ISBN: 978-628-95207-8-1. ISSN: 2414-6390. Digital Object Identifier: 10.18687/LACCEI2024.1.1.1652

ANN-based Prototype for the Prediction of CO_2 Pollution Levels: Eco-Logica

Guillermo Augusto Durán-Boneth[1], Rosanna Costaguta[2], and Dewar Rico-Bautista[1]
[1]*Universidad Francisco de Paula Santander, Ocaña, Colombia, {gaduranb, dwricob}@ufpso.edu.co*
[2]*Universidad Nacional de Santiago del Estero, Argentina, rosanna@unse.edu.ar*

Abstract*– *Eco-Logica is the development of a prototype based on artificial neural networks that allows the prediction and visualization of CO_2 pollution levels. The lack of control of CO_2 pollution levels produces impacts that negatively affect people's quality of life. For the development of the prototype, a research methodology with a quantitative approach is applied, which aims to analyze data associated with the pollution produced by carbon dioxide. The prototype, using regression techniques, makes predictions assuming pollution as a target variable, which is time dependent. Additionally, the neural network model is trained using datasets consulted from national government databases, whose information is freely accessible and usable. The information processed by the network allows us to build reports that are rendered graphically in the prototype, and thus monitor the pollution levels. To assess the quality of the predictions, the coefficient of determination, known as R-squared (R^2), is used, resulting in a value of 0.87117. From this, it can be concluded that the proposed model adequately describes the data's variability. Furthermore, cross-validation is performed using the standard deviation of R-squared, yielding a value of 0.0042, which is a positive indication that the model is not overfitting.

Keywords-- Artificial neural networks, AI prediction, Regression techniques, CO_2 pollution levels.

I. INTRODUCTION

Carbon dioxide emissions, known as CO_2, are one of the main factors responsible for global warming and its negative impact on air quality has serious consequences on people's quality of life [1], [2]. Therefore, it is vitally important to understand the behavior of CO_2 and how it affects our planet to take effective measures to reduce its impact [3]. The following is a brief overview of the impact of CO_2 on our planet.

The regulation and monitoring of pollution levels, and CO_2 emissions, has always been a major issue in today's societies [4], [5], [6]. As the growth of the world population continues to increase considerably, it brings with it a significant increase in the motorization of populations, especially in urban sectors [7], [8], [9], [10].

Considering that these emissions are one of the most frequent causes of decrease in air quality and cause of respiratory diseases [11], [12], [13], [14], the regulation and monitoring of emissions produced by internal combustion vehicles are both necessary and important. Despite the wide dissemination of information on pollution especially CO_2 pollution, there are still many regions of the world that lack monitoring measures to control pollution levels [15].

In addition, it is important to recognize that monitoring and regulating CO_2 emissions not only benefits the environment but can also have a positive impact on the economy and society [16]. For example, the promotion of cleaner and more sustainable technologies can create green jobs and reduce the long-term costs associated with environmental pollution [17], [18].

For the reasons set out in the previous paragraphs, governments have issued guidelines for organizations in different sectors to establish controls on the levels of CO_2 produced [19], [20].

The article is organized as follows. First, the introduction where the problem situation, the research question, and the justification are presented. Second, the methodology, where the design of the research process to obtain and test the proposed solution is presented, and the third part, the results, together with the respective discussion.

II. RELATED WORKS

The research technique used in this article in the literature review was a systematic mapping [21], [22] and further developed by [23], [24], [25], [26], [27], [28], [29], [30], [31]. It is divided into 4 steps (i) generation of research questions, (ii) selection of search criteria for primary studies, (iii) inclusion and exclusion criteria, and (iv) data extraction and study mapping, see Fig. 1.

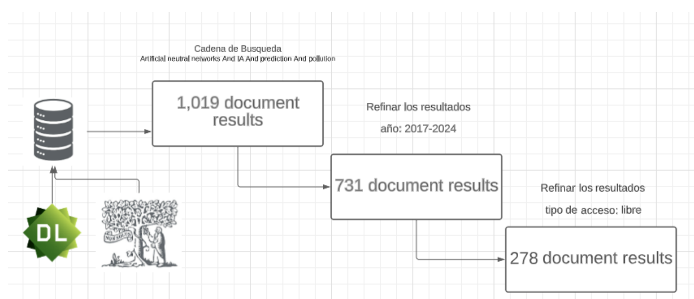

Fig. 1 Systematic mapping.

