Grenze International Journal of Engineering and Technology, June Issue

Real Time Noise Pollution Prediction in a City using Machine Learning and IoT

A.Manoj Pavan Sai Kumar[1], S.V.V.D.Jagadeesh[2], G.Gagan Siddarth[3], K. Manasa Phani Sri[4] and V.Rajesh[5]
[1-5]Lakireddy Bali Reddy College of Engineering, Mylavaram, India
Email: manojpavansai76@gmail.com, jagadeesh.s@lbrce.ac.in, ggagan2121@gmail.com, manasaphanisrikoya3559@gmail.com, rajeshvallepu7@gmail.com

***Abstract*—This study introduces a novel approach to tackle urban noise pollution in Vijayawada city by leveraging Machine Learning (ML) and Internet of Things (IoT) technologies. The aim is to predict real-time noise pollution levels, with a particular emphasis on the impact of road traffic. The methodology integrates four regression models (Decision Tree, Gradient Boosting, Support Vector, Linear Regression with Grid Search) and three forecasting models (ARIMA, Holt-Winters Exponential Smoothing, SARIMA), utilizing data collected from IoT acoustic sensors across diverse landscape configurations. This research underscores the importance of ML and IoT in addressing urban noise pollution challenges and implications for urban planning and policy making, acknowledging both strengths and limitations of the proposed approach. In Existing models there is no real time data used for prediction. But this case study includes real time data collected from IOT sound sensor which give better results.**

***Index Terms*—Urban noise pollution, Machine Learning, Internet of Things(IoT), Real-time prediction, Regression models, Forecasting models, Road traffic impact, Acoustic sensors, Urban planning, Policy implications.**

I. Introduction

The rapid urbanization of cities worldwide has brought about numerous challenges, with noise pollution emerging as a significant issue. As urban areas continue to expand and become more densely populated, effectively managing noise pollution is crucial for the well-being of residents and the sustainability of urban development. This study explores a novel approach to addressing this challenge by using Machine Learning (ML) and the Internet of Things (IoT) to predict real-time noise pollution levels with in Vijayawada city. Significant traffic congestion and safety issues have arisen due to the increasing demand for transportation and construction, potentially harming one's physical health, general well-being, and even mortality. DeCoensel, Brown and Tomerini claim that noise pollution can be caused by variations in the frequency, duration, and intensity of noise events resulting from traffic. Overexposure to sound in the workplace, school, home, and other settings can have long-term consequences on human development and health, as well as impact performance and other concurrent activities.

Accurate road traffic data analysis is necessary for creating effective traffic management plans that reduce the negative effects of traffic on the neighborhood When assessing the environmental quality and population health of urban areas, it is imperative to consider noise pollution resulting from vehicular traffic. Continuous and

Grenze ID: 01.GIJET.10.2.644

comprehensive data on ambient noise levels can be gathered by strategically placing IoT-enabled noise sensors throughout urban areas. These sensors, which can measure a range of sound frequencies and intensities, are the cornerstone of an extensive system designed to track and forecast noise pollution in real time. Combining this system with state-of-the-art machine learning algorithms empowers city planners and citizens to proactively tackle the adverse impacts of excessive noise on public health and urban quality of life.

TABLE I. GENERALIZED NOISE LEVELS

Category of Area/Zone	Limits in dB(A) Leq Day Time (from 6.00 a.m. to 10.00 p.m.)	Limits in dB(A) Leq Night Time (from 6.00 a.m. to 10.00 p.m.)
Industrial Area	75	75
Commercial Area	65	55
Residential Area	55	45
Silence Zone	50	40

The system not only provides instant insights into noise levels but also forecasts potential future trends. The development of efficient noise reduction strategies, such as traffic management strategies, noise barriers, and vehicle emission limits, can be guided by accurate noise measurement and analysis.

Many benefits can arise from these initiatives, such as enhanced living conditions for locals, a decrease in stress and related health issues, and a more sustainable urban environment. The current study explores the effects of several variables, including the distance from the road to the receiving point, the type of landscape, the time of day, and the presence of heavy vehicles.

