ISSN: 2995-3286

Research Article

International Journal of Media and Networks

Enhancing Air Quality Monitoring Through Spatial Classification in Central Zone of Peninsular Malaysia Using Artifical Neural Network as Prediction Tools

Mohd Suzairi Mohd Shafii[1*] and Hafizan Juahir[2]

[1]*East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia*

[2]*Faculty of Bioresource and Food Industry, Universiti Sultan Zainal Abidin, Besut Campus, 22200, Besut, Terengganu, Malaysia*

[*]Corresponding Author
Mohd Suzairi Mohd Shafii, East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Gong Badak Campus, 21300 Kuala Terengganu, Terengganu, Malaysia.

Submitted: 2024, Mar 21 **Accepted**: 2024, Apr 12 **Published**: 2024, Apr 24

Citation: Shafi'i, M. S. M., Juahir, H. (2024). Enhancing Air Quality Monitoring Through Spatial Classification in Central Zone of Peninsular Malaysia Using Artifical Neural Network as Prediction Tools. *Int J Med Net, 2*(4), 01-11.

Abstract
Ambient air monitoring plays a crucial role in the effective implementation of air quality management systems. This practice entails systematically and over a long period assessing and quantifying specific pollutants in the outdoor environment. However, the high cost of acquiring sufficient equipment for comprehensive air monitoring poses a challenge. Thus, this study proposes that spatial classification could be a viable approach to reducing monitoring stations while still obtaining adequate air monitoring data. The objective of this study was to examine the predictive performance of artificial neural networks (ANNs) in spatial classification to support air quality monitoring. By implementing ANN in this study, the MLP-FF-ANN model successfully distinguished air quality samples in the HPC, MPC, and LPC regions. Particularly notable were the positive outcomes achieved with a configuration of ten hidden nodes, resulting in an R^2 value of 0.7982, as well as the lowest RMSE (0.3799) and MR (0.1950). Additionally, the MLP-FF-ANN model demonstrated commendable performance, achieving an average correct classification rate of 76.38%. These findings suggest that air quality monitoring based on clustered data can effectively maintain data quality.

Keywords: Artificial Neural Networks, Air Monitoring, MLP-FF-ANN.

1. Introduction

The rapid pace of urbanization and industrialization has significant impacts on various aspects, including social, economic, and environmental domains. These impacts are evident in increased energy and water consumption, elevated pollution levels, and the degradation of land and forests, all of which have consequential effects on human well-being [1]. Several factors contribute to air pollution, including transportation, manufacturing, power generation, commerce, urban areas with agricultural economies, wood burning, particulate matter, fires, and volcanic eruptions [2-4].

In recent years, Malaysia has been grappling with a severe crisis of air pollution. Due to its status as a developing economy with a robust manufacturing sector and heavy dependence on automobiles, the country is particularly vulnerable to this issue [5]. Furthermore, Malaysia also contends with significant pollution stemming from slash-and-burn practices and forest fires in its neighboring country, Indonesia, intensifying the predicament [6,7]. The consequent haze has diverse social and economic ramifications, especially concerning human health. The Ministry of Health Malaysia has reported a significant correlation between daily mean levels of air pollutants and respiratory fatalities as well as natural deaths. Previous studies have demonstrated a robust association between exposure to PM_{10}, N_{O2}, and CO and increased hospital admissions for respiratory diseases [8,9]. The economic implications of air pollution on health are substantial. According to the report "The Health and Economic Impacts of Ambient Air Quality in Malaysia," the total cost of healthcare for diseases caused by air pollution, as well as the loss of productivity due to illness, is estimated to be RM303 billion. A study conducted by Li et al. (2020) found that air pollution increases the incidence of respiratory diseases and worsens overall health, leading to higher healthcare expenses [10].

The economic implications of air pollution are significant, necessitating the prioritization of measures aimed at improving air quality and establishing effective monitoring and management protocols. Air quality monitoring plays a crucial role in understanding the extent of air pollution and its impacts on human health and the environment. This involves the measurement of atmospheric pollutants and holds substantial importance in global efforts towards environmental conservation [11]. In Malaysia, collaborative endeavors between the Department of Environment (DOE) and Alam Sekitar Malaysia Sdn Bhd (ASMA) are essential for conducting monitoring and data collection activities. The primary objective of these initiatives is to provide real-time information to the public regarding the concentrations of major pollutants, as highlighted by Hawari et al. (2019) and Mohd Shafie et al. (2022). A significant milestone was achieved in 1989 with the establishment of the Recommended Malaysian Air Quality Guidelines (RMAQG). Subsequently, in 1993, the Malaysian Air Quality Index (MAQI) was introduced to complement the existing regulatory framework. In 2015, an improved guideline featuring a three-tier implementation approach was introduced. The New Malaysia Ambient Air Quality Standards (NMAAQS) form the basis for calculating the Air Pollutant Index (API). The API is determined through the monitoring of key air pollutants, including CO, N_{O2}, S_{O2}, $PM_{2.5}$ and PM_{10} with its value derived from assessing the highest concentration of these pollutants detected over a specified timeframe.

Air quality monitoring systems typically rely on large, stationary instruments that are associated with high installation and maintenance costs. Consequently, these systems exhibit limited coverage, which hampers accurate monitoring and timely interventions. As a potential solution, there is a suggestion to develop an affordable real-time air quality monitoring system [12]. However, instead of pursuing the development of new physical instruments, recent attention has been directed towards the utilization of artificial intelligence (AI) for the management and mitigation of air pollution. AI represents a critical tool in environmental protection endeavors, assisting regulatory bodies in the identification of effective mitigation strategies aimed at minimizing public exposure to air pollutants [13,14]. Moreover, AI's capability to navigate complex interactions among various air quality parameters enables more precise forecasting of air pollutant concentrations [15]. Therefore, the present study endeavors to examine the predictive efficacy of artificial neural networks (ANNs) in spatial classification, with potential implications for enhancing air quality monitoring practices.

