

# Revisão Sistemática em Sensoriamento de Qualidade de Ar via Lógica Fuzzy e Aprendizado de Máquina

Vagner A. Seibert<sup>1</sup>, Rafael R. Bastos<sup>1</sup>, Gabriel R. Silva<sup>1</sup>,  
Adenauer Yamin<sup>1</sup>, Renata H. R. Reiser<sup>1</sup>, Giancarlo Lucca<sup>2</sup>, Helida Santos<sup>3</sup>

<sup>1</sup> Centro de Desenvolvimento Tecnológico – Universidade Federal de Pelotas (UFPel)  
Pelotas – RS – Brasil

{vaseibert, rrbastos, adenauer, grosilva, reiser}@inf.ufpel.br

<sup>2</sup>PGEEC – Universidade Católica de Pelotas (UCPEL), Pelotas-RS, Brasil

giancarlo.lucca@ucpel.edu.br

<sup>3</sup>Centro de Ciências Computacionais - Universidade Federal do Rio Grande  
Rio Grande - RS - Brasil.

helida@furg.br

**Abstract.** *This work proposes a systematic review of the literature about the state of the art of air quality sensors research. This study intends to analyze methodologies integrating machine learning techniques and fuzzy algorithms for sensor calibration and prediction, attempting to demonstrate the main differences and advantages of the approaches. One of the main concerns of this work is to highlight the most used techniques and algorithms, as well as elucidate in which context each one of these methods improves the applications, aiming to achieve a better understanding of this area.*

**Resumo.** *Este trabalho propõe uma revisão sistemática da literatura sobre o estado da arte da pesquisa de sensores de qualidade do ar. Este estudo pretende analisar metodologias que integram técnicas de aprendizado de máquina e algoritmos fuzzy para calibração e previsão de sensores, tentando demonstrar as principais diferenças e vantagens das abordagens. Uma das principais preocupações deste trabalho é destacar as técnicas e algoritmos mais utilizados, bem como elucidar em que contexto cada um destes métodos melhora as aplicações, visando alcançar um melhor entendimento desta área.*

## 1. Introdução

O sensoriamento é definido como o conjunto de instrumentos utilizados para coletar informações remotamente [Campbell and Wynne 2011]. Essas informações podem ser referentes a diversas questões do nosso cotidiano, como por exemplo: ao clima [Yang et al. 2013], oceanos [Kumari et al. 2019], iluminação [Pandharipande and Caicedo 2015], ou, como um dos focos desse trabalho, a qualidade do ar [Zhang and Srinivasan 2020].

A Organização Mundial da Saúde<sup>1</sup> (OMS) reportou que a baixa qualidade de ar foi a causa de 4.2 milhões de mortes em 2016 [Organization 2016]. A existência de

---

<sup>1</sup><https://www.who.int>

poluentes no ar de ambientes fechados, como ozônio, monóxido de carbono, dióxido de carbono, dióxido de nitrogênio, *particulate matter*, e componentes orgânicos voláteis torna evidente o quão crítico a qualidade de ar é para a saúde humana.

Embora sua importância em um âmbito geral, a qualidade de ar possui peculiaridades em seu tratamento dependendo da região no qual monitora-se. Tendo isso em mente, esse trabalho visa explorar como esse problema é tratado nos ambientes fechados (*indoors*) das regiões urbanas.

Para um adequado monitoramento da qualidade de ar, é necessário a respectiva qualidade do sensoriamento, este que é avaliada por meio do campo chamado validação de sensores. Devido a inerente suscetibilidade à falhas presente nos sensores, diversos métodos para detectar esses problemas já foram propostos, como por exemplo abordagens estatísticas [Kerschen et al. 2004], redes neurais [Upadhyaya and Eryurek 1992] e até mesmo lógica *fuzzy* [Li et al. 2012].

