferentes regiões do Brasil, mostrou que é possível generalizar modelos com certos desfechos mesmo em contextos de grande desigualdade socioeconômica. Esses resultados apontam para a relevância de se desenvolver modelos de ML generalizados que possam ser aplicados em diferentes populações. Por fim, o terceiro artigo apresentou uma análise de transferência do aprendizado em modelos de ML entre hospitais das diferentes regiões brasileiras, concluindo que essa técnica ainda precisa ser mais aprimorada e explorada. Em conjunto, os resultados evidenciam a importância de se avançar no desenvolvimento de técnicas de generalização em modelos de ML em saúde, sobretudo em países que enfrentam grandes desafios em relação à desigualdade e ao acesso aos serviços de saúde.

Referências

1 RUSSELL, S. J. *Artificial intelligence a modern approach*. [S.l.]: Pearson Education, Inc., 2010. Citado na página 16.

2 ALPAYDIN, E. *Machine learning: the new AI*. [S.l.]: MIT press, 2016. Citado na página 16.

3 BEAM, A. L.; KOHANE, I. S. Big data and machine learning in health care. *Jama*, American Medical Association, v. 319, n. 13, p. 1317–1318, 2018. Citado na página 16.

4 JORDAN, M. I.; MITCHELL, T. M. Machine learning: Trends, perspectives, and prospects. *Science*, American Association for the Advancement of Science, v. 349, n. 6245, p. 255–260, 2015. Citado na página 16.

5 LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group UK London, v. 521, n. 7553, p. 436–444, 2015. Citado na página 16.

6 OBERMEYER, Z.; EMANUEL, E. J. Predicting the future—big data, machine learning, and clinical medicine. *The New England journal of medicine*, NIH Public Access, v. 375, n. 13, p. 1216, 2016. Citado 3 vezes nas páginas 16, 37 e 59.

7 SIDEY-GIBBONS, J. A.; SIDEY-GIBBONS, C. J. Machine learning in medicine: a practical introduction. *BMC medical research methodology*, Springer, v. 19, n. 1, p. 1–18, 2019. Citado 4 vezes nas páginas 16, 19, 59 e 69.

8 RAJKOMAR, A.; DEAN, J.; KOHANE, I. Machine learning in medicine. *New England Journal of Medicine*, Mass Medical Soc, v. 380, n. 14, p. 1347–1358, 2019. Citado 4 vezes nas páginas 16, 17, 37 e 59.

9 CHIAVEGATTO-FILHO, A. D. P.; SANTOS, H. G. D.; NASCIMENTO, C. F. do; MASSA, K.; KAWACHI, I. Overachieving municipalities in public health: a machine-learning approach. *Epidemiology*, LWW, v. 29, n. 6, p. 836–840, 2018. Citado na página 16.

10 TOPOL, E. J. High-performance medicine: the convergence of human and artificial intelligence. *Nature medicine*, Nature Publishing Group US New York, v. 25, n. 1, p. 44–56, 2019. Citado na página 16.

11 DONALDSON, M. S.; CORRIGAN, J. M.; KOHN, L. T. *et al.* To err is human: building a safer health system. National Academies Press, 2000. Citado na página 16.

12 MIOTTO, R.; LI, L.; KIDD, B. A.; DUDLEY, J. T. Deep patient: an unsupervised representation to predict the future of patients from the electronic health records. *Scientific reports*, Springer, v. 6, n. 1, p. 1–10, 2016. Citado na página 16.

13 NAYLOR, C. D. On the prospects for a (deep) learning health care system. *Jama*, American Medical Association, v. 320, n. 11, p. 1099–1100, 2018. Citado na página 16.

14 CARRILLO-LARCO, R. M.; CAR, L. T.; PEARSON-STUTTARD, J.; PANCH, T.; MIRANDA, J. J.; ATUN, R. Machine learning health-related applications in low-income and middle-income countries: a scoping review protocol. *BMJ open*, British Medical Journal Publishing Group, v. 10, n. 5, p. e035983, 2020. Citado na página 16.

15 MORSE, S. S.; MAZET, J. A.; WOOLHOUSE, M.; PARRISH, C. R.; CARROLL, D.; KARESH, W. B.; ZAMBRANA-TORRELIO, C.; LIPKIN, W. I.; DASZAK, P. Prediction and prevention of the next pandemic zoonosis. *The Lancet*, Elsevier, v. 380, n. 9857, p. 1956–1965, 2012. Citado na página 17.

16 HASTIE, T.; TIBSHIRANI, R.; FRIEDMAN, J. H.; FRIEDMAN, J. H. *The elements of statistical learning: data mining, inference, and prediction*. [S.l.]: Springer, 2009. v. 2. Citado 2 vezes nas páginas 17 e 37.

17 LEE, N. T. Detecting racial bias in algorithms and machine learning. *Journal of Information, Communication and Ethics in Society*, Emerald Publishing Limited, v. 16, n. 3, p. 252–260, 2018. Citado 2 vezes nas páginas 18 e 69.

18 BARTNECK, C.; LÜTGE, C.; WAGNER, A.; WELSH, S. *An introduction to ethics in robotics and AI*. [S.l.]: Springer Nature, 2021. Citado na página 18.

19 GÉRON, A. *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow*. [S.l.]: "O'Reilly Media, Inc.", 2022. Citado na página 19.

20 YU, F.; CROSO, G. S.; KIM, T. S.; SONG, Z.; PARKER, F.; HAGER, G. D.; REITER, A.; VEDULA, S. S.; ALI, H.; SIKDER, S. Assessment of automated identification of phases in videos of cataract surgery using machine learning and deep learning techniques. *JAMA network open*, American Medical Association, v. 2, n. 4, p. e191860–e191860, 2019. Citado na página 19.

21 LAGHI, A. Cautions about radiologic diagnosis of covid-19 infection driven by artificial intelligence. *The Lancet Digital Health*, Elsevier, v. 2, n. 5, p. e225, 2020. Citado na página 19.

22 YU, K.-H.; LEE, T.-L. M.; YEN, M.-H.; KOU, S.; ROSEN, B.; CHIANG, J.-H.; KOHANE, I. S. *et al.* Reproducible machine learning methods for lung cancer detection using computed tomography images: Algorithm development and validation. *Journal of medical Internet research*, JMIR Publications Inc., Toronto, Canada, v. 22, n. 8, p. e16709, 2020. Citado na página 19.

23 XU, Y.; BAHADORI, M. T.; SEARLES, E.; THOMPSON, M.; JAVIER, T.-S.; SUN, J. Predicting changes in pediatric medical complexity using large longitudinal health records. In: AMERICAN MEDICAL INFORMATICS ASSOCIATION. *AMIA Annual Symposium Proceedings*. [S.l.], 2017. v. 2017, p. 1838. Citado 2 vezes nas páginas 19 e 59.

24 LIU, D.; SHIN, W.-Y.; SPRECHER, E.; CONROY, K.; SANTIAGO, O.; WACHTEL, G.; SANTILLANA, M. Machine learning approaches to predicting no-shows in pediatric medical appointment. *NPJ digital medicine*, Nature Publishing Group UK London, v. 5, n. 1, p. 50, 2022. Citado 2 vezes nas páginas 19 e 59.

25 GOLAS, S. B.; SHIBAHARA, T.; AGBOOLA, S.; OTAKI, H.; SATO, J.; NAKAE, T.; HISAMITSU, T.; KOJIMA, G.; FELSTED, J.; KAKARMATH, S. *et al.* A machine learning model to predict the risk of 30-day readmissions in patients with heart failure: a retrospective analysis of electronic medical records data. *BMC medical informatics and decision making*, BioMed Central, v. 18, n. 1, p. 1–17, 2018. Citado 2 vezes nas páginas 19 e 59.

26 RAVAUT, M.; SADEGHI, H.; LEUNG, K. K.; VOLKOVS, M.; KORNAS, K.; HARISH, V.; WATSON, T.; LEWIS, G. F.; WEISMAN, A.; POUTANEN, T. *et al.* Predicting adverse outcomes due to diabetes complications with machine learning using administrative health data. *NPJ digital medicine*, Nature Publishing Group, v. 4, n. 1, p. 1–12, 2021. Citado 2 vezes nas páginas 19 e 59.

27 GOTO, T.; CAMARGO, C. A.; FARIDI, M. K.; FREISHTAT, R. J.; HASEGAWA, K. Machine learning–based prediction of clinical outcomes for children during emergency department triage. *JAMA network open*, American Medical Association, v. 2, n. 1, p. e186937–e186937, 2019. Citado 2 vezes nas páginas 19 e 59.

28 KO, S.; CHOI, J.; AHN, J. Gves: machine learning model for identification of prognostic genes with a small dataset. *Scientific Reports*, Nature Publishing Group, v. 11, n. 1, p. 1–8, 2021. Citado 2 vezes nas páginas 19 e 59.

29 ZHONG, Z.; YUAN, X.; LIU, S.; YANG, Y.; LIU, F. Machine learning prediction models for prognosis of critically ill patients after open-heart surgery. *Scientific Reports*, Nature Publishing Group, v. 11, n. 1, p. 1–10, 2021. Citado 2 vezes nas páginas 19 e 59.

30 ROSE, S. Mortality risk score prediction in an elderly population using machine learning. *American journal of epidemiology*, Oxford University Press, v. 177, n. 5, p. 443–452, 2013. Citado 2 vezes nas páginas 19 e 59.

31 PARIKH, R. B.; MANZ, C.; CHIVERS, C.; REGLI, S. H.; BRAUN, J.; DRAUGELIS, M. E.; SCHUCHTER, L. M.; SHULMAN, L. N.; NAVATHE, A. S.; PATEL, M. S. *et al.* Machine learning approaches to predict 6-month mortality among patients with cancer. *JAMA network open*, American Medical Association, v. 2, n. 10, p. e1915997–e1915997, 2019. Citado 2 vezes nas páginas 19 e 59.

32 SANTOS, H. G. d.; NASCIMENTO, C. F. d.; IZBICKI, R.; DUARTE, Y. A. d. O.; FILHO, P. C.; DIAS, A. Machine learning para análises preditivas em saúde: exemplo de aplicação para predizer óbito em idosos de são paulo, brasil. *Cadernos de Saúde Pública*, SciELO Public Health, v. 35, p. e00050818, 2019. Citado 2 vezes nas páginas 19 e 59.

33 MOLL, M.; QIAO, D.; REGAN, E. A.; HUNNINGHAKE, G. M.; MAKE, B. J.; TAL-SINGER, R.; MCGEACHIE, M. J.; CASTALDI, P. J.; ESTEPAR, R. S. J.; WASHKO, G. R. *et al.* Machine learning and prediction of all-cause mortality in copd. *Chest*, Elsevier, v. 158, n. 3, p. 952–964, 2020. Citado 2 vezes nas páginas 19 e 59.

34 ZHANG, A.; LIPTON, Z. C.; LI, M.; SMOLA, A. J. Dive into deep learning. *arXiv preprint arXiv:2106.11342*, 2021. Citado na página 19.

35 GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016. Citado na página 20.

36 GUO, J.; LI, B. The application of medical artificial intelligence technology in rural areas of developing countries. *Health equity*, Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, v. 2, n. 1, p. 174–181, 2018. Citado na página 20.

37 ZHUANG, F.; QI, Z.; DUAN, K.; XI, D.; ZHU, Y.; ZHU, H.; XIONG, H.; HE, Q. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, IEEE, v. 109, n. 1, p. 43–76, 2020. Citado 2 vezes nas páginas 21 e 59.

38 TORREY, L.; SHAVLIK, J. Transfer learning. In: *Handbook of research on machine learning applications and trends: algorithms, methods, and techniques*. [S.l.]: IGI global, 2010. p. 242–264. Citado 2 vezes nas páginas 21 e 59.

39 PAN, S. J.; YANG, Q. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, IEEE, v. 22, n. 10, p. 1345–1359, 2010. Citado 2 vezes nas páginas 21 e 59.

40 Citado 2 vezes nas páginas 21 e 59.

41 KABOLI, M. A review of transfer learning algorithms. Technische Universität München, 2017. Citado 2 vezes nas páginas 21 e 59.

42 TAN, C.; SUN, F.; KONG, T.; ZHANG, W.; YANG, C.; LIU, C. A survey on deep transfer learning. In: SPRINGER. *Artificial Neural Networks and Machine*

Learning–ICANN 2018: 27th International Conference on Artificial Neural Networks, Rhodes, Greece, October 4-7, 2018, Proceedings, Part III 27. [S.l.], 2018. p. 270–279. Citado 2 vezes nas páginas 21 e 59.