Dybå et al. [1], provide a list of important databases in the field of computer science. The databases with the most important conferences and journals were selected. ACM Digital Library, IEEE Xplore, ScienceDirect, and Scopus were selected; each of them uses a search engine with different and varied mechanisms.

Using the use of diverse literature bases we were able to search with our search chain with the keywords once having the results we proceeded to refine these searches to be able to arrive at the most updated results in the subject, and that possess free access for this form to guide or to orient the solution of the raised problem. Going from a chain of data with 1019 and once limiting the time of 7 years of the antiquity of the data and that possess free access to the information.

The literature search and analysis that provides the scientific basis for this research was carried out using a specific set of words, also known as keywords, which facilitate the identification of relevant studies in the context of contamination using AI.

A. Main keywords

- "Artificial Neural Network (ANN): Studies are identified that focus on the use and application of artificial neural networks, studies that are key to the purpose of this research.
- "AI Prediction": The literature search using this term allows us to explore and learn about different AI techniques applied in environmental variable contexts, CO2 being one of them, thus providing a broader view of different predictive approaches.
- "Pollution Prediction": Finally, "pollution prediction" identifies studies focused on pollution prediction, highlighting the relevance of addressing only predictive models in environmental monitoring.

B. Keyword Combination:

- "Artificial Neural Networks" AND "IA Prediction": This search provided studies that determine ANN as the focus.
- "IA Prediction" AND "Pollution": The combination of "IA Prediction" and "Pollution" provided different investigations that directly address the relationship between pollutants, emphasizing environmental variables and predictions with artificial intelligence.
- "Artificial Neural Networks" AND "Pollution Prediction": This search string allowed filtering those research studies where the use and application of ANNs are highlighted accurately to predict pollution levels.

C. Relevance of recent references in research

During the process of reviewing the state of the art, it was decided to filter research that had been published recently, prioritizing those references that ranged between 2019 and 2024, to ensure that the knowledge consulted was up to date and relevant to the current state of the problem.

D. Practical Applicability and Current Methodologies.

Recent research claims to be able to analyze research methodologies and approaches, bearing in mind that these two aspects may change over time.

Considering the approach of prioritizing recently published research, the result was that more than 88% of the referents meet this condition, thus contributing to the quality and relevance of the research, see Table I.

TABLE I
YEAR OF PUBLICATION OF THE STATE OF THE ART

YEAR	QUANTITY	%
2024	6	9.83%
2023	23	37.70%
2022	8	13.11%
2021	8	13.11%
2020	7	11.47%
2019	2	3.27%
< 2019	7	11.47%

III. METHODOLOGY

To predict contamination levels during the research, data captured from the Colombian National Government databases, whose information is freely accessible and usable, were used. Additionally, for the interpretation of the information captured from the repository, consulted literature is used to establish a criterion for classification and evaluation of the data. The technique used is Regression [32], considering that the main purpose of this research is to be able to make a prediction or estimate of a numerical value, which represents in terms of Artificial Intelligence, the target variable. The results provide an estimate of CO_2 pollution levels as a function of time [33], [34]. By clearly understanding the objective of the research, it is possible to identify the key factors and the relationship between them (CO_2 as a function of time).

For the construction of the ANN-based prototype, the following phases were established:

1) Define the objective: The central objective of the prototype is established, to achieve the estimation and monitoring of CO_2 levels.
2) Data capture: To carry out the research with real data, an API capable of establishing a connection with the national government databases is built to query such data.
3) Data analysis and preparation: The consulted data must be subjected to an analysis stage, in which an attempt is made to discard information not relevant to the research and to determine possible adjustments to it.
4) ANN model design: The design of the neural network, concerning the hyperparameters, is based on the quantity and quality of the data [35], [36].
5) ANN training: Key phase for the design of the model. Training and tests on the designed model are established to identify possible over-adjustments.
6) Model evaluation: The performance of the proposed model is evaluated, for which the most appropriate technique must be defined considering the characteristics of the data and the network [37], [38].

7) Reports: It identifies which are the reports whose implementation in the prototype contributes in a better way to reading and interpreting in a better way the consulted data, covering in this way the monitoring of the same.
8) Integration: The respective integration between the model and the proposed reports that will fulfill the prediction and monitoring functions is carried out.

A. Applied technique for prediction.

In the field of deep learning, different types of Neural Network models can be found, such as Artificial Neural Networks (with an emphasis on the Multilayer Perceptron model, MLP), Recurrent Neural Networks (RNN), and Convolutional Neural Networks (CNN). Each of these models has its advantages, depending on the characteristics of the problem at hand. However, CNN and RNN tend to be even more complex models. CNN and RNN have gained dominance in classification tasks, commonly in image classification tasks [46].

