



DESENVOLVIMENTO DE ESTAÇÃO VIRTUAL DE MONITORAMENTO DA QUALIDADE DO AR COM REDE NEURAL ARTIFICIAL: UMA APLICAÇÃO DIDÁTICA COM DADOS REAIS

DOI: 10.37702/2175-957X.COBENGE.2025.6406

Autores: ITALO PINTO RODRIGUES, JONATHAN DA SILVA SALDANHA, FLAVIA COELHO ROCHA, BIANCA DE CÁSSIA CARDOSO BERALDO, JESSICA FRANCINE DA SILVEIRA PAIVA, MARIA EDUARDA LACERDA DE ALMEIDA, ANA CLAUDIA SILVA DE ALMEIDA, SAMANTHA GRISOL DA CRUZ NOBRE

Resumo: A poluição atmosférica é responsável por milhões de mortes prematuras no mundo, o que torna essencial o monitoramento contínuo da qualidade do ar. No entanto, o alto custo de instalação e manutenção de estações físicas limita sua abrangência espacial, especialmente em países em desenvolvimento. Este artigo apresenta os resultados de uma experiência pedagógica conduzida na disciplina de Concepção e Gestão de Projetos de Engenharia, do curso de Engenharia Ambiental do Centro Universitário de Volta Redonda, cujo objetivo foi desenvolver uma estação virtual baseada em redes neurais artificiais. Utilizaram-se dados históricos de PM₁₀ (2004-2024) de três estações em Volta Redonda/RJ (Belmonte, Retiro e Vila Santa Cecília). A rede, com três camadas ocultas de 30 neurônios cada, obteve MRSE de 55% e correlação moderada ($R = 0,486$), demonstrando viabilidade da abordagem e potencial para formação acadêmica, embora indique a necessidade de modelos mais robustos para fins operacionais.

Palavras-chave: Monitoramento do ar, Estação Virtual, Redes Neurais Artificiais

DESENVOLVIMENTO DE ESTAÇÃO VIRTUAL DE MONITORAMENTO DA QUALIDADE DO AR COM REDE NEURAL ARTIFICIAL: UMA APLICAÇÃO DIDÁTICA COM DADOS REAIS

1 INTRODUÇÃO

A poluição do ar constitui um dos principais riscos ambientais para a saúde humana. Estima-se que, somente na Europa, mais de 300 mil mortes prematuras por ano estejam associadas à exposição prolongada a poluentes atmosféricos (Samad *et al.*, 2023). O monitoramento da qualidade do ar, portanto, é fundamental para a formulação de políticas públicas eficazes, mitigação de impactos e conscientização da população. Contudo, os elevados custos de instalação e manutenção de estações físicas de monitoramento limitam sua cobertura espacial, especialmente em regiões urbanas de países em desenvolvimento (Vu *et al.*, 2019a).

Nesse contexto, surgem as estações virtuais de monitoramento da qualidade do ar, fundamentadas em técnicas de modelagem e inteligência artificial, como alternativas viáveis para expandir a capacidade de vigilância ambiental sem a necessidade de infraestrutura física intensiva (Samad *et al.*, 2023). Essas soluções digitais são capazes de estimar concentrações de poluentes com base em dados meteorológicos, geográficos e históricos, oferecendo suporte técnico para tomada de decisão e planejamento urbano.

Paralelamente, a formação de engenheiros ambientais demanda estratégias pedagógicas que promovam o desenvolvimento de competências técnicas, reflexivas e colaborativas. Entre essas abordagens, a aprendizagem baseada em projetos (Project-Based Learning – PBL) destaca-se por aproximar os estudantes de situações reais, estimulando a resolução de problemas complexos e o trabalho em equipe (Frank; Lavy; Elata, 2003). As recentes confirmam os efeitos positivos do PBL no desempenho acadêmico, nas habilidades cognitivas superiores e na atitude dos alunos frente à aprendizagem (Zhang; Ma, 2023).

Neste artigo, relata-se a experiência pedagógica desenvolvida na disciplina “Concepção e Gestão de Projetos de Engenharia”, do curso de Engenharia Ambiental do Centro Universitário de Volta Redonda. O projeto propôs que os estudantes participassem ativamente da criação de uma estação virtual de monitoramento da qualidade do ar, simulando todas as etapas de um projeto de engenharia — desde a concepção até a validação. O projeto extrapolou o escopo da disciplina, transformando-se posteriormente em um trabalho de iniciação científica e Trabalho de Conclusão de Curso (TCC).

O objetivo deste artigo é apresentar as etapas do desenvolvimento da estação virtual e os principais resultados obtidos até o momento, destacando o potencial formativo dessa abordagem no contexto da educação em engenharia ambiental.

2 FUNDAMENTAÇÃO TEÓRICA

A poluição atmosférica representa um dos maiores desafios ambientais contemporâneos, sendo responsável por milhões de mortes prematuras no mundo. Estima-se que mais de 7 milhões de pessoas morram anualmente devido à exposição prolongada a partículas inaláveis finas (PM_{2.5}), que penetram profundamente nos pulmões e atingem a corrente sanguínea, causando doenças respiratórias e cardiovasculares (Liu *et al.*, 2019;

Zaman *et al.*, 2021). A Agência Europeia de Meio Ambiente aponta que, somente em 2014, cerca de 428 mil mortes prematuras ocorreram na Europa devido à exposição ao PM_{2.5} (Liu *et al.*, 2019).

