

UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ

JOÃO LUIZ MIRANDA MEYER

**UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA PREVISÃO DE
INTERAÇÕES CAUSADAS PELA POLUIÇÃO ATMOSFÉRICA**

PONTA GROSSA

2022

JOÃO LUIZ MIRANDA MEYER

**UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA PREVISÃO DE
INTERNAÇÕES CAUSADAS PELA POLUIÇÃO ATMOSFÉRICA**

**Use of Artificial Neural Networks for Predicting Hospitalization Caused by Air
Pollution**

Trabalho de conclusão de curso de graduação
apresentada como requisito para obtenção do título
de Bacharel em Engenharia Química da Universidade
Tecnológica Federal do Paraná (UTFPR).
Orientador(a): Profa. Dra. Yara de Souza Tadano.

PONTA GROSSA

2022



[4.0 Internacional](https://creativecommons.org/licenses/by-nc-sa/4.0/)

Esta licença permite remixe, adaptação e criação a partir do trabalho, para fins não comerciais, desde que sejam atribuídos créditos ao(s) autor(es) e que licenciem as novas criações sob termos idênticos. Conteúdos elaborados por terceiros, citados e referenciados nesta obra não são cobertos pela licença.

JOÃO LUIZ MIRANDA MEYER

**UTILIZAÇÃO DE REDES NEURAS ARTIFICIAIS PARA PREVISÃO DE
INTERAÇÕES CAUSADAS PELA POLUIÇÃO ATMOSFÉRICA**

Trabalho de Conclusão de Curso de Graduação
apresentado como requisito para obtenção do título
de Bacharel em Engenharia Química da Universidade
Tecnológica Federal do Paraná (UTFPR).

Data de aprovação: 18 de novembro de 2022

Yara de Souza Tadano
Doutorado em Engenharia Mecânica
Universidade Tecnológica Federal do Paraná

Hugo Valadares Siqueira
Doutorado em Engenharia Elétrica
Universidade Tecnológica Federal do Paraná

Elis Regina Duarte
Doutorado em Engenharia Química
Universidade Tecnológica Federal do Paraná

PONTA GROSSA

2022

AGRADECIMENTOS

Gostaria primeiramente agradecer aos meus pais, Wagner de Carvalho Meyer e Mariângela Miranda Meyer, por todo apoio dado durante a trajetória da vida. Espero um dia poder retribuir da mesma forma. Sempre lembrarei tudo que vocês fizeram por mim.

A minha noiva Fernanda Silva Ribas, por todos os momentos que estivemos juntos e pelo apoio dado sempre que precisei. Obrigado pelo carinho e consideração que você sempre teve por mim.

Agradeço a minha orientadora Prof.(a) Dr.(a) Yara de Souza Tadano, pela sabedoria com que me guiou nesta trajetória, pelas oportunidades e confiança que sempre depositou em mim.

Aos meus amigos que fiz durante essa época de faculdade, que sempre estiveram presentes nos momentos bons e ruins. Em especial Felipe Nishihara, João Eduardo Levandoski, Leonardo Barboza, Luciano Martins e Pedro Afonso Maldonado. Sempre levarei comigo os aprendizados e momentos bons que vivemos juntos.

Agradeço a todos os professores que fizeram parte da minha graduação, pelos aprendizados e experiências vividas.

Aos meus colegas de sala.

A Secretaria do Curso, pela cooperação.

Ao CNPq pelo fomento dado durante minhas pesquisas de iniciações científicas.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

RESUMO

A poluição atmosférica é considerada uma das principais causas de problemas respiratórios para a população urbana, devido ao alto volume de fontes poluidoras, como os automóveis e indústrias. Estudos na literatura frequentemente avaliam o impacto dos poluentes na saúde por meio de modelos de regressão estatística, porém recentemente, o uso de redes neurais artificiais vem se destacando, devido à sua eficiência em problemas de mapeamento. O objetivo deste trabalho foi realizar uma comparação entre duas arquiteturas de redes neurais, a MLP (do inglês - *Multilayer Perceptron*) e ELM (do inglês - *Extreme Learning Machine*), para a previsão do número de internações por doenças respiratórias (J00-J99 da Classificação Internacional de Doenças - CID-10) causadas pelo material particulado de diâmetro aerodinâmico menor ou igual à 10 μm . Como entrada, foram consideradas ainda variáveis meteorológicas (temperatura e umidade). Como estudo de caso considerou-se a cidade de São Paulo. Ambas as redes neurais foram implementadas utilizando a linguagem de programação Python versão 3.7 com o auxílio das bibliotecas *Scikit-learn* e *Numpy*. Os efeitos da poluição na saúde podem ocorrer algum tempo após a exposição. Desta forma, é essencial modelar os efeitos na saúde no dia da exposição à poluição (*lag0*) até sete dias após a exposição (*lag7*). As métricas utilizadas para avaliação das redes foram o Erro Quadrático Médio (MSE - do inglês *Mean Square Error*) e Erro Médio Absoluto (MAE - do inglês *Mean Absolute Error*). Os melhores resultados apresentados pelo MAE para a MLP testada foi para o *lag 0*, apresentando valor de 34,38, enquanto para a ELM o melhor resultado foi para o *lag 7*, apresentando um valor de 40,35. Para o caso do MSE, a MLP apresentou melhor resultado para o *lag 0*, apresentando o valor de 1.820,18, enquanto para a ELM, o melhor resultado foi para o *lag 2*, com o valor de 2.585,73. As redes neurais mostraram ser uma ferramenta de previsão robusta, sendo que a MLP teve uma capacidade melhor de realizar previsões. Foi possível observar uma limitação da capacidade de generalização das redes, sendo justificável pela complexidade do problema e por depender de outras variáveis (idade, gênero, condição social, entre outras) que não foram consideradas no trabalho.

Palavras-chave: ELM; MLP; Previsão; Material Particulado; Python.

ABSTRACT

Atmospheric pollution is considered one of the main causes of respiratory problems for the urban population, due to the high volume of polluting sources, such as automobiles and industries. Studies in the literature often assess the impact of pollutants on health through statistical regression models, but recently, the use of artificial neural networks has been highlighted, due to its efficiency in mapping problems. The objective of this work was to compare two neural network architectures, the MLP (Multilayer Perceptron) and ELM (Extreme Learning Machine), to predict the number of hospitalizations due to respiratory diseases (J00-J99 of International Classification of Diseases – CID-10) caused by particulate matter with an aerodynamic diameter less than or equal to 10 μm . As input, meteorological variables (temperature and humidity) were also considered. As a case study, the city of São Paulo was considered. Both neural networks were implemented using the programming language Python version 3.7 with the help of the Scikit-learn and Numpy libraries. The health effects of pollution may occur sometime after exposure. Therefore, it is essential to model health effects on the day of pollution exposure (lag0) to seven days after exposure (lag7). The metrics used to evaluate the networks were Mean Square Error (MSE) and Mean Absolute Error (MAE). The best results presented by MAE for the tested MLP were for lag 0, presenting a value of 34.38, while for ELM the best result was for lag 7, presenting a value of 40.35. For the case of the MSE, the MLP presented the best result for lag 0, with a value of 1,820.18, while for the ELM, the best result was for lag 2, with a value of 2,585.73. The neural networks proved to be a robust forecasting tool, and the MLP had a better ability to perform forecasts. It was possible to observe a limitation of the generalization capacity of the networks, which is justifiable by the complexity of the problem and because it depends on other variables (age, gender, social status, among others) that were not considered in the work.

Keywords: ELM; MLP; Prediction; Particulate Matter; Python.

LISTA DE FIGURAS

Figura 1 – Regiões de deposição do MP para diferentes diâmetros	16
Figura 2 – Representação dos cálculos da etapa de propagação	18
Figura 3 – Arquitetura de uma rede MLP.....	20
Figura 4 – Representação do plano de erro em uma situação de dois pesos sinápticos (w_1 e w_2).....	21
Figura 5 – Arquitetura de uma ELM	22
Figura 6 – Modelo Sub-ajustado.	24
Figura 7 – Modelo Sobre-ajustado.....	25
Figura 8 – Modelo ajustado.	25
Figura 6 – Esquema da Validação Cruzada do tipo <i>k-fold</i>	27
Figura 7 – Esquema das variáveis dentro da rede	29
Figura 8 – Mapa da estação de monitoramento do bairro Cerqueira César, em São Paulo.....	30
Figura 9 – Pesquisa por base de dados.	34

LISTA DE GRÁFICOS

Gráfico 1 – Frequência de RNA usadas nos artigos da RB	35
Gráfico 2 – Frequência de variáveis de entrada usadas nos artigos da RBS. ...	35
Gráfico 3 - Frequência de variáveis de saída usadas nos artigos da RBS.....	36
Gráfico 4 – Valores observados versus valores ajustados para a MLP	40
Gráfico 5 – Valores observados versus valores ajustados para a ELM.	40

LISTA DE QUADROS

Quadro 1 – Variáveis qualitativas e seus respectivos valores.....	29
--	-----------

LISTA DE TABELAS

Tabela 1 - MAE e MSE para MLP com ambas as funções de ativação.	37
Tabela 2 - MAE e MSE para ELM com ambas as funções de ativação.	38

LISTA DE EQUAÇÕES

Equação 1 – Somatório entre a multiplicação dos pesos e seus respectivos pesos.....	28
Equação 2 – Multiplicação entre o vetor X e a matriz W	32
Equação 3 – Sistema Linear utilizando a inversa de Moore–Penrose.....	32
Equação 4 - <i>Mean Square Error</i>.....	42
Equação 5 - <i>Mean Absolute Error</i>.....	42

