



III MCSM

Modelagem Computacional no
Sertão Mineiro



PPGMCS
Programa de Pós-Graduação em
Modelagem Computacional e Sistemas

APLICAÇÃO DE REDES NEURAIAS NA PREDIÇÃO DE EPISÓDIOS DE POLUIÇÃO POR OZÔNIO TROPOSFÉRICO A PARTIR DE DADOS SINTÉTICOS DE MONITORAMENTO ATMOSFÉRICO

Daniel Oliveira Paixão¹ - engsisdaniel.profissional@gmail.com

Karen Pinheiro Couto² - karenpinheiroeu@gmail.com

Ricardo Pereira C. Santos³ - januariaricardo@gmail.com

¹Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

²Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

³Universidade Estadual de Montes Claros, UNIMONTES - Montes Claros, MG, Brazil

Abstract. O ozônio troposférico (O_3) representa um dos principais poluentes atmosféricos associados a impactos negativos à saúde humana e ao ecossistema vegetal. A predição de suas concentrações constitui um desafio relevante em razão das relações não lineares entre as variáveis envolvidas. O presente projeto busca desenvolver e avaliar um modelo de Rede Neural Artificial (RNA) Multilayer Perceptron (MLP) para a estimativa da concentração de O_3 a partir de medições horárias de monóxido de carbono (CO), dióxido de nitrogênio (NO_2) e temperatura ambiente. Os experimentos foram conduzidos sobre uma base de dados sintética, composta por 5.000 amostras geradas aleatoriamente a partir do seguinte modelo matemático com ruído gaussiano aditivo $\varepsilon \sim N(0, 5)$. Os resultados indicam que a abordagem é viável para capturar as relações funcionais entre os poluentes, abrindo perspectivas para aplicações em sistemas de monitoramento ambiental em tempo real. Os resultados sugerem que o modelo alcançou desempenho próximo ao limite teórico de erro irreduzível.

Keywords: Ozônio Troposférico; Redes Neurais Artificiais; Multilayer Perceptron; Poluição Atmosférica; Monitoramento Ambiental.

1. INTRODUÇÃO

A poluição atmosférica constitui um desafio de escala global, resultante de múltiplos fatores de natureza econômica, industrial e energética (VILLALBA, 2025). Dentre os poluentes de maior relevância, destaca-se o ozônio troposférico (O_3), cuja presença em concentrações elevadas tem suscitado crescente preocupação científica. Sua formação secundária é decorrente de reações fotoquímicas complexas envolvendo variáveis, como dióxido de nitrogênio (NO_2), monóxido de carbono (CO) e radiação solar (CAMPOS NETO, NELO, 2026).

De acordo com Villalba (2025), a preocupação pela crescente elevação do O_3 na atmosfera justifica-se por um conjunto de características intrínsecas a esse composto: sua capacidade de atingir regiões geograficamente distantes das fontes primárias de emissão, seu papel como agente do efeito estufa e, sobretudo, seus comprovados efeitos adversos à saúde humana e aos ecossistemas vegetais.

Estudos conduzidos ao longo de 14 anos em cidades dos Estados Unidos da América (EUA) concluíram que elevadas concentrações de ozônio troposférico estão associadas a maiores taxas de mortalidade nas populações expostas (LIMA, FELIX, CARDOSO, 2021).

A elevação das concentrações de O_3 também causa danos ao ecossistema vegetal e afeta a fauna terrestre. Segundo Santos (2024), a exposição ao O_3 pode comprometer processos biológicos essenciais e provocar efeitos adversos nas plantas, impactando a produção agrícola e reduzindo os rendimentos de forma modesta, mas com consequências economicamente relevantes.

Diante desse cenário, o desenvolvimento de modelos preditivos capazes de estimar as concentrações de O_3 com antecedência é de fundamental importância para subsidiar políticas públicas de controle da qualidade do ar e sistemas de alerta precoce. As Redes Neurais Artificiais (RNA), como as arquiteturas do tipo Multilayer Perceptron (MLP), têm demonstrado potencial na modelagem de fenômenos ambientais, constituindo uma alternativa aos modelos físico-químicos tradicionais.

