MIKAILALSYS
Journal of Multidisciplinary Sciences

e-ISSN : 2987-2286
p-ISSN : 2987-3924

Index: Harvard, Boston, Sydney, Copernicus, Sinta 4, Scilit, Semantic, Google, etc

https://doi.org/10.58578/mikailalsys.v4i3.11950

Air Quality Classification Using the Backpropagation Neural Network Algorithm on DKI Jakarta Province ISPU Data

Dila Sari & Syafriandi
Universitas Negeri Padang, Indonesia
dilasari2258@gmail.com

Article Info:

Submitted:	Revised:	Accepted:	Published:
Jul 15, 2026	Aug 12, 2026	Aug 24, 2026	Aug 29, 2026

Abstract

Air quality is a critical environmental determinant of public health and quality of life. In Indonesia, air quality is commonly monitored using the Air Pollution Standard Index (ISPU), which classifies conditions according to pollutant concentrations. This study develops an air quality classification model using a backpropagation neural network and the DKI Jakarta ISPU dataset. The input variables comprised PM_{10}, $PM_{2.5}$, SO_2, CO, O_3, and NO_2, while air quality was classified into three categories: Good, Moderate, and Unhealthy. The analytical procedure included data cleaning, min–max normalization, partitioning the dataset into 80% training and 20% testing subsets, and tuning the learning rate and number of hidden neurons. Model performance was evaluated using a confusion matrix, accuracy, Cohen's kappa, sensitivity, and specificity. The optimal model, configured with a learning rate of 0.5 and six hidden neurons, achieved an accuracy of 97.67% and a kappa value of 0.933, indicating almost perfect agreement between predicted and observed classifications. High sensitivity and specificity across all categories further confirmed the model's ability to distinguish air quality levels accurately. These findings demonstrate the effectiveness of the backpropagation neural network for classifying air quality based on DKI Jakarta ISPU data and provide a methodological basis for developing data-driven air quality monitoring and public health decision-support systems.

Keywords: Air Pollution; Air Quality Classification; Artificial Neural Network; Backpropagation Neural Network; ISPU

INTRODUCTION

Air quality is an environmental factor that directly affects human health and the community's quality of life. Increased transportation, industrial activity, and energy generation, alongside population growth, have led to rising air pollutant emissions in urban areas. According to the World Health Organization (WHO, 2024), air pollution is one of the major environmental risk factors contributing to various diseases and premature deaths globally. Exposure to air pollutants impacts health; consequently, air quality monitoring and prediction have become crucial components of environmental management (Agbehadji & Obagbuwa, 2024). Furthermore, Zhou et al. (2024) state that exposure to high concentrations of air pollutants can increase the risk of respiratory disorders and cardiovascular diseases, as well as degrade environmental quality. Therefore, air quality monitoring is crucial for supporting public health protection and sustainable environmental management.

In Indonesia, air quality is monitored using the Air Pollutant Standard Index (ISPU), which is based on several key pollutant parameters: PM10, PM2.5, sulfur dioxide (SO_2), carbon monoxide (CO), ozone (O_3), and nitrogen dioxide (NO_2). These parameters serve to characterize air pollution levels and air quality conditions in a given area (Ihsan et al., 2025; Irawan et al., 2017). ISPU data plays a crucial role in communicating air quality conditions to the public and informing environmental policy decisions. The use of ISPU data for air quality classification has also been applied to data from DKI Jakarta, utilizing these specific pollutant parameters (Ramadhan & Triayudi, 2024).

Advancements in artificial intelligence have fostered the use of various machine learning methods in the environmental sector, including for air quality classification. One widely used method is the Artificial Neural Network (ANN). According to Cabaneros et al. (2019), ANNs possess the ability to learn complex nonlinear relationships without requiring assumptions about the specific form of the relationship between variables. This capability makes ANNs well-suited for air quality data, which is typically influenced by multiple pollutant parameters.

One of the most popular learning algorithms for ANNs is Backpropagation. This algorithm operates through a feedforward process to generate an output and a backward propagation process to update network weights based on the resulting error. According to Xiaogang & Xin (2022), this learning process enables the neural network to gradually minimize prediction errors until an optimal model is achieved. Backpropagation was selected for this study due to its ability to learn complex nonlinear relationships, handle multiple input variables simultaneously, and achieve high accuracy across various classification problems. The capacity of neural networks to learn complex nonlinear patterns has made them a widely used method in air quality research (Nandi et al., 2023). Agbehadji & Obagbuwa (2024) also reported that neural network and deep learning approaches deliver strong performance when handling complex environmental data.