On the other hand, MLP is often more commonly used in regression tasks rather than classification, unlike CNN and RNN. Although the latter two models can also perform regression, the complexity of MLP models varies depending on the dataset's specific features. An important factor to consider is the training time, in which MLP has a significant advantage as it largely depends on the number of configured neurons [47].

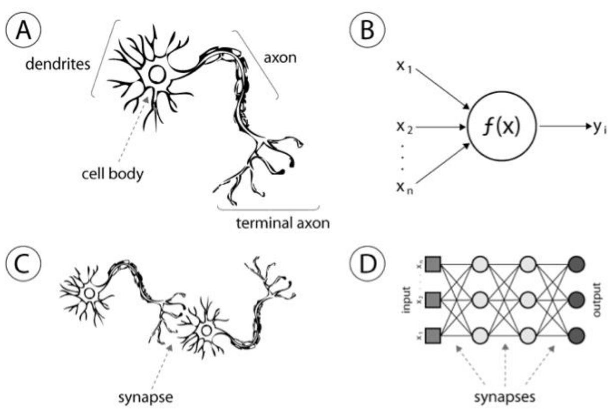

Fig. 2 Biological Neuron and Artificial Neural Network.

However, given the purpose of the current problem, which is to predict pollution levels, meaning predicting or estimating any numerical value, it is necessary to use a model that generates good results based on these characteristics. Therefore, a model with significant advancements in classification would not be the best choice for a regression task. Due to these considerations, it has been determined that MLP is the best choice in this case, supported by multiple studies concluding that MLP is the model that yields the best results in estimating numerical values [48]. It's also worth mentioning that the dataset being worked with does not have a significant level of complexity, and the number of variables it contains further allows for simplification in the model development.

An artificial neural network (ANN) is based on processing units that interconnect tightly, which are called neurons since they mimic the behavior of a biological neuron. These neurons, or processing units, are also capable of processing and transmitting signals, just like biological neurons.

Below, a biological neuron (A, C) and a representation of an artificial neural network (B, D) are illustrated and detailed, see Fig. 2.

A simple Multilayer Perceptron (MLP) model is proposed. The hyperparameters have been selected based on the data complexity and the nature of the task at hand, and are described as follows:

- **Hidden Layers (hidden_layer_sizes)**: The model consists of a single hidden layer with 29 neurons, as the data does not exhibit a high degree of complexity.
- **Activation Function (activation)**: The activation function used in all layers, including the hidden layer, is ReLU (Rectified Linear Unit). ReLU tends to mitigate the vanishing gradient problem, which can allow for faster and more stable network training.
- **Solver or Optimizer (solver)**: The optimizer used is 'lbfgs' (Limited-memory Broyden-Fletcher-Goldfarb-Shanno). lbfgs has fewer hyperparameters to tune compared to some other optimizers, simplifying the hyperparameter selection process.
- **Maximum Number of Iterations (max_iter)**: Model training will halt after 500 iterations.
- **Random State (random_state)**: The "random_state" value is set to 115 based on observed behavior during training sessions.
- **Regularization (alpha)**: L2 regularization (also known as Ridge regularization) is applied with an alpha value of 0.046 to prevent overfitting.

As an optimization method, also known as a solver, 'lbfgs' is set, along with a maximum limit of 500 iterations to guarantee a reasonable training time, thus avoiding possible overfitting. Finally, the model has been trained with the previously prepared data, to minimize the error between the model predictions and the real values of CO_2 levels. On the other hand, to split the dataset (X and y) into training and testing sets, there has been an alternating between 10% and 30% for testing, with the remainder for training, resulting in assigning 10% of the data for testing and the remaining 90% for training, which has yielded better performance, see Fig. 3.

```
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.1, random_state=9)

mlp = MLPRegressor(hidden_layer_sizes=(55), activation='relu',
                   solver='lbfgs', max_iter=500, random_state=115, alpha=0.046)
```

Fig. 3 Hyperparameters configured.

The training sessions were conducted in the collaborative development environment Google Colab, as it offers free access to computational resources, such as GPUs (Graphics Processing Units), which are beneficial for machine learning tasks.

During the training period, more advanced techniques were employed, such as Regularization, which aims to prevent overfitting and enhance the model's generalization capability. Additionally, simpler approaches were used, such as iterating over the model while alternating the random state value, in pursuit of obtaining the optimal behavior for this hyperparameter.

