Comparison of Regression Model and Artificial Neural Network Model in Noise Prediction in a Mixed Area of Dhaka City

VUBAN CHOWDHURY

SAGUPTH ALAM ZARIF

MUBASHIR SHABAB LASKAR

Department of Civil & Environmental Engineering

Islamic University of Technology

2021

Comparison of Regression Model and Artificial Neural Network Model in Noise Prediction in a Mixed Area of Dhaka City

VUBAN CHOWDHURY (160051048)

SAGUPTH ALAM ZARIF (160051024)

MUBASHIR SHABAB LASKAR (160051079)

A THESIS SUBMITTED FOR THE DEGREE OF BACHELOR OF SCIENCE IN CIVIL ENGINEERING (TRANSPORTATION)

Department of Civil & Environmental Engineering

Islamic University of Technology

2021

APPROVAL

This is to certify that the dissertation entitled "**Comparison of Regression Model and Artificial Neural Network Model in Noise Prediction in a Mixed Area of Dhaka City**", by Vuban Chowdhury, Sagupth Alam Zarif and Mubashir Shabab Laskar has been approved fulfilling the requirements for the Bachelor of Science Degree in Civil Engineering.

Supervisor:
Tajkia Syeed Tofa
Assistant Professor
Department of Civil and Environmental Engineering
Islamic University of Technology (IUT)
Board Bazar, Gazipur, Bangladesh.

DECLARATION

We declare that the undergraduate research work reported in this thesis has been performed by us under the supervision of Assistant Professor Tajkia Syeed Tofa. We have taken appropriate precautions to ensure that the work is original and has not been plagiarized. We can also make sure that the work has not been submitted for any other purpose (except for publication).

Vuban Chowdhury
ID: 160051048
August, 2021

Sagupth Alam Zarif

Sagupth Alam Zarif
ID: 160051024
August, 2021

Mubashir Shabab Laskar

Mubashir Shabab Laskar
ID: 160051042
August, 2021

ACKNOWLEDGEMENT

"In the name of Allah, Most Gracious, Most Merciful"

All praises to Allah (SWT) for giving us the opportunity to complete this research work. We would especially like to thank our parents who have provided us with the strength and dedication needed to complete this thesis.

It is a proud privilege to acknowledge our sincere and deepest sense of gratitude to our respected supervisor Tajkia Syeed Tofa, Assistant Professor of Civil and Environmental Engineering (IUT) for his untiring efforts, inspiration, invaluable guidance and constructive criticisms throughout the period.

We would also like to express gratitude to all of our departmental faculty members and staff for their help and support throughout our undergraduate period.

The authors are thankful to the department of Civil and Environmental Engineering for providing all the necessary equipment and accessories that were needed to collect primary data that were used in the research work.

DEDICATION

We wholeheartedly dedicate this thesis to our parents, who for many years, sacrificed their time, commitment, and livelihood to ensure that we could follow our passion through education. They have always been the driving force in our lives and careers and we would not have realized our potential without their support. We will be forever indebted to them.

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 (Leq). 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 (Leq). 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.

TABLE OF CONTENTS

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Introduction

1.1 Introduction to noise

In our rapidly expanding environment, noise is one of the developing problems. Noise is becoming a serious concern for industrial corporations, residential areas & commercial areas. When an unwanted sound occurs in an area is called noise. This is basically a personal opinion. The object of this part is to discuss the concept of noise, problems of noise & health hazards of noise.

The major sources of noise are:

1. Industrial noise
2. Traffic noise
3. Community noise

Above all of those mentioned sources, almost 70% noise is created through vehicle noise. Vehicle noise is the combined noise of all the cumulative value of vehicles & honking of the cumulative vehicles. The major concern of the study is to create a model for a mixed area to know accuracy of sound level at a particular time.

Harmful effects of noise on human beings

- Reduces work efficiency
- Induces loss of hearing ability
- May cause cardiac arrest

- Increases the cholesterol of blood
- Causes uneasiness
- Cause nausea & headache

1.2 Fundamentals of noise

Any elastic medium which can transmit a pressure wave results in sound. To be able to hear sound there must air or elastic medium at ear. The magnitude of the pressure variations is proportional to the loudness of the sound. The number of pressure cycles per second determines whether we will hear high pitch sound or low pitch sound.

1.2.1Physical Properties of sound

A device which can detect small pressure variations in the sound field, it will produce an electric signal proportional to the sound pressure. The unit of sound level is dB . The range of audible sound pressure range is very wide from 0 dB to 134 dB.

dB is the logarithmic ratio which defines the sound pressure level as follows:

$L_{eq} = L_{50}+(L_{90}-L_{10})^2/60$

This logarithmic scale has some advantage over linear scale. The advantages are:

1. The linear would lead to some enormous & unwieldly data.
2. The ear responds not linearly but stimulus to logarithmic

1.2.2 Sound sources

- Point Source
- Line Source
- Plane source

Point Source: A sound source can be considered as a point source, if its dimensions are small in relation to the distance to the receiver and it radiates an equal amount of energy in all directions. Typical point sources are industrial plants, aircraft and individual road vehicles. The sound pressure level decreases 6 dB whenever the distance to a point source is doubled.

Line Source: A line source may be continuous radiation, such as from a pipe carrying a turbulent fluid, or may be composed of a large number of point sources so closely spaced that their emission may be considered as emanating from a notional line connecting them. The sound pressure level decreases 3 dB, whenever the distance to a line source is doubled.

Plane Source: A plane source can be described as follows. If a piston source is constrained by hard walls to radiate all its power into an elemental tube to produce a plane wave, the tube will contain a quantity of energy numerically equal to the power output of the source. In the ideal situation there will be no attenuation along the tube. Plane sources are very rare and only found in duct systems.

1.3 NOISE MEASUREMENT TECHNIQUES & INSTRUMENTS

Noise measuring devices typically use a sensor to receive the noise signals emanating from a source. The sensor, however, not only detects the noise from the source, but also any ambient background noise. Thus, measuring the value of the detected noise is inaccurate, as it includes the ambient background noise. Many different types of instruments are available to measure sound levels and the most widely used are sound level meters

1.3.1 Noise measurement instruments

- Sound level Meter
- Velocity speed Gun
- Measurement tape
- Digital camera & tripod
- GPS

1.3.2 Steps of Measurement System

Sound levels were measured in A-weighted decibels at an interval of 15 seconds for a period of 5 minutes, then equivalent noise levels were measured using the formula: $L_{eq}=L_{50}+(L_{90}-L_{10})^2/60$

Velocity speed gun was used to collect 5 instantaneous speed data for each class of vehicles (light, medium & heavy) within 5 minutes. Then they were averaged to find the time mean speed for each class of vehicles.