Various studies have utilized machine learning, neural network, and deep learning approaches for air quality analysis. Agarwal et al. (2020) demonstrated that Artificial Neural Networks could be used to predict air quality by employing pollutant parameters as input variables. Ayus et al. (2023) also showed that machine learning and deep learning methods could effectively predict air pollution. Furthermore, Gilik et al. (2022) implemented architectures based on Convolutional Neural Networks and Long Short-Term Memory for air quality prediction, while Zaini et al. (2022) demonstrated the applicability of deep learning approaches to air quality forecasting. Mottahedin et al. (2024) also applied machine learning within air quality prediction and control systems. These results indicate that neural network-based approaches hold potential for learning complex patterns in air quality data.

Beyond prediction, the neural network approach can also be applied to air quality classification. Zhu *et al.* (2025) demonstrated that neural network-based models are capable of classifying air quality based on air pollutant parameters. Sari & Purwaningsih (2022) also showed that neural network algorithms can classify air quality indices with a high level of accuracy. However, research specifically examining the classification of air quality categories using ISPU data and the Backpropagation Neural Network algorithm remains relatively limited, particularly regarding air quality data in Indonesia. Given this context, this study utilizes DKI Jakarta ISPU data from January 2024 to February 2025 comprising six pollutant parameters (PM10, PM2.5, SO_2, CO, O_3, and NO_2) to develop an air quality classification model using the Backpropagation Neural Network algorithm and to evaluate its performance based on accuracy, Kappa, sensitivity, and specificity.

METHODS

1. Research Data

The data used in this study consists of the Air Pollutant Standard Index (ISPU) for DKI Jakarta for the period from January 2024 to February 2025, obtained from official air quality data sources of the DKI Jakarta Provincial Government. The variables used are presented in Table 1.

Table 1. Research Variables

Variable	Keterangan
Y	Category (Good, Moderate, Unhealthy)
X1	PM10 (Particulate matter ≤ 10 μm)
X2	PM2.5 (Particulate matter ≤ 2,5 μm)
X3	SO_2 (Sulfur Dioxide)
X4	CO (Carbon Monoxide)
X5	O_3 (Ozone)
X6	NO_2 (Nitrogen Dioxide)

2. Analysis Stages

a. Data Cleaning

At this stage, the data is examined to identify missing values, duplicate entries, and data type inconsistencies. Records containing missing values are removed so that only complete data is used in the modeling process.

b. Data Normalization

Since the Backpropagation algorithm is sensitive to differences in data scales, all input variables are normalized using the Min-Max Normalization method, ensuring data values fall within the range of 0 to 1. Normalization is performed using the following equation:

$$x' = \frac{x - x_{min}}{x_{max} - x_{min}} \quad (1)$$

where:

x : initial data value

x_{min} : minimum value of the variable

x_{max} : maximum value of the variable

x' : normalized value

3. Variable Transformation

The air quality category variable is converted into dummy variables using the one-hot encoding method. The categories Good, Moderate, and Unhealthy are each represented as binary vectors so that they can be processed by an artificial neural network.

4. Training and Testing Data Split

The normalized data is subsequently divided into training and testing sets using the holdout method, with 80% allocated for training and 20% for testing. This proportion is commonly used in machine learning to provide sufficient data for model training while retaining an independent subset for model evaluation (Joseph, 2022).

5. Construction of the Backpropagation Neural Network Model

An air quality classification model is constructed using an Artificial Neural Network (ANN) with the Backpropagation algorithm. The network architecture consists of an input layer, a hidden layer, and an output layer. The training process begins with the feedforward stage, which involves propagating signals from the input layer to the hidden layer and the output layer (Geron, 2019). The input value for the j-th neuron in the hidden layer is calculated using the following equation:

$$z_j = \sum_{i=1}^{p} w_{ij} x_i + b_j \quad (2)$$

where x_i represents the input value, w_{ij} is the weight connecting the input neuron to the hidden neuron, and b_j is the bias. Furthermore, for the classification process, the input values are transformed into the 0–1 range using the sigmoid activation function (Purnawansyah et al., 2021), formulated as:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (3)$$

The resulting output is then compared with the actual target to obtain the error value. In this study, the error function used to achieve convergence (Mostesinos Lopez et al., 2022) is the Sum of Squared Error (SSE), formulated as follows:

$$SSE = \frac{1}{2} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \quad (4)$$

where y_i is the actual value and $\hat{y}_i$ is the network's predicted value. This error value is then propagated back to the preceding layers via backpropagation to update the network's weights and biases until a model with minimal error is obtained or a state of convergence is reached.