43 YANG, Q.; ZHANG, Y.; WENYUAN, D.; PAN, S. J. *Transfer Learning*. [S.l.]: Cambridge University Press, 2020. Citado na página 22.

44 LEVIN, R.; CHEREPANOVA, V.; SCHWARZSCHILD, A.; BANSAL, A.; BRUSS, C. B.; GOLDSTEIN, T.; WILSON, A. G.; GOLDBLUM, M. Transfer learning with deep tabular models. *arXiv preprint arXiv:2206.15306*, 2022. Citado na página 22.

45 FAWCETT, T. An introduction to roc analysis. *Pattern recognition letters*, Elsevier, v. 27, n. 8, p. 861–874, 2006. Citado na página 33.

46 WIENS, J.; SARIA, S.; SENDAK, M.; GHASSEMI, M.; LIU, V. X.; DOSHI-VELEZ, F.; JUNG, K.; HELLER, K.; KALE, D.; SAEED, M. *et al.* Do no harm: a roadmap for responsible machine learning for health care. *Nature medicine*, Nature Publishing Group US New York, v. 25, n. 9, p. 1337–1340, 2019. Citado na página 37.

47 AZAD, T. D.; EHRESMAN, J.; AHMED, A. K.; STAARTJES, V. E.; LUBELSKI, D.; STIENEN, M. N.; VEERAVAGU, A.; RATLIFF, J. K. Fostering reproducibility and generalizability in machine learning for clinical prediction modeling in spine surgery. *The Spine Journal*, Elsevier, v. 21, n. 10, p. 1610–1616, 2021. Citado na página 37.

48 YANG, J.; SOLTAN, A. A.; CLIFTON, D. A. Machine learning generalizability across healthcare settings: insights from multi-site covid-19 screening. *npj Digital Medicine*, Nature Publishing Group UK London, v. 5, n. 1, p. 69, 2022. Citado 2 vezes nas páginas 37 e 62.

49 ROTH, J. A.; BATTEGAY, M.; JUCHLER, F.; VOGT, J. E.; WIDMER, A. F. Introduction to machine learning in digital healthcare epidemiology. *Infection Control & Hospital Epidemiology*, Cambridge University Press, v. 39, n. 12, p. 1457–1462, 2018. Citado na página 37.

50 HILL, B. L.; BROWN, R.; GABEL, E.; RAKOCZ, N.; LEE, C.; CANNESSON, M.; BALDI, P.; LOOHUIS, L. O.; JOHNSON, R.; JEW, B. *et al.* An automated machine learning-based model predicts postoperative mortality using readily-extractable preoperative electronic health record data. *British journal of anaesthesia*, Elsevier, v. 123, n. 6, p. 877–886, 2019. Citado na página 41.

51 XIE, F.; CHAKRABORTY, B.; ONG, M. E. H.; GOLDSTEIN, B. A.; LIU, N. *et al.* Autoscore: a machine learning–based automatic clinical score generator and its application to mortality prediction using electronic health records. *JMIR medical informatics*, JMIR Publications Inc., Toronto, Canada, v. 8, n. 10, p. e21798, 2020. Citado na página 41.

52 CHI, S.; GUO, A.; HEARD, K.; KIM, S.; FORAKER, R.; WHITE, P.; MOORE, N. Development and structure of an accurate machine learning algorithm to predict inpatient mortality and hospice outcomes in the coronavirus disease 2019 era. *Medical care*, Wolters Kluwer Health, v. 60, n. 5, p. 381, 2022. Citado na página 41.

53 PETERSON, D. J.; OSTBERG, N. P.; BLAYNEY, D. W.; BROOKS, J. D.; HERNANDEZ-BOUSSARD, T. Machine learning applied to electronic health records: identification of chemotherapy patients at high risk for preventable emergency department visits and hospital admissions. *JCO Clinical Cancer Informatics*, Wolters Kluwer Health, v. 5, p. 1106–1126, 2021. Citado na página 41.

54 JORGE, A. M.; SMITH, D.; WU, Z.; CHOWDHURY, T.; COSTENBADER, K.; ZHANG, Y.; CHOI, H. K.; FELDMAN, C. H.; ZHAO, Y. Exploration of machine learning methods to predict systemic lupus erythematosus hospitalizations. *Lupus*, SAGE Publications Sage UK: London, England, v. 31, n. 11, p. 1296–1305, 2022. Citado na página 41.

55 EDGCOMB, J. B.; SHADDOX, T.; HELLEMANN, G.; III, J. O. B. Predicting suicidal behavior and self-harm after general hospitalization of adults with serious mental illness. *Journal of psychiatric research*, Elsevier, v. 136, p. 515–521, 2021. Citado na página 41.

56 BARAK-CORREN, Y.; CASTRO, V. M.; NOCK, M. K.; MANDL, K. D.; MADSEN, E. M.; SEIGER, A.; ADAMS, W. G.; APPLEGATE, R. J.; BERNSTAM, E. V.; KLANN, J. G. *et al.* Validation of an electronic health record–based suicide risk prediction modeling approach across multiple health care systems. *JAMA network open*, American Medical Association, v. 3, n. 3, p. e201262–e201262, 2020. Citado 2 vezes nas páginas 41 e 43.

57 WILIMITIS, D.; TURER, R. W.; RIPPERGER, M.; MCCOY, A. B.; SPERRY, S. H.; FIELSTEIN, E. M.; KURZ, T.; WALSH, C. G. Integration of face-to-face screening with real-time machine learning to predict risk of suicide among adults. *JAMA network open*, American Medical Association, v. 5, n. 5, p. e2212095–e2212095, 2022. Citado na página 41.

58 BONDE, A.; VARADARAJAN, K. M.; BONDE, N.; TROELSEN, A.; MURATOGLU, O. K.; MALCHAU, H.; YANG, A. D.; ALAM, H.; SILLESEN, M. Assessing the utility of deep neural networks in predicting postoperative surgical complications: a retrospective study. *The Lancet Digital Health*, Elsevier, v. 3, n. 8, p. e471–e485, 2021. Citado na página 41.

59 TOMAŠEV, N.; HARRIS, N.; BAUR, S.; MOTTRAM, A.; GLOROT, X.; RAE, J. W.; ZIELINSKI, M.; ASKHAM, H.; SARAIVA, A.; MAGLIULO, V. *et al.* Use of deep learning to develop continuous-risk models for adverse event prediction from electronic health records. *Nature Protocols*, Nature Publishing Group UK London, v. 16, n. 6, p. 2765–2787, 2021. Citado 2 vezes nas páginas 41 e 43.

60 SANZ, H.; REVERTER, F.; VALIM, C. Enhancing svm for survival data using local invariances and weighting. *BMC bioinformatics*, Springer, v. 21, p. 1–20, 2020. Citado na página 41.

61 KUMAR, M.; ANG, L. T.; HO, C.; SOH, S. E.; TAN, K. H.; CHAN, J. K. Y.; GODFREY, K. M.; CHAN, S.-Y.; CHONG, Y. S.; ERIKSSON, J. G. *et al.* Machine learning–derived prenatal predictive risk model to guide intervention and prevent the progression of gestational diabetes mellitus to type 2 diabetes: Prediction model development study. *JMIR diabetes*, JMIR Publications Toronto, Canada, v. 7, n. 3, p. e32366, 2022. Citado na página 41.

62 YUAN, K.-C.; TSAI, L.-W.; LEE, K.-H.; CHENG, Y.-W.; HSU, S.-C.; LO, Y.-S.; CHEN, R.-J. The development an artificial intelligence algorithm for early sepsis diagnosis in the intensive care unit. *International journal of medical informatics*, Elsevier, v. 141, p. 104176, 2020. Citado na página 41.

63 JEFFERIES, J. L.; SPENCER, A. K.; LAU, H. A.; NELSON, M. W.; GIULIANO, J. D.; ZABINSKI, J. W.; BOUSSIOS, C.; CURHAN, G.; GLIKLICH, R. E.; WARNOCK, D. G. A new approach to identifying patients with elevated risk for fabry disease using a

machine learning algorithm. *Orphanet Journal of Rare Diseases*, Springer, v. 16, p. 1–8, 2021. Citado na página 41.

64 WANG, S.; LI, M.; NG, S. B. Research on infant health diagnosis and intelligence development based on machine learning and health information statistics. *Frontiers in Public Health*, Frontiers Media SA, v. 10, 2022. Citado na página 41.

65 TRINHAMMER, M.; MERRILD, A. H.; LOTZ, J.; MAKRANSKY, G. Predicting crime during or after psychiatric care: Evaluating machine learning for risk assessment using the danish patient registries. *Journal of psychiatric research*, Elsevier, v. 152, p. 194–200, 2022. Citado na página 41.

66 ROTHMAN, K. J.; GREENLAND, S.; LASH, T. L. *et al. Modern epidemiology*. [S.l.]: Wolters Kluwer Health/Lippincott Williams & Wilkins Philadelphia, 2008. v. 3. Citado na página 47.

67 FRANÇA, E.; LANSKY, S. Mortalidade infantil neonatal no brasil: situação, tendências e perspectivas. *Anais*, p. 1–29, 2016. Citado na página 47.

68 UNICEF-DATA. 2020. Disponível em: ⟨https://data.unicef.org/topic/child-survival/neonatal-mortality/.html#fn48⟩. Citado na página 47.

69 SHARROW, D.; HUG, L.; LIU, Y.; YOU, D. Levels and trends in child mortality. *New York: United Nations Inter-agency Group for Child Mortality Estimation*. Disponível em: ⟨https://www.unicef.org/reports/levels-and-trends-child-mortality-report-2020/.html#fn48⟩. Citado na página 47.

70 BATISTA, A. F.; DINIZ, C. S.; BONILHA, E. A.; KAWACHI, I.; FILHO, A. D. C. Neonatal mortality prediction with routinely collected data: a machine learning approach. *BMC pediatrics*, BioMed Central, v. 21, n. 1, p. 1–6, 2021. Citado na página 47.

71 LI, Z.; KARLSSON, O.; KIM, R.; SUBRAMANIAN, S. Distribution of under-5 deaths in the neonatal, postneonatal, and childhood periods: a multicountry analysis in 64 low-and middle-income countries. *International journal for equity in health*, Springer, v. 20, n. 1, p. 1–11, 2021. Citado na página 47.

72 YOU, D.; HUG, L.; EJDEMYR, S.; IDELE, P.; HOGAN, D.; MATHERS, C.; GERLAND, P.; NEW, J. R.; ALKEMA, L. *et al.* Global, regional, and national levels and trends in under-5 mortality between 1990 and 2015, with scenario-based projections to 2030: a systematic analysis by the un inter-agency group for child mortality estimation. *The Lancet*, Elsevier, v. 386, n. 10010, p. 2275–2286, 2015. Citado na página 47.

73 SVS - SECRETARIA DE VIGILÂNCIA EM SAÚDE. 2021. Disponível em: ⟨http://svs.aids.gov.br/dantps/centrais-de-conteudos/paineis-de-monitoramento/mortalidade/infantil-e-fetal/.html#fn48⟩. Citado na página 47.

74 PGIODS IBGE. Disponível em: ⟨https://pgiods.ibge.gov.br/index.html?mapid=187/.html#fn48⟩. Citado na página 47.

75 HUG, L.; ALEXANDER, M.; YOU, D.; ALKEMA, L.; CHILD, U. I.-a. G. for. National, regional, and global levels and trends in neonatal mortality between 1990 and 2017, with scenario-based projections to 2030: a systematic analysis. *The Lancet Global Health*, Elsevier, v. 7, n. 6, p. e710–e720, 2019. Citado na página 47.

76 DOURADO, I.; MEDINA, M. G.; AQUINO, R. The effect of the family health strategy on usual source of care in brazil: data from the 2013 national health survey (pns 2013). *International journal for equity in health*, BioMed Central, v. 15, n. 1, p. 1–10, 2016. Citado na página 47.

77 LEAL, M. d. C.; SZWARCWALD, C. L.; ALMEIDA, P. V. B.; AQUINO, E. M. L.; BARRETO, M. L.; BARROS, F.; VICTORA, C. Reproductive, maternal, neonatal and child health in the 30 years since the creation of the unified health system (sus). *Ciencia & saude coletiva*, SciELO Public Health, v. 23, p. 1915–1928, 2018. Citado na página 47.