A. Related Work

Research on noise pollution prediction the use of device getting to know and IoT encompasses quite a number modern procedures that leverage superior technology for powerful tracking and forecasting. IoT primarily based totally structures install networks of low-value sensors to collect continuous, real-time noise facts throughout diverse city locations. This facts is then processed the use of device getting to know algorithms consisting of regression fashions and aid vector machines to expect destiny noise levels. These predictive fashions regularly comprise extra facts reassets like climate situations and visitors styles to decorate accuracy. Spatiotemporal evaluation the use of Geographic Information Systems (GIS) in addition aids in visualizing noise distribution and figuring out hotspots. have proven that those included structures can substantially enhance noise control strategies, supplying precious insights for city planners and policymakers to increase focused interventions for noise reduction. This frame of labor highlights the capacity of mixing device getting to know and IoT to deal with the demanding situations Studies of noise pollutants in cutting-edge cities.

B. Problem Statement

The implications of noise pollution for readmission risks and mitigation strategies have been extensively researched by scholars due to recent studies emphasizing the issue as a global concern. Research has examined several variables both individually and collectively in an attempt to determine the predictive significance of sound pollution, decibel levels, and vehicle flow. Insights into the intricate dynamics of noise pollution are provided by this literature review, which also serves as a roadmap for the development of targeted interventions to mitigate its adverse effects. IoT sound sensors that provide real-time data were used to create a unique dataset tailored for the study to predict the impact of traffic noise in the city. To This research is important because it has the potential to transform urban noise management by providing decision-makers with practical information to implement targeted interventions. Cities can enhance their capacity to create more sustainable urban environments by harnessing the potential of ML and IoT technologies.

II. METHODOLOGY AND SOLUTION APPROACH

This section aims to describe the procedure for gathering the data and configuring the settings required for training and testing the prediction models. To ensure that the models have the necessary data to predict traffic noise accurately, it is essential to create the dataset and optimize the model's parameters. In Figure 1

A. IoT Design

Simulations were used in the case study to assess how estimates of sound event characteristics are impacted by considering realistic power distributions in vehicle noise. The objective is to collect relevant sound data,

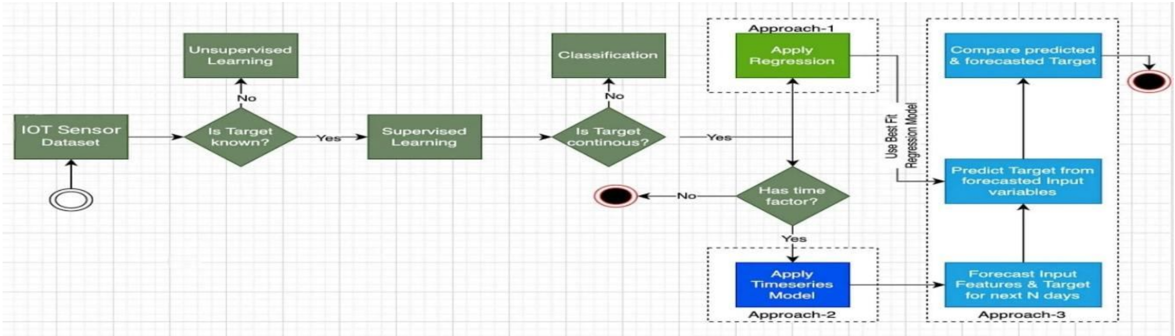

Figure 1. Flow Chart on Methodology

including noise levels, frequencies, timestamps, and locations, by strategically placing sound sensors in areas of interest, such as city streets or public spaces.

Sound Sensor works on a signal that is electrical and can be understood by a micro controller. VCC, GND, Digital Out, and Analog Out are the four pins on the sound sensor module. Digital output is provided by the DO pin, whereas analog readings are provided by the AO pin. The pinout of the sound sensor is illustrated in the IoT Connectivity Diagram. The power supply pin, or VCC, can be connected to a 3.3V or 5V supply. The supply voltage affects the analog output in various ways. AOUT receives analog readings directly from the sensor, DOUT is the digital output pin that indicates sound detection, and GND is the ground pin. In Figure 3.

IoT connection is a Establishment of connection between the sound sensors that have been deployed and an IoT platform. Each sensor should be equipped with communication capabilities to transmit data to the central IoT platform. This connection can be set up using various communication protocols. Configure the IoT platform to receive and manage data from the sound sensors, including setting up the necessary channels, protocols, and security measures for secure data transmission to One of the popular of IoT cloud platform is called **ThingSpeak.** In Figure 2.

