



Performance Evaluation of ANN Models for Prediction

Mohammad Khrisat^{1*}, Ziad Alqadi²

¹ Department of Mechatronic Engineering, Faculty of Engineering Technology, Balqa Applied University, Amman 15008, Jordan

² Department of Electrical Engineering, Faculty of Engineering Technology, Balqa Applied University, Amman 15008, Jordan

* Correspondence: Mohammad Khrisat (mkhrisat@bau.edu.jo)

Received: 01-02-2023

Revised: 02-13-2023

Accepted: 03-06-2023

Citation: M. Khrisat and Z. Alqadi, "Performance evaluation of ANN models for prediction," *Acadlore Trans. Mach. Learn.*, vol. 2, no. 1, pp. 13-20, 2023. <https://doi.org/10.56578/ataiml020102>.



© 2023 by the authors. Published by Acadlore Publishing Services Limited, Hong Kong. This article is available for free download and can be reused and cited, provided that the original published version is credited, under the CC BY 4.0 license.

Abstract: One of the biggest problems that humans are faced with today is pollution and climate change. Pollution is not a new phenomenon and remains a leading cause of diseases and deaths. Mining, industrialization, exploration and urbanization caused global pollution, whose burdens are shared by developed and undeveloped countries alike. Awareness and stricter laws in the developed countries have contributed to environmental protection. Although all countries have paid attention to pollution, the impact and severity of its long-term consequences are being felt. There is a cause-and-effect link between the pollution of air, water and soil and the environment. This research aimed to prove that the main function of the philosophy of science is to have a functional understanding of knowledge, which views knowledge as a tool for prediction. Prediction is the function or mission of science or the goal that must be achieved if the scientific project is successful. In other words, prediction is the final harvest of description and interpretation. In addition, science is primarily concerned with the prediction of events that have occurred in the universe. A mature prediction is what science provides to validate scientific models. This paper introduced the concepts of using machine learning techniques to enhance the prediction process results. Pollution data set and the negative effects of polluted air data were used. We built, trained and tested various models in order to find the optimal model, which could enhance the results of the prediction process.

Keywords: Pollution; Prediction methods; Artificial neural networks; Pollution measurement regression analysis

1. Introduction

Pollution, also called environmental pollution, is the presence of harmful substances in the environment that has a poisonous effect. These harmful substances are called pollutants, which may cause health problems to humans and creatures.

In a scientific way, one might define pollution as the emergence of substances, which may surround living organisms and eventually will cause harm to them [1].

One of the most useful fertilizers for soil is animal waste. However, on the other hand, if the animal waste was disposed in drains, it would lead to diseases and epidemics [2].

It is known that the primary cause of pollution is humankind. Pollution explosion is blamed on a list of reasons, including industrial revolution, technological advancements, misuse of natural resources, and population expansions. Pollution is responsible for a lot of deaths. It was stated that pollution caused 50,000 deaths annually, accounting for about 2% of the total [3, 4].

Tobacco or cigarette smoke kill about three million people annually. If smoking continues to exist, the deaths will increase to 10 million annually [5].

Air pollution is the presence of harmful substances in the air, which can be classified into two parts [6, 7]. The first part is natural sources of pollution, such as dust. The second part includes industrial sources, car emissions, and fuel-driven electric power generators, which emit large amounts of harmful fine particles and gases into the air [8].

Large industrial cities are one of the most affected areas of air pollution. In addition, the developed countries do not have the capabilities to eliminate pollution. Some of the elements that cause air pollution are shown in Figure 1 [9] and discussed in the list below.