A digital camera was used to record traffic volume for non-motorized, light, medium & heavy vehicles for 5 minutes at each location. Traffic density was measured using the formula: Density= Traffic Volume/Time mean speed

A measuring tape was used to measure the effective road width from the edge to the center of the road.

The typical floor height of different types of buildings were observed from BNBC and the total numbers of floor of the buildings surrounding our locations of interest were counted.

1.3.3 Noise Measurement System

- Sound level meter should be at least at a distance of 0.5 m from the body of the observer.
- Reflections from the body of the observer can cause an error of up to 6 dB at frequencies around 400 Hz.
- Sound level meter should be at a height of 1.2 –1.5 m from the floor level.
- Preferred position from near buildings and windows is 1 –2 m away.
- Outdoor measurements to be made at least 3.5 m away from other reflecting structures.
- Within the room measurement should be made in the Free Field zone.

.

CHAPTER 2

Literature Review

Literature Review

2.1 Overview

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 the environmental impacts of noise from traffic in future urban projects. Many studies have established distinct traffic noise modelling and prediction of noise levels in various metropolitan locations based on field observations of various road noise descriptors and traffic noise characteristics.

2.2 Traffic Conditions

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. Communication difficulties, 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).

2.3 Roads & Highways Conditions

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. However, collecting noise measurements on high-speed highways can be expensive, time-consuming and dangerous. In other words, when designing new highways, traffic noise models should provide a comfortable living environment in industrial, commercial and residential areas.

2.4 Health deterioration Factor

Noise is a stressor that has undoubtedly a negative impact on human health, particularly after prolonged exposure. Psychological and Physiological discomfort, as well as a disruption of organism's homeostasis and increase in allostatic load, all contribute to these negative health impacts. Hearing loss, high blood pressure, mental disorder, insomnia, productivity loss, and an overall decline in quality of life may all be caused by the increase of noise level (Singh & Davar, 2004). Moreover, Continual noise sets off the body's acute stress response. The risks grow as the noise intensity and exposure time increase. The findings of the study revealed that high noise intensity has an impact on residents of megacities around the world. Additionally, it is estimated that over 200 million people around the world are affected by the impacts of noise pollution (Mirzaei et al., 2012).

2.5 Environmental assessment

Environmental health inequalities can occur not just as a result of various exposure levels, but also as a consequence of varying levels of vulnerabilities to the detrimental health impacts of noise. Increased susceptibility to noise related health impacts may be caused by chronic conditions with less healthy lifestyle. Despite the fact that noise pollution is a chronic and silent killer, virtually little effort has been taken to mitigate it. It has become a threat to one's quality of life, like other forms of pollution. It is to blame for the increased rate of deafness around the world (Singh & Davar, 2004). So, the authority must assist in efficient urban environmental management through more effective environmental policies. It may compare predicted noise to the existing noise on different roads to tackle unnecessary causes of noise pollution such as excessive honking or traffic law violations. Zoning of proposed urban areas is essential for predicting the possible levels of noise that are created to save its future inhabitants from the hazards of noise pollution. Studies reveal that some groups are linked to higher levels of environmental noise or it can be said that they have higher prevalences of noise exposure when looking at exposure variations between those groups. People's socioeconomic circumstances influence where and how they can afford to live as well as those who are socially disadvantaged end up living and working in the lowest conditions. As a result, some socioeconomic groups may dwell in places that are more polluted than others. Furthermore, generalizations are difficult to make due to large methodological variations between studies. To measure noise exposure, different approaches can be applied, different populations can be studied, and different social indicators can be used. This theory is supported by the fact that several studies show that more advantaged people are not as likely than less advantaged people to suffer from serious noise related effects, even if the advantaged people reside in noisier locations.

2.6 Noise Mapping

Noise mapping is simply a visual portrayal of noise levels at places that forms the shape of a contour map. The map is usually installed on a layout of the area with contours, colored to show noise levels with the availability of low or high frequencies. Information with details can be obtained using noise mapping. It is feasible to improve the quality and efficiency of noise effect research by using GIS to map noise effects. Noise map can be utilized to facilitate the formulation of noise-control policies as well as enforcing noise control; outline a cost-benefit analysis to help areas seeking to mitigate noise; determine the primary sources of noise; provide policymakers with a clear illustration of noise exposure; use theory to investigate the impact of environmental improvement strategies (Tsai et al., 2009). It strengthens the execution of regional or national strategies to reduce new noise resources and secure new noise-sensitive places. During the enforcement procedure, it takes note of noise reduction techniques and their efficiency, changes noise trends in the environment and establishes a research area for the study of noise impacts on the human body. Usually geographic information systems (GIS) is used to analyze the collected data. To explore the current noise of different regions, noise maps need to be generated. The noise levels of various land use zones, such as commercial, residential, cultural as well as educational districts are to be analyzed and compared to applicable noise regulations.

2.7 Studies in Developed Countries

Noise pollution has become a serious problem in metropolitan areas which is affecting people's well-being, efficiency, and health along with animal behavior and habitat. The European Commission passed a directive demanding major cities to collect data to develop such local action projects on noise exposure charts, identifying it more as a critical issue (Maisonneuve et al., 2009). Several international broadcasts have emphasized the necessity of public participation in achieving long-term development. A research (Maisonneuve et al., 2009) has been published where citizens of France, Belgium and Netherlands contributed and suggested a new approach to assess the noise level that involves citizens from different cities and is based on citizen science and participatory sensing concepts. Using GPS-enabled smart phones integrated noise sensors, it enables the citizens to assess their own exposure to noise in their daily lives. The geo-localized noise measurements and consumer-generated information can spontaneously be sent and published online with people to contribute with urban noise mapping.