Through the use of **ThingSpeak**, data can be sent over a private or public channel to the cloud for storage by sensors, devices, and online platforms. Though public channels can be used for data sharing, **ThingSpeak** stores data in private channels by default. Ensure that the sound sensors consistently transmit real-time data, including timestamps, frequency spectrum, sensor location, and sound intensity, to the IoT platform. Conduct checks to ensure that the sound data collected is accurate and intact. This may involve verifying sensor readings, addressing corrupted or missing data, and maintaining data consistency. In order to facilitate accurate temporal analysis of the sound data, **ThingSpeak** synchronize timestamps across all sensors.

To layout an IoT dataset for predicting noise pollutants, diverse vital elements want to be protected to make sure correct and complete statistics collection. The dataset ought to embody Distance, which refers back to the proximity of noise reassets to size points, and Time, shooting temporal versions in noise levels. Landscape functions ought to be recorded to account for herbal and synthetic obstacles that have an effect on sound propagation. Road Surface kinds ought to be detailed, as one of the kind substances generate various noise levels. The Vehicles in line with hour (Vehicles/h) metric is essential for knowledge site visitors density, even as the Speed Limit gives insights into anticipated car noise profiles. Additionally, the Percentage of Heavy Vehicles is vital to differentiate among noise contributions from popular and heavy-obligation vehicles. Together, those elements allow a strong prediction version for noise pollutants the use of IoT, leveraging real-time and context-conscious statistics for more desirable accuracy and responsiveness in city noise management.

TABLE II. DESCRIPTION ABOUT IOT DATASET FROM SENSOR

Features	Values	Number of Observations	Target Variable
Distance	15,30,45,60,75,90(m)	6480	Noise
Time	Day, Night		
Landscape	None, Tree, Wall		
Road surface	Asphaltic Concrete, Uneven surface		
Vehicles/h	10,20,40,50,100,200,400,500,1000,2000		
Speed limit	60,80,100,120,140(km/h)		
Percentage of heavy vehicles	5,10,20(%)		

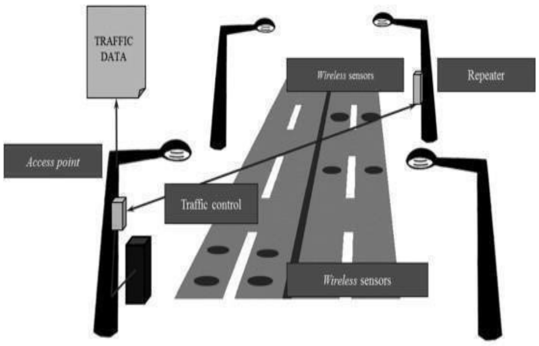

Figure 2. IoT Sensors Connectivity

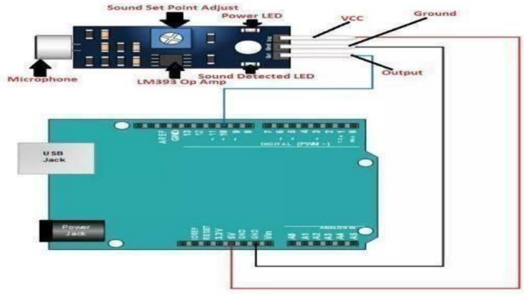

Figure 3. Sound Sensor Connectivity

B. Model Preparation

i. Decision Tree Regressor

A Decision Tree Regressor operates with the aid of using splitting the information into subsets primarily based totally at the values of enter features, developing a tree-like shape of selection nodes and leaf nodes. Each selection node represents a check on a feature, and every leaf node represents a anticipated value. The splits are made to decrease the variance inside the subsets, efficiently partitioning the information in a manner that reduces prediction errors. Decision Tree Regressors are intuitive, smooth to interpret, and may seize non-linear relationships.

ii. Support Vector Regressor

Support Vector Regressor (SVR) is a machine learning model used for regression tasks, which involves predicting a continuous output variable. SVR is based on the principles of Support Vector Machines (SVM), primarily used for classification. In SVR, the model attempts to fit the best line within a threshold margin that captures the maximum number of data points. It aims to minimize prediction error while also controlling model complexity to avoid overfitting.

iii. Gradient Boosting Regressor

Gradient Boosting Regressor is a effective gadget gaining knowledge of method used for regression tasks. It builds an ensemble of susceptible prediction models, usually choice bushes, in a sequential manner. Each new tree is educated to accurate the mistakes of the preceding bushes with the aid of using minimizing a loss function. The set of rules combines the predictions from all of the bushes to provide a final, sturdy prediction model. Gradient Boosting is thought for its excessive accuracy and capacity to address diverse forms of data, making it powerful for complicated datasets

iv. Linear Regression with Grid Search

Linear Regression with Grid Search is a technique used to optimize the overall performance of a linear regression version with the aid of using systematically looking through a predefined set of hyperparameters. Linear regression pursuits to version the connection among a established variable and one or greater impartial variables with the aid of using becoming a linear equation. It works with the aid of using exhaustively checking out all viable mixtures inside a particular variety and comparing every version the use of cross-validation.

