Comparison of regression model and artificial neural network model in noise prediction in a mixed area of Dhaka City

Chowdhury V.[1,*], Zarif S.[1], Tofa T.[1,2] and Laskar M.[1]

[1] Department of Civil and Environmental Engineering (CEE), Islamic University of Technology (IUT), Board Bazar, Gazipur, Bangladesh.

[2] Department of Civil Engineering, Military Institute of Technology (MIST), Mirpur Cantonment, Dhaka-1216, Bangladesh.

*corresponding author e-mail: vubanchowdhury@iut-dhaka.edu

Abstract

The equivalent noise levels regularly exceed acceptable limits within Dhaka city, the capital of Bangladesh, especially in the mixed urban areas (where trips are generated to serve commercial, residential, and industrial demands). The study aims to assess the noise level in mixed urban areas, build noise prediction models and allow scopes for ensuring sustainable environmental management. Two traffic noise prediction models were assessed: a regression model and an artificial neural network (ANN) model to predict the equivalent noise level (L_{eq}). Traffic and noise level data were collected from two mixed urban areas, statistical analyses were performed to describe the existing trends and to evaluate both model's responses in predicting equivalent noise level (L_{eq}). The ANN model (coefficient of determination: 0.82) showed better performance than the regression model (coefficient of determination: 0.70). The predicted equivalent noise levels from the ANN model were compared to acceptable limits to display the extent of noise pollution using GIS. The traffic noise models can assist in environmental impact assessment to protect the communities susceptible to the adversities of noise pollution.

Keywords: Noise pollution, Equivalent noise level, Prediction model, Regression, Artificial Neural Network.

1. Introduction

Any form of unwanted sound is known as noise. The inhabitants of Dhaka city are exposed to severe levels of noise pollution. The resulting health hazards have physiological and psychological consequences. Chowdhury et al. (2010) identified motorized traffic as the major source of noise pollution among the various sources such as construction activities, public gatherings, concerts, etc. They also stated that noise pollution in Dhaka was not a significant issue in the 70s and early 80s but with the influx of motorized vehicles and urbanization, noise pollution has become intolerable. Noise prediction models can assist in formulating more effective environmental policies, identifying areas with noise pollution problems, and assessing environmental impacts of noise from traffic in future urban projects. Machine learning models can be a feasible way of predicting noise levels and assessing their environmental impacts on urban areas. The benchmarking of an artificial neural network model and conventional regression model to predict noise showed that the artificial neural network performed better in Villa S. Giovanni, Italy (Cirianni & Leonardi, 2015). Assessments of urban noise levels from traffic using various soft computing approaches showed the superiority of the Neural Network approach but it was concluded that the greater predictability comes with higher complexity and resource consumption. (Tomić et al., 2016). A three-variable neural network model for predicting highway traffic noise from traffic parameters relevant to India proved to be accurate (Kumar et al., 2014). Although such models integrate a wide variety of parameters, they need experimental data to be trained. As a result, every model is influenced by the conditions of the region of data collection and is unique to a certain region or country (Tomić et al. 2016). A few studies have analyzed the ability of statistical approaches to predict equivalent noise levels in Bangladesh (Alam et al. 2006, Tanvir & Rahman, 2011). However, the predictability of deep learning approaches has not been analyzed. Such noise prediction models can be highly accurate in Dhaka city if trained within its unique traffic conditions e.g.: The tendency of drivers to change lanes frequently, negligence to traffic regulations. In this study, a regression model and an artificial neural network model are proposed for predicting equivalent noise levels (L_{eq}) in Dhaka. The abundance of bicycles & rickshaws (a three-wheeled non-motorized passenger cart) in Dhaka prompted the consideration of the proportion of non-motorized vehicles as a variable in this study. However, the other variables considered in this study (barrier height, road width, traffic volume, traffic speed, traffic density) were previously used in traffic noise modelling (Hamad et al., 2017, Cirianni & Leonardi, 2015). It was observed that the mixed area around Ramna at the central part of Dhaka city experiences the highest levels of Noise pollution (Tanvir & Rahman, 2011). The roads in such areas simultaneously serve the traffic demands of industries, residences and commercial establishments, leading to greater levels of noise. Hence, collecting data from a mixed area would train the models for a wide range of noise levels and traffic conditions.

