

## Relação entre hospitalizações por doenças respiratórias e poluição atmosférica por meio do uso da rede neural Perceptron de Múltiplas Camadas

## Relationship between hospitalizations for respiratory diseases and air pollution through the use of the Multi-Layer Perceptron neural network

### RESUMO

O objetivo foi realizar uma Revisão Bibliográfica Sistema (RBS) relacionada à poluição atmosférica, saúde humana e redes neurais artificiais. Em seguida, a rede neural MLP (do inglês - *Multi-Layer Perceptron*) foi utilizada para prever o número de internações por doenças respiratórias causadas pelo MP10 na cidade de São Paulo. A RBS foi realizada em 6 bases de dados, utilizando as palavras-chave “*air pollutants*”, “*neural network*” e “*health*”, sendo os anos de publicações restritos entre 2009 e janeiro de 2020. A MLP foi implementada utilizando a linguagem de programação Python versão 3.7 com o auxílio da biblioteca *Scikit-learn*. Foi possível concluir que apenas uma pequena parcela dos artigos encontrados foi selecionada para a RBS, mostrando ser um tema ainda pouco explorado no meio acadêmico. As previsões encontradas pela MLP apresentam valores próximos aos observados. Além disso, a MLP se mostrou uma ferramenta eficaz para esse tipo de previsão, podendo ser explorado em trabalhos futuros.

**PALAVRAS-CHAVE:** Inteligência artificial. Saúde. Emissão de poluentes.

### ABSTRACT

The objective was to carry out a Systemic Bibliographic Review (SBR) related to atmospheric, human health and artificial neural networks. Then, the MLP neural network (in English - *Multi-Layer Perceptron*) was used to predict the number of hospitalizations for respiratory diseases caused by MP<sub>10</sub> in the city of São Paulo. The RBS was carried out in 6 databases, using the keywords “*air pollutants*”, “*neural network*” and “*health*”, with the years of publications being restricted between 2009 and January 2020. MLP was implemented using the language Python programming version 3.7 with the aid of the *Scikit-learn* library. It was possible to conclude that only a small portion of the articles found was obtained for RBS, showing that it is a topic that is still little explored in academia. Those found by MLP present values to those observed. In addition, an MLP has proven to be an effective tool for this type of forecast, and can be explored in future work.

**KEYWORDS:** Artificial intelligence. Health. Pollutant emissions.

João Luiz Miranda Meyer  
[Joao\\_lmm@hotmail.com](mailto:Joao_lmm@hotmail.com)  
Universidade Tecnológica Federal do Paraná,  
Ponta Grossa, Paraná, Brasil

Yara de Souza Tadano  
[yaratadano@utfpr.edu.br](mailto:yaratadano@utfpr.edu.br)  
Universidade Tecnológica Federal do Paraná,  
Ponta Grossa, Paraná, Brasil

**Recebido:** 19 ago. 2020.

**Aprovado:** 01 out. 2020.

**Direito autorial:** Este trabalho está licenciado sob os termos da Licença Creative Commons-Atribuição 4.0 Internacional.



## INTRODUÇÃO

O alto número de poluentes lançados diariamente na atmosfera, por meio de veículos que utilizam combustíveis fósseis ou indústrias que não possuem um controle efetivo de seus resíduos atmosféricos, pode impactar diretamente na saúde humana. Segundo Arbex et al. (2012), as principais doenças causadas por poluentes atmosféricos são câncer de pulmão, asma e doença pulmonar obstrutiva crônica. Isso é intensificado em grandes centros urbanos, onde há uma maior concentração dessas fontes emissoras, sendo esse problema agravado para a população local.

Vários podem ser os poluentes, segundo o Ministério do Meio Ambiente (2020), os mais comuns são CO, CO<sub>2</sub>, CH<sub>4</sub>, SO<sub>2</sub>, Hidrocarbonetos (HC), NO e NO<sub>2</sub>. Além desses, há também os poluentes que são constituídos por material sólido ou líquido de pequenas dimensões, que acabam ficando suspensos no ar e entrando pelas vias aéreas para o pulmão, sendo denominados de material particulado (MP). Alguns exemplos de MP são poeiras, fuligens originadas da queima de combustíveis fósseis, emissões decorrentes da pavimentação de vias e emissões da amônia na agricultura. Suas concentrações servem como indicadores para a qualidade do ar; quanto maior a concentração, pior para a saúde humana. O foco deste artigo ficará relacionado ao impacto causado especificamente pelo MP<sub>10</sub>, ou seja, partículas com diâmetro aerodinâmico menor que 10 µm.