The sample selected for this research is obtained from the Colombian National Government databases. Based on this data, the training and validation of the proposed ANN regression model are carried out. The quality and representativeness of the sample of records depends on the completeness and accuracy of the data provided by the state; for this same reason, the size of the sample also depends on the data that have been reported to the national government. On the other hand, the quality and adequacy of the sample size can be determined by the influence it has on the performance of the model through techniques such as cross-validation. To determine whether the influence of the sample is adequate, the model must be reliable, and for this, it must be satisfactory with statistical evaluations, such as, the importance that the model does not tend towards randomness or overfitting, which can be corroborated through the interpretation of the results given by the cross-validation [39]

The cross-validation approach is generally one of the most reliable among the different validation techniques, and it consists of randomly dividing the research dataset into some subsets that are called k-fold, parts of the subsets are used for testing, and the remaining for training. Finally, the number of subsets(k) determines the number of iterations, where in each cycle mathematical calculations are applied to each subset. The value of the parameter k can be any positive integer value, and the most common choice for practical purposes is usually k=5, which represents the above definition [40], see Fig. 4.

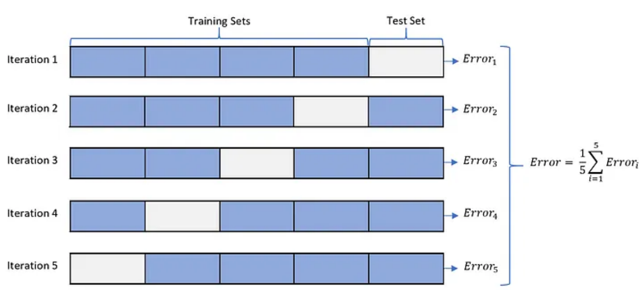

Fig. 4 Cross Validation: Source: [40].

IV. RESULTS

An analysis is carried out to determine the tools or techniques that would allow a clear and precise visualization of the data already obtained through the API, thus covering as a first objective, to monitor the data. It is important to mention that this type of report is aimed at people who know the subject since the goal is to implement reports that give added value to the data in question.

A. *Histograms to show recorded average contamination.*

A histogram-type report is proposed, being a clear and accurate data visualization technique, specifically when it is desired to carry out historical data readings. It allows us to evaluate the behavior of contamination levels and to make respective comparisons between different periods. Additionally, it generates the possibility of being able to identify peaks in pollution levels for each of the sites that report CO_2 data, see Fig. 5.

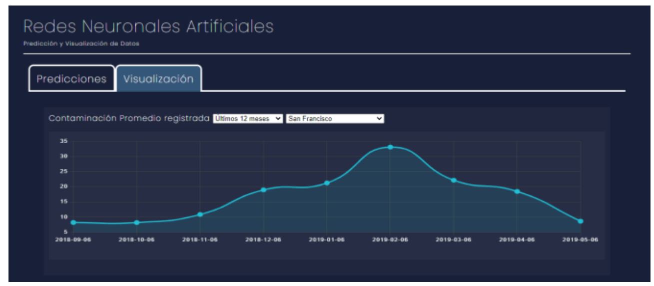

Fig. 5 Example of a histogram report about CO_2 data.

B. *Pie charts to show average contamination per station.*

To read the pollution levels of all stations over the last month, it is necessary to identify a technique that provides clarity and reflects the information accurately. The pie chart allows us to visualize the information understandably and to make comparisons between all the stations simultaneously, which generates added value to the information. Additionally, it allows to discriminate between different data sets for an even more specialized visualization, see Fig. 6.

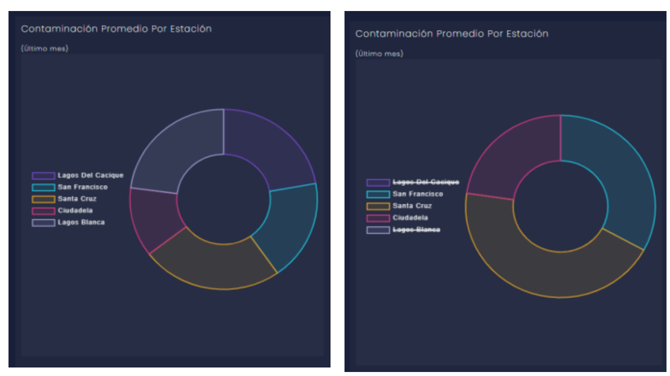

Fig. 6 Example of a pie chart report to make comparisons.