2. Materials and Methods

2.1 Study Area and Air Quality Data

The Department of Environment (DOE) of Malaysia has classified eleven continuous air quality monitoring (CAQM) stations in the central zone of Peninsular Malaysia. These stations are in the central zone, surrounded by a combination of residential, industrial, heavy traffic, and rural areas [16,17]. The selection of these station locations was based on factors such as historical and current monitoring data, representativeness, accessibility, availability of support services, security, and topography [18]. According to Elias et al., 2023, the central zone is an area in Malaysia experiencing rapid development and a high-density population [19]. The rapid growth, uncontrolled industrial development, biomass, and fossil fuel combustion, increasing number of motor vehicles, and human activities such as land clearing, road and highway development, and residential expansion with a growing population in the central zone are responsible for the emissions and release of various types of pollutants in the area.

Table 1 presents a comprehensive list of the CAQM stations in the central zone. For this study, the Department of Environment (DOE) of Malaysia provided secondary air quality data, encompassing a five-year period from January 1, 2017, to December 31, 2021. The study utilizes daily data on various air quality parameters, including carbon monoxide (CO), sulfur dioxide (SO_2), nitrogen dioxide (NO_2), particulate matter (PM_{10} and $PM_{2.5}$), and ozone (O_3), as well as daily Air Pollutant Index (API) readings.

No.	State	Location	Coordinates	Zone	Classification
1.	Kuala Lumpur	Batu Muda	03° 12' 44.78" N, 101° 40' 56.02" E	Central	Suburban
2.		Cheras	03° 06' 22.44" N, 101° 43' 04.50" E		Urban
3.	Putrajaya	Putrajaya	02° 54' 53.33" N, 101° 41' 24.17" E		Sub Urban
4.	Selangor	Kuala Selangor	03° 19' 16.70" N, 101° 15' 22.47" E		Rural
5.		Petaling Jaya	03° 07' 59.40" N, 101° 36' 28.83" E		Suburban
6.		Shah Alam	03° 06' 16.98" N, 101° 33' 22.39" E		Urban
7.		Klang	03° 00' 53.60" N, 101° 24' 47.19" E		Suburban
8.		Banting	02° 49' 00.08" N, 101° 37' 23.36" E		Suburban
9.	Negeri Sembilan	Nilai	02° 49' 18.09" N, 101° 48' 41.34" E		Suburban
10.		Seremban	02° 43' 24.17" N, 101° 58' 06.58" E		Urban
11.		Port Dickson	02° 26' 28.97" N, 101° 52' 00.68" E		Suburban

Table 1: CAQM Stations in Central Zone of Peninsular Malaysia

2.2 Statistical Analysis Methods

2.2.1 Descriptive Analysis

Univariate statistics were employed to examine the minimum, maximum, mean, median, and standard deviation values of each air quality parameter. The outcomes of this analysis will provide significant insights into the ambient air quality within the study area. Subsequently, these findings were assessed against the New Malaysia Ambient Air Quality Standards (NMAAQS), as depicted in Table 2.

Pollutants	Averaging Time	Ambient Air Quality Standard	
		ppm	$\mu g/m^3$ / *mg/m^3
Ozone, O_3	1-Hour	0.090	180
	8-Hour	0.050	100
Carbon Monoxide, CO	1-Hour	26.2	30*
	8-Hour	8.75	10*
Nitrogen Dioxide, NO_2	1-Hour	0.150	280
	24-Hour	0.037	70
Sulfur Dioxide, SO_2	1-Hour	0.095	250
	24-Hour	0.030	80
Particulate Matter, PM_{10}	24-Hour		100
	1-Year		40
Particulate Matter, $PM_{2.5}$	24-Hour		35
	1-Year		15

Table 2: New Malaysia Ambient Air Quality Standards (DOE, 2020)

2.2.2 Hierarchical Agglomerative Cluster Analysis (HACA)

Hierarchical Agglomerative Cluster Analysis (HACA) is extensively employed for identifying air pollution characteristics in air quality monitoring stations based on their location. Previous research conducted by Lu et al. (2011), Austin et al. (2013), Azid et al. (2015), Isiyaka & Azid (2015), Song et al. (2016), and Liu et al. (2018) has demonstrated the effectiveness of HACA in categorizing air pollution characteristics using air quality monitoring stations [20-25]. This clustering analysis technique is commonly used in data mining to establish a hierarchical structure of clusters, which is typically depicted in a dendrogram, a tree diagram [26,27]. It identifies similarity patterns within the dataset and presents them as homogeneous subsets to reveal relationships between observations [27-29].

In this study, HACA was employed to analyze the air pollutant index (API) data from each monitoring station as a variable for classifying spatial air quality into clusters with high homogeneity within the class and high heterogeneity between classes. Ward's method utilizes squared Euclidean distance as a dissimilarity measure between individuals (Strauss & von Maltitz, 2017). The Euclidean distance in this study is calculated using the following equation (Eq.1):

$$(D_{link} / D_{max}) \times 100 \quad \ldots\ldots\ldots\ldots\ldots\ldots \quad \text{(Eq.1)}$$

where D_{link} represents the linkage distance and D_{max} denotes the maximal distance. The quotient is multiplied by 100 to standardize the linkage distance depicted on the y-axis [30-33].

The results of the hierarchical clustering are presented in the dendrogram, which provides a visual summary of the clustering process. The dendrogram serves as an image of the groups and their proximity [31,34,35].

2.2.3 Artificial Neural Network for Spatial Classification

Artificial neural networks (ANNs) have emerged as powerful problem-solving tools, attracting considerable attention across diverse fields, including social media analysis, medical diagnosis, aerospace, defense, route optimization, robotics, stock market prediction, weather forecasting, and air quality assessment. Structured through interconnected layers of inputs, hidden units, and outputs, ANNs form complex networks capable of learning patterns and making predictions [36,37]. Both supervised and unsupervised classification methods extensively employ ANNs, leveraging their capacity for data-driven learning and modeling [38]. This study focuses on utilizing a multi-layer perceptron feed-forward artificial neural network (MLP-FF-ANN) for spatial classification, implemented using JMP10 software. Optimization of the network architecture, including the determination of the optimal number of hidden nodes and neurons within each layer, was accomplished through iterative experimentation [39]. Furthermore, a backpropagation algorithm was employed to minimize prediction errors during training [35,39]. The accuracy of the predictive model in classification will be evaluated based on performance metrics such as the coefficient of determination (R^2) and root mean square error (RMSE) [40]. A higher R^2 value and a lower RMSE indicate improved predictive capabilities of the model, demonstrating its efficacy in spatial classification tasks.

