Noise estimation using an artificial neural network in the urban area of Jaen, Cajamarca

Wendy Díaz[1], Anali Tarrillo[1], Candy Ocaña[1,2], Lenin Quiñones[1,2]
[1]Escuela Académica Profesional de Ingeniería Forestal y Ambiental, Universidad Nacional de Jaén, Perú
[2]Instituto de Ciencia de Datos, Universidad Nacional de Jaén, Perú

Article Info

Article history:

Received Aug 25, 2022
Revised Nov 2, 2022
Accepted Nov 27, 2022

Keywords:

Artificial intelligence
Expert system
Information technology
Noise pollution
Urban traffic

ABSTRACT

Jaen is a city in constant urban growth which generates an increase in vehicular traffic and active noise pollution. The research presents the development of an artificial neural network (ANN) to estimate the noise produced by vehicular traffic in the urban area of the city. Consequently, information was collected from two investigations coded as T1 and T2, for which a matrix of 10 variables was elaborated with 210 and 273 data respectively. Random random sampling was performed to divide the data matrix into 80% (training) and 20% (validation). Weka software and the multi-layer perceptron (MLP) training algorithm were used to model the ANN. An ANN for T1 with 6-19-1 architecture and an ANN for T2 with 6-15-1 architecture were obtained. The performance of the ANNs was evaluated using the correlation coefficient (R), coefficient of determination (R2) and root mean square error (RMSE). The results show that the MLP networks are able to estimate the sound pressure level with values of R=0.9927, R2=0.9854 and RMSE=0.7313 for T1, R=0.9989, R2=0.9978, and RMSE=0.1515 for T2.

This is an open access article under the CC BY-SA license.

Corresponding Author:

Candy Ocaña
Escuela Académica Profesional de Ingeniería Forestal y Ambiental
Instituto de Ciencia de Datos, Universidad Nacional de Jaén
Calle Higinio Ortiz # 511, Jaén, Cajamarca, Perú
Email: candy.ocana@unj.edu.pe

1. INTRODUCTION

Noise pollution generates excessive noise that causes annoyance to living organisms. This noise is considered one of the critical environmental problems because it endangers the health of the population; causing affectations such as cardiovascular diseases, hearing problems, sleep disorders and adverse social behaviour [1], [2], in addition to this, [3] found that this type of pollution influences the incidence and severity of COVID-19, because it generates high levels of cortisol, weakening the immune system. Vehicle traffic is one of the main sources of this pollution in cities [4], [5]. In general, there are two sets of factors that influence noise annoyance: i) related to the physical characteristics of the sound (type of noise, level, duration, and frequency spectrum), the time of day it occurs and the exposure and ii) related to the individual, including physiological, psychological, and social characteristics that affect the subjective perception of noise [5], [6].

Vehicles are predominantly sources of low and medium frequency noise, which has a high penetrating power and propagates with low dissipative absorption over long distances. The continuous growth of the car fleet has progressively increased the need for special attention to urban traffic noise, which not only increases in line with the growth of residential, industrial and commercial areas, but also causes adverse impact of noise emissions on people [7], [8]. Predicting the level of noise produced by urban transport is an essential aspect of mitigating environmental pollution. Therefore, it is necessary to have appropriate and specific mathematical tools (models) that can reproduce or simulate different acoustic scenarios for use in assessing and planning urban planning activities [9], [10]. Artificial neural networks (ANNs) have recently emerged as an important area of research, not only for their general ability to process noise data, but also to learn and store it, specifically this computational model allows prediction and optimisation of traffic noise descriptors [4], [11], [12].

Studies have proposed method for assessing and predicting noise in the environment. These prediction methods are mainly classified into three groups: physical propagation models, traditional statistical methods and machine learning methods [13], [14]. Deep learning derives from the study of ANNs, which in many areas of data science have demonstrated a remarkable ability to learn complex, non-linear relationships between sets of variables [15], [16]. The ANN is inspired by the nature of real dynamic systems emulating the human brain, where neurons are layered and interconnected by mathematical functions. Each neuron receives a weighted signal from the previous layer, which is processed to learn from the examples provided by training algorithms [17]. Common neural network models include multilayer perceptron (MLP), convolutional neural networks (CNN), and recurrent neural networks (RNN) [18]. These algorithms iteratively update the model parameters until the error between the actual value of the output variable and the experimental one is minimised [4], [13], [19], [20].