O material particulado é classificado conforme seu diâmetro aerodinâmico: partículas com menos de 10 µm (PM₁₀) podem alcançar os brônquios, enquanto as com menos de 2,5 µm (PM_{2.5}) atingem os alvéolos pulmonares (Liu *et al.*, 2019; Zaman *et al.*, 2021). O PM₁₀ é frequentemente utilizado como indicador da qualidade do ar, sendo medido em estações de monitoramento fixas ou estimado por meio de sensores de baixo custo. Esses sensores, como o Nova SDS011, apresentam viabilidade para uso em redes amplas de monitoramento, especialmente quando calibrados com algoritmos de aprendizado de máquina (Liu *et al.*, 2019).

Apesar das vantagens dos sensores de referência, seu alto custo limita a abrangência espacial das medições. Por isso, a modelagem baseada em dados e a construção de estações virtuais tornaram-se alternativas promissoras, utilizando dados meteorológicos, satelitais e históricos para estimar concentrações de poluentes em regiões sem monitoramento físico (Fong *et al.*, 2020; Yu *et al.*, 2022).

As redes neurais artificiais (RNA), sobretudo aquelas baseadas em arquiteturas profundas como LSTM (Long Short-Term Memory), têm se mostrado eficazes na modelagem da qualidade do ar, por sua capacidade de capturar relações não lineares e padrões temporais complexos (Chang *et al.*, 2020). Modelos como redes recorrentes (RNN), regressão por vetores de suporte (SVR) e florestas aleatórias (RF) também são amplamente utilizados na estimativa de concentrações de PM₁₀ e PM_{2.5} (Yu *et al.*, 2022; Zaman *et al.*, 2021).

Estudos recentes demonstram que técnicas de aprendizado profundo e modelos ensemble, como o Deep Ensemble Machine Learning (DEML), superam modelos estatísticos tradicionais, fornecendo maior precisão na previsão de poluentes, inclusive em áreas rurais e suburbanas (Fong *et al.*, 2020; Yu *et al.*, 2022). Transfer learning e métodos de agregação de modelos, por sua vez, auxiliam na melhoria da performance preditiva mesmo em regiões com escassez de dados históricos (Fong *et al.*, 2020).

3 TRABALHOS RELACIONADOS

A previsão da qualidade do ar tem atraído crescente atenção da comunidade científica, sobretudo com o avanço de técnicas de aprendizado de máquina aplicadas à modelagem ambiental. Um dos estudos de maior relevância nesse contexto é o trabalho de Samad *et al.* (2023), que propuseram um modelo para substituir estações físicas de monitoramento por estações virtuais, utilizando algoritmos de aprendizado de máquina como Regressão Linear, k-NN, Árvores de Decisão, Random Forest, XGBoost, e redes neurais artificiais. O estudo obteve alta acurácia preditiva (R^2 superior a 0,9) na estimativa das concentrações de PM_{2.5}, PM₁₀, NO₂, CO e O₃ em Lahore, Paquistão, destacando o potencial dessas tecnologias como alternativas de baixo custo e alta eficiência para monitoramento ambiental contínuo.

Diversos trabalhos têm seguido essa linha de pesquisa. Chang *et al.* (2020) propuseram um modelo agregado de LSTM (ALSTM) que combina dados de diferentes tipos de estações (urbanas, industriais, rurais) para aprimorar a previsão de PM_{2.5} em Taiwan. Essa abordagem multirigim demonstrou significativa melhoria nos índices de erro (RMSE, MAE), superando modelos individuais como SVR e GBTR.

Drewil e Al-Bahadili (2022) desenvolveram um modelo baseado em LSTM otimizado por Algoritmos Genéticos, aplicado à previsão de PM₁₀, PM_{2.5}, CO e NO_x. O foco do trabalho foi melhorar a performance preditiva por meio da otimização de hiperparâmetros, alcançando resultados mais rápidos e precisos que modelos tradicionais de redes neurais e de aprendizado de máquina puro.

Além disso, Czernecki *et al.* (2021) avaliaram a eficácia de diversos algoritmos — como regressão linear, Random Forest, XGBoost e redes neurais — na previsão de curto prazo de PM₁₀ e PM_{2.5} em aglomerações urbanas da Polônia. O XGBoost se destacou como o modelo mais preciso, especialmente em contextos com alta variabilidade sazonal e meteorológica.

Vu *et al.* (2019b) apresentam uma aplicação robusta de técnicas de Inteligência Artificial, especificamente o modelo de Random Forest, para estimar concentrações diárias de material particulado fino (PM_{2.5}) em alta resolução espacial (1 km²) na cidade de Lima, Peru. A pesquisa integra dados de sensoriamento remoto por satélite (AOD-MAIAC), variáveis meteorológicas oriundas de modelos de previsão (ECMWF e WRF-Chem), dados de uso do solo e medições de estações de monitoramento terrestre para treinar e validar o modelo preditivo. Entre as 16 variáveis utilizadas no modelo estão: profundidade óptica de aerossóis (AOD), radiação solar incidente, temperatura, umidade relativa, pressão atmosférica, altura da camada limite planetária (PBL), componentes do vento (u e v), cobertura de nuvens, albedo, NDVI, densidade populacional, elevação, distância das vias principais, percentagem de urbanização e PM_{2.5} simulado por modelo químico. O uso de IA permite superar a escassez de dados locais ao capturar padrões complexos e não lineares entre esses preditores ambientais e os níveis de poluição, fornecendo assim estimativas confiáveis mesmo em regiões com cobertura limitada de sensores. A abordagem mostrou desempenho satisfatório ($R^2 = 0,70$) e gera uma base histórica de dados de qualidade do ar essencial para estudos epidemiológicos e formulação de políticas públicas voltadas à saúde ambiental.