The objectives of this study were to compare the accuracies of an ANN model and a regression model in predicting L_{eq}, test the superior model in a new area and map the level of noise pollution in a mixed area using the predictions from the superior model.

2. Methodology

2.1. Study area

Environmental Conservation Rules 1997 (ECR'97) labels a "Mixed Area" as an area that is used for residential, commercial & industrial purposes. Two mixed urban areas situated at the center of Dhaka city were selected for collecting data, namely- Ramna & Dhanmondi. The data from 81 sampling stations in Ramna were used for model training, primary testing and statistical analysis. The data from 7 sampling stations in Dhanmondi were used as a secondary testing dataset.

2.2. Data collection and tabulation

Sampling stations were set up at different arterials and collectors with an uninterrupted flow within 9 AM to 5PM from November 2020 to January 2021. Geographic coordinates (WGS 1984), time mean speed, traffic volume, traffic density, barrier height, road width and A-weighted equivalent noise level (L_{eq}) data were collected at each station. A digital camera recorded the two-way traffic volume, which was later classified into non-motorized vehicles and three categories of motorized vehicles- light vehicles, medium vehicles and heavy vehicles according to s. 1.2 of *The motor vehicles ordinance, 1983* (Mllr). To measure the time-mean speed for each class of motorized vehicles, a Bushnell speed gun was used. A measuring tape was used to determine road width. The average A-weighted noise level (L_{eq}) was recorded using the "B&K Precision 735" sound level meter at each sampling station. Traffic density was computed using the following formula,

$$\text{Traffic Density} = \frac{\text{Hourly Traffic volume}}{\text{Time} - \text{mean speed}}$$

Determination of Pearson correlation coefficients (r) of the independent variables assisted in understanding their contribution to equivalent noise level (L_{eq}).

2.3. Artificial Neural Network Model

An artificial Neural Network combines the architecture of the human brain with statistical learning models to predict one or more dependent variables from a set of independent variables. "It is a multilayer perceptron (MLP) that involves a simple interconnected system of nodes or neurons" (Ahmed & Pradhan, 2019, p. 7). Multi-layer feed-forward (MLF) network architecture and the Bayesian-Regularization (BR) training algorithm was chosen to train the model using the neural network toolbox of MATLAB. The following input variables were selected for the ANN model by the forward selection method:

NMVR = proportion of non-motorized vehicles, *LV* = Volume of light vehicles, *MV* = Volume of medium vehicles, *TV* = Total volume, *LS* = Time mean speed of light vehicles, *AS* = Time mean speed of all vehicles, *TD* = Total density

This variable selection method prioritizes the predictive accuracy of the model. It is initiated with a model containing one independent variable which produces the highest predictive accuracy. Then the combinations of all the remaining variables with the first variable are tested to acquire the most accurate two-variable model. The process is repeated until the addition of new variables doesn't improve the accuracy of the model (Anderson & Bro, 2010). The proposed network (Figure 1) consists of an input layer, 2 hidden layers (12 neurons in each hidden layer). The TANSIG (tangent sigmoid) activation function was used in each of the hidden layers and a PURELIN (linear) activation function was used in the output layer. Every Neuron had a weight (w_{ij}) and a bias (b) associated with them. After each iteration, the weight and the bias were updated to minimize the mean square error (MSE). The model was trained using 85% (69 samples) of the data from Ramna and primarily tested on the remaining 15% (12 samples).

