



RESEARCH ARTICLE

OPEN ACCESS

AGENT- AI BASED FRAMEWORK FOR GREEN ROUTE: ENHANCING TRAVEL USING BEST PATHFINDING WITH SSFS AND SE-TDT-SVM THROUGH AQI-AWARE ROUTING

Alamelu R M¹, Christy Eunaicy J I², Jayapratha C³, Govindaprabhu G B⁴, Mahalakshmi Priya R⁵

¹Assistant Professor, Department of Computer Science, Sri sarada Niketan College for Women, Amaravathi pudur, Karaikudi, Tamilnadu, India.

²Assistant Professor, Department of Computer Application, Thiagarajar College, Madurai, Tamilnadu, India.

³Professor and Head of the Department, Department of Computer Science, Karpaga Vinayaga College of Engineering and Technology, Madhuranthagam, Tamilnadu, India.

^{4,5}Research Scholar, Madurai Kamaraj University (MKU), Madurai, Tamilnadu, India.

¹<https://orcid.org/0000-0003-2157-8795>, ²<https://orcid.org/0000-0003-3371-7422>, ³<https://orcid.org/0000-0002-3365-1956>,

⁴<https://orcid.org/0000-0002-2297-3597>, ⁵<https://orcid.org/0000-0002-0591-3744>

E-mail: alamuinfor@gmail.com, prabhupri.pp@gmail.com, g.p.priya2707@gmail.com

ARTICLE INFO

Article History

Received on: March 08, 2026

Reviewed on: April 08, 2026

Accepted on: July 01, 2026

Published on: July 31, 2026

Keywords:

AQI Prediction,
Dynamic Path Prediction,
AQI Pre-Processing,
SSFS,
SE-TDT-SVM,
AQI-Aware Routing.

ABSTRACT

Air pollution poses significant health risks in urban environments. It necessitating accurate Air Quality Index (AQI) prediction and route planning. In urban environments, it is becoming increasingly important. The importance of managing and mitigating air quality has been highlighted in a various recent study. This work presents a novel approach integrating real-time air quality data with Agent based AI techniques. It predicts AQI levels and recommend low-pollution travel routes in major Indian cities. It proposes a SALP Swarm based Feature Selection (SSFS) algorithm to identify key pollutant indicators. This also proposed a Stacked Ensemble of Tuned Decision Tree and Support Vector Machine (SE-TDT-SVM) for AQI prediction. AQI-Aware Routing (AAR) algorithm optimizes path selection based on both air quality and distance metrics. The SE-TDT-SVM model achieved a prediction accuracy of 94.5% with an error rate of 5.5% on the feature-selected dataset. The AAR algorithm successfully generated multiple path options, balancing travel distance and AQI exposure. This integrated approach offers a robust solution for urban air quality management and health-conscious route planning. The potential applications in smart city initiatives and public health interventions.



Copyright ©2026 by authors and Galileo Institute of Technology and Education of the Amazon (ITEGAM). This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).

I. INTRODUCTION

A daily AQI is an index used to assess the status of the atmosphere air. AQI is used to offer individuals with knowledge about how their air quality impacts health. From the report of WHO, air pollution accounts to around 1.3 million deaths per year worldwide. Pollutants discharged into the air are responsible for various harmful effects. Besides acid rain, global warming, and photo-chemical smog, there have been other harmful significance over the past few decades [1]. The health impacts of air pollution have made the prediction and optimal path selection of the AQI increasingly in urban environments. The recent studies have highlighted the urgency of managing and mitigating air quality. The Lancet Planetary Health emphasized that air pollution is one of environmental factors in global disease burdens [2]. Approximately 4.5 million premature deaths worldwide were attributed to exposure to fine PM2.5 alone in 2019. It shows the importance of accurate AQI prediction for public health awareness. Furthermore, [3] reports that exposure to high levels of air pollution during daily commutes. It can result in increased blood pressure and reduced lung function. It illustrates the importance of real-time AQI prediction and optimal path selection to minimize exposure during travel. Indian air quality standards index pollutants according to their effects on the atmosphere within a given state or country. The atmosphere is polluted by a variety of gases. Depending on the pollutants measured, there is a different level of pollution. Each pollutant's index is calculated by computing all available data points for each region. AQIs represent air quality in very clear color representations so that a layman can easily comprehend good or bad air quality.

The air quality index can be used to identify major pollutants and locations seriously affected by those pollutants. A number of advanced techniques are used to extract knowledge from data in order to determine the heavily affected area within a region. Air pollution source and state are revealed by the results of the analysis. Models based on agent-based interactions have emerged as powerful tools for simulating complex systems. In the context of air quality and routing, agents can represent vehicles, pedestrians, cyclists, or stationary entities like sensors. By modeling individual behaviors, ABMs enable the study of emergent patterns from interactions among agents and between agents and dynamic environmental conditions, such as air quality variations. For example, an ABM could simulate commuter preferences in route selection depending on variables like travel time, AQI levels, and weather conditions. Additionally, ABMs can predict how large-scale environmental changes, such as increased pollution in an urban center or green corridors, will affect travel patterns. By using simulations, researchers and policymakers can forecast the effectiveness of interventions and ensure their effectiveness.

Objectives

1. To develop an advanced AQI prediction model (SE-TDT-SVM). It integrates real-time data with ML techniques for accurate air quality across major Indian cities.
2. To create a novel Agent Based AQI-Aware Routing (AAR) that optimizes path selection based on both air quality and distance metrics.
3. To implement a comprehensive system that associates AQI prediction with dynamic path selection for urban navigation.