Em contraste com essas abordagens sofisticadas, este estudo adota uma rede neural simples, aplicada a dados reais obtidos pelo Instituto Estadual do Ambiente do Rio de Janeiro (INEA/RJ). O objetivo é validar o conceito de uma Estação Virtual de Monitoramento da Qualidade do Ar no contexto da formação de engenheiros ambientais, priorizando o aprendizado prático e a viabilidade técnica com recursos computacionais acessíveis.

4 REDES NEURAIS ARTIFICIAIS

As redes neurais artificiais são compostas por unidades chamadas neurônios artificiais, que operam a partir de um conjunto de entradas, associadas a pesos, além de um termo de viés e uma função de ativação não linear, conforme representado na Figura 1. Durante o processo de aprendizagem da rede, esses parâmetros — pesos e bias — são sistematicamente ajustados para reduzir a discrepância entre a resposta esperada e a resposta obtida pela rede (Haykin, 2016, 2000).

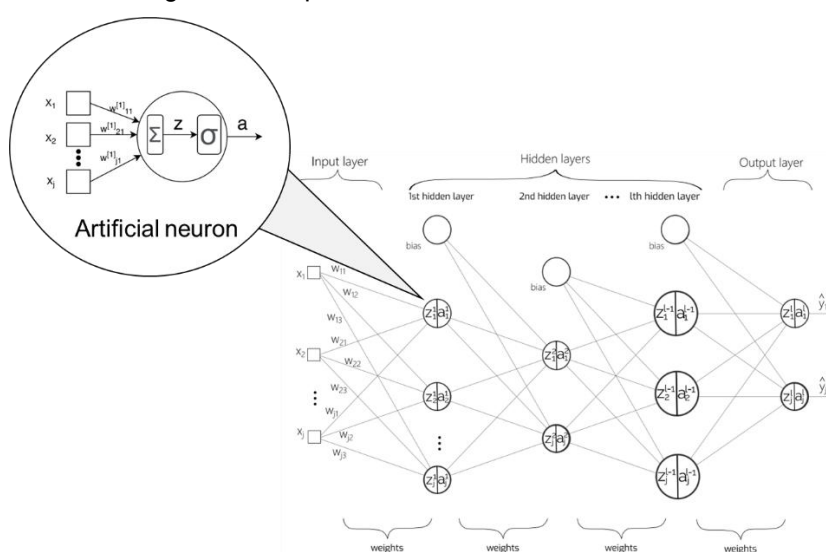
As Redes neurais artificiais (RNA) podem ser expressas matematicamente pelas Equações 1 e 2 (Haykin, 2016, 2000):

$$a_j^{[l]} = \sigma(z_j^{[l]}) \quad (1)$$

$$z_j^{[l]} = \sum_{k=1}^m w_{jk}^{[l]} x_j^{[i]} + b_j^{[l]} \quad (2)$$

onde a é a saída do neurônio, σ representa a função de ativação, w é o valor do peso, x é o vetor que contém as entradas de terminada camada, e b representa a bias. m representa o número de entradas da RNA, o sobrescrito $[l]$ denota a $l^{n-ésima}$ camada, o sobrescrito (i) denota a $i^{n-ésima}$ amostra, o sobrescrito j denota qual é a entrada, e o sobrescrito k denota qual é o neurônio.

Figura 1 – Arquitetura da Rede Neural Artificial.



Fonte: Adaptado de Haykin (2016, 2000).

As Redes Neurais Artificiais (RNAs) podem ser organizadas em diferentes arquiteturas, nas quais são incorporadas funções de ativação e algoritmos de aprendizado. A estrutura de uma RNA é composta por camadas distintas — de entrada, ocultas e de saída —, cada uma com uma quantidade específica de neurônios. O número de neurônios na camada de entrada depende diretamente da quantidade de variáveis de entrada do problema. Já a quantidade de camadas ocultas, bem como o número de neurônios em cada uma delas, pode variar conforme as necessidades da aplicação. Por fim, a dimensão da camada de saída é definida de acordo com o número de saídas que se deseja obter (Haykin, 2016, 2000).

O Perceptron de Múltiplas Camadas (MLP) é uma das arquiteturas de rede neural artificial mais utilizadas na resolução de problemas em diversas áreas do conhecimento (Haykin, 2016, 2000). A flexibilidade dessa estrutura foi significativamente ampliada com o desenvolvimento do algoritmo de retropropagação, que permite à rede aprender e representar, de forma eficaz, a relação entre os dados de entrada e os resultados esperados (Haykin, 2016, 2000).