Considerando o exposto até esse ponto, o trabalho tem como objetivo geral realizar uma Revisão Sistemática da Literatura (RSL) [Keele et al. 2007] dos principais estudos relacionados a sistemas de monitoramento que fazem uso de lógica *fuzzy* e aprendizado de máquina para realizar inferências.

O trabalho está organizado da seguinte forma. Na Seção 2 apresentamos a metodologia utilizada na RSL, que consiste na conceituação do que é uma RSL, as questões de pesquisa, quais foram as palavras-chave e *strings* de busca utilizadas, quais foram as bibliotecas digitais consultadas, e discutiremos os critérios de inclusão e exclusão. A análise dos artigos selecionados está discutida na Seção 3, onde serão descritos em maior detalhe os critérios utilizados para selecionar os trabalhos mais relevantes para a área. Por fim, apresentamos as conclusões na Seção 4, no qual será analisado como os trabalhos respondem às questões de pesquisa propostas.

## **2. Metodologia de Pesquisa**

Nessa seção iremos abordar os principais temas da revisão realizada. Inicialmente é apresentado uma visão geral sobre RSL, após, para cada passo realizado detalhamos as informações necessárias.

### **2.1. Revisão Sistemática da Literatura**

Este trabalho visa fazer uso da revisão sistemática como metodologia de pesquisa, que consiste em uma modalidade de segue protocolos específicos de busca com o intuito de conferir alguma logicidade a um grande corpus documental, trazendo uma prospecção criteriosa do tema de pesquisa abordado [Keele et al. 2007], bem como conceder um estudo facilmente reproduzível e evitando duplicação de pesquisas [Galvão and Ricarte 2019].

A revisão sistemática implementada nesse trabalho considera seguintes etapas:

1. Definição da questão de pesquisa
2. Escolha das palavras-chave e construção das strings de busca
3. Seleção das bibliotecas digitais
4. Formulação dos critérios de inclusão e exclusão de trabalhos
5. Definição das questões de avaliação de qualidade do trabalho
6. Construção do formulário de extração de dados

Cabe destacar por fim, que para auxílio da condução deste trabalho, foi utilizada uma ferramenta que auxilia no processo de RSL. Precisamente, consideramos o Parsifal<sup>2</sup>, um conhecido software que automatiza o processo de coletar e gerar informações sobre os estudos selecionados.

## 2.2. Questão de pesquisa, palavras-chave e string de busca

O primeiro passo de uma RSL compreende a definição de uma questão de pesquisa. Ela é responsável por nortear todo o estudo. Tendo esse ponto em questão, esse trabalho visa responder a seguinte Questão de Pesquisa (QP):

- Como sistemas sensoriais de controle de qualidade de ar fazem uso de métodos baseados em lógica fuzzy e em aprendizado de máquina?

Uma vez definida a questão de pesquisa, é preciso criar *queries* para buscar trabalhos nos repositórios. As palavras-chaves definidas foram as seguintes: Air Quality, Sensors, Machine Learning e Fuzzy.

Com base nessas palavras-chave, uma string de busca foi definida, com o intuito de responder a questão de pesquisa:

- “Sensors” AND “Air Quality” AND “Machine Learning” AND “Fuzzy”

## 2.3. Bibliotecas Digitais

Devido à reputação e reconhecimento no âmbito acadêmico, bem como a relevante quantidade de artigos científicos publicados, as seguintes bibliotecas digitais foram selecionadas para a realização da revisão sistemática:

- Springer Link - <https://link.springer.com>
- ACM Digital Library - <https://dl.acm.org>
- Scopus - <https://scopus.com>
- Science Direct - <https://sciencedirect.com>

Após a inserção da *string* de busca nas bibliotecas digitais, foram encontradas uma quantidade excessiva de resultados. Afim de remover resultados irrelevantes para o trabalho, alguns filtros foram implementados. Precisamente, para a biblioteca Springer Link, considerou-se apenas os artigos de conferência dentro da área da ciência da computação. ACM e Scopus tiveram os resultados limitados a apenas anais. Science Direct teve os resultados limitados a artigos de revisão ou pesquisa dentro da área da Ciência da Computação, com acesso aberto. Como filtro geral, foram selecionados apenas trabalhos pertencentes aos últimos cinco anos cujo idioma é o inglês.