Several dimensions of these findings are relevant in practice. By applying these routes, air pollution is mitigated and public health is safeguarded. In urban India, poor air quality causes numerous health complications. Using this innovative model, elevated pollution levels can be identified. In turn, enhancing atmospheric conditions through the planning of green corridors. Consequently, healthcare expenditures decrease as pollution-related illnesses decline. An urban environment that is more sustainable and economically viable. By incorporating environmental considerations into navigation, the study exemplifies how to integrate smart technologies within urban planning. Such interdisciplinary efforts can foster sustainable and health-promoting urban development. Nguyen et al. [4] predicted PM_{2.5}, which could affect air quality. A genetic algorithm (GA) was used in the research work to predict PM_{2.5}. It assisted feature selection and removed outliers to enhance prediction accuracy. LSTM-based encoder-decoder models were able to predict accurately PM_{2.5} concentrations. As a result of collecting data for forecasting air pollution, Suganya and Meyyappan [5] proposed guidelines for avoiding and regulating air pollution. In two phases, the research was conducted. Preprocessing the selected dataset was the first step. The second phase involved forecasting air quality levels. Using sensor data, Ditsuhi et al. [6] reviewed the air pollution forecasting studies in smart cities using machine learning algorithms. Instead of using simple machine learning techniques, innovative and sophisticated techniques were applied to conclude the research. Prediction of air quality was influenced by weather conditions, spatial characteristics, and temporal characteristics. Multiple weather stations and air quality metering stations were used by Trenchevski et al. [7] to predict potential improvements in air pollution. Over a number of years, air quality, pollution, and weather data, including temperature, humidity, and wind characteristics, were collected. Work involved features extraction, redundancy removal, training multiple regression models, and hyper parameter optimization. Mahajan et al. [8] developed the CAR (Clean Air Routing) algorithm to build a healthy optimal route suggestion system between the source and destination. The road network was navigated with the least exposure to PM_{2.5} by using a weight function. A pollution consideration may cause an alternative pathway to exceed the minimum distance. Zahmatkesh et al. [9] suggest that AI-based systems select routes with lower AQI values. Optimizing the path from origin to destination uses Dijkstra's algorithm.

Using machine learning algorithms, the study determines. The comprehensive atmospheric data to estimate PM_{2.5} concentrations in urban areas. Using Auto Regression, PM_{2.5} levels will be predicted. Many studies focus on single pollutants, such as PM_{2.5} [10]. AQI prediction and path selection need to consider multiple pollutants simultaneously. The five models were compared using RMSE and MAPE. Both locations showed better performance with bidirectional GRU. IoT-based air quality, humidity, and temperature alert system proposed by P. Tripathi [11]. Particulate matter (PM₁₀ and PM_{2.5}) and carbon monoxide (CO) are measured in real time. Real-time data updates are provided by IoT. It generates alerts based on temperature, humidity, and air quality data using fuzzy logic. A study by Chandar et al. [12] used historical AQI data of decade to determine levels, cyclicity, and seasonality. The models used included Random Forest (RF), Support Vector Regression (SVR), XGBoost (XGB), Neural Network (NN), and Long Short-Term Memory (LSTM). Model performance is measured by Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). XGBoost outperforms other models with MAE scores of 7.991 (7-day) and 8.126 (30-day). By combining the strengths of numerous base learners, Singh et al. [13] improve prediction accuracy. Based on decomposed time series components, hybrid ensemble learning uses base learners. Using ensemble aggregation, the final forecast is generated. Results show that the proposed approach outperforms traditional forecasting models and individual base learners for the targeted pollutants. According to Priyanka et al. [14], air quality is monitored and a path with the least amount of pollution is then suggested. In the model, air pollution was instantly reduced for vulnerable populations. In order to calculate the shortest path between source and destination, the Dijkstra algorithm was used. The air quality index was predicted using environmental monitoring data and meteorological data by Zhou et al. [15]. Using a non-linear autoregressive neural network model, an air quality assessment model was enhanced. Using other standard machine learning algorithms, the model was demonstrated to be feasible and to perform well. Khan et al. [16] investigated bidirectional variations of recurrent neural networks, long short-term memories, and gated recurrent units. Various atmospheric and air pollution variables are represented by 15 variables. The following points are the research gap of the work:

- **Integration of Real-time and Predictive Models:** There is currently a lack of integration between real-time AQI data and predictive models for path selection in studies like Nguyen et al. [4] and Zhou et al. [15]. For route planning, it is necessary to combine up-to-date air quality information with forecasted data.
- **Multi-pollutant Approach:** Many studies focus on single pollutants, such as PM_{2.5} [10]. AQI prediction and path selection need to consider multiple pollutants simultaneously.

- **Dynamic Path Adjustment:** It is lacking a system that can dynamically adjust paths based on changing air quality conditions in real time, despite the routing algorithms proposed by Mahajan et al. [8]
- **Advanced Machine Learning Techniques:** It would be interesting to explore more advanced ensemble methods or hybrid models combining the strengths of different algorithms to improve ML.
- **Long-term Forecasting and Trend Analysis:** According to Chandar et al., long-term AQI predictions are necessary for better urban planning and policymaking.

II. METHODOLOGY

In Figure 1, the work is intended to monitor and optimize the air quality index forecast. To determine the pollutant levels over time, real-time air pollution data from India's states is acquired. By analyzing and predicting air quality, AQI based on historical and live data, a model can be developed to circumvent the prediction problem. In the work, low AQI will be used as the basis for path selection.

