III SIMPOSIO BRASILEIRO DE BIOMETEOROLOGIA HUMANA
26 a 28 de agosto de 2026

MODELO DE MACHINE LEARNING PARA PREVISÃO DE MORTALIDADE CARDIOVASCULAR NOS DIAS DE INVERNO NA CIDADE DE PORTO ALEGRE

Márcio do Carmo Pinheiro (1); Anderson Spohr Nedel (2); Fúlvio Borges Nedel (3)
(1) Doutor, Universidade Federal da Fronteira Sul, marcio.pinheiro@uffs.edu.br
(2) Doutor, Universidade Federal da Fronteira Sul, anderson.nedel@uffs.edu.br
(3) Doutor, Universidade Federal de Santa Catarina, fulvionedel@gmail.com

RESUMO
Avaliamos a influência das variáveis ambientais dos dias de frio intenso na mortalidade de idosos por doenças cardiovasculares na cidade de Porto Alegre/RS. Dados climatológicos diários calculados a partir dos registros da Estação Meteorológica automática Porto Alegre/Jardim Botânico foram associados à mortalidade de idosos fornecida pelo DATASUS. As médias históricas da mortalidade cardiovascular de idosos nos meses de julho superam a mortalidade dos meses de verão em mais de 40%. O atraso do tempo de resposta entre o início do frio intenso e o registro dos óbitos por DCVs é bastante difuso, mas o impacto mostra-se levemente mais relevante 4 dias após a ocorrência do frio extremo. Um modelo de previsão da elevação da mortalidade nesses dias extremos a partir de valores diários de velocidade do vento, temperatura e umidade do ar previu com boa taxa de acertos (> 75%) a média móvel de sete dias, embora não detecte oscilações extremas com resolução temporal diária.

Palavras-chave: Idosos, doenças cardiovasculares, aprendizado de máquina, biometeorologia.

1. INTRODUÇÃO

As condições atmosféricas exercem influência significativa sobre a saúde humana, especialmente entre grupos populacionais mais vulneráveis, como idosos, crianças e indivíduos portadores de doenças crônicas. Estudos recentes demonstram que períodos de frio intenso e de calor extremo podem elevar substancialmente os riscos de hospitalizações e óbitos, em razão do aumento da sobrecarga fisiológica imposta aos mecanismos de termorregulação e ao sistema cardiovascular (ARBUTHNOTT; HJALMARS; ÅSTRÖM, 2020; NICK et al., 2022). Embora os impactos das ondas de calor tenham recebido crescente atenção nas últimas décadas, especialmente em função das mudanças climáticas, os episódios de frio intenso continuam como um importante fator de risco para a saúde em diversas regiões do mundo e, no contexto brasileiro, especialmente na região Sul. Pesquisas recentes mostram que essas situações estão associadas ao aumento da morbimortalidade, especialmente entre a população idosa, grupo cuja capacidade de adaptação fisiológica às variações térmicas tende a ser reduzida (FERREIRA et al., 2019; KRUGER; NEDEL, 2022). Nesse contexto, o desenvolvimento de ferramentas capazes de antecipar períodos de maior risco torna-se uma estratégia relevante para a gestão em saúde pública.

Neste trabalho, investigamos a influência das condições meteorológicas de inverno sobre a mortalidade de idosos por doenças cardiovasculares na cidade de Porto Alegre/RS, utilizando dados observacionais de saúde e meteorologia entre os anos 2000 e 2025. Além de avaliar a relação entre eventos de frio intenso e o aumento da mortalidade, buscamos desenvolver um modelo preditivo baseado em aprendizado de máquina capaz de antecipar elevações na taxa de óbitos, contribuindo para o aperfeiçoamento de estratégias de vigilância e mitigação dos impactos dos extremos climáticos sobre a saúde da população.

2. OBJETIVO

Neste trabalho, buscamos avaliar o impacto de variáveis meteorológicas sobre a mortalidade de idosos por DCVs nos períodos de inverno em Porto Alegre/RS entre os anos de 2001 a 2024. Nessa análise, buscamos identificar o tempo de resposta entre o início do evento meteorológico (frio intenso) e o registro dos óbitos por DCVs e procuramos selecionar as variáveis meteorológicas mais relevantes nessa associação. A partir desses primeiros resultados, o objetivo central desse trabalho é desenvolver um modelo de previsão das elevações do número de óbitos de idosos por DCVs usando algoritmos de aprendizado de máquina (ML),

que possa ser acoplado aos dados de previsão do tempo, a fim de gerar alertas que permitam aos sistemas públicos a mitigação dos impactos dos eventos climáticos mais extremos.