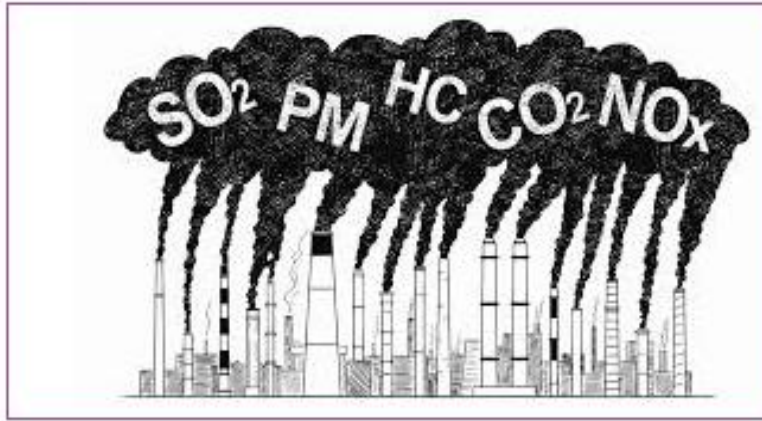


Figure 1. Air pollution

- 1) Fine particles, generated by petroleum-driven vehicles, factories that release unfiltered smoke of chemicals, and desert dust.
- 2) Carbon dioxide, with factories as its primary source.
- 3) Nitrogen dioxide, generated from burning fuels.
- 4) Ozone, a gas composed of three atoms of oxygen, exists both in the upper atmosphere of the earth and on the ground. For ozone at the ground level, it's harmful to people and the environment, and is the main ingredient in "smog" [10].
- 5) Carbon monoxide, generated from automobile emissions in highly populated cities, bush fires and volcanoes.
- 6) Cigarette smoke, which produces fine Particulate Matter (PM), is the most dangerous element of air pollution for health (Medical News Today).
- 7) Lead, the most toxic metal, is considered a major pollutant. Lead content exceeds 64,000 parts per million in the dust of houses and exceeds 3,000 parts per million in the air of streets.

Polluted air means the surrounding air has large quantities of pollutants, causing harm to humans, animals, plants or materials. Nowadays air pollution is a big killer. In 2015, pollution caused 6.4 million deaths globally, with 2.8 million deaths of them caused by household air pollution [11]. Data for the year 2015 below shows the percentages in global deaths caused by air pollution [11]:

- 19% of all cardiovascular deaths
- 24% of ischemic heart disease deaths
- 21% of stroke deaths
- 23% of lung cancer deaths

In addition, the polluted air caused neurodevelopmental disorder in children [12, 13] and neurodegenerative diseases in adults [14].

With large economic and human losses caused by air pollution, the environment predictive and forecasting technology is needed.

2. Regression Analysis Model

Air pollution must be put under control. Or by 2030 the air will become so polluted that it will be necessary to use an oxygen kit to breathe easily. Increasingly serious air pollution will also lead to premature negative effects. Human exposure to air toxins will increase to a large extent if air pollution is not controlled. Analysis of historical data sets showed that Pollution Parameter (PM) levels were associated with negative effects. This paper used an equation to forecast future target levels, with Regression Analysis Model (RAM) and Artificial Neural Networks (ANNs) with deep learning as efficient prediction tools and techniques [15, 16].

The RAM was used to formulate the association between the PMs and the negative pollution effects based on Eq. (1).

$$t = C_0 + C_1a_1 + \dots + C_da_d + \epsilon \quad (1)$$

where, t - response. Dependent variable, observation and a_i -predictor. Independent variable, explanatory variable and C_d -coefficient and ϵ - random error noise.

The regression coefficients were easily calculated using MATLAB software, then the obtained equations were applied to calculate the target values.

3. Deep Machine Learning

Deep Machine Learning (DML) is an Artificial Intelligence (AI) technique (Figure 2), which can be used for various applications, including prediction applications. By using the DML, the prediction gets better results with more data, bigger models and more computation [17, 18].

ANN models can easily solve the prediction problem [19] by increasing the number of neurons in the input layer [20], or by adding extra hidden layers [21], which leads to more computations. Therefore, the Mean Square Error (MSE) between the targets and the calculated outputs is shown in Figure 3.