The research proposes the use of a MLP ANN as a model for predicting urban environmental noise in the city of Jaén. It was proposed to use the MLP network type for its versatility and prediction capacity. The knowdlege discovery in databases (KDD) method was used for the development of the network, as a sequence of ordered steps that allowed accurate information to be obtained. The modelling was carried out in Weka software, introducing data obtained from environmental monitoring of research carried out in the same city, with authors between 2016-2020; for training and validation we opted for the division of the data (80-20%) in order to avoid the complexity of the model and therefore an overfitting of the model (overfitting).

2. METHOD

Research on sound evaluation in the urban area of Jaén for the period 2016-2020 was taken into account. The databases were reviewed from institutional repositories of Peruvian universities and were subjected to evaluation considering standardisation criteria in the information considered by the authors. We opted for those data sources that consider the sound pressure level of the vehicle fleet and that consider the variables (inclusion and exclusion criteria): name of road, location coordinates of sampling points, time and date of data collection, maximum sound pressure level (Lmax), minimum sound pressure level (Lmin), number of motokar per time unit and sampling point, number of linear moto per time unit and sampling point, number of cars per time unit and sampling point, equivalent continuous sound pressure level (LAeqT). The ANNs were developed using the KDD method, in the free software Weka and with the backpropagation learning algorithm (80% of the data for training and 20% for validation).

To select independent variables influencing the equivalent continuous sound pressure level, the CorrelationAtributeEval evaluator attribute was used, prioritising input variables with significance values greater than or equal to 0.1. The correlation coefficient (R), the coefficient of determination (R^2) and the root mean square error (RMSE) are selected as criteria for evaluating model performance. To visualise the relationship between the actual and the protonistic sound pressure level (LAeqT), the variables used in Weka were simulated in SPSS based on linear regression. Figure 1 (in Appendix) shows the methodological flow for the development of the proposed neural network.

3. RESULTS AND DISCUSSION

3.1. Collection of noise pollution data from the urban area of Jaen

The institutional repositories of the country's universities were consulted during the period 2016-2020. Five data sources were preliminarily evaluated (Table 1), of which two met the inclusion and exclusion criteria previously detailed: i) sound pressure level by the vehicle fleet in the city of Jaén, from December 2018 to February 2019 (T1) and ii) assessment of vehicular noise pollution based on the supreme decree N°085-2003-PCM regulation of environmental quality standards for noise carried out in the province of Jaen, Department of Cajamarca, 2016 (T2). It was verified that the data collected and considered in the study were obtained by sound level meters calibrated by the National Institute of Quality (INACAL) of Peru.

3.2. Artificial neural network for estimating noise pollution in the urban area of Jaen

3.2.1. Attribute selection for ANN, T1 and T2

The data obtained for both T1 and T2 were divided into 80% for training and 20% for validation. The CorrelationAttributeEval attribute evaluator and the ranker search method in Weka software were used with 80% of the data for both T1 and T2, obtaining the importance values for each input variable.

Tables 2 and 3 show the nine importance values for each variable evaluated for T1 and T2, however, in order to propose the number of input variables, the values of importance <0.1 were chosen, reducing the input variables for the models to six.

Table 1. Sources of literature review

Name of research	Author	University
Evaluation of sound pressure levels in commercial establishments in the urban area of the city of Jaen, based on supreme decree N°085-2003-PCM	Gianela Olivera Zurita Kiara Belkiss Silva Vega	National University of Jaen
Environmental Quality Standards (ECAS) for noise in the main university higher education centres in the city of Jaen	Felipe Nery Silva Cabrera	National University of Jaen
Sound pressure levels in the markets of the city of Jaen, Cajamarca -2019	Katiri Tatiana Estela Carranza Jefferson Jair Goicochea Pérez	National University of Jaen
Evaluation of vehicular noise pollution based on the Supreme Decree N°085-2003-PCM Regulation of Environmental Quality Standards for Noise carried out in the province of Jaén, department of Cajamarca, 2016	Cintia Karely Cruzado Ancajima Yanira Susana Soto Medina	National University of Jaen
Sound pressure level by the vehicle fleet in the city of Jaén, December 2018 to February 2019	Elser Burga Mendoza	National University of Jaen