SUMÁRIO

1.	INTRODUÇÃO	13
1.1.	OBJETIVO GERAL.....	13
1.2.	OBJETIVO ESPECÍFICO.....	14
2.	REFERENCIAL TEÓRICO.....	15
2.1.	Material Particulado	15
2.2.	Redes Neurais Artificiais	16
2.2.1.	Perceptron de Múltiplas Camadas	17
2.2.2.	Máquina de Aprendizado Extremo.....	21
2.2.3.	Separação dos Dados	23
2.2.4.	Treinamento sobre-ajustado e sub-ajustado	24
2.2.5.	Validação Cruzada.....	26
3.	METODOLOGIA	28
3.1.	Trabalhos Relacionados	28
3.2.	Banco de dados.....	28
3.3.	Implementação computacional da MLP	31
3.4.	Implementação computacional da ELM	32
4.	RESULTADOS.....	34
4.1.	Trabalhos Relacionados	34
4.2.	Desempenho das redes	36
5.	CONCLUSÃO	42
	REFERÊNCIAS.....	43

1. INTRODUÇÃO

Segundo a Organização Mundial da Saúde (OMS, 2020), por volta de 7 milhões de pessoas morrem anualmente por alguma doença originada pela poluição atmosférica. Um dos poluentes que impacta diretamente na saúde humana é o material particulado (MP) (POLEZER et al., 2018). Vários estudos já apontaram esta relação, como por exemplo o trabalho de Kassomenos et al. (2011), o qual analisou o impacto de vários poluentes (CO, SO₂, NO₂, O₃, MP₁₀) na saúde utilizando redes neurais artificiais e identificou o MP₁₀ como o poluente com maior impacto para internações causadas por doenças cardiorrespiratórias. Outro estudo que também chegou a conclusões semelhantes foi o estudo de Zhu et al. (2013), o qual fez uma meta-análise que identificou a relação entre o MP₁₀ e o aumento de hospitalizações e mortes causadas pela obstrução crônica pulmonar.

Frequentemente, utilizam-se modelos de regressão estatística para análise dos impactos da poluição na saúde (TADANO et al., 2012; LAZARRI, 2013; VANOS, HEBBERN e CAKMAK, 2014). Outros estudos recentes mostram que as Redes Neurais Artificiais (RNA) são ferramentas robustas para mapeamento e previsão de impactos de poluição atmosférica na saúde (POLEZER et al., 2018; ARAUJO et al., 2020, TADANO et al., 2016; KASSOMENOS et al., 2011).

1.1. OBJETIVO GERAL

Desta forma, o objetivo deste trabalho foi realizar a previsão de internações por doenças respiratórias (Classificação Internacional de Doenças – CID-10 J00 – J99) causadas pelo material particulado com até 10 µm de diâmetro aerodinâmico (MP₁₀), na cidade de São Paulo, estado de São Paulo, Brasil, utilizando as RNA denominadas de Perceptron de Múltiplas Camadas (MLP – do inglês Multilayer Perceptron) e Máquina De Aprendizado Extremo (ELM – do inglês Extreme Learning Machine). Ambas as redes foram escolhidas pois são bastante conhecidas na literatura, a MLP pelos seus bons resultados e por ser amplamente usada, e a ELM por se tratar de uma rede com solução analítica, demandando baixo poder computacional (HAYKIN, 2015; HUANG, ZHU e SIEW, 2004). A principal justificativa para o estudo é analisar o poder de previsão das redes neurais para o problema de

poluição atmosférica relacionada com o número de internações por doenças respiratórias, e assim avaliar sua capacidade preditiva em cenários futuros.

1.2. OBJETIVO ESPECÍFICO

Para que seja possível alcançar os objetivos gerais, foram implementadas ambas as redes via programação em Python, realizou-se variações dos parâmetros das redes (quantidade de neurônios na camada escondida, função de ativação e *k fold cross-validation*) e comparou-se os desempenhos por métricas diferentes (MAE e MSE).

2. REFERENCIAL TEÓRICO

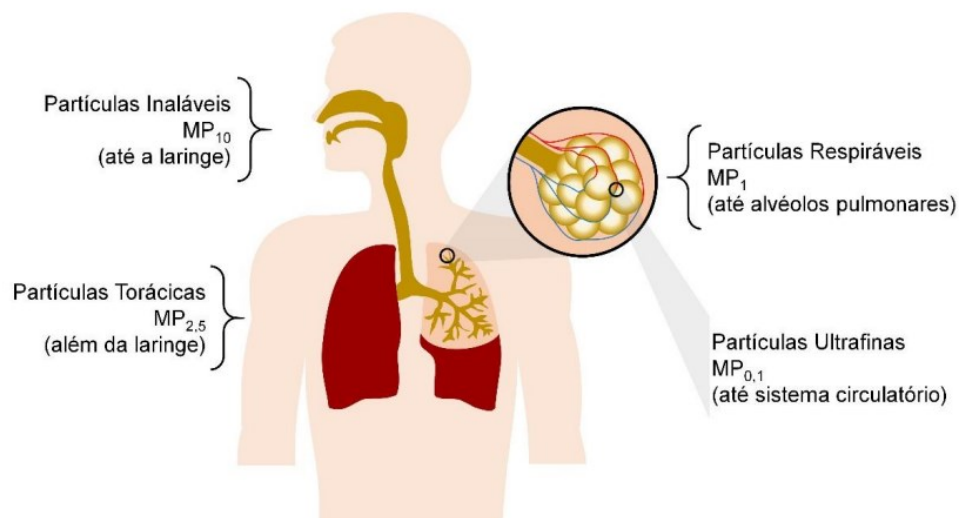
Nesta seção, explanou-se sobre Material Particulado e Redes Neurais Artificiais. Primeiro, falando sobre o que é Material Particulado e como são classificados. Em seguida, o que são Redes Neurais Artificiais e como elas são utilizadas para análises e previsões. Em seguida, os termos MLP, ELM, treinamento sobre-ajustado, sub-ajustado e validação cruzada foram explicados.

2.1. Material Particulado

Vários podem ser os poluentes prejudiciais para a saúde humana, segundo o Ministério do Meio Ambiente (2020), os mais comuns são CO, CO₂, CH₄, SO₂, Hidrocarbonetos (HC), NO e NO₂. Além desses, há também os poluentes que são constituídos por material sólido ou líquido de pequenas dimensões, que acabam ficando suspensos no ar, denominado material particulado (MP). O MP pode ser originado de fontes naturais (vulcões, queimadas naturais, entre outros) como antropogênicas (combustíveis fósseis, processos industriais, ressuspensão de poeiras, entre outros) (FREITAS E SOLCI, 2009). A classificação do MP pode ser feita pelo tamanho de partícula, composição química ou tipo da fonte (primária ou secundária).

Quanto ao tamanho, o MP é classificado como: Partículas Inaláveis Finas (MP_{2,5}) - cujo diâmetro aerodinâmico é menor ou igual a 2,5 µm, as quais penetram profundamente no sistema respiratório, podendo atingir os alvéolos pulmonares; Partículas Inaláveis (MP₁₀) - diâmetro aerodinâmico é menor ou igual a 10 µm, e podem ficar retidas na parte superior do sistema respiratório ou penetrar mais profundamente, alcançando os alvéolos pulmonares; Partículas Totais em Suspensão (PTS) - diâmetro aerodinâmico é menor ou igual a 50 µm; uma parte dessas partículas é inalável e pode causar problemas à saúde, outra parte pode afetar desfavoravelmente a qualidade de vida da população, interferindo nas condições estéticas do ambiente e prejudicando as atividades normais da comunidade (CETESB, 2021). A Figura 1 mostra as regiões do sistema respiratório humano em que o MP com diferentes diâmetros se deposita. Quanto mais finas as partículas, mais profundamente no sistema respiratório ela chega, podendo causar problemas respiratórios, cardiovasculares e até a morte para exposições de longo prazo.

Figura 1 – Regiões de deposição do MP para diferentes diâmetros aerodinâmicos



Fonte: Brito, Sodré e Almeida (2018).

A composição do MP varia com sua origem. Fontes naturais possuem composição semelhante ao material de origem, enquanto os provenientes da desagregação de rochas e solo são comumente ricos em minerais, matéria orgânica e eventualmente, partículas biológicas, como pólenes, esporos, bactérias, vírus e detritos de plantas (BRITO, SODRÉ E ALMEIDA, 2018).

Os MP originados direto da fonte são classificados como primários (poeiras do solo, por exemplo), enquanto os secundários são aqueles formados na atmosfera de maneira secundária (BRITO, SODRÉ E ALMEIDA, 2018).

2.2. Redes Neurais Artificiais

As redes neurais artificiais (RNA) são modelos computacionais capazes em relacionar dados de entrada e de saída, com o intuito de fornecer uma generalização de padrões e, assim, realizar previsões, mapeamento não-linear ou então, classificações. Outras possíveis tarefas de sua aplicação são, segundo SILVA, SPATTI E FLAUZINO (2018):

- Aproximação universal de funções;
- Reconhecimento de padrões;
- Identificação e controle de processos;
- Previsão de séries temporais;

- Otimização de sistemas.

Uma RNA é composta por várias unidades de processamento (neurônios artificiais) que são conectadas por ligações que estão associadas a determinado peso. As unidades fazem operações apenas sobre seus dados locais, que são entradas recebidas pelas suas conexões. O comportamento inteligente de uma RNA vem das interações entre as unidades de processamento da rede (HAYKIN, 2015; SILVA, SPATTI E FLAUZINO, 2018).

A propriedade mais importante das redes neurais é a capacidade em aprender de seu ambiente e, com isso, ir melhorando o seu desempenho. Isso é feito por meio de um processo de ajustes em seus pesos, fase conhecida como treinamento. O aprendizado ocorre quando a rede neural atinge uma solução generalizada para uma classe de problemas (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015).

Denomina-se algoritmo de aprendizado a um conjunto de regras para a solução de um problema de aprendizado. Há vários tipos de algoritmos de aprendizado específicos para determinados modelos de RNA, sendo diferenciados entre si principalmente pelo modo como os pesos são ajustados. Alguns exemplos desses algoritmos são: *Backpropagation*, *Batch Gradient Descent*, *Mini-batch Gradient Descent*, *Stochastic Gradient Descent* e o *Broyden-Fletcher-Goldfarb-Shanno (BFGS)*.