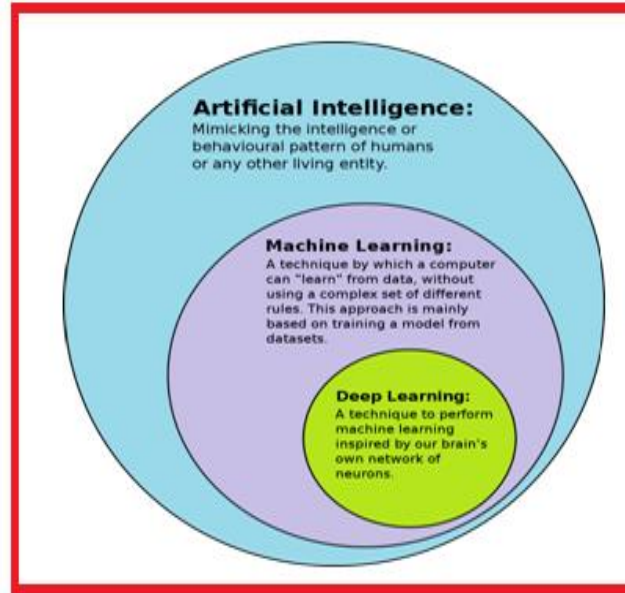


Figure 2. DML as a part of AI

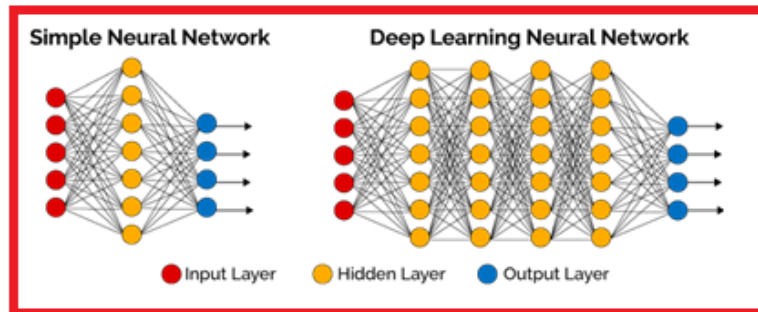


Figure 3. ANN using the DLM

4. Methodology

This paper selected a pollution input data set, which was composed of a 2D matrix with 8 rows and 508 columns. Each column represented the PM values, such as temperature, relative humidity, carbon monoxide, sulfur dioxide, nitrogen dioxide, hydrocarbons, ozone and so on. The target data was a 2D matrix with 3 rows and 508 columns, and each row represented the values of the negative effects of pollution, such as total mortality, respiratory mortality, cardiovascular mortality and so on. Figure 4 shows samples of the input data set.

A MATLAB code was used to calculate regression coefficients. Then the regression equation for each of the three targets was used to calculate the predicted outputs.

This paper created and tested various Feed Forward ANN (FFANN) and calculated the MSE for each FFANN in order to select the optimal FFANN (with the minimum MSE). Then a MATLAB code was used to create, train and test different ANNs with various architectures (Figure 5).

This paper created and tested various Cascade Feed Forward ANN (CFFANN) and calculated the MSE for each CFFANN in order to select the optimal FFANN (with the minimum MSE). Then a MATLAB code was used to create, train and test different ANNs with various architectures (Figure 6).

	Inputs							
	a1	a2	a3	a4	a5	a6	a7	a8
	72.3800	29.2000	11.5100	3.3700	9.6400	45.7900	6.6900	72.7200
	67.1900	67.5100	8.9200	2.5900	10.0500	43.9000	6.8300	49.6000
	62.9400	61.4200	9.4800	3.2900	7.8000	32.1800	4.9800	55.6800
	72.4900	58.9900	10.2800	3.0400	13.3900	40.4300	9.2500	55.1600
	74.2500	34.8000	10.5700	3.3900	11.9000	48.5300	9.1500	66.0200
	67.8800	66.7800	7.9900	2.5700	10.1100	48.6100	8.8000	44.0100
	74.2000	35.1600	10.5100	2.3500	9.4900	38.7600	7.4200	47.8300
	74.8800	55.8600	9.8400	3.3800	10.7200	45.1900	10.6800	43.6000
	64.1700	52.6800	6.6100	1.5000	5.7700	31.6500	5.8500	24.9900
	67.0900	50.9400	7.9000	2.5600	7.4300	40.2000	7.6300	40.4100
	Targets							
t1	183.6300	191.0500	180.0900	184.6700	173.6000	183.7300	171.7900	177.9500
t2	11.9000	10.7500	9.3300	9.5400	8.2700	7.5500	9.1200	7.7600
t3	97.8500	104.6400	94.3600	98.0500	95.8500	95.9800	88.6300	90.8500