78 SILVA, E. S. d. A. d.; PAES, N. A. Bolsa família programme and the reduction of child mortality in the municipalities of the brazilian semiarid region. *Ciência & Saúde Coletiva*, SciELO Brasil, v. 24, p. 623–630, 2019. Citado na página 47.

79 LIMA, S. S. d.; BRAGA, M. C.; VANDERLEI, L. C. d. M.; LUNA, C. F.; FRIAS, P. G. Avaliação do impacto de programas de assistência pré-natal, parto e ao recém-nascido nas mortes neonatais evitáveis em pernambuco, brasil: estudo de adequação. *Cadernos de Saúde Pública*, SciELO Brasil, v. 36, 2020. Citado na página 47.

80 CHAO, F.; YOU, D.; PEDERSEN, J.; HUG, L.; ALKEMA, L. National and regional under-5 mortality rate by economic status for low-income and middle-income countries: a systematic assessment. *The Lancet Global Health*, Elsevier, v. 6, n. 5, p. e535–e547, 2018. Citado na página 47.

81 GUINSBURG, R.; SANUDO, A.; KIFFER, C. R. V.; MARINONIO, A. S. S.; COSTA-NOBRE, D. T.; ARECO, K. N.; KAWAKAMI, M. D.; MIYOSHI, M. H.; BANDIERA-PAIVA, P.; BALDA, R. d. C. X. *et al.* Annual trend of neonatal mortality and its underlying causes: population-based study–são paulo state, brazil, 2004–2013. *BMC pediatrics*, BioMed Central, v. 21, n. 1, p. 1–9, 2021. Citado na página 47.

82 COLLINS, G. S.; MOONS, K. G. Reporting of artificial intelligence prediction models. *The Lancet*, Elsevier, v. 393, n. 10181, p. 1577–1579, 2019. Citado na página 47.

83 FAN, H.; LI, L.; GILBERT, R.; O'CALLAGHAN, F.; WIJLAARS, L. A machine learning approach to identify cases of cerebral palsy using the uk primary care database. *The Lancet*, Elsevier, v. 392, p. S33, 2018. Citado na página 47.

84 MOREIRA, M. W.; RODRIGUES, J. J.; CARVALHO, F. H.; CHILAMKURTI, N.; AL-MUHTADI, J.; DENISOV, V. Biomedical data analytics in mobile-health environments for high-risk pregnancy outcome prediction. *Journal of Ambient Intelligence and Humanized Computing*, Springer, v. 10, n. 10, p. 4121–4134, 2019. Citado na página 47.

85 MOREIRA, M. W.; RODRIGUES, J. J.; MARCONDES, G. A.; NETO, A. J. V.; FURTADO, V. Predicting neonatal condition at birth through ensemble learning methods in pregnancy care. In: SBC. *Anais do XVIII Simpósio Brasileiro de Computação Aplicada à Saúde*. [S.l.], 2018. Citado na página 47.

86 PODDA, M.; BACCIU, D.; MICHELI, A.; BELLÙ, R.; PLACIDI, G.; GAGLIARDI, L. A machine learning approach to estimating preterm infants survival: development of the preterm infants survival assessment (pisa) predictor. *Scientific reports*, Nature Publishing Group, v. 8, n. 1, p. 1–9, 2018. Citado na página 47.

87 SINGHA, A. K.; PHUKAN, D.; BHASIN, S.; SANTHANAM, R. Application of machine learning in analysis of infant mortality and its factors. *Work Pap*, p. 1–5, 2016. Citado na página 47.

88 HOUWELING, T. A.; KLAVEREN, D. van; DAS, S.; AZAD, K.; TRIPATHY, P.; MANANDHAR, D.; NEUMAN, M.; JONGE, E. de; BEEN, J. V.; STEYERBERG, E. *et al.* A prediction model for neonatal mortality in low-and middle-income countries: an analysis of data from population surveillance sites in india, nepal and bangladesh.

International journal of epidemiology, Oxford University Press, v. 48, n. 1, p. 186–198, 2019. Citado na página 47.

89 NAIMI, A. I.; PLATT, R. W.; LARKIN, J. C. Machine learning for fetal growth prediction. *Epidemiology (Cambridge, Mass.)*, NIH Public Access, v. 29, n. 2, p. 290, 2018. Citado na página 47.

90 KUHLE, S.; MAGUIRE, B.; ZHANG, H.; HAMILTON, D.; ALLEN, A. C.; JOSEPH, K.; ALLEN, V. M. Comparison of logistic regression with machine learning methods for the prediction of fetal growth abnormalities: a retrospective cohort study. *BMC pregnancy and childbirth*, BioMed Central, v. 18, n. 1, p. 1–9, 2018. Citado na página 47.

91 MOREIRA, M. W.; RODRIGUES, J. J.; FURTADO, V.; MAVROMOUSTAKIS, C. X.; KUMAR, N.; WOUNGANG, I. Fetal birth weight estimation in high-risk pregnancies through machine learning techniques. In: IEEE. *ICC 2019-2019 IEEE International Conference on Communications (ICC)*. [S.l.], 2019. p. 1–6. Citado na página 47.

92 DEY, A.; HAY, K.; AFROZ, B.; CHANDURKAR, D.; SINGH, K.; DEHINGIA, N.; RAJ, A.; SILVERMAN, J. G. Understanding intersections of social determinants of maternal healthcare utilization in uttar pradesh, india. *PLoS One*, Public Library of Science San Francisco, CA USA, v. 13, n. 10, p. e0204810, 2018. Citado na página 47.

93 AVALOS, L. A.; FLANAGAN, T.; LI, D.-K. Preventing perinatal depression to improve maternal and child health—a health care imperative. *JAMA pediatrics*, American Medical Association, v. 173, n. 4, p. 313–314, 2019. Citado na página 47.

94 DINIZ, C. S. G.; BONILHA, E. de A.; ALY, C. M. C.; FIORETTI-FOSCHI, B.; NIY, D. Y.; BATISTA, A. F. de M. *Idade gestacional em dias*. Harvard Dataverse, 2022. Disponível em: ⟨https://doi.org/10.7910/DVN/PP2VVF⟩. Citado na página 48.

95 WHO, W. H. O. Making every baby count: audit and review of stillbirths and neonatal deaths. World Health Organization, 2016. Citado na página 49.

96 MBOYA, I. B.; MAHANDE, M. J.; MOHAMMED, M.; OBURE, J.; MWAMBI, H. G. Prediction of perinatal death using machine learning models: a birth registry-based cohort study in northern tanzania. *BMJ open*, British Medical Journal Publishing Group, v. 10, n. 10, p. e040132, 2020. Citado na página 54.

97 SHUKLA, V. V.; EGGLESTON, B.; AMBALAVANAN, N.; MCCLURE, E. M.; MWENECHANYA, M.; CHOMBA, E.; BOSE, C.; BAUSERMAN, M.; TSHEFU, A.; GOUDAR, S. S. *et al.* Predictive modeling for perinatal mortality in resource-limited settings. *JAMA network open*, American Medical Association, v. 3, n. 11, p. e2026750–e2026750, 2020. Citado na página 54.

98 ROCHA, R.; ATUN, R.; MASSUDA, A.; RACHE, B.; SPINOLA, P.; NUNES, L.; LAGO, M.; CASTRO, M. C. Effect of socioeconomic inequalities and vulnerabilities on health-system preparedness and response to covid-19 in brazil: a comprehensive analysis. *The Lancet Global Health*, Elsevier, v. 9, n. 6, p. e782–e792, 2021. Citado 2 vezes nas páginas 54 e 59.

99 PAIM, J.; TRAVASSOS, C.; ALMEIDA, C.; BAHIA, L.; MACINKO, J. The brazilian health system: history, advances, and challenges. *The Lancet*, Elsevier, v. 377, n. 9779, p. 1778–1797, 2011. Citado na página 59.

100 MASSUDA, A.; HONE, T.; LELES, F. A. G.; CASTRO, M. C. D.; ATUN, R. The brazilian health system at crossroads: progress, crisis and resilience. *BMJ global health*, BMJ Specialist Journals, v. 3, n. 4, p. e000829, 2018. Citado na página 59.

101 CASTRO, M. C.; CARVALHO, L. R. de; CHIN, T.; KAHN, R.; FRANÇA, G. V.; MACÁRIO, E. M.; OLIVEIRA, W. K. de. Demand for hospitalization services for covid-19 patients in brazil. *MedRxiv*, Cold Spring Harbor Laboratory Press, p. 2020–03, 2020. Citado na página 59.

102 BIGONI, A.; MALIK, A. M.; TASCA, R.; CARRERA, M. B. M.; SCHIESARI, L. M. C.; GAMBARDELLA, D. D.; MASSUDA, A. Brazil's health system functionality amidst of the covid-19 pandemic: An analysis of resilience. *The Lancet Regional Health-Americas*, Elsevier, v. 10, p. 100222, 2022. Citado na página 59.

103 BOCCIA, S.; RICCIARDI, W.; IOANNIDIS, J. P. What other countries can learn from italy during the covid-19 pandemic. *JAMA internal medicine*, American Medical Association, v. 180, n. 7, p. 927–928, 2020. Citado na página 59.

104 BLUMENTHAL, D.; FOWLER, E. J.; ABRAMS, M.; COLLINS, S. R. *Covid-19—implications for the health care system*. [S.l.]: Mass Medical Soc, 2020. 1483–1488 p. Citado na página 59.

105 WALKER, P. G.; WHITTAKER, C.; WATSON, O. J.; BAGUELIN, M.; WINSKILL, P.; HAMLET, A.; DJAFAARA, B. A.; CUCUNUBÁ, Z.; MESA, D. O.; GREEN, W. *et al.* The impact of covid-19 and strategies for mitigation and suppression in low-and middle-income countries. *Science*, American Association for the Advancement of Science, v. 369, n. 6502, p. 413–422, 2020. Citado na página 59.

106 BORISOV, V.; LEEMANN, T.; SESSLER, K.; HAUG, J.; PAWELCZYK, M.; KASNECI, G. Deep neural networks and tabular data: A survey. *IEEE Transactions on Neural Networks and Learning Systems*, IEEE, 2022. Citado na página 59.

107 WICHMANN, R. M.; FERNANDES, F. T.; FILHO, A. D. P. C. Improving the performance of machine learning algorithms for health outcomes predictions in multicentric cohorts. *Scientific Reports*, Nature Publishing Group UK London, v. 13, n. 1, p. 1022, 2023. Citado 2 vezes nas páginas 60 e 61.

108 GORISHNIY, Y.; RUBACHEV, I.; KHRULKOV, V.; BABENKO, A. Revisiting deep learning models for tabular data. *Advances in Neural Information Processing Systems*, v. 34, p. 18932–18943, 2021. Citado na página 61.

109 HÄYRINEN, K.; SARANTO, K.; NYKÄNEN, P. Definition, structure, content, use and impacts of electronic health records: a review of the research literature. *International journal of medical informatics*, Elsevier, v. 77, n. 5, p. 291–304, 2008. Citado na página 66.

110 JIANG, X.; COFFEE, M.; BARI, A.; WANG, J.; JIANG, X.; HUANG, J.; SHI, J.; DAI, J.; CAI, J.; ZHANG, T. *et al.* Towards an artificial intelligence framework for data-driven prediction of coronavirus clinical severity. *Computers, Materials & Continua*, Computers, Materials and Continua (Tech Science Press), v. 63, n. 1, p. 537–551, 2020. Citado na página 66.

111 HAIDER, A. H.; CHANG, D. C.; EFRON, D. T.; HAUT, E. R.; CRANDALL, M.; CORNWELL, E. E. Race and insurance status as risk factors for trauma mortality. *Archives of Surgery*, American Medical Association, v. 143, n. 10, p. 945–949, 2008. Citado na página 69.

112 JOBIN, A.; IENCA, M.; VAYENA, E. The global landscape of ai ethics guidelines. *Nature Machine Intelligence*, Nature Publishing Group, v. 1, n. 9, p. 389–399, 2019. Citado na página 69.