Outras definições importantes são quanto ao seu aprendizado. Aprendizado Supervisionado é quando se utiliza um agente externo que indica à rede, a resposta esperada para o padrão de entrada. Aprendizado Não Supervisionado, quando não existe um agente externo mostrando a resposta desejada para os padrões de entrada. Aprendizado por Reforço, quando um agente externo avalia a resposta fornecida pela rede (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015).Dentre as diversas RNAs existentes, neste trabalho serão utilizadas a MLP e a ELM.

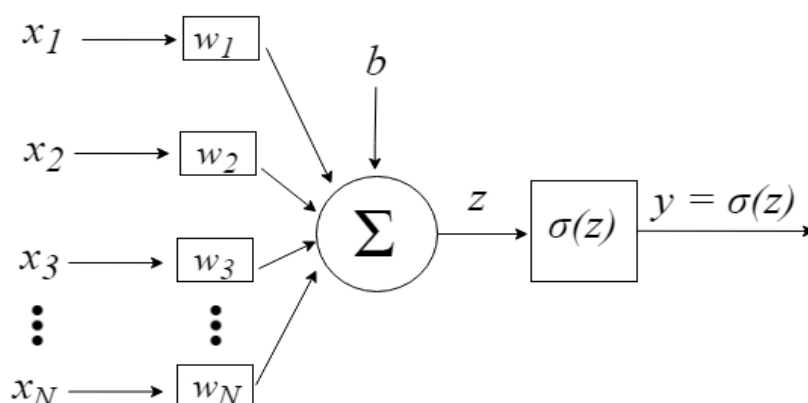
2.2.1. Perceptron de Múltiplas Camadas

A MLP é caracterizada pela presença de pelo menos uma camada intermediária (escondida), estando entre a camada de entrada e a respectiva camada neural de saída . O método de treinamento é feito de forma supervisionado e iterativa, necessitando de um alto poder computacional para o seu treinamento,

dependendo do tamanho da base de dados usada e de quantos neurônios são utilizados. Os pesos são, geralmente, gerados inicialmente de forma aleatória e atualizados a cada iteração (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015).

O fluxo de informações utilizadas na rede se inicia na camada de entrada, percorrendo em seguida, as camadas intermediárias e sendo finalizado na camada neural de saída. O modelo de aprendizagem de redes MLP consiste em duas etapas por meio das diferentes camadas da rede: a propagação e a retropropagação (que também são bastante conhecidas pelos seus termos em inglês *forward* e *backward*). Na etapa de propagação, um padrão de ativação é aplicado aos neurônios artificiais da camada de entrada da rede e seu efeito se propaga através da rede, camada por camada. Na última camada, um conjunto de saídas é produzido como estimativa do valor real pela rede. Deve-se salientar que, durante a etapa de propagação, os pesos sinápticos da rede são todos fixos. A Figura 2 mostra uma representação da etapa de propagação.

Figura 2 – Representação dos cálculos da etapa de propagação



Fonte: Leite (2018)

Os dados de entrada no neurônio são representados pelo vetor $x = [x_1, x_2, x_3, \dots, x_N]$, que são os dados das variáveis utilizadas. Ao chegarem ao neurônio, são multiplicados pelos respectivos pesos sinápticos, que são os elementos do vetor $w = [w_1, w_2, w_3, \dots, w_N]$, resultando em um valor z , conhecido como potencial de ativação, de acordo com a Equação 1:

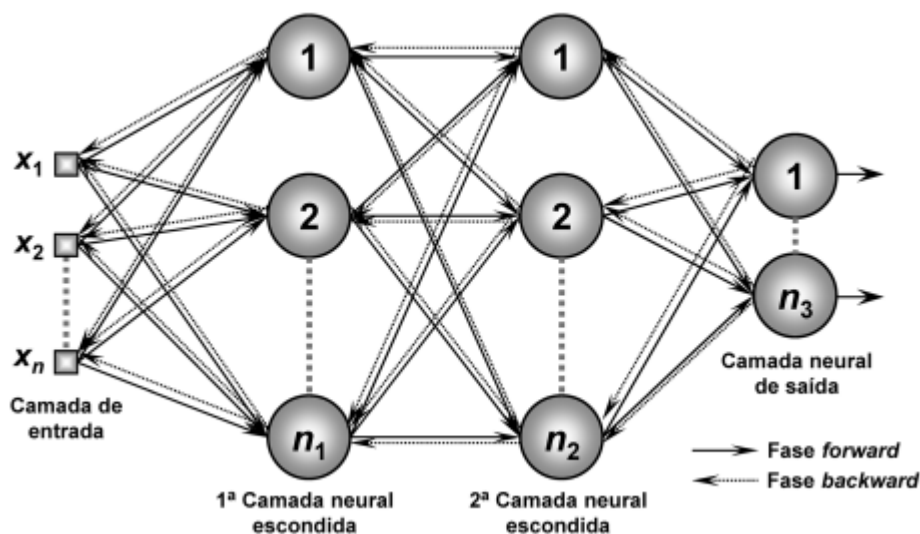
$$z = \sum_{i=1}^N x_i w_i + b \quad (1)$$

O potencial de ativação então servirá de argumento para a função de ativação, que tem como função fornecer capacidade não-linear para as redes neurais. As funções de ativação podem ser de dois tipos: parcialmente diferenciáveis (degrau, degrau bipolar) ou totalmente diferenciáveis (logística, tangente hiperbólica, gaussiana). As funções parcialmente diferenciáveis são menos recomendadas, pois não é possível calcular a derivada da função erro, sendo as funções totalmente diferenciáveis mais adequadas (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015).

Durante a etapa de retropropagação, os pesos sinápticos, que são parâmetros da rede, são ajustados de acordo com uma regra de correção do erro. A resposta real da rede é subtraída de uma resposta desejada para produzir um sinal de erro. A derivada da função erro é então propagado para trás, contra a direção das conexões sinápticas, sendo os pesos sinápticos ajustados para fazer com que a resposta real da rede se aproxime da resposta desejada (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015; NIED, 2007). A forma que será feita essa otimização dependerá da regra usada. Para este trabalho, utilizou-se o algoritmo de treinamento conhecido como Broyden-Fletcher-Goldfarb-Shanno (BFGS). Ele é, geralmente, usado para mínimos não lineares, ajudando a determinar a direção inicial de descida do gradiente (NAWI, RANSING e RANSING, 2006).

A Figura 3 representa uma arquitetura de uma rede MLP, sendo representadas também as etapas *forward* e *backward*.

Figura 3 – Arquitetura de uma rede MLP

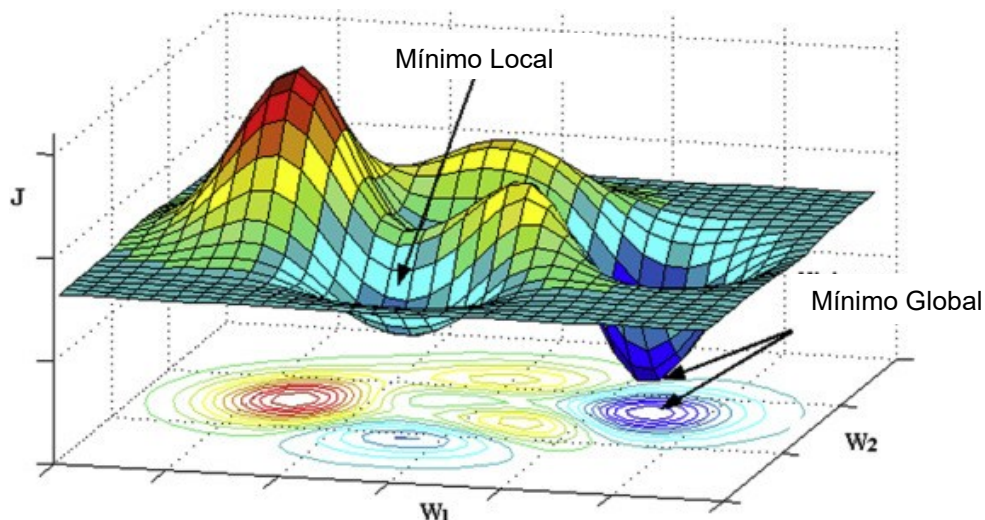


Fonte: Silva, Spatti e Flauzino (2018).

É possível perceber, pela Figura 3, que o número de neurônios em cada camada pode variar, assim como o número de camadas escondidas. Aumentar muito a quantidade de neurônios pode causar, em alguns casos, demora de aprendizado da rede devido ao alto volume de processamento de dados e, também, problema de sobre-ajuste da rede (SILVA, SPATTI E FLAUZINO, 2018).

O erro gerado após o término do treinamento da RNA dependerá dos valores iniciais dos seus pesos sinápticos e, também, por sua taxa de aprendizagem, os quais serão determinantes se o erro permanecerá em um mínimo local (menor erro para aquela parte do plano) ou um mínimo global (menor erro a ser alcançado pela generalização da rede). Por depender das condições iniciais, faz-se necessário que a rede seja testada várias vezes com pesos sinápticos iniciais diferentes.