Table 2. T1, 80% attribute selection

Variable	Importance values
Lmax	0.981
Lmin	0.7142
CoordenadaUTM	0.2799
NombredelaVía	0.2799
Hora	0.155
Motokar	0.1481
Carros	0.0868
MotoLineal	0.0536
Fecha	-0.0327

Table 3. T2, 80%, attribute selection

Variable	Importance values
Lmax	0.8943
Lmin	0.3823
MotoLineal	0.1882
Motokar	0.1818
CoordenadasUTM	0.1543
NombredelaVía	0.1475
Carros	0.0788
Hora	0.0655
Fecha	0.051

3.2.2. Training of ANNs

For T1, a 6-19-1 architecture ANN was obtained; six (6) input neurons, a hidden layer with nineteen (19) nodes and an output layer with the dependent variable LAeqT, Figure 2(a); on the contrary, for T2, a 6-15-1 architecture ANN was obtained; six (6) input neurons, a hidden layer with fifteen (15) nodes and an output layer with the dependent variable LAeqT, Figure 2(b). The fit statistics used for the training set of ANN-T1 and ANN-T2 were R, R^2 and RMSE (Table 4); acceptable values were obtained for both models. For T1 and T2 the R and R^2 presented values close to unity showing a good performance of the model; unlike the values obtained for the RMSE, which showed values close to unity and therefore a higher error rate, this due to the rate of learning in the training process (high learning rates will make the training converge faster, but the fit of the trainable parameters will be less accurate and will result in higher error rates) [21]. The results of the training of the networks are shown in Figures 3 and 4, where the results of the dispersion of predicted LAeqt values and actual LAeqt are graphically represented. The relationship between the observed LAeqt and the estimated LAeqt, for model T1, presents a line of fit of equation $y=2.48+1.01\ x$, with coefficient of determination $R^2=0.98$, on the contrary, model T2 presents a line of fit of equation $y=13.05+0.83\ x$, with $R^2=0.92$.

3.3. Validation of the artificial neural network

For the validation of the ANNs, the same adjustment statistics used in the training stage were used, as shown in Table 5. However, in this case 20% of the data from each investigation was used. For both models' good values of R and R2 were obtained due to their closeness to unity; differing in the RMSE, as T2 presented a more positive value compared to T1, but for both models' acceptable values were obtained.

