

PRONÓSTICO DE LAS CONCENTRACIONES DE MATERIAL PARTICULADO EN EL AIRE (PM₁₀) UTILIZANDO REDES NEURONALES ARTIFICIALES: CASO ESTUDIO EN EL DISTRITO DE ATE, LIMA

Jhojan Pool Rojas Quincho^{a,b,*}, Elvis Anthony Medina Dionicio^b

RESUMEN

La presente investigación tuvo como objetivo evaluar el desempeño del modelo de Redes Neuronales Artificiales (RNA) para pronosticar las concentraciones de PM₁₀ en el aire, para lo cual se hizo un caso estudio para el distrito de Ate, Lima. Para ello se desarrolló distintas arquitecturas de RNA usando como datos de entrada a los registros de contaminantes del aire y variables meteorológicas obtenidas de la Estación de Monitoreo de la Calidad del Aire "ATE" y datos simulados del modelo WRF-CHEM. Las diferentes arquitecturas de RNA pasaron por un proceso de entrenamiento y verificación, y su desempeño se evaluó mediante el Error Cuadrático Medio (ECM), la precisión (BIAS) y el coeficiente de determinación (R²). Se determinó que la arquitectura que tiene un mejor desempeño tuvo 19 neuronas en la capa oculta, con valores de 0,0230 para el ECM, 0,5308 para la BIAS y 0,823 para el R², asimismo, esta puede brindar pronósticos hasta con 6 horas de antelación. Este estudio puede contribuir a la implementación de Sistemas de Alertas Tempranas (SAT) sobre posibles incrementos en el aire de las concentraciones de PM₁₀.

Palabras clave: PM₁₀, Redes Neuronales Artificiales, RNA, Lima, contaminación del aire, modelamiento de la calidad del aire.

FORECAST OF THE CONCENTRATIONS OF PARTICULATE MATTER IN THE AIR (PM₁₀) USING ARTIFICIAL NEURAL NETWORKS: CASE STUDY IN THE DISTRICT OF ATE, LIMA.

ABSTRACT

The aim of this research was to evaluate the performance of the Artificial Neural Network (ANN) model to predict the concentrations of PM₁₀ in the air, for which a case study was made for the district of Ate, Lima. For this, different ANN architectures were developed using as input data the records of air pollutants and meteorological variables obtained from the Air Quality Monitoring Station "ATE" and simulated data from the WRF-CHEM model. The different ANN architectures went through a training and verification process, and their performance was evaluated using the Mean Square Error (MSE), precision (BIAS)

^a Facultad de Ingeniería Geológica, Minera, Metalúrgica y Geográfica, Universidad Nacional Mayor de San Marcos. Lima-11, Perú, 11-0058, E-mail: jhojan_rojas@icloud.com

^b Servicio Nacional de Meteorología e Hidrología del Perú. Jr. Cahuide 785, Jesús María, Lima, Perú

and determination coefficient (R^2). It was determined that the architecture that has a better performance had 19 neurons in the hidden layer, with values of 0,0230 for the ECM, 0,5308 for the BIAS and 0,823 for the R^2 , likewise, it can provide forecasts up to 6 hours in advance. This study can contribute to the implementation of Early Warning Systems (SAT) on possible increases in the air of PM_{10} concentrations.

Key words: PM_{10} , Artificial Neural Networks, ANN, Lima, air pollution, air quality modeling.

INTRODUCCIÓN

El material particulado con un diámetro menor a 10 micras (PM_{10}) es uno de los contaminantes atmosféricos ampliamente asociados a problemas de salud^{1,2}; puede estar compuesto por diferentes elementos químicos, como por ejemplo cadmio, plomo, carbono, arsénico³. La Organización Mundial de la Salud (OMS) ha catalogado a la contaminación del aire como un causante de cáncer en las personas⁴; asimismo, otras investigaciones han encontrado relación entre el PM_{10} y afectaciones a la salud como asma, infecciones respiratorias y enfermedades cardiovasculares⁵. En los últimos años se ha hecho más frecuente a nivel internacional el desarrollo de estudios para modelar las concentraciones de PM_{10} utilizando las bondades de las Redes Neuronales Artificiales (RNA)⁶; la cual es un modelo inspirado en el sistema nervioso biológico ya que se compone de neuronas artificiales (nodos), colocados en capas y conectados entre sí que son capaces de aprender de los patrones que se les presentan y de los errores cometidos en sus procesos de aprendizaje, identificando patrones nunca antes vistos⁷. Esta técnica ha mostrado buenos resultados para pronosticar concentraciones de material particulado; siendo los perceptrones multicapa, las arquitecturas de RNA más comúnmente aplicadas en ciencias atmosféricas debido a su capacidad para representar relaciones funcionales entre predictores y predictandos⁸. Franceschi *et al.*⁹ utilizaron las RNA para pronosticar concentraciones de PM_{10} y $PM_{2,5}$ en Bogotá, Colombia obteniendo indicadores de desempeño aceptables. Este tipo de modelos han sido usados también en Chile¹⁰ e Italia¹¹.