Figura 4 – Representação do plano de erro em uma situação de dois pesos sinápticos (w_1 e w_2)



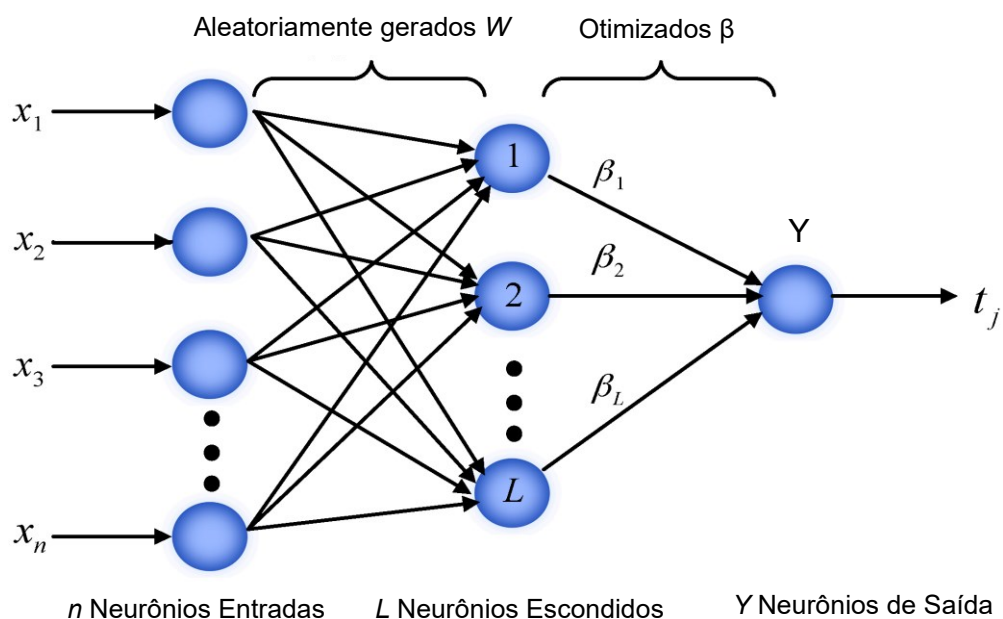
Fonte: Issaadi (2019).

Observa-se na Figura 4, um caso ilustrativo, em que o eixo J representa o valor do erro encontrado em relação aos eixos dos pesos w_1 e w_2 , ou seja, há um conjunto de pesos que irão representar a melhor generalização da rede, sendo esses, os pesos que irão gerar menor erro (mínimo global), que na imagem podem ser vistos na cor azul escura. Porém, para chegar nessa situação, é necessário que os pesos iniciais estejam próximos e convirjam para esse valor, caso contrário irão convergir para um mínimo local.

2.2.2. Máquina de Aprendizado Extremo

A arquitetura da ELM é similar à da MLP e apresenta apenas uma camada oculta. A principal diferença é a forma de treinamento da rede. Enquanto a MLP utiliza um processo iterativo, a ELM é feita de maneira analítica e apenas na camada de saída, o que acarreta grande economia de esforço computacional (HUANG, ZHU e SIEW, 2004).

Figura 5 – Arquitetura de uma ELM



Fonte: Adaptado de Li et al. (2017).

Como observado na Figura 5, a estrutura da ELM é muito parecida com a da MLP. Os dados de entrada são representados pelo vetor X , a matriz W com os pesos entre a camada de entrada e a camada intermediária, o vetor β com os pesos entre a camada intermediária e a camada de saída Y . A matriz W é gerada de maneira aleatória e não se altera até o término do cálculo da rede. Essa matriz pode ser gerada de uma distribuição uniforme no intervalo $[-1,1]$. Portanto, o objetivo do treinamento da ELM é encontrar o vetor de pesos β , baseado na matriz de saída Y e na matriz de pesos aleatórios W , por meio da resolução de um sistema linear (HUANG, ZHU e SIEW, 2004).

Para que seja possível encontrar os valores dos neurônios da camada intermediária, que para este caso está representado por L , é necessário fazer uma multiplicação entre o vetor X e a matriz W , sendo o resultado dessa multiplicação, o argumento da função de ativação para cada elemento do vetor L . Esse cálculo pode ser visto na Equação 2:

$$L_i = [x_1, x_2, x_3, \dots, x_N] \times \begin{bmatrix} w_{11} & \dots & w_{n1} \\ \vdots & \ddots & \vdots \\ w_{1L} & \dots & w_{nL} \end{bmatrix} \rightarrow L = \begin{bmatrix} f(L_1) \\ f(L_2) \\ \vdots \\ f(L_{L-1}) \\ f(L_L) \end{bmatrix} \quad (2)$$

Uma vez determinados os valores do vetor L , pode-se calcular os valores do vetor β . Para isto, é preciso resolver o sistema linear representado pela Equação 3:

$$L \beta = Y \rightarrow \beta = L^t Y \quad (3)$$

sendo L^t a inversa generalizada de *Moore–Penrose* da matriz L . Caso fosse utilizada a inverso padrão, os algoritmos ficariam limitados a problemas em que essa inversa existisse. A inversa generalizada diminui algumas exigências da inversa tradicional, como por exemplo, a matriz não necessita ser quadrada (HUANG, ZHU e SIEW, 2004; PACHECO, 2017). Não há iteração no treinamento da ELM. Os valores obtidos para o vetor β terão forte influência dos pesos iniciais gerados na matriz W , sendo necessário treinar a rede várias vezes para que se possa avaliar qual obtém menor erro. Com os valores do vetor β , pode-se encontrar a saída estimada da rede da mesma maneira que na MLP pela etapa *forward*.

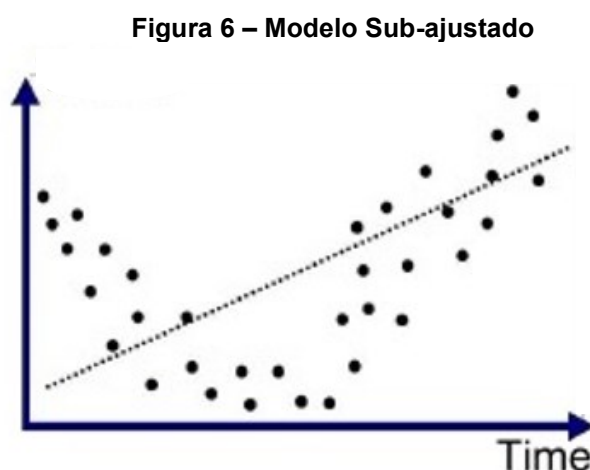
2.2.3. Separação dos Dados

Uma etapa bastante importante do processo é saber quais dados utilizar e como fazer essa distribuição. Para isso, é necessário dividir os dados em 3 subconjuntos: treinamento, validação e teste. Em todas as etapas, é responsabilidade de quem está fazendo a análise dizer quais são os dados de entrada e quais são os dados de saída da rede.

Os dados de treinamento serão utilizados para fazer os ajustes nos pesos, sendo necessária uma grande porção para essa etapa. Já os dados de validação serão utilizados para avaliar qual dos modelos treinados irá gerar o menor sinal de erro menor. Caso o erro seja menor, os pesos são atualizados, caso contrário, não. A etapa de teste é quando se terá noção sobre os resultados da rede, os dados de entrada serão utilizados para obter uma resposta da rede e, então, comparar com os dados reais (HAYKIN, 2015).

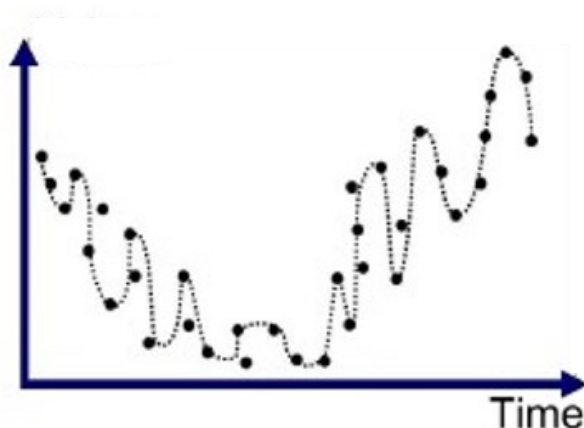
2.2.4. Treinamento sobre-ajustado e sub-ajustado

O treinamento é considerado sub-ajustado (também conhecido pelo seu termo em inglês, *underfitting*) ocorre quando o modelo não conseguiu aprender suficiente sobre os dados. O sub-ajuste leva à um erro elevado tanto nos dados de treino quando nos dados de teste. A Figura 6 – Modelo Sub-ajustado Figura 6 mostra um exemplo gráfico de uma rede sub-ajustada:



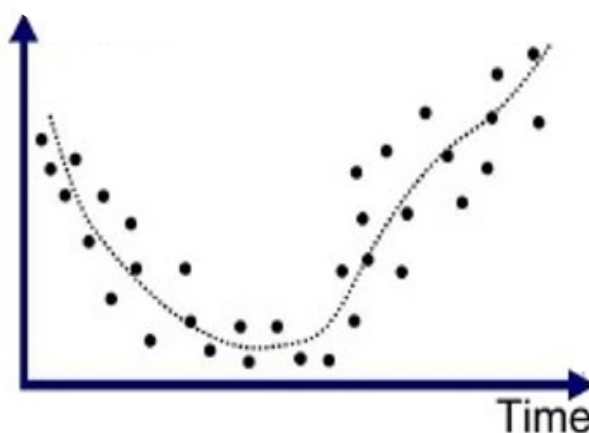
Fonte: Branco (2022).

Quando o treinamento é sobre-ajustado (também conhecido pelo seu termo em inglês, *overfitting*), ela aprende demais sobre os dados. Neste caso, o modelo se mostra adequado apenas para os dados de treino, como se o modelo tivesse apenas decorado os dados de treino e não fosse capaz de generalizar para outros dados nunca vistos antes. Quando isso acontece, os dados de treino apresentam resultados muito bons, enquanto a performance da fase de teste cai drasticamente. A Figura 6 – Modelo Sub-ajustado Figura 7 mostra um exemplo gráfico de uma rede sobre-ajustada.

Figura 7 – Modelo Sobre-ajustado

Fonte: Branco (2022).

O comportamento gráfico esperado como resposta de uma rede sem estar sofrendo o sobre-ajuste e o sub-ajuste pode ser visto na Figura 8. É importante salientar que analisar graficamente não é a melhor forma para avaliar se a rede está sofrendo com esses problemas, sendo necessário uma análise por meio das métricas de erro.

Figura 8 – Modelo ajustado

Fonte: Branco (2022).