6. Model Evaluation

Model performance evaluation is conducted using a confusion matrix, which presents the relationship between predicted results and actual data (Tharwat, 2018). The performance metrics employed include accuracy, Kappa, sensitivity, and specificity. According to Geron (2019), accuracy measures the proportion of data correctly classified relative to the entire test dataset. Additionally, the Kappa value measures the level of agreement between classification results and actual data, accounting for agreement occurring by chance. To complement the model evaluation, sensitivity and specificity are also used to indicate the model's ability to recognize and distinguish between each classification category (Chicco & Jurman, 2020).

RESULTS

1. Data Cleaning

Table 2. Total Data

Initial data	2125
After cleaning	1940

Based on Table 2, the initial dataset consisted of 2,125 observations. Following the data cleaning process, 1,940 observations were obtained for use in the analysis. A total of 185 observations were excluded because they did not meet the required data criteria. This process was carried out to ensure that the data used for modeling was complete and consistent, thereby supporting the training and testing of the Backpropagation Neural Network model.

2. Data Normalization

Table 3. Data Normalization

PM10	PM2.5	SO_2	CO	O_3	NO_2
0.313	0.603	0.710	0.103	0.089	0.151
0.251	0.634	0.710	0.029	0.080	0.141

PM10	PM2.5	SO_2	CO	O_3	NO_2
0.246	0.359	0.710	0.959	0.080	0.135
0.078	0.336	0.623	0.191	0.188	0.227
0.229	0.443	0.623	0.147	0.312	0.184

Based on Table 3, all input variables namely PM10, PM2.5, SO_2, CO, O_3, and NO_2 have been normalized using the Min-Max method, ensuring all values fall within the 0–1 range. Normalization was performed to standardize the scales of variables that originally had different value ranges. A uniform scale makes the data more suitable for the Backpropagation Neural Network learning process and helps ensure a more stable model training process.

3. Variable Transformation

Table 4. Variable Transformation

Category	Good	Moderate	Unhealthy
Good	1	0	0
Moderate	0	1	0
Unhealthy	0	0	1

In Table 4, air quality categories are converted from text to numerical values to enable processing by the Backpropagation Neural Network model. Each category is assigned a distinct numerical combination, allowing the model to differentiate between the "Good," "Moderate," and "Unhealthy" categories during the training process.

4. Training and Testing Data Split

Tabel 5. Data Split

Training	Testing
1553	387

Based on Table 5, the majority of the data (1,553 records) is used to train the model to recognize air quality patterns, while the remaining 387 records are used to test the model's ability to provide accurate classifications.

5. Backpropagation Model Development

Tabel 6. Hasil Pengujian Kombinasi Learning Rate dan Hidden Neuron

Learning Rate \ Hidden	1	2	3	4	5	6	7
0.001	86.56	94.32	87.60	95.61	88.11	NA	88.11
0.005	86.82	NA	89.15	96.12	NA	89.92	94.32
0.01	86.82	NA	NA	NA	NA	88.89	88.11
0.05	82.43	94.32	87.60	87.60	89.92	90.44	88.63
0.1	86.82	88.63	89.41	NA	89.41	NA	95.61
0.5	82.17	93.02	95.61	88.37	96.12	97.67	88.63

Based on Table 6, the test results indicate that the combination of learning rate and the number of hidden neurons affects the performance of the Backpropagation Neural Network model. The accuracy values obtained range from 82.17% to 97.67%. The highest accuracy of 97.67% was achieved with a combination of a 0.5 learning rate and 6 hidden neurons; consequently, this configuration was selected as the best model. Additionally, several parameter combinations yielded an "NA" result, indicating that the training process failed to converge before reaching the maximum iteration limit (stepmax), preventing the model from producing an evaluation value.

6. Model Evaluation

Tabel 7. Model Evaluation Results

Evaluation	Value
Accuracy	97,67%
Kappa	0,933

Based on Table 7, the Backpropagation Neural Network model achieved an accuracy of 97.67%, indicating that the model is capable of classifying air quality with a very high level of precision. Furthermore, the Kappa value of 0.933 falls into the "Almost Perfect Agreement" category, demonstrating a very high level of agreement between the model's classification results and the actual data.

Tabel 8. Model Sensitivity and Specificity Values

Kelas	Sensitivity	Specificity
Good	0,9348	0,9912
Moderate	0,9838	0,9494
Unhealthy	0,9697	0.9943

Based on Table 8, the Sensitivity and Specificity values for all categories exceed 93%, indicating that the model possesses excellent capability in recognizing and distinguishing between each air quality category. The highest Sensitivity was achieved in the MODERATE category at 98.38%, while the highest Specificity was achieved in the UNHEALTHY category at 99.43%. These results demonstrate that the Backpropagation Neural Network model is capable of classifying air quality with a very high level of precision.

