ISSN: 2458-8989

Natural and Engineering Sciences

NESciences, 2025, 10 (3): 382-393 doi: 10.28978/nesciences.1811131

# Use of Support Vector Machines (SVM) for Classifying Pollution Sources in Urban Environments

Rishabh Bhardwaj [1*], Rakhi Jha [2], Dr. Mercy Paul Selvan [3], Dr. Koushik Sar [4], M. Sunil Kumar [5], Lakshman Singh[6]

[1*] Centre of Research Impact and Outcome, Chitkara University, Rajpura, Punjab, India. E-mail: rishabh.bhardwaj.orp@chitkara.edu.in

[2] Assistant Professor, Department of Computer Science & IT, ARKA JAIN University, Jamshedpur, Jharkhand, India. E-mail: rakhi.j@arkajainuniversity.ac.in

[3] Professor, Department of Computer Science and Engineering, Sathyabama Institute of Science and Technology, Chennai, India. E-mail: mercypaulselvan.cse@sathyabama.ac.in

[4] Assistant Professor, Department of Agronomy, Institute of Agricultural Sciences, Siksha 'O' Anusandhan (Deemed to be University), Bhubaneswar, Odisha, India.
E-mail: koushiksar@soa.ac.in

[5] Assistant Professor, Department of Mechanical Engineering, Faculty of Engineering and Technology, JAIN (Deemed-to-be University), Ramnagar, Karnataka, India.
E-mail: sunilkumar.m@jainuniversity.ac.in

[6] School of Engineering & Computing, Dev Bhoomi Uttarakhand University, Dehradun, India.
E-mail: ce.lakshman@dbuu.ac.in

## Abstract

This study investigates the application of Support Vector Machines (SVM) to classify major air pollution sources in Bengaluru, India, by integrating routinely collected air quality, meteorological, and land-use data from 2021 to 2023. The main objective is to assess whether commonly available datasets can accurately distinguish between vehicular, industrial, domestic, and biomass burning sources. Air pollutant concentrations ($PM_{2.5}$, $PM_{10}$, $NO_2$, $SO_2$, CO, $O_3$) were combined with meteorological parameters, satellite-derived land-cover indices (NDVI, NDBI), and urban activity datasets to develop feature vectors for classification. Data preprocessing ensured quality control, synchronisation, and normalisation, while principal component analysis reduced dimensionality. An SVM with a radial basis function kernel was trained and evaluated using stratified cross-validation, with model stability improved through auxiliary Support Vector Regression (SVR) for temporal smoothing. The classifier achieved an overall accuracy of

*Corresponding Author: Rishabh Bhardwaj, E-mail: rishabh.bhardwaj.orp@chitkara.edu.in*

70% (Cohen's kappa: 0.59), with best performance for biomass burning (F1-score: 0.78) and industrial emissions (F1-score: 0.68), and moderate success in differentiating vehicular (F1-score: 0.63) and domestic (F1-score: 0.64) sources. Predictor importance analysis revealed that road density, wind-adjusted pollutant concentrations, and land-cover indices were most influential. Spatial and temporal validation demonstrated consistency with external ground-truth activities. The findings suggest that SVM, supplemented by routine datasets, provides a robust, cost-effective alternative to traditional source apportionment for urban air quality management, with potential for real-time application in rapidly growing cities.

**Keywords:**

*Air pollution source, machine learning, meteorological data, support vector machines, urban air quality.*

**Article history:**

Received: 27/06/2025, Revised: 15/08/2025, Accepted: 15/09/2025, Available online: 12/12/2025

## Introduction

The safety and wellness of the public, as well as the environment, face critical challenges from urban air pollution. The rising urban population, increasing industrial activity, and the increasing number of motor vehicles all contribute to the rampant increase of pollutants. The pollutants include particulate matter, nitrogen oxides, sulfur dioxide, carbon monoxide, and ozone. These pollutants combine to worsen the air quality, which exacerbates respiratory and cardiovascular diseases as well as exacerbates climate change (Sowmya et al., 2023).

Studies in the recent past have explored the impact of pollution-related emissions and their relation to greenhouse gas dynamics. In aquatic ecosystems, the diurnal changes have an amplifying effect on climate change impacts (Chandravanshi & Neetish, 2023). Thus, identifying key pollution sources is vitally important for focused pollution alleviation measures and corrective actions. Aquatic ecosystem pollution by heavy metals and its impacts on ecology and human health are other examples of urban pollution (Salayev et al., 2025). The risks of urban pollution are further increased by surface and groundwater pollutants such as landfill leachates and other wastes from the city (Sredić et al., 2024).