Figure 4. Samples of the input data set

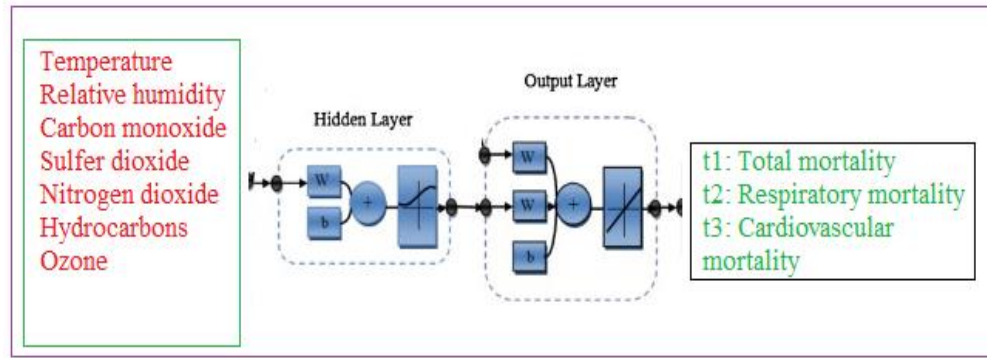


Figure 5. Basic architecture of the FFANN

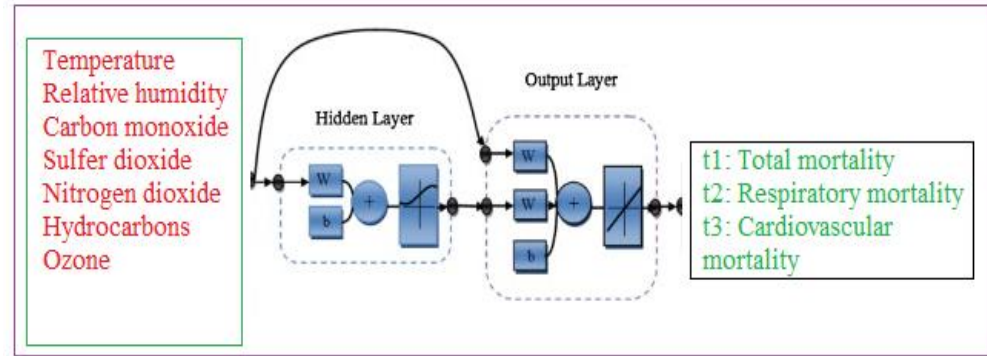


Figure 6. Basic architecture of the CFFANN

5. Result and Discussion

A MATLAB code was written and implemented using the input data set to find the regression coefficients, which are shown in Figure 7. Then the coefficients were used to form the equation for each target. Figure 8 shows the error plot between the targets and the predicted outputs. The maximum and minimum errors (Table 1) were high, and thus the MSE (applying Eq. (2)) between the targets and the predicted outputs was also high.

$$MSE = \frac{1}{n} \sum (t - \hat{t})^2 \quad (2)$$

where, $(t - \hat{t})^2$ - the square of the difference between actual and predicted.