Durante o processo de aprendizagem, a rede recebe um conjunto de entradas, que são processadas sequencialmente por seus neurônios até atingir a camada de saída, gerando uma resposta. A primeira camada oculta realiza um ajuste nos dados de entrada para que seus valores fiquem entre -1 e 1, promovendo a padronização das escalas de entrada. Esse processo, conhecido como normalização pela média (mean normalization), contribui para o

desempenho da rede ao facilitar o tratamento dos dados, conforme descrito na Equação 3 (Haykin, 2016, 2000).

$$x_{norm} = \frac{x_i - \frac{1}{N} \sum_{i=1}^N x_i}{\max(x) - \min(x)} \quad (3)$$

onde N é o número de amostras disponíveis, e x é a $i^{n-ésima}$ amostra de entrada. O cálculo de $z_j^{[l]}$ na primeira camada é dado pela Equação 4.

$$Z^{[l]} = W^{[l]} X_{norm} + B^{[l]} \quad (4)$$

Em que os símbolos representados por letras maiúsculas são matrizes. A saída da primeira camada é calculada com a função de ativação como mostrada na Equação 5.

$$A^{[l]} = \sigma(Z^{[l]}) \quad (5)$$

Nas etapas subsequentes do processamento em uma Rede Neural Artificial, a saída gerada por uma camada passa a ser utilizada como entrada para a camada seguinte. Esse encadeamento sequencial garante o fluxo contínuo de informações entre as camadas, permitindo o aprendizado e a transformação progressiva dos dados ao longo da rede, conforme representado pela Equação 6.

$$z_j^{[l+1]} = \sum_{k=1}^m w_{jk}^{[l+1]} a_j^{(i)[l]} + b_j^{[l+1]} \quad (6)$$

Na camada de saída, a saída estimada $\hat{y}_j$ é obtida pela Equação 7:

$$\hat{y}_j = a_j^{[l+1]} = z_j^{[l+1]} \quad (7)$$

Ao lidar com uma função linear, a saída produzida pela rede neural é comparada ao valor desejado, gerando um sinal de erro (E). Esse sinal de erro é então retropropagado pela rede, camada por camada, com o objetivo de ajustar os pesos e os bias. Esse procedimento ocorre de forma iterativa até que sejam atendidos determinados critérios de parada. Tais critérios podem incluir um número pré-definido de épocas (iterações), a obtenção de um índice de desempenho específico ou a combinação de ambos. O processo de aprendizado é encerrado assim que esses critérios são atingidos, indicando que a rede alcançou um nível satisfatório de precisão ou desempenho (Haykin, 2016, 2000).

Neste estudo, o algoritmo de treinamento adotado foi o método de Levenberg-Marquardt (LM) (Haykin, 2016, 2000). Esse método busca um equilíbrio entre a estabilidade do método do gradiente e a rapidez da aproximação de Gauss-Newton. A avaliação do desempenho da Rede Neural Artificial é realizada por meio da métrica do erro quadrático médio (E_{MSE}), apresentada na Equação 8, que expressa a média das diferenças elevadas ao quadrado entre os valores previstos pela rede e os valores reais. (Haykin, 2016, 2000).

$$E_{MSE} = \frac{1}{N} \sum_{i=1}^N [(y_i - \hat{y}_i)^2] \quad (8)$$

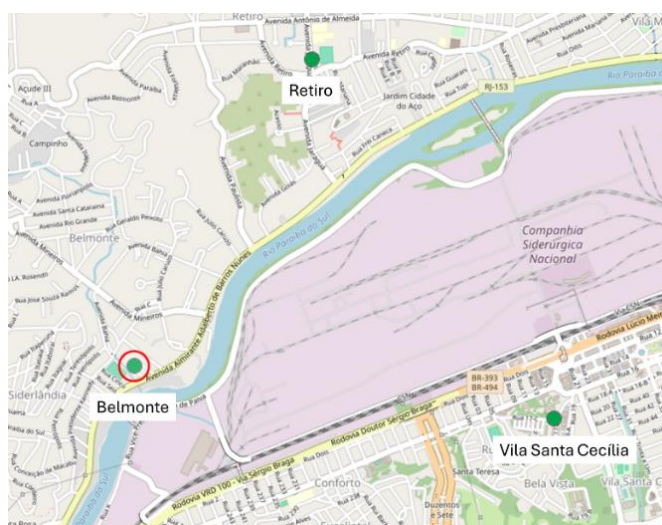
onde N é o número de amostras, y_i representa o valor real da amostra e $\hat{y}_i$ representa o valor estimado pela RNA.

Durante o processo de aprendizado de uma Rede Neural Artificial, o conjunto de dados é segmentado em três subconjuntos distintos (Haykin, 2016, 2000): i) conjunto de treinamento, utilizado para calcular os erros e ajustar os pesos e bias da rede a cada iteração do processo de aprendizagem; ii) conjunto de validação, empregado para reduzir o risco de sobreajuste (overfitting), auxiliando na escolha do melhor modelo; e iii) conjunto de teste, destinado a estimar o erro de generalização da rede, ou seja, sua capacidade de produzir resultados precisos com dados não vistos anteriormente.

5 METODOLOGIA

Neste estudo, foram utilizadas três estações de monitoramento da qualidade do ar localizadas na cidade de Volta Redonda (RJ): Belmonte, Retiro e Vila Santa Cecília, conforme Figura 2. A escolha dessas estações justifica-se pela proximidade geográfica e pela cobertura urbana homogênea, o que permite minimizar efeitos climáticos ou topográficos diferenciados.

Figura 2 – Posição das estações.



Fonte: Adaptada de INEA (2025).

O objetivo principal da modelagem foi utilizar as séries temporais de PM10 obtidas nas estações de Belmonte e Retiro como variáveis preditoras para estimar os valores correspondentes na estação Vila Santa Cecília. Esta abordagem simula a funcionalidade de uma estação virtual, substituindo medições físicas por inferências baseadas em padrões de dados ambientais.