Apêndice A – Suplemento Artigo 2

Table 4 – Distribution of characteristics of train set

Characteristics	**N=807.932**	**Missing**
Mother's age (years)	28.0 (6.66)	<0.1%
Mother's education		<0.1%
0 - No education	983 (<1%)	
1 - Fundamental	17,122 (2%)	
2 - Fundamental 2	130,104 (16%)	
3 - High school	409,648 (51%)	
4 - Incomplete superior	47,686 (6%)	
5 - Complete superior	198,822 (25%)	
Color		¡0.1%
1 - White	433,244 (54%)	
2 - Black	50,821 (6%)	
3 - Yellow	10,247 (1%)	
4 - Brown	306,778 (38%)	
5 - Indigenous	3,827 (<1%)	
Mode of delivery		<0.1%
1 - Vaginal	341,769 (42%)	
2 - Cesarean	466,084 (58%)	
Place of delivery		<0.1%
1 - Hospital	802,454 (>99%)	
2 - Other health facility	1,941 (<1%)	
3 - Residence	2,954 (<1%)	
4 - Others	558 (<1%)	
Time of gestation (weeks)	39 (2.11)	<0.1%
Birth weight (grams)	3,180 (551)	<0.1%
Congenital Anomaly		<0.1%
1 - Yes	14,054 (2%)	
2 - No	793,298 (98%)	
1st Apgar	8.55 (3.23)	<0.1%
5th Apgar	9.58 (2.93)	<0.1%

Note: Median (Standard Deviation)

Table 5 – Predictive performance for neonatal mortality on the test set for vulnerable subgroups with XGBoost model for the capitals of Brazil states, 2016 (Teenage mothers).

Region	**State**	**Capital**	**Neonatal mortality rate**	**AUC**	**Precision**	**Recall**	**F1**	**AUPRC**	**Brier**
North	Acre	Rio Branco	8.61	0.880	0.667	0.200	0.308	0.314	0.007
	Amapá	Macapá	14.50	0.921	0.800	0.381	0.516	0.605	0.010
	Amazonas	Manaus	9.34	0.948	0.667	0.390	0.492	0.479	0.008
	Pará	Belém	19.11	0.945	0.788	0.472	0.591	0.645	0.012
	Rondônia	Porto Velho	19.40	0.756	0.700	0.350	0.467	0.664	0.016
	Roraima	Boa Vista	14.44	0.966	0.818	0.563	0.667	0.692	0.008
	Tocantins	Palmas	26.41	0.967	0.833	0.667	0.741	0.731	0.012
Notheast	Alagoas	Maceió	13.31	0.966	0.760	0.487	0.593	0.647	0.009
	Bahia	Salvador	17.48	0.986	0.756	0.654	0.701	0.776	0.010
	Ceará	Fortaleza	12.89	0.965	0.654	0.735	0.692	0.744	0.008
	Maranhão	São Luís	22.67	0.972	0.698	0.667	0.682	0.738	0.014
	Paraíba	João Pessoa	8.30	0.939	0.692	0.615	0.640	0.642	0.006
	Pernambuco	Recife	17.39	0.956	0.692	0.652	0.671	0.667	0.011
	Piauí	Teresina	25.13	0.982	0.853	0.617	0.716	0.791	0.012
	Rio Grande do Norte	Natal	13.22	0.919	0.478	0.458	0.468	0.588	0.014
	Sergipe	Aracaju	20.62	0.956	0.840	0.525	0.646	0.723	0.012
Midwest	Distrito Federal	Brasília	11.06	0.920	0.544	0.794	0.646	0.707	0.010
	Goiás	Goiânia	25.62	0.955	0.721	0.607	0.660	0.653	0.016
	Mato Grosso	Cuiabá	20.37	0.961	0.833	0.454	0.588	0.689	0.013
	Mato Grosso do Sul	Campo Grande	10.62	0.943	0.384	0.455	0.417	0.465	0.014
Southest	Espírito Santo	Vitória	8.00	0.975	0.500	0.600	0.545	0.583	0.008
	Minas Gerais	Belo Horizonte	11.35	0.923	0.567	0.607	0.586	0.584	0.010
	Rio de Janeiro	Rio de Janeiro	11.12	0.959	0.737	0.600	0.661	0.647	0.007
	São Paulo	São Paulo	13.72	0.957	0.729	0.609	0.664	0.709	0.008
South	Paraná	Curitiba	10.06	0.945	0.611	0.688	0.647	0.587	0.008
	Rio Grande do Sul	Porto Alegre	11.99	0.995	0.571	0.842	0.681	0.654	0.009
	Santa Catarina	Florianópolis	5.12	1.000	0.667	1.000	0.800	1.000	0.003

Note: AUC = area under the receiver operating characteristic curve; Recall = sensibility; Precision = positive predictive value; F1 = F1 score; AUPRC = area under prediction-recall curve; Brier = Brier Score.

Table 6 – Predictive performance for neonatal mortality on the test set for vulnerable subgroups with XGBoost model for the capitals of Brazil states, 2016 (Low education mothers).

Region	**State**	**Capital**	**Neonatal mortality rate**	**AUC**	**Precision**	**Recall**	**F1**	**AUPRC**	**Brier**
North	Acre	Rio Branco	7.88	0.917	0.565	0.448	0.500	0.425	0.007
	Amapá	Macapá	11.14	0.920	0.760	0.396	0.521	0.565	0.008
	Amazonas	Manaus	9.80	0.932	0.810	0.379	0.516	0.574	0.007
	Pará	Belém	16.01	0.932	0.754	0.369	0.495	0.543	0.012
	Rondônia	Porto Velho	19.45	0.963	0.771	0.409	0.534	0.686	0.014
	Roraima	Boa Vista	10.53	0.916	0.941	0.593	0.727	0.678	0.005
	Tocantins	Palmas	14.28	0.951	0.818	0.409	0.545	0.664	0.010
Notheast	Alagoas	Maceió	11.55	0.941	0.681	0.454	0.544	0.556	0.009
	Bahia	Salvador	19.58	0.971	0.706	0.599	0.648	0.709	0.013
	Ceará	Fortaleza	13.11	0.953	0.620	0.696	0.656	0.691	0.009
	Maranhão	São Luís	15.92	0.957	0.561	0.582	0.571	0.646	0.014
	Paraíba	João Pessoa	11.40	0.955	0.679	0.545	0.605	0.639	0.008
	Pernambuco	Recife	18.04	0.960	0.700	0.491	0.578	0.642	0.013
	Piauí	Teresina	21.50	0.969	0.800	0.541	0.646	0.733	0.013
	Rio Grande do Norte	Natal	13.35	0.953	0.662	0.586	0.622	0.656	0.010
	Sergipe	Aracaju	17.82	0.938	0.688	0.458	0.550	0.574	0.013
Midwest	Distrito Federal	Brasília	9.21	0.913	0.482	0.643	0.551	0.570	0.010
	Goiás	Goiânia	19.83	0.941	0.672	0.554	0.608	0.677	0.014
	Mato Grosso	Cuiabá	12.99	0.944	0.680	0.515	0.586	0.602	0.009
	Mato Grosso do Sul	Campo Grande	11.51	0.930	0.568	0.467	0.512	0.564	0.010
Southest	Espírito Santo	Vitória	6.39	0.949	0.600	0.462	0.522	0.568	0.005
	Minas Gerais	Belo Horizonte	9.40	0.953	0.531	0.607	0.567	0.580	0.008
	Rio de Janeiro	Rio de Janeiro	10.38	0.942	0.647	0.565	0.603	0.584	0.008
	São Paulo	São Paulo	9.98	0.959	0.664	0.560	0.608	0.651	0.007
South	Paraná	Curitiba	7.50	0.952	0.614	0.540	0.574	0.531	0.006
	Rio Grande do Sul	Porto Alegre	11.15	0.955	0.509	0.602	0.552	0.528	0.011
	Santa Catarina	Florianópolis	7.57	0.990	0.783	0.750	0.766	0.823	0.005

Note: AUC = area under the receiver operating characteristic curve; Recall = sensibility; Precision = positive predictive value; F1 = F1 score; AUPRC = area under prediction-recall curve; Brier = Brier Score.

Table 7 – Predictive performance for neonatal mortality on the test set for vulnerable subgroups with XGBoost model for the capitals of Brazil states, 2016 (Non-white mothers).

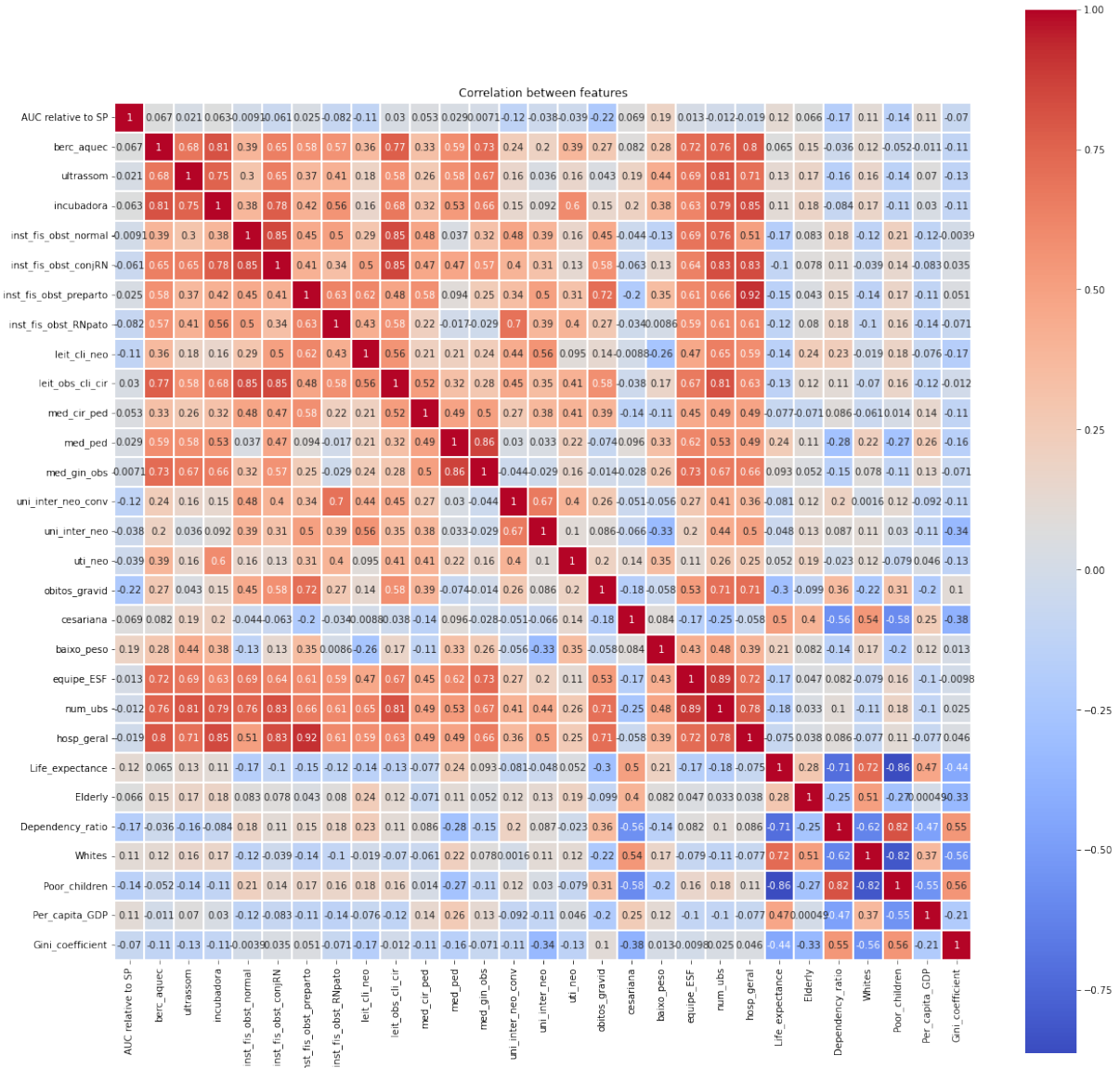
Region	**State**	**Capital**	**Neonatal mortality rate**	**AUC**	**Precision**	**Recall**	**F1**	**AUPRC**	**Brier**
North	Acre	Rio Branco	2.52	0.928	0.571	0.431	0.491	0.457	0.007
	Amapá	Macapá	11.50	0.944	0.762	0.436	0.555	0.601	0.007
	Amazonas	Manaus	8.45	0.940	0.802	0.491	0.609	0.642	0.005
	Pará	Belém	14.98	0.954	0.742	0.446	0.558	0.603	0.011
	Rondônia	Porto Velho	16.42	0.960	0.802	0.468	0.591	0.696	0.011
	Roraima	Boa Vista	8.66	0.955	0.824	0.575	0.677	0.670	0.005
	Tocantins	Palmas	12.78	0.962	0.827	0.573	0.667	0.735	0.007
Notheast	Alagoas	Maceió	10.65	0.942	0.671	0.495	0.570	0.564	0.008
	Bahia	Salvador	14.58	0.977	0.710	0.654	0.681	0.734	0.009
	Ceará	Fortaleza	11.35	0.959	0.657	0.636	0.647	0.666	0.008
	Maranhão	São Luís	14.28	0.965	0.674	0.636	0.655	0.661	0.010
	Paraíba	João Pessoa	11.03	0.965	0.667	0.503	0.573	0.623	0.008
	Pernambuco	Recife	15.36	0.964	0.712	0.541	0.615	0.658	0.010
	Piauí	Teresina	15.86	0.968	0.811	0.529	0.640	0.703	0.009
	Rio Grande do Norte	Natal	11.26	0.940	0.641	0.558	0.597	0.588	0.008
	Sergipe	Aracaju	16.41	0.949	0.742	0.496	0.594	0.638	0.011
Midwest	Distrito Federal	Brasília	8.20	0.939	0.504	0.675	0.577	0.609	0.008
	Goiás	Goiânia	14.79	0.950	0.656	0.545	0.595	0.606	0.011
	Mato Grosso	Cuiabá	11.35	0.972	0.760	0.547	0.636	0.673	0.007
	Mato Grosso do Sul	Campo Grande	9.67	0.941	0.588	0.484	0.531	0.551	0.008
Southest	Espírito Santo	Vitória	4.10	0.920	0.682	0.429	0.526	0.519	0.003
	Minas Gerais	Belo Horizonte	8.76	0.950	0.566	0.592	0.579	0.607	0.008
	Rio de Janeiro	Rio de Janeiro	9.59	0.956	0.672	0.576	0.620	0.617	0.007
	São Paulo	São Paulo	8.37	0.957	0.647	0.560	0.600	0.611	0.006
South	Paraná	Curitiba	7.85	0.951	0.655	0.559	0.603	0.629	0.006
	Rio Grande do Sul	Porto Alegre	10.43	0.948	0.523	0.648	0.579	0.553	0.010
	Santa Catarina	Florianópolis	4.92	0.992	0.800	0.727	0.762	0.797	0.004