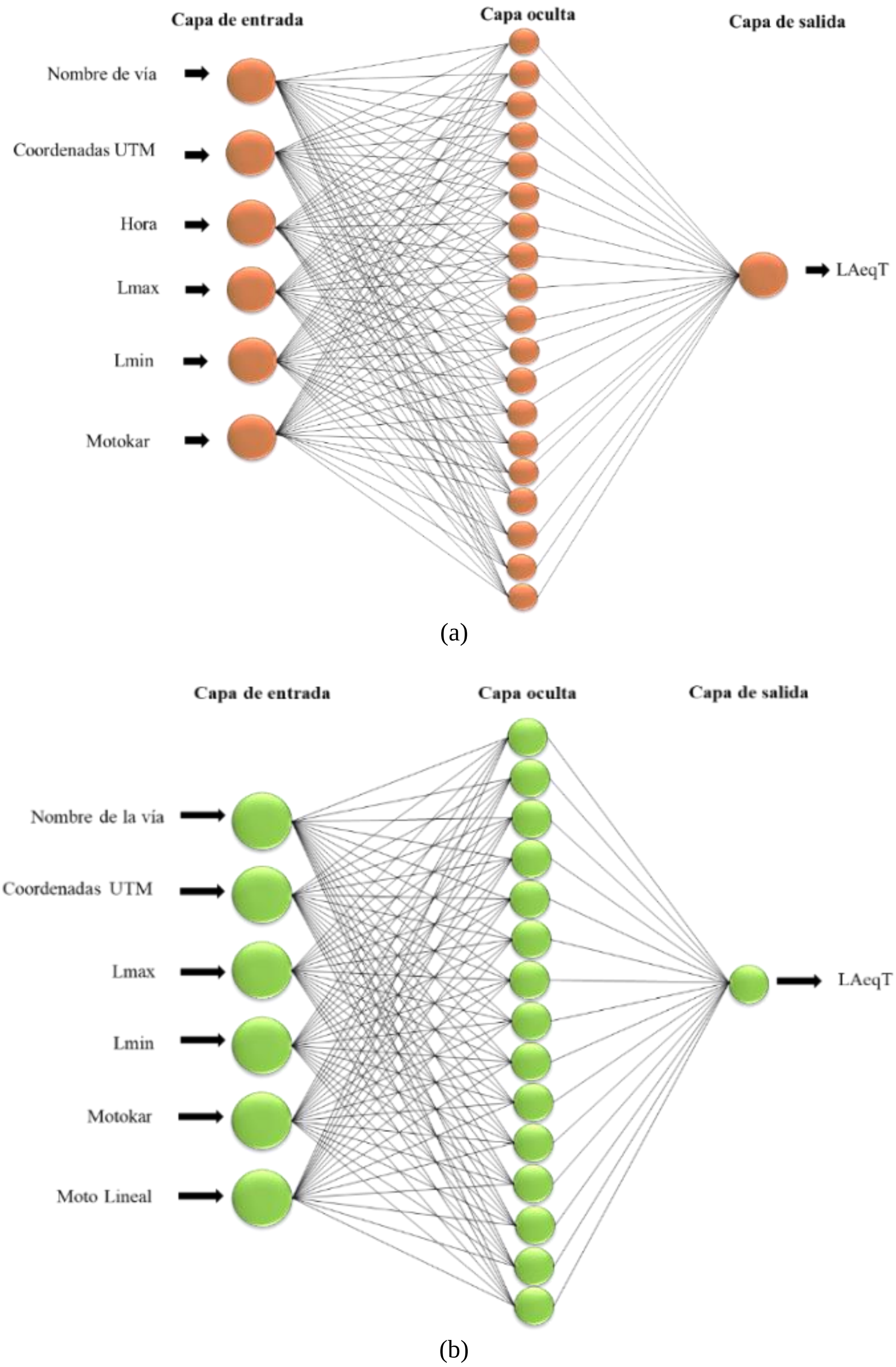

Figure 2. Validation of the ANNs (a) architecture of T1-TNA and (b) architecture of T2-TNNA