Caustic charges exist in traditional ways of carrying out source apportionment-receptor modelling or chemical speciation since they require intensive sampling, are very costly, and need laboratory tests that may not be possible to conduct on a regular basis in many Indian cities. Therefore, most monitoring setups consider just pollutant concentration data, which cannot easily separate superimposed sources or meteorological factors affecting dispersion and chemical transformation. This breeds a gap in the practical, real-time means of using extant monitoring and environmental data for source classification. In addition to the issues brought by air pollution, the role that digitisation, its related systems, and information technologies play in eco-sustainability has been increasingly recognised in the related fields. (Shimazu, 2024) pointed out that sustainable supply chain management can be improved with configurational strategies and digitisation, further emphasising the wider opportunity of data-driven and intelligent methods for eco-issues. Similarly, (Anaya Menon & Srinivas, 2023) illustrate how cloud analytics and AI can motivate cross-sector collaboration for climate issues, further emphasising the growing importance of computational tools for eco and sustainability issues.

Over the last decade, machine learning has advanced the state of the art in pattern recognition, especially in understanding the overcomplex and incredibly high-dimensional data. In particular, studies in traffic engineering have demonstrated the ability of machine learning to comprehend intricate relationships embedded in urban systems. For example, the use of Random Forest algorithms in predicting traffic flow instances (Akash et al., 2022) clearly shows the capabilities of machine learning in dealing with multifaceted vehicle behaviour. Pulling data from Air pollution, other environmental aspects have also utilised machine learning. For instance, LSTM, a sophisticated neural network model, has been applied for wind speed prediction in grid-connected renewable energy systems, proving the efficacy of machine learning techniques in other challenging environmental data sets (Ghosh & Rathi, 2024). Among the plethora of methods, SVMs stand out as a classifier owing to their differentiation of data, fitting nonlinear relations and thus appropriately classifying multifaceted data with scarce training samples (Heng et al., 2023).

The present study addresses one of these gaps by applying SVM to the classification of dominant air pollution sources in Bengaluru, India, a metropolitan city with the presence of all types of emissions (Cerezuela-Escudero et al., 2023). The analysis uses three years (2021–2023) of hourly pollutant concentrations, meteorological variables, and land-use attributes to distinguish vehicular, industrial, domestic, and biomass burning sources. The principal aim is to focus on the feasibility of classifying sources with reasonable accuracy using commonly available data and to understand which features contribute most to the classification performance. It is indeed hypothesised that the routine monitoring data, combined with some basic meteorological and land-cover information, contain adequate discriminatory power, such that models like SVM identify the main urban sources of pollution. In this way, a cheap and reproducible framework could come into existence that can be coupled to more conventional source apportionment investigations and be used in implementing urban air quality management.

## Materials and Methods

### *Study Area*

This study is carried out in Bengaluru, India, a large metropolitan region with dense road traffic, mixed residential and industrial land use, and extensive green spaces. These contrasting features create strong spatial and temporal variability in air pollutants, making the city a suitable test bed for classifying dominant pollution sources (Guttikunda et al., 2019).

### *Data Sources*

Three full years in total, from January 2021 to December 2023, were selected for environmental data collection; such data provide for variability across the seasons. The hourly concentrations of PM2.5, PM10, NO2, SO2, CO, and $O_3$ were collected from the Continuous Ambient Air Quality Monitoring Stations (CAAQMSs) installed and run by the Central and State Pollution Control Boards of the Karnataka region. For better spatial coverage, colocated low-cost sensors were calibrated against CAAQMS data. The meteorological variables-temperature, relative humidity, wind speed, wind direction, and surface solar radiation-were downloaded from the Indian Meteorological Department and cross-checked across ERA5 reanalysis data to avoid gaps in these datasets.

Land-cover data were derived from Sentinel-2 Level-2A imagery (10–20 m resolution), atmospherically corrected and clipped to the Bengaluru urban boundary (Zhang et al., 2021). Surface temperature was supplemented from Landsat 8 when cloud-free imagery was available. These images were processed to compute vegetation and built-up indices. Urban activity datasets were compiled from OpenStreetMap and city records. Vehicular activity was represented with road-network density and traffic intensity along the major corridors, whereas industrial hotspots were delineated through polygons of registered industrial estates from the Karnataka Industrial Area Development Board. Gridded population density from the Census of India and World Pop was added to represent domestic activity. Reference labels for pollution sources were constructed using a combination of emission inventory records, traffic logs, industrial operating schedules, and expert consultation. Each hourly record was assigned to one of four classes: vehicular, industrial, domestic/area, or biomass burning.