DISCUSSION

The research results demonstrate that the Backpropagation Neural Network (BPNN) algorithm is capable of classifying air quality based on six pollutant parameters: PM10, PM2.5, SO_2, CO, O_3, and NO_2. Following data cleaning and normalization, the model was developed using training and testing datasets. Testing results indicate that the model configuration influences classification performance. The optimal configuration was achieved using a learning rate of 0.5 and six hidden neurons, yielding an accuracy of 97.67%. Conversely, other configurations resulted in "NA" values because the training process failed to converge within the specified iteration limit. These findings highlight the critical role of selecting an appropriate learning rate and network structure in enabling the BPNN to effectively learn the relationship between pollutant parameters and air quality categories.

The optimal model also yielded a Kappa value of 0.933, indicating a very high level of agreement between the predicted results and the actual categories. Sensitivity and specificity values exceeding 93% across all categories further demonstrate that the model not only possesses high overall accuracy but is also capable of effectively distinguishing between air quality categories. The highest sensitivity was observed in the "Moderate" category (98.38%), highlighting the model's excellent ability to identify data belonging to that category. Meanwhile, the highest specificity was found in the "Unhealthy" category (99.43%), demonstrating the model's superior capability in distinguishing the "Unhealthy" category from the others. These results align with the findings of Sari dan Purwaningsih (2022), who achieved high accuracy using a neural network for air quality index classification. These findings also support the research Agarwal et al. (2020) and Gilik et al. (2022), which demonstrated that neural network-based approaches are capable of capturing complex patterns in air quality data.

The high classification performance demonstrates that this study contributes to the application of neural network methods for air quality analysis, specifically regarding classification based on ISPU data in Indonesia. The resulting BPNN model can serve as a supporting tool for the automated identification of air quality categories based on pollutant parameters, potentially aiding in monitoring efforts and the dissemination of air quality information to the public. However, the study's findings are currently limited to data from January 2024 to February 2025 and six pollutant parameters. Furthermore, the failure of certain configurations to converge indicates a need for further model development and optimization. Future research could utilize longer observation periods and broader geographical coverage, incorporate additional pollutant parameters, and compare the BPNN with other machine learning or deep learning methods to assess the model's robustness and generalization capabilities.

CONCLUSION

This study demonstrates that the Backpropagation Neural Network (BPNN) algorithm is capable of classifying air quality based on six pollutant parameters—PM10, PM2.5, SO_2, CO, O_3, and NO_2—using DKI Jakarta's ISPU (Air Pollution Standard Index) data. The optimal configuration was achieved with a learning rate of 0.5 and six hidden neurons, yielding an accuracy of 97.67% and a Kappa value of 0.933. High sensitivity and specificity values across all categories indicate that the model effectively recognizes and distinguishes between the "Good," "Moderate," and "Unhealthy" categories.

The results of this study contribute to the application of neural network approaches for ISPU-based air quality classification, specifically by testing BPNN parameter configurations to achieve a high-performance model. These findings suggest that BPNN serves as an effective machine learning approach for modeling the relationship between pollutant parameters and air quality categories. Future research should utilize longer data periods and broader geographical coverage, while also considering additional pollutant parameters. Further development could involve comparing BPNN with other machine learning or deep learning methods and optimizing the network architecture to achieve a model with superior generalization capabilities.

REFERENCES

Agarwal, S., Sharma, S., Suresh, R., Rahman, M. H., Vranckx, S., Maiheu, B., Blyth, L., Janssen, S., Gargava, P., Shukla, V. K., & Batra, S. (2020). Air quality forecasting using artificial neural networks with real time dynamic error correction in highly polluted regions. *Science of the Total Environment, 735*, Article 139454. https://doi.org/10.1016/j.scitotenv.2020.139454

Agbehadji, I. E., & Obagbuwa, I. C. (2024). Systematic review of machine learning and deep learning techniques for spatiotemporal air quality prediction. *Atmosphere, 15*(11), Article 1352. https://doi.org/10.3390/atmos15111352

Ayus, I., Natarajan, N., & Gupta, D. (2023). Comparison of machine learning and deep learning techniques for the prediction of air pollution: A case study from China. *Asian*

Journal of Atmospheric Environment, 17, Article 4. https://doi.org/10.1007/s44273-023-00005-w

Cabaneros, S. M., Calautit, J. K., & Hughes, B. R. (2019). A review of artificial neural network models for ambient air pollution prediction. *Environmental Modelling & Software, 119*, 285–304. https://doi.org/10.1016/j.envsoft.2019.06.014

Chicco, D., & Jurman, G. (2020). The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation. *BMC Genomics, 21*, Article 6. https://doi.org/10.1186/s12864-019-6413-7

Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems* (2nd ed.). O'Reilly Media.