Table 4. Values of statistics used in RNA-T1 and RNA-T2 training

Statistic	RNA-T1	RNA-T2
R	0.9861	0.9606
R^2	0.9723	0.9227
RMSE	0.7952	0.99

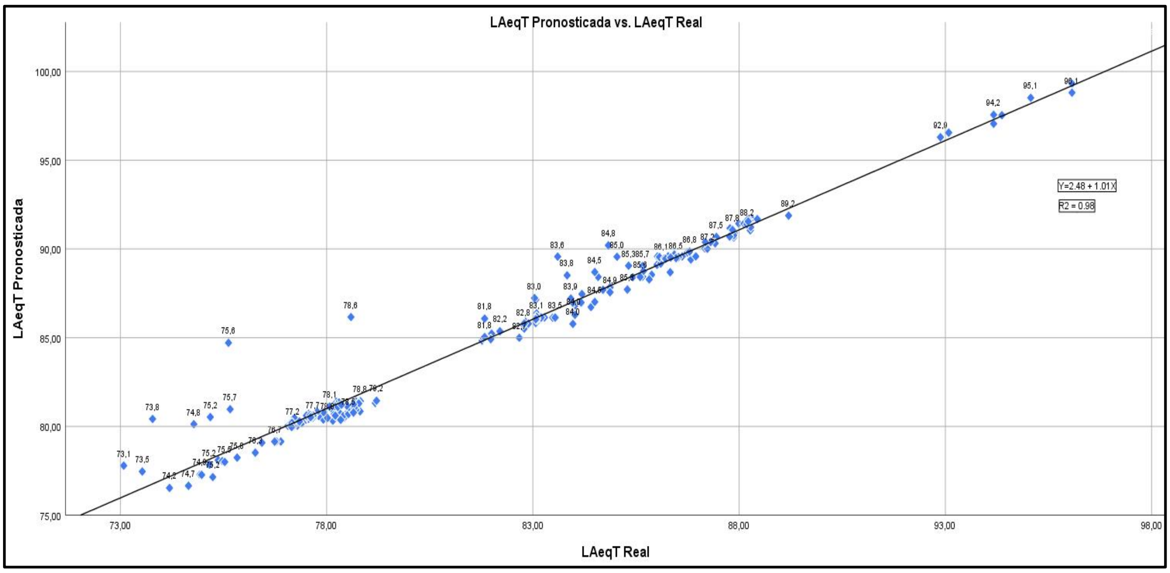

Figure 3. Predicted LAeqT vs. actual LAeqT of T1

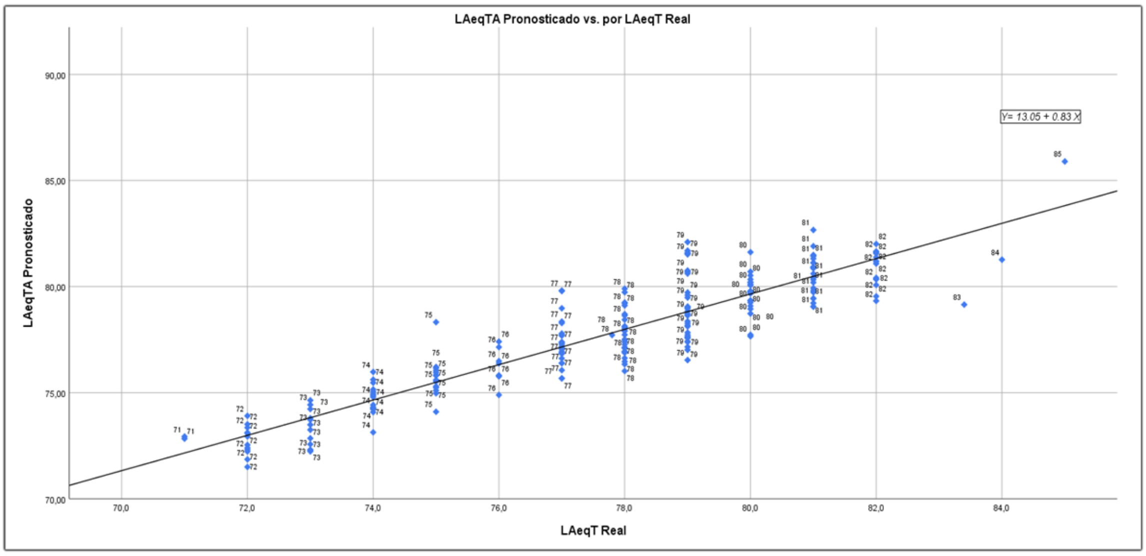

Figure 4. Predicted LAeqT vs. actual LAeqT of T2

Table 5. Values of statistics used in validation of RNA-T1 and RNA-T2

Statistic	RNA-T1	RNA-T2
R	0.9927	0.9989
R^2	0.9854	0.9978
RMSE	0.7313	0.1515

The performance of the neural network was obtained through the R, R^2 and RMSE statistics. The networks obtained show good performance in both the training and validation stages, giving values close to unity, which demonstrates a positive relationship between the data obtained by the network and the data provided. It should be noted that the validation values were more significant; obtaining R=0.9927 and R^2=0.9854, RMSE=0.7313 at T1 and R=0.9989 and R^2=0.9978, RMSE=0.1515 at T2. If these results are compared with those obtained by different authors such as Mansourkhaki *et al.* [22] who obtained an R=0.992 and R^2=0.983, RMSE=0.1515 at T2 and the case of Sequeira *et al.* [23] who obtained R=0.995, R^2=0.991, and RMSE=0.44; it can be said that the networks obtained for both T1 and T2 show high efficiency for noise prediction. It can also be affirmed that MLP ANNs for both T1 and T2 show high efficiency for the prediction of the equivalent continuous sound pressure level (LAeq); these results are analogous to those obtained by Genaro García where he compared the efficiency of an ANN with mathematical models (RSLS 90 and technical guidelines for noise impact assessment (Criterion)). Obtaining that the ANN has a better performance and by Chen *et al.* [24] comparing two types of neural networks, including the MLP network and the radial basis function (RBF) network for predicting traffic noise, where it was shown that the MLP network performed better than the RBF network in predicting noise level [25].