C. *Bubble chart to show the relation between CO_2 and humidity.*

Considering that other factors facilitate the understanding of the data, it is determined that humidity is a key factor that should be considered during the monitoring process. A bubble graph is proposed to corroborate that the CO_2 contamination data are consistent concerning the humidity readings made by each of the stations, see Fig. 7.

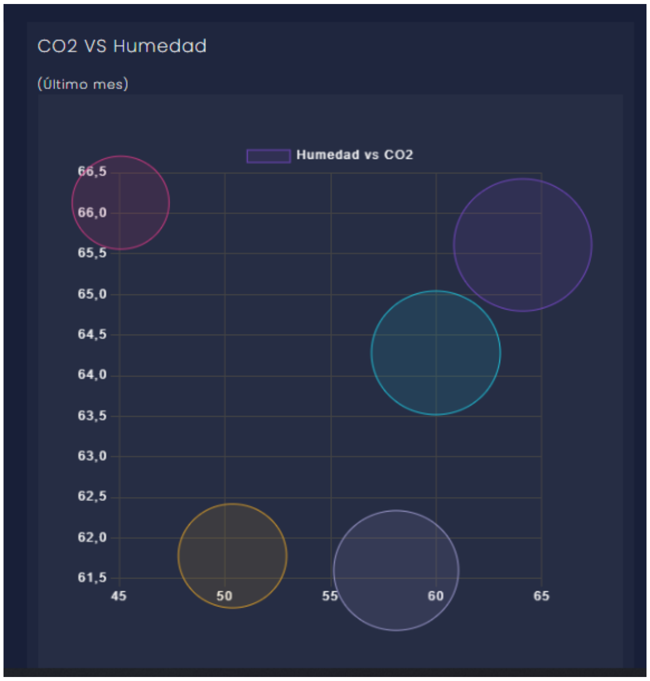

Fig. 7 Example of a bubble chart report about values of CO_2 and humidity.

D. Gauge chart to show the predictions.

To implement the proposed ANN model, it is necessary to propose a tool that allows a simple interaction with the model and that, at the same time, is capable of adequately reflecting the predictions made, this being the most important indicator.

Gauges, or gauge charts, are a powerful tool that facilitates the understanding and interpretation of the predictions generated by ANN. In our case, the gauge chart shows a specific measurement of its reference limits, providing a representation of the estimated pollution levels, see Fig. 8.

Fig. 8 Example of a gauge chart report to show the final prediction.

V. Conclusions

In general, the results of this research are positive; the correct approach and development of each of the objectives set out have effectively guided each of the proposed activities.

The application of Artificial Neural Networks (ANNs) has proven to be very useful in the prediction of pollutant emissions, particularly CO_2. These neural networks are capable of processing large amounts of data quickly and efficiently, allowing the identification of complex patterns and the prediction of future trends. Moreover, the advantage of ANNs is their ability to learn through feedback on the results, which allows them to improve their accuracy and performance. Around the same perspective, several research has been carried out that have employed the use of machine learning as a key tool to address the contamination problem. This has allowed the integration of new technologies in the monitoring and control of pollutants, which has significantly improved the efficiency and accuracy in the detection of emissions and the assessment of environmental impact.

This is why the implementation of this technique in the proposed problem generates an added value in the technological solution, giving way to further research to deepen all the benefits that it can provide. In summary, the use of Artificial Neural Networks is presented as an essential tool in the fight against air pollution and CO_2 emissions, allowing a more accurate and effective prediction of pollutant levels and helping to take preventive measures to reduce their impact on human health and the environment.

The exhaustive analytical study of the state of the art, focused on the use of AI techniques with special emphasis on ANN, has allowed the construction of solid bases that support the Eco-Logic proposal. The selection and use of keywords, in strategic combinations of each of the terms, together with the relevance of recent references, made it possible to approach the problem from an updated and pertinent perspective.

On the other hand, the construction of the ANN-based prototype for monitoring and predicting CO_2 pollution levels has culminated in a successful process. Considering that, the implementation of the OSINT technique for the search and use of open sources of information, together with the appropriate selection of the type of ANN, led to the development of a functional prototype, with different tools and specialized reports.

Likewise, the quantitative evaluation with the use of the coefficient of determination (R^2), properly applied to the ANN model developed, clearly indicated a positive result of 86.49%, thus understanding that the model can explain more than 85% of the CO_2 pollution data. Furthermore, this value provides a solid indicator of the ability acquired by the prototype to predict CO_2 pollution levels.