CHAPTER 3

METHODOLOGY

3.1. Workflow

The sequence of tasks that has been undertaken in order to fulfill the aims and objectives of the study consists of 8 steps as shown in figure 1. The first 2 steps consist of data collection and analysis, the 3rd and 4th steps are comprised of model building, the 5th and 6th steps include result evaluation, comparison and secondary testing and the 7th and 8th steps are made up of mapping and visualizing the status of noise pollution using the output from the best model.

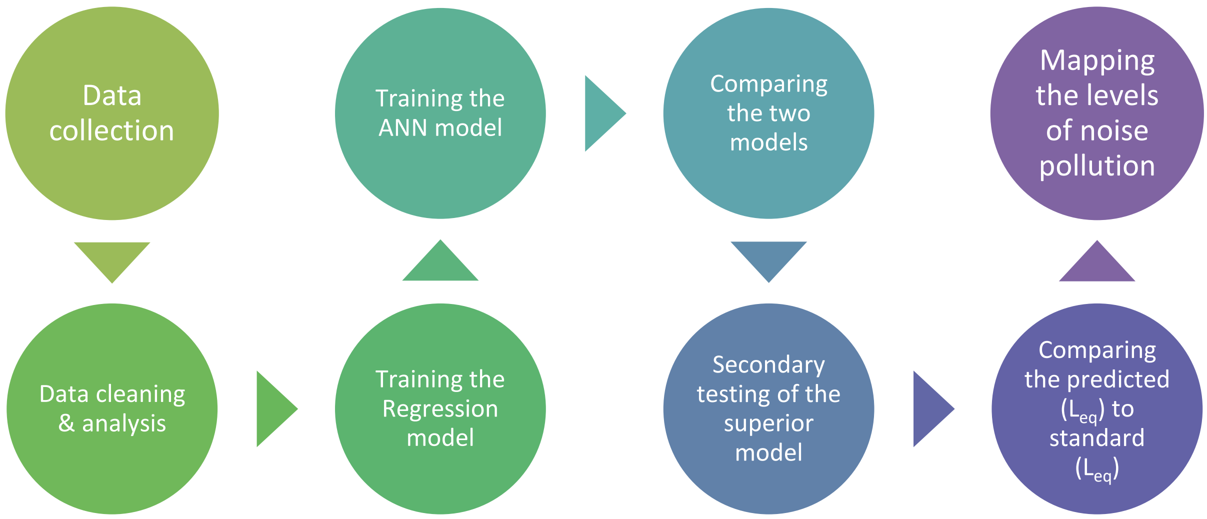

Figure 1: Workflow

3.2. Study area

After identifying the urban areas that are the most widely affected by sound pollution within the city, it was imperative that a definite boundary was outlined for the study area. In delimiting the area to be used for data collection, the variety of traffic conditions and the range of noise levels prevalent within the area was considered to be the selecting criteria. These two factors were prioritized as the introduction of a wider variety in the dataset would maximize the performance of the models, allowing them to be trained for a wider range of conditions.

Mixed urban areas at the center of the city were identified as areas that experience the highest magnitude of noise levels from the literature review. Moreover, according to Environmental Conservation Rules 1997 (ECR'97) a "Mixed Area" is an area that is mainly used for residential purposes but is also used for commercial & industrial purposes. Two mixed urban areas situated at the center of Dhaka city were selected for collecting data which are Ramna & Dhanmondi. These two areas fulfill both the selecting criteria. 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.

Ramna lies at the central part of Dhaka city and the boundary which was selected has Madhubag on its north, Begum Bazar on its south, Kalabagan on its west and Bashabo on its east. The area comprises arterials, collectors, distributors and local roads. The Ramna Park is placed at the center of the area while Dhaka University and Dhaka Medical College make up the southern part of the area within the boundary. A part of the arterial network was, however, inaccessible due to the construction work that was going on for the Mass Rapid Transit project. These are typically identified as silent zones by Environmental Conservation Rules 1997 (ECR'97). According to Environmental Conservation Rules 1997 (ECR'97), both the areas can be sub-divided into residential zones, commercial zones, industrial zones and silent zones.

Figure 1 shows the location and boundary of the study area used for collecting data used for Primary testing.