4. CONCLUSION

An ANN model was developed for the estimation of the equivalent continuous sound pressure level (LAeqT) using the MLP etwork type and algorithm, using the variables that contributed most to the estimation of the dependent variable during the training stage. For T1-Burga Mendoza they were: name of road, UTM coordinates, time, Lmax, Lmin, and motokar, for T2-Cruzado Ancajima and Soto Medina they were: name of road, UTM coordinates, Lmax, Lmin, Motokar, and Moto linear. A structure of 6-19-1 was obtained for T1-Burga Mendoza and 6-15-1 for T2-Cruzado Ancajima and Soto Medina. The validation results of the work show that the network created for T1 is capable of estimating the sound pressure level with R=0.9927 and R^2=0.9854 and for T2 with R=0.9989 and R^2=0.9978.

ACKNOWLEDGEMENTS

The authors would like to thank the Universidad Nacional de Jaén for the facilities to carry out the research. To the professionals Elser Burga Mendoza, Yanira Soto Medina and Cintia Cruzado Ancajima for providing us with their research data.

APPENDIX

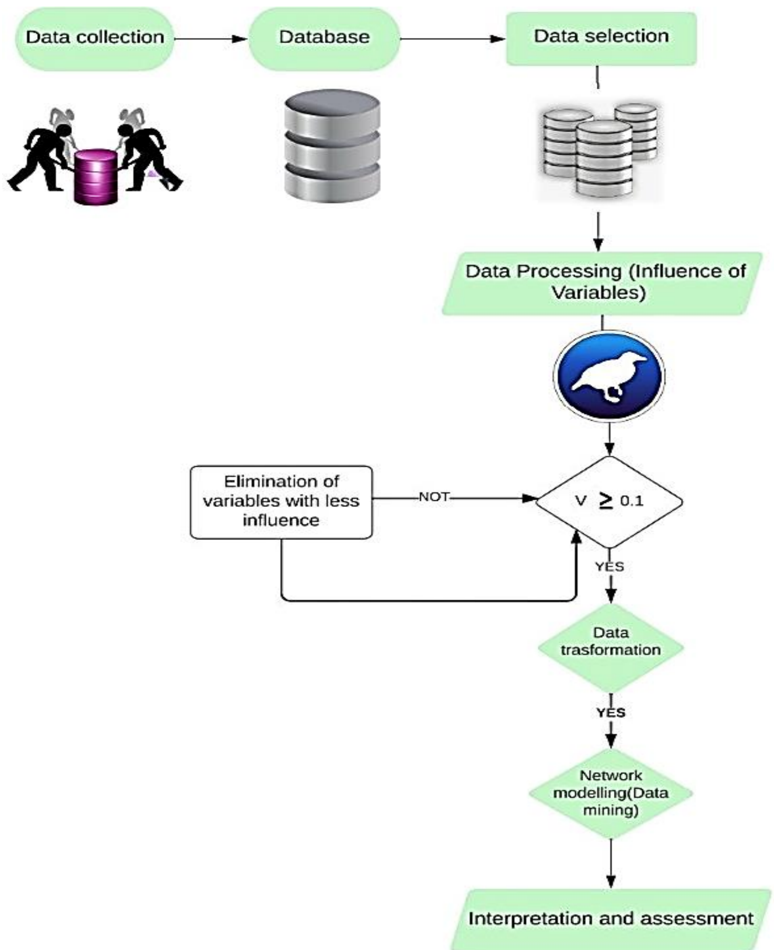

Figure 1. Method

REFERENCES

[1] S. V. Kuznetsov *et al.*, "Emotional artificial neural network (EANN)-based prediction model of maximum A-weighted noise pressure level," *Noise Mapping*, vol. 9, no. 1, pp. 1–9, Jan. 2022, doi: 10.1515/noise-2022-0001.

[2] P. E. Baffoe, A. A. Duker, and E. V. S-Kwarteng, "Assessment of health impacts of noise pollution in the Tarkwa Mining Community of Ghana using noise mapping techniques," *Global Health Journal*, vol. 6, no. 1, pp. 19–29, Mar. 2022, doi: 10.1016/j.glohj.2022.01.005.

[3] J. Díaz, J. A-López-Bueno, D. Culqui, C. Asensio, G. S-Martínez, and C. Linares, "Does exposure to noise pollution influence the incidence and severity of COVID-19?," *Environ Res*, vol. 195, p. 110766, Apr. 2021, doi: 10.1016/j.envres.2021.110766.