En el Área Metropolitana de Lima-Callao (AMLC), el crecimiento demográfico, ha afectado negativamente a la calidad de aire. El AMLC ha sido considerada como una de las ciudades más contaminadas por material particulado en Latinoamérica^{12,13}. Según informes del Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI)¹⁴ uno de los principales contaminantes que excede con frecuencia su respectivo Estándar de Calidad Ambiental para Aire (ECA-aire) es el PM_{10} . De acuerdo con Silva *et al.*¹², el AMLC posee una mala calidad del aire con respecto al PM_{10} , encontrándose concentraciones promedio diarias de $80 \mu\text{g}/\text{m}^3$, siendo las zonas norte y este las más contaminadas (concentraciones entre 90 y $110 \mu\text{g}/\text{m}^3$) y la zona centro la menos contaminada (concentraciones entre 50 y $70 \mu\text{g}/\text{m}^3$). De todas las estaciones de la Red de Monitoreo del AMLC, la que registra las mayores concentraciones de PM_{10} durante los meses de verano y otoño (período enero-mayo) es la Estación de Monitoreo de Calidad del Aire (EMCA) ATE ubicada en el distrito de Ate. Dada la versatilidad que tienen las RNA, esta investigación busca usar datos registrados de estaciones y salidas de modelos para usarlas como entrada a las RNA, debido a que estas pueden abrir una ventana

de posibilidades para la implementación de Sistemas de Alertas Tempranas sobre las posibles elevadas concentraciones de PM_{10} . Por ello, el presente estudio se enfoca en evaluar el uso de las RNA para predecir las concentraciones de PM_{10} en el distrito de Ate.

PARTE EXPERIMENTAL

Base de datos de contaminantes y variables meteorológicas

Los datos horarios de la EMCA ATE fueron proporcionados por el Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). La base de datos de contaminantes del aire y variables meteorológicas tuvieron registros de PM_{10} , material particulado con diámetro menor a 2,5 micras ($PM_{2.5}$), dióxido de nitrógeno (NO_2), temperatura del aire (T), humedad relativa (HR), velocidad del viento (VV), dirección del viento (DV) y radiación solar (RS). También fue entregado los datos provenientes de las simulaciones del modelo Weather Research and Forecasting coupled with Chemistry (WRF-CHEM), estos consistieron de datos de PM_{10_WRF} , $PM_{2.5_WRF}$, dióxido de azufre (SO_{2_WRF}), NO_{2_WRF} , monóxido de nitrógeno (NO_WRF), monóxido de carbono (CO_WRF), T_WRF , presión atmosférica en superficie ($PSFC_WRF$), componente zonal del viento a 10 metros (U_{10_WRF}) y componente meridional del viento a 10 metros (V_{10_WRF}). El periodo solicitado correspondió del 2010 al 2018 con la finalidad de analizar y entender el comportamiento de cada variable.

Ubicación de la estación de monitoreo

La EMCA ATE se encuentra situada dentro de la Plaza de Armas de “Vitarte”, perteneciente a la zonificación urbana tipo residencial, rodeada de áreas principalmente comerciales y de uso especial (educación y salud). Por su lado noroeste recorre la carretera central, vía principal de acceso al AMLC, la cual une Lima con los distintos departamentos del interior del País. La ubicación de la EMCA es mostrada en la figura 1.