Note: AUC = area under the receiver operating characteristic curve; Recall = sensibility; Precision = positive predictive value; F1 = F1 score; AUPRC = area under prediction-recall curve; Brier = Brier Score.

Table 8 – Variables of correlation analysis

Variable	**Description**
$\text{berc}_a quec$	proportion of heated cribs per 1,000 live births in 2016
ultrassom	proportion of ultrasound machines in use per 1,000 live births in 2016
incubadora	proportion of incubator machines in use per 1,000 live births in 2016
$\text{inst}_f is_o bst_n ormal$	proportion of physical facilities for obstetrics and neonatology hospital environment Newborn Normal in use per 1,000 births in 2016
$\text{inst}_f is_o bst_c onjRN$	proportion of physical obstetrics and neonatology facilities hospital environment NB set in use per 1,000 live births in 2016
$\text{inst}_f is_o bst_p reparto$	proportion of physical obstetrics and neonatology facilities in hospital environment Antepartum in use per 1,000 live births in 2016
$\text{inst}_f is_o bst_R npato$	proportion of physical facilities for obstetrics and neonatology hospital environment Pathological RN in use per 1,000 live births in 2016
$\text{leit}_c li_n eo$	proportion of neonatal clinical beds per 1,000 live births in 2016
$\text{leit}_o bs_c li_c ir$	proportion of obstetric surgical beds per 1,000 live births in 2016
$\text{med}_c ir_p ed$	proportion of pediatric surgeons per 1,000 live births in 2016
$\text{med}_p ed$	proportion of pediatricians per 1,000 live births in 2016
$\text{med}_g in_o bs$	proportion of gynecologists and obstetricians per 1,000 live births in 2016
$\text{uni}_i nter_n eo_c onv$	proportion of conventional neonatal intermediate unit per 1,000 live births in 2016
$\text{uni}_i nter_n eo$	proportion of neonatal intermediate unit per 1,000 live births in 2016
$\text{uti}_n eo$	proportion of neonatal ICUs per 1,000 live births in 2016
$\text{obitos}_g ravid$	proportion of maternal deaths per 1,000 live births in 2016
cesariana	proportion of caesarean sections per 1,000 live births in 2016
$\text{baixo}_p eso$	proportion of children with low birth weight per 1,000 live births in 2016
$\text{equipe}_E SF$	proportion of ESF teams per 1,000 live births in 2016
$\text{num}_u bs$	proportion of UBS per 1,000 live births in 2016
$\text{hosp}_g eral$	proportion of general hospitals per 1,000 live births in 2016
$\text{Life}_e xpectance$	life expectancy at birth, 2010 Census
Elderly	percentage of elderly people in the resident population by municipality, 2010 Census
$\text{Dependence}_r atio$	dependency ratio according to municipality, 2010 Census
Whites	percentage of whites in the resident population by municipality, 2010 Census
$\text{Poor}_c hildren$	percentage of children in low-income households, 2010 Census
$\text{Per}_c apita_G DP$	GDP per capita, 2010 Census
$\text{Gini}_c oef ficient$	Gini index of per capita household income by municipality, 2010 Census

Figure 7 – Correlation analysis.



Apêndice B – Suplemento Artigo 3

Table 9 – Distribution of demographic characteristics comparing data from all 18 hospitals and training data for TL

Characteristics	**Death**		**Total**
	No	**Yes**	
	Mean (SD)	*Mean (SD)*	*Mean (SD)*
Age (years) - All	55.2 (17.0)	66.7 (15.1)	58.4 (17.3)
Age (years) - HC hospital	55.5 (15.9)	66.2 (13.6)	59.3 (16.0)
Hospital time - All	13.2 (17.3)	16.4 (16.5)	14.2 (17.1)
Hospital time - HC hospital	19.1 (17.2)	17.0 (12.6)	18.3 (15.7)
Male (%) - All	53.3	60	55.1
Male (%) - HC hospital	52.3	62.0	55.8
Race - white (%) - All	68.3	50.6	62.1
Race - white (%) - HC hospital	61.2	60.5	60.9
Race - black/mixed/Asian (%) - All	31.7	49.5	38
Race - black/mixed/Asian (%) – HC hospital	38.8	39.5	39.1

Table 10 – Set of optimized hyperparameters.

% outcome	N	Place	Algorithm	Hyperparameters
0.42	1776	Hospital Santa Casa São Paulo - SP	XGBoost	(colsample_bytree=0.85, gamma=0.55, learning_rate=0.035, min_child_weight=3.0, n_estimators=150, n_jobs=-2, random_state=42, reg_alpha=1.5, scale_pos_weight=3, subsample=0.25)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate: 0.06051021240691517, l2_leaf_reg: 4.0, border_count: 32, silent: True,max_depth: 6, n_estimators: 150, random_state: 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42,solver='sgd')
0.36	1500	Hospital de Clínicas da USP São Paulo - SP	XGBoost	(colsample_bytree=0.85, gamma=0.55, learning_rate=0.035, min_child_weight=3.0, n_estimators=150, n_jobs=-2,random_state=42, reg_alpha=1.5, scale_pos_weight=3,subsample=0.25)
			LightGBM	(class_weight='balanced', colsample_bytree=0.6165041429483651, n_estimators=10, num_leaves=112, random_state=42, reg_alpha=0.9547804466890908, reg_lambda=0.26832475371053754, scale_pos_weight=1)
			Catboost	(learning_rate= 0.06051021240691517, l2_leaf_reg= 4.0, border_count= 32, silent= True, max_depth= 6, n_estimators= 150, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42, solver='sgd')
0.07	1359	Hospital Português da Bahia Salvador - BA	XGBoost	(colsample_bytree=0.85, gamma=0.55, learning_rate=0.035, min_child_weight=3.0, n_estimators=150, n_jobs=-2, random_state=42, reg_alpha=1.5, scale_pos_weight=3, subsample=0.25)
			LightGBM	(colsample_bytree=0.9622926542536769, n_estimators=671, num_leaves=91, random_state=42, reg_alpha=0.9739479466464522, reg_lambda=0.017096689947744492, scale_pos_weight=3)
			Catboost	(learning_rate= 0.046093902470756815, l2_leaf_reg= 7.0, border_count= 20, silent= True, max_depth= 4, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=10, hidden_layer_sizes=6, learning_rate_init=0.1, max_iter=100, random_state=42)
0.13	845	Hospital UNIMED Fortaleza - CE	XGBoost	(colsample_bytree=0.35, gamma=0.7000000000000001, learning_rate=0.015, max_depth=6, min_child_weight=3.0, n_jobs=-2, random_state=42, reg_alpha=1.25, scale_pos_weight=3, subsample=0.7)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate= 0.046093902470756815, l2_leaf_reg= 7.0, border_count= 20, silent= True, max_depth= 4, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=10, hidden_layer_sizes=6, learning_rate_init=0.1, max_iter=100, random_state=42)
0.39	539	Hospital Regional de Luiziania Luiziania - GO	XGBoost	(colsample_bytree=0.3, gamma=0.75, learning_rate=0.025, max_depth=7, min_child_weight=2.0, n_estimators=200, n_jobs=-2, random_state=42, reg_alpha=1.25, subsample=0.25)
			LightGBM	(class_weight='balanced', colsample_bytree=0.9987948839600422, n_estimators=935, num_leaves=135, random_state=42, reg_alpha=0.869355916285306, reg_lambda=0.08174694195215582, scale_pos_weight=9)
			Catboost	(learning_rate= 0.030217575111747233, l2_leaf_reg= 3.0, border_count= 10, silent= True, max_depth= 3, n_estimators-= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42, solver='sgd')
0.1	456	Hospital Moinhos de Vento Porto Alegre - RS	XGBoost	(colsample_bytree=0.8, gamma=0.8, learning_rate=0.245, max_depth=5, min_child_weight=2.0, n_estimators=75, n_jobs=-2, random_state=42, reg_alpha=1.25, subsample=0.4)
			LightGBM	(class_weight='balanced', colsample_bytree=0.7868594263341625, n_estimators=803, num_leaves=111, random_state=42, reg_alpha=0.8820998106981124, reg_lambda=0.08885184806583746, scale_pos_weight=1)
			Catboost	(learning_rate= 0.4167439198625154, l2_leaf_reg= 8.0, border_count= 20, silent= True, max_depth= 3, n_estimators= 125, random_state= 42)
			MLP	(alpha=0.01, batch_size=25, hidden_layer_sizes=11, learning_rate_init=0.1, max_iter=600, random_state=42, solver='sgd')