[4] V. Nourani, H. Gökçekuş, I. K. Umar, and H. Najafi, "An emotional artificial neural network for prediction of vehicular traffic noise," *Science of The Total Environment*, vol. 707, p. 136134, Mar. 2020, doi: 10.1016/j.scitotenv.2019.136134.

[5] L. B-Moncayo, J. L-Naranjo, M. Chávez, I. P-García, and C. Garzón, "A machine learning approach for traffic-noise annoyance assessment," *Applied Acoustics*, vol. 156, pp. 262–270, Dec. 2019, doi: 10.1016/j.apacoust.2019.07.010.

[6] J. Tomić, N. Bogojević, and D. Š. Pavlović, "Application of Soft Computing Techniques in Prediction of Road Traffic Noise Levels," in *2018 26th Telecommunications Forum (TELFOR)*, Nov. 2018, pp. 1–7, doi: 10.1109/TELFOR.2018.8611931.

[7] I. Deryabin, "Traffic lights as a factor in the regulation of traffic noise," *Transportation Research Procedia*, vol. 63, pp. 436–440, Jan. 2022, doi: 10.1016/j.trpro.2022.06.032.

[8] C. Guarnaccia, J. Quartieri, and C. Tepedino, "A hybrid predictive model for acoustic noise in urban areas based on time series analysis and artificial neural network," *AIP Conference Proceedings*, vol. 1836, no. 1, p. 020069, Jun. 2017, doi: 10.1063/1.4982009.

[9] F. Cirianni and G. Leonardi, "The application of a neural network on a study of noise pollution in urban transport: a case in Villa S. Giovanni," Jan. 2004, pp. 559–570.

[10] A. S. R. Subramanian, T. Gundersen, and T. A. Adams, "Modeling and Simulation of Energy Systems: A Review," *Processes*, vol. 6, no. 12, Dec. 2018, doi: 10.3390/pr6120238.

[11] L. J. R. Nunes, A. M. Rodrigues, L. M. E. F. Loureiro, L. C. R. Sá, and J. C. O. Matias, "Energy Recovery from Invasive Species: Creation of Value Chains to Promote Control and Eradication," *Recycling*, vol. 6, no. 1, Mar. 2021, doi: 10.3390/recycling6010021.

[12] Kumar, "Traffic noise prediction and optimization using response surface method (RSM)," *Arabian Journal of Geosciences*, pp. 1–16, Oct. 2021, doi: 10.1007/s12517-021-08547-5.

[13] X. Zhang, M. Zhao, and R. Dong, "Time-Series Prediction of Environmental Noise for Urban IoT Based on Long Short-Term Memory Recurrent Neural Network," *Applied Sciences*, vol. 10, no. 3 Jan. 2020, doi: 10.3390/app10031144.

[14] U. Masood, H. Farooq, and A. Imran, "A Machine Learning Based 3D Propagation Model for Intelligent Future Cellular Networks," in *2019 IEEE Global Communications Conference (GLOBECOM)*, Dec. 2019, pp. 1–6, doi: 10.1109/GLOBECOM38437.2019.9014187.

[15] A. S. Tenney, M. N. Glauser, C. J. Ruscher, and Z. P. Berger, "Application of Artificial Neural Networks to Stochastic Estimation and Jet Noise Modeling," *AIAA Journal*, vol. 58, no. 2, pp. 647–658, 2020, doi: 10.2514/1.J058638.

[16] E. Turajlic, A. Begović, and N. Škaljo, "Application of Artificial Neural Network for Image Noise Level Estimation in the SVD domain," *Electronics*, vol. 8, no. 2, Feb. 2019, doi: 10.3390/electronics8020163.

[17] V. K. Mishra, K. Anand, and A. Bhardwaj, "Clustering assisted artificial neural network for handling noisy big data: An application for estimation of parameter in combined mode conduction and radiation heat transfer," *Heat Transfer*, vol. 51, no. 6, pp. 5386–5416, 2022, doi: 10.1002/htj.22552.

[18] R. Vinayakumar, M. Alazab, K. P. Soman, P. Poornachandran, A. Al-Nemrat, and S. Venkatraman, "Deep Learning Approach for Intelligent Intrusion Detection System," *IEEE Access*, vol. 7, pp. 41525–41550, 2019, doi: 10.1109/ACCESS.2019.2895334.