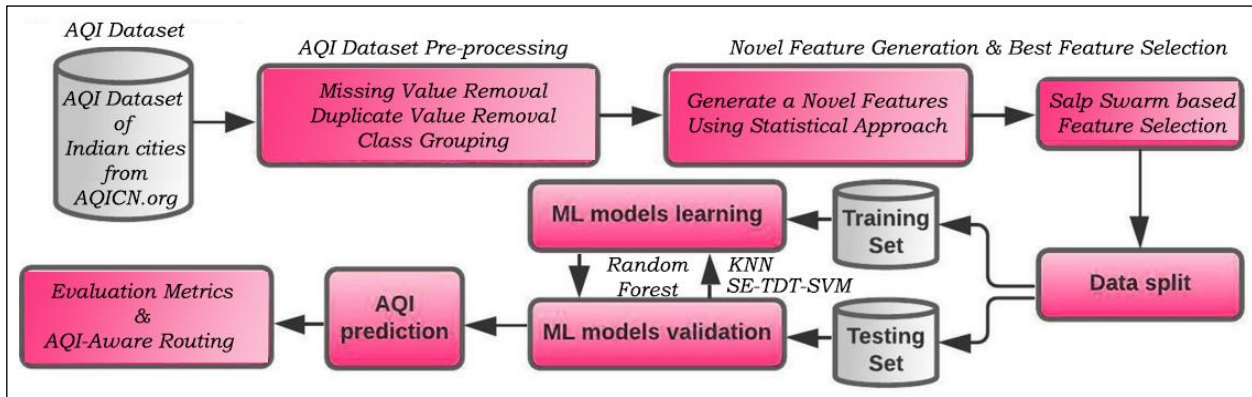


Figure 1: Proposed Framework.

Source: Authors, (2026).

II.1 DATA GATHERING AND CLEANING

Data was collected from aqicn.org, an aggregator of data contributed by government agencies, private companies, and the community. Various locations globally, including major Indian cities, can access this platform for comprehensive and up-to-date air quality information. A 2-year period of historical measurements of air quality is included in the dataset, which covers the period 2022 to 2024. In the study, major Indian cities are examined, such as Jaipur, Chennai, Kochi, etc. Urban air quality conditions are represented geographically by numerous locations on aqicn.org. AQI prediction models require appropriate data in a suitable format. It consists of attributes such as Particulate Matter (PM10), Particulate Matter (PM2.5), Ozone (O3), Sulphur dioxide (SO2), Nitrogen dioxide (NO2), and Carbon monoxide (CO) that are necessary for calculating an area's air quality index. Air Quality Index China (AQICN) data is a foundational component for the AQI-Aware Routing (AAR) algorithm and Agent-Based AI framework. Data from multiple sources, including government agencies, private companies, and community-driven sensor networks, is aggregated by AQICN. With this rich dataset, the algorithm can evaluate pollutant concentrations such as PM2.5, PM10, NO2, SO2, CO, and O3. Its diversity ensures broad coverage, which is crucial for modeling air quality trends and providing real-time AQI predictions. Table 1 displays a sample of data on air pollutants. The AQI dataset for the specific date and city contains air pollutants. In accordance with the 2016 US-EPA Air Quality Index (AQI) standard, there are six levels of air pollution: 0-50 (Good), 51-100 (Moderate), 101-150 (Unhealthy for Sensitive Groups), 151-200 (Unhealthy), 201-300 (Very Unhealthy), and 300+ (Hazardous).

Table 1: Sample Data.

Date	PM _{2.5}	PM ₁₀	O ₃	NO ₂	SO ₂	CO	City
10/1/2024	89	49	23	18	6	8	Jaipur
10/2/2024	91	58	19	16	7	5	Jaipur
9/1/2024	54	39	16	7	8	4	Jaipur
7/2/2024	64	32	18	11	6	6	Jaipur
9/3/2024	51	68	5	2	1	5	Chennai
8/4/2024	57	52	5	1	2	5	Chennai
9/5/2024	48	28	5	2	2	5	Chennai

Source: Authors, (2026).

Air quality is satisfactory with little or no risk at Good levels. The moderate levels may pose a moderate health concern to those who are very sensitive. Sensitive groups can be affected by levels categorized as Unhealthy. Those with sensitive needs may be more affected by unhealthy levels. Hazardous levels pose serious health risks to everyone, while Very Unhealthy levels carry health warnings for the entire population. Depending on the severity of the AQI, cautionary statements advise children, adults, and those with respiratory conditions to limit or avoid outdoor activities [17]. AQI dataset size is 119198 x 8.

Pre-processing removes 45934 missing and 0 duplicate data, resulting in a total of 45934 missing and 0 duplicate data. Before pre-processing, the AQI dataset is 119198 x 8; after eliminating missing and duplicate values, it becomes 73264 x 8. There are six levels of pollution in the AQI: Good, Satisfactory, Moderate, Poor, Very Poor, and Severe. According to ambient concentrations of pollutants, the pollution levels are classified based on potential health impacts. Six pollutants (PM₁₀, PM_{2.5}, NO₂, SO₂, CO, O₃) are sub-indexed and health breakpoints are developed. The calculation of AQI is based on a sub-index. Equation 1 calculates AQI based on the worst sub-index.

$$I = \frac{I_{x_{hi}} - I_{x_{lo}}}{Cn_{hi} - Cn_{lo}} (Cp - Cn_{lo}) + I_{x_{lo}}$$