Trabalhos como o de Campos Neto e Nelo (2026) apontam que existem diversas variáveis para determinar a concentração de ozônio troposférico em uma região específica, como a amplitude térmica, temperatura máxima, altura da camada limite, dióxido de nitrogênio (NO_2), radiação e recorte temporal. Neste trabalho, entretanto, iremos nos restringir a utilizar uma base sintética de uma estação de monitoramento ambiental que coleta dados horários de três poluentes para prever a concentração de ozônio (O_3) no ar (em $\mu g/m^3$). As variáveis disponíveis são de concentração de CO (mg/m^3), NO_2 ($\mu g/m^3$) e temperatura ambiente ($^{\circ}C$).

Nesse contexto, este trabalho tem como objetivo desenvolver e avaliar um modelo MLP para a predição da concentração de ozônio troposférico a partir de dados horários de CO, NO_2 e temperatura. O estudo utiliza uma base de dados sintética gerada a partir de uma função matemática ruidosa, permitindo a avaliação controlada do desempenho do modelo.

2. Referencial Teórico

2.1 Ozônio Troposférico e seus Impactos

O ozônio (O_3) é um gás presente naturalmente na atmosfera terrestre. Existem dois tipos de ozônio: o estratosférico (25 e 35 Km de altitude) e o troposférico (10 a 15 Km de altitude). O primeiro é considerado benéfico por absorver grande parte da radiação ultravioleta solar, protegendo o ecossistema terrestre (SARTORI, 2026).

Por outro lado, o ozônio troposférico é classificado como um poluente tóxico e um dos principais componentes da névoa fotoquímica observada em regiões metropolitanas com intenso fluxo de veículos. Ainda segundo Sartori (2026), esse poluente é formado por reações envolvendo gases emitidos principalmente por atividades humanas, sob ação da radiação solar.

A poluição atmosférica aumentou no período de 2010 a 2019, segundo Santos (2024). Esta expansão se deu principalmente após a elevação térmica e da emissão de gases precursores do O_3 . Para avaliação do cenário mundial, dos 20 maiores países, 12 apresentaram crescimentos nos índices de ozônio troposférico.

“A poluição atmosférica é reconhecida como um dos principais fatores de risco ambiental para a saúde humana, responsável por aproximadamente 7 milhões de mortes prematuras anualmente em escala global” (Oliveira et al., 2026, p.3). Evidências científicas indicam que o ozônio troposférico, em interação com outras partículas atmosféricas, apresenta potencial para desencadear danos ao sistema respiratório, podendo evoluir para processos inflamatórios crônicos e lesões ao sistema cardiovascular (Oliveira et al., 2026).

Apesar de indivíduos sem comorbidades lidarem com efeitos leves do O₃ no organismo, em grupos vulneráveis com doenças pré-existentes, como asmáticos, cardiopatas, idosos e outros, a exposição ao poluente pode desencadear reações adversas graves.

A elevação das concentrações de ozônio troposférico também pode causar danos ao ecossistema vegetal e afetar a fauna terrestre. Para Santos (2024), a exposição ao O₃ pode comprometer processos biológicos essenciais e provocar efeitos adversos nas plantas, impactando a produção agrícola e reduzindo os rendimentos de forma comedida, mas com consequências economicamente relevantes.

2.2 Redes Neurais Artificiais e o Modelo MLP

As Redes Neurais Artificiais (RNAs) são modelos computacionais inspirados no funcionamento do sistema nervoso biológico, compostos por unidades de processamento interconectadas denominadas neurônios artificiais. Uma de suas características mais relevantes é a capacidade de aprender a partir de exemplos. “A etapa de aprendizagem consiste de um processo iterativo de ajustes dos pesos sinápticos, que ao final do processo guardam o conhecimento que a rede adquiriu do ambiente externo” (SOARES, TEIVE, 2015, p. 2).

Ainda segundo Soares e Teive (2015), a capacidade de modelagem não linear das RNAs é essencial para identificar e assimilar padrões frequentemente mascarados por ruídos nos dados, característica relevante para o desenvolvimento do presente trabalho.