% outcome	N	Place	Algorithm	Hyperparameters
0.24	449	Hospital UNIMED Rio de Janeiro - RJ	XGBoost	(colsample_bytree=0.35, gamma=0.7000000000000001, learning_rate=0.015, max_depth=6, min_child_weight=3.0, n_jobs=-2,random_state=42, reg_alpha=1.25, scale_pos_weight=3, subsample=0.7)
			LightGBM	(colsample_bytree=0.9622926542536769, n_estimators=671, num_leaves=91, random_state=42, reg_alpha=0.9739479466464522, reg_lambda=0.017096689947744492, scale_pos_weight=3)
			Catboost	(learning_rate= 0.030217575111747233, l2_leaf_reg= 3.0, border_count= 10, silent= True, max_depth= 3, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=1e-05, batch_size=10, hidden_layer_sizes=11, learning_rate_init=0.03, max_iter=600, random_state=42)
0.36	296	Hospital Universitário Clementino Fraga Filho	XGBoost	(colsample_bytree=0.3, gamma=1.0, learning_rate=0.125, max_depth=2, min_child_weight=3.0, n_estimators=250, n_jobs=-2, random_state=42, reg_alpha=1.0, scale_pos_weight=3, subsample=0.9)
			LightGBM	(colsample_bytree=0.9622926542536769, n_estimators=671, num_leaves=91, random_state=42, reg_alpha=0.9739479466464522, reg_lambda=0.017096689947744492, scale_pos_weight=3)
			Catboost	(learning_rate= 0.4264289139310363, l2_leaf_reg= 2.0, border_count= 200, silent= True, max_depth= 5, n_estimators= 225, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=10, hidden_layer_sizes=6, learning_rate_init=0.1, max_iter=100, random_state=42)
0.2	281	Rio de Janeiro - RJ Hospital Santa Lúcia Brasília - DF	XGBoost	(colsample_bytree=0.75, gamma=0.6000000000000001, learning_rate=0.205, max_depth=2, min_child_weight=5.0, n_estimators=175, n_jobs=-2, random_state=42, reg_alpha=1.0, subsample=0.7)
			LightGBM	(class_weight='balanced', colsample_bytree=0.6165041429483651, n_estimators=10, num_leaves=112, random_state=42, reg_alpha=0.9547804466890908, reg_lambda=0.26832475371053754, scale_pos_weight=1)
			Catboost	(learning_rate= 0.3138583088615541, l2_leaf_reg= 9.0, border_count= 10, silent= True, max_depth= 12, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42, solver='sgd')
0.15	247	Hospital Santa Júlia Manaus - AM	XGBoost	(colsample_bytree=0.4, gamma=0.8500000000000001, learning_rate=0.195, max_depth=2, min_child_weight=3.0, n_estimators=125, n_jobs=-2, random_state=42, reg_alpha=1.25, subsample=0.95)
			LightGBM	(class_weight='balanced', colsample_bytree=0.6165041429483651, n_estimators=10, num_leaves=112, random_state=42, reg_alpha=0.9547804466890908, reg_lambda=0.26832475371053754, scale_pos_weight=1)
			Catboost	(learning_rate= 0.030217575111747233, l2_leaf_reg= 3.0, border_count= 10, silent= True, max_depth= 3, n_estimators= 50, random_state= 42)
			MLP	(alpha=0.01, batch_size=15, hidden_layer_sizes=8, learning_rate_init=0.1, max_iter=500, random_state=42, solver='sgd')
0.07	148	Hospital Santa Catarina Blumenau - SC	XGBoost	(colsample_bytree=0.4, gamma=0.7000000000000001, learning_rate=0.045, min_child_weight=3.0, n_estimators=25, n_jobs=-2, random_state=42, reg_alpha=0.5, scale_pos_weight=3, subsample=0.35)
			LightGBM	(class_weight='balanced', colsample_bytree=0.9987948839600422, n_estimators=935, num_leaves=135, random_state=42, reg_alpha=0.869355916285306, reg_lambda=0.08174694195215582, scale_pos_weight=9)
			Catboost	(learning_rate= 0.408900899040920696, l2_leaf_reg= 8.0, border_count= 50, silent= True, max_depth= 13, n_estimators= 175, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42, solver='sgd')
0.14	124	Hospital São Francisco Mogiguaçu - SP	XGBoost	(colsample_bytree=0.9, gamma=0.8, learning_rate=0.145, max_depth=4, min_child_weight=5.0, n_estimators=175, n_jobs=-2, random_state=42, reg_alpha=0.75, scale_pos_weight=3, subsample=0.75)
			LightGBM	(class_weight='balanced', colsample_bytree=0.7922228674844964, n_estimators=1200, num_leaves=87, random_state=42, reg_alpha=0.7041340569061305, reg_lambda=0.4349979536508063, scale_pos_weight=3)
			Catboost	(learning_rate= 0.046093902470756815, l2_leaf_reg= 7.0, border_count= 20, silent= True, max_depth= 4, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.01, batch_size=5, hidden_layer_sizes=7, learning_rate_init=0.1, max_iter=100, random_state=42, solver='lbfgs')

% outcome	N	Place	Algorithm	Hyperparameters
0.37	112	Hospital de Clínicas da UFPE Recife - PE	XGBoost	(colsample_bytree=0.2, gamma=0.8500000000000001, learning_rate=0.055, max_depth=6, min_child_weight=3.0, n_estimators=50, n_jobs=-2, random_state=42, reg_alpha=1.25, subsample=0.5)
			LightGBM	(colsample_bytree=0.885217808605771, n_estimators=142, num_leaves=39, random_state=42, reg_alpha=0.2298357960090034, reg_lambda=0.9759998977389169, scale_pos_weight=5)
			Catboost	(learning_rate= 0.3079946686059291, l2_leaf_reg= 2.0, border_count= 100, silent= True, max_depth= 12, n_estimators= 225, random_state= 42)
			MLP	(alpha=0.1, batch_size=15, hidden_layer_sizes=9, learning_rate_init=0.01, max_iter=100, random_state=42)
0.37	91	Hospital Escola da UFPel Pelotas - RS	XGBoost	(colsample_bytree=0.35, gamma=0.7000000000000001, learning_rate=0.015, max_depth=6, min_child_weight=3.0, n_jobs=-2, random_state=42, reg_alpha=1.25, scale_pos_weight=3, subsample=0.7)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8432996648684546, n_estimators=1067, num_leaves=148, random_state=42, reg_alpha=0.059936189081400526, reg_lambda=0.8609304819134347, scale_pos_weight=3)
			Catboost	(learning_rate= 0.057355876381734484, l2_leaf_reg= 4.0, border_count= 10, silent= True, max_depth= 9, n_estimators= 100, random_state= 42)
			MLP	(activation='tanh', alpha=1e-06, batch_size=20, hidden_layer_sizes=10, learning_rate_init=0.01, max_iter=800, random_state=42, solver='sgd')
0.46	78	Hospital de Urgências de Trindade Trindade - GO	XGBoost	(colsample_bytree=0.9, gamma=0.8, learning_rate=0.145, max_depth=4, min_child_weight=5.0, n_estimators=175, n_jobs=-2, random_state=42, reg_alpha=0.75, scale_pos_weight=3, subsample=0.75)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate= 0.38797660062196027, l2_leaf_reg= 1.0, border_count= 5, silent= True, max_depth= 11, n_estimators= 50, random_state= 42)
			MLP	(alpha=1e-05, batch_size=20, hidden_layer_sizes=8, learning_rate_init=0.03, random_state=42, solver='sgd')
0.29	73	Hospital Universitário Walter Cantídio Fortaleza - CE	XGBoost	(colsample_bytree=0.6, gamma=0.55, learning_rate=0.295, max_depth=7, min_child_weight=1.0, n_estimators=275, n_jobs=-2, random_state=42, reg_alpha=1.0, subsample=0.85)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate= 0.38797660062196027, l2_leaf_reg= 1.0, border_count= 5, silent= True, max_depth= 11, n_estimators= 50, random_state= 42)
			MLP	(alpha=0.01, batch_size=10, hidden_layer_sizes=9, learning_rate_init=0.01, max_iter=700, random_state=42, solver='lbfgs')
0.23	56	Hospital Evangélico de Vila Velha Vila Velha - ES	XGBoost	(colsample_bytree=0.3, gamma=1.0, learning_rate=0.125, max_depth=2, min_child_weight=3.0, n_estimators=250, n_jobs=-2, random_state=42, reg_alpha=1.0, scale_pos_weight=3, subsample=0.9)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate= 0.11398638732673426, l2_leaf_reg= 6.0, border_count= 5, silent= True, max_depth= 7, n_estimators= 175, random_state= 42)
			MLP	(alpha=0.01, batch_size=25, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=800, random_state=42, solver='sgd')
0.34	47	Hospital Universitário Getúlio Vargas Manaus - AM	XGBoost	(colsample_bytree=0.6, gamma=0.55, learning_rate=0.295, max_depth=7, min_child_weight=1.0, n_estimators=275, n_jobs=-2, random_state=42, reg_alpha=1.0, subsample=0.85)
			LightGBM	(class_weight='balanced', colsample_bytree=0.8480475990706564, n_estimators=10, num_leaves=81, random_state=42, reg_alpha=0.6517423997488939, reg_lambda=0.8839266254726155, scale_pos_weight=9)
			Catboost	(learning_rate= 0.3138583088615541, l2_leaf_reg= 9.0, border_count= 10, silent= True, max_depth= 12, n_estimators= 50, random_state= 42)
			MLP	(activation='tanh', alpha=0.1, batch_size=20, hidden_layer_sizes=5, learning_rate_init=0.03, max_iter=500, random_state=42, solver='sgd')

Table 11 – Local test result of optimized models with all variables.

N	Place	Algorithm	AUROC (CI95%)	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
1776	Hospital Santa Casa São Paulo - SP	XGBoost	0.722 [0.68-0.77]	0.517	0.815	0.633	0.640	0.462	0.778	-0.008
		LightGBM	0.691 [0.65-0.74]	0.467	0.919	0.619	0.580	0.256	0.816	-0.166
		Catboost	0.726 [0.68-0.77]	0.643	0.570	0.604	0.640	0.776	0.718	0.141
		MLP	0.701 [0.66-0.74]	0.576	0.516	0.544	0.610	0.731	0.681	0.088
1500	Hospital de Clínicas da USP São Paulo - SP	XGBoost	0.804 [0.76-0.85]	0.565	0.814	0.667	0.710	0.652	0.863	0.146
		LightGBM	0.775 [0.73-0.82]	0.626	0.677	0.651	0.660	0.776	0.812	0.131
		Catboost	0.805 [0.76-0.85]	0.724	0.522	0.607	0.710	0.890	0.77	0.265
		MLP	0.809 [0.77-0.85]	0.708	0.528	0.605	0.710	0.879	0.770	0.272
1359	Hospital Português da Bahia Salvador - BA	XGBoost	0.865 [0.79-0.94]	0.31	0.321	0.316	0.380	0.947	0.95	0.030
		LightGBM	0.871 [0.79-0.95]	0.417	0.357	0.385	0.460	0.963	0.953	0.080
		Catboost	0.871 [0.80-0.94]	0.667	0.143	0.235	0.370	0.995	0.940	0.187
		MLP	0.872 [0.00-1.00]	1.000	0.000	0.000	0.000	1.000	0.000	0.079
845	Hospital UNIMED Fortaleza - CE	XGBoost	0.946 [0.91-0.98]	0.667	0.727	0.696	0.730	0.946	0.959	0.220
		LightGBM	0.848 [0.77-0.92]	0.284	0.758	0.413	0.470	0.715	0.952	-0.644
		Catboost	0.933 [0.88-0.98]	0.667	0.242	0.356	0.650	0.982	0.897	0.387
		MLP	0.911 [0.00-1.00]	1.000	0.000	0.000	0.000	1.000	0.000	0.241
539	Hospital Regional de Luiziania Luiziania - GO	XGBoost	0.719 [0.64-0.80]	0.702	0.879	0.780	0.790	0.413	0.684	0.133
		LightGBM	0.706 [0.62-0.79]	0.680	0.859	0.759	0.790	0.365	0.622	-0.080
		Catboost	0.752 [0.68-0.83]	0.672	0.889	0.765	0.830	0.318	0.645	0.135
		MLP	0.737 [0.66-0.81]	0.743	0.758	0.750	0.830	0.587	0.607	0.097
456	Hospital Moinhos de Vento Porto Alegre - RS	XGBoost	0.919 [0.87-0.97]	0.667	0.286	0.400	0.550	0.984	0.924	0.276
		LightGBM	0.916 [0.86-0.97]	0.538	0.500	0.518	0.490	0.951	0.944	0.244
		Catboost	0.857 [0.79-0.93]	0.375	0.214	0.273	0.350	0.959	0.915	-0.040
		MLP	0.904[0.79-0.97]	0.667	0.429	0.522	0.480	0.976	0.938	0.137