## 2.4. Critérios de inclusão/exclusão

Para a seleção de artigos, um critério de inclusão (CI) foi definido:

- Trabalhos relacionados em *surveys* ou *reviews*, cujos tópicos estejam relacionados ao uso de Lógica Fuzzy ou Aprendizado de Máquina no contexto de sensoria-mento de qualidade de ar.

Para a remoção dos artigos, consideramos os seguintes Critérios de Exclusão (CE):

- **CE1** - Leitura dos títulos - Se o título do artigo apresenta relação com o tema.
- **CE2** - Leitura do abstract - Se os tópicos abordados são relevantes.
- **CE3** - Leitura da conclusão - Se as conclusões tiveram um impacto significativo.

---

<sup>2</sup><https://parsif.al>

## 2.5. Avaliação da qualidade e extração de dados

Para mensurar a qualidade dos trabalhos escolhidos após a aplicação dos critérios de inclusão/exclusão, foram definidas as seguintes perguntas:

1. O trabalho é referente a sensoriamento da qualidade de ar?
2. Faz uso de Lógica Fuzzy?
3. Faz uso de aprendizado de máquina?
4. O algoritmo proposto é reproduzível?
5. Possui dataset aberto?

Cada uma dessas perguntas possui uma resposta binária (sim ou não) e uma respectiva nota associada (0 ou 10). A nota final do trabalho corresponde a média dessas respostas.

Na extração de dados dos trabalhos selecionados, optou-se pelas questões:

1. Qual o principal algoritmo utilizado pelo trabalho?
2. Qual o tipo de modelo que este algoritmo se enquadra?
3. Os dados do trabalho são oriundos de onde?
4. Quais os componentes considerados para a simulação?

## 2.6. Condução da revisão

Uma vez que os passos adotados na RSL foram expostos, essa subseção explicita como ocorreu a condução da revisão. A primeira fase da revisão foi a aplicação da string de busca nas bibliotecas digitais selecionadas, que obteve um total de 181 artigos, cujo 8 foram escolhidos para a leitura integral, conforme descrito na Tabela 1.

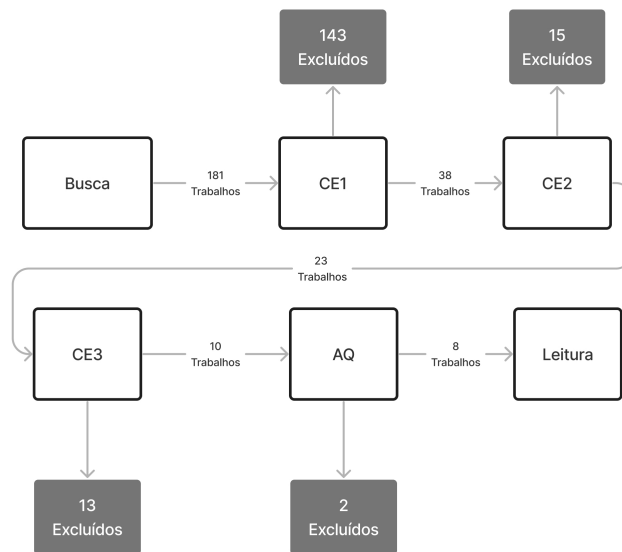
**Tabela 1. Trabalhos obtidos por biblioteca digital**

<b>Biblioteca Digital</b>	<b>Obtidos</b>	<b>Selecionados</b>
Springer Link	111	1
ACM Digital Library	23	3
Scopus	9	2
ScienceDirect	38	2
Total	181	8

A seleção destes oito trabalhos foi o resultado do processo da aplicação dos critérios de exclusão/inclusão, bem como a avaliação da qualidade dos artigos. Após a aplicação do Critério de exclusão 1, reduzimos o número de artigos para 38. Após o CE2, restaram 23 trabalhos. O CE3 reduziu mais uma vez o número para dez. Por fim, a avaliação da qualidade atribuiu uma nota de 0 a 50 para cada trabalho, e quaisquer um com nota menor do que 40 foi removido, o que resultou na eliminação de dois artigos, deixando oito para a etapa de leitura integral.