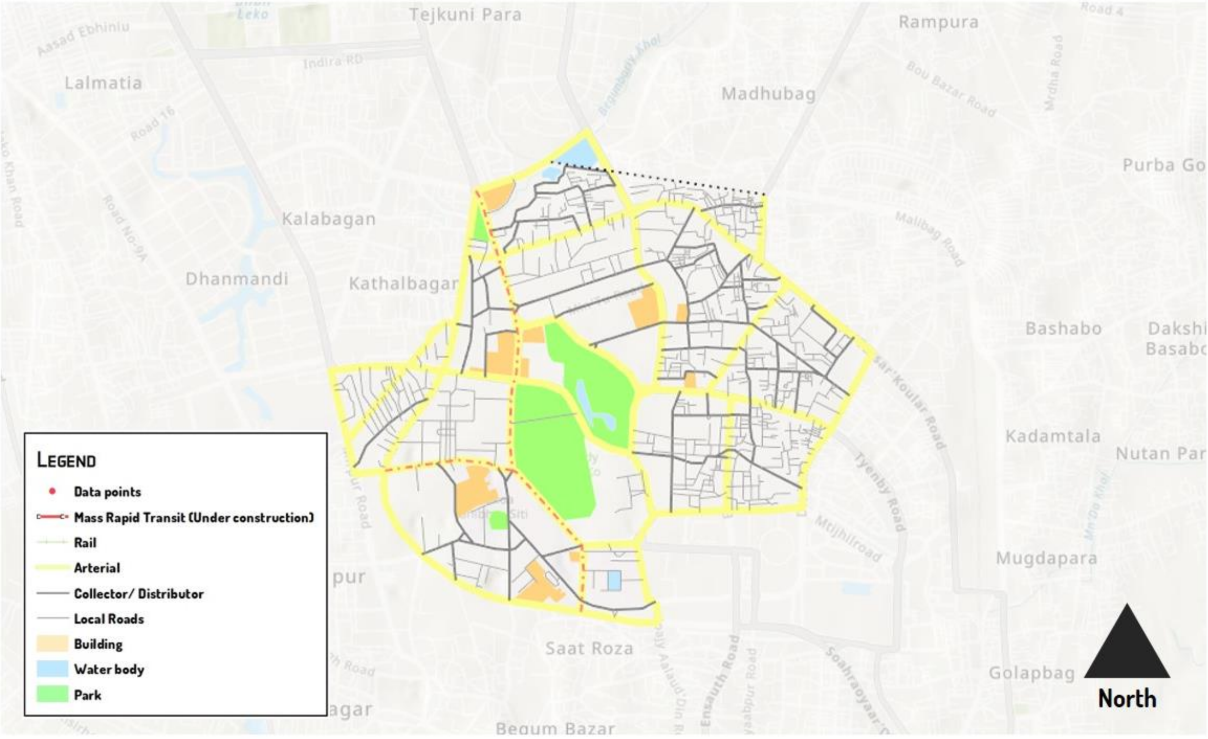

Figure 2: The location of the primary study area

3.2. Data collection and tabulation

Sampling stations were set up at different arterials and collectors with an uninterrupted flow. They were set up between 9 AM to 5 PM from November 2020 to January 2021. Locations were selected carefully such that interference from sound from sources was minimum as the presence of such noise would lead to a loss of predictive accuracy. Moreover, the sampling stations in Dhanmondi were selected with a view to incorporating a wide range of values for different traffic parameters (volume, density, speed, road width, barrier height) to allow for a fair evaluation in the second testing phase. Data were collected for all the dependent variables and the independent variables that were initially considered to be incorporated in the models. The variable consideration was mostly guided by a literature review. However, non-motorized vehicles are unique to the traffic of Dhaka city and data on such vehicles were collected due to the abundance of such vehicles in both of the study areas.

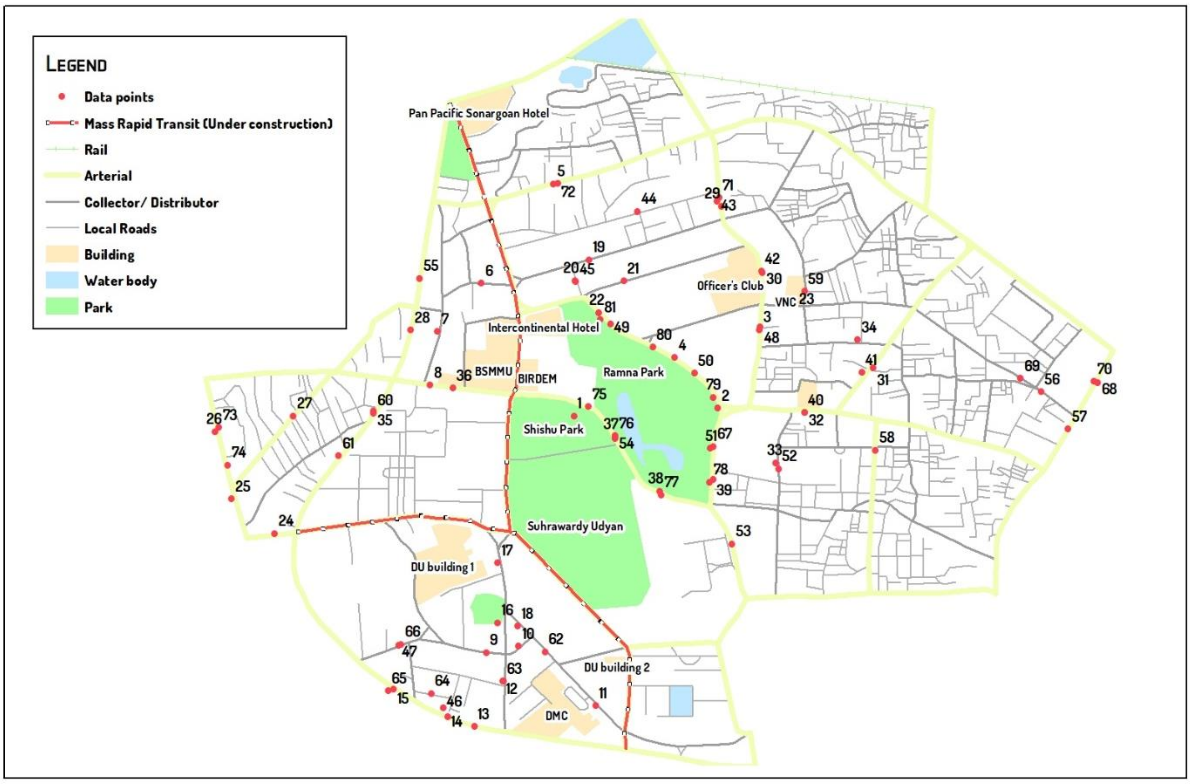

Figure 3: The sampling stations at Ramna

The Geographic coordinates (WGS 1984) for each of the sampling stations were recorded using Google Maps to aid in visualizing the results from the models using maps and to keep a record of the locations from which data were collected making sure that the sampling stations are evenly distributed throughout the area. Figure 2 shows the distribution of sampling stations within the study area.

Apart from the geographic coordinates, the field measurements included recording traffic volume, time-mean speed, barrier height, road width and A-weighted equivalent noise level (Leq) at each station. The measurements of A-weighted noise level (Leq), traffic volume and time-mean speed were conducted simultaneously. A digital camera recorded the two-way traffic volume for 15 minutes which was then extrapolated to hourly traffic volume. The videos from the digital camera were carefully scrutinized later on in order to classify the traffic volume 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, a Bushnell speed gun was used. 5 measurements for each of the previously mentioned motorized vehicles were recorded. The measurements were taken from both lanes to account for the speed of the two-way traffic. The proportion of non-motorized vehicles (NMVR) was calculated as the ratio of non-motorized vehicle volume to total volume:

$$NMVR = \frac{Volume\ of\ non-motorized\ vehicles}{Total\ volume}$$

The barriers that were present at the sampling stations comprised of buildings and walls. The heights of the walls were measured using a measuring tape and the height of the buildings were measured using BNBC. The average height of each floor for different categories of the building was recorded and the following formula was used to note down the heights of the buildings.

$$Building\ Height = Typical\ Floor\ Height\ \times Number\ of\ Floors$$

A measuring tape was used to determine road width when the traffic flow was absent. The average A-weighted noise level (Leq) was recorded using the “B&K Precision 735” sound level meter at each sampling station. 20 measurements of A-weighted noise level (Leq) were recorded within a 5-minute period. Each measurement was recorded after an interval of 15 seconds. The following formula was used to obtain the final equivalent noise level (Leq) for each of the sampling stations:

$$L_{eq} = L_{50} + \frac{(L_{10} - L_{90})^2}{60}$$

Traffic density was computed for all vehicles as well as the individual categories of non-motorized vehicles using the following formula from the extrapolated traffic volume and the measured time mean speed.

$$TD = \frac{Hourly\ Traffic\ Volume}{Time\ mean\ Speed}$$

All the data from each of the independent variables and the dependent variable were tabulated on Microsoft Excel spreadsheets. The data were cleaned to get rid of outliers and prepare them for feeding into the models. Determination of Pearson correlation coefficients (r) of the independent variables assisted in understanding their contribution to equivalent noise level (Leq). Table 1 shows all the 15 variables that were considered in the study before filtering out specific variables for each of the models.