In short, the identification of different types of neural networks, the application of techniques for open-source data collection, and the rigorous validations applied in real contexts suggest significant contributions throughout this research. The evident potential of the prototype built for future improvements and applications in more case studies associated with environmental monitoring is highlighted, which has been supported by all the traceability made to its effectiveness in predicting and monitoring CO_2 pollution levels.

When trying to apply the prototype in other case studies, for example in other regions of the country, it is recommended to keep the same dimension of the information. Keep within the data set supplied to the ANN the same environmental variables with which it has been trained, such as humidity and CO_2 contamination, among others. This facilitates the generalization of the prototype to different study cases and its applicability.

On the other hand, it is suggested the implementation of a strategy that allows a continuous updating of the data, to guarantee the relevance of the predictions made by the prototype in the long term. Therefore, the incorporation of updated data can help the model to identify new variations in the reported contamination levels.

Also, considering the geographic variety of the Colombian territory, it is recommended to perform validations using the coefficient of determination (R^2) to ensure the correct effectiveness of the prototype about the new data provided.

REFERENCES

[1] R. R. Moussa, "Reducing carbon emissions in Egyptian roads through improving the streets quality," *Environ Dev Sustain*, vol. 25, no. 5, pp. 4765–4786, May 2023, doi: 10.1007/s10668-022-02150-8.

[2] S. Davoodi, H. Vo Thanh, D. A. Wood, M. Mehrad, V. S. Rukavishnikov, and Z. Dai, "Machine-learning predictions of solubility and residual trapping indexes of carbon dioxide from global geological storage sites," *Expert Syst Appl*, vol. 222, p. 119796, Jul. 2023, doi: 10.1016/j.eswa.2023.119796.

[3] D. Castells-Quintana, E. Dienesch, and M. Krause, "Air pollution in an urban world: A global view on density, cities and emissions," *Ecological Economics*, vol. 189, p. 107153, Nov. 2021, doi: 10.1016/j.ecolecon.2021.107153.

[4] T.-L. Chen, C.-H. Lai, Y.-C. Chen, Y.-H. Ho, A. Y. Chen, and T.-C. Hsiao, "Source-oriented risk and lung-deposited surface area (LDSA) of ultrafine particles in a Southeast Asia urban area," *Science of The Total Environment*, vol. 870, p. 161733, Apr. 2023, doi: 10.1016/j.scitotenv.2023.161733.

[5] M. Bilgili, A. Ozbek, A. Yildirim, and E. Simsek, "Artificial neural network approach for monthly air temperature estimations and maps," *J Atmos Sol Terr Phys*, vol. 242, p. 106000, Jan. 2023, doi: 10.1016/j.jastp.2022.106000.

[6] M. C. B. Franco, P. J. G. Delgado, and J. J. V. Moreno, "Evaluaciön De La Calidad Del Aire En El Casco Urbano Del Municipio De Vijes – Valle," in *2019 Congreso Colombiano y Conferencia Internacional de Calidad de Aire y Salud Pública (CASP)*, IEEE, Aug. 2019, pp. 1–5. doi: 10.1109/CASAP48673.2019.9364073.

[7] M. Soruma and M. Woldeamanuel, "The level of air quality at public transport stations: The Case of Torhailoch-Ayat main road in Addis Ababa," *J Transp Health*, vol. 24, p. 101328, Mar. 2022, doi: 10.1016/j.jth.2021.101328.

[8] M. M. Sepadi and V. Nkosi, "Health Risk Assessment of Informal Food Vendors: A Comparative Study in Johannesburg, South Africa," *Int J Environ Res Public Health*, vol. 20, no. 3, p. 2736, Feb. 2023, doi: 10.3390/ijerph20032736.

[9] A. F. Scagliotti, D. H. Margarit, M. V. Reale, and G. A. Jorge, "Influence of settings and predictors in neural network model performance: a Buenos Aires air quality case," *Procedia Comput Sci*, vol. 212, pp. 348–357, 2022, doi: 10.1016/j.procs.2022.11.019.

[10] L. A. Rodríguez-Camargo, R. J. Sierra-Parada, and L. C. Blanco-Becerra, "Análisis espacial de las concentraciones de PM2,5 en Bogotá según los valores de las guías de la calidad del aire de la Organización Mundial de la Salud para enfermedades cardiopulmonares, 2014-2015," *Biomédica*, vol. 40, no. 1, pp. 137–152, Mar. 2020, doi: 10.7705/biomedica.4719.