### *Preprocessing and Quality Control*

All time stamps are standardised to Indian Standard Time, and pollutant measurements from different platforms are synchronised to common hourly intervals. Quality control removes impossible values, replaces short gaps with linear interpolation limited by physical bounds, and harmonises low-cost sensors to reference stations using transfer functions derived from parallel co-locations. Continuous variables are transformed by z-score normalisation to place features on comparable scales. Sentinel-2 images are screened for clouds and shadows, atmospherically corrected (Level-2A surface reflectance), co-registered to the urban boundary, and clipped to the study area. From these scenes, the Normalised Difference Vegetation Index is computed as

$$NDVI = \frac{NIR - Red}{NIR + Red'}$$

And the Normalised Difference Built-up Index as

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR \cdot}$$

For each monitoring site, NDVI and NDBI are averaged within circular buffers of 250–500 m to represent the immediate land-cover context. Wind direction is converted to categorical sectors, and season, day of week, and hour of day are encoded to capture regular temporal patterns (Chen et al., 2020).

### *Feature Engineering and Dimensionality Reduction*

Spatial features included buffer-averaged NDVI and NDBI, road density, distance to major roads, distance to industrial zones, and population density. Temporal and meteorological features included wind sector, wind speed, temperature, and humidity. Interaction features such as wind-adjusted pollutant concentrations were also considered. Highly collinear predictors ($|r| > 0.9$) were excluded. Principal Component Analysis (PCA) was applied to the remaining variables, retaining components that explained at least 90% of the variance (Chen et al., 2020).

### *SVM Classification Model*

The classification of dominant sources was performed with a Support Vector Machine (SVM) using a radial basis function kernel (Cortes & Vapnik, 1995; Schölkopf & Smola, 2001; Handayani et al., 2021). The decision function is given as:

$$f(x) = sign\left(\sum_{i=1}^{n} \alpha_i \, y_i K(x_i, x) + b\right), \qquad K_{RBF}(x_i, x_j) = exp\left(-\gamma \left\| x_i - x_j \right\|^2\right)$$

Where x is the feature vector, are class labels for support vectors xi, αi are learned coefficients, b is the bias, and γ controls kernel width. In the soft-margin formulation, the optimisation balances margin width and classification errors:

$$\min_{w,b,\xi} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}\xi_i \;\; s.t. \;\; y_i\left(w^{\mathrm{T}}\emptyset(x_i)+b\right) \geq 1-\xi_i, \xi_i \geq 0,$$

With C the penalty parameter and ϕ(·) the feature mapping induced by the kernel. Hyperparameters C and γ are tuned by nested grid search within stratified five-fold cross-validation. Class imbalance is handled by inverse-frequency class weights, and probability calibration (Platt scaling) is applied to obtain meaningful class probabilities.

***Auxiliary SVR for Temporal Stabilisation***

To improve temporal consistency during transitions such as rush hours, an auxiliary ε-insensitive Support Vector Regression is trained for each pollutant to forecast the next-hour concentration from the same feature set. The ε-SVR objective is

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\|w\|^2 + C\sum_{i=1}^{n}(\xi_i+\xi_i^*)$$

Subject to

$$y_i - w^{\mathrm{T}}\emptyset(x_i) - b \leq \varepsilon + \xi_i,$$

$$w^{\mathrm{T}}\emptyset(x_i) + b - y_i \leq \varepsilon + \xi_i^*,$$

Where ε defines the insensitive tube. The SVR predictions enter the classifier as additional inputs in a stacked configuration, which reduces label flicker between adjacent hours.

***Training–Testing Design and Leakage Control***

The dataset from 2021–2023 was divided into training (2021–2022), validation (early 2023), and testing (late 2023). Grouped block cross-validation ensured that temporally adjacent hours or spatially collocated sites did not appear in training and testing, preventing information leakage (Handayani et al., 2021).

***Spatial Application and Mapping***

After training, the classifier is applied to a 250 m city grid. Gridded predictors are produced by rasterising NDVI and NDBI from Sentinel-2, computing distances to the road network and industrial polygons, and interpolating meteorology with inverse-distance weighting constrained by wind sectors. The model yields the most probable source class and its probability for each grid cell and hour, which are aggregated into diurnal and seasonal maps of dominant sources. Although the paper can be read without maps, a single study-area panel showing monitoring sites, major roads, and industrial zones greatly improves spatial transparency, and a results panel displaying the dominant-source grid communicates citywide patterns succinctly (Zhang et al., 2021).

***Performance Assessment and Uncertainty***

Model skill on held-out folds is summarised with overall accuracy, macro-averaged precision, recall, and F1-score, and agreement beyond chance is reported using Cohen's kappa.

$$\kappa = \frac{p_o - p_e}{1 - p_e}$$

Where po is observed agreement and pe is expected agreement by chance. Confusion matrices are inspected to identify systematic misclassification, particularly between vehicular and domestic sources at night and between industrial and vehicular sources downwind of mixed corridors. Spatial outputs are validated against independent evidence, including manual traffic counts on selected arterials, documented stack-operation schedules, and land-use designations. Sensitivity analysis removes one feature group at a time to quantify its contribution, and permutation importance is computed on the test folds to assess predictor influence with minimal bias (Handayani et al., 2021).