Foram considerados dados diários do período entre 2004 e 2024, extraídos do banco de dados do Instituto Estadual do Ambiente (INEA/RJ). Apenas foram selecionadas amostras com registros simultâneos e completos nas três estações, totalizando 36.315 amostras válidas. Amostras com datas divergentes ou dados ausentes em qualquer uma das estações foram descartadas.

Para fins ilustrativos, as Figuras 3, 4 e 5 apresentam a média móvel a cada 100 amostras das concentrações de PM₁₀ em cada estação, evidenciando a variação temporal ao longo de duas décadas de medições.

Foi implementada uma rede neural artificial simples, composta por uma camada de entrada com dois neurônios (valores normalizados das estações Belmonte e Retiro), uma ou mais camadas ocultas com função de ativação sigmoide, e uma camada de saída com um único neurônio para prever os valores da estação Vila Santa Cecília. Os hiperparâmetros de configuração da RNA utilizada podem ser encontrados no Quadro 1.

Quadro 1: Hiperparâmetros da RNA.

Hiperparâmetros	Valor
Algoritmo de treinamento	Levenberg-Marquardt
Quantidade de camadas	5, sendo 1 camada de entrada 3 camadas ocultas e 1 camada de saídas
Neurônios (por camada)	[2 30 30 30 1]
Função de ativação	Tangente hiperbólica

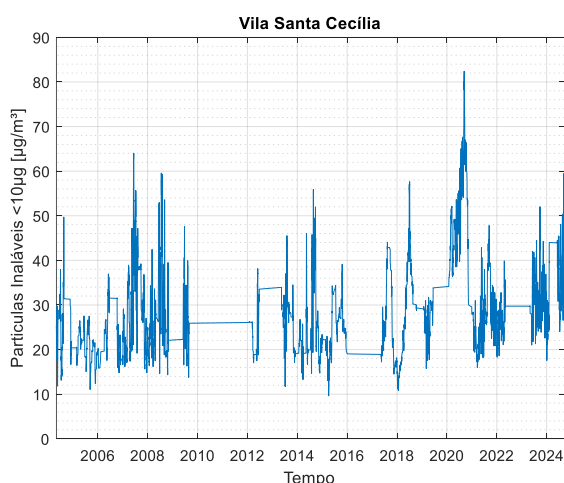
Fonte: Elaborado pelos autores (2025).

A quantidade de camadas ocultas e neurônios por camada foi determinada de forma empírica, buscando o melhor desempenho preditivo conforme testes sucessivos.

Das 36.315 amostras, 12.105 (cerca de 33%) foram selecionadas para o treinamento do modelo, subdivididas em 70% para treinamento, 15% para validação e 15% para teste.

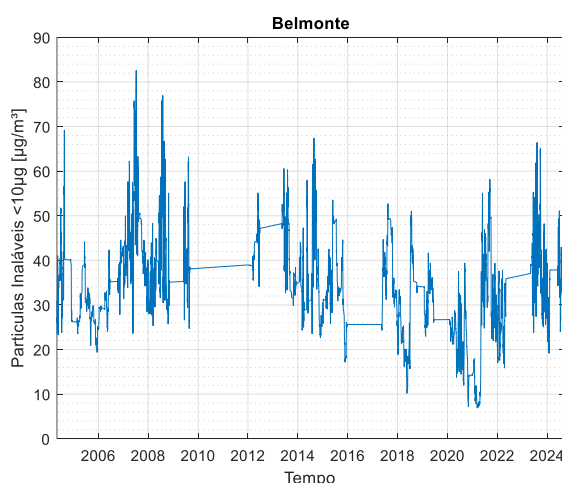
Diferentemente do estudo de Samad *et al.* (2023), que empregou uma variedade de algoritmos de aprendizado supervisionado (como Random Forest, XGBoost e redes neurais profundas) aplicados a diversas estações distribuídas por uma metrópole, este trabalho adota uma abordagem simplificada, com foco em uma rede neural artificial elementar e aplicação em uma área geográfica restrita.

Figura 3 – Medições de PM₁₀ na Vila Santa Cecília.



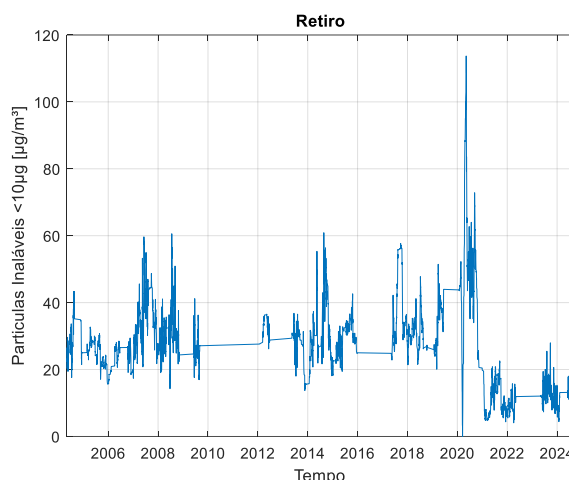
Fonte: Elaborada pelos autores (2025).

Figura 4 – Medições de PM₁₀ no Belmonte.



Fonte: Elaborada pelos autores (2025).