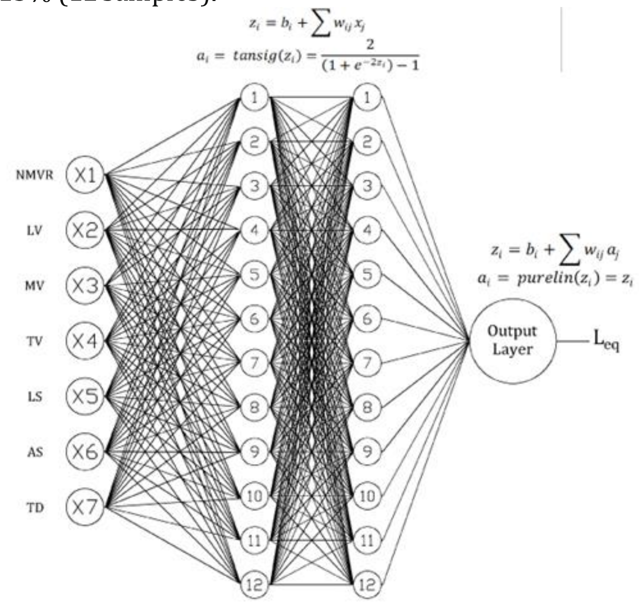

Figure 1. Artificial Neural Network model Architecture

2.4. Ridge Regression Model

Ridge regression is widely used for parameter assessment and approximation to address the collinearity issue regularly emerging in numerous linear regression models (McDonald, 2019). Ridge regression was chosen as the ideal regression model to fit the collected data because of the existing multicollinearity among variables (assessed by correlation analysis). Similar to the ANN model, the relevant independent variables were selected using the forward selection method. 90% (73 samples) of the data from Ramna were randomly selected as training data (10 cross-validation subsets) and 10% (8 samples) of the data as primary testing data. Cross-validation segregates a certain sample of data into corresponding subsets, performing an analysis on the training subset and validating the performed analysis on the validation subset (El-Habil & Almghari, 2011). The function "GridSearchCV" was used to select the degree of bias (α=1050), which is a hyperparameter used to reduce standard errors. The equivalent sound level (L_{eq}) was predicted as a function of the following variables:

L_{eq}= 0.106 NMVR - 0.0445 LV - 0.06 MV + 0.061 TV + 0.009 TD + 0.177 LS

2.4. Model comparison, statistical analyses and GIS mapping

The accuracies of the two models were analyzed and compared for identifying the superior model which was tested on the secondary testing dataset from Dhanmondi. Statistical analyses were performed to explain the variation of equivalent noise levels for different categorical and quantitative variables. The predicted noise levels were compared to the standard limits set according to Rule-12, Schedule-4 of ECR '97 to map the level of noise pollution in different parts of the study area using ArcGIS Pro 1.2.

3. Results & Discussion

To analyze and compare the predictive accuracies of the ANN model and the ridge regression model (RR), their performances were evaluated based on their root mean square error (RMSE), coefficient of determination (R^2) and adjusted R^2. (Table 1).

Table 1. Comparison of the predictive accuracies of ANN and RR model in noise prediction in Ramna

Dataset	Training set		Testing set		All data	
Model	**ANN**	**RR**	**ANN**	**RR**	**ANN**	**RR**
RMSE	1.42	1.80	1.23	1.89	1.40	1.81
R^2	0.80	0.69	0.90	0.71	0.82	0.70
Adjusted R^2					0.80	0.67

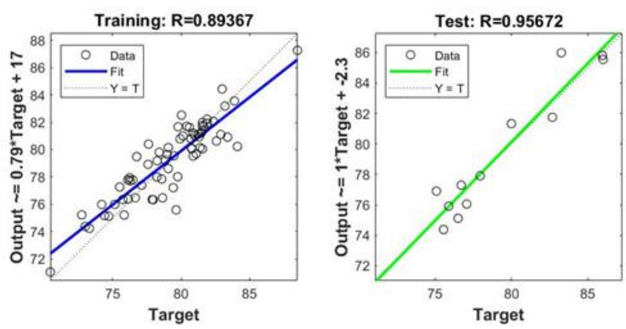

Figure 2. Comparison between observed L_{eq} and predicted L_{eq} from the ANN model for training and testing data

The ANN model outperforms the RR model in all the evaluation criteria with an overall R^2 of 0.82. Figure 2 shows a plot of the observed L_{eq} and predicted outputs from the ANN model along with their correlation coefficient (R). A close observation of the comparison of the two models shows that the superiority of the ANN model is most prominent in its ability to fit 90% (R^2= 0.90) of the primary testing data compared to the 71% (R^2 = 0.71) fit by the RR model. The ability to handle unseen data patterns is a distinct feature of the ANN model defined as its generalization capability (Urolagin et al., 2012). The trained ANN model's generalization ability was further validated by its predictions in a different area. Table 2 shows a comparison of the observed L_{eq} and the L_{eq} predicted by the ANN model in Dhanmondi. The model was able to generalize and explain 83% (R^2=0.83) of the variability in the secondary dataset with an RMSE of 1.17.