Coefficients		
t1	t2	t3
220.3519	15.3448	134.3318
-0.8580	-0.1126	-0.6681
-0.2183	-0.0158	-0.1815
1.3875	0.0779	1.0582
2.4728	-0.0530	3.4070
-0.9017	-0.0227	-0.6651
0.1590	0.0216	0.0169
0.7085	0.0199	0.2887
0.0725	0.0183	0.0169

Figure 7. Obtained regression coefficients

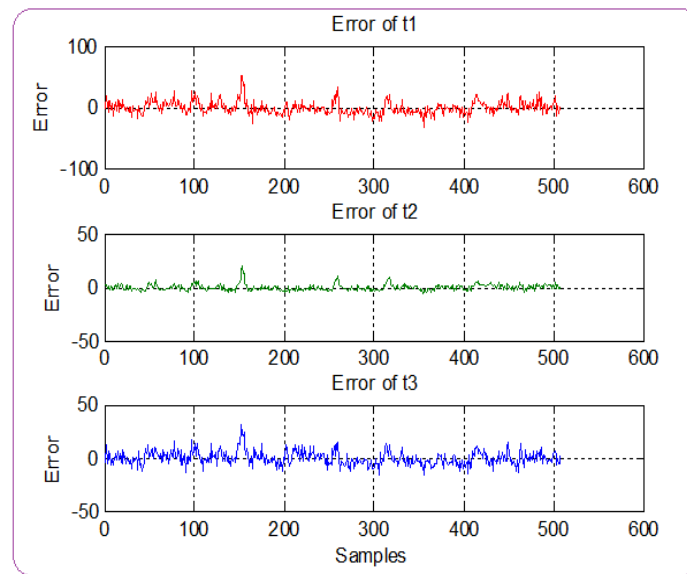


Figure 8. Errors between the targets and the predicted outputs (regression)

Table 1. Errors between the targets and the predicted outputs

Target	t1	t2	t3
Max. error	51.9156	20.3733	31.6502
Min. error	-32.9672	-5.1315	-16.5542

This paper created, tested and implemented different complicated FFANN and CFFANN models to decrease the MSE value. Figure 9 shows the obtained errors using the CFFANN with 5 layers (8-16-8-9-3 neurons), with linear activation function for the output layer and Tansig activation function for all other layers. Figure 10 shows the obtained errors using the FFANN with the same architecture.

According to Figures 9 and 10, when the CFFANN was used for prediction purposes, the MSE value rapidly decreased, thus enhancing the prediction results.

From Table 2 we can see the following facts [22, 23]:

- Use of the ANN enhanced the prediction results, compared with regression analysis.
- The CFFANN gave better results for any ANN architecture.
- Although the DML process required a lot of time (minutes and in some cases hours), it was important and vital in improving the results of the prediction process, taking into account that this sentence was only carried out once.
- Increase of the number of layers decreased the MSE.
- It was better to use 2-layer ANN with a big number of neurons in the input layer. Thus, this paper increased the number of neurons in the input layer, which decreased the MSE for both types of ANN, the FFANN and the CFFANN. This can be seen in Figure 11 [19, 20].

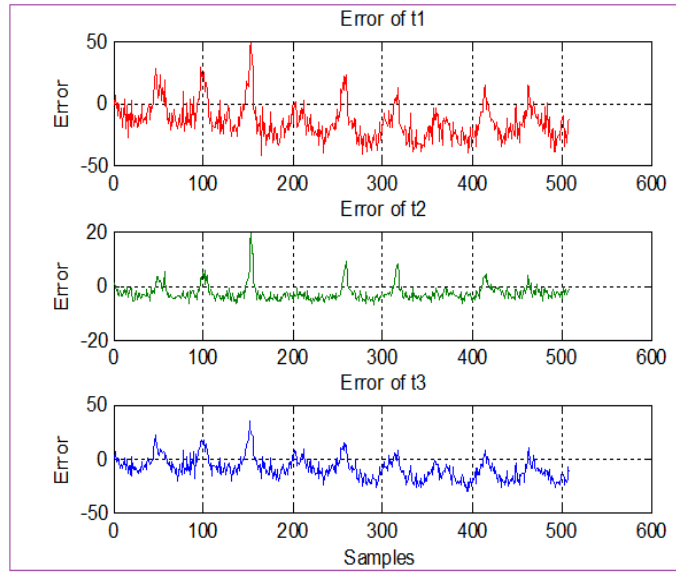


Figure 9. Errors using the CFFANN (MSE=9.09967e-006)