Table 1. All the variables (Independent and dependent) considered in the study

Variable	Abbreviation	Variable	Abbreviation
Barrier Height (m)	BH	Medium Density	MD
Road Width (m)	RW	Total Density	TD
Proportion of non-motorized vehicles	NMVR	Light Velocity	LS
Light Volume	LV	Medium Velocity (Kph)	MS
Medium Volume	MV	Heavy Velocity (Kph)	HS
Heavy Volume	HV	Average Velocity	AS
Total Volume	TV	Equivalent noise level	L_{eq}
Light Density	LD		

3.2. 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). Every Neuron had a weight (wij) and a bias (b) associated with them. Multi-layer feed-forward (MLF) network architecture and the Bayesian-Regularization (BR) training algorithm were chosen to train the model in this study using the neural network toolbox of MATLAB. The training algorithm is the method used for calculating and optimizing the weight for each node in the network. This algorithm typically requires more time but can result in good generalization for difficult, small or noisy datasets. 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 hyperparameters such as the number of hidden layers and the number of neurons in each layer were selected by trial and error. A small number of neurons may result in under-fitting of the data that is supplied to the model while a model with a large number of neurons may result in a very complex model that overfits the collected data. The structure that yielded the highest accuracy for primary training and the testing data set was adopted and selected for further testing. The proposed network (Figure 3) 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. After each iteration, the weight and the bias attached to each neuron 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). The summary from the "nntrain" tool on MATLAB showed that the training was completed within 2 seconds.

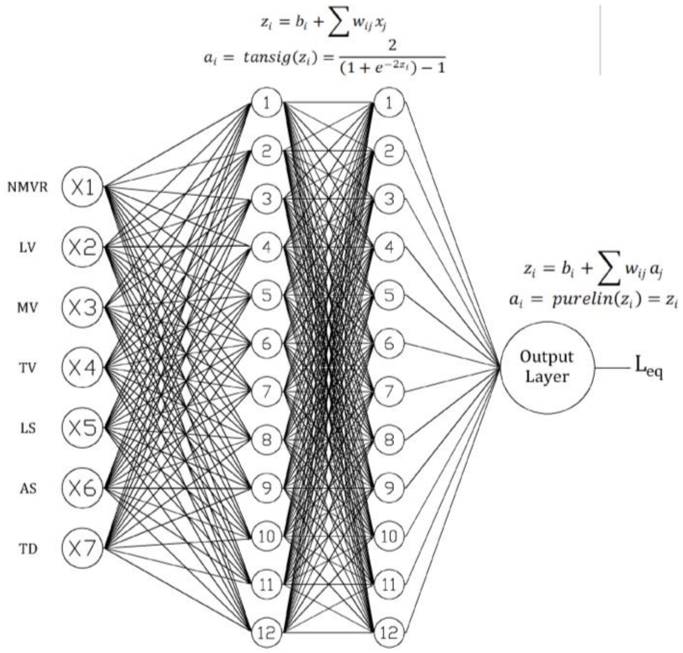

Figure 4: Artificial Neural Network Model

3.4. 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 (Leq) was predicted as a function of the following variables:

$$Leq = 0.106\,NMVR - 0.0445\,LV - 0.06\,MV + 0.061\,TV + 0.009\,TD + 0.177\,LS$$

3.5. Model comparison, statistical analyses and GIS mapping

The accuracies of the two models were analyzed and compared based on the coefficient of determination (R2), adjusted coefficient of determination and Root mean square error (RMSE) for the training set, testing set and the entire dataset (training and testing set combined). The dataset used for primary training and testing was comprised of the data collected from Ramna. The superior model determined from this comparison was tested on the secondary testing dataset from Dhanmondi.

Statistical analyses were performed to explain the variation of equivalent noise level (Leq) for different categorical and quantitative variables. The variation of equivalent noise level with respect to the categorical variables was visualized using Microsoft Excel. The Pearson correlation coefficients of different variables with respect to the equivalent noise level (Leq) explained the contribution of each of these variables to the acoustic condition of a mixed urban area. A correlation matrix was formed in order to assess this contribution.

The predicted noise levels were compared to the standard limits set according to Rule-12, Schedule-4 of ECR '97 to determine the extent to which the noise levels at different sampling stations exceeded the standards set by the aforementioned legal instrument. Each of the sampling stations was considered to be a part of one of the four zones (Residential, commercial, industrial, Silent) mentioned in ECR '97 and the predicted noise level at a sampling station from the superior model was compared to the standard noise level of the zone to which it is assigned. The levels of noise pollution at different parts of the study area were mapped using ArcGIS Pro 1.2 based on the extent of exceedance. The "optimized hotspot analysis" tool and the IDW tool was used together to form a hotspot map where the warmer regions represent areas with a higher level of noise pollution and the cooler regions represent a lower level of noise pollution.

CHAPTER 4

RESULTS AND DISCUSSION

4.1. Model comparison

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 (R2) and adjusted R2. The coefficient of determination expresses how much more accurate the predictions from a model are with respect to the mean of the observed values. Adjusted R2, besides performing the same evaluation, takes into account the size of the dataset. Therefore, adjusted R2 is a more accurate measure of a model's accuracy when comparing two models that deal with different sizes of the dataset.