O Multilayer Perceptron (MLP) é uma das arquiteturas de Redes Neurais Artificiais consolidadas, também inspirada na organização funcional do sistema nervoso biológico. Sua estrutura é composta por ao menos três camadas sequenciais: a camada de entrada, responsável por receber e transmitir os dados ao restante da rede; uma ou mais camadas ocultas, nas quais ocorre o processamento das representações intermediárias por meio de funções de ativação não lineares; e a camada de saída, que produz o resultado final da predição (SOUZA et al., 2025).

Redes neurais com diversas camadas intermediárias de processamento são classificadas como parte do aprendizado profundo ou Deep Learning (DL), um ramo do aprendizado de máquina dedicado ao uso de arquiteturas de elevada profundidade, capazes de extrair padrões hierárquicos e crescentemente mais complexos a partir dos dados fornecidos como entrada.

“A RNA MLP, por sua vez, é um tipo de rede feedforward que permite a modelagem de múltiplas transições simultaneamente por meio de um processo iterativo de treinamento supervisionado” (NERY et al., 2025, p.1). Nesse cenário, ao longo do processo de treinamento, os neurônios presentes nas camadas internas processam e propagam sinais entre si, possibilitando que a atualização dos parâmetros da rede se dê de maneira progressiva e contínua, até que determinadas condições de encerramento sejam satisfeitas, como o esgotamento do número de épocas definido ou a estabilização do erro abaixo de um patamar aceitável (NERY et al., 2025).

2.3 Aplicações de MLP na Predição da Qualidade do Ar

A literatura registra aplicações bem-sucedidas de redes neurais na predição de poluentes atmosféricos (SOUZA et al., 2025). Modelos MLP é um dos empregados para estimar concentrações de O_3 em contextos urbanos e industriais, superando o desempenho de modelos físico-químicos em cenários com alta variabilidade dos dados. A capacidade das RNA de aprender representações não lineares a partir de dados históricos as torna especialmente adequadas para a modelagem de fenômenos atmosféricos complexos.

3. Metodologia

3.1 Descrição do Problema e Base de Dados

O problema investigado consiste na estimativa da concentração de ozônio troposférico (O_3 , em $\mu g/m^3$), simulada a partir de dados equivalentes aos obtidos em uma estação de monitoramento ambiental. Para isso foram consideradas três variáveis preditoras: concentração de CO (mg/m^3), concentração de NO_2 ($\mu g/m^3$) e temperatura ambiente ($^{\circ}C$). As faixas de variação das entradas são: CO entre 0,5 e 4,0 mg/m^3 ; NO_2 entre 20 e 120 $\mu g/m^3$; e temperatura entre 15 e 35 $^{\circ}C$.

A base de dados utilizada é sintética, composta por 5.000 amostras geradas aleatoriamente a partir do seguinte modelo matemático com ruído gaussiano aditivo:

$$O_3 = 5 \cdot CO + 0,8 \cdot NO_2 - 0,5 \cdot T + 0,03 \cdot T^2 - 0,01 \cdot CO \cdot NO_2 + \varepsilon$$

Em que $\varepsilon \sim N(0,5)$ representa o ruído estocástico do fenômeno real. Essa formulação captura relações não lineares e interações entre variáveis, reproduzindo a complexidade do comportamento real do ozônio troposférico. A Tabela 1 apresenta a amostra obtida do experimento.

Table 1: Base para concentração de ozônio

CO (mg/m^3)	NO_2 ($\mu g/m^3$)	Temperatura ($^{\circ}C$)	O_3 ($\mu g/m^3$)
0,5	20	15	11,2
1,0	40	18	35,7
1,5	55	20	53,4
2,0	70	22	75,8
2,5	85	25	97,2
3,0	100	28	110,5
3,5	110	30	128,9
0,8	30	32	28,3
2,2	60	35	67,1
4,0	120	33	149,6

3.2 Ferramentas Computacionais

A implementação foi realizada em Python 3, com o suporte das seguintes bibliotecas: numpy, para a geração das amostras sintéticas e cálculo do ruído gaussiano; pandas, para a

organização e inspeção da base de dados; e scikit-learn, para o pré-processamento, treinamento da MLP e cálculo das métricas de desempenho. De acordo com Pedregosa et al. (2011), o Python se tornou uma escolha atraente para o desenvolvimento algorítmico e análise de dados graças à sua maturidade em ferramentas científicas e sua natureza interativa. Os autores ressaltam que a linguagem transcendeu a academia, ganhando forte espaço na indústria. É nesse contexto que opera o Scikit-learn, uma biblioteca que disponibiliza recursos de ponta para machine learning por meio de uma estrutura de uso simplificado e integrada nativamente ao Python.