3. MÉTODO

O número diário de óbitos de idosos (população com idade igual ou superior a 60 anos) por doenças cardiovasculares em Porto Alegre foi obtido a partir dos microdados disponibilizados pelo Departamento de Informática do SUS (DATASUS). Para o download dos dados brutos e o devido pré-processamento, utilizamos a biblioteca *microdatasus*, disponibilizada na linguagem de programação estatística *R* (R CORE TEAM, 2025) por Saldanha, Bastos e Barcellos (2019). A sequência do processamento, a obtenção dos dados meteorológicos, a pré-análise e a construção do modelo foram realizadas na linguagem *Python* com as funções da biblioteca *pandas*.

O *download* dos dados meteorológicos foi realizado diretamente a partir do banco de dados históricos do Instituto Nacional de Meteorologia (INMET, 2026), de onde foram extraídos dados horários de temperatura, umidade relativa do ar, velocidade e direção do vento medidos na Estação Meteorológica automática Porto Alegre/Jardim Botânico (A801), que entrou em operação em 22 de setembro de 2000.

3.1. Tratamento dos dados meteorológicos

A partir dos dados meteorológicos, calculamos o índice de sensação térmica *Wind Chill* (WC) horário definido por Osczevski e Bluestein (2005). Com o objetivo de buscar correlações entre esses dados e os dados diários do DATASUS, reduzimos os dados horários em dados diários médios, máximos e mínimos. Considerando os quatro dados meteorológicos de interesse nesse estudo, observou-se apenas 130 dias com alguma medida não registrada (1,4%), de um total de 9222 dias no período de outubro de 2000 a dezembro de 2025. A fim de manter a completeza dos dados e observando o baixo número de medidas não registradas, optamos por preencher essas medidas por interpolação linear em vez de simplesmente excluí-las, o que levaria à perda de medidas realizadas para algumas das variáveis em dias nos quais apenas um dos dados não foi registrado (somente em 77 dias, ou 0,8% do total, não houve registro de temperatura). A variação da temperatura média diária ao longo de toda a série histórica analisada neste trabalho é mostrada na Figura 1.

3.2. Tratamento dos dados de mortalidade

Os dados brutos obtidos diretamente do DATASUS trazem os registros de cada óbito individualmente. Usando funções da biblioteca *pandas*, os registros foram filtrados pela causa básica registrada (DCV), município de residência e idade igual ou superior a 60 anos e, em seguida, agrupados pela data do registro.

Uma vez que a população idosa no Brasil aumentou consideravelmente no período analisado, fez-se necessário normalizar os dados de mortalidade, calculando o número de óbitos para cada 100 mil habitantes acima de 60 anos em POA, valor que denominamos taxa de óbitos. As estimativas populacionais anuais desagregadas por faixa etária para o município de Porto Alegre foram obtidas junto ao Departamento de Economia e Estatística do Rio Grande do Sul (DEE/SPGG), por meio do portal demográfico PopVis. Para o período de 2000 a 2024, utilizamos a série histórica recomendada pelo próprio DEE, elaborada pelo Ministério da Saúde/Rede Interagencial de Informações para a Saúde (RIPSA), convergente com as projeções populacionais do Instituto Brasileiro de Geografia e Estatística (IBGE) por grupo etário.

Um gráfico da evolução da população idosa de POA entre 2000 e 2024 é mostrado na Figura 2(a). A partir das estimativas do DEE, calculadas para 1º de julho de cada ano, ajustamos polinômios *spline* cúbicos com nós em 2005, 2013, 2019, 2020, 2021 e 2022, a fim de gerar estimativas diárias para todo o período. A proximidade dos nós no final da série histórica fez-se necessária a fim de captar os desvios da tendência de longo prazo ocasionados pela pandemia de COVID-19. Um polinômio de grau quatro também foi ajustado entre os anos 2000 e 2019 e extrapolado para os próximos 5 anos a fim de mostrar o desvio da tendência de longo prazo dos dados após o início da pandemia.