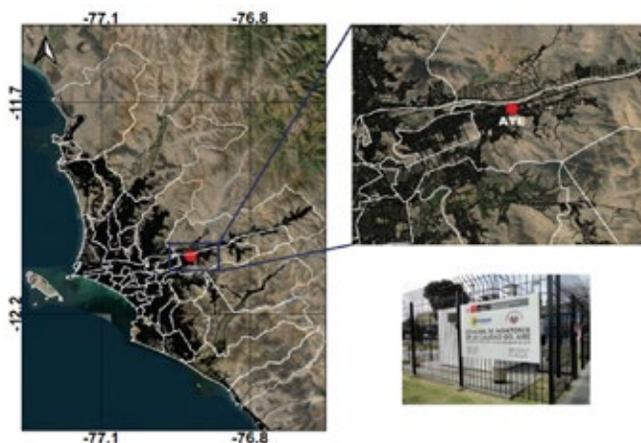


Figura 1. Mapa de ubicación de la EMCA ATE

Minería de datos

Usando la función TimeVariation del paquete Openair (software R) se determinó el patrón de comportamiento promedio horario de cada una de las variables analizadas incluida el comportamiento del PM₁₀, con el objetivo de entender su tendencia, patrón o variación durante el día, mes y año. En la figura 2 se muestra el comportamiento promedio horario del PM₁₀ durante el periodo 2010-2018.

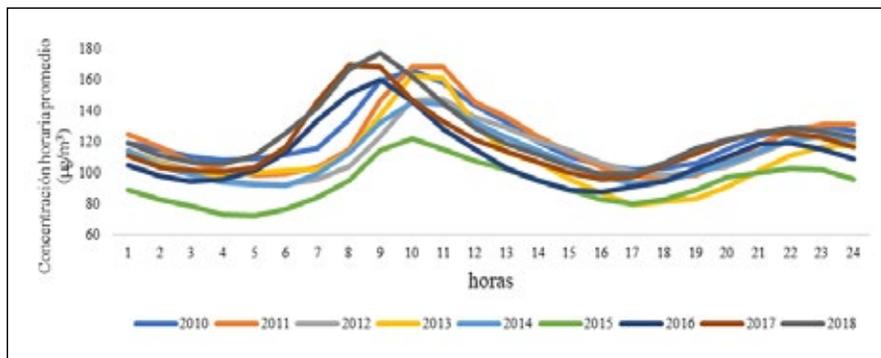


Figura 2. Variación de las concentraciones horarias promedio de PM10 en los años del 2010 al 2018

Posteriormente al análisis visual de cada variable, se procedió a buscar patrones de comportamiento que con la estadística convencional no se pueden determinar, esto con la finalidad de obtener nuevas variables para la construcción de patrón de entrada en la etapa de entrenamiento de la RNA 15,10. Las variables incluidas son las siguientes:

- Hora (HO): esta variable fue establecida a partir de las 24 horas del día (0-23).
- Día (D): Esta variable fue establecida a partir de los 7 días de la semana (1-7).
- Mes (M): Esta variable fue establecida a partir de los 12 meses del año (1-12).
- Estación (C_E): esta variable fue establecida a partir de las estaciones del año. Asignando valores numéricos de la siguiente manera, verano (1 – a partir del 21 de diciembre al 20 de marzo), otoño (2 – a partir del 21 de marzo al 21 de junio), primavera (3 – a partir del 22 de junio al 21 de setiembre) e invierno (4 – a partir del 22 de setiembre al 20 de diciembre).
- Concentración horaria (C_H): la variable fue obtenida, asignando etiquetas de nocturno (0) y diurno (1), es decir desde las 6:00 hasta las 18:00 horas (diurno) y desde las 19 hasta las 5:00 horas del día siguiente (nocturno).

Se incorporó más variables que representen la relación o asociación de los datos meteorológicos y de contaminantes del aire₁₀. Las variables obtenidas se muestran a continuación:

- Promedio de la temperatura horaria máxima del día anterior (MTND).
- Temperatura horaria máxima entre las 19:00 horas del día anterior y 18:00 horas del presente día (TMND).

- Diferencia entre la temperatura horaria máxima y mínima del día anterior (DMTN).
- Diferencia entre la temperatura horaria máxima y mínima de entre las 19:00 horas del día anterior y 18:00 horas del presente día (DMMT).
- Humedad relativa horaria promedio mínima del día anterior (MPHN).
- Humedad relativa horaria promedio mínima entre las 19:00 horas del día anterior y 18 horas del presente día (MPPHN).