4. Resultados

O modelo MLP foi treinado sobre uma base sintética de 5.000 amostras geradas a partir da função matemática com ruído gaussiano $\varepsilon \sim N(0, 5)$, divididas em 4.000 amostras para treinamento e 1.000 para teste, na proporção 80/20. As variáveis preditoras — CO, NO₂ e temperatura — foram normalizadas pelo método Min-Max antes da alimentação à rede, procedimento que contribui para a estabilidade e a velocidade de convergência do algoritmo de otimização L-BFGS. Ao treinar um modelo preditivo, é necessário quantificar quão distantes as previsões estão da realidade. Para isso, foram utilizadas as seguintes métricas: MSE (Erro Quadrático Médio), RMSE (Raiz do Erro Quadrático Médio), MAPE (Erro Percentual Absoluto Médio) e o MSE teórico mínimo, que representa o erro irreduzível — o piso teórico que nenhum modelo consegue superar, pois decorre do ruído intrínseco do fenômeno, independentemente da qualidade do modelo.

A Tabela 2 apresenta as métricas de desempenho obtidas no conjunto de testes.

Table 2: Métricas de desempenho obtidas

Métrica	Valor
MSE	26,0014 ($\mu\text{g}/\text{m}^3$) ²
RMSE	5,099 $\mu\text{g}/\text{m}^3$
MAPE	6,26%
MSE teórico mínimo	25,00 ($\mu\text{g}/\text{m}^3$) ²

Os resultados indicam que o modelo alcançou desempenho próximo ao limite teórico de erro irreduzível, dado pela variância do ruído gaussiano adicionado à função geradora ($\sigma^2 = 25,00(\mu\text{g}/\text{m}^3)^2$). O MSE obtido de 26,0014 ($\mu\text{g}/\text{m}^3$)² supera esse limite mínimo por apenas 4,01%, indicando que a rede aprendeu de forma eficaz as relações funcionais entre os poluentes, sem capturar o componente estocástico do ruído — comportamento esperado e desejável em um modelo bem ajustado.

O RMSE de aproximadamente 5,10 $\mu\text{g}/\text{m}^3$ deve ser interpretado à luz da faixa de variação da variável-alvo. O erro médio percentual de 6,26% confirma que, em termos relativos, as previsões apresentam desvio médio inferior a 7% em relação aos valores reais, o que representa acurácia satisfatória para a modelagem de fenômenos atmosféricos com ruído intrínseco.

5. Conclusão

A análise das previsões no conjunto de teste indica desempenho consistente ao longo da faixa de variação da variável-alvo, com RMSE de 5,10 $\mu\text{g}/\text{m}^3$ e MAPE de 6,26%, sugerindo

que o modelo generaliza adequadamente para amostras não vistas durante o treinamento.

Esses resultados indicam que a arquitetura MLP com camadas ocultas de dimensões (10, 8, 4), função de ativação ReLU e regularização L2 ($\alpha = 10^{-4}$) é adequada para a tarefa de predição da concentração de ozônio troposférico a partir das variáveis de entrada consideradas, constituindo uma base sólida para extensões futuras com dados reais de estações de monitoramento ambiental.

Agradecimentos

Os autores agradecem à FAPEMIG (grant n OET-00243-26), UFVJM, Centro Universitário UniÚnica e ao Programa de Pós Graduação em Modelagem Computacional e Sistemas (PPGMCS/UNIMONTES).