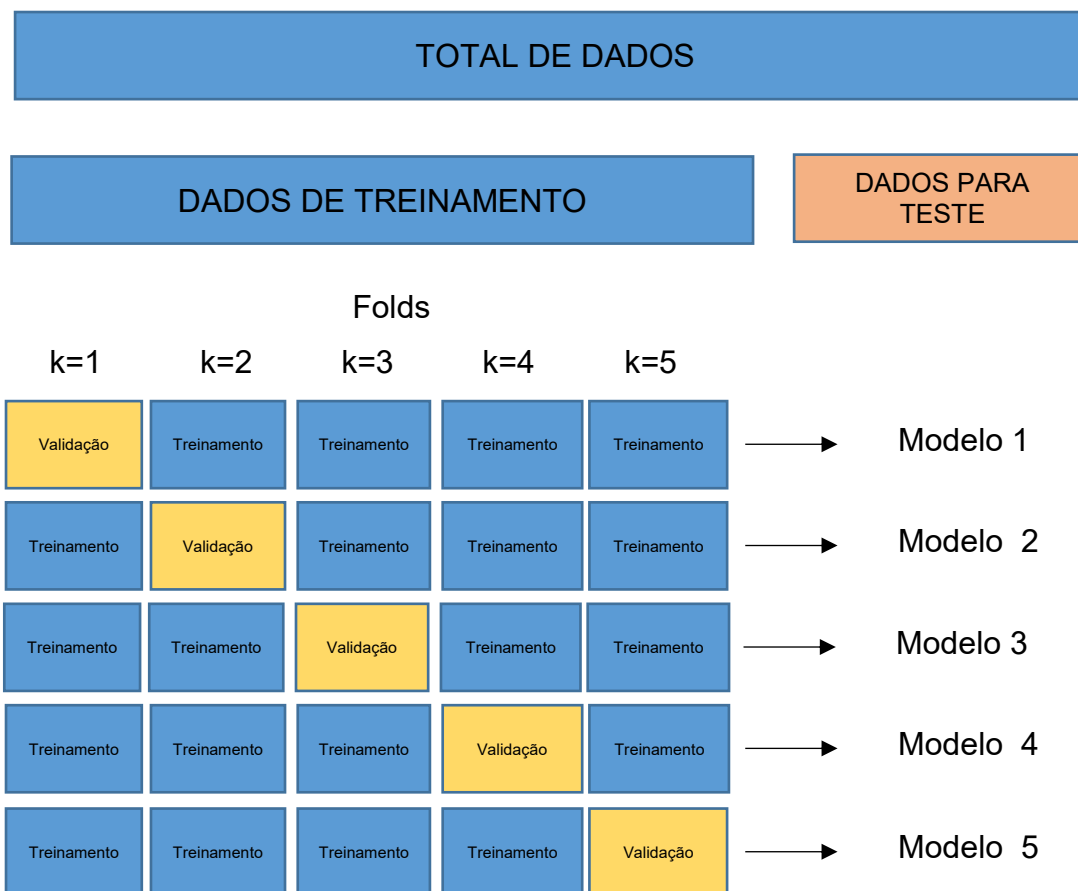
Para contornar esses problemas, é necessário que seja realizado o processo de validação cruzada.

2.2.5. Validação Cruzada

A validação cruzada, também conhecida pelo seu termo em inglês *cross-validation*, é uma técnica para avaliar a capacidade de generalização de um modelo, a partir de um conjunto de dados. Ela se faz necessária em problemas em que o objetivo da modelagem é a predição. Procura-se estimar o quão preciso é um modelo na prática, ou seja, o seu desempenho em um novo conjunto de dados. Preferivelmente, o modelo precisa ser avaliado em amostras que não foram usadas para construção ou ajuste do modelo, com a finalidade de fornecer um senso imparcial de eficácia do modelo. Os dados de treinamento são usados para criação do modelo, enquanto os de validação são usados para ajustar o desempenho (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015).

Há vários tipos de validação, dentre eles pode-se destacar: *K-fold Cross-validation*, Método *Holdout*, *Cross-validation Leave-P-Out*, *Bootstrap*, entre outros. Neste trabalho, o foco foi a validação *K-fold*. Para essa validação, divide-se o conjunto total de dados em três subconjuntos mutuamente exclusivos, um para treinamento (estimação dos parâmetros), um para validação (ajuste de parâmetros) e outro para teste. Os dados de teste serão usados apenas no final, quando a rede já estiver treinada e ajustada, e não terão nenhuma influência durante a etapa de treinamento e validação.

Figura 9 – Esquema da Validação Cruzada do tipo *k-fold*



Fonte: Autoria própria (2022).

Na Figura 9, pode-se observar a divisão dos dados da validação cruzada do tipo *k-fold*. Neste método, os dados são divididos em k subconjuntos (neste caso, 5), de tal forma que, a cada iteração, um dos k subconjuntos é usado como conjunto de validação e os outros subconjuntos $k-1$ são colocados juntos para formar um conjunto de treinamento. A estimativa de erro é calculada com base em todas as k tentativas para obter a eficácia total do modelo (SILVA, SPATTI E FLAUZINO, 2018; HAYKIN, 2015). Após a etapa de treinamento, para este exemplo, tem-se 5 modelos treinados, que serão utilizados na fase de teste, sendo utilizado aquele que obtiver menor erro.

3. METODOLOGIA

Nesta seção, será abordado o desenvolvimento do trabalho em si, a parte de revisão bibliográfica sistemática sobre o tema, informações sobre os dados usados, construção do código, parâmetros de rede e métricas para avaliação de desempenho.

3.1. Trabalhos Relacionados

Para melhor conhecimento sobre o tema estudado e identificação dos principais trabalhos relacionados, foi realizada uma revisão bibliográfica sobre os trabalhos relacionados ao tema. A pesquisa de artigos foi realizada em seis 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. Utilizou-se a palavra *health* para que fosse possível abranger todos os problemas de saúde causados pelos poluentes atmosféricos, e não apenas os problemas respiratórios. Os anos de publicações dos artigos ficaram 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 concentração 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.

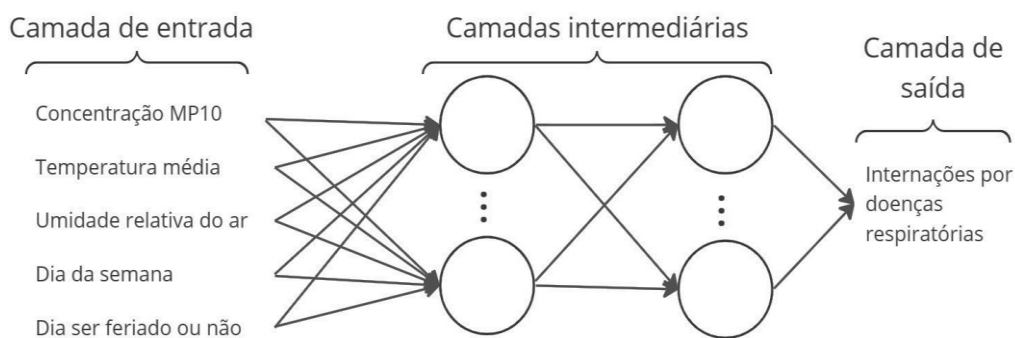
Durante a pesquisa manual, foram selecionados artigos que remetessem ao tema pelo seu título, sendo realizada a leitura dos seus resumos, para que fosse possível ter certeza sobre o tema abordado em cada artigo. Para obter um panorama das publicações no tema, informações específicas foram observadas: RNA usada, principais variáveis e *outputs* (saída da rede).

3.2. Banco de dados

Como estudo de caso, considerou-se os dados de 01 de janeiro de 2014 a 31 de dezembro de 2016 para a cidade de São Paulo, SP, Brasil, consistindo em 1.009 dados diários. Foi feito um tratamento prévio na base para retirar dados faltantes. As variáveis utilizadas neste trabalho foram: concentração de MP₁₀,

temperatura média, umidade relativa do ar, o dia da semana e o dia ser feriado ou não, já que os dados de internações frequentemente possuem um padrão definido, com menos internações sendo observadas em dias de semana e dias de não feriado (TADANO et al., 2016; ARAUJO et al., 2020). Na Figura 10, podemos ver um esquema das variáveis de entrada e saída em um esquema de uma rede neural.

Figura 10 – Esquema das variáveis dentro da rede



Fonte – Autoria própria (2022).

As variáveis de dias são qualitativas, ou seja, podem ser separadas em diferentes categorias que se distinguem por alguma característica não numérica. Portanto, foi necessário atribuir valores para elas, que podem ser vistos no Quadro 1.

Quadro 1 – Variáveis qualitativas e seus respectivos valores

Variável	Valor atribuído
Domingo	1
Segunda-feira	2
Terça-feira	3
Quarta-feira	4
Quinta-feira	5
Sexta-feira	6
Sábado	7
Não feriado	0
Feriado	1

Fonte: Autoria própria (2022).

Para o período estudado, existiam doze estações de monitoramento, das quais somente quatro monitoravam as concentrações de MP₁₀. Dentre estas quatro

(Classificação Internacional de Doenças - CID-10: J00- J99) (DATASUS, 2018) da cidade de São Paulo como um todo.

Algo que deve ser considerado é o efeito da poluição atmosférica na saúde após alguns dias de exposição ao poluente, sendo comum trabalhar com defasagens de até sete dias (TADANO et al., 2012; POLEZER et al., 2018; ARAUJO et al., 2020).

Por isso, neste trabalho, foram realizadas análises considerando o efeito na saúde devido à exposição ao MP₁₀ no mesmo dia da exposição (*lag* 0) até 7 dias após a exposição (*lag* 7).

3.3. Implementação computacional da MLP

A MLP foi implementada utilizando a linguagem de programação Python versão 3.7. Usou-se, também, a biblioteca *Scikit-learn*, a qual possibilitou a implementação da rede por meio da ferramenta MLP *Regressor*, e a biblioteca *Pandas* para se trabalhar com os dados. Para tanto, os parâmetros utilizados foram:

- Quantidade de camadas escondidas: 1;
- Quantidade de neurônios na camada escondida: variação de 5 à 500, alterando de 5 em 5;
- Quantidade de execuções: 2;
- Número máximo de iterações: 1000;
- Taxa de aprendizagem: 0,01;
- Função de ativação: Logística, Tangente hiperbólica;
- Valor do erro permitido para interromper o treinamento: 10^{-4} ;
- Algoritmo de treinamento: Broyden–Fletcher–Goldfarb–Shanno algorithm (BFGS);
- Validação: K fold *Cross-Validation*, $k=5$.