A busca dos artigos foi realizada até o dia 2 de Maio de 2023, através da busca avançada presente nas bibliotecas digitais. A Figura 1 ilustra o fluxograma dos passos que compuseram a execução desta RSL, desde a obtenção dos artigos até a sua leitura integral.

Após a leitura integral dos artigos, o processo de extração de dados foi conduzido, e o produto oriundo desse processo foi resumido na Tabela 2, no qual são relacionadas as principais características de cada trabalho.



**Figura 1. Fluxograma das etapas da RSL**

### 3. Relação dos Trabalhos sobre Sensoriamento de Qualidade de Ar

Esta seção discutirá os principais projetos que compõem a RSL, estes trabalhos que foram selecionados após a aplicação dos critérios de exclusão e inclusão, bem como o corte feito após a avaliação de qualidade.

Em [Alhasa et al. 2018] foi apresentada uma solução mirrada em sensores de baixo custo para qualidade de ar, por meio de um sistema de inferência Neuro-Fuzzy adaptativa. O trabalho demonstra a obtenção de uma alta taxa de correlação linear da calibração entre o sensor utilizado e o instrumento de referência. A performance de diversos modelos de calibração foram comparadas, tendo o método *Artificial Neural Fuzzy Inference System* (ANFIS) como o mais promissor dentre eles.

O trabalho feito em [Ferreira et al. 2022] propõe uma alternativa para predição de qualidade de ar usando rede neural *Fuzzy Adaptive Resonance Theory Map* (ARTMAP). O sistema provou ser uma boa alternativa na previsão de componentes de ar em ambientes fechados, sendo possível obter múltiplas previsões futuras usando esse método.

Uma abordagem utilizando *Deep Multi-task Learning* é adotada por [Wang et al. 2018], com o intuito de prever a qualidade de ar. Uma abordagem similar foi adotada por [Chen et al. 2019], no contexto do monitoramento de zonas urbanas. A primeira obra demonstra a superioridade comparada a modelos rasos e nove outras *baselines*, enquanto o segundo mostra que uma abordagem utilizando *Gated Recurrent Unit* (GRU) e *Long Short-Term Memory* (LSTM) é capaz de realizar uma predição confiável de até 24 horas.

Outro exemplo de sistema de inferência neuro-fuzzy adaptativa pode ser encontrado em [Bhardwaj and Pruthi 2020], porém o diferencial desse método é a utilização

**Tabela 2. Extração de Dados**

Article	Algorithm	Model Type	Data Origin	Compounds
1	Linear Regression	Classic	Gas Sensors	CO, CO <sub>2</sub> , NH <sub>3</sub> , (CH <sub>3</sub> ) <sub>2</sub> CO
2	ANFIS	Neuro Fuzzy	Sensors	PM <sub>2.5</sub>
3	ARTMAP	Neuro Fuzzy	Sensors	PM <sub>2.5</sub>
4	ANFIS	Neuro Fuzzy	Low Cost Sensors	O <sub>3</sub> , NO <sub>2</sub> , CO
5	Residual GRU	Deep Learning	Open Dataset	O <sub>3</sub> , NO <sub>2</sub> , PMs
6	PANDA	Deep Learning	AQ Station	Weather, AQI, POI
7	LSTM/GRU	Neural Network	Open Dataset	PM <sub>2.5</sub>
8	SARIMA and Prophet	Statistical	Open Dataset	PSO <sub>2</sub> , NO <sub>2</sub> , SPM, RSPM