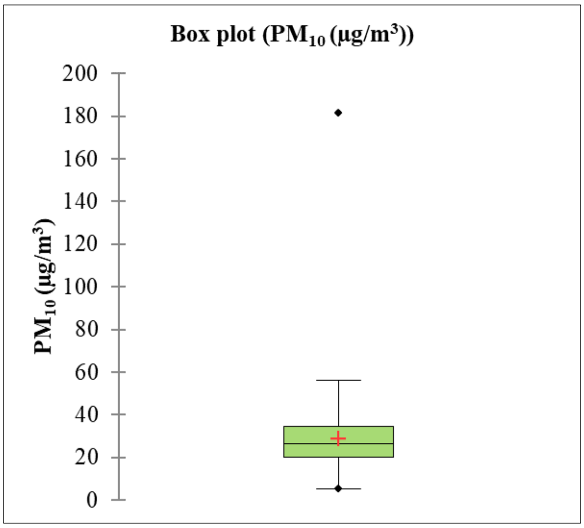
3. Results and Discussion

3.1 Descriptive Analysis

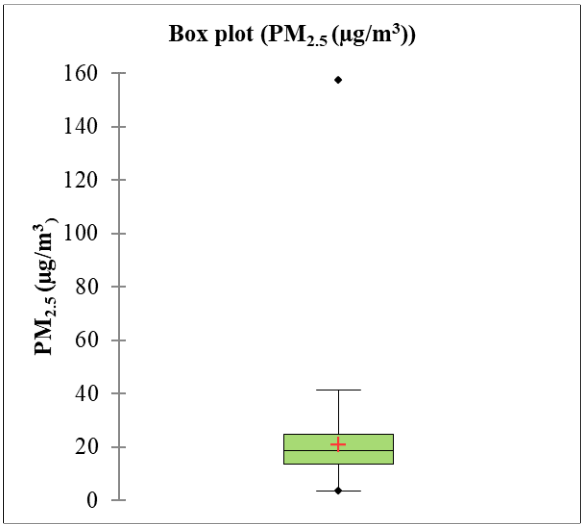
Based on the results shown in Table 3, the highest maximum concentrations of PM_{10}, $PM_{2.5}$, S_{O2}, N_{O2}, O_3, and CO are recorded as 181.50 $\mu g/m^3$, 157.70 $\mu g/m^3$, 0.01 ppm, 0.05 ppm, 0.06 ppm, and 2.24 ppm, respectively. Both PM_{10} and $PM_{2.5}$ exhibit maximum values that exceed the National Ambient Air Quality Standards (NMAAQS) approved levels, whereas the concentrations of other pollutants are lower than the approved values. The Air Pollutant Index (API) reading reached its highest level of 227 within a five-year span, indicating a very unhealthy air quality status (API: 201-300). The mean values for all six air quality parameters at all stations did not exceed the approved concentration limits for air pollutants based on NMAAQS. Figure 1 (i)-(vii) illustrates the distribution of the collected data.

Statistic	PM_{10} (µg/m³)	$PM_{2.5}$ (µg/m³)	SO_2 (ppm)	NO_2 (ppm)	O_3 (ppm)	CO (ppm)	API
Minimum	5.50	3.68	0.00	0.00	0.00	0.11	21.00
Maximum	181.50	157.70	0.01	0.05	0.06	2.24	227.00
1st Quartile	20.19	13.76	0.00	0.01	0.01	0.53	55.00
Median	26.61	18.64	0.00	0.01	0.02	0.67	61.00
3rd Quartile	34.53	24.82	0.00	0.02	0.02	0.87	67.00
Mean	**29.13**	**20.87**	**0.00**	**0.01**	**0.02**	**0.71**	**62.70**
Variance (n-1)	202.90	144.18	0.00	0.00	0.00	0.07	218.46
Std dev. (n-1)	14.24	12.01	0.00	0.01	0.01	0.26	14.78
Averaging Period NMAAQS	24hrs 100	24hrs 35	1hr 0.095	1hr 0.150	1hr 0.090	1hr 26.2	

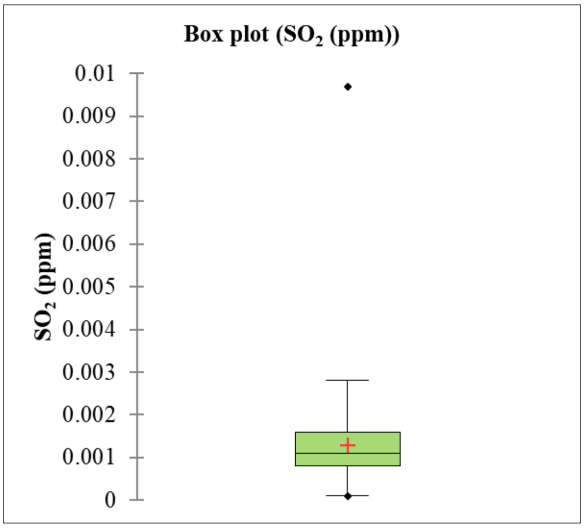
Table 3: Summary of Descriptive Analysis for Eleven CAQM Stations



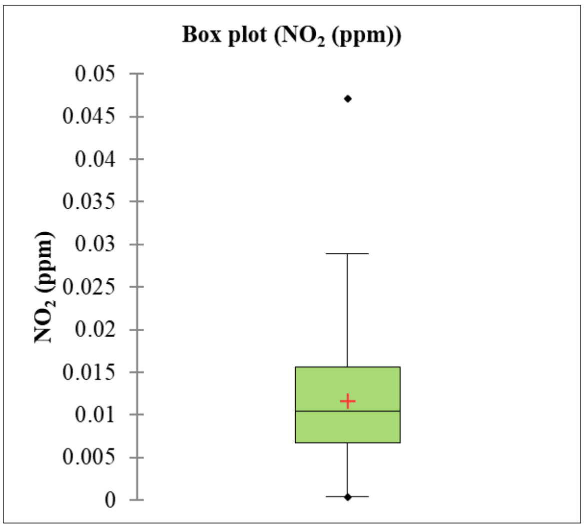
(i)



(ii)



(iii)



(iv)