Um indicativo para relacionar poluição atmosférica e saúde humana pode ser por meio do número de internação hospitalares relacionadas às doenças respiratórias. É possível prever o número de internações com o uso de Redes Neurais Artificiais (RNA), que utiliza dados como parâmetros e identifica suas relações, obtendo-se previsões ou então classificações. O tipo de RNA a ser usado dependerá do problema, neste artigo será utilizado a rede Perceptron de Múltiplas Camadas (MLP - do inglês *Multi-Layer Perceptron*).

O objetivo deste artigo é realizar uma Revisão Bibliográfica Sistema (RBS) referente ao tema e utilizar uma MLP para prever o número de internações por doenças respiratórias causadas pelo MP<sub>10</sub> na cidade de São Paulo.

## MATERIAL E MÉTODOS

As pesquisas de artigos, para a parte da RBS, foram realizadas em 6 bases de dados (*IEEE, PubMed, Science Direct, ACM digital Library, Springer Link, Scopus*). As palavras-chave utilizadas foram “*air pollutants*”, “*neural network*” e “*health*”, com o conectivo “*AND*” sendo usado entre elas. Os anos de publicações dos artigos ficou entre 2009 e janeiro de 2020. Para os resultados encontrados em cada banco de dados, foi necessário realizar uma pesquisa manual dos artigos relacionados à emissão de poluentes e seus danos à saúde, podendo ser por meio do número de internações hospitalares, índice de qualidade do ar ou algum outro parâmetro que permitisse analisar a influência dos poluentes na saúde humana. Buscou-se artigos que utilizaram alguma RNA como meio para encontrar a relação entre saúde e poluição do ar.

Durante a pesquisa manual, foram selecionados artigos que remetessem ao tema pelo seu título, sendo realizada a leitura dos seus resumos, para que fosse possível ter certeza sobre o tema abordado em cada artigo. Para obter um

panorama das publicações no tema, informações específicas foram observadas: RNA usada, principais variáveis e outputs.

Após o conhecimento e aprofundamento no tema, a RNA mais clássica e usada no tema abordado (MLP) foi implementada utilizando a linguagem de programação Python versão 3.7. Além disso, usou-se a biblioteca *Scikit-learn*, a qual possibilitou a implementação da rede. Utilizou-se a ferramenta MLP *Regressor*, responsável pela criação propriamente da rede. Para tanto, os parâmetros utilizados foram:

- a) Quantidade de camadas escondidas: 1;
- b) Quantidade de neurônios na camada escondida: 5;
- c) Número máximo de iterações: 200;
- d) Taxa de aprendizagem: 0,001;
- e) Função de ativação: Logística;
- f) Valor do erro permitido para interromper o treinamento:  $10^{-4}$ ;
- g) Otimizador dos pesos: *Broyden-Fletcher-Goldfarb-Shanno algorithm* (BFGS).

O otimizador de pesos BFGS foi escolhido devido à melhor convergência dos pesos quando comparados a outros. O BFGS consegue convergir melhor neste caso por usar uma quantidade de dados relativamente pequena. Os dados foram divididos em 70% treinamento, 15% validação e 15% teste. Para avaliar a qualidade dos resultados foram utilizados os valores de Erro Quadrático Médio (MSE - do inglês *Mean Square Error*) e Erro Médio Absoluto (MAE - do inglês *Mean Absolute Error*), apresentados nas equações (1) e (2):

$$MSE = \frac{1}{N_{td}} \sum_{t=1}^N (rt - yt)^2 \quad (1)$$

$$MAE = \frac{1}{N_{td}} \sum_{t=1}^N |rt - yt| \quad (2)$$

Sendo  $rt$  o valor real observado,  $yt$  o output da rede e  $N_{td}$  o número total de dados utilizados na fase de teste.

O tipo de validação utilizada para este trabalho foi *k-fold*. Os dados de treinamento e validação foram divididos em 5, variando para que fosse possível analisar qual possuía um melhor treinamento da rede.

Os dados de poluição e meteorológicos foram obtidos pela Companhia Ambiental do Estado de São Paulo (CETESB) (2010), contendo um total de 1068 dados, que datam de 1º de janeiro de 2014 até 31 de dezembro de 2016. São Paulo é, de acordo com KACHBA et al. (2020) o maior centro financeiro e comercial da América do Sul, possuindo, segundo o IBGE (2019), 12,176,866 habitantes. A cidade possui mais de 8 milhões de veículos circulando diariamente e, segundo a Organização Mundial da Saúde (OMS) (2017), o dobro da concentração de poluentes para os seus padrões. Além disso, segundo Araujo et al. (2020), a cidade de São Paulo possui uma área total de 1521,11 km<sup>2</sup>, sendo 968.32 km<sup>2</sup> de área urbana.