**Artigos:** 1: [Kumar et al. 2020]; 2: [Bhardwaj and Pruthi 2020]; 3: [Ferreira et al. 2022]; 4: [Alhasa et al. 2018]; 5: [Wang et al. 2018]; 6: [Chen et al. 2019]; 7: [Wang et al. 2019]; 8: [Samal et al. 2019]

de uma abordagem evolutiva para superar o problema de ótimos locais, como Particle Swarm Optimization (PSO) e Genetic Algorithm (GA), que auxiliam na otimização dos parâmetros dos algoritmos neuro-fuzzy utilizados, algoritmos esses baseados no ANFIS.

Uma abordagem para a predição de qualidade de ar usando Recurrent Neural Networks (RNNs) é apresentado por [Wang et al. 2019], onde é utilizado um modelo baseado em Gated Recurrent Long Short-Term Memory (GRLSTM), que faz uso de redes neurais duplamente recursivas para a predição. Os resultados desse trabalho mostram uma boa predição, embora a acurácia não seja muito alta.

No contexto de predição de série temporal usando Internet of Things (IoT), temos [Kumar et al. 2020], que faz uso de um modelo linear em conjunto com um arranjo de sensores gerados por IoT, sendo capaz de prever a qualidade de ar do próximo dia.

Por fim, [Samal et al. 2019] faz uso dos modelos Seasonal Auto Regressive Integrated Moving Average (SARIMA), bem como Prophet, um modelo preditivo desenvolvido pelo Facebook, para atingir a predição de series temporais de qualidade de ar. Ambos os métodos proveem uma boa qualidade de acurácia, todavia a melhor abordagem é o modelo Prophet em transformação logarítmica, demonstrando as menores métricas de erros.

#### 4. Conclusão

Com o aumento da poluição, a qualidade de ar tornou-se uma preocupação cada vez maior em nossa sociedade, acarretando no interesse em desenvolver melhores soluções para otimizar mecanismos de sensoriamento. Esta revisão sistemática apresenta o atual estado da arte de algoritmos para otimização de sensoriamento de qualidade de ar. Dos 181 trabalhos encontrados, oito são discutidos em uma análise mais detalhada, elucidando as principais características de cada método.

Embora ainda haja alguns trabalhos recentes embasados em métodos clássicos e estatísticos, as duas principais abordagens destacadas nos trabalhos foram *Deep Learning* e *Neuro Fuzzy*, dos grupos centrais com referência à questão de pesquisa desse trabalho. Os métodos embasados em *deep learning* trazem como vantagem, além de uma boa performance em termos de predição, soluções capazes de modelar o problema considerando uma gama maior de fatores, como apresentado por [Chen et al. 2019], que faz uso de diversos outros atributos além das leituras dos sensores de qualidade de ar.

Os métodos baseados em *Neuro Fuzzy* elucidaram as vantagens proporcionadas pelo foco em expandir os parâmetros de entrada, fazendo uso de funções de pertinência, bem como transformações no espaço de dados para habilitar melhores resultados nas predições. Além disso, métodos *fuzzy* tem a inerente vantagem com relação a maneira como o conhecimento é representado, facilitando potencialmente a interpretabilidade dos resultados.

Como trabalho futuro seria interessante explorar com maior ênfase uma das áreas majoritárias presentes nesse trabalho. Uma revisão mais detalhada sobre os métodos focados em *deep learning* poderia aprofundar as comparações entre os métodos utilizados, destacando as nuances e particularidades de cada algoritmo, particularmente com o foco na performance. O mesmo poderia ser aplicado seguindo a direção *Neuro Fuzzy*, onde poderia-se explorar melhor os benefícios que a abordagem *Fuzzy* pode acrescentar à modelagem de qualidade de ar.