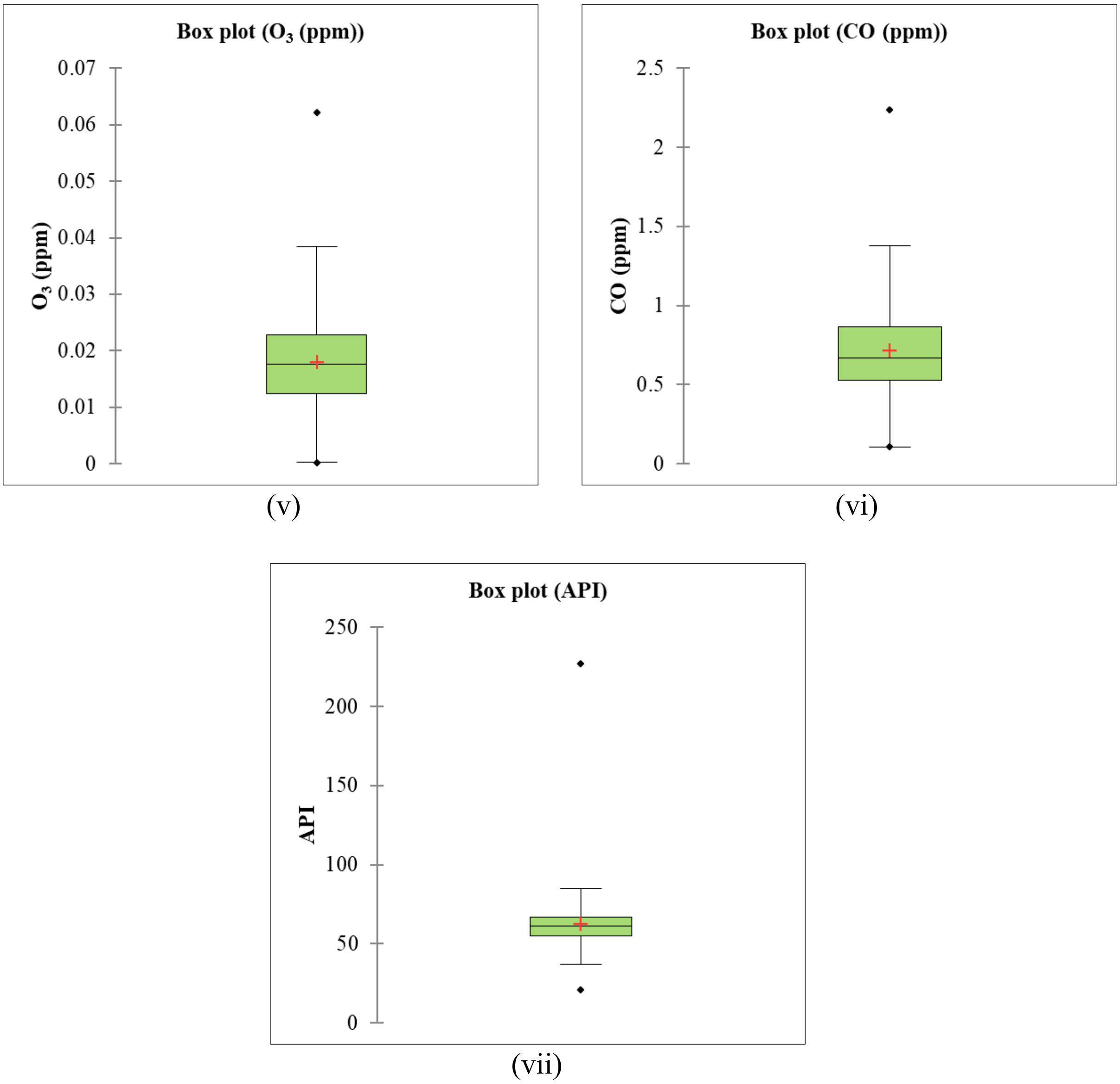


Figure 1: (i)-(ii). Summary of Data Distributions for Five Years Sampling from Eleven CAQM Stations

3.2 Spatial Classification Based on Air Pollutant Index (API) by HACA

Eleven CAQM stations in the central zone of Peninsular Malaysia have been categorized into three distinct clusters using the HACA method. As shown in Figure 4.1, the dendrogram illustrates three separate clusters of monitoring stations, each exhibiting similar characteristics within their respective cluster. These clusters are referred to as the Low Pollution Cluster (LPC), Moderate Pollution Cluster (MPC), and High Pollution Cluster (HPC). According to Table 4, the LPC includes the Batu Muda, Kuala Selangor, Port Dickson, and Seremban stations, with the highest recorded API readings in each station being 203, 198, 213, and 178, respectively. Batu Muda and Port Dickson stations were classified as having a very unhealthy status, while the others were categorized as unhealthy. The average API for the LPC was found to be 59.04, indicating a moderate air quality status over the course of five years. The MPC consists of six stations: Banting, Cheras, Nilai, Petaling Jaya, Putrajaya, and Shah Alam. The highest recorded API readings in these stations over the past five years were 195, 189, 212, 209, 205, and 221, respectively, with an average of 65.09. Banting and Cheras stations were classified as unhealthy, while the remaining stations were categorized as very unhealthy. Overall, the MPC cluster indicates a moderate air quality status. Klang station was the sole station classified as HPC, with the highest recorded API of 227, signifying a very unhealthy status. The average API for Klang station was 69.62, and it also experienced a moderate air quality status over the past five years. Figure 3 presents a comparison of the average API values within the three significant clusters.