REFERENCES

- Campos Neto, Antonio; Nelo, Humberto C. S. (2026), “Modelagem da concentração de ozônio troposférico para a região Nordeste com uso de aprendizado de máquina”, *Sociedade e Território*, Natal, v. 38, n. 1, p. 275-285, jan./abr. 2026. Seção Especial – Prêmio José Lacerda 2025. ISSN 2177-8396. Disponível em: <https://periodicos.ufrn.br/sociedadeeterritorio/article/view/42149/22118>. Acesso em: 26 maio 2026.
- Lima, Manoel J. A.; Felix, Erika P.; Cardoso, Arnaldo A. (2021), “Aplicações e implicações do ozônio na indústria, ambiente e saúde”, *Química Nova*, v. 44, n. 9, p. 1151-1158, 2021. DOI: 10.21577/0100-4042.20170759. Disponível em: <https://quimicanova.sbq.org.br/pdf/RV2020-0461>. Acesso em: 26 maio 2026.
- Nery, Lilliane M. et al. (2025), “Estimativa da evapotranspiração com redes neurais artificiais e dados espectrais: aplicação em uma unidade de conservação”, *Revista Brasileira de Meteorologia*, v. 40, 2025. DOI: 10.1590/0102-778640250003. Disponível em: <https://doi.org/10.1590/0102-778640250003>. Acesso em: 26 maio 2026.
- Oliveira, Tatiane S. et al. (2026), “Qualidade do ar em espaços públicos de recreação e planejamento urbano: um estudo de caso em Capanema, Pará”, *Editora Científica Digital*, v. 1, 2026. DOI: 10.37885/260321406. ISBN 978-65-6155-119-9. Disponível em: <https://downloads.editoracientifica.com.br/articles/260321406.pdf>. Acesso em: 26 maio 2026.
- Pedregosa, Fabian et al. (2011), “Scikit-learn: Machine Learning in Python”, *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Disponível em: <https://www.jmlr.org/papers/volume12/pedregosa11a/pedregosa11a.pdf>. Acesso em: 26 maio 2026.
- Santos, Guilherme C. (2024), *Ozônio troposférico com referência à proteção da vegetação no Estado de São Paulo: 2018-2022*. Dissertação (Mestrado em Qualidade Ambiental) – Universidade Federal de Uberlândia, Uberlândia, 2024. Disponível em: <http://doi.org/10.14393/ufu.di.2024.295>. Acesso em: 26 maio 2026.
- Sartori, Daniele G. P. (2026), “Medições de ozônio troposférico no campus central da UTFPR, Curitiba, PR”, *Aurum Editora*, [S. l.], p. 73–105, 2026. DOI: 10.63330/aurumpub.041-006. Disponível em: <https://aurumpublicacoes.com/index.php/editora/article/view/1130>. Acesso em: 26 maio 2026.
- Soares, Daniel G; Teive, Raimundo C. G. (2015), “Estudo comparativo entre as redes neurais artificiais MLP e RBF para previsão de cheias em curto prazo”, *Revista de Informática Teórica e Aplicada*, [S. l.], v. 22, n. 2, p. 67–86, 2015. DOI: 10.22456/2175-2745.53236. Disponível em: <https://seer.ufrgs.br/index.php/rita/article/view/RITA-VOL22-NR2-67>. Acesso em: 26 maio 2026.
- Souza, Geovana R. P. et al. (2025), “Predição de poluentes atmosféricos via métodos de aprendizado de máquina”, in *XIX Brazilian e-Science Workshop*, 2025, Fortaleza. *Proceedings of the XIX Brazilian e-Science Workshop (BreSci)*. Disponível em: <https://sol.sbc.org.br/index.php/bresci/article/view/36909/36695>. Acesso em: 26 maio 2026.
- Villalba, Isabelle C. C. (2025), *Aplicação de redes neurais artificiais para avaliar a qualidade do ar com base em dados meteorológicos, geográficos e de poluentes atmosféricos*. Trabalho de Conclusão de Curso (Bacharelado em Engenharia Química) – Universidade Estadual Paulista, Instituto de Química, Araraquara, 2025. Disponível em: <https://repositorio.unesp.br/entities/publication/ff7ec5d6-fff8-457d-b1bf-adf87d85c970>. Acesso em: 26 maio 2026.