II.2 NOVEL FEATURE GENERATION

i.**Ratio of coarse to fine particulate matter (ROC):** This ratio indicates the proportion of larger particles (PM₁₀, diameter ≤ 10 μm) to smaller particles (PM_{2.5}, diameter ≤ 2.5 μm) in the air. Higher ratios indicate particles, whereas lower ratios indicate finer particles, which are generally considered more harmful.

$$ROC = PM_{10}/PM_{2.5}$$

ii.**Ratio of nitrogen dioxide to sulfur dioxide (RNDSD):** This compares levels of two common gaseous pollutants. Generally, if the ratio is higher, combustion sources (e.g., vehicle emissions) are contributing more pollution, while a lower ratio is indicating coal combustion or industrial pollution.

$$RNDSD = NO_2 / SO_2$$

iii.**Ratio of ozone to nitrogen dioxide (ROND):** This ratio provides insight into the formation of photochemical smog. A reaction between NO_x (including NO₂) and volatile organic compounds produces ozone. A higher ratio could indicate more advanced smog formation or more intense sunlight.

$$ROND = O_3 / NO_2$$

iv.**Ratio of carbon monoxide to fine particulate matter (RCM):** The comparison is between a gaseous pollutant (CO) and fine particles. An elevated ratio might indicate more incomplete combustion (e.g., vehicle emissions in cold weather), while a lowered ratio might indicate more secondary particle formation.

$$RCM = CO / PM_{2.5}$$

v.**Ratio of sulfur dioxide to coarse particulate matter (RSD):** A gaseous pollutant (SO₂) is compared to a larger particle (PM₁₀). Higher ratios indicate more industrial or coal-burning pollution than dust.

$$RSD = SO_2 / PM_{10}$$

vi.**Total Pollutants:** This sum gives a rough idea of pollution levels, but it does not account for the different units or health impacts of different pollutants.

$$Total\ Pollutants = PM_{2.5} + PM_{10} + O_3 + NO_2 + SO_2 + CO$$

vii.**Average Pollutants:** It provides a metric for averaging all measured pollutants, but has the same limitations as the metric for total pollutants.

$$Average\ Pollutants = (PM_{2.5} + PM_{10} + O_3 + NO_2 + SO_2 + CO) / 6$$

viii.**Interaction of particulate matter and ozone (IOMO):** It shows that these two pollutants can cause respiratory issues in synergistic ways.

$$IOMO = PM_{2.5} \times O_3$$

ix.**Interaction of nitrogen dioxide and carbon monoxide (IONCM):** The product emphasizes the presence of two pollutants commonly found in vehicle emissions in combination.

$$IONCM = NO_2 \times CO$$

x.**Interaction of particulate matter and sulfur dioxide (IOPMSD):** The product demonstrates the combination of pollutants found in industrial emissions and dust.

$$ONCM = PM_{10} \times SO_2$$

xi.**Combined of PM, ozone, and nitrogen dioxide (CPMOND):** In this triple product, there is potential for complex interactions among these three pollutants.

$$PM_{2.5} \times O_3 \times NO_2$$

xii.**Max Pollutant** This identifies the highest level among all measured pollutants, potentially highlighting the most problematic pollutant.

$$Max\ Pollutants = \max(PM_{2.5}, PM_{10}, O_3, NO_2, SO_2, CO)$$

xiii.**Minimum pollutant level:** This identifies the lowest level among all measured pollutants.

$$Min\ Pollutants = \min(PM_{2.5}, PM_{10}, O_3, NO_2, SO_2, CO)$$

xiv.**Range of pollutant levels:** By calculating the spread between the highest and lowest levels of pollution, we can see how pollution levels vary.

$$Range\ Pollutants = Max\ Pollutants - Min\ Pollutants$$

xv. **Variance Pollutants:** This measures the spread of pollutant levels around their mean (μ), providing insight into how much the levels fluctuate.

$$\text{Variance Pollutants} = \frac{1}{5} ((PM_{2.5} - \mu)^2) + (PM_{10} - \mu)^2 + (O_3 - \mu)^2 + (NO_2 - \mu)^2 + (SO_2 - \mu)^2 + (CO - \mu)^2),$$

Where μ is the mean of the pollutants.

xvi. **Normalized_PM2.5:** By normalizing the PM2.5 level against its maximum value, it is easier to compare different time periods or locations.

$$\text{Normalized_PM2.5} = \max(PM2.5)/PM2.5$$

II.3 SALP SWARM BASED FEATURE SELECTION ALGORITHM

By using feature selection, AQI data is evaluated to identify and eliminate inappropriate, redundant, and unnecessary attributes that are not relevant to a predictive model, which can reduce its accuracy. Based on the swarming behavior of salps in the ocean, the Salp Swarm Algorithm (SSA) is a bio-inspired optimization technique. A salp's collective behavior in navigation and foraging is mimicked in order to solve various optimization problems, both single- and multi-objective. SSAs operate by creating a population of salps, each representing a potential solution. Each salp's position is updated iteratively based on its distance from the leader salp and the location of the food source. Position updates for the i -th salp are defined as follows:

$$x_i^j = \begin{cases} F_j + c_1 ((ub_j - lb_j)c_2 + lb_j), & \text{if } i = 1 \\ \frac{1}{2}(x_i^j + x_{i-1}^j), & \text{if } i > 1 \end{cases}$$