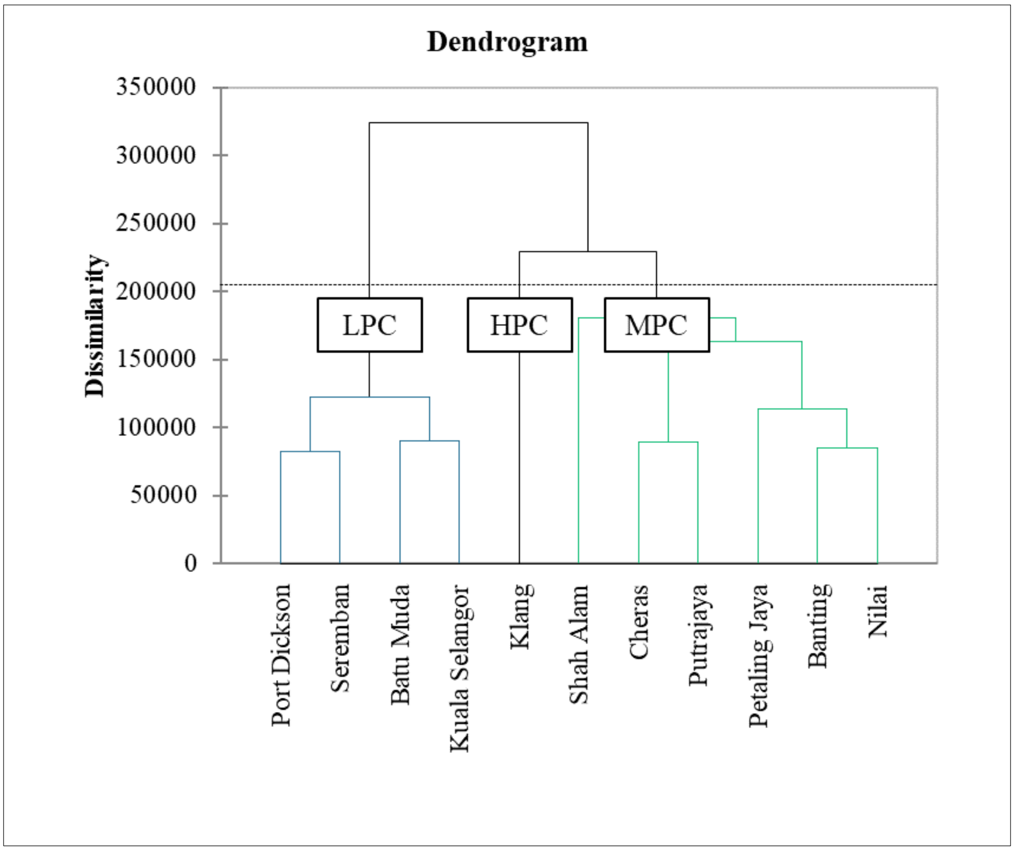


Figure 2: Classification of CAQM in Central Zone

Cluster	LPC	MPC	HPC
Average API	59.04	65.09	69.62
Number of stations by cluster	4	6	1
	Batu Muda	Banting	Klang
	Kuala Selangor	Cheras	
	Port Dickson	Nilai	
	Seremban	Petaling Jaya	
		Putrajaya	
		Shah Alam	

Table 4: List of CAQM Stations Based on Cluster Performed by HACA

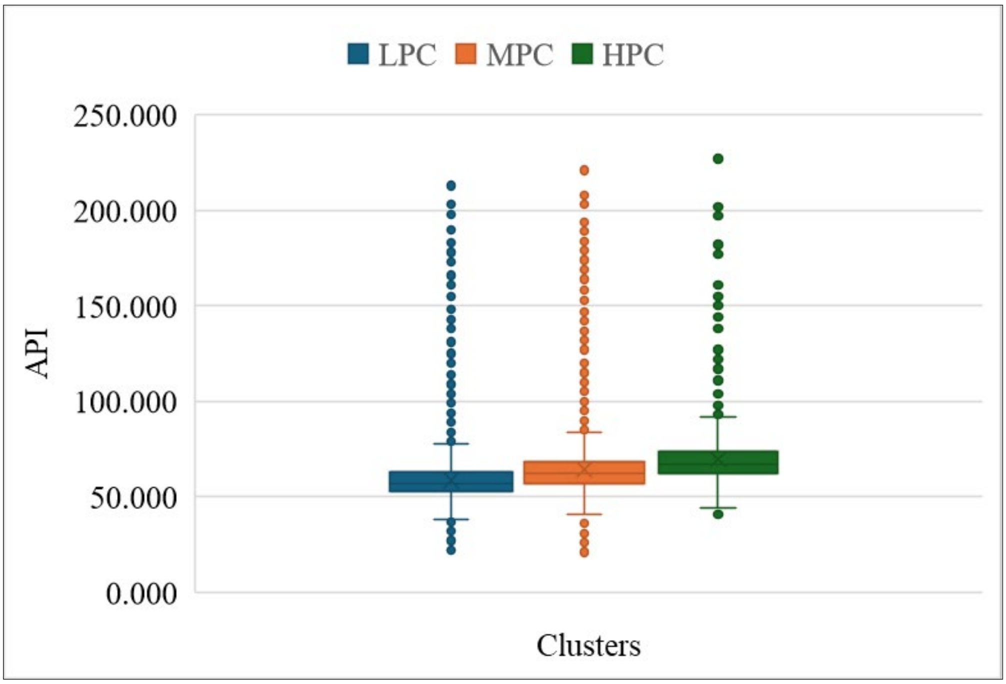


Figure 3: Comparison of API Among the Clusters

3.2.3 Spatial Distribution of Air Quality Parameters Using MLP-FF-ANN Model

The objective of this analysis was to identify spatial variables related to air quality by utilizing fifteen (15) hidden nodes to generate fifteen (15) network structures ranging from 1 to 15. Six air quality parameters were included in this analysis using raw data. The summarized findings can be found in Table 5. Based on these findings, the ML-FF-ANN model successfully differentiated between air quality samples in the HPC, MPC, and LPC regions. In this analysis, ten hidden nodes were determined to be optimal since increasing the number of hidden nodes beyond this point resulted in a decline in prediction performance, as shown in Figure 4 (i)-(iii). The model produced the most favourable results with the configuration of ten hidden nodes, with an R^2 value of 0.7982, as well as the lowest RMSE (0.3799) and MR (0.1950) compared to other configurations.

The MLP-FF-ANN model effectively differentiated between each group, achieving an average correct classification rate of 76.38%, as shown in Table 6. The performance of the receiver operating characteristic (ROC) is depicted in Figure 5, based on the area under the ROC curve (AUC). The spatial distribution model displayed a normal distribution with area values of 0.9151, 0.8862, and 0.9812 for HPC, MPC, and LPC, respectively. According to Deary & Griffiths (2021) and Bekkar et al. (2013), a value higher than 0.9 signifies excellence, while the range of 0.8 – 0.9 is considered very good. Consequently, the MLP-FF-ANN model demonstrated excellence as a classifier parameter, with an average value surpassing 0.9 across all clusters.