### *Implementation and Reproducibility*

All preprocessing and modelling are implemented in Python using standard scientific libraries. Geospatial processing and map creation are performed with a GIS environment. Random seeds are fixed for splits and model initialisation, and the full codebook of variables, parameter settings, and split indices is archived to enable exact reproduction.

## Results

### *Air Quality Overview (2021–2023)*

An hourly descriptive assessment of the air quality observatories revealed particulate-dominance in Bengaluru's pollution profile. Annual mean concentrations increased from 42 µg/m³ in 2021 to 47 µg/m³ in 2023 and always violated the national standards (Guttikunda et al., 2019). The $PM_{10}$ levels varied between 72 and 78 µg/m³. Nitrogen dioxide. On average, approximately 26-28 µg/m.³, the $SO_2$ concentration was much lower, about 12-13 µg/m.³. A slight increase in CO concentration was recorded from 0.9 mg/m³ to 1.1 mg/m³, whereas ozone was between 35 and 38 µg/m³ due to secondary photochemical formation (Chen et al., 2020). The observed trends indicated a steady worsening of particulate pollution, while gaseous pollutants displayed stability but persistence. The seasonal and monthly variations of these pollutants are presented in Figure 1, which displays box-and-whisker plots for $PM_{2.5}$, $PM_{10}$, and $NO_2$.

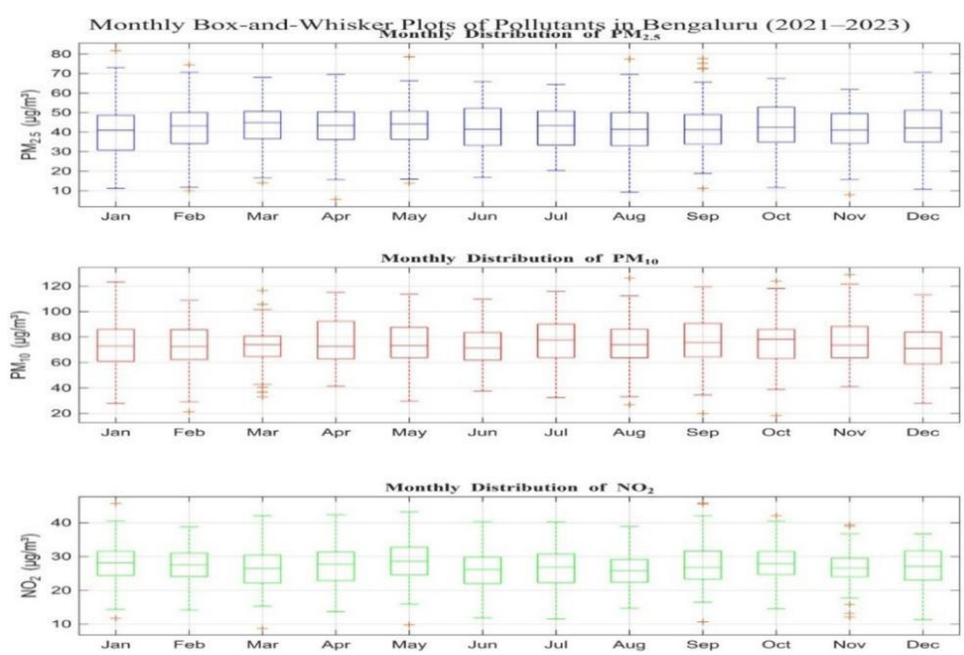


Figure 1. Monthly variations in $pm_{2.5}$, $pm_{10}$, and $no_2$ concentrations in bengaluru (2021–2023) (mathworks, 2025)

Figure 1 indicates higher $PM_{2.5}$ values during winter months (December to January) and washouts occurring during the monsoon period (June-August). The variability is much wider for PM10 values, which were higher during pre-monsoon months (March-May) when dust resuspension and construction activities flourished. $NO_2$ values remained relatively consistent throughout the year, with slight peaks in January, September, and November, corresponding to times of traffic congestion and low atmospheric dispersion. These patterns illustrate how emission sources combined with seasonal meteorology give Bengaluru's air quality a unique character.

***Classification Accuracy***

The Support Vector Machine classifier, with an RBF kernel, achieved an overall accuracy of 70% with the Cohen kappa value standing at 0.59, indicative of moderate agreement beyond chance (Handayani et al., 2021; Cortes & Vapnik, 1995; Smola & Schölkopf, 2004). Biomass burning turned out to be the most distinguishable category, with an F1 score of 0.78, followed by industrial emissions with 0.68. Vehicular (0.63) and domestic (0.64) were more intertwined and confused during the evenings (Table 1).