Finalmente se construyó un patrón de entrada con 38 variables, el cual es mostrado en la tabla 1, junto con sus valores mínimo, promedio, máximo, conteo y su respectiva unidad.

Tabla 1. Variables que conformaron el patrón de entrada para el entrenamiento de las RNA

Parámetros Entradas	Mínimo	Media	Máximo	Recuento	Unidades
PM _{2.5}	4,08	45,33	276,8	3924	µg/m ³
Ratio PM _{2.5/10}	0,035	0,3987	0,9934	3924	-
NO ₂	0	32,65	137,46	3924	µg/m ³
T	13,3	20,38	29,1	3924	° C
HR	42	76,5	100	3924	%
VV	0,1	1,2	3,3	3924	m/s
DV	21	238,44	347	3924	°
RS	0	587,81	3419,64	3924	W/m ²
PM _{10_WRF}	4,73	48,47	201,61	3924	µg/m ³
PM _{2.5_WRF}	3,06	40,89	178,9	3924	µg/m ³
Ratio PM _{2.5/10-WRF}	0,6	0,8	0,8876	3924	-
SO _{2-WRF}	0,0008164	0,007261418	0,0310053	3924	µg/m ³
NO _{2_WRF}	0,0154325	0,208509847	0,988407	3924	µg/m ³
NO_WRF	0,000001	0,031500848	0,4280814	3924	µg/m ³
CO_WRF	0,2661	1,902243479	7,8237629	3924	µg/m ³
T_WRF	13,6577698	21,01975796	27,8684631	3924	° C
PSFC_WRF	94862,86719	95319,81709	96483,88281	3924	hPa
U _{10_WRF}	-5,771975	-0,749438696	4,7972865	3924	°
V _{10_WRF}	-1,5429325	1,967574255	7,3669128	3924	°
ws_WRF	0,03938043	3,097598915	8,329012752	3924	m/s
wd_WRF	0,289666505	180,9486104	358,9540763	3924	°
Angulo_WRF	-89,92420412	-18,86790338	89,98832141	3924	°
HO	0	-	23	3924	Horas.
D	1 (lunes)	-	7 (Domingo)	3924	día
M	1 (enero)	-	12 (diciembre)	3924	meses
C_E	1 (verano)	-	4 (primavera)	3924	estaciones del año
C_H	1 (día)	-	2 (noche)	3924	horarios
KM	1	1,773	2	3924	Clúster
MTND	16,5	24,2	29,1	3924	° C
TMND	16,5	24,19	29,1	3924	° C
DMTN	2,1	6,3	8,8	3924	° C
DMMT	1,5	6,26	9,7	3924	° C
MPHN	42	60,64	87	3924	%
MPPHN	42	60,66	87	3924	%
wsp	0,94	1,98	1,61	3924	m/s
wspn	0,92	1,19	1,47	3924	m/s
wdi	0,00000076	1	1,999	3924	-
Salida					
PM ₁₀	13,57	118,2446507	676,3	3924	µg/m ³

Entrenamiento y Validación de las Redes Neuronales Artificiales

Normalización de las RNA

Para la configuración de las RNA, se utilizó el Software MATLAB R2008a el cual tiene el toolbox (nntool). Previo a la configuración, la base de datos pasó por un proceso de normalización utilizando la ecuación que se indica a continuación:

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

En donde X' es el valor de la variable normalizada, x es el valor de la variable sin normalizar, x_{min} es el valor mínimo de la variable sin normalizar y x_{max} es el valor máximo de la variable sin normalizar. Con este proceso se obtuvo valores entre 0 y 1, los cuales no estaban afectadas por unidades (adimensionales).

Arquitectura de la red neuronal artificial

No existe un criterio determinado para la selección de la arquitectura de las RNA¹⁶. En este estudio, se determinó la arquitectura óptima para el pronóstico del PM₁₀, en base a lo realizado en investigaciones relacionadas (Tabla 2).