No. of Hidden Nodes	R^2	RMSE	Misclassification Rate
[10,1,1]	0.7870	0.3878	0.2010
[10,2,1]	0.7908	0.3853	0.2000
[10,3,1]	0.7867	0.3881	0.2022
[10,4,1]	0.7936	0.3830	0.1990
[10,5,1]	0.7964	0.3807	0.1964
[10,6,1]	0.7966	0.3809	0.1963
[10,7,1]	0.7963	0.3815	0.1954
[10,8,1]	0.7971	0.3806	0.1951
[10,9,1]	0.7955	0.3819	0.1954
[10,10,1]	**0.7982**	**0.3799**	**0.1950**
[10,11,1]	0.7977	0.3804	0.1966
[10,12,1]	0.7971	0.3805	0.1962
[10,13,1]	0.7966	0.3810	0.1966

[10,14,1]	0.7976	0.3802	0.1953
[10,15,1]	0.7965	0.3810	0.1960

Table 5: The Prediction Performance of Spatial Classification (HPC, MPC and LPC).

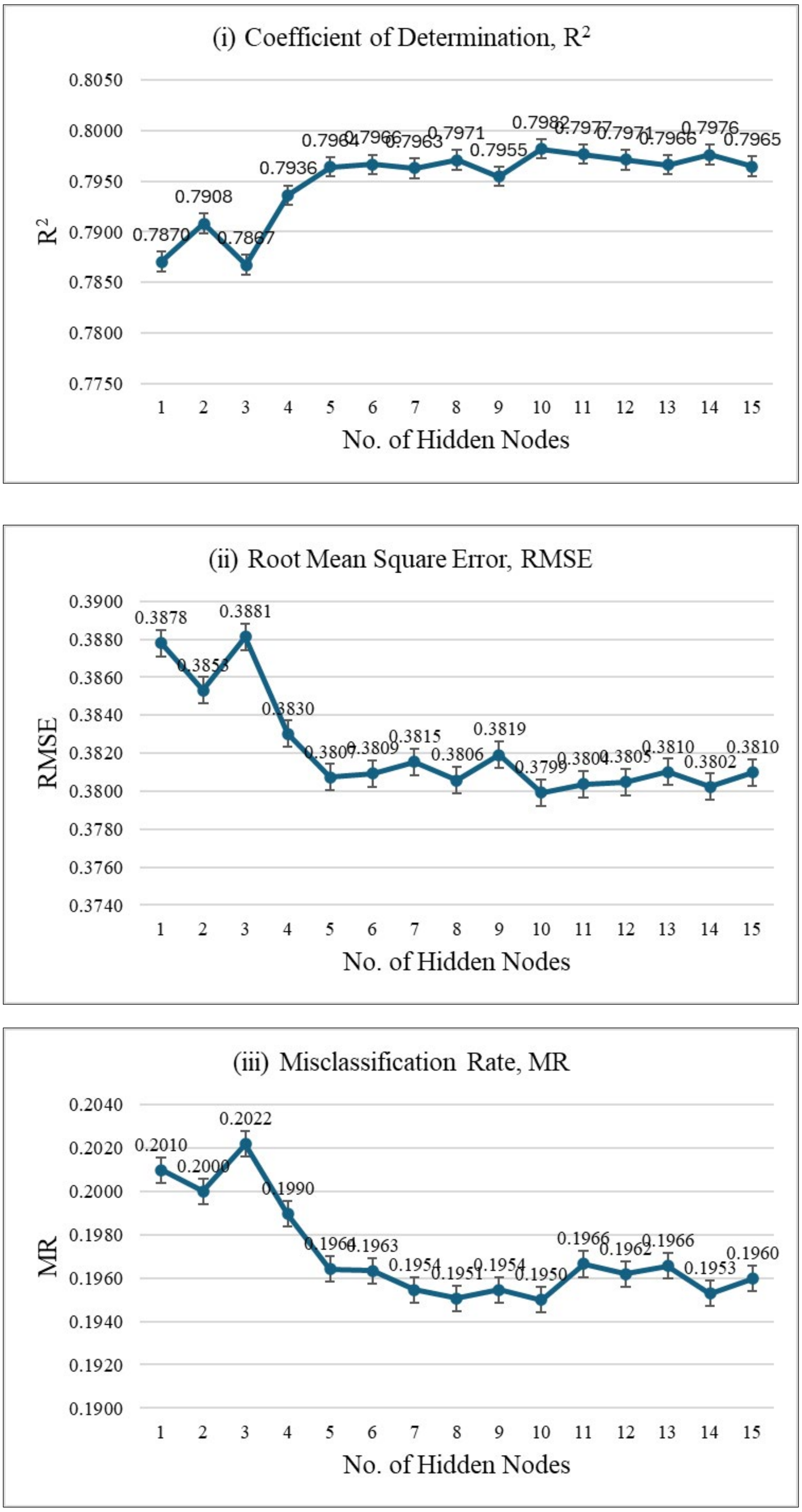


Figure 4: (i)-(iii). The Performance of the MLP-FF-ANN Model Obtained Using (i) R^2 Value, (ii) RMSE and (iii) MR

Sampling Clusters	Clusters assigned by the ANN			Total	% Correct
	HPC	LPC	MPC		
HPC	2842	195	1244	4281	66.39
LPC	6	9137	490	9633	94.85
MPC	1289	692	4191	6172	67.90
Total	4137	10024	5925	20086	76.38

Table 6: Classification Matrix for the MLP-FF-ANN Model Depicting Spatial Variation Across the Three Clusters