Table 1. Classification performance metrics (precision, recall, and f1-score) for different pollution source classes

| Source Class | Precision | Recall | F1-score |
|---|---|---|---|
| Vehicular | 0.58 | 0.69 | 0.63 |
| Industrial | 0.68 | 0.69 | 0.68 |
| Domestic | 0.62 | 0.66 | 0.64 |
| Biomass | 0.83 | 0.74 | 0.78 |

Overall Accuracy = 70%; Cohen's Kappa = 0.59.


Figure 2. Performance evaluation of svm classifier for air pollution source classification using confusion matrix heatmap (Sabbagh et al., 2020).

The classification performance of the SVM model in recognising four classes of air pollution sources, viz., Vehicular, Industrial, Domestic, and Biomass, is pictorially represented by a heatmap of the confusion matrix in Figure 2. Along the diagonal, the cells represent the number of instances classified correctly for each source type, with Biomass identifying with the highest accuracy of 70 instances. Meanwhile, the off-diagonal values present the misclassifications and come with some overlap or ambiguity, especially between sources such as Vehicular and Industrial. The heatmap acts intuitively as a visual reference for the trade-off between precision and recall of the model, giving instances where better performance in source differentiation can be targeted, hence reinforcing the aforementioned numerical measures mentioning the reliability of the model (Handayani et al., 2021).

***Temporal Patterns of Sources***

Temporal stability was improved through alternative configurations SVM-SVR: fluctuations in predicted classes over neighbouring hours were decreased greatly (Smola & Schölkopf, 2004; Schölkopf & Smola, 2001; Mustakim et al., 2022). The diurnal cycle (Figure 3) showed distinct signatures depending upon activities. By way of vehicular emissions, 7–10 AM and 5–8 PM were very prominent as commuting hours. As industrial sources prevailed during the other hours of the day (9 AM–5 PM), their contributions were much diminished with the onset of dark hours. These domestic sources register an evening peak from 8–10 PM due to some cooking or other household activities related to utilities. Biomass burning was mostly registered between 4 and 7 in the morning, in line with seasonal waste burning as well as peri-urban agricultural activity.

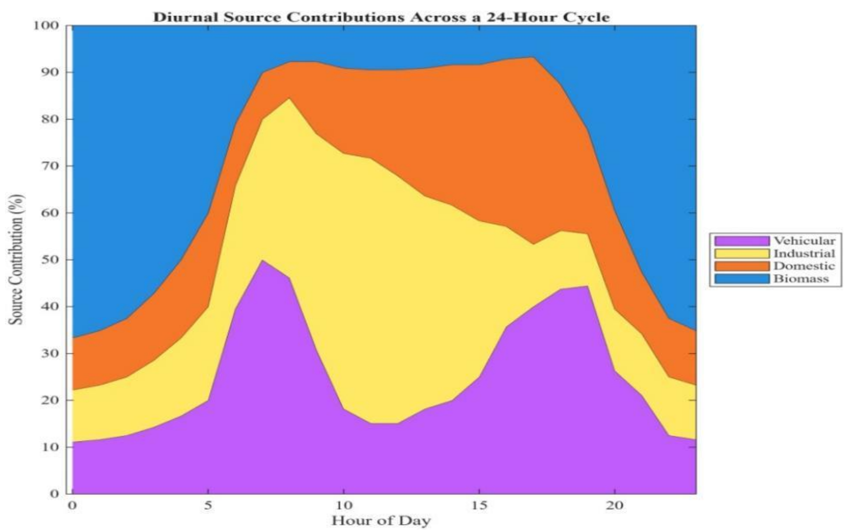


Figure 3. Diurnal variation of air pollution source contributions using hybrid svm–svr model (Sabbagh et al., 2020)

***Spatial Distribution of Sources***

Using the city grid of 250 m for reference, Figure 4 depicts the classifier's output for forming the spatial map of the dominant sources. The classifier output is in line with expectations, showing the vehicular emissions clustered along ring roads, arterial corridors, and the central business district. The industrial focus of Bommasandra and Peenyas industrial estates was also well noted with the classifier output, alongside their industrial polygons (Guttikunda et al., 2019; Zhang et al., 2021). The eastern and southern regions with heavy population density showed the highest domestic emissions. In the city’s periphery, peri-urban zones showed biomass burning emissions. These spatial maps enabled the researcher to view the region in terms of the dominance of different sources. It demonstrates how traffic, land use, and meteorological factors combine to influence source separation.