Tabla 2. Referencias para la determinación de la arquitectura de una red neuronal artificial para pronóstico de PM₁₀

Referencia	Cantidad de Patrones de entrenamiento	# de capas	# Capas ocultas	# Capas salidas	# Neuronas en la Capa Oculta	Tipo de Funciones de Activación
(Espinosa et al., 2017) ¹⁷	8	3	1	1	30	Logsig - purelin
(Park et al., 2017) ¹⁸	3	3	1	1	-	Tangent -purelin
(Pérez y Gramsh, 2016) ¹⁹	13	3	1	1	8	-
(Feng et al., 2015) ²⁰	10	3	1	1	8	Sigmoid - purelin
(Fernando et al., 2011) ²¹	5	3	1	1	4	-
(Cortina, 2012) ¹⁵	N	3	1	1	N/2	Sigmoid - purelin

Validación del modelo de RNA

Para la validación de la red neuronal, se utilizó información de la concentración promedio horario del PM₁₀ del periodo 2017 (febrero) como variable objetivo y como variables predictoras a los datos de la estación de monitoreo de la calidad de aire ATE, los datos del modelo WRF-CHEM y los patrones encontrados después del análisis de minería de datos.

Evaluación del desempeño del modelo RNA

Para evaluar el desempeño de los modelos de RNA, se emplearon tres indicadores: el coeficiente de correlación (R²), la raíz del error cuadrático medio (ECM) y BIAS. Cuanto mayor es el R² y menor la ECM, mejor será el efecto de predicción de los modelos¹⁸. Los índices se pueden calcular mediante las siguientes ecuaciones:

$$R^2 = \frac{(\sum_{i=1}^N (y_p^i - \bar{y}_p)(y_o^i - \bar{y}_o))^2}{\sum_{i=1}^N (y_p^i - \bar{y}_p)^2 \sum_{i=1}^N (y_o^i - \bar{y}_o)^2} \dots (2)$$

$$ECM = \sqrt{\frac{\sum_{i=1}^N (y_p^i - y_o^i)^2}{N}} \dots (3)$$

Donde y_pⁱ y y_oⁱ son los i-ésimos valores predichos y observados; y_p y y_o son el promedio del valor previsto y observado; N es el número de muestra.

RESULTADOS Y DISCUSIÓN

Arquitectura óptima de la red neuronal

La arquitectura optima (tabla 3), fue determinada en función de las investigaciones revisadas. Como algoritmo de aprendizaje, se adoptó el descenso de gradiente con el algoritmo de retropropagación (Backpropagation) ampliamente utilizado^{21,15}. En adición, se utilizó la función “Logsig” y “purelin” como las funciones de transferencia de la capa oculta y capa de salida respectivamente.

Tabla 3. Arquitectura óptima de la red neuronal utilizada para cada hora de antelación

Patrones de entrenamiento	# de capas	Capas ocultas	Capa de salida	Neuronas en la capa oculta	Funciones de activación	Algoritmo de entrenamiento y validación
38	3	1	1	19	Logsig - purelin	Levenberg - Marquardt

Estimación de la concentración promedio de PM_{10}

Se estimó las concentraciones de PM_{10} para cada hora de antelación (1 a 24 horas) usando la arquitectura óptima. El desempeño de cada una, se evaluó mediante los indicadores BIAS, ECM y R^2 (Tabla 4).

Tabla 4. Estimación del PM_{10} con 1 a 24 horas de antelación en la etapa de validación.

Horas de antelación	BIAS	ECM	R^2
1	0,5308	0,0230	0,823
3	0,6997	0,0265	0,78
6	1,1932	0,0345	0,812
9	1,2580	0,0355	0,660
12	1,2757	0,0357	0,637
15	1,4544	0,0381	0,687
18	1,4652	0,0383	0,641
21	2,4767	0,0498	0,663
24	2,7350	0,0523	0,502

Los modelos que obtuvieron mayor ajuste y menor error para la estimación del PM_{10} , fueron los que se elaboraron con 1, 3 y 6 horas de antelación. En la Figura 3, se observa la correlación entre las salidas del modelo neuronal con mejor desempeño para la estimación de la concentración del PM_{10} con 1 hora de antelación y los datos observados del PM_{10} , obteniendo un coeficiente de correlación (R^2) de 0,823, un ECM de 0,0230 y un sesgo (BIAS) de 0,5308. Asimismo, se aprecia las series temporales de los datos observados y modelados para el mismo periodo.