## Agradecimentos

Os autores agradecem as seguintes agências de fomento: CAPES, CNPq (309160/2019-7; 311429/2020-3, 3305805/2021-5, 150160/2023-2), PqG/FAPERGS (21/2551-0002057-1) e FAPERGS/CNPq (23/2551-0000126-8), PRONEX (16/2551-0000488-9).

## Referências

- Alhasa, K. M., Mohd Nadzir, M. S., Olalekan, P., Latif, M. T., Yusup, Y., Iqbal Faruque, M. R., Ahamad, F., Abd. Hamid, H. H., Aiyub, K., Md Ali, S. H., et al. (2018). Calibration model of a low-cost air quality sensor using an adaptive neuro-fuzzy inference system. *Sensors*, 18(12):4380.
- Bhardwaj, R. and Pruthi, D. (2020). Evolutionary techniques for optimizing air quality model. *Procedia Computer Science*, 167:1872–1879.
- Campbell, J. B. and Wynne, R. H. (2011). *Introduction to remote sensing*. Guilford Press.
- Chen, L., Ding, Y., Lyu, D., Liu, X., and Long, H. (2019). Deep multi-task learning based urban air quality index modelling. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(1):1–17.
- Ferreira, W. d. A. P., Grout, I., and da Silva, A. C. R. (2022). Application of a fuzzy art-map neural network for indoor air quality prediction. In *2022 International Electrical Engineering Congress (iEECON)*, pages 1–4. IEEE.
- Galvão, M. C. B. and Ricarte, I. L. M. (2019). Revisão sistemática da literatura: conceituação, produção e publicação. *Logeion: Filosofia da informação*, 6(1):57–73.
- Keele, S. et al. (2007). Guidelines for performing systematic literature reviews in software engineering. Technical report, Technical report, Ver. 2.3 EBSE Technical Report. EBSE.
- Kerschen, G., De Boe, P., Golinval, J.-C., and Worden, K. (2004). Sensor validation using principal component analysis. *Smart materials and structures*, 14(1):36.
- Kumar, R., Kumar, P., and Kumar, Y. (2020). Time series data prediction using iot and machine learning technique. *Procedia computer science*, 167:373–381.

- Kumari, C. U., Samiappan, D., Kumar, R., and Sudhakar, T. (2019). Fiber optic sensors in ocean observation: A comprehensive review. *Optik*, 179:351–360.
- Li, H., Monti, A., and Ponci, F. (2012). A fuzzy-based sensor validation strategy for ac motor drives. *IEEE Transactions on Industrial Informatics*, 8(4):839–848.
- Organization, W. H. (2016). *Ambient air pollution: a global assessment of exposure and burden of disease*. World Health Organization.
- Pandharipande, A. and Caicedo, D. (2015). Smart indoor lighting systems with luminaire-based sensing: A review of lighting control approaches. *Energy and Buildings*, 104:369–377.
- Samal, K. K. R., Babu, K. S., Das, S. K., and Acharaya, A. (2019). Time series based air pollution forecasting using sarima and prophet model. In *proceedings of the 2019 international conference on information technology and computer communications*, pages 80–85.
- Upadhyaya, B. R. and Eryurek, E. (1992). Application of neural networks for sensor validation and plant monitoring. *Nuclear Technology*, 97(2):170–176.
- Wang, B., Kong, W., and Guan, H. (2019). Air quality forecasting based on gated recurrent long short-term memory model. In *Proceedings of the ACM Turing Celebration Conference-China*, pages 1–9.
- Wang, B., Yan, Z., Lu, J., Zhang, G., and Li, T. (2018). Deep multi-task learning for air quality prediction. In *Neural Information Processing: 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13–16, 2018, Proceedings, Part V 25*, pages 93–103. Springer.
- Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., and Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature climate change*, 3(10):875–883.
- Zhang, H. and Srinivasan, R. (2020). A systematic review of air quality sensors, guidelines, and measurement studies for indoor air quality management. *Sustainability*, 12(21):9045.