Foi necessário fazer uma varredura dos parâmetros das redes para que fosse possível avaliar quais parâmetros obteriam melhor resultado (quantidade de neurônios e funções de ativação). O otimizador de pesos BFGS foi escolhido devido à melhor convergência dos pesos quando comparados a outros algoritmos, tendo

sido realizado testes prévios, resultando em valores muito distantes do esperado para outras otimizadores (*Backpropagation*, *Stochastic Gradiente Descente*, *Stochastic Gradiente-based Optimizer*).

Os dados foram divididos em 70% treinamento, 15% validação e 15% teste. Essa divisão foi escolhida por ser amplamente usada na literatura (HAYKIN, 2015) 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 respectivamente nas Equações 4 e 5. Ambas as métricas foram utilizadas, pois são parâmetros amplamente usados (BOTCHKAREV, 2018)

$$\text{MSE} = \frac{1}{N_{td}} \sum_{t=1}^N (r_t - y_t)^2 \quad (2)$$

$$\text{MAE} = \frac{1}{N_{td}} \sum_{t=1}^N |r_t - y_t| \quad (3)$$

sendo r_t o valor real observado, y_t a saída da rede e N_{td} o número total de dados utilizados na fase de teste. O MAE será uma média aritmética absoluta do erro, enquanto o MSE irá punir mais os erros por sua fórmula estar elevando ao quadrado, sendo uma boa métrica para situações em que o erro pode ser um problema grave. Para ambas as métricas, quanto menor o valor, melhor será o resultado.

Foi necessário realizar uma normalização dos dados para não permitir que uma variável tivesse maior impacto que outras devido a diferença de seus valores, sendo normalizados entre o intervalo +1 e -1. Após o treinamento da rede e obtenção dos resultados da rede, foi necessário desnormalizar os resultados para que fossem testados com os valores reais e obtenção do MSE e MAE (MICROSOFT, 2022).

3.4. Implementação computacional da ELM

Para a implementação da ELM, utilizou-se também programação Python, versão 3.7. Foi usada, predominantemente, a biblioteca *Numpy* e *Pandas*, o que facilitou as multiplicações de matrizes e possibilitou a obtenção da matriz inversa generalizada de Moore-Penrose, necessária para a solução da rede. A variação dos

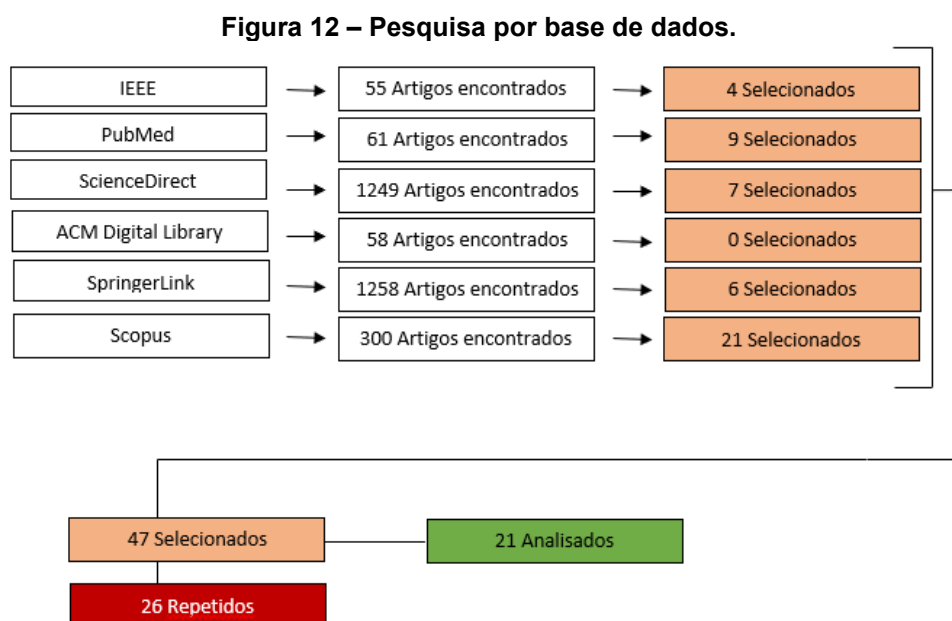
parâmetros ocorreu da mesma forma que a MLP, não sendo considerados os parâmetros não existentes na ELM (taxa de aprendizagem, número máximo de iterações e valor para interromper o treinamento). As métricas de erro também foram MSE e MAE para avaliação de performance da rede.

4. RESULTADOS

Nesta seção, os resultados estarão organizados em duas seções principais. A primeira tratará dos artigos relacionados ao tema do trabalho, apresentando uma breve descrição desses artigos e destacando os principais pontos de contribuição. A segunda tratará do desempenho das redes neurais utilizadas no trabalho, comparando os resultados obtidos com diferentes modelos e parâmetros.

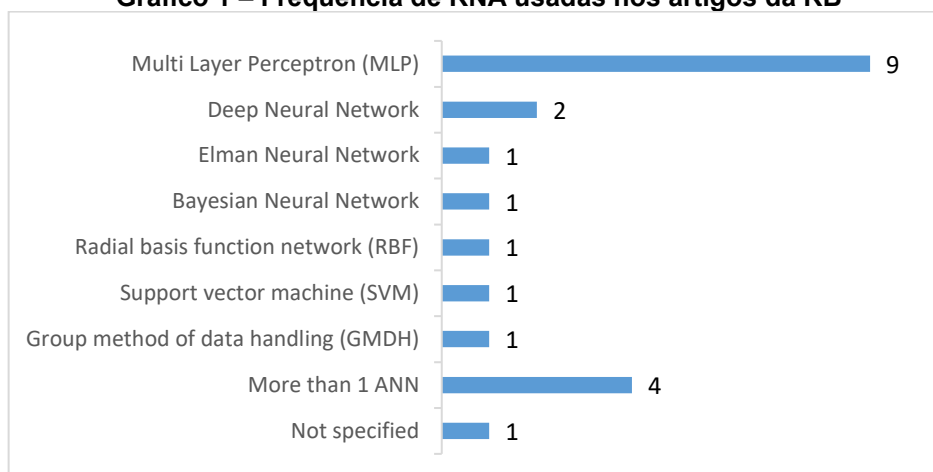
4.1. Trabalhos Relacionados

Na revisão bibliográfica foram encontrados, somando todas as bases de dados, um total de 2.981 artigos, porém, foram selecionados apenas 21 artigos (0,7% do total) que, de alguma forma, relacionavam os três temas (poluição atmosférica, saúde humana e RNA). A Figura 12 mostra quantos artigos foram encontrados em cada base de dados e quantos foram selecionados. Foi possível observar uma grande quantidade de artigos repetidos (26) entre as bases.



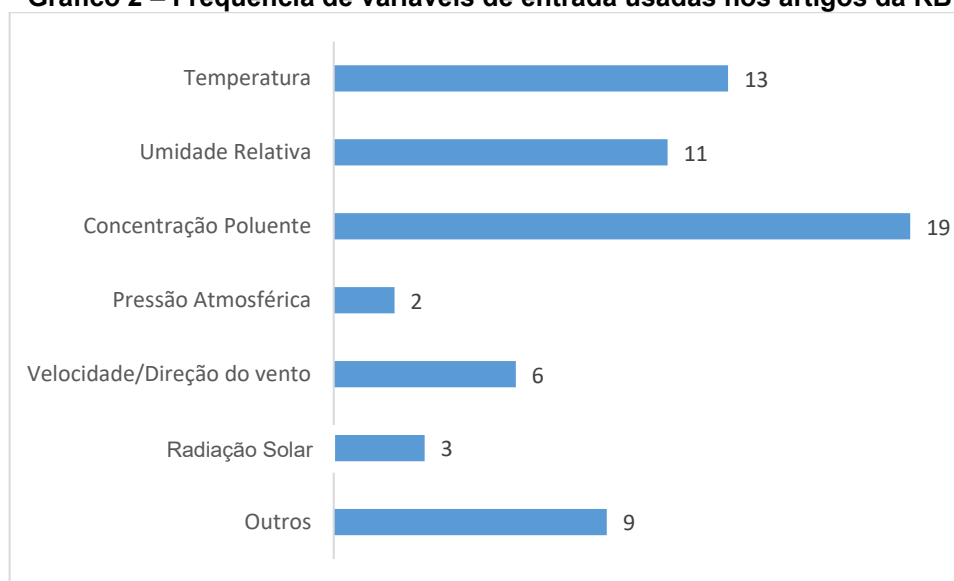
Fonte: Autoria própria (2022)

Entre as RNA usadas, houve uma predominância no uso da MLP, contabilizando 9 artigos dentre os 21 selecionados. O Gráfico 1 mostra as relações entre a frequência em que cada RNA apareceu durante a pesquisa.

Gráfico 1 – Frequência de RNA usadas nos artigos da RB

Fonte: Autoria própria (2022)

No Gráfico 2, são mostradas as principais variáveis utilizadas pelos autores e os parâmetros de entrada das RNA.

Gráfico 2 – Frequência de variáveis de entrada usadas nos artigos da RB

Fonte: Autoria própria (2022)

Observa-se que os dados de temperatura, umidade e concentração dos poluentes são os mais utilizados dentre os dados de entrada, por serem aqueles que mais afetam diretamente a saúde humana. Por esta razão, escolheu-se utilizar estas variáveis de entrada neste estudo.

No Gráfico 3, são mostradas as principais variáveis calculadas na camada de saída dos artigos analisados. Observa-se o destaque nas análises de

hospitalizações por doenças respiratórias, efeito à saúde que está mais relacionado à poluição atmosférica.

Gráfico 3 - Frequência de variáveis de saída usadas nos artigos da RB



Fonte: Autoria própria (2022).