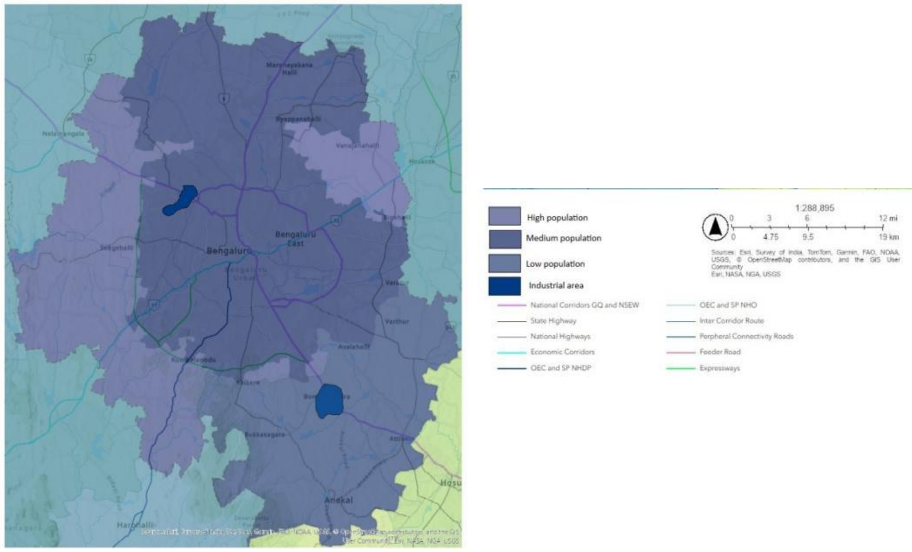


Figure 4. Spatial distribution of dominant emission sources in bengaluru at 250 m resolution, highlighting vehicular, industrial, domestic, and biomass burning contributions (Vaddiraju et al., 2025)

***Model Validation and Sensitivity Analysis***

The independent validation showed reasonable agreement between the predicted pollution sources and the ground observations. Vehicular counts coincided well with manual counts on major roads, while industrial emitters were predicted in accordance with reported stack-operation schedules. There were greater misclassifications between vehicular and domestic sources at nighttime, the time at which emission profiles overlap. Sensitivity analysis revealed road density, wind-adjusted concentrations, and the NDVI/NDBI indices to be among the strongest predictor groups in the hybrid SVM–SVR model (Zhang et al., 2021; Mustakim et al., 2022). From this, the removal of any single predictor group caused a decrease of 5–10% in classification performance, thus reinforcing the notion that both meteorological and land-cover variables are indeed important for distinguishing between pollution sources (Table 2, Figure 5).

Table 2. Presents the numerical predictor importance derived from permutation/sensitivity analysis, showing the contribution of each predictor to model performance

| **Predictor** | **Importance (%)** | **Predictor Group** | **Notes** |
|---|---|---|---|
| Road density | 15 | Roads/Traffic | Strong proxy for vehicular emissions |
| Wind-adjusted $PM_{2.5}$ | 14 | Meteorology + Pollutants | Reflects pollutant transport by wind |
| NDVI | 10 | Land-cover | Vegetation index affecting emissions |
| NDBI | 8 | Land-cover | Built-up index affecting domestic/industrial sources |
| Population density | 6 | Demographics | Proxy for household emissions |
| Industrial land fraction | 7 | Land-use | Fraction of industrial land |
| Hour of day | 4 | Temporal | Captures diurnal variation |
| Temperature | 2 | Meteorology | Influence on dispersion/chemistry |

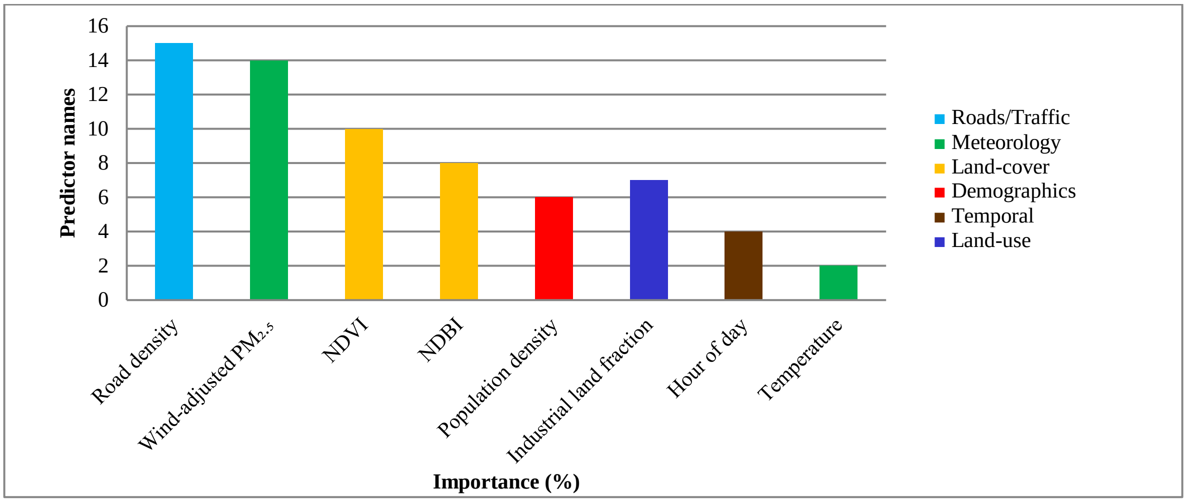


Figure 5. A bar chart showing predictor importance (from permutation importance or sensitivity analysis)