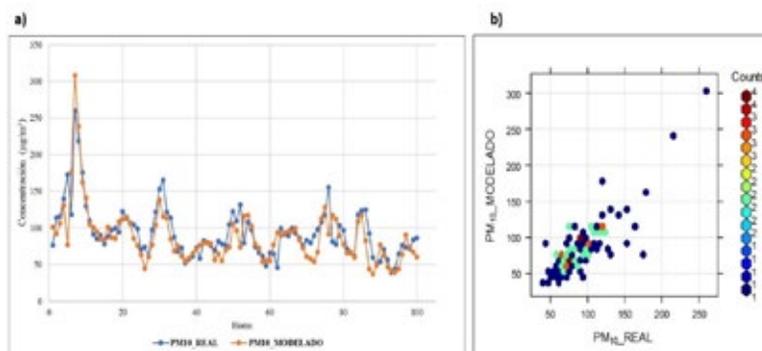


Figura 3. Estimación de las concentraciones del PM_{10} con 1 hora de antelación. a) Serie de tiempo de datos observados y modelados, b) Correlación de los datos observados y modelados

En la estimación de la concentración del PM_{10} con 3 horas de antelación, se obtuvo un coeficiente de correlación (R^2) de 0,78, un ECM de 0,0265 y un sesgo (BIAS) de 0,6997. Asimismo, en la figura 4 se aprecia las series temporales de los datos observados y modelados para el mismo periodo.

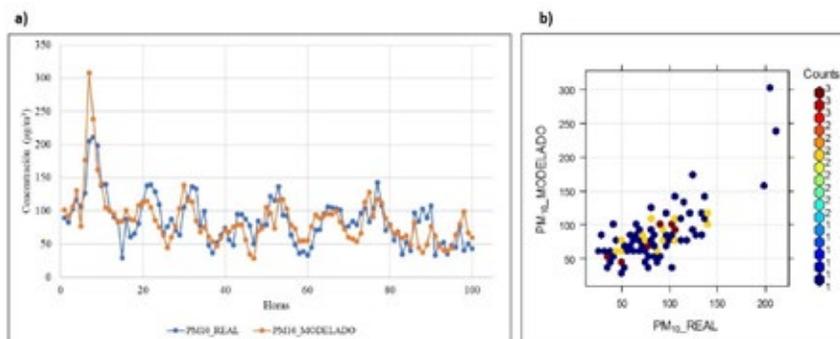


Figura 4. Estimación de las concentraciones de PM_{10} con 3 horas de anticipación. a) Serie de tiempo de datos observados y modelados, b) Correlación de los datos observados y modelado.

Para la estimación de la concentración del PM_{10} con 6 horas de antelación se obtuvo un coeficiente de correlación (R^2) de 0,812, un ECM de 0,0345 y un sesgo (BIAS) de 1,1932. Asimismo, se aprecian las series temporales de los datos observados y modelados para el mismo periodo (Figura 5).

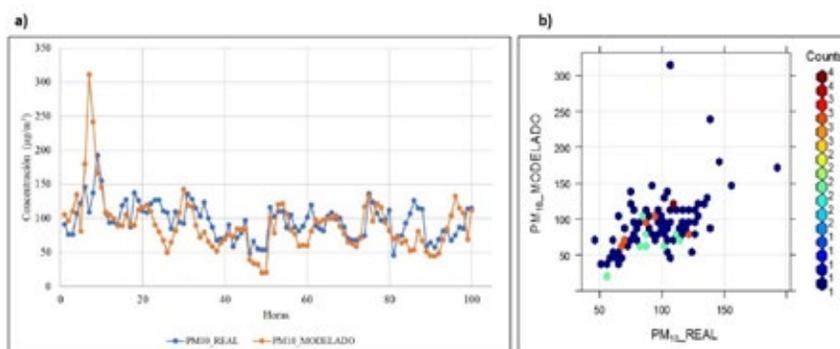


Figura 5. Estimación de las concentraciones de PM_{10} con 6 horas de anticipación. a) Serie de tiempo de datos observados y modelados, b) Correlación de los datos observados y modelados

La precisión de una estimación con redes neuronales es muy afectada por la arquitectura, asimismo, no existe un criterio determinado para la selección de esta ¹⁶, por lo que los investigadores se basan en las experiencias para ajustarlas a sus realidades. Tal cual queda en evidencia cuando se ve que al aumentar la cantidad de horas de antelación para la estimación del PM_{10} de 3 a 6, el R^2 incrementa de 0.78 a 0.812, Por lo cual, se tuvo que experimentar y a la vez investigar más sobre estudios parecidos y con resultados aceptables para poder así elaborar la Tabla 2. Otra de las consideraciones que se tuvo en la investigación fue el de encontrar asociaciones entre las variables meteorológicas y las concentraciones de PM_{10} , obteniendo nuevas variables para el entrenamiento de la red neuronal, lo que conlleva a que la red neuronal diseñada obtenga un mejor desempeño al igual que Pérez y Reyes¹⁰.