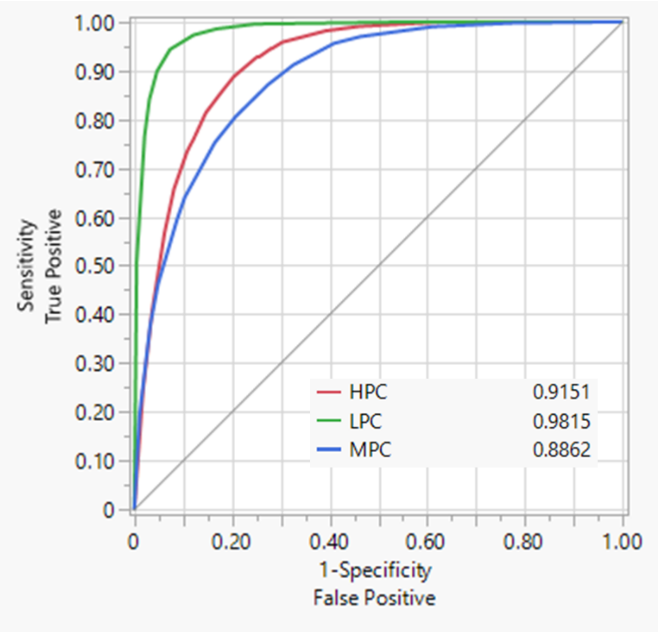


Figure 5: Receiver Operating Characteristics (ROC) for Spatial Distribution of Air Quality Parameters in Central Zone of Peninsular Malaysia

4. Conclusion

In conclusion, this study presents a comprehensive examination of the efficacy of artificial neural networks in generating reliable models for spatial classification. The ML-FF-ANN model effectively differentiated air quality samples in the HPC, MPC, and LPC regions. Particularly notable were the impressive outcomes achieved using a configuration of ten hidden nodes, resulting in a significantly high R^2 value of 0.7982, as well as the lowest RMSE (0.3799) and MR (0.1950). Additionally, the MLP-FF-ANN model displayed commendable performance, attaining an average correct classification rate of 76.38%. The exceptional performance as a classifier parameter is further underscored by an average value surpassing 0.9 across all clusters. These findings highlight the potential of artificial neural networks in spatial classification tasks and emphasize their applicability in addressing complex environmental challenges. By facilitating spatial classification, this technology could assist the authorized agency in effectively managing air monitoring while minimizing costs and time.

Acknowledgment

The author expresses heartfelt gratitude to the Department of Environment of Malaysia and the East Coast Environmental Research Institute (ESERI), Universiti Sultan Zainal Abidin, Malaysia, for their invaluable guidance and support throughout the completion of this research project.

References

1. Indraj, I., & Warpa, V. (2024). Assessment of the seasonal trends of air pollution: A case study of Gurugram city, Haryana, India. *Journal of Air Pollution and Health, 9*(1), 29-40.
2. Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and health impacts of air pollution: a review. *Frontiers in public health, 8*, 505570.
3. Rahman, E. A., Hamzah, F. M., Latif, M. T., & Dominick, D. (2022). Assessment of PM2. 5 Patterns in Malaysia Using the Clustering Method. *Aerosol and Air Quality Research, 22*(1), 210161.
4. Nawaz, M. O., Henze, D. K., Anenberg, S. C., Ahn, D. Y., Goldberg, D. L., Tessum, C. W., & Chafe, Z. A. (2023). Sources of air pollution-related health impacts and benefits of radially applied transportation policies in 14 US cities. *Frontiers in Sustainable Cities, 5,* 1102493.
5. Akhtar, R. (2023). Impact of Air Pollution on Human Health.
6. Yusof, K. M. K. K., Azid, A., Samsudin, M. S., & Jamalani,