v. ARIMA

ARIMA, which stands for Auto-Regressive Integrated Moving Average, is a famous statistical technique used for time collection forecasting. It combines 3 components: the autoregressive (AR) part, which fashions the dependency among an statement and numerous lagged observations; the integrated (I) part, which entails differencing the facts to make it stationary; and the transferring common (MA) part, which fashions the dependency among an statement and a residual mistakess from a transferring common version carried out to lagged observations. ARIMA fashions are effective due to the fact they could manage numerous kinds of facts styles and trends, making them appropriate for forecasting in numerous domain names like finance, economics, and environmental science. The version calls for cautious parameter choice and tuning, generally denoted as ARIMA (p, d, q), in which p is the order of the AR part, d is the diploma of differencing, and q is the order of the MA part.

vi. Holt-Winters Exponential Smoothing

Holt-Winters Exponential Smoothing is a forecasting technique that extends simple exponential smoothing to capture trends and seasonality in time series data. It involves three components: level (representing the average value of the series), trend (reflecting the direction and rate of change), and seasonality (depicting the periodic fluctuations). By updating these components iteratively, Holt-Winters method effectively predicts future values by considering both recent observations and historical patterns. The method is particularly useful for short to medium-term forecasting where trends and seasonal variations play a significant role. It provides a flexible and intuitive approach to handle time series data, making it widely adopted in various fields such as finance, economics, and supply chain management.

vii. SARIMA

Seasonal Autoregressive Integrated Moving Average (SARIMA) is a complicated time collection forecasting approach used to version and are expecting records with seasonal patterns. SARIMA includes 3 principal components: Seasonal (S), Autoregressive (AR), and Moving Average (MA). The seasonal element captures periodic fluctuations withinside the records, whilst the autoregressive element fashions the connection among an commentary and some of lagged observations. The transferring common element represents the connection among an commentary and a residual blunders from a transferring common version implemented to lagged observations. By accounting for seasonality in conjunction with autoregressive and transferring common components, SARIMA can offer correct forecasts for time collection records

C. Model Evaluation

i. MSE (MSE (Mean Square Error)

The Mean Squared Error (MSE) metric measures the average of the squared variances between the estimated and actual values. A lower mean square error (MSE) in noise pollution prediction models indicates a higher level of accuracy, as the model's predictions closely align with the actual values.

ii. RMSE (Root Mean Square Error)

On comparing the actual values in the same units as the response variable to the model's predictions, RMSE offers insight into the typical deviation. For assessing the model's actual predictive performance, this makes it a useful metric.

iii. R-SQUARE

The R-squared (R2) value measures the extent to which the independent variable(s) in a regression model explain the variance in the dependent variable. It is also referred to as the coefficient of determination. In noise pollution prediction models, a higher R-squared value indicates a stronger agreement between the model and the observed data, suggesting a higher level of correlation between the variables under study.

$$\mathrm{MSE} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2 \qquad \mathrm{R}^2 = 1 - \frac{\sum(y_i - \hat{y}_i)^2}{\sum(y_i - \bar{y})^2}$$

$$\mathrm{RMSE} = \sqrt{\mathrm{MSE}} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{y}_i)^2}$$

Where: y_i - true values of y
$\hat{y}_i$ - predicted values of y
$\bar{y}$ - average value of y

TABLE III. MODEL EVALUATION

MODEL	RMSE	ACCURACY	PREDICTED NOISE
Decision Tree Regressor	2.18	0.95	66.67
Gradient Decent Regressor	2.15	0.97	69.70
Linear Regressor with Grid Search	5.04	0.93	74.68
Support Vector Regressor	2.15	0.96	70.30
ARIMA	10.26	0.89	62.45
Holt-Winters Exponential Smoothing	9.43	0.90	63.80
SARIMA	10.46	0.88	62.70