Table 1 shows the results obtained from the training set, testing set and the entire data set (training and testing combined) from the data that were collected from Ramna, the primary study area. The comparison allowed us to choose the model for secondary testing and noise mapping. Before evaluating the results from the comparison, it is necessary to understand that the Ridge regression model was trained using 90% of the data and tested on the remaining 10% while the artificial neural network model had 85% of the data for training and the remaining 15% for testing. Despite the smaller size of the training dataset, the ANN model was still able to fit the data with great accuracy due to the Bayesian regularization training algorithm which can deal with relatively smaller or noisier datasets.

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
$\mathbf{R^2}$	0.80	0.69	0.90	0.71	0.82	0.70
Adjusted $\mathbf{R^2}$					0.80	0.67

The ANN model outperforms the RR model in all the evaluation criteria with an overall R2 of 0.82. Figure 5 shows a plot of the observed Leq and predicted outputs from the ANN model along with their correlation coefficient (R) and Figure 6 shows the results from the "nntrain" tool of MATLAB. The epoch was set to 1000 but the model was able to minimize the mean-square error (MSE) within 173 iterations i.e.; the weights and biases of the model were adjusted 173 times.

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% (R2 = 0.90) of the primary testing data compared to the 71% (R2 = 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 generalization capability is determined by the complexity of the model; an overly complex or simple model would lead to a low R2. The complexity of the model relies on the size and nature of the dataset. Therefore, a model that does not lead to over-fitting or under-fitting is one that produces high accuracy in predicting unseen data.

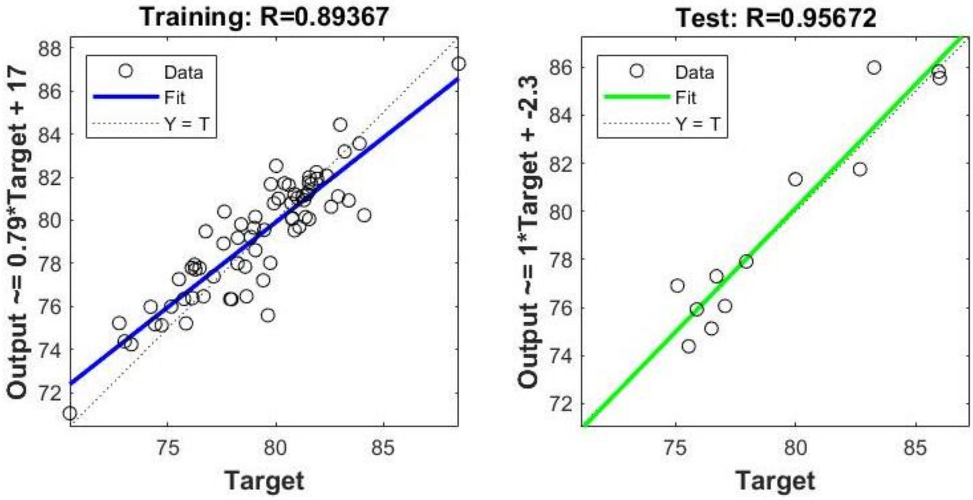

Figure 5

Figure 5: Comparison between observed Leq and predicted Leq from the ANN model for training and testing data

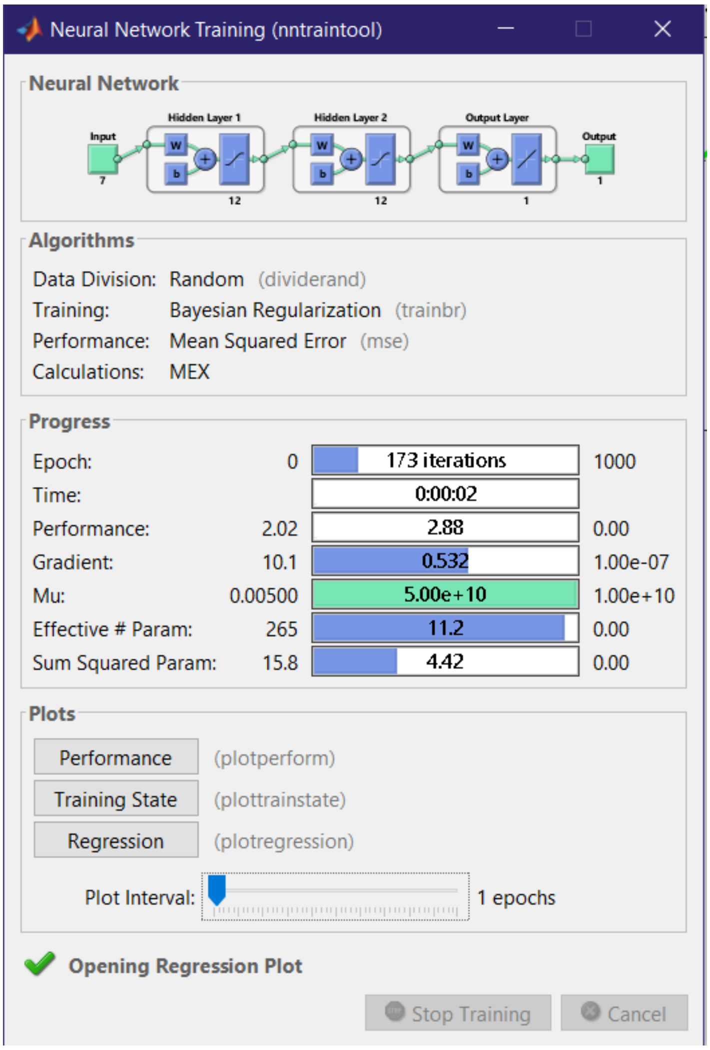

Figure 6: Summary of the training of the ANN model on MATLAB

4.2. Secondary testing

The results from Ramna showed that the ANN model was able to generalize with 90% accuracy. However, whether the ANN model would be able to replicate the accuracy in a different area cannot be ascertained from these results. Hence, the trained ANN model's generalization ability was further validated by its predictions in a different area. The dataset from Dhanmondi consisted of 7 samples. The 7 samples comprised of equivalent noise levels within a range of 75 dBA to 82 dBA.

Table 2 shows a comparison of the observed Leq and the Leq predicted by the ANN model in Dhanmondi, the secondary study area. The model was able to generalize and explain 83%

(R2=0.83) of the variability in the secondary dataset with an RMSE of 1.17. The comparison of the observed Leq and the predicted Leq suggests that the model was relatively less accurate in predicting the sound levels from 77 dBA to 80 dBA.