En la etapa de validación de cada arquitectura óptima para cada hora de antelación, se encontró de que a medida que incrementaban las horas de estimación el ECM y el BIAS, incrementaban y el R^2 disminuía (lo cual quedó más claro a partir del uso de 9 horas de antelación), como se puede observar en la Tabla 4. Este comportamiento es consistente a lo descrito en varias investigaciones que tienen el fin de realizar predicciones de las concentraciones del PM_{10} ^{15,22,17,21}, pues las concentraciones del PM_{10} provienen de muchas fuentes de emisión y estas pueden sufrir cambios al pasar las horas por efecto de la meteorología¹².

Los resultados de la estimación pudieron verse afectados por la cantidad de datos disponibles para el entrenamiento de la red neuronal ^{10,23,24,15}, limitante que se encontró en el desarrollo de esta investigación. Sin embargo, las redes diseñadas son capaces de describir las tendencias de las concentraciones del PM_{10} como se puede visualizar en los resultados, de las cuales las que obtuvieron mayor ajuste y menos error, fueron las de 1, 3 y 6 horas de antelación respectivamente.

CONCLUSIONES

- La red neuronal diseñada para el pronóstico de la concentración del PM_{10} para distintas horas de antelación en el distrito de Ate, tiene la capacidad de pronosticar las concentraciones de PM_{10} con un buen desempeño (bajo error cuadrático medio y sesgo BIAS así como también un alto ajuste), hasta con 6 horas de antelación. Para una estimación de 9 a 24 horas de antelación, se encontró que las salidas no muestran un buen ajuste, sin embargo, las tendencias obtenidas son capaces de describir si las concentraciones de PM_{10} incrementarán o disminuirán.
- La arquitectura más óptima para la estimación de las concentraciones del PM_{10} en el distrito de Ate, posee: 3 capas; 38 variables de entrada (inputs); 19 neuronas en la capa oculta; una variable objetivo (target); utilizando una función de activación sigmooidal (capa oculta) y una lineal (capa de salida).

AGRADECIMIENTOS

Al SENAMHI por proporcionar los datos observados y modelados para introducirlos a las RNA. A los ingenieros Erik Gregorio Merino Navarro y David Fernando Vargas La Rosa por su apoyo en la revisión exhaustiva del presente artículo científico.

REFERENCIAS BIBLIOGRÁFICAS

1. Terzi E, Argyropoulos G, Bougatioti A, Mihalopoulos N, Nikolaou K, Samara C. Chemical composition and mass closure of ambient PM_{10} at urban sites. *Atmospheric Environment*. 2010; 44(18): 2231–2239.
2. Pallarés S, Gómez E, Martínez A, Jordán MM. The relationship between indoor and outdoor levels of PM_{10} and its chemical composition at schools in a coastal region in Spain. *Heliyon*. 2019; 5(8):e02270. doi: 10.1016/j.heliyon.2019.e02270.
3. Xue H, Liu G, Zhang H, Hu R, Wang X. Similarities and differences in PM_{10} and $PM_{2.5}$ concentrations, chemical compositions and sources in Hefei City, China. *Chemosphere*. 2019; 220: 760–765.
4. Organización Mundial de la Salud (OMS). Contaminación del aire doméstico y salud [Internet]. [citado 20 de octubre del 2022]. Disponible en: <https://www.who.int/es/news-room/fact-sheets/detail/household-air-pollution-and-health>
5. Wu X, Chen B, Wen T, Habib A, Shi G. Concentrations and chemical compositions of PM_{10} during hazy and non-hazy days in Beijing. *J Environ Sci (China)*. 2020;87:1-9.
6. Biancofiore F, Busilacchio M, Verdecchia M, Tomassetti B, Aruffo E, Bianco S et al. Recursive neural network model for analysis and forecast of PM_{10} and $PM_{2.5}$. *Atmos Pollut Res*. 2017; 8(4): 652–659.
7. Durão RM, Mendes MT, João Pereira M. Forecasting O_3 levels in industrial area surroundings up to 24 h in advance, combining classification trees and MLP models. *Atmos Pollut Res*. 2016; 7: 961–970.
8. Gardner MW, Dorling SR. Regression modelling of hourly $NO(x)$ and NO_2 concentrations in urban air in London. *Atmospheric Environment*. 1997; 31(24): 4081–4094.
9. Franceschi F, Cobo M, Figueredo M. Discovering relationships and forecasting PM_{10} and $PM_{2.5}$ concentrations in Bogotá Colombia, using Artificial Neural Networks, Principal Component Analysis, and k-means clustering. *Atmos Pollut Res*. 2018; 9(5): 912–922.
10. Perez P, Reyes J. Prediction of maximum of 24 h average of PM_{10} concentrations 30 h in advance in Santiago, Chile. *Atmos Environ*. 2002; 36(28): 4555–4561.
11. Campanile L, Cantiello P, Iacono M, Lotio R, Marulli F y Mastroianni M. Applying Machine Learning to Weather and Pollution Data Analysis for a Better Management of Local Areas: The Case of Napoli, Italy. In Proceedings of the 6th International Conference on Internet of Things, Big Data and Security; 22-24 April 2021. p. 354–363. doi: 10.5220/0010540003540363.