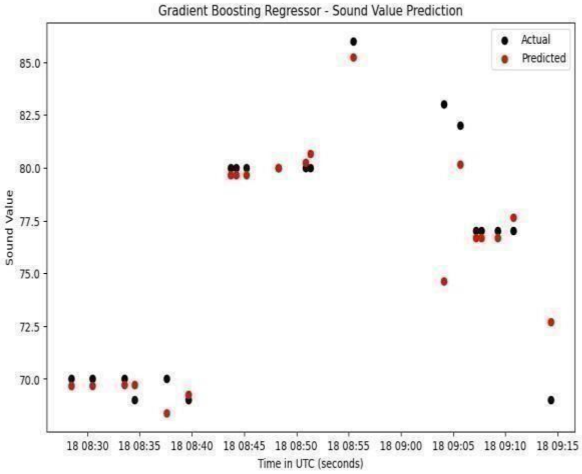

Figure 4. Gradient Boosting Regressor

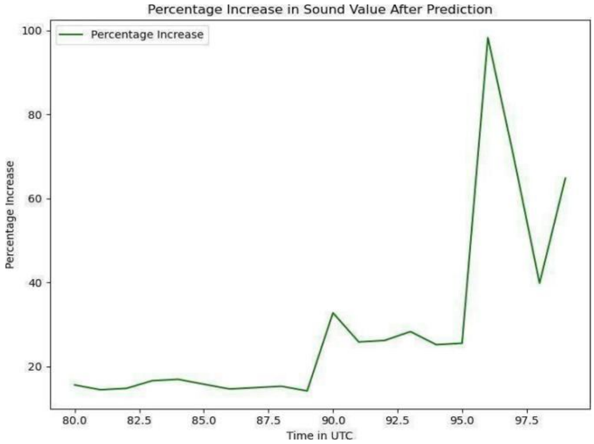

Figure 5. Percentage Increase of Noise

III. CONCLUSIONS

Through an analysis of contributing factors, we investigated methods to estimate the levels of noise pollution generated by road traffic in this study. Taking various factors into consideration Our results highlight the complexity of noise pollution dynamics and underscore the importance of comprehensive modelling approaches in understanding and predicting its impacts. The **Gradient Boosting Regressor** model should be selected based on the specific characteristics of the dataset, taking into account the interpretability of the results, and aiming for a predictive accuracy of **97.18%**. The fact that a **29.51%** increase in noise levels is predicted for the upcoming years annually. The results of this study can be utilized by environmental scientists, policymakers, and urban planners to more effectively address the issue of noise pollution and work towards creating more vibrant and sustainable communities.

REFERENCES

[1] J. Shin, Y. Kim, S. Yoon, and K. Jung, ”Contextual-CNN: A Novel Architecture Capturing Unified Meaning for Sentence Classification,” 2018 IEEE International Conference on Big Data and Smart Computing(Big Comp), Shanghai, China, 2018, pp. 491-494, doi: 10.1109/BigComp.2018.00079.

[2] W. Lu, Y. Duan, and Y. Song, ”Self-Attention-Based Convolutional Neural Networks for Sentence Classification,” 2020 IEEE 6th International Conference on Computer and Communications (ICCC), Chengdu, China, 2020, pp. 2065-2069, doi: 10.1109/ICCC51575.2020.9345092.

[3] J. Yang and J. Yang, ”Aspect Based Sentiment Analysis with Self Attention and Gated Convolutional Networks,” 2020 IEEE 11th International Conference on Software Engineering and Service Science (ICSESS),Beijing, China, 2020, pp. 146-149, doi: 10.1109/ICSESS49938.2020.9237640.

[4] Feng, X. Zhang, and X. Song, ”Unrestricted Attention May Not Be All You Need–Masked Attention Mechanism Focuses Better on Relevant Parts in Aspect-Based Sentiment Analysis,” in IEEE Access, vol. 10,pp. 8518-8528, 2022, doi: 10.1109/ACCESS.2022.3142178.

[5] Pamungkas, E.W., Basile, V., Patti, V. ”Towards multidomain and multilingual abusive language detection: a survey.” Pers Ubiquit Comput 27, 17–43 (2023). https://doi.org/10.1007/s00779-021-01609-1 .

[6] R. Amrutha and K. R. Bindu, ”Detecting Hate Speech in Tweets Using Different Deep Neural Network Architectures,” 2019 International Conference on Intelligent Computing and Control Systems (ICCS), Madurai, India, 2019, pp. 923-926, doi: 10.1109/ICCS45141.2019.9065763.

[7] Asudani, D.S.,Nagwani, N.K. Singh, P. ”Impact of word embedding models on text analytics in deep learning environment: a review.” Artif Intell Rev 56,10345–10425(2023).https://doi.org/10.1007/s10462-02310419-1.