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		

4.3. Statistical analyses

Statistical analyses show the distribution of Leq with respect to two categorical variables: time and day of the week. The results, interpretation of the results of the analyses along with their interpretations are presented. The correlation coefficients of each of the independent variables with respect to the other variables are presented in a correlation matrix for evaluating their contribution to the predictions from the models.

The visualization (Figure 6) of variation of Leq with respect to time of the day reveals that the highest recorded Leq (88.5 dBA) falls within the interval of 2 PM – 3 PM. The field observations also suggested the same as this was usually the peak hour of traffic. Moreover, the equivalent noise levels also deviated the most (σ=5.76) during this interval. This can be explained by the field observations as the traffic was concentrated in the comparatively busier areas that fall within the commercial zones and industrial zones while the silent zones and residential zones had very little flow in this interval.

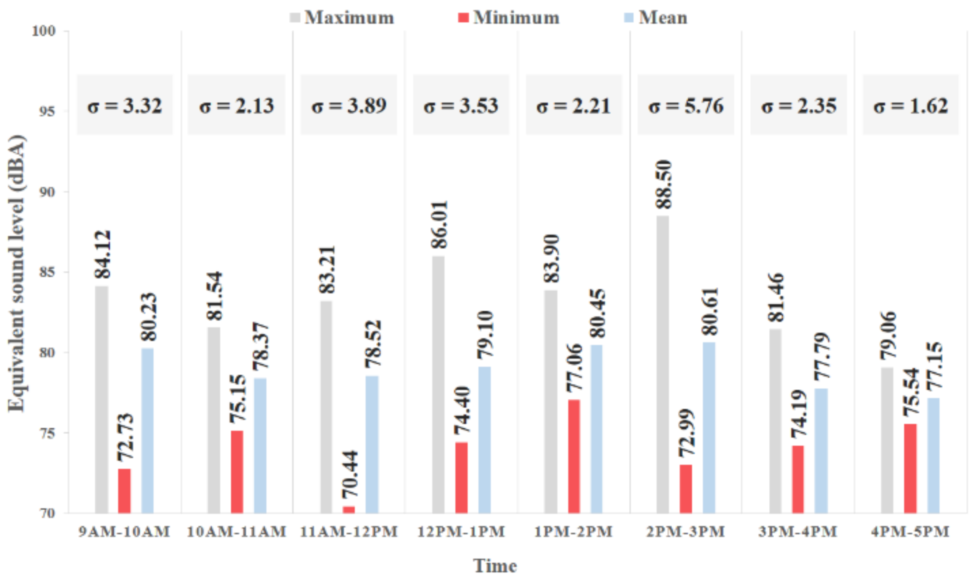

Figure 7: Figure 3. Hourly variation of Leq in Ramna

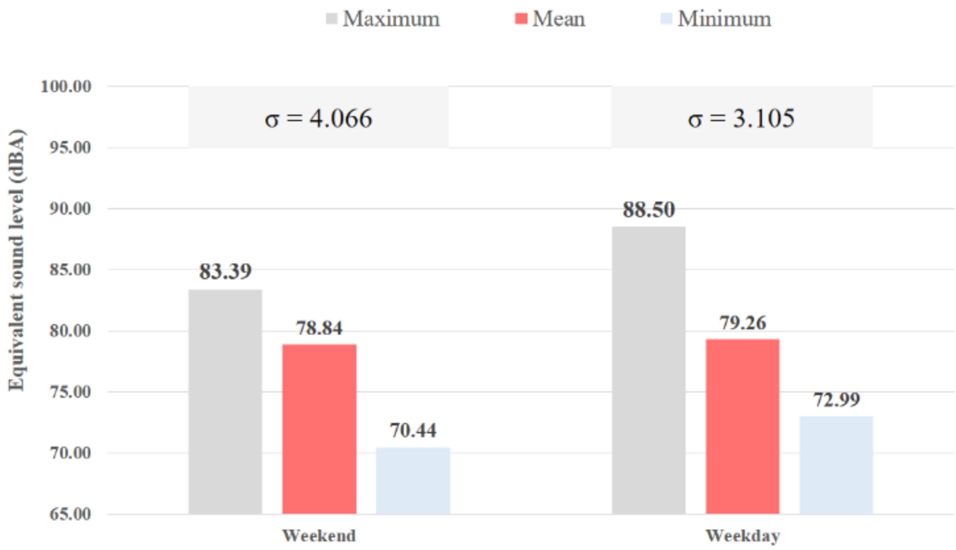

Figure 8: Variation of Leq on weekdays and weekends in Ramna

From figure 8, it is evident that the weekdays experience the highest levels of noise while the observed equivalent noise levels show a higher deviation (σ=4.07) from their mean on weekends. This is analogous to the field observations as a higher number of vehicles were observed on weekdays as compared to that on weekends.

From table 3, it is evident that NMVR was the only independent variable that showed a high negative correlation (r = -0.58) with Leq 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 Leq. The negative correlations of NMVR with AS (r= -0.60) and TV (r= -0.66) support the field observations. The independent variable that showed the highest positive correlation (r=0.76) with L_{eq} was total volume. This was also similar to the observations made during data collection as the arterials with the highest number of vehicles experienced the highest levels of noise.

Table 3. Correlation matrix of all the variables considered in the study

	BH	**RW**	**NMVR**	**LV**	**MV**	**HV**	**TV**	**LD**	**MD**	**TD**	**LS**	**MS**	**HS**	**AS**	**L_{eq}**
BH	1.00	0.05	0.11	-0.07	-0.01	-0.17	-0.07	-0.04	0.01	-0.03	-0.22	-0.18	-0.10	-0.21	-0.08
RW	0.05	1.00	-0.32	0.50	0.54	0.46	0.58	0.46	0.48	0.53	0.47	0.47	0.52	0.45	0.59
NMVR	0.11	-0.32	1.00	-0.62	-0.61	-0.27	-0.66	-0.55	-0.50	-0.56	-0.55	-0.59	-0.53	-0.60	-0.58
LV	-0.07	0.50	-0.62	1.00	0.87	0.01	0.97	0.97	0.79	0.93	0.41	0.37	0.37	0.38	0.73
MV	-0.01	0.54	-0.61	0.87	1.00	0.09	0.95	0.86	0.97	0.95	0.30	0.31	0.38	0.30	0.67