[19] R. M-Medina, R. P-Gálvez, A. Guadix, and E. M. Guadix, "Artificial neuronal network modeling of the enzymatic hydrolysis of horse mackerel protein using protease mixtures," *Biochemical Engineering Journal*, vol. 105, pp. 364–370, Jan. 2016, doi: 10.1016/j.bej.2015.10.009.

[20] Z. Zhang, X. Deng, and L. Zheng, "A review on varying-parameter convergence differential neural network," *Neurocomputing*, vol. 490, pp. 54–65, Jun. 2022, doi: 10.1016/j.neucom.2022.03.026.

[21] J. M-Oreja and J. Gozalvez, "Predicción de la velocidad del tráfico basada en redes neuronales convolucionales," *Avances en arquitectura y tecnología de computadores. Actas de las Jornadas SARTECO 2017*, pp. 677–684, Sep. 2017, doi: 10.5281/zenodo.996051.

[22] A. Mansourkhaki, M. Berangi, M. Haghiri, and M. Haghani, "A neural network noise prediction model for Tehran urban roads," *Journal of Environmental Engineering and Landscape Management*, vol. 26, no. 2, Jun. 2018, doi: 10.3846/16486897.2017.1356327.

[23] M. E. Sequeira, V. H. Cortínez, and A. P. Azzurro, "Caracterización Acústica de Ambientes Urbanos Mediante el Uso de Redes Neuronales," *Mecánica Computacional*, vol. 37, no. 5, Nov. 2019.

[24] L. Chen, T. Liu, B. Tang, H. Xiang, and Q. Sheng, "Modelling traffic noise in a wide gradient interval using artificial neural networks," *Environmental Technology*, vol. 42, no. 23, pp. 3561–3571, Oct. 2021, doi: 10.1080/09593330.2020.1734098.

[25] A. Mansourkhaki, M. Berangi, and M. Haghiri, "Comparative Application of Radial Basis Function and Multilayer Perceptron Neural Networks to Predict Traffic Noise Pollution in Tehran Roads," *J. Ecol. Eng.*, vol. 19, no. 1, pp. 113–121, Jan. 2018, doi: 10.12911/22998993/79411.

BIOGRAPHIES OF AUTHORS

Wendy Díaz she received his degree in Forestry and Environmental Engineering from the National University of Jaen (UNJ) in Peru in 2021 and is now a Master's student in Environmental Management at the National University of Trujillo (UNT) in Peru. She can be contacted at email: wendy.diaz@est.unj.edu.pe.

Anali Tarrillo received the degree of Forestry and Environmental Engineer from the Universidad Nacional de Jaén in 2021. Currently, she works as an engineer in the private company ZURSAN INGENIERIA Y CONSTRUCCIÓN EIRL in the area of Safety, Occupational Health and Environment. Her current research interests include artificial intelligence and deep learning. She can be contacted at email: anali.tarrillo@est.unj.edu.pe.

Candy Ocaña formed by the Faculty of Agricultural Sciences of the National University of Cajamarca, profession forest engineer, master's degree in public management and completed studies of the master's degree in Forestry and Natural Resource Management, currently a doctoral student in Economics of Natural Resources. With knowledge and application of research methodologies for the elaboration and execution of scientific projects, administration of research projects and public investment. Work and professional experience in undergraduate university teaching at the University of Cajamarca, currently teaching in the Assistant category at the National University of Jaen (UNJ). She can be contacted at email: candy.ocana@unj.edu.pe.

Lenin Quiñones Bachelor's degree in Mathematics, Master's degree in Computer Science and Ph.D in Industrial Engineering. Work and professional experience in undergraduate university teaching in Chile and Peru and postgraduate in the Master of Science with mention in Environmental Management of the Faculty of Ecology of the Universidad Nacional de San Martín. Currently teaching undergraduate and graduate courses at the National University of Jaen. The lines of research that I develop: Construction and use of mathematical models for Engineering, application of Artificial Intelligence, computation as a tool for solving research problems and teaching of Mathematics. Organizer and speaker at national and international conferences, thesis advisor, reviewer of scientific journals, author of scientific articles and books. He can be contacted at email: lenin.quinones@unj.edu.pe.