[11] W. Xie, A. Chapman, and T. Yan, "Do Environmental Regulations Facilitate a Low-Carbon Transformation in China's Resource-Based Cities?," *Int J Environ Res Public Health*, vol. 20, no. 5, p. 4502, Mar. 2023, doi: 10.3390/ijerph20054502.

[12] C. Tao *et al.*, "Time-sensitive prediction of NO2 concentration in china using an ensemble machine learning model from multi-source data," *Journal of Environmental Sciences*, vol. 137, pp. 30–40, Mar. 2024, doi: 10.1016/j.jes.2023.02.026.

[13] Y. Tadokoro *et al.*, "Artificial-intelligence-assisted mass fabrication of nanocantilevers from randomly positioned single carbon nanotubes," *Microsyst Nanoeng*, vol. 9, no. 1, p. 32, Mar. 2023, doi: 10.1038/s41378-023-00507-1.

[14] G. C. Spyropoulos, P. T. Nastos, K. P. Moustris, and K. J. Chalvatzis, "Transportation and Air Quality Perspectives and Projections in a Mediterranean Country, the Case of Greece," *Land (Basel)*, vol. 11, no. 2, p. 152, Jan. 2022, doi: 10.3390/land11020152.

[15] D. Majumdar *et al.*, "Managing future air quality in megacities: Emission inventory and scenario analysis for the Kolkata Metropolitan City, India," *Atmos Environ*, vol. 222, p. 117135, Feb. 2020, doi: 10.1016/j.atmosenv.2019.117135.

[16] Y. Liu, B. Huang, H. Guo, and J. Liu, "A big data approach to assess progress towards Sustainable Development Goals for cities of varying sizes," *Commun Earth Environ*, vol. 4, no. 1, p. 66, Mar. 2023, doi: 10.1038/s43247-023-00730-8.

[17] Q. Zhuang *et al.*, "Impact of global urban expansion on the terrestrial vegetation carbon sequestration capacity," *Science of The Total Environment*, vol. 879, p. 163074, Jun. 2023, doi: 10.1016/j.scitotenv.2023.163074.

[18] H. Zhang *et al.*, "Improving predictions of shale wettability using advanced machine learning techniques and nature-inspired methods: Implications for carbon capture utilization and storage," *Science of The Total Environment*, vol. 877, p. 162944, Jun. 2023, doi: 10.1016/j.scitotenv.2023.162944.

[19] L. A. Rodríguez-Camargo, R. J. Sierra-Parada, and L. C. Blanco-Becerra, "Análisis espacial de las concentraciones de PM2,5 en Bogotá según los valores de las guías de la calidad del aire de la Organización Mundial de la Salud para enfermedades cardiopulmonares, 2014-2015," *Biomédica*, vol. 40, no. 1, pp. 137–152, Mar. 2020, doi: 10.7705/biomedica.4719.

[20] J. Ren *et al.*, "A field study of CO2 and particulate matter characteristics during the transition season in the subway system in Tianjin, China," *Energy Build*, vol. 254, p. 111620, Jan. 2022, doi: 10.1016/j.enbuild.2021.111620.

[21] K. Petersen, S. Vakkalanka, and L. Kuzniarz, "Guidelines for conducting systematic mapping studies in software engineering: An update," *Inf Softw Technol*, vol. 64, pp. 1–18, Aug. 2015, doi: 10.1016/j.infsof.2015.03.007.

[22] K. Petersen, R. Feldt, S. Mujtaba, and M. Mattsson, "Systematic Mapping Studies in Software Engineering," Jun. 2008. doi: 10.14236/ewic/EASE2008.8.

[23] D. Rico-Bautista, G. Maestre-Gongora, and C. D. Guerrero, "Smart University:IoT adoption model," in *2020 Fourth World Conference on Smart Trends in Systems, Security and Sustainability (WorldS4)*, IEEE, Jul. 2020, pp. 821–826. doi: 10.1109/WorldS450073.2020.9210369.

[24] D. Rico-Bautista, Y. Medina-Cardenas, Y. Areniz-Arevalo, E. Barrientos-Avendano, G. Maestre-Gongora, and C. D. Guerrero, "Smart University: Big Data adoption model," in *2020 9th International Conference On Software Process Improvement (CIMPS)*, IEEE, Oct. 2020, pp. 52–60. doi: 10.1109/CIMPS52057.2020.9390151.