Os dados da série histórica do número de óbitos e da taxa de óbitos são mostrados na Figura 1, junto com os dados de temperatura média diária. As tendências de longo prazo, bem como as sazonalidades, foram obtidas por decomposição da série temporal utilizando o método *STL* do módulo *statsmodels.tsa.seasonal* do *Python*. Os dados e as sazonalidades apresentados estão suavizados com uma média móvel de sete dias a fim de reduzir o ruído. A tendência dos dados de mortalidade era de queda até 2019, mas se inverte a partir desse ano. Mesmo assim, a taxa de óbitos no final da série é aproximadamente 60% daquela observada em outubro de 2000. Um outro ponto bastante claro quando se observa as sazonalidades é a forte correlação entre os dias mais frios do ano e a mortalidade de idosos por DCVs. A correlação de Pearson entre os dados suavizados de temperatura média e a taxa de óbitos é -0,79.

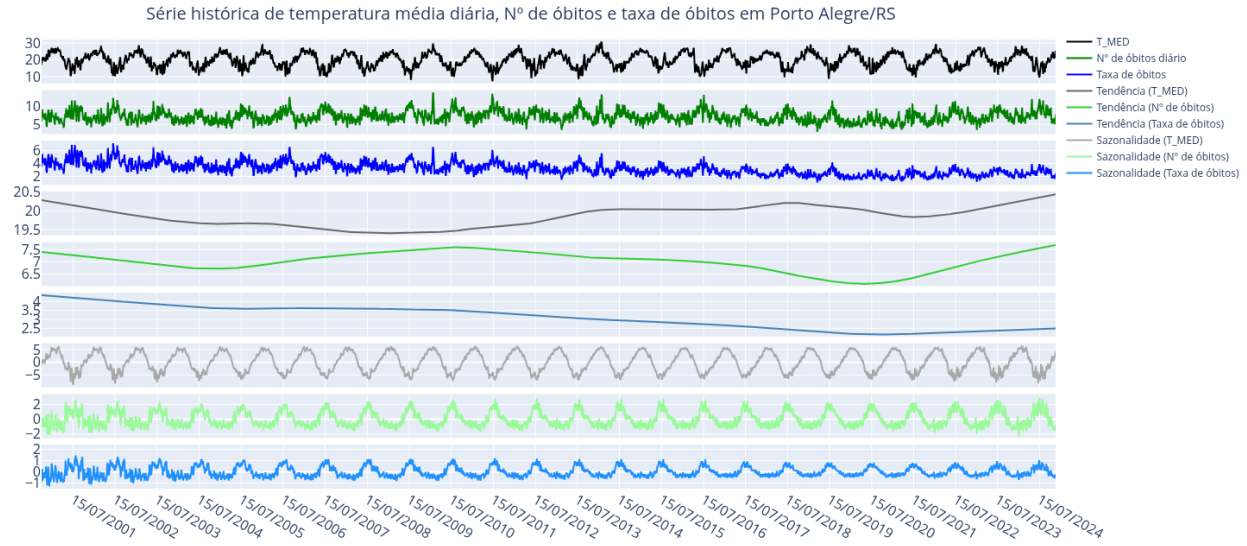


Figura 1 - Série histórica dos dados analisados. Os dados e as sazonalidades estão suavizados com uma média móvel semanal.

3.3. Construção do modelo de previsão da elevação da taxa de óbitos nos dias mais frios

A partir da pré-análise dos dados e da forte correlação encontrada entre os meses mais frios do ano e a mortalidade de idosos em POA, buscamos construir um modelo que nos permita prever as elevações do número de óbitos de idosos por DCVs usando algoritmos de aprendizado de máquina (ML) e que possa ser acoplado aos dados de previsão do tempo, a fim de gerar alertas que permitam aos sistemas públicos a mitigação dos impactos dos eventos climáticos mais extremos.

Diversos modelos preditivos foram implementados em *Python* utilizando a biblioteca *scikit-learn* e o algoritmo *Random Forest Regressor* (RFR) para regressão supervisionada. Embora tenhamos dados diários médios, máximos e mínimos de temperatura, umidade, velocidade e direção do vento em todo o período, a inserção de todos esses dados nos modelos resultou em sobreajuste (*overfitting*). Da mesma forma, os testes mostraram melhor capacidade de previsão quando os modelos consideraram vários *lags* em cada dado meteorológico, uma vez que as correlações entre esses dados e a taxa de óbitos não apontam para um atraso claro entre o extremo climatológico e a resposta dos dados de saúde.