N	Place	Algorithm	AUROC (CI95%)	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
449	Hospital UNIMED Rio de Janeiro - RJ	XGBoost	0.804 [0.73-0.88]	0.517	0.469	0.492	0.560	0.864	0.840	0.066
		LightGBM	0.782 [0.70-0.86]	0.552	0.500	0.525	0.480	0.874	0.849	0.066
		Catboost	0.793 [0.00-1.00]	1.000	0.125	0.222	0.000	1.000	0.000	0.174
		MLP	0.700 [0.60-0.80]	0.444	0.375	0.407	0.470	0.854	0.815	-0.263
296	Hospital Universitário Clementino Fraga Filho Rio de Janeiro - RJ	XGBoost	0.771 [0.66-0.88]	0.548	0.719	0.622	0.720	0.667	0.809	0.043
		LightGBM	0.776 [0.67-0.88]	0.600	0.750	0.667	0.720	0.719	0.837	0.059
		Catboost	0.787 [0.69-0.89]	0.625	0.625	0.625	0.720	0.789	0.789	0.029
		MLP	0.739 [0.63-0.85]	0.667	0.438	0.528	0.640	0.877	0.735	0.140
281	Hospital Santa Lúcia Brasília - DF	XGBoost	0.825 [0.73-0.92]	0.636	0.412	0.500	0.590	0.941	0.865	0.208
		LightGBM	0.777 [0.66-0.89]	0.400	0.588	0.476	0.530	0.779	0.883	-0.100
		Catboost	0.753 [0.63-0.88]	0.625	0.294	0.400	0.470	0.956	0.844	0.105
		MLP	0.754 [0.64-0.87]	0.312	0.294	0.303	0.450	0.838	0.826	-0.242
247	Hospital Santa Júlia Manaus - AM	XGBoost	0.814 [0.68-0.95]	0.500	0.083	0.143	0.520	0.984	0.849	0.202
		LightGBM	0.785 [0.64-0.93]	0.375	0.500	0.429	0.400	0.841	0.898	-0.256
		Catboost	0.853 [0.00-1.00]	1.000	0.083	0.154	0.000	1.000	0.000	0.189
		MLP	0.788 [0.66-0.95]	0.625	0.417	0.500	0.620	0.952	0.896	0.214
148	Hospital Santa Catarina Blumenau - SC	XGBoost	0.929 [0.00-1.00]	1.000	0.000	0.000	0.000	1.000	0.000	-0.492
		LightGBM	0.976 [0.00-1.00]	0.600	1.000	0.750	0.000	0.952	0.000	0.055
		Catboost	0.952 [0.00-1.00]	1.000	0.333	0.500	0.000	1.000	0.000	0.332
		MLP	0.897 [0.00-1.00]	0.000	0.000	0.000	0.000	0.952	0.000	-0.328
124	Hospital São Francisco Mogiguaçu - SP	XGBoost	0.897 [0.76-1.00]	0.375	0.600	0.462	0.700	0.849	0.933	-0.008
		LightGBM	0.903 [0.79-1.00]	0.333	0.800	0.471	0.540	0.758	0.962	-0.447
		Catboost	0.952 [0.00-1.00]	1.000	0.200	0.333	0.000	1.000	0.000	0.351
		MLP	0.733 [0.55-0.90]	0.182	0.400	0.250	0.260	0.727	0.889	-1.647

N	Place	Algorithm	AUROC (CI95%)	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
112	Hospital de Clínicas da UFPE Recife - PE	XGBoost	0.803 [0.00-1.00]	1.000	0.167	0.286	0.000	1.000	0.000	0.114
		LightGBM	0.758 [0.58-0.94]	0.526	0.833	0.645	0.700	0.591	0.867	-0.043
		Catboost	0.883 [0.75-1.00]	0.778	0.583	0.667	0.860	0.909	0.800	0.446
		MLP	0.716 [0.53-0.90]	0.500	0.500	0.500	0.690	0.727	0.727	-0.157
91	Hospital Escola da UFPel Pelotas - RS	XGBoost	0.806 [0.63-0.98]	0.615	0.800	0.696	0.760	0.722	0.867	0.099
		LightGBM	0.700 [0.49-0.91]	0.714	0.500	0.588	0.590	0.889	0.762	-0.022
		Catboost	0.772 [0.57-0.97]	0.714	0.500	0.588	0.720	0.889	0.762	0.079
		MLP	0.633 [0.38-0.89]	0.750	0.300	0.429	0.610	0.944	0.708	-0.051
78	Hospital de Urgências de Trindade Trindade - GO	XGBoost	0.706 [0.48-0.93]	0.600	0.923	0.727	0.710	0.273	0.750	-0.072
		LightGBM	0.636 [0.00-1.00]	0.542	1.000	0.703	0.000	0.000	0.000	-0.145
		Catboost	0.741 [0.54-0.95]	0.667	0.615	0.640	0.820	0.636	0.583	0.132
		MLP	0.650 [0.41-0.87]	0.615	0.615	0.615	0.660	0.545	0.545	-0.636
73	Hospital Universitário Walter Cantídio Fortaleza - CE	XGBoost	0.510 [0.22-0.81]	0.500	0.167	0.250	0.420	0.938	0.750	-0.122
		LightGBM	0.703 [0.00-1.00]	0.273	1.000	0.429	0.000	0.000	0.000	-0.957
		Catboost	0.469 [0.12-0.88]	0.500	0.167	0.250	0.420	0.938	0.750	-0.099
		MLP	0.562 [0.24-0.82]	0.286	0.333	0.308	0.320	0.688	0.733	-0.669
56	Hospital Evangélico de Vila Velha Vila Velha - ES	XGBoost	0.673 [0.26-1.00]	0.333	0.250	0.286	0.560	0.846	0.786	0.170
		LightGBM	0.500 [0.00-1.00]	0.235	1.000	0.381	0.000	0.000	0.000	-0.389
		Catboost	0.654 [0.25-1.00]	0.500	0.250	0.333	0.570	0.923	0.800	0.058
		MLP	0.423 [0.01-0.84]	0.333	0.250	0.286	0.420	0.846	0.786	-0.471
47	Hospital Universitário Getúlio Vargas Manaus - AM	XGBoost	0.900 [0.00-1.00]	1.000	0.600	0.750	0.000	1.000	0.000	0.333
		LightGBM	0.500 [0.00-1.00]	0.333	1.000	0.500	0.000	0.000	0.000	-0.125
		Catboost	0.860 [0.66-1.00]	0.667	0.400	0.500	0.790	0.900	0.750	0.357
		MLP	0.640 [0.34-0.96]	0.333	0.200	0.250	0.460	0.800	0.667	0.000

Table 12 – Local test result of optimized models with boruta variables.

N	Place	Algorithm	AUROC	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
1776	Hospital Santa Casa São Paulo - SP	XGBoost	0.726	0.527	0.805	0.637	0.640	0,487	0.779	0.003
		LightGBM	0.681	0.563	0.584	0.573	0.590	0.679	0.679	-0.063
		Catboost	0.722	0.630	0.538	0.581	0.630	0.776	0.703	0.134
		MLP	0.711	0.593	0.507	0.546	0.610	0.753	0.683	0.117
1500	Hospital de Clínicas da USP São Paulo - SP	XGBoost	0.780	0.525	0.783	0.628	0.680	0.607	0.834	0.090
		LightGBM	0.758	0.413	0.969	0.579	0.630	0.234	0.932	-0.242
		Catboost	0.781	0.652	0.453	0.535	0.660	0.866	0.740	0.220
		MLP	0.781	0.624	0.547	0.583	0.670	0.817	0.765	0.221
1359	Hospital Português da Bahia Salvador - BA	XGBoost	0.856	0.360	0.321	0.340	0.330	0.958	0.95	0.033
		LightGBM	0.869	0.393	0.393	0.393	0.360	0.955	0.955	-0.055
		Catboost	0.836	0.000	0.000	0.000	0.000	1.000	0.000	0.068
		MLP	0.891	0.389	0.250	0.304	0.360	0.971	0.946	0.110
845	Hospital UNIMED Fortaleza - CE	XGBoost	0.722	0.429	0.273	0.333	0.370	0.946	0.897	-0.032
		LightGBM	0.658	0.217	0.151	0.179	0.270	0.919	0.879	-0.180
		Catboost	0.87	0.000	0.000	0.000	0.000	1.000	0.000	-0.014
		MLP	0.695	0.400	0.061	0.105	0.310	0.986	0.876	0.057
539	Hospital Regional de Luiziania Luiziania - GO	XGBoost	0.619	0.629	0.960	0.760	0.710	0.111	0.636	-0.117
		LightGBM	0.667	0.611	1.000	0.759	0.000	0.000	0.000	0.005
		Catboost	0.666	0.667	0.869	0.754	0.750	0.318	0.606	0.064
		MLP	0.674	0.692	0.748	0.718	0.770	0.476	0.545	0.057
456	Hospital Moinhos de Vento Porto Alegre - RS	XGBoost	0.876	0.444	0.286	0.348	0.520	0.959	0.922	0.223
		LightGBM	0.900	0.400	0.143	0.210	0.460	0.976	0.909	0.179
		Catboost	0.865	0.500	0.143	0.222	0.460	0.984	0.910	0.188
		MLP	0.911	0.538	0.500	0.518	0.600	0.951	0.944	0.222

N	Place	Algorithm	AUROC	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
449	Hospital UNIMED Rio de Janeiro - RJ	XGBoost	0.775	0.765	0.406	0.531	0.590	0.961	0.839	0.203
		LightGBM	0.739	0.313	0.812	0.452	0.470	0.447	0.885	-0.623
		Catboost	0.765	1.000	0.125	0.222	0.000	1.000	0.000	0.128
		MLP	0.800	1.000	0.156	0.270	0.000	1.000	0.000	0.203
296	Hospital Universitário Clementino Fraga Filho Rio de Janeiro - RJ	XGBoost	0.751	0.537	0.688	0.603	0.650	0.667	0.792	0.044
		LightGBM	0.756	0.590	0.719	0.648	0.640	0.719	0.820	-0.031
		Catboost	0.750	0.594	0.594	0.594	0.670	0.772	0.772	0.122
		MLP	0.728	0.654	0.531	0.586	0.660	0.842	0.762	0.124
281	Hospital Santa Lúcia Brasília - DF	XGBoost	0.755	0.600	0.529	0.562	0.560	0.912	0.886	0.136
		LightGBM	0.755	0.562	0.529	0.545	0.420	0.897	0.884	-0.081
		Catboost	0.739	0.600	0.353	0.444	0.500	0.941	0.853	0.015
		MLP	0.810	0.667	0.353	0.462	0.570	0.956	0.855	0.203
247	Hospital Santa Júlia Manaus - AM	XGBoost	0.778	0.333	0.167	0.222	0.400	0.936	0.855	0.067
		LightGBM	0.760	0.417	0.417	0.417	0.380	0.889	0.889	-0.162
		Catboost	0.827	0.000	0.000	0.000	0.000	0.984	0.000	0.036
		MLP	0.734	0.400	0.167	0.235	0.420	0.952	0.857	-0.137
148	Hospital Santa Catarina Blumenau - SC	XGBoost	0.885	0.333	0.333	0.333	0.490	0.952	0.952	-0.017
		LightGBM	0.905	0.250	0.333	0.286	0.550	0.929	0.951	-0.110
		Catboost	0.889	0.500	0.333	0.400	0.540	0.976	0.953	0.054
		MLP	0.936	0.333	0.333	0.333	0.460	0.952	0.952	-0.362
124	Hospital São Francisco Mogiguaçu - SP	XGBoost	0.970	0.800	0.800	0.800	0.840	0.970	0.970	-0.009
		LightGBM	0.816	0.417	1.000	0.588	0.000	0.788	0.000	-0.200
		Catboost	0.915	0.667	0.400	0.500	0.540	0.970	0.914	0.293
		MLP	0.818	0.273	0.600	0.375	0.360	0.758	0.926	-0.946

N	Place	Algorithm	AUROC	Precision	Recall	F1	AUPRC	Specificity	NPV	Brier
112	Hospital de Clínicas da UFPE Recife - PE	XGBoost	0.786	0.600	0.500	0.545	0.710	0.818	0.750	0.234
		LightGBM	0.805	0.556	0.833	0.667	0.730	0.636	0.875	0.074
		Catboost	0.833	0.800	0.667	0.727	0.820	0.909	0.833	0.146
		MLP	0.773	0.533	0.667	0.593	0.730	0.682	0.789	-0.014
91	Hospital Escola da UFPel Pelotas - RS	XGBoost	0.756	0.500	0.800	0.615	0.760	0.556	0.833	0.022
		LightGBM	0.761	0.450	0.900	0.600	0.660	0.389	0.875	-0.321
		Catboost	0.778	0.556	0.500	0.526	0.600	0.778	0.737	0.143
		MLP	0.750	0.500	0.500	0.500	0.650	0.722	0.722	-0.345
78	Hospital de Urgências de Trindade Trindade - GO	XGBoost	0.601	0.500	0.769	0.606	0.720	0.091	0.250	-0.390
		LightGBM	0.542	0.542	1.000	0.703	0.000	0.000	0.000	-0.131
		Catboost	0.559	0.500	0.615	0.552	0.680	0.273	0.375	-0.006
		MLP	0.503	0.545	0.462	0.500	0.590	0.545	0.462	-0.783
73	Hospital Universitário Walter Cantídio Fortaleza - CE	XGBoost	0.323	0.167	0.167	0.167	0.230	0.688	0.688	-0.267
		LightGBM	0.734	0.273	1.000	0.429	0.000	0.000	0.000	-0.936
		Catboost	0.573	0.000	0.000	0.000	0.000	1.000	0.000	-0.079
		MLP	0.500	0.286	0.333	0.308	0.280	0.688	0.733	-0.863
56	Hospital Evangélico de Vila Velha Vila Velha - ES	XGBoost	0.769	0.667	0.500	0.571	0.530	0.923	0.857	0.208
		LightGBM	0.500	0.235	1.000	0.381	0.000	0.000	0.000	-0.389
		Catboost	0.808	0.000	0.000	0.000	0.000	1.000	0.000	0.019
		MLP	0.500	0.500	0.250	0.333	0.350	0.923	0.800	-0.297
47	Hospital Universitário Getúlio Vargas Manaus - AM	XGBoost	0.720	1.000	0.400	0.571	0.000	1.000	0.000	0.166
		LightGBM	0.500	0.333	1.000	0.500	0.000	0.000	0.000	-0.125
		Catboost	0.600	0.500	0.200	0.286	0.430	0.900	0.692	-0.141
		MLP	0.720	1.000	0.200	0.333	0.000	1.000	0.000	0.037