M. A. (2017). An overview of transboundary haze studies: The underlying causes and regional disputes on Southeast Asia region. *Malaysian Journal of Fundamental and Applied Sciences, 13*(4), 747-753.
7. Chew, Y. J., Ooi, S. Y., Pang, Y. H., & Wong, K. S. (2022). A Review of forest fire combating efforts, challenges and future directions in Peninsular Malaysia, Sabah, and Sarawak. *Forests, 13*(9), 1405.
8. Sofwan, N. M., Mahiyuddin, W. R. W., Latif, M. T., Ayub, N. A., Yatim, A. N. M., Mohtar, A. A. A., ... & Sahani, M. (2021). Risks of exposure to ambient air pollutants on the admission of respiratory and cardiovascular diseases in Kuala Lumpur. *Sustainable Cities and Society, 75,* 103390.
9. Rozita, W. M. W., Zamtira, S., Mohd, F. I., Mohd, T. L., Nurul, I. A., Muhammad, I. A. W., & Mazrura, S. (2022). Air Pollution, Cardiovascular and Respiratory Admissions in Klang Valley, Malaysia-Finding the Effects. *Nature Environment and Pollution Technology, 21*(4), 1565-1573.
10. Li, L., Du, T., & Zhang, C. (2020). The impact of air pollution on healthcare expenditure for respiratory diseases: evidence from the people's republic of china. *Risk Management and Healthcare Policy,* 1723-1738.
11. Ranganathan, R. H., Balusamy, S., Partheeban, P., Mani, C., Sridhar, M., & Rajasekaran, V. (2023). Air quality monitoring and analysis for sustainable development of solid waste dump yards using smart drones and geospatial technology. *Sustainability, 15*(18), 13347.
12. Hawari, H. F., Zainal, A. A., & Ahmad, M. R. (2019). Development of real time internet of things (IoT) based air quality monitoring system. *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS), 13*(3), 1039-1047.
13. Hassan, R., & Min, L. I. (2010). Urban air pollution forecasting using artificial intelligence-based tools. In *Air Pollution.*
14. Schürholz, D., Kubler, S., & Zaslavsky, A. (2020). Artificial intelligence-enabled context-aware air quality prediction for smart cities. *Journal of Cleaner Production, 271*, 121941.
15. Liang, Y. C., Maimury, Y., Chen, A. H. L., & Juarez, J. R. C. (2020). Machine learning-based prediction of air quality. applied sciences, 10(24), 9151. Liang, Y. C., Maimury, Y., Chen, A. H. L., & Juarez, J. R. C. (2020). *Machine learning-based prediction of air quality. applied sciences, 10*(24), 9151.
16. Afroz, R., Hassan, M. N., & Ibrahim, N. A. (2003). Review of air pollution and health impacts in Malaysia. *Environmental research, 92*(2), 71-77.
17. Sentian, J., Herman, F., Yih, C. Y., & Wui, J. C. H. (2019). Long-term air pollution trend analysis in Malaysia. *International Journal of Environmental Impacts, 2*(4), 309-324.
18. Sahak, N., Asmat, A., & Yahaya, N. Z. (2022). Spatio-Temporal Air Pollutant Characterization for Urban Areas. *Journal of Geoscience and Environment Protection, 10*(1), 218-237.
19. Elias, M. S., Hashim, A., Bahrudin, N. F. D., Sapiee, N. A., Paulus, W., Azman, M. A., ... & Abdullah, I. M. (2023, July). Assessment and sources identification of air quality pollution in Klang Valley, Kuala Lumpur, Malaysia. In *IOP Conference Series: Materials Science and Engineering* (Vol. 1285, No. 1, p. 012017). IOP Publishing.
20. Lu, W. Z., He, H. D., & Dong, L. Y. (2011). Performance assessment of air quality monitoring networks using principal component analysis and cluster analysis. *Building and Environment, 46*(3), 577-583.
21. Austin, E., Coull, B. A., Zanobetti, A., & Koutrakis, P. (2013). A framework to spatially cluster air pollution monitoring sites in US based on the PM2. 5 composition. *Environment international, 59,* 244-254.
22. Azid, A., Juahir, H., Toriman, M. E., Endut, A., Abdul Rahman, M. N., Amri Kamarudin, M. K., ... & Yunus, K. (2016). Selection of the most significant variables of air pollutants using sensitivity analysis. *Journal of Testing and Evaluation, 44*(1), 376-384.
23. Isiyaka, H. A., & Azid, A. (2015). Air quality pattern assessment in Malaysia using multivariate techniques. *Malaysian Journal of Analytical Sciences, 19*(5), 966-978.
24. Song, J., Guang, W., Li, L., & Xiang, R. (2016). Assessment of air quality status in Wuhan, China. *Atmosphere, 7*(4), 56.
25. Liu, J., Li, W., & Wu, J. (2018). A framework for delineating the regional boundaries of PM2. 5 pollution: A case study of China. *Environmental pollution, 235,* 642-651.
26. Sasirekha, K., & Baby, P. (2013). Agglomerative hierarchical clustering algorithm-a. *International Journal of Scientific and Research Publications, 83*(3), 83..
27. Zulkepli, N. F. S., Noorani, M. S. M., Razak, F. A., Ismail, M., & Alias, M. A. (2020). Cluster analysis of haze episodes based on topological features. *Sustainability, 12*(10), 3985.
28. McKenna Jr, J. E. (2003). An enhanced cluster analysis program with bootstrap significance testing for ecological community analysis. *Environmental Modelling & Software, 18*(3), 205-220.
29. Ibarra-Berastegi, G., Sáenz, J., Ezcurra, A., Ganzedo, U., de Argandoña, J. D., Errasti, I., ... & Polanco-Martínez, J. (2009). Assessing spatial variability of SO2 field as detected by an air quality network using self-organizing maps, cluster, and principal component analysis. *Atmospheric Environment, 43*(25), 3829-3836.
30. Singh, K. P., Malik, A., Mohan, D., & Sinha, S. (2004). Multivariate statistical techniques for the evaluation of spatial and temporal variations in water quality of Gomti River (India)—a case study. *Water research, 38*(18), 3980-3992.
31. Singh, K. P., Malik, A., & Sinha, S. (2005). Water quality assessment and apportionment of pollution sources of Gomti river (India) using multivariate statistical techniques—a case study. *Analytica Chimica Acta, 538*(1-2), 355-374.
32. Shrestha, S., & Kazama, F. (2007). Assessment of surface water quality using multivariate statistical techniques: A case study of the Fuji river basin, Japan. *Environmental Modelling & Software, 22*(4), 464-475.
33. Lau, J., Hung, W. T., & Cheung, C. S. (2009). Interpretation of air quality in relation to monitoring station's surroundings. *Atmospheric Environment, 43*(4), 769-777.
34. Forina, M., Armanino, C., & Raggio, V. (2002). Clustering with dendrograms on interpretation variables. *Analytica*

Chimica Acta, 454*(1), 13-19.

35. Juahir, H., Zain, S. M., Yusoff, M. K., Hanidza, T. T., Armi, A. M., Toriman, M. E., & Mokhtar, M. (2011). Spatial water quality assessment of Langat River Basin (Malaysia) using environmetric techniques. *Environmental monitoring and assessment, 173,* 625-641.
36. Taşpınar, F., & Bozkurt, Z. (2014). Application of artificial neural networks and regression models in the prediction of daily maximum PM10 concentration in Düzce, Turkey. *Fresenius Environ. Bull, 23,* 2450-2459.
37. Yadav, V. I. B. H. A., & Nath, S. A. T. Y. E. N. D. R. A. (2017). Prediction of air quality using artificial neural network techniques: a review. *Pollution Research, 36*(3), 242-244.
38. Bączkiewicz, A., Wątróbski, J., Sałabun, W., & Kołodziejczyk, J. (2021). An ANN model trained on regional data in the prediction of particular weather conditions. *Applied Sciences, 11*(11), 4757.
39. Arhami, M., Kamali, N., & Rajabi, M. M. (2013). Predicting hourly air pollutant levels using artificial neural networks coupled with uncertainty analysis by Monte Carlo simulations. *Environmental Science and Pollution Research, 20,* 4777-4789.
40. Sarkar, A., & Kumar, R. (2012). Artificial neural networks for event based rainfall-runoff modeling. *Journal of Water Resource and Protection, 4*(10), 891.
41. Department of Environment (DOE) (2020) Malaysia Environmental Quality Report 2020. Department of Environment, Putrajaya, Malaysia. (n.d.).
42. Goel, A., Goel, A. K., & Kumar, A. (2023). The role of artificial neural network and machine learning in utilizing spatial information. *Spatial Information Research, 31*(3), 275-285.
43. Mohd Shafie, S. H., Mahmud, M., Mohamad, S., Rameli, N. L. F., Abdullah, R., & Mohamed, A. F. (2022). Influence of urban air pollution on the population in the Klang Valley, Malaysia: a spatial approach. *Ecological Processes, 11*(1), 3.
44. Shakhov, V. (2024). An Approach for Design of Air Quality Monitoring Systems. In *E3S Web of Conferences* (Vol. 496, p. 04002). EDP Sciences.
45. Strauss, T., & Von Maltitz, M. J. (2017). Generalising Ward's method for use with Manhattan distances. *PloS one, 12*(1), e0168288.