Considerando como métrica o número de eventos em que o modelo foi capaz de prever acertadamente taxas de óbitos acima da mediana do inverno analisado, as melhores previsões foram obtidas a partir do modelo alimentado com os valores diários de WC médio (WC_MED), máxima umidade relativa do ar (U_MAX) e valor mínimo de velocidade do vento (V_MIN), simultaneamente com os valores de 1 a 7 dias anteriores aos registros dos óbitos (*lags* de 1 a 7 dias na resposta). As correlações de Pearson entre essas variáveis e a taxa de óbitos para cada *lag* são mostradas na Figura 2(b).

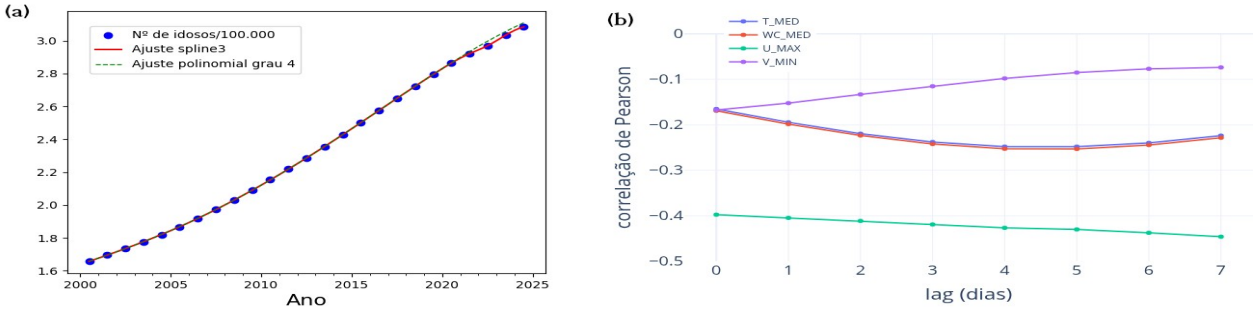


Figura 2 - (a) População de idosos de POA entre os anos 2000 e 2024 (azul). Ajuste *spline* cúbico com nós em 2005, 2013, 2019, 2020, 2021 e 2022 (vermelho). Polinômio de grau quatro ajustado entre os anos 2000 e 2019 e extrapolado para os próximos 5 anos a fim de mostrar o desvio da tendência de longo prazo dos dados após o início da pandemia de COVID-19 (verde). (b) Correlações de Pearson entre as variáveis climatológicas usadas no modelo e a taxa de óbitos para cada *lag*.

4. RESULTADOS

O modelo final foi construído usando como dados de treino todos os dias de junho, julho e agosto entre os anos de 2001 e 2023, a fim de que pudesse ser testado com os dados do inverno de 2024, ano limite para o qual os dados de mortalidade estão consolidados. Os resultados do modelo são mostrados na Figura 3. O modelo conseguiu prever taxa de óbitos acima da mediana do inverno de 2024 em 76% dos dias em que isso ocorreu, com resultado falso positivo em 31% das previsões. Nessa mesma linha, dos 24 dias em que a

taxa de óbitos superou a mediana do inverno em 10%, o modelo conseguiu prever 18 (75%), com resultado falso positivo em 22% das previsões. Em ambos os casos, o erro médio quadrático entre os valores previstos e os dados suavizados foi de aproximadamente 7%.

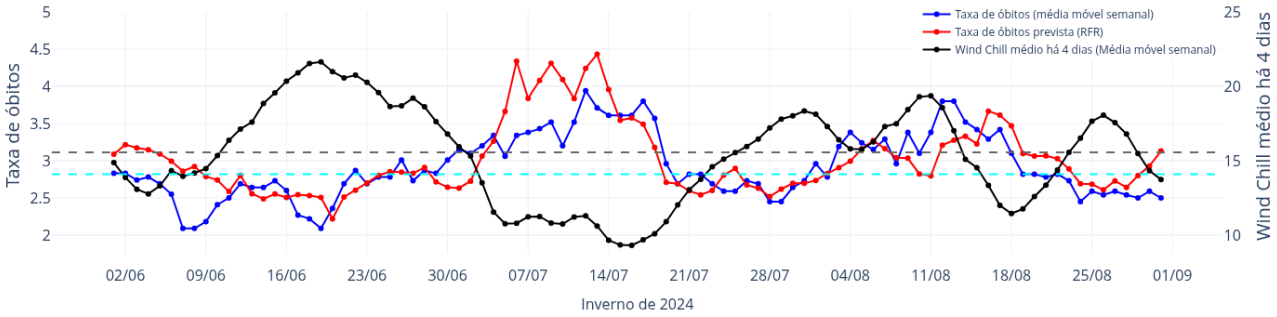


Figura 3 - Taxa de óbitos diária suavizada (azul), taxa de óbitos prevista pelo modelo treinado com o algoritmo de ML RFR (vermelho) e índice de conforto térmico *wind chill* (WC) com lag de 4 dias, também suavizado (preto). As linhas tracejadas marcam as medianas dos valores de WC_MED (cinza) e da taxa de óbitos (ciano) para o inverno de 2024.