[25] D. Rico-Bautista *et al.*, "Smart University: a vision of technology adoption," *Revista Colombiana de Computación*, vol. 22, no. 1, pp. 44–55, Jun. 2021, doi: 10.29375/25392115.4153.

[26] D. Rico-Bautista, "Conceptual framework for smart university," *J Phys Conf Ser*, 2019.

[27] D. Rico-Bautista *et al.*, "Smart university: Strategic map since the adoption of technology," *RISTI - Revista Iberica de Sistemas e Tecnologias de Informacao*, vol. 2020, no. E28, 2020.

[28] D. Rico-Bautista, Y. Medina-Cardenas, L. A. Coronel-Rojas, F. Cuesta-Quintero, G. Maestre-Gongora, and C. D. Guerrero, "Smart University: Key Factors for an Artificial Intelligence Adoption Model," 2021, pp. 153–166. doi: 10.1007/978-981-33-4565-2_10.

[29] D. Rico-Bautista *et al.*, "Smart University: Key Factors for a Cloud Computing Adoption Model," *Lecture Notes in Networks and*

Systems, vol. 334, pp. 85–93, 2022, doi: 10.1007/978-981-16-6369-7_8.

[30] D. Rico-Bautista *et al.*, *Bibliometric Analysis on the Smart University Concept*, vol. 579. 2023. doi: 10.1007/978-981-19-7663-6_14.

[31] D. Rico-Bautista, C. A. Collazos, C. D. Guerrero, G. Maestre-Gongora, and Y. Medina-Cárdenas, "Latin American Smart University: Key Factors for a User-Centered Smart Technology Adoption Model," in *Sustainable Intelligent Systems*, 2021, pp. 161–173. doi: 10.1007/978-981-33-4901-8_10.

[32] S. Pandya, H. Ghyvat, K. Kotecha, and P. Gope, "Linear Regression and Artificial Neural Network (ANN)-based Approaches to Predict Air Pollution," in *Encyclopedia of Sensors and Biosensors*, Elsevier, 2023, pp. 497–511. doi: 10.1016/B978-0-12-822548-6.00073-X.

[33] H.-J. Kim and Y.-H. Cho, "Optimization of supply air flow and temperature for VAV terminal unit by artificial neural network," *Case Studies in Thermal Engineering*, vol. 40, p. 102511, Dec. 2022, doi: 10.1016/j.csite.2022.102511.

[34] H. Jian, "Assessment of hydroelectric potential under climate change and hydrological parameters based on soft-computing: A case study," *Energy Reports*, vol. 9, pp. 3035–3047, Dec. 2023, doi: 10.1016/j.egyr.2022.11.197.

[35] Q. Zhuang *et al.*, "Impact of global urban expansion on the terrestrial vegetation carbon sequestration capacity," *Science of The Total Environment*, vol. 879, p. 163074, Jun. 2023, doi: 10.1016/j.scitotenv.2023.163074.

[36] M. A. Zaidan *et al.*, "Intelligent Calibration and Virtual Sensing for Integrated Low-Cost Air Quality Sensors," *IEEE Sens J*, vol. 20, no. 22, pp. 13638–13652, Nov. 2020, doi: 10.1109/JSEN.2020.3010316.

[37] H. Alzahrani, M. Arif, A. K. Kaushik, M. Q. Rana, and H. M. Aburas, "Evaluating the effects of indoor air quality on teacher performance using artificial neural network," *Journal of Engineering, Design and Technology*, vol. 21, no. 2, pp. 604–618, Mar. 2023, doi: 10.1108/JEDT-07-2021-0372.

[38] F. Carotenuto *et al.*, "Long-Term Performance Assessment of Low-Cost Atmospheric Sensors in the Arctic Environment," *Sensors*, vol. 20, no. 7, p. 1919, Mar. 2020, doi: 10.3390/s20071919.

[39] E. Lopez, J. Etxebarria-Elezgarai, J. M. Amigo, and A. Seifert, "The importance of choosing a proper validation strategy in predictive models. A tutorial with real examples," *Anal Chim Acta*, vol. 1275, p. 341532, Sep. 2023, doi: 10.1016/j.aca.2023.341532.

[40] K. Demertzis, K. Kostinakis, K. Morfidis, and L. Iliadis, "An interpretable machine learning method for the prediction of R/C buildings' seismic response," *Journal of Building Engineering*, vol. 63, p. 105493, Jan. 2023, doi: 10.1016/j.jobe.2022.105493.