HV	-0.17	0.46	-0.27	0.01	0.09	1.00	0.14	-0.04	0.05	0.09	0.40	0.42	0.55	0.44	0.37
TV	-0.07	0.58	-0.66	0.97	0.95	0.14	1.00	0.95	0.89	0.97	0.41	0.39	0.44	0.40	0.76
LD	-0.04	0.46	-0.55	0.97	0.86	-0.04	0.95	1.00	0.83	0.96	0.22	0.23	0.28	0.22	0.67
MD	0.01	0.48	-0.50	0.79	0.97	0.05	0.89	0.83	1.00	0.94	0.13	0.09	0.25	0.10	0.59
TD	-0.03	0.53	-0.56	0.93	0.95	0.09	0.97	0.96	0.94	1.00	0.23	0.20	0.33	0.19	0.69
LS	-0.22	0.47	-0.55	0.41	0.30	0.40	0.41	0.22	0.13	0.23	1.00	0.78	0.62	0.89	0.59
MS	-0.18	0.47	-0.59	0.37	0.31	0.42	0.39	0.23	0.09	0.20	0.78	1.00	0.69	0.94	0.53
HS	-0.10	0.52	-0.53	0.37	0.38	0.55	0.44	0.28	0.25	0.33	0.62	0.69	1.00	0.67	0.59
AS	-0.21	0.45	-0.60	0.38	0.30	0.44	0.40	0.22	0.10	0.19	0.89	0.94	0.67	1.00	0.58
L_{eq}	-0.08	0.59	-0.58	0.73	0.67	0.37	0.76	0.67	0.59	0.69	0.59	0.53	0.59	0.58	1.00

4.4. Noise pollution and GIS mapping

The standard noise levels for the four different zones of an urban area, to which the predicted noise levels from the ANN model were compared, are stated below.

1) **Silent zones**- 45 dBA 3) **Commercial area**- 70 dBA

2) **Industrial area**- 75 dBA 4) **Residential area**- 50 dBA

The comparison of predicted equivalent noise levels from the ANN model to the standard limits from ECR ’97 reveals the extent of noise pollution in Ramna (Figure 9). 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.

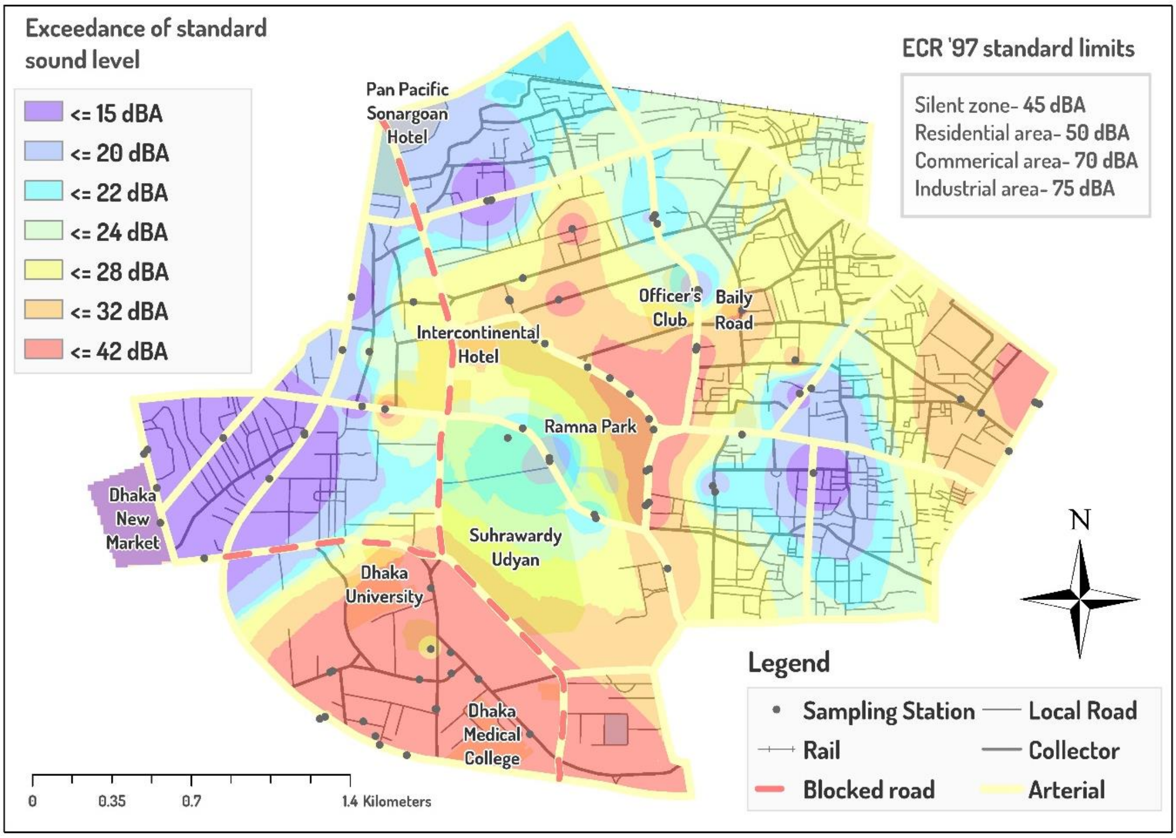

Figure 9: The extent of noise pollution in Ramna

The higher magnitude of exceedance of noise level indicates a higher level of noise pollution with respect to the standards limit. However, it does not necessarily indicate a higher equivalent noise level (Leq). The silent zones, despite being the areas with the lowest levels of noise still had a higher level of pollution due to the lower standard for those areas.

4.5. Limitations

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.

Moreover, this study only considered traffic variables that are associated with an uninterrupted flow. A model consisting of traffic variables that include interrupted flow would result in a more complex but comprehensive model. The flow interruption is a distinct feature of the roads of Dhaka and this is the future scope of the study that has been conducted. Although interrupted flow has been considered to be a part of a study that developed a prediction model for the traffic of Dhaka, the incorporation of machine learning models and interrupted flow in the same study has not been conducted yet, according to the authors' knowledge.

CHAPTER 5

CONCLUSION

The ANN model was found to be superior compared to the RR model in predicting the equivalent noise level (Leq) 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 Leq 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.

CHAPTER 6

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.