Figura 5 – Medições de PM10 no Retiro.



Fonte: Elaborada pelos autores (2025).

Enquanto Samad *et al.* (2023) enfatizam a substituição completa de estações físicas por modelos preditivos complexos baseados em atributos meteorológicos e múltiplos poluentes, este estudo explora o conceito de prova de conceito em ambiente acadêmico, utilizando exclusivamente dados históricos de PM10 para simular uma estação virtual. A metodologia aqui proposta busca ser uma ferramenta formativa para estudantes de engenharia ambiental, mimetizando as etapas reais de concepção, implementação e validação de projetos científicos e tecnológicos.

A avaliação do desempenho do modelo foi realizada por meio do Mean Relative Squared Error (MRSE), uma métrica que quantifica o erro relativo quadrático médio entre os valores previstos e os reais. A equação utilizada para o cálculo do MRSE está descrita a seguir:

$$MRSE(\%) = \frac{1}{n_o} \sum_{o=1}^{n_o} \left[\sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (\hat{y}_i)^2}} \right] \quad (9)$$

onde n_o representa o número total de saídas, e N a quantidade de dados utilizada. Valores mais próximos de zero indicam melhor desempenho, sinalizando que o modelo é capaz de reproduzir adequadamente os padrões observados nos dados reais

6 RESULTADOS

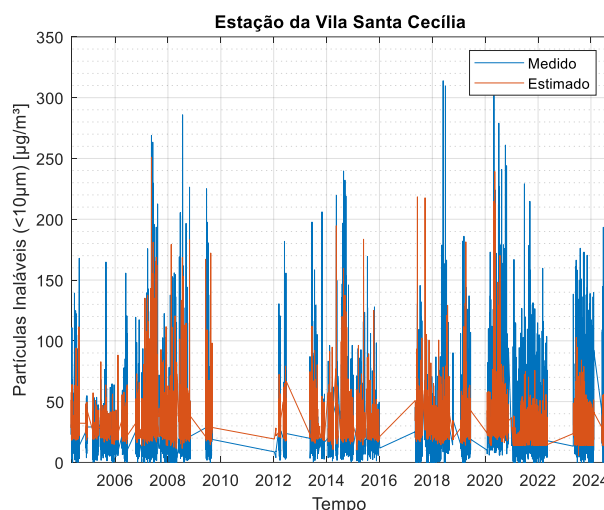
Após o treinamento da rede neural artificial, composta por uma estrutura relativamente simples, mas funcional, foi possível prever, com desempenho moderado, as concentrações de PM10 na estação Vila Santa Cecília, a partir das medições simultâneas realizadas nas estações Belmonte e Retiro.

A rede neural foi configurada com três camadas ocultas, cada uma contendo 30 neurônios com função de ativação sigmoide. A camada de entrada contou com 2 neurônios, correspondentes às variáveis de entrada (valores normalizados de PM10 das estações Belmonte e Retiro), e a camada de saída foi composta por 1 neurônio, responsável pela predição dos valores correspondentes à estação Vila Santa Cecília.

O desempenho do modelo foi avaliado por meio do Mean Relative Squared Error (MRSE), que resultou em 55%, indicando que o modelo foi capaz de capturar tendências gerais, mas apresentou limitações ao representar com precisão os valores extremos.

A Figura 6 exibe a série temporal com as concentrações medidas e estimadas de PM₁₀ ao longo dos anos de 2004 a 2024. A análise gráfica mostra que o modelo consegue seguir os padrões gerais de variação, mas tende a subestimar os picos de concentração, o que é comum em modelos de complexidade limitada sem ajuste fino de parâmetros.

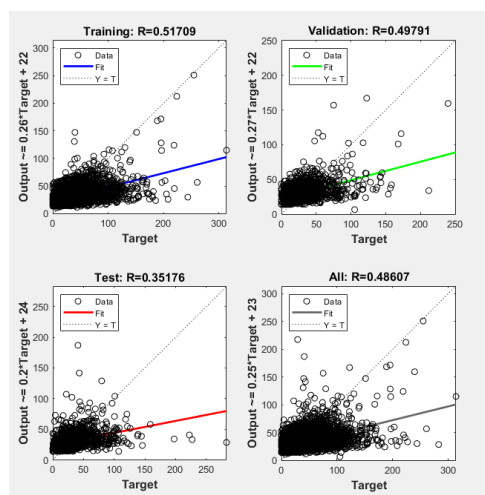
Figura 6 – Comparação entre os valores reais (Medidos) o da Estação Virtual (Estimado).



Fonte: Elaborada pelos autores (2025).

A Figura 7 apresenta os gráficos de dispersão dos valores reais (target) versus os valores previstos (output) para os conjuntos de treinamento, validação, teste e o conjunto total. Os coeficientes de correlação linear (R) obtidos foram: i) Treinamento: $R = 0,51709$; ii) Validação: $R = 0,49791$; iii) Teste: $R = 0,35176$; Conjunto Total: $R = 0,48607$.

Figura 7 – Comparação entre os valores reais (Medidos) o da Estação Virtual (Estimado).



Fonte: Elaborada pelos autores (2025).