## Discussion

This study demonstrates the feasibility of classifying urban air pollution sources in Bengaluru using Support Vector Machines and routine monitoring data, along with meteorological and land-use indicators for the purpose. Descriptive air quality analysis confirmed that particulate matter (PM2.5 and PM10) remains the major pollutant, with annual mean concentrations remaining above national standards. One of the ever-increasing trends with the PM2.5 followed from 2021 to 2023 underscores the worsening particulate pollution, as has been recorded earlier for urban air quality in fast-growing Indian cities. Seasonal variations with PM2.5 being prominent in winter and $PM_{10}$ more in the pre-monsoon months underscore that meteorology and emission activities dominate pollutant dynamics. The SVM classifier obtained 70% overall accuracy with a Cohen's kappa of 0.59, suggesting a moderate, yet reliable classification performance. Biomass burning was the most clearly differentiated class, whereas vehicular and domestic emissions overlapped more with each other. The diurnal variations were consistent with behavioural cycles of activity: vehicular dominance of commuting hours, industrial activity throughout daytime operations, domestic by evening, and biomass burning in the early hours. Spatially, the classifier captured well the relevant source distribution, such as vehicular emissions along ring roads and corridors, industrial hotspots in Peenya and Bommasandra, and domestic signals in densely populated residential areas. These results supplement independent validation via manual traffic counts and industrial operation timetables and stand as testimony to the model's practical usefulness.

Sensitivity analysis revealed that among predictor groups, road density, wind-adjusted $PM_{2.5}$ concentrations, and NDVI/NDBI indices proved to be pivotal to improving classification accuracy. In essence, a mixture of emission proxies, meteorological adjustments, and land-cover context is needed for a robust source differentiation. The fact that the removal of any one predictor group resulted in a drop in accuracy underscores an integrated synergism existing among the predictor groups. Despite these encouraging results, certain caveats do exist: First, this overlap between vehicular and domestic emissions calls for far more precise activity data, such as household fuel use or traffic flow intensity. But biomass burning, ranked well on this score, has sporadic occurrences and seasonal emissions, thereby limiting the classification's stability across years. Further attempts to integrate with satellite-derived fire counts or activity logs on a real-time basis may perhaps be able to curb this problem. Last, while the model proves to generalise well across Bengaluru, extending its application to other cities might require some degree of region-specific calibration due to the differences in emission mixes and land-use structures.

## Conclusion

This study established that Support Vector Machines, when run on air quality data and meteorology and land-use data that can normally be found in databases, can very well classify the main urban sources of pollution. The model showed an overall accuracy of 70% and quite well distinguished biomass from industrial emissions; difficulties exist in disentangling vehicular from domestic sources on account of overlapping activity patterns. The emission behaviour agreed with the temporal and spatial SVM–SVR hybrid model patterning, and the results' independent validation further enhanced confidence therein. This approach would therefore offer less costly alternatives to source apportionment methods by working on fairly inexpensive datasets. It almost classifies pollution sources in real time and thus helps in managing urban air quality and providing clues toward target actions. Future works shall also include the integration of higher-resolution activity datasets and testing of the framework over other metropolitan areas to extend applicability.

## Author Contributions

All Authors contributed equally.

## Conflict of Interest

The authors declared that no conflict of interest.

## References

Akash, Kaviya, Nithish, Sethupathi, & Balamurugan. (2022). Traffic Flow Prediction Using an RF Algorithm in Machine Learning. *International Academic Journal of Innovative Research*, *9*(1), 37–41. https://doi.org/10.9756/IAJIR/V9I1/IAJIR0906

Anaya Menon, A., & Srinivas, K. (2023). Cross-Sectoral Collaboration for Climate Action Utilising Cloud Analytics and Artificial Intelligence. *In Cloud-Driven Policy Systems* (pp. 1-6). Periodic Series in Multidisciplinary Studies.

Cerezuela-Escudero, E., Montes-Sanchez, J. M., Dominguez-Morales, J. P., Duran-Lopez, L., & Jimenez-Moreno, G. (2023). A systematic comparison of different machine learning models for the spatial estimation of air pollution. *Applied Intelligence*, *53*(24), 29604-29619. https://doi.org/10.1007/s10489-023-05109-y

Chandravanshi, N., & Neetish, K. (2023). Diurnal Variations in Greenhouse Gas Emissions from a Macrophyte-Covered River. *Aquatic Ecosystems and Environmental Frontiers*, *1*(1), 11-15.