Gilik, A., Öğrenci, A. S., & Özmen, A. (2022). Air quality prediction using CNN+LSTM-based hybrid deep learning architecture. *Environmental Science and Pollution Research, 29*(8), 11920–11938. https://doi.org/10.1007/s11356-021-16227-w

Ihsan, I. M., Ma'rufatin, A., Zahroh, N. F., Ikhsan, I. N., Suwedi, N., Pratama, R. A., Adhi, R. P., Handika, R., Lusia, A., Nishihashi, M., Terao, Y., Hashimoto, S., Nara, H., & Mukai, H. (2025). Air quality assessment based on real-time continuous monitoring: Particulate and nitrogen dioxide concentrations in South Tangerang. *Jurnal Teknologi Lingkungan, 26*(1), 97–104. https://doi.org/10.55981/jtl.2025.2887

Irawan, A., Sutomo, A. H., & Sukandarrumidi. (2017). Indeks standar pencemar udara dan faktor meteorologi pada kejadian ISPA. *Berita Kedokteran Masyarakat, 33*(5), 225–232. https://doi.org/10.22146/bkm.12669

Joseph, V. R. (2022). Optimal ratio for data splitting. *Statistical Analysis and Data Mining: The ASA Data Science Journal, 15*(4), 531–538. https://doi.org/10.1002/sam.11583

Montesinos López, O. A., Montesinos López, A., & Crossa, J. (2022). *Multivariate statistical machine learning methods for genomic prediction*. Springer. https://doi.org/10.1007/978-3-030-89010-0

Mottahedin, P., Chahkandi, B., Moezzi, R., Fathollahi-Fard, A. M., Ghandali, M., & Gheibi, M. (2024). Air quality prediction and control systems using machine learning and adaptive neuro-fuzzy inference system. *Heliyon, 10*(21), Article e39783. https://doi.org/10.1016/j.heliyon.2024.e39783

Nandi, B. P., Singh, G., Jain, A., & Tayal, D. K. (2024). Evolution of neural network to deep learning in prediction of air, water pollution and its Indian context. *International Journal of Environmental Science and Technology, 21*(1), 1021–1036. https://doi.org/10.1007/s13762-023-04911-y

Purnawansyah, Haviluddin, Darwis, H., Azis, H., & Salim, Y. (2021). Backpropagation neural network with combination of activation functions for inbound traffic prediction. *Knowledge Engineering and Data Science, 4*(1), 14–28. https://doi.org/10.17977/um018v4i12021p14-28

Ramadhan, D. P., & Triayudi, A. (2024). Jakarta air quality classification based on air pollutant standard index using C4.5 and Naïve Bayes algorithms. *SAGA: Journal of Technology and Information System, 2*(4), 311–327. https://doi.org/10.58905/saga.v2i4.395

Sari, E. N., & Purwaningsih, E. (2022). Air quality index classification using neural network algorithms: Klasifikasi index kualitas udara menggunakan algoritma neural network. *SYSTEMATICS, 4*(3), 473–481. https://doi.org/10.35706/sys.v4i3.7722

Tharwat, A. (2021). Classification assessment methods. *Applied Computing and Informatics, 17*(1), 168–192. https://doi.org/10.1016/j.aci.2018.08.003

World Health Organization. (2024, October 24). *Ambient (outdoor) air pollution*. https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Li, X., & Lu, X. (2022). Air quality index prediction based on multilayer perceptron. *Academic Journal of Computing & Information Science, 5*(10), 101–105. https://doi.org/10.25236/AJCIS.2022.051016

Zaini, N., Ean, L. W., Ahmed, A. N., & Malek, M. A. (2022). A systematic literature review of deep learning neural network for time series air quality forecasting. *Environmental Science and Pollution Research, 29*(4), 4958–4990. https://doi.org/10.1007/s11356-021-17442-1

Zhou, S., Wang, W., Zhu, L., Qiao, Q., & Kang, Y. (2024). Deep-learning architecture for PM2.5 concentration prediction: A review. *Environmental Science and Ecotechnology, 21*, Article 100400. https://doi.org/10.1016/j.ese.2024.100400

Zhu, J., Zhang, Z., Gu, W., Zhang, C., Xu, J., & Li, P. (2025). Air quality prediction using neural networks with improved particle swarm optimization. *Atmosphere, 16*(7), Article 870. https://doi.org/10.3390/atmos16070870