Where F_j is the position of the food source, ub_j and lb_j are the upper and lower bounds of the j -th dimension, c_1 and c_2 are random numbers uniformly distributed between 0 and 1. In model construction, feature selection refers to selecting a subset of relevant features. A salp's position can be treated as a possible subset of features for this purpose by using SSA. In order to select features, first initialize a population of salps randomly within the search space. In each salp, the position in each dimension corresponds to a specific feature. Determine whether a salp is fit by evaluating its classification accuracy, often based on a machine learning model trained with the selected features. Calculate the position of each salp using the SSA mathematical model. Follower salps adjust their positions based on the leader's positioning, while the leader moves towards the best solution. Converge the algorithm (that is, stabilize the salp positions or reach the maximum number of iterations). Optimal subset of features is represented by the best salp at the end of optimization. Feature quality is evaluated using the fitness function. For feature selection, the following fitness function is commonly used:

$$\text{Fitness} = \alpha \cdot \text{ErrorRate} + (1 - \alpha) \cdot \frac{|S|}{|F|}$$

Where Error Rate is the classification error rate, S is the number of selected features, F is the total number of features, and α is the weight factor that balances accuracy and feature reduction. Swarm Algorithms are robust and efficient ways of selecting features in machine learning. SSA improves the performance of machine learning models while reducing computational complexity by leveraging the swarming behavior of salps.

Input: Feature set (AQI Data), Population size N , Max iterations T

Output: Best subset of features

Algorithm: SSA_Feature_Selection

```

1. Initialize population  $X$  with  $N$  salps
2. Define fitness function
3. Evaluate fitness of initial population
4. Set food source to the best solution in initial population
5. For  $t = 1$  to  $T$  do
6. For  $i = 1$  to  $N$  do
7. If  $i == 1$  then
    Update leader salp's position using food source
Else
    Update follower salp's position using previous salp
End If
8. Convert continuous positions to binary
9. Evaluate fitness of salp
10. If salp's fitness is better than food source fitness then
    Update food source
End If
End For
End For
11. Return the best subset of features (position of food source)
End Algorithm

```

Algorithm 1: SSA_Feature_Selection.

Source: Authors, (2026).

As binary strings encoded as binary salps, each bit represents a potential solution (1 indicates a selected feature, 0 indicates a non-selected feature). Assign random binary strings to a population of salps to represent different subsets of features. A fitness function should be used to evaluate the fitness of each salp. If the salp selects a subset of features, the fitness function could be the accuracy of a machine learning model trained on the subset. Based on the SSA equations, update the salp positions. A binary string is flipped to explore different subsets of features in the context of feature selection. As a result of a predetermined number of iterations, select the salp with the highest fitness value as the optimal subset of features. Utilize the optimal subset of features identified by the SSA to train the final machine learning model.

II.4 STACKED ENSEMBLE OF TUNED DECISION TREE AND SVM MODEL

In the classification techniques, historical data and live data are used to predict a given pollution level. For AQI prediction, the best feature selection algorithm is determined by the feature importance score and selection time. SE-TDT-SVM, a stacked ensemble of Tuned Decision Trees and SVMs, is developed for AQI prediction. As each node contains a classifier, the stacked ensemble model resembles a tree. Various machine learning models are combined into an ensemble model. Based on the output of the base classifier, the meta classifier predicts the final class label. An architecture based on stacked layers is proposed in the work. In Decision Trees (DTs) internal nodes represent attributes being tested, branches represent the results of the tests, and leaf nodes represent class labels. To learn a tree from the DAQI dataset, an attribute value test can be used to recursively separate subsets as shown in equation 6. Recursion ends when all subsets at a node have the same target value. Sorting from the root to a leaf node allows it to classify instances.

$$DT_{root} \xleftarrow{\text{best feature}} x_i, \text{ where } x_i \in S_{Train} \forall i = 1, 2, \dots, n$$

AQI attributes specified by the top node of the tree are tested by the leaves to determine the pollution level based on attribute selection measures. By building the finest boundary that isolates the n-dimensional space into different class labels, the SVM Prediction (Meta) Model algorithm classifies a new instance in the future. A hyperplane denotes this boundary. A hyperplane is created by finding the vector points that play a vital role. The extreme points are called support vectors. AQI prediction is performed using a nonlinear SVM.

$$S_{Test} = \{(x_1, y_1), (x_2, y_2) \dots (x_n, y_n)\} | x_i \in S_{Test}, y_i \in \{\text{very poor, severe, satisfactory, poor, moderate, good}\}$$

Through STrain and STest, the SVM discovers a distinct hyperplane in the feature set in order to classify the pollution level in Eq7. The DT and SVM models are stacked to produce an efficient classification model that combines individual decision-making. When training with a large dataset, the SVM takes a long time to process and train. The main drawback of SVMs is their long training time. Training huge datasets requires an effective Decision Tree-based data selection and Support Vector Classification approach. A decision tree was built during the training phase to reduce the size of the SVM training dataset. With DT, the number of data dimensions is reduced and features are selected and generated. A stacked prediction model (DT and SVM) is used to classify pollution levels using the AQI dataset. The training set STrain is divided into n sections and taken from the selected features set FB of the AQI dataset. In the nth section, the model makes predictions based on the fitted base classifier model DT. The training set STrain is divided into sections according to this technique. SVM prediction models are trained using the same technique for training and testing sets STrain and STest. The training set is divided into two parts: training and validation. The predictions from the validation set are used to create a new model by combining them with those from the testing set STest. The meta model classifier SVM is then initialized with the kernel function and the parameters are tuned for class label validation. To predict the AQI class label, the STest is used. The parameters of the stacked classifier are shown in Table 2.