Table 2. Results of testing the ANN model on the secondary dataset from Dhanmondi

Observed L_{eq} (dBA)	Predicted L_{eq} (dBA)	RMSE	R^2
82.68	83.04	**1.17**	**0.83**
82.61	81.54		
78.45	77.14		
78.19	78.55		
75.79	74.64		
75.15	75.60		
78.40	80.63		

The highest recorded L_{eq} (88.5 dBA) falls within the interval of 2 PM – 3 PM where the equivalent noise levels deviated the most (σ=5.76) (Figure 3). From figure 4, it is evident that the weekdays experience the highest levels of noise and the observed equivalent noise levels show a higher deviation (σ=4.07) from their mean on weekends.

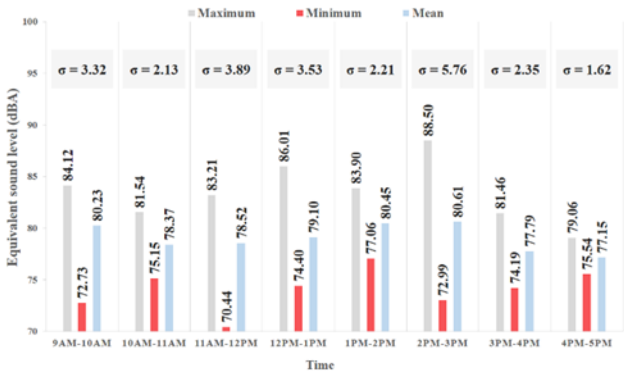

Figure 3. Hourly variation of L_{eq} in Ramna

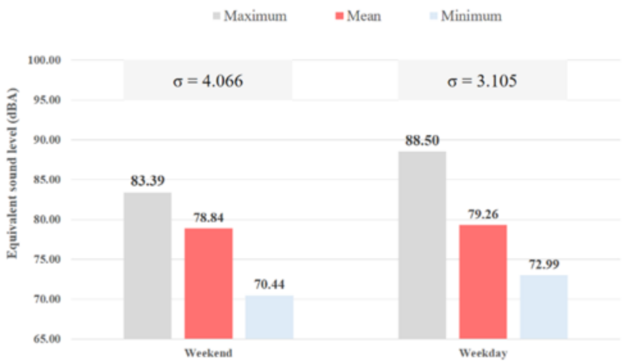

Figure 4. Variation of L_{eq} on weekdays and weekends in Ramna

NMVR was the only independent variable that showed a high negative correlation (r = -0.58) with L_{eq} as non-motorized vehicles produce low levels of noise. Moreover, field observations suggested that a high proportion of non-motorized vehicles on busy roads slows down traffic and allows a lower number of vehicles to pass within a certain period which results in a lower L_{eq}. The negative correlations of NMVR with AS (r= -0.60) and TV (r= -0.66) support the field observations. The comparison of predicted equivalent noise levels from the ANN model to the standard limits from ECR '97 for different zones- 45 dBA (silent zones), 50 dBA (residential area), 70 dBA (commercial area) and 75 dBA (industrial area) reveal the extent of noise pollution in Ramna (Figure 5). Noise pollution is the highest in the silent zones that are on the east of Ramna Park, west of Dhaka Medical College and south of Officer's Club and the standard limits were exceeded the least on the east of Dhaka New Market which is a commercial area.

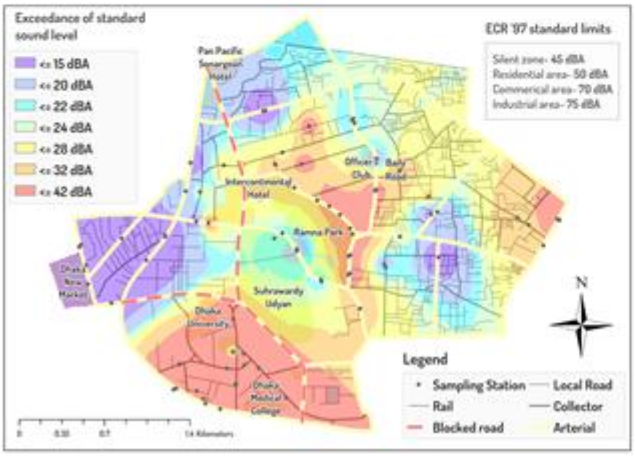

Figure 5. The extent of noise pollution in Ramna

The dataset used in this study is smaller compared to some of the studies that used similar approaches to predict equivalent noise levels (Cirianni & Leonardi, 2006; Hamad et al., 2017; Ahmed & Pradhan, 2019). Despite the restricted database, the training data from Ramna allowed the ANN model to be trained for a wide range of conditions as the area is used for multi-dimensional purposes. This helped the model predict equivalent noise levels without any significant loss of accuracy in both areas.