5. CONCLUSÕES

O inverno rigoroso da cidade de Porto Alegre/RS mostra ter correlação importante com o aumento da mortalidade de idosos registrado nessa época do ano. As médias históricas da mortalidade cardiovascular nos meses de julho superam a mortalidade dos meses de verão em mais de 40%. O atraso do tempo de resposta entre o início do frio intenso e o registro dos óbitos por DCVs é bastante difuso, mas o impacto mostra-se levemente mais relevante 4 dias após a ocorrência do frio extremo. O modelo de previsão da elevação do registro de óbitos quando da ocorrência de evento meteorológico extremo no inverno de Porto Alegre prevê com boa taxa de acertos (> 75%) a média móvel de sete dias, mas não é capaz de prever oscilações diárias extremas. Uma expansão do modelo para a região metropolitana e para outras macrorregiões do RS será relizada nos próximos meses.

REFERÊNCIAS BIBLIOGRÁFICAS

ARBUTHNOTT, K. G.; HAJAT, S. The health effects of hotter summers and heat waves in the population of the United Kingdom: **a review of the evidence. Environmental Health**, v. 16, Suppl. 1, 119, 2017.

FERREIRA, L. C. M.; NOGUEIRA, M. C.; PEREIRA, R. V. B. et al. Ambient temperature and mortality due to acute myocardial infarction in Brazil: an ecological study of time-series analyses. **Scientific Reports**, v. 9, 13790, 2019.

INSTITUTO NACIONAL DE METEOROLOGIA (INMET). Banco de dados meteorológicos para ensino e pesquisa: dados históricos. Brasília: **INMET**, 2026. Disponível em: https://portal.inmet.gov.br/dadoshistoricos. Acesso em: 28 maio 2026.

KRÜGER, E. L.; NEDEL, A. S. Investigating the relationship between climate and hospital admissions for respiratory diseases before and during the COVID-19 pandemic in Brazil. **Sustainability**, v. 15, n. 288, p. 1–15, 2022.

NICK, L. M.; NEDEL, A. S.; ALONSO, M. F.; MARQUES, J. Q.; FREITAS, R. A. P. Relationship between meteorological variables and pneumonia in children in the Metropolitan Region of Porto Alegre, Brazil. **International Journal of Biometeorology**, v. 66, p. 2301–2308, 2022.

OSCZEVSKI, R.; BLUESTEIN, M. The new wind chill equivalent temperature chart. **Bulletin of the American Meteorological Society,** v. 86, n. 10, p. 1453–1458, 2005.

PEDREGOSA, F. et al. Scikit-learn: machine learning in Python. **Journal of Machine Learning Research**, v. 12, p. 2825–2830, 2011.

R CORE TEAM. R: A language and environment for statistical computing. Vienna: **R Foundation for Statistical Computing**, 2025. Disponível em: https://www.r-project.org/. Acesso em: 28 maio 2026.

RIO GRANDE DO SUL. Secretaria de Planejamento, Governança e Gestão. Departamento de Economia e Estatística (DEE). PopVis: **Portal Demográfico – estimativas populacionais por município, sexo e faixa etária**. Porto Alegre: DEE/SPGG, [s.d.]. Disponível em: PopVis – Portal Demográfico . Acesso em: 29 maio 2026.

SALDANHA, R. F.; BASTOS, R. R.; BARCELLOS, C. Microdatasus: pacote para download e pré-processamento de microdados do Departamento de Informática do SUS (DATASUS). **Cadernos de Saúde Pública**, v. 35, n. 9, 2019. DOI: 10.1590/0102-311X00032419.

AGRADECIMENTOS

Os autores agradecem o apoio para o desenvolvimento deste trabalho do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), do Departamento de Ciência e Tecnologia, da Secretaria de Ciência, Tecnologia, Inovação e Complexo da Saúde, do Ministério da Saúde (Decit/SECTICS/MS), projeto nº 444178/2023-6.