Table 2: Parameters of SE-TDT-SVM.

Parameters	Default Value	Tuned Value
Decision Tree (Level 0)		
Max Depth	1	4
Min Samples Leaf	1	5
Min Weight Fraction Leaf	0.0	0.3
Splitter	Random	Best
SVM (Level 1)		
C	1.0	1.0
Coef0	0.1	0.5
ToL	0.0001	0.01
Decision function Shape	OVR	OVO
Kernel	Linear	RBF
Degree	1	4
Gamma	Auto	Scale

Source: Authors, (2026).

This algorithm combines the strengths of multiple machine learning models to predict AQI using the Stacked Ensemble of Tuned Decision Trees and Support Vector Machines (SE-TDT-SVM).

The Decision Tree (DT) excels at handling non-linear relationships and identifying significant features, while the Support Vector Machine (SVM) excels at classification tasks involving complex decision boundaries. With SE-TDT-SVM, DTs preprocess and simplify the dataset by reducing its dimensionality, while SVMs refine the predictions by learning optimal hyperplanes. With high-dimensional datasets like AQICN, ensemble design enhances model accuracy and reduces overfitting. Furthermore, SE-TDT-SVM is parameter tuned to ensure each base model operates at its optimal level. On the feature-selected dataset, this hybrid approach consistently outperforms single algorithms. In dynamic routing systems, SE-TDT-SVM provides a computationally efficient, scalable solution for AQI prediction.

Input: AQI dataset DAQI, Training set S_{Train} , Testing set S_{Test}

Output: Predicted AQI class labels for S_{Test}

Algorithm: SE-TDT-SVM

1. Split the AQI dataset DAQI into training set S_{Train} and testing set S_{Test} .
2. Compute the feature importance scores.
3. Select the best features F_B based on the importance scores.
4. Divide S_{Train} into n sections.
5. For each section i from 1 to n :
 - a. Use $n-1$ sections for training the Decision Tree (DT).
 - b. Predict the AQI class labels for the remaining i -th section.
6. Collect predictions for all sections to form the intermediate dataset $SDT-Pred$.
7. Divide S_{Train} into training and validation sets.
8. Use $SDT-Pred$ and validation predictions as features to train the SVM model.
9. Tune the SVM parameters using the validation set.
10. Use the trained Decision Tree to reduce the dimensions of S_{Test} .
11. Apply the SVM model to predict the AQI class labels for S_{Test} .
12. Output the predicted AQI class labels for S_{Test} .

End Algorithm

Algorithm 2: SE-TDT-SVM.

Source: Authors, (2026).

II.5 DYNAMIC PATH PREDICTION WITH AQI-AWARE ROUTING (DPPAAR) USING AGENT BASED AI

An AQI-Aware Routing (AAR) algorithm for optimal path suggestion is presented in this work, which incorporates both air quality and distance metrics. AAR involves several phases, as shown in Figure 2. For a particular location, the model involves phases such as location profiling and data collection, discovering intermediate cities, computing distances and directions, generating multiple paths, and generating shortest distances with low AQI. [

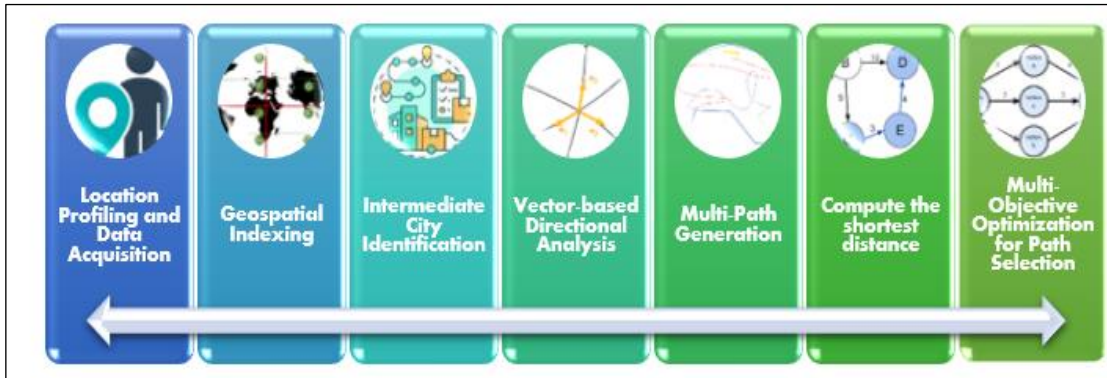


Figure 2: Dynamic Path Selection Process.

Source: Authors, (2026).

During the first phase, 'Location Profiling and Data Acquisition,' let $L = [l_1, l_2, \dots, l_n]$ represent the set of locations, where each l_i represents (ID, NAME, CITY, STATE, LAT, LON). We utilized the AQICN.org API to obtain real-time AQI data in JSON format, which is then parsed and stored as $AQI(l_i, t) = f(PM_{2.5}, PM_{10}, NO_2, SO_2, CO, O_3)$. The next step is to implement a novel R-tree spatial indexing structure R to efficiently organize cities according to their geographical coordinates. Path finding can be significantly optimized with $O(\log n)$ time complexity in nearest neighbor queries. Based on the latitude point information of the specific station, cities are ordered based on their state. By ordering the cities, the optimal path can be recommended. The next phase is an Intermediate City Identification. Given a source s and destination d , we define the set of potential intermediate cities I as: $I = \{l_i \in L \mid \min(\text{lat}(s), \text{lat}(d)) \leq \text{lat}(l_i) \leq \max(\text{lat}(s), \text{lat}(d)) \wedge \min(\text{lon}(s), \text{lon}(d)) \leq \text{lon}(l_i) \leq \max(\text{lon}(s), \text{lon}(d))\}$. Using the ordered location dataset, the intermediate cities between the user's source and destination are computed. A location dataset is then searched and indexed for the source and destination cities. For a possible path, intermediate cities are acquired between the source and destination cities. Vector-based Directional Analysis is the next phase, which uses vectors to determine the direction between cities. For any two cities a and b , we calculate their direction vector as follows: $v(a, b) = (\text{lon}(b) - \text{lon}(a), \text{lat}(b) - \text{lat}(a))$. The angle θ between $\rightarrow v(a, b)$ and the positive x-axis determines the cardinal and intercardinal directions. There are four main directions - east, west, north, and south. On all paths, directions are named the same way.