As variáveis utilizadas foram: concentração do MP<sub>10</sub>, temperatura ambiente média, umidade relativa do ar, dia da semana e se o dia é feriado ou não. Segundo

Tadano, Siqueira e Alves (2016), mesmo não sendo variáveis climáticas, saber o dia da semana e de feriados influencia no número de internações, por isso foi considerado. Os dados de internação foram obtidos pelo Sistema Nacional de Saúde (DATASUS) (2020). Dias que possuíam dados faltantes foram excluídos para evitar problemas na hora de serem usados.

Foi necessário realizar uma normalização dos dados para que não ocorresse interferência devido à diferença entre os valores de uma variável e outra, sendo normalizados entre +1 e -1. Após o treinamento da rede e obtenção dos outputs, foi necessário desnormalizar os resultados para que fossem testados com os valores reais e obtenção do MSE e MAE.

## RESULTADOS E DISCUSSÃO

Durante a pesquisa dos artigos da RBS foram encontrados, somando todas as bases de dados, um total de 2984 artigos, porém, no final foram selecionados apenas 21 artigos que de alguma forma relacionavam os três temas (poluição atmosférica, saúde humana e RNA).

Foi possível observar, por meio da RBS, que as variáveis temperatura, umidade relativa do ar, concentração de poluentes e direção/velocidade do vento são as mais usadas pelos pesquisadores. É importante salientar que não existe uma relação direta quanto maior o número de variáveis, melhor o resultado. Isso acaba gerando a necessidade de seleção de variáveis, baseado no conhecimento físico do problema, gerando um melhor desempenho da rede. Além disso, a RBS ainda mostrou uma predominância no uso da MLP como a RNA mais usada pelos pesquisadores, contabilizando 8 artigos dentre os 21 selecionados, e um foco na previsão de internações por doenças respiratórias, 9 artigos, e na identificação do índice de qualidade do ar para a saúde humana, 6 artigos.

Como estudo de caso para aplicação da MLP, considerou-se a cidade de São Paulo por ser a maior cidade brasileira e ter uma alta concentração de poluentes atmosféricos, segundo Kachba et al. (2020).

A tabela 1 mostra os resultados do MSE e MAE obtidos das 5 diferentes divisões dos dados de treinamento e validação.

Tabela 1 – MSE e MAE para cada teste.

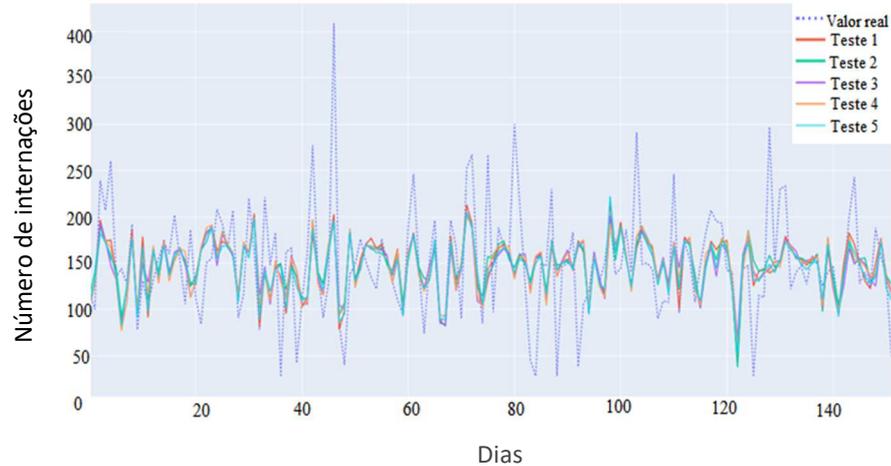
Teste	MSE	MAE
Teste 1	2606.796	36.714
Teste 2	2674.037	37.667
Teste 3	2730.485	38.685
Teste 4	2717.111	38.373
Teste 5	2589.953	36.621

Fonte: Autoria própria

Os menores valores, tanto de MSE quanto do MAE, são do teste 5, sendo esse o melhor resultado. A figura 1 mostra os valores reais (em tracejado) junto

com os 5 testes realizados pela MLP. Quanto mais sobrepostos estiverem os gráficos dos valores calculados com o real, melhor o resultado.

Figura 1 – Número de internações hospitalares observadas versus estimadas.