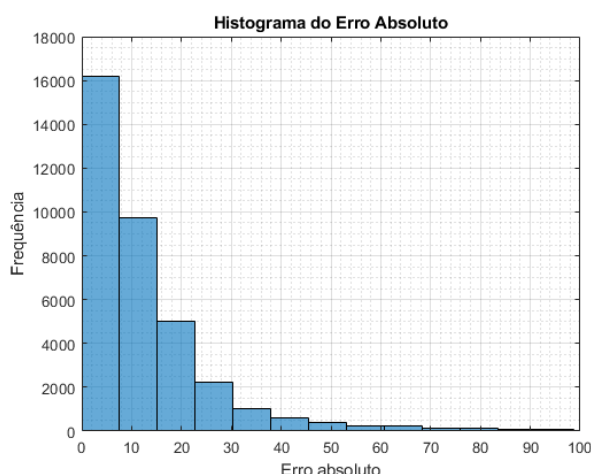
Esses valores indicam uma correlação positiva moderada, com desempenho um pouco melhor nos conjuntos de treinamento e validação do que no conjunto de teste, o que pode ser interpretado não como um sobreajuste, mas como uma limitação na capacidade do modelo

em capturar padrões complexos dos dados. Isso sugere que uma arquitetura neural mais sofisticada ou o uso de variáveis adicionais poderia melhorar significativamente o desempenho preditivo.

Os resultados obtidos mostram que, mesmo com uma arquitetura neural relativamente simples, é possível obter previsões coerentes e com padrão reconhecível. Contudo, o desempenho observado é inferior ao de abordagens mais sofisticadas, como a de Samad *et al.* (2023), que utilizaram múltiplos algoritmos (Random Forest, XGBoost, redes profundas) e conjuntos de dados com múltiplas variáveis meteorológicas, atingindo R^2 superiores a 0,9.

Para quantificar essa limitação, a Figura 8 mostra o histograma do erro absoluto entre os valores medidos e estimados. Nota-se que aproximadamente 44,63% das amostras apresentam erro inferior a $7,6 \mu\text{g}/\text{m}^3$, indicando que o modelo é capaz de estimar com boa precisão a maior parte dos dados da série temporal. No entanto, a cauda longa à direita no histograma sugere a ocorrência de erros mais elevados em menor frequência, com valores que ultrapassam os $90 \mu\text{g}/\text{m}^3$. Esses erros estão concentrados, principalmente, nos períodos de maior concentração de PM_{10} , como observado a partir de 2020.

Figura 8 – Histograma dos erros absolutos.



Fonte: Elaborada pelos autores (2025).

Dessa forma, os resultados indicam que há um campo promissor de pesquisa, e que, com a adoção de modelos mais robustos e variáveis complementares, pode-se alcançar resultados mais competitivos e próximos aos da literatura internacional.

7 CONSIDERAÇÕES FINAIS

Este estudo teve como objetivo avaliar a viabilidade de uma estação virtual de monitoramento da qualidade do ar, por meio da aplicação de uma rede neural artificial simples, utilizando dados reais de PM_{10} obtidos em três estações da cidade de Volta Redonda/RJ. A modelagem utilizou os dados das estações Belmonte e Retiro como entrada para estimar os valores correspondentes à estação Vila Santa Cecília.

Os resultados obtidos, com MRSE de 55% e correlações moderadas (R variando entre 0,35 e 0,52), demonstram que, mesmo com uma arquitetura de rede relativamente básica, é possível identificar padrões relevantes nas séries temporais de poluentes. Isso reforça o potencial das estações virtuais como alternativa viável, especialmente em contextos com restrições de infraestrutura para monitoramento ambiental físico.

Contudo, os limites do modelo empregado — como a ausência de variáveis meteorológicas, a simplicidade da arquitetura e a ausência de técnicas de regularização e otimização de hiperparâmetros — influenciaram diretamente na acurácia preditiva. Em comparação com estudos como o de Samad et al. (2021), que utilizaram algoritmos avançados e múltiplas variáveis, os resultados aqui apresentados destacam a necessidade de abordagens mais sofisticadas para aplicações práticas em larga escala.

Como perspectivas futuras, recomenda-se o uso de redes neurais profundas, algoritmos ensemble, e a integração de variáveis climáticas, topográficas e urbanas. Além disso, o uso deste modelo em contextos pedagógicos mostrou-se eficaz para a formação de engenheiros ambientais, ao proporcionar uma vivência prática de concepção e validação de soluções baseadas em dados reais.

AGRADECIMENTOS

Os autores agradecem ao Centro Universitário de Volta Redonda (UniFOA) pelo apoio institucional e financeiro (93978/17/RPE).

REFERÊNCIAS

CHANG, Yue-Shan; CHIAO, Hsin-Ta; ABIMANNAN, Satheesh; HUANG, Yo-Ping; TSAI, Yi-Ting; LIN, Kuan-Ming. An LSTM-based aggregated model for air pollution forecasting. **Atmospheric Pollution Research**, [s. l.], v. 11, n. 8, p. 1451–1463, ago. 2020. <https://doi.org/10.1016/j.apr.2020.05.015>.

CZERNECKI, Bartosz; MAROSZ, Michał; JEDRUSZKIEWICZ, Joanna. Assessment of Machine Learning Algorithms in Short-term Forecasting of PM10 and PM2.5 Concentrations in Selected Polish Agglomerations. **Aerosol and Air Quality Research**, [s. l.], v. 21, n. 7, p. 200586, 2021. <https://doi.org/10.4209/aaqr.200586>.