12. Silva J, Rojas J, Norabuena M, Molina C, Toro R, Leiva M. Particulate matter levels in a South American megacity the metropolitan área of Lima-Callao, Perú. *Environ Monit Assess.* 2017; 189(12): 1-18.
13. IQAir. World air quality report 2021 [Internet]. [citado 20 de octubre del 2022]. Disponible en: <https://www.iqair.com/world-most-polluted-cities/world-air-quality-report-2021-en.pdf>
14. Servicio Nacional de Meteorología e Hidrología del Perú (SENAMHI). Evaluación de la Calidad del Aire en Lima Metropolitana 2015 [Internet]. Lima, Perú; 2016 [Citado 10 de enero 2022]. Disponible en: <https://www.senamhi.gob.pe/load/file/01403SENA-7.pdf>
15. Cortina M. Aplicación de técnicas de inteligencia artificial a la estimación de contaminantes atmosféricos [Tesis doctoral]. Madrid: Universidad Politécnica de Madrid; 2012.
16. Perez P, Reyes J. Prediction of particulate air pollution using neural techniques. *Neural Comput Appl.* 2001; 10: 167-171.
17. Espinoza A, May O, Balan I, Reyes J, Pérez I, Bassam A. Modelado de partículas PM10 y PM2.5 mediante redes neuronales artificiales sobre clima tropical de San Francisco de Campeche, México. *Quimi Nova.* 2017; 9(40): 1025-1034.
18. Park S, Kim M, Kim M, Namgung HG, Kim KT, Cho KH, et al. Predicting PM10 concentration in Seoul metropolitan subway stations using artificial neural network (ANN). *J Hazard Mater.* 2018; 341:75-82.
19. Perez P, Gramsch E. Forecasting hourly PM2.5 in Santiago de Chile with emphasis on night episodes. *Atmos Environ.* 2016; 124(Pt A):22-27.
20. Feng X, Qi L, Zhu Y, Hou J, Jin L, Wang J. Artificial neural networks forecasting of PM2.5 pollution using air mass trajectory based geographic model and wavelet transformation. *Atmos Environ.* 2015; 107: 118-128.
21. Fernando H, Mammarella M, Grandoni G, Fedele P, Di Marco R, Dimitrova R, Hyde P. Forecasting PM10 in metropolitan areas: Efficacy of neural networks. *Environ Pollut.* 2012; 163: 62-63.
22. Díaz L, Ortega J, Fu J, Reed G, Chow J, Watson J, Moncada J. A hybrid ARIMA and artificial neural networks model to forecast particulate matter in urban áreas: The case of Temuco, Chile. *Atmos Environ.* 2008; 42: 8331-8340.
23. Infante S, Ortega J, Cedeño F. Estimación de datos faltantes en estaciones meteorológicas de Venezuela vía un modelo de redes neuronales. *Revista de Climatología.* 2008; 8: 51-70.
24. Alvarado S, Silva C, Cáceres D. Modeling critical episodes of air pollution by PM₁₀ in Santiago, Chile. Comparison of the predictive efficiency of parametric and non-parametric statistical models. *Gac Sanit.* 2010; 24 (6): 466-472.