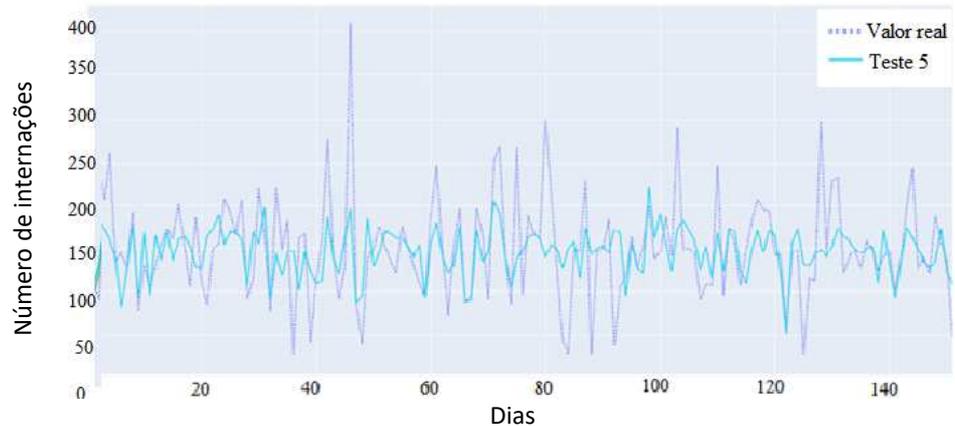


Fonte: Autoria própria.

Como é possível observar pela figura 1, os valores obtidos pela MLP possuem uma tendência de seguir os valores reais, ficando distantes apenas nos picos.

A figura 2 compara os valores reais com os valores ajustados pelo teste 5, que foi o que obteve melhor desempenho.

Figura 2 - Número de internações hospitalares observadas versus estimadas pelo teste 5.



Fonte: Autoria própria.

## CONCLUSÃO

A RBS mostrou ao total 2984 resultados, sendo selecionados apenas 21 artigos, o que indica uma área de conhecimento ainda pouco explorada no ambiente acadêmico. Houve uma predominância no uso da MLP como rede neural, sendo usado principalmente na tentativa de previsão de internações por doenças respiratórias e no índice de qualidade do ar para a saúde humana. Além disso, é importante notar que as variáveis mais usadas pelos pesquisadores foram temperatura, umidade relativa do ar e concentração de poluentes, sendo as mesmas utilizadas neste trabalho. Os resultados obtidos pela MLP mostraram uma

tendência de aproximação com os dados observados, sendo assim, a utilização da MLP para esse tipo de previsão é válida. Para trabalhos futuros há a possibilidade do uso de mais variáveis de entrada e o uso de outros tipos de redes neurais.

### AGRADECIMENTOS

Agradeço ao CNPq (Conselho Nacional de Desenvolvimento Científico e Tecnológico) pelo apoio financeiro fornecido neste ano em que participei deste projeto por meio do Programa de Bolsas de Iniciação em Desenvolvimento Tecnológico e Inovação (PIBITI).

### REFERÊNCIAS

ARAUJO, L. N.; BELOTTI, J. T.; ALVES, T. A.; TADANO, Y. de S., & SIQUEIRA, H. *Ensemble method based on Artificial Neural Networks to estimate air pollution health risks. Environmental Modelling & Software*. v. 123 p. 1-11. 2020

ARBEX, M.A.; SANTOS, U.P.; MARTINS, C.; SALDIVA, P.H.N.; PEREIRA, A.A.; BRAGA, A.L.F. A poluição do ar e o sistema respiratório. *Jornal Brasileiro Pneumologia*. v. 38(5), p. 643-655. 2012.

BRASIL. Ministério do Meio Ambiente. Poluentes Atmosféricos. 2020. Disponível em: <https://www.mma.gov.br/cidades-sustentaveis/qualidade-do-ar/poluentes-atmosf%C3%A9ricos.html>. Acesso em: 10 jun. 2020.

CETESB - Companhia Ambiental do Estado de São Paulo. 2010. Disponível em: <https://cetesb.sp.gov.br/ar/publicacoes-relatorios/>. Acesso em: 10 jun. 2020.

DATASUS - Departamento de Informática do Sistema Único de Saúde. 2020. <http://www2.datasus.gov.br/DATASUS/%20index.php?area=02>. Acesso em: 10 jun. 2020.

IBGE - Instituto Brasileiro de Geografia e Estatística. 2019. Disponível em: <https://www.ibge.gov.br/estatistica-s-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?%t%resultados>. Acesso em: 10 jun. 2020.

KACHBA, Y.; CHIROLI, D. M. de G.; BELOTTI, T. J.; ALVES, T. A.; TADANO, Y.S. & SIQUEIRA, H. *Artificial Neural Networks to Estimate the Influence of Vehicular Emission Variables on Morbidity and Mortality in the Largest Metropolis in South America. Sustainability*, v. 12(7), p. 1-15, mar. 2020.

OMS – Organização Mundial da Saúde. Evolution of WHO Air Quality Guidelines: WHO Regional Office for Europe. 2017.

TADANO, Y.S.; SIQUEIRA, H.V.; ALVES, T.A. *Unorganized machines to predict hospital admissions for respiratory diseases. IEEE LATIN AMERICAN CONFERENCE. anais...* p. 1–6. 2016.