O artigo com maior número de citações, de RAM et al. (2015), foi encontrado na base de dados IEEE. A pesquisa foi realizada nos Estados Unidos e procurou prever o número de casos de emergência por asma. Não foi especificado qual RNA utilizou, mas utilizou como dados de entrada, a concentração do poluente (CO, NO₂, MP_{2,5}) e um banco de dados do Twitter com tweets que continham palavras-chave relacionadas à asma. Pré-processamento e seleção de variáveis (*backward feature selection algorithm*) foram realizados. O periódico de publicação foi o *Journal of Biomedical and Health Informatics*.

4.2. Desempenho das redes

A Tabela 1 e Tabela 2 mostram os resultados do MSE e MAE obtidos para todos os *lags*, sendo mostrado entre parênteses a quantidade de neurônios na camada escondida e em fundo verde os melhores resultados.

Tabela 1 - MAE e MSE para MLP com ambas as funções de ativação

	MLP Tangente Hiperbólica		MLP Logística	
	MAE	MSE	MAE	MSE
Lag 0	34,38 (10 Neurônios)	1820,18 (330 Neurônios)	36,53 (10 Neurônios)	2085,48 (10 Neurônios)
Lag1	35,93 (70 Neurônios)	1951,79 (70 Neurônios)	38,66 (5 Neurônios)	2238,44 (5 Neurônios)
Lag 2	37,43 (5 Neurônios)	2209,15 (310 Neurônios)	38,86 (15 Neurônios)	2378,40 (15 Neurônios)
Lag 3	38,28 (355 Neurônios)	2151,13 (355 Neurônios)	39,39 (5 Neurônios)	2333,16 (5 Neurônios)
Lag 4	37,32 (315 Neurônios)	2116,61 (280 Neurônios)	40,47 (15 Neurônios)	2473,58 (20 Neurônios)
Lag 5	37,82 (485 Neurônios)	2188,64 (180 Neurônios)	39,40 (15 Neurônios)	2433,24 (15 Neurônios)
Lag 6	36,70 (275 Neurônios)	2155,27 (60 Neurônios)	37,46 (10 Neurônios)	2071,95 (10 Neurônios)
Lag 7	36,40 (230 Neurônios)	2139,58 (230 Neurônios)	37,54 (5 Neurônios)	2220,32 (5 Neurônios)

Fonte: Autoria própria (2022).

Tabela 2 - MAE e MSE para ELM com ambas as funções de ativação

	ELM Tangente Hiperbólica		ELM Logística	
	MAE	MSE	MAE	MSE
Lag 0	43,31 (110 Neurônios)	3111,70 (110 Neurônios)	40,36 (100 Neurônios)	2666,65 (90 Neurônios)
Lag 1	41,20 (175 Neurônios)	2961,55 (175 Neurônios)	42,03 (95 Neurônios)	3001,98 (120 Neurônios)
Lag 2	39,90 (135 Neurônios)	2585,73 (135 Neurônios)	42,36 (155 Neurônios)	3163,37 (155 Neurônios)
Lag 3	45,35 (80 Neurônios)	3455,28 (80 Neurônios)	48,67 (215 Neurônios)	3978,12 (215 Neurônios)
Lag 4	44,06 (155 Neurônios)	3368,17 (155 Neurônios)	42,94 (130 Neurônios)	3190,37 (130 Neurônios)
Lag 5	44,59 (130 Neurônios)	3256,35 (130 Neurônios)	43,29 (110 Neurônios)	3206,25 (110 Neurônios)
Lag 6	43,11 (130 Neurônios)	3287,03 (100 Neurônios)	40,58 (125 Neurônios)	2887,40 (125 Neurônios)
Lag 7	44,46 (110 Neurônios)	3289,24 (210 Neurônios)	40,35 (130 Neurônios)	2783,83 (130 Neurônios)

Fonte – Autorial própria (2022).

Pelas Tabelas 2 e 3, quando se compara os resultados de cada defasagem (*lag*), a MLP alcançou menores erros para *lag 0* utilizando a função tangente hiperbólica (MAE – 34,38 e MSE – 1820,18). Para a ELM, os menores erros foram obtidos o *lag 2* também utilizando a tangente hiperbólica. Observa-se, então, um melhor resultado da MLP quando comparada a ELM (MAE – 39,90 e MSE – 2585,73). Os erros obtidos para a tangente hiperbólica são menores quando comparado aos obtidos pela função logística, porém a diferença não é tão significativa, sendo os valores de ambas as funções próximas.

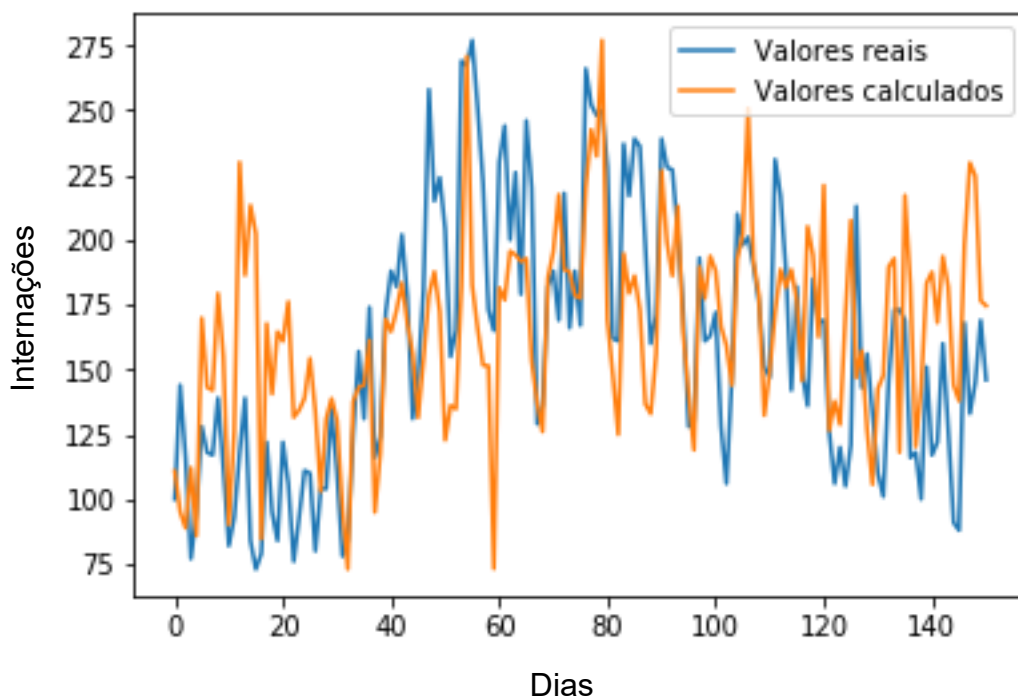
A quantidade ótima de neurônios utilizados na camada escondida variou bastante para cada *lag*, sendo necessário uma quantidade menor para a MLP se levar em conta o menor MAE, 10 neurônios, enquanto para a ELM foram necessários 135 neurônios. Além disso, a função tangente hiperbólica mostrou ser mais eficaz para esse tipo de previsão.

Os altos valores obtidos para o MSE são explicados pela diferença entre o valor real e a resposta da rede ser elevado ao quadrado, fazendo sentido os valores obtidos. Além disso, o MSE irá aumentar ainda mais quando houverem previsões que ficaram muito distantes dos valores reais, ou seja, ele penalizará resultados que estiverem muito distantes do real.

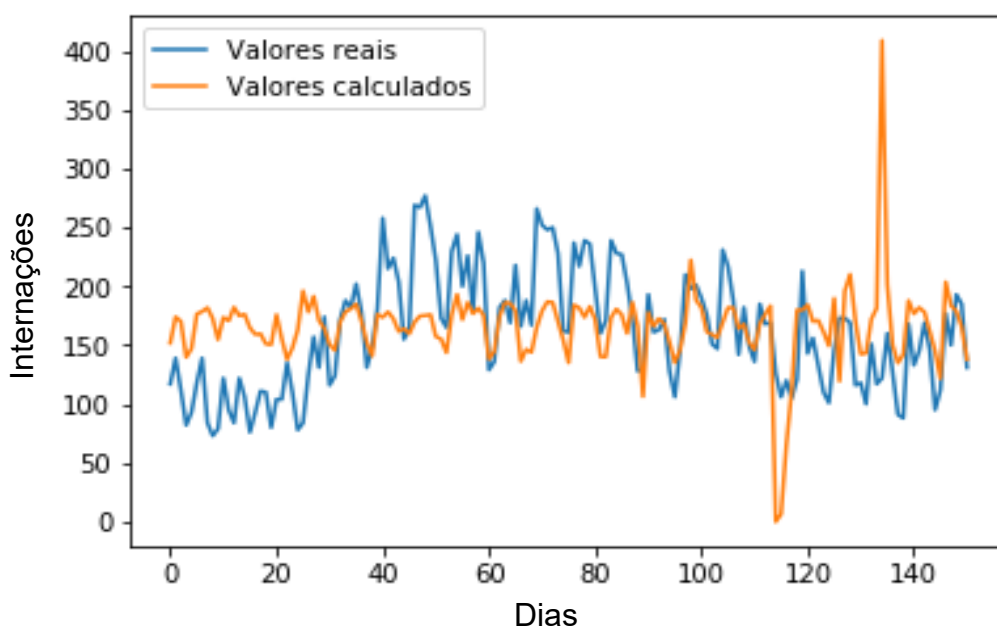
Alguns estudos epidemiológicos de poluição do ar também apresentaram melhor desempenho com uso da MLP, como Polezer et al. (2018). Neste estudo foi analisado o poluente $MP_{2,5}$ para a cidade de Curitiba – PR, estudo no qual comparou três RNA (MLP, ELM e ESN) sendo feita uma variação na quantidade de neurônios na camada intermediária em todos eles, com o melhor resultado ficando para a MLP para o caso de 50 neurônios na camada escondida no *lag* 1, com um valor de MSE de 54,05. Os dados utilizados foram de setembro de 2008 a julho de 2011 e de setembro de 2014 a setembro de 2015.