4. Conclusion

The ANN model was found to be superior compared to the RR model in predicting the equivalent noise level (L_{eq}) in Dhaka city. The secondary testing suggests that its accuracy can be replicated in a different area within the city. The mapping of noise pollution reveals an urgency to address the noise pollution problem in the silent zones of Ramna as the standard limit of L_{eq} is exceeded the most in these zones. ANN models would allow planners to foresee the environmental impacts of noise and assist in fast decision-making in the presence of time and budget constraints. The inclusion of the proportion of non-motorized vehicles as a variable in this study optimized the models for Dhaka city. However, a more elaborate data collection campaign with the integration of interrupted traffic flow would allow scope for a more inclusive noise prediction model to be built.

References

Ahmed, A. A., & Pradhan, B. (2019). Vehicular traffic noise prediction and propagation modelling using neural networks and geospatial information system. *Environmental monitoring and assessment*, *191*(3), 1-17.

Alam, J. B., Rahman, M. M., Dikshit, A. K., & Khan, S. K. (2006). Study on traffic noise level of Sylhet by multiple regression analysis associated with health hazards. *Iran. J. Environ. Health. Sci. Eng.*, 3(2), 71-78.

Andersen, C. M., & Bro, R. (2010). Variable selection in regression-a tutorial. *Journal of Chemometrics*, *24*(11–12), 728–737. https://doi.org/10.1002/cem.1360

Cai, M., Zou, J., Xie, J., & Ma, X. (2015). Road traffic noise mapping in Guangzhou using GIS and GPS. *Applied Acoustics*, *87*, 94–102. https://doi.org/10.1016/j.apacoust.2014.06.005

Chowdhury, S. C., Razzaque, M. M., Helali, M. M., & Bodén, H. (2010). Assessment of noise pollution in Dhaka city. In *17th International Congress on Sound and Vibration, Cairo, Egypt, 2010-07-18-2010-07-22.*

Cirianni, F., & Leonardi, G. (2015). Artificial neural network for traffic noise modelling. *ARPN Journal of Engineering and Applied Sciences*, *10*(22), 10413–10419.

El-Habil, A. M., & Almghari, K. I. A. (2011). Remedy of multicollinearity using Ridge regression. *Journal of Al Azhar University-Gaza (Natural Sciences), 13, 119-134.*

Environment Conservation Rules 1997 (Mef) s. 4.12 (Bangl)

Hamad, K., Ali Khalil, M., & Shanableh, A. (2017). Modeling roadway traffic noise in a hot climate using artificial neural networks. *Transportation Research Part D: Transport and Environment*, *53*, 161–177. https://doi.org/10.1016/j.trd.2017.04.014

Kumar, P., Nigam, S. P., & Kumar, N. (2014). Vehicular traffic noise modeling using artificial neural network approach. *Transportation Research Part C: Emerging Technologies*, *40*, 111–122. https://doi.org/10.1016/j.trc.2014.01.006

Kyasville. (2007). *NCSS User Guide 1*. 335-339

McDonald. (2019). *Ridge Regression*. 1-2

Razzaque, M. M., Chowdhury, S. C., Helali, M. M., & Bodén, H. (2010). On the impacts of noise pollution in Dhaka. *17th International Congress on Sound and Vibration 2010, ICSV 2010*, *4*(July), 3068–3074.

Tanvir, S., & Rahman, M. M. (2011). Development of interrupted flow traffic noise prediction model for Dhaka City. *Bangladesh University of Engineering and Technology*, *4*, 131-138.

Tomić, J., Bogojević, N., Pljakić, M., & Šumarac-Pavlović, D. (2016). Assessment of traffic noise levels in urban areas using different soft computing techniques. *The Journal of the Acoustical Society of America*, *140*(4), EL340–EL345. https://doi.org/10.1121/1.4964786

The Motor Vehicles Ordinance 1983 (Mllr) s. 1.2 (Bangl.)

Urolagin, S., Prema, K. V, & Reddy, N. V. S. (2012). *Generalization Capability of Artificial Neural Network*. 171–178.