It is not always possible to locate a place accurately using only the four directions. There are clearly sub-directions and scaled directions. Using zone points, this work calculates the distance between two points. Based on the route, the direction-finding offers possible paths between the locations. Multi-Path Generation Using Modified A* Algorithm, this work proposes a modified A* algorithm that incorporates AQI levels into the heuristic function. For each node n , we define: $f(n) = g(n) + h(n) + \alpha \cdot q(n)$. An AQI level at n is related to the distance from the start node to 'n', and the estimated distance from n to the goal is related to the AQI level at n , and the weighting factor is related to the AQI level at n . In order to compute the best path, a multi-path must be generated for the specific location (Tirupati to Bhopal). A dissimilar path is determined by grouping together different directions for a given location after estimating different directions. Therefore, for a specific source and destination, numerous paths exist. The best path is determined by analyzing these various paths. The next step is to calculate the shortest distance between the multipaths, the distance from the source city to the destination city through the intermediate cities should also be determined. As a result, a distance is computed for each path in the multipath set. By computing the distance, the shortest path among the multipaths can be found. To identify the less polluting path for traveling, the AQI for the paths is also calculated after estimating the distance among the multipath cities. This work introduces an AQI exposure metric E for a path $P = \{l_1, l_2, \dots, l_n\}$: $E(P) = \sum_{i=1}^{k-1} (AQI(l_i) + AQI(l_{i+1})) / 2 \cdot d(l_i, l_{i+1})$. AQI levels and distance traveled are both taken into account in this metric. Finally, a Multi-Objective Optimization for Path Selection recommendation is computed based on less AQI and shortest distance between source and destination cities. In order to select the final path, we formulate it as a multi-objective optimization problem, where $F(P) = (D(P), E(P))$, subject to $P \in$ feasible paths from s to d . In this case, $D(P)$ represents the total distance of path P . An optimization is used to find non-dominated solutions, thereby offering users a range of paths that are both distance-efficient and AQI-efficient.

Input: Set of Locations

Output: Return Shortest Path

Algorithm: AQI-Aware Routing (AAR) Algorithm

1. Define $L = \{l_1, l_2, \dots, l_n\}$ with each location l_i represented by $(ID, NAME, CITY, STATE, LAT, LON)$.
2. Acquire AQI data from aqicn.org, store it as: $AQI(l_i, t) = f(PM2.5, PM10, NO2, SO2, CO, O3)$
3. Build an R-tree spatial indexing structure $\{R\}$ to efficiently organize cities by their coordinates.
4. Order the cities based on state using the latitude point of the station information.
5. Define the source city s and destination city d .
6. Identify the set of potential intermediate cities I as: $I = \{l_i \in L \mid \min(lat(s), lat(d)) \leq lat(l_i) \leq \max(lat(s), lat(d)) \text{ and } \min(lon(s), lon(d)) \leq lon(l_i) \leq \max(lon(s), lon(d))\}$
7. For any two cities a and b , calculate the direction vector: $v(a, b) = (lon(b) - lon(a), lat(b) - lat(a))$
8. Determine the cardinal and inter-cardinal directions based on the angle θ between $v(a, b)$ and the positive.
9. Use a modified A* algorithm incorporating AQI levels into the heuristic function: $f(n) = g(n) + h(n) + \alpha \cdot q(n)$
10. Generate multiple paths from the source to the destination, grouping directions to obtain paths.
11. Compute the AQI exposure metric E for a path $P = \{l_1, l_2, \dots, l_n\}$:

$$E(P) = \sum_{i=1}^{k-1} \frac{AQI(l_i) + AQI(l_{i+1})}{2} \cdot d(l_i, l_{i+1})$$
13. Formulate the final path selection as a multi-objective optimization problem: $\text{minimize } F(P) = (D(P), E(P)) \text{ subject to } P \in \text{feasible paths from } s \text{ to } d$
14. Return Shortest Path

End Algorithm

Algorithm 3: AQI-Aware Routing (AAR) Algorithm.

Source: Authors, (2026).

By combining real-time air quality data with predictive AQI modeling, AQI-Aware Routing (AAR) stands out. Among the unique aspects:

1. Integration of Dynamic Air Quality Index: AQI predictions are constantly updated in AAR. It allows real-time route adjustments.
2. An algorithm that considers multiple air pollutants at the same time provides a more comprehensive analysis of air quality.