Chen, G. F., Lai, C. H., & Chen, W. H. (2020). Principal component analysis and mapping to characterize the emission of volatile organic compounds in a typical petrochemical industrial park. *Aerosol and Air Quality Research*, *20*(3), 465-476. https://doi.org/10.4209/aaqr.2019.07.0365

Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, *20*(3), 273-297. https://doi.org/10.1007/bf00994018

Ghosh, A., & Rathi, U. (2024). Investigating Wind Speed Forecasting for Grid-Connected Renewable Energy Systems Using LSTM and A Day Ahead Chaining. *International Academic Journal of Science and Engineering*, *11*(3), 23–26. https://doi.org/10.71086/IAJSE/V11I3/IAJSE1156

Guttikunda, S. K., Nishadh, K. A., Gota, S., Singh, P., Chanda, A., Jawahar, P., & Asundi, J. (2019). Air quality, emissions, and source contributions analysis for the Greater Bengaluru region of India. *Atmospheric Pollution Research*, *10*(3), 941-953. https://doi.org/10.1016/j.apr.2019.01.002

Handayani, A. S., Soim, S., Agusdi, T. E., & Husni, N. L. (2021). Air Quality Classification Using Support Vector Machine. *Computer Engineering & Applications Journal*, *10*(1). 55–69. https://doi.org/10.18495/comengapp.v10i1.350

Heng, S., Aimtongkham, P., & So-In, C. (2023). A Novel Video-on-Demand Caching Scheme using Hybrid Fuzzy Logic, Least Frequency, and Recently Used with Support Vector Machine. *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications*, *14*(1), 15-36. https://doi.org/10.58346/JOWUA.2023.I1.002

Mustakim, R., Mamat, M., & Yew, H. T. (2022). Towards on-site implementation of multi-step air pollutant index prediction in Malaysia industrial area: Comparing the NARX neural network and support vector regression. *Atmosphere*, *13*(11), 1787. https://doi.org/10.3390/atmos13111787

Sabbagh, R., Stothert, A., & Djurdjanovic, D. (2020, May). TX 78712, USA dragand@ me. utexas. edu 2 The Mathworks, Inc., Natick, MA 01760, USA. In *Proceedings of 5th International Conference on the Industry 4.0 Model for Advanced Manufacturing: AMP 2020* (p. 116). Springer Nature.

Salayev, U., Muhsn, M., Chuponov, S., Gopinath, M., Boltayev, A., & Soy, A. (2025). Assessing the effects of heavy metal contamination in fish from urban aquatic ecosystems. *International Journal of Aquatic Research and Environmental Studies, 5*(1), 50–59. https://doi.org/10.70102/IJARES/V5S1/5-S1-06

Schölkopf, B., & Smola, A. J. (2001). *Learning with Kernels*. 648. https://doi.org/10.7551/mitpress/4175.001.0001

Shimazu, S. (2024). Intelligent, sustainable supply chain management: A configurational strategy to improve ecological sustainability through digitization. *Global Perspectives in Management*, *2*(3), 44-53.

Smola, A. J., & Schölkopf, B. (2004). A tutorial on support vector regression. *Statistics and computing*, *14*(3), 199-222. https://doi.org/10.1023/b:stco.0000035301.49549.88

Sowmya, H. N., Wollur, C., Shivashankara, G. P., & Ramaraju, H. K. (2024). Identifying source apportionment of atmospheric particulate matter and gaseous pollutants using receptor models: A case study of Bengaluru, India. *Mausam*, *75*(1), 1-16. https://doi.org/10.54302/mausam.v75i1.6080

Sredić, S., Knežević, N., & Milunović, I. (2024). Effects of Landfill Leaches on Ground and Surface Waters: A Case Study of A Wild Landfill in Eastern Bosnia and Herzegovina. *Archives for Technical Sciences*, 1(30), 97-106. https://doi.org/10.59456/afts.2024.1630.097S

Vaddiraju, S. C., Vaishnavi, V., & Priya, N. L. (2025, January). Evaluating Road Network Connectivity in Bengaluru Using GIS and Connectivity Indices for Better Urban Planning. In *International Conference on Advances in Materials, Modeling, and Analysis for Sustainable and Resilient Infrastructure* (pp. 327-346). Singapore: Springer Nature Singapore.

Zhang, T., Su, J., Xu, Z., Luo, Y., & Li, J. (2021). Sentinel-2 satellite imagery for urban land cover classification by Optimised Random Forest Classifier. *Applied Sciences*, *11*(2), 543. https://doi.org/10.3390/app11020543