Entretanto, em outros, a ELM apresentou melhores resultados (Tadano et al., 2016). O trabalho comparou três RNA (ELM, ESN, MLP) e utilizou dados da cidade de Campinas – SP, datados de 1 de janeiro de 2007 a 30 de setembro de 2008. As variáveis usadas foram concentração do MP_{10} , temperatura, umidade relativa e o número de hospitalizações por doenças respiratórias. Foram feitas variações na quantidade de neurônios na camada escondida, tendo o melhor resultado para a ELM com 30 neurônios na camada escondida, alcançando um MSE de 35,8. Esta variação na RNA que melhor se ajusta a cada base de dados enfatiza a importância de se aplicar mais de uma RNA em qualquer problema estudado.

Os Gráficos 4 e Gráfico 5 apresentam de forma visual, os melhores resultados da fase de teste de cada rede neural proposta (MLP e ELM, respectivamente), sendo possível comparar de forma visual os resultados obtidos pela rede e os valores reais observados.

Gráfico 4 – Valores observados versus valores ajustados para a MLP

Fonte – Autoria própria (2022).

Gráfico 5 – Valores observados versus valores ajustados para a ELM

Fonte – Autoria própria (2022).

É possível observar graficamente uma tendência de comportamento da saída produzida pela rede seguir os mesmos padrões que o valores reais, mesmo que em alguns pontos houvesse diferença significativa nos valores. Isso indica que de fato as variáveis de entrada possuem relação com as saídas, mas que ainda há a necessidade de utilização de outras variáveis para melhor generalização das redes.

Salienta-se que, este é um problema de mapeamento não-linear estático, em que a saída (internações por doenças respiratórias) depende de várias outras variáveis, além das consideradas neste estudo, como idade, histórico familiar, fatores genéticos, entre outros (Araujo et al., 2020), sendo assim, é esperada uma diferença entre os valores ajustados e observados, os quais não coincidem para parte significativa das amostras (Tadano et al., 2016; Polezer et al., 2018; Araujo et al., 2020).

5. CONCLUSÃO

A revisão bibliográfica mostrou ao total 2981 resultados, sendo selecionados apenas 21 artigos. Houve uma predominância no uso da MLP como rede neural, sendo o seu uso 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. As RNAs provaram ser uma ferramenta importante de previsão. Isso possibilita medidas preventivas por parte de governantes com o intuito de desafogar os sistemas de saúde em períodos críticos de poluição do ar ou outras situações, como a da pandemia da COVID-19. Para ambas as métricas de erro, a MLP se mostrou ser uma RNA mais robusta para este estudo. Para trabalhos futuros, recomenda-se utilizar dados hospitalares da mesma região em que foram extraídos os dados dos poluentes, para que seja possível verificar o desempenho das redes para esses dados. Além disso, utilizar variáveis que não foram consideradas para este trabalho por não terem sido possível obtê-las, mas que possuem importante relação com poluentes atmosféricos e internação por doenças respiratórias, sendo elas: quantidade de veículos circulando, velocidade do vento, quantidade de pessoas circulando.

REFERÊNCIAS

- ARAUJO, L. N. et al. (2020). *Ensemble method based on Artificial Neural Networks to estimate air pollution health risks*. **Environmental Modelling and Software**, vol.123, 104567.
- BOTCHKAREV, A. (2018). *Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics: Properties and Typology*. **Interdisciplinary Journal of Information, Knowledge, and Management**, v 14, 45-79.
- BRANCO, H. (2022). **Overfitting e underfitting em Machine Learning**. Disponível em: <<https://abracd.org/overfitting-e-underfitting-em-machine-learning/>>. Acesso em: 3 dez. 2022.
- BRITO, G. F. S.; SODRÉ, F. F.; ALMEIDA, F. V. (2018). O Impacto do Material Particulado na Qualidade do Ar. *Revista Virtual de Química*, v. 10 n. 5.
- CETESB - **Companhia Ambiental do Estado de São Paulo** (2021). <https://cetesb.sp.gov.br/ar/poluentes/>, março.
- CETESB - **Companhia Ambiental do Estado de São Paulo** (2018). <https://cetesb.sp.gov.br/ar/qualar/>, outubro.
- DataSUS - **Departamento de Informática do Sistema Único de Saúde** (2018). <http://www2.datasus.gov.br/DATASUS/%20index.php?area=02>, dezembro.
- FREITAS, A. M. e SOLCI, M. C. (2009). Caracterização do MP₁₀ e MP_{2,5} e distribuição por tamanho de cloreto, nitrato e sulfato em atmosfera urbana e rural de Londrina. **Química Nova**, vol. 32, páginas 1750-1754.
- HAYKIN, S. (2015). **Neural Networks and Learning Machines**, Prentice Hall, 3 rd edition.
- HUANG, G. B.; ZHU, Q. Y.; SIEW, C. K. (2004). *Extreme learning machine: a new learning scheme of feedforward neural networks*. **IEEE International Joint Conference on Neural Networks**, v. 2, p. 985–990.
- IBGE - **Instituto Brasileiro de Geografia e Estatística** (2019). <https://www.ibge.gov.br/estatisticas-novoportal/sociais/populacao/9103-estimativas-de-populacao.html?1/4&t1/4resultados>, julho.
- ISSAADI, S., ISSAADI, W., & KHIREDDINE, A. (2019). **New intelligent control strategy by robust neural network algorithm for real time detection of an optimized maximum power tracking control in photovoltaic systems**. *Energy*, 115881. doi:10.1016/j.energy. 2019.115881.
- KASSOMENOS, P. et al. (2011). *Identifying the contribution of physical and chemical stressors to the daily number of hospital admissions implementing an artificial neural*

network model. Air Qual Atmos Health 4, 263–272. <https://doi.org/10.1007/s11869-011-0139-2>

LAZZARI, A. R. (2013). Comparação de técnicas estatísticas para analisar a relação entre doenças respiratórias e concentrações de poluentes atmosféricos. **Ciência e Natura**, v. 35, n. 1, p. 98-105.

LEITE, T. M. (2018). Redes Neurais, Perceptron Multicamadas e o Algoritmo Backpropagation. **Medium**. Disponível em: <<https://medium.com/ensina-ai/redes-neurais-perceptron-multicamadas-e-o-algoritmo-backpropagation-eaf89778f5b8>> . Acesso em: 05, nov. 2022.

LI, Y. et al. (2017) *A Novel Online Sequential Extreme Learning Machine for Gas Utilization Ratio Prediction in Blast Furnace.. Sensors* v.17 p.1847.

MICROSOFT. (2022). Componente Normalizar Dados. Disponível em: <<https://learn.microsoft.com/pt-br/azure/machine-learning/component-reference/normalize-data>> . Acesso em: 05, nov. 2022.

NIED, A. (2007). **Treinamento de redes neurais artificiais baseado em sistemas de estrutura variável com taxa de aprendizado adaptativa**. 2007. Tese (Doutorado) - Universidade Federal de Minas Gerais.

NAWI, N. M., RANSING, M. R., e RANSING, R. S. (2006). *An Improved Learning Algorithm Based on The Broyden-Fletcher-Goldfarb-Shanno (BFGS) Method For Back Propagation Neural Networks. Sixth International Conference on Intelligent Systems Design and Applications*, 2006, pp. 152-157, doi: 10.1109/ISDA.2006.95.

OMS - Organização Mundial da Saúde (2020) **“Air Pollution”**, https://www.who.int/health-topics/air-pollution#tab=tab_1, maio.

PACHECO, A. (2017). A máquina de aprendizado extremo – ELM. **Medium**. Disponível em: <<http://computacaointeligente.com.br/algoritmos/maquina-de-aprendizado-extremo/>>. Acesso em: 05, nov. 2022.

Python (2019). <https://phyton.org>.

POLEZER et al. (2018). *Assessing the impact of PM 2.5 on respiratory disease using artificial neural networks. Environmental Pollution*, vol. 235, pages 394-403.

RAM, S., ZHANG, W., WILLIAMS, M., & PENGETNZE, Y. (2015). *Predicting Asthma-Related Emergency Department Visits Using Big Data. IEEE Journal of Biomedical and Health Informatics*, 19(4), 1216–1223. doi:10.1109/jbhi.2015.2404829

SHAIKH. R. (2018). *Cross Validation Explained: Evaluating estimator performance. Medium*. Disponível em: <<http://computacaointeligente.com.br/algoritmos/maquina-de-aprendizado-extremo/>>. Acesso em: 05, nov. 2022.

SILVA, I. D., SPATTI, D. H., FLAUZINO, R. D. (2016), Redes Neurais Artificiais para Engenharia e Ciências Aplicadas, **Artliber**, 2a edição.

TADANO, Y. S. et al. (2021) *Dynamic model to predict the association between air quality, COVID-19 cases, and level of lockdown*. **Environmental Pollution**, vol. 268, 115920.

TADANO, Y. S. et al. (2012). “*Methodology to assess air pollution impact on human health using the Generalized Linear Model with Poisson regression*”, In: *Air Pollution – Monitoring, Modelling and Health*, Edited by D. M. Khare, Intech, England.

TADANO, Y. S. et al. (2016). *Unorganized machines to predict hospital admissions for respiratory diseases*. 2016 **IEEE Latin American Conference on Computational Intelligence** (LA-CCI).

VANOS, J., HEBBERN, C., CAKMAK, S. (2014). Risk assessment for cardiovascular and respiratory mortality due to air pollution and synoptic meteorology in 10 Canadian cities. **Environmental Pollution**, n. 185, p. 322–332.

ZHU, R. et al. (2013) *The Relationship between Particulate Matter (PM10) and Hospitalizations and Mortality Of Chronic Obstructive Pulmonary Disease: A Meta-Analysis*, COPD: **Journal of Chronic Obstructive Pulmonary Disease**, 10:3, 307-315, DOI: 10.3109/15412555.2012.744962