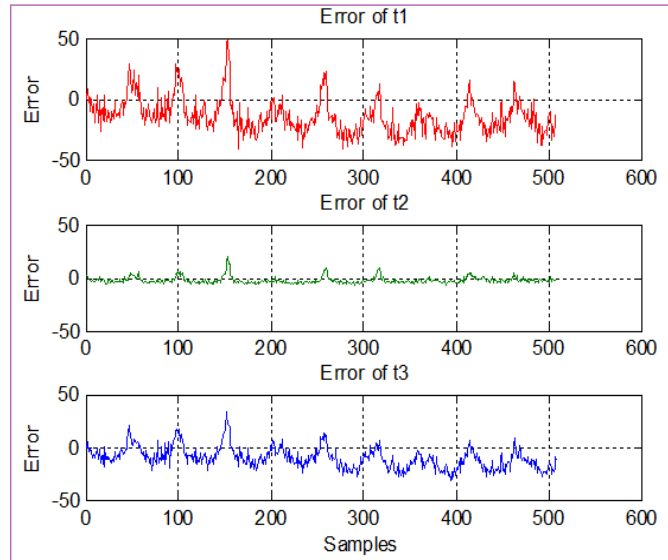


Figure 10. Errors using the FFANN (MSE=0.000101754)

Several experiments were implemented to check the effects of applying the DML. Various compacted ANN models were created and tested. Table 2 shows the obtained experimental results [23, 24]:

Table 2. Obtained results using various ANN models

ANN architecture					MSE	
Layer 1/neurons	Layer 2/neurons	Layer 3/neurons	Layer 4 /neurons	Output layer /neurons	FFANN	CFFANN
8	-	-	-	3	0.000806971	0.000744657
16	-	-	-	3	0.000528262	0.000434399
24	-	-	-	3	0.000278335	0.000240503
32	-	-	-	3	0.000190689	0.000185831
64	-	-	-	3	4.45135e-005	4.37043e-005
128	-	-	-	3	1.66451e-007	9.27851e-008
8	9	-	-	3	0.000452378	0.000279326
8	16	-	-	3	0.000279576	0.00012127
8	16	9	-	3	0.000132483	3.82815e-005
8	16	8	9	3	0.000101754	9.09967e-006