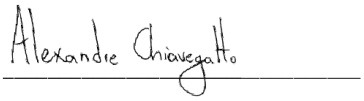
Anexo A – Declaração de Ciência e Aceitação de utilização de artigo em caso de coautoria

Declaração de ciência e aceitação de utilização de artigo em caso de coautoria

Os autores do manuscrito intitulado *"Generalization of machine learning models: A systematic review of the literature for health applications"*, ainda não submetido a periódico científico: *Keisyanne de Araújo Moura, Helena Silveira Schuch, Alexandre Dias Porto Chiavegatto Filho*, por meio deste suficiente instrumento, declaram, para os devidos fins, que concordam com a utilização do manuscrito como resultado da Tese de doutorado intitulada *"Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde"*, apresentada pela aluna *Mariane Furtado Borba* ao programa de pós-graduação em epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. Os autores declaram ainda que aceitam a condição de que o referido manuscrito não seja utilizado em nenhuma outra Tese ou Dissertação.

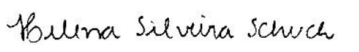
São Paulo, 14 de abril de 2023.

Assinaturas:



Alexandre Dias Porto Chiavegatto Filho

Keisyanne de Araújo Moura



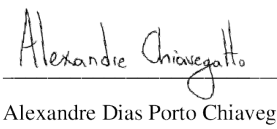
Helena Silveira Schuch

Declaração de ciência e aceitação de utilização de artigo em caso de coautoria

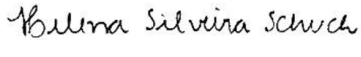
Os autores do manuscrito intitulado *"Population-based validation of machine learning models for neonatal mortality prediction"*: *Helena Silveira Schuch, Hellen Geremias dos Santos, André Filipe de Moraes Batista, Alexandre Dias Porto Chiavegatto Filho*, por meio deste suficiente instrumento, declaram, para os devidos fins, que concordam com a utilização do artigo como resultado da Tese de doutorado intitulada *"Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde"*, apresentada pela aluna *Mariane Furtado Borba* ao programa de pós-graduação em epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. Os autores declaram ainda que aceitam a condição de que o referido manuscrito não seja utilizado em nenhuma outra Tese ou Dissertação.

São Paulo, 20 de abril de 2023.

Assinaturas:



Alexandre Dias Porto Chiavegatto Filho



Helena Silveira Schuch

Hellen Geremias dos Santos

Carmen Simone Grilo Diniz

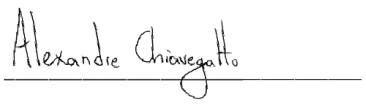
André Filipe de Moraes Batista

Declaração de ciência e aceitação de utilização de artigo em caso de coautoria

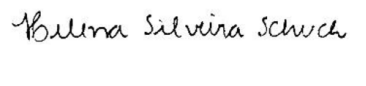
Os autores do manuscrito intitulado *"Generalization of transfer learning algorithms for tabular healthcare data"*: *Helena Silveira Schuch, Alexandre Dias Porto Chiavegatto Filho*, por meio deste suficiente instrumento, declaram, para os devidos fins, que concordam com a utilização do artigo como resultado da Tese de doutorado intitulada *"Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde"*, apresentada pela aluna *Mariane Furtado Borba* ao programa de pós-graduação em epidemiologia da Faculdade de Saúde Pública da Universidade de São Paulo. Os autores declaram ainda que aceitam a condição de que o referido manuscrito não seja utilizado em nenhuma outra Tese ou Dissertação.

São Paulo, 14 de abril de 2023.

Assinaturas:



Alexandre Dias Porto Chiavegatto Filho



Helena Silveira Schuch

Anexo B – Lattes

Alexandre Dias Porto Chiavegatto Filho

Bolsista de Produtividade em Pesquisa do CNPq - Nível 2

Endereço para acessar este CV: http://lattes.cnpq.br/5517850224634709

ID Lattes: **5517850224634709**

Última atualização do currículo em 11/04/2023

Possui graduação em Economia pela USP, doutorado em Saúde Pública pela USP e pós-doutorado na Universidade Harvard. É Professor Livre Docente do Departamento de Epidemiologia da Faculdade de Saúde Pública da USP e orientador dos programas de pós-graduação de Saúde Pública, Bioinformática e Epidemiologia da USP. Atuou como professor convidado (2016 e 2020) e pesquisador visitante (2017 e 2019) na Universidade Harvard. Em 2020, recebeu o Prêmio Abril e Dasa de Inovação Médica. Em 2022, recebeu o Prêmio Excelência para Novas Lideranças em Pesquisa da USP. Atualmente é editor científico da Revista de Saúde Pública, presidente da Comissão de Pesquisa e Inovação da FSP/USP, e membro e Topic Driver da iniciativa Artificial Intelligence for Health (AI4H) da Organização Mundial da Saúde (OMS). Nos últimos anos, tem sido o Pesquisador Principal de projetos de inteligência artificial em saúde financiados pela FAPESP, CNPq, Microsoft, Secretaria Estadual de Saúde de São Paulo e Fundação Lemann. É o diretor do Laboratório de Big Data e Análise Preditiva em Saúde (Labdaps) da FSP/USP. É também o coordenador da rede IACOV-BR (Inteligência Artificial para Covid-19 no Brasil), que tem como objetivo desenvolver algoritmos de machine learning para o diagnóstico e prognóstico de covid-19 nas cinco regiões brasileiras. Tem experiência em pesquisas na área de saúde pública, com ênfase em estatísticas de saúde e machine learning. **(Texto informado pelo autor)**

Identificação

Nome Alexandre Dias Porto Chiavegatto Filho

Nome em citações bibliográficas Chiavegatto Filho, A.D.P.;CHIAVEGATTO FILHO, A. D. P.;Chiavegatto Filho, Alexandre Dias Porto;Filho, Alexandre Dias Porto Chiavegatto;Chiavegatto Filho, Alexandre DP;Chiavegatto, Alexandre Dias Porto;Chiavegatto Filho, Alexandre;Filho, Alexandre Chiavegatto;CHIAVEGATTO FILHO, ALEXANDRE D P;CHIAVEGATTO FILHO, A D P;CHIAVEGATTO FILHO, ALEXANDRE D. P.;DIAS PORTO CHIAVEGATTO FILHO, ALEXANDRE;Chiavegatto Filho, A. D. P.;CHIAVEGATO FILHO, ALEXANDRE D. P.;CHIAVEGATTO FILHO, A;CHIAVEGATTO FILHO, ALEXANDRE D.P.

Lattes iD http://lattes.cnpq.br/5517850224634709

Endereço

Endereço Profissional Faculdade de Saúde Pública.
Av. Dr. Arnaldo, 715
Cerqueira César
01246904 - São Paulo, SP - Brasil
Telefone: (11) 30617914
URL da Homepage: https://www.fsp.usp.br/labdaps

Formação acadêmica/titulação

2007 - 2010 Doutorado em Saúde Pública (Conceito CAPES 6).
Universidade de São Paulo, USP, Brasil.
com **período sanduíche** em Harvard School of Public Health (Orientador: Ichiro Kawachi).
Título: Efeito da desigualdade de renda na mortalidade do Município de São Paulo., Ano de obtenção: 2010.
Orientador: Sabina Léa Davidson Gotlieb.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: desigualdade de renda; epidemiologia; mortalidade; São Paulo.
Grande área: Ciências da Saúde
Grande Área: Ciências da Saúde / Área: Saúde Coletiva / Subárea: Saúde Pública.

2008 - 2014 Graduação em Economia.
Universidade de São Paulo, USP, Brasil.

2002 - 2006 Graduação em Nutrição.
Universidade de São Paulo, USP, Brasil.

Pós-doutorado e Livre-docência

2018 Livre-docência.
Universidade de São Paulo, USP, Brasil.
Título: Machine learning em estatísticas de saúde: desafios e mudanças estruturais, Ano de obtenção: 2018.

2011 - 2012 Pós-Doutorado.
Harvard University, HARVARD, Estados Unidos.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Grande área: Ciências da Saúde
Grande Área: Ciências da Saúde / Área: Saúde Coletiva.

Mariane Furtado Borba

Endereço para acessar este CV: http://lattes.cnpq.br/7255563415314720

ID Lattes: **7255563415314720**

Última atualização do currículo em 17/04/2023

Doutoranda em Epidemiologia na Faculdade de Saúde Pública da Universidade de São Paulo. Possui graduação em Economia pela Universidade Federal de Pelotas e mestrado em Economia Aplicada pela mesma instituição. Tem experiência em pesquisas sobre avaliação de políticas públicas intersetoriais. Atualmente desenvolve projetos como pesquisadora interdisciplinar junto ao Laboratório de Big Data e Análise Preditiva em Saúde (LABDAPS) da Faculdade de Saúde Pública da USP na área de análise preditiva em saúde com a aplicação de machine learning. **(Texto informado pelo autor)**

Identificação

Nome Mariane Furtado Borba

Nome em citações bibliográficas FURTADO, M.;BORBA, M. F.;FURTADO, MARIANE;FURTADO BORBA, MARIANE

Lattes iD http://lattes.cnpq.br/7255563415314720

Endereço

Endereço Profissional Universidade de São Paulo, Faculdade de Saúde Pública.
Faculdade de Saúde Pública
Cerqueira César
01246904 - São Paulo, SP - Brasil
Telefone: (11) 30618049

Formação acadêmica/titulação

2019 Doutorado em andamento em Epidemiologia (Conceito CAPES 5).
Universidade de São Paulo, USP, Brasil.
Título: Análise da generalização de algoritmos de machine learning e suas aplicações na otimização de decisões em saúde
Orientador: Alexandre Chiavegatto Filho.
Coorientador: André Filipe de Moraes Batista.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Palavras-chave: Generalization; Machine learning; Predictive Models; Health Decisions.
Grande área: Ciências da Saúde
Grande Área: Ciências da Saúde / Área: Saúde Coletiva / Subárea: Epidemiologia.

2016 - 2018 Mestrado em Organizações e Mercados (Conceito CAPES 4).
Universidade Federal de Pelotas, UFPEL, Brasil.
Título: Efeitos do Programa Primeira Infância Melhor sobre a proficiência em Matemática e Português em alunos do ciclo de alfabetização, Ano de Obtenção: 2018.
Orientador: Felipe Garcia Ribeiro.
Bolsista do(a): Coordenação de Aperfeiçoamento de Pessoal de Nível Superior, CAPES, Brasil.
Setores de atividade: Pesquisa e desenvolvimento científico.

2010 - 2014 Graduação em Ciências Econômicas.
Universidade Federal de Pelotas, UFPEL, Brasil.
Título: Efeitos da Poluição Atmosférica sobre a Saúde da População: Um estudo para as regiões metropolitanas brasileiras.
Orientador: Rodrigo Nobre Fernandez.

2009 interrompida Graduação interrompida em 2010 em Meteorologia.
Universidade Federal de Pelotas, UFPEL, Brasil.
Ano de interrupção: 2010

2007 - 2008 Ensino Médio (2º grau).
Colégio Estadual Dom João Braga, DJB, Brasil.

2006 - 2007 Ensino Médio (2º grau).
Escola Estadual Osvaldo Camargo, EEOC, Brasil.

1999 - 2005 Ensino Fundamental (1º grau).
Escola Estadual Osvaldo Camargo, EEOC, Brasil.

1998 - 1999 Ensino Fundamental (1º grau).
EMEF Amélia Schemes, EMEFAS, Brasil.