Utilizing a new weighting system for air pollution and travel distance, AAR maintains a balance between these objectives. Through Agent-Based AI, AQI-Aware Routing transforms pathfinding into an adaptive, dynamic process. AI-driven autonomous agents make path selection decisions using real-time AQI data and predictive models like SE-TDT-SVM. Multiple factors are balanced by agents to evaluate path options:

- **AQI Exposure Metric (E(P)):** Agents calculate potential exposure to pollutants on each route by leveraging the AAR algorithm's AQI exposure metric,

$$E(P) = \sum_{i=1}^{k-1} \frac{AQI(l_i) + AQI(l_{i+1})}{2} \cdot d(l_i, l_{i+1}),$$

where l_i represents locations along the route

- **Travel Distance (D(P)):** AAR uses a multi-objective optimization approach to assess route lengths and tradeoff between minimal distance and low AQI exposure.
- **Real-Time Adjustments:** Agents adaptively adjust path recommendations based on AQI conditions or traffic dynamics as they are changed.

Consider a delivery agent traveling from source s to destination d . It uses AAR to identify intermediate cities $I=\{l_i \in L\}$, calculates directional vectors, and evaluates paths using real-time AQI predictions. The agent recalculates the optimal path using updated data whenever conditions change, such as an increase in AQI near an industrial zone.

III. RESULTS AND DISCUSSION

The experimental simulation environment is Windows -10, a CPU of 2.80GHz, 8GB of memory, and Python programming language. Python was used to implement SSA-based feature selection, utilizing the NumPy library for efficient numerical computations. For stability, the algorithm was run 30 times: 15 out of 22 original features were selected; PM2.5, NO2, and O3 were selected most frequently; Average computational time: 0.9 seconds. With SSA-based feature selection, the model's predictive performance was significantly improved while the feature space was reduced. As a result, we were able to increase the efficiency of our AQI prediction model as well as gain insight into the most influential pollutants for AQI determination in the studied Indian cities. AQI Prediction computes the following evaluation metrics.

III.1 ACCURACY AND ERROR RATE

A machine learning algorithm's accuracy can be determined based on the number of AQI samples that are classified appropriately in relation to the total number of AQI samples [18]. The proposed classifier's accuracy and error rate are shown in Table 3.

Table 3: Evaluation Metrics.

Datasets	Classifiers	Accuracy	Error Rate	Precision	Recall	F1-Score
Processed Dataset	SE-TDT-SVM	90.2	9.8	91.6	92.7	92.14
	Random Forest	82.3	17.7	80.1	83.6	81.81
	KNN	79.4	20.6	77.8	79.1	78.44
Feature Selected Dataset	SE-TDT-SVM	94.5	5.5	90.8	93.2	91.98
	Random Forest	86.5	13.5	84.6	87.3	85.92
	KNN	81.9	18.1	80.6	81.1	80.84

Source: Authors, (2026).

The accuracy result is shown in figure 3. Error rates (ERRs) are calculated by dividing the number of incorrect AQI predictions by the total number of datasets analyzed.

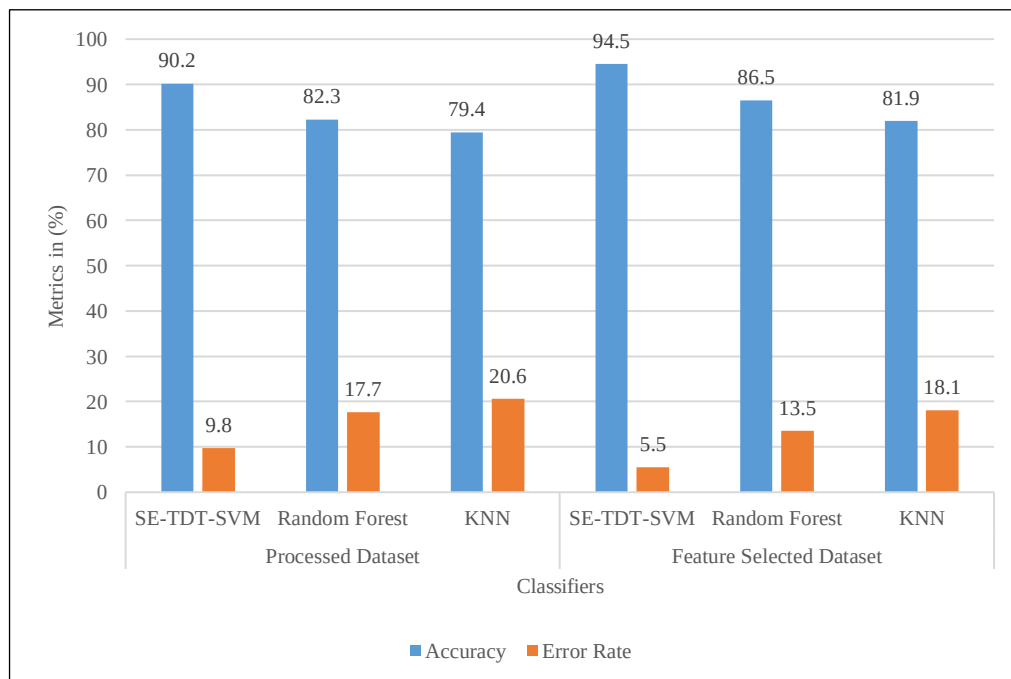


Figure 3: Accuracy and Error Rate.

Source: Authors, (2026).

In figure 4, three classifiers are compared using two datasets: Processed and Feature Selected. For each classifier, it displays Precision, Recall, and F1-Score. A nearly perfect score was achieved with SE-TDT-SVM across both datasets. In the Feature Selected dataset, Random Forest and KNN also show good performance. According to the figure, all three metrics generally show higher values after feature selection, illustrating how feature selection impacts classifier performance.