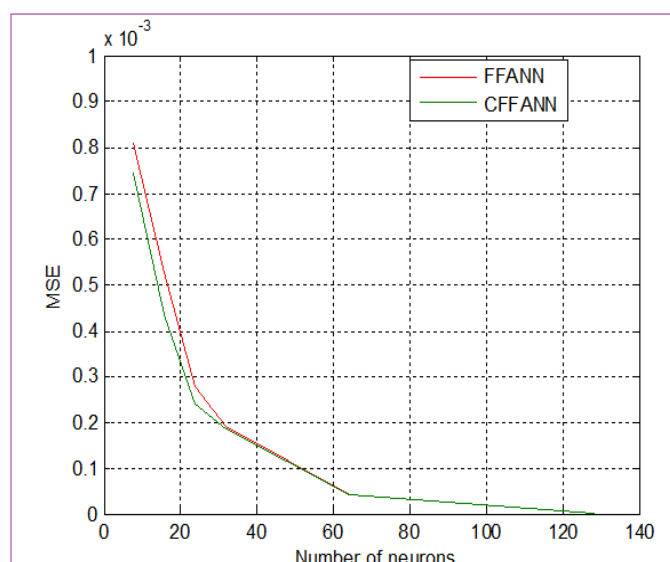


Figure 11. MSE as a function of number of neurons in the input layer

6. Conclusion

When the regression models were used, the negative effects of pollution parameters were easily predicted. These models required a short time, but the obtained MSE was always high. Thus, this paper used the DML models to enhance the prediction results. Using the ANN with complicated architecture increased the computation and rapidly decreased the MSE between the targets and the predicted outputs. It was shown that the best choice was to use the CFFANN with two layers, because the increase of the number of neurons rapidly decreased the MSE, making the prediction tool more accurate.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] A. S. Brown, R. J. C. Brown, P. J. Coleman, C. Conolly, A. J. Sweetman, K. C. Jones, D. M. Butterfield, D. Sarantaris, B. J. Donovan, and I. Roberts, "Twenty years of measurement of polycyclic aromatic hydrocarbons (PAHs) in UK ambient air by nationwide air quality networks," *Environ. Sci.: Proc. Impacts*, vol. 15, no. 6, pp. 1199-1215, 2013. <https://doi.org/10.1039/c3em00126a>.
- [2] A. Utku and M. A. Akcayol, "Deep learning based prediction model for the next purchase," *Adv. Electr. Comput. En.*, vol. 20, no. 2, pp. 35-44, 2020. <https://doi.org/10.4316/AECE.2020.02005>.
- [3] G. Adamkiewicz, A. R. Zota, M. P. Fabian, T. Chahine, R. Julien, J. D. Spengler, and J. I. Levy, "Moving environmental justice indoors: Understanding structural influences on residential exposure patterns in low-income communities," *Am. J. Public Health*, vol. 101, no. 1, pp. S238-S245, 2011. <https://doi.org/10.2105/AJPH.2011.300119>.
- [4] M. Milicevic, M. Baranovic, and K. Zubrinic, "Application of machine learning algorithms for the query performance prediction," *Adv. Electr. Comput. Eng.*, vol. 15, no. 3, pp. 33-44, 2015. <https://doi.org/10.4316/AECE.2015.03005>.
- [5] K. Kampouropoulos, F. Andrade Rengifo, A. García Espinosa, and J. L. Romeral Martínez, "A combined methodology of adaptive neuro-fuzzy inference system and genetic algorithm for short-term energy forecasting," *Adv. Electr. Comput. En.*, vol. 14, no. 1, pp. 9-14, 2014. <https://doi.org/10.4316/AECE.2014.01002>.
- [6] K. Agay-Shay, M. Friger, S. Linn, A. Peled, Y. Amitai, and C. Peretz, "Air pollution and congenital heart defects," *Environ. Res.*, vol. 124, pp. 28-34, 2013. <https://doi.org/10.1016/j.envres.2013.03.005>.