DREWIL, Ghufan Isam; AL-BAHADILI, Riyadh Jabbar. Air pollution prediction using LSTM deep learning and metaheuristics algorithms. **Measurement: Sensors**, [s. l.], v. 24, p. 100546, dez. 2022. <https://doi.org/10.1016/j.measen.2022.100546>.

FONG, Iat Hang; LI, Tengyue; FONG, Simon; WONG, Raymond K.; TALLÓN-BALLESTEROS, Antonio J. Predicting concentration levels of air pollutants by transfer learning and recurrent neural network. **Knowledge-Based Systems**, [s. l.], v. 192, p. 105622, mar. 2020. <https://doi.org/10.1016/j.knosys.2020.105622>.

FRANK, Moti; LAVY, Ilana; ELATA, David. Implementing the Project-Based Learning Approach in an Academic Engineering Course. **International Journal of Technology and Design Education**, [s. l.], v. 13, n. 3, p. 273–288, out. 2003. <https://doi.org/10.1023/A:1026192113732>.

HAYKIN, Simon S. **Neural networks and learning machines**. Third edition, Indian subcontinent adaptation. Delhi Chennai: Pearson, 2016.

HAYKIN, Simon S. **Redes Neurais**. 2. ed. [S. l.]: Bookman, 2000.

INEA. AQMIS. 2025. **SIGQAr - Sistema Integrado de Gestão da Qualidade do Ar**. Disponível em: <https://ei.weblakes.com/INEAPublico/ViewerMap>. Acesso em: 19 jul. 2025.

LIU, Hai-Ying; SCHNEIDER, Philipp; HAUGEN, Rolf; VOGT, Matthias. Performance Assessment of a Low-Cost PM2.5 Sensor for a near Four-Month Period in Oslo, Norway. **Atmosphere**, [s. l.], v. 10, n. 2, p. 41, 22 jan. 2019. <https://doi.org/10.3390/atmos10020041>.

SAMAD, A.; GARUDA, S.; VOGT, U.; YANG, B. Air pollution prediction using machine learning techniques – An approach to replace existing monitoring stations with virtual monitoring stations. **Atmospheric Environment**, [s. l.], v. 310, p. 119987, out. 2023. <https://doi.org/10.1016/j.atmosenv.2023.119987>.

VU, Bryan N.; SÁNCHEZ, Odón; BI, Jianzhao; XIAO, Qingyang; HANSEL, Nadia N.; CHECKLEY, William; GONZALES, Gustavo F.; STEENLAND, Kyle; LIU, Yang. Developing an Advanced PM2.5 Exposure Model in Lima, Peru. **Remote Sensing**, [s. l.], v. 11, n. 6, p. 641, 16 mar. 2019a. <https://doi.org/10.3390/rs11060641>.

VU, Bryan N.; SÁNCHEZ, Odón; BI, Jianzhao; XIAO, Qingyang; HANSEL, Nadia N.; CHECKLEY, William; GONZALES, Gustavo F.; STEENLAND, Kyle; LIU, Yang. Developing an Advanced PM2.5 Exposure Model in Lima, Peru. **Remote Sensing**, [s. l.], v. 11, n. 6, p. 641, 16 mar. 2019b. <https://doi.org/10.3390/rs11060641>.

YU, Wenhua; LI, Shanshan; YE, Tingting; XU, Rongbin; SONG, Jiangning; GUO, Yuming. Deep Ensemble Machine Learning Framework for the Estimation of PM2.5 Concentrations. **Environmental Health Perspectives**, [s. l.], v. 130, n. 3, mar. 2022. DOI 10.1289/ehp9752. Disponível em: <https://ehp.niehs.nih.gov/doi/10.1289/EHP9752>. Acesso em: 1 maio 2025.

ZAMAN, Nurul Amalin Fatihah Kamarul; KANNIAH, Kasturi Devi; KASKAOUTIS, Dimitris G.; LATIF, Mohd Talib. Evaluation of Machine Learning Models for Estimating PM2.5 Concentrations across Malaysia. **Applied Sciences**, [s. l.], v. 11, n. 16, p. 7326, 9 ago. 2021. <https://doi.org/10.3390/app11167326>.

ZHANG, Lu; MA, Yan. A study of the impact of project-based learning on student learning effects: a meta-analysis study. **Frontiers in Psychology**, [s. l.], v. 14, p. 1202728, 17 jul. 2023. <https://doi.org/10.3389/fpsyg.2023.1202728>.

DEVELOPMENT OF A VIRTUAL AIR QUALITY MONITORING STATION USING AN ARTIFICIAL NEURAL NETWORK: A DIDACTIC APPLICATION WITH REAL DATA

Abstract: Air pollution is responsible for millions of premature deaths worldwide, making continuous air quality monitoring essential. However, the high costs of installing and maintaining physical monitoring stations limit their spatial coverage, especially in developing countries. This paper presents the results of a pedagogical experience conducted in the Engineering Project Design and Management course of the Environmental Engineering program at the Centro Universitário de Volta Redonda. The project aimed to develop a virtual monitoring station using artificial neural networks. Historical PM10 data (2004–2024) from three stations in Volta Redonda/RJ (Belmonte, Retiro, and Vila Santa Cecília) were used. The network, with three hidden layers of 30 neurons each, achieved an MRSE of 55% and a moderate correlation ($R = 0.486$), demonstrating the feasibility of the approach for educational purposes and highlighting the need for more robust models for operational applications.

Keywords: Air monitoring, Virtual station, Artificial neural network.