- [7] R. C. Suganthe, R. S. Latha, M. Geetha, and G. R. Sreekanth, "Diagnosis of Alzheimer's disease from brain magnetic resonance imaging images using deep learning algorithms," *Adv. Electr. Comput. En.*, vol. 20, no. 3, pp. 57-64, 2020. <https://doi.org/10.4316/AECE.2020.03007>.
- [8] F. Al Jallad, E. Al-Katheeri, and M. Al-Omar, "Concentrations of particulate matter and their relationships with meteorological variables," *Sustain. Environ. Res.*, vol. 23, no. 3, pp. 191-198, 2013.
- [9] O. A. Schipor, W. Wu, W. T. Tsai, and R. D. Vatavu, "Software architecture design for spatially-indexed media in smart environments," *Adv. Electr. Comput. En.*, vol. 17, no. 2, pp. 17-22, 2017. <https://doi.org/10.4316/AECE.2017.02003>.
- [10] H. Wang, M. Naghavi, C. Allen, et al., "Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980-2015: A systematic analysis for the Global Burden of Disease Study 2015," *The Lancet*, vol. 388, no. 10053, pp. 1459-1544, 2016. [https://doi.org/10.1016/S0140-6736\(16\)31012-1](https://doi.org/10.1016/S0140-6736(16)31012-1).
- [11] J. A. A. N. Asad, Z. Alqadi, I. Shayeb, Q. Jaber, and J. Al-Azzeh, "Simple procedures to create HSCS," *Int. J. Eng. Res. Man.*, vol. 7, no. 5, pp. 6-10, 2020.
- [12] A. Al-Qaisi, A. Manasreh, A. Sharadqeh, and Z. Alqadi, "Digital color image classification based on modified local binary pattern using neural network," *Int. J. Comm. Antenna Pro.*, vol. 9, no. 6, pp. 403-408, 2019. <https://doi.org/10.15866/irecap.v9i6.18425>.
- [13] S. Mohammad, P. Khrisat, and A. Ziad, "Color Images Classifier Optimization," *Int. J. Eng. Res. Rec. Manage.*, vol. 5, no. 3, pp. 6-14, 2021.
- [14] A. Al-Hasanat, H. Alasha'ary, K. Matrouk, Z. Al-Qadi, and H. Al-Shalabi, "Experimental investigation of training algorithms used in back propagation artificial neural networks to apply curve fitting," *Eur. J. Sci. Res.*, vol. 121, no. 4, pp. 328-335, 2014.
- [15] R. A. Zneit, Z. AlQadi, and M. A. Zalata, "A methodology to create a fingerprint for RGB color image," *Int. J. Comput. Sci. Mobile Com.*, vol. 16, no. 1, pp. 205-212, 2017.
- [16] A. A. Moustafa, Z. A. Alqadi, and E. A. Shahroury, "Performance evaluation of artificial neural networks for spatial data analysis," *WSEAS Transactions Com.*, vol. 10, no. 4, pp. 115-124, 2011. <https://doi.org/10.5555/2001184.2001186>.
- [17] J. Al-Azzeh, Z. Alqadi, and M. Abuzalata, "Performance analysis of artificial neural networks used for color image recognition and retrieving," *Int. J. Com. Sci. Mobile Com.*, vol. 8, no. 2, pp. 20-33, 2017.
- [18] K. Matrouk, H. Alasha'ary, A. Al-Hasanat, Z. Al-Qadi, and H. Al-Shalabi, "Investigation and Analysis of ANN Parameters," *Eur. J. Sci. Res.*, vol. 121, no. 2, pp. 217-225, 2014.
- [19] A. Hindi, M. O. Dwairi, and Z. Alqadi, "Analysis of procedures used to build an optimal fingerprint recognition system," *Int. J. Com. Sci. Mobile Com.*, vol. 9, no. 2, pp. 21-37, 2020.
- [20] O. Ulkir, G. Akgun, A. Nasab, and E. Kaplanoglu, "Data-driven predictive control of a pneumatic ankle foot orthosis," *Adv. Electr. Comput. En.*, vol. 21, no. 1, pp. 65-74, 2021. <https://doi.org/10.4316/AECE.2021.01007>.
- [21] A. Al-Rasheed, "Identification of important features and data mining classification techniques in predicting employee absenteeism at work," *Int. J. Elec. Com. Eng.*, vol. 11, no. 5, pp. 4587-4596, 2021. <https://doi.org/10.11591/ijece.v11i5>.
- [22] I. Koyuncu, "Implementation of high speed tangent sigmoid transfer function approximations for artificial neural network applications on FPGA," *Adv. Electr. Comput. En.*, vol. 18, no. 3, pp. 79-86, 2018. <https://doi.org/10.4316/AECE.2018.03011>.
- [23] W. D. Gong, D. Y. Guo, and Y. Q. Liang, "Prediction model of coal and gas outburst based on rough set-unascertained measure theory," *J. Eng. Technol. Sci.*, vol. 50, no. 6, pp. 758-777, 2018. <https://doi.org/10.5614/j.eng.technol.sci.2018.50.6.2>.
- [24] C. B. Rabah, G. Coatrieux, and R. Abdelfattah, "Boosting up source scanner identification using wavelets and convolutional neural networks," *Trait. Signal.*, vol. 37, no. 6, pp. 881-888, 2020. <https://doi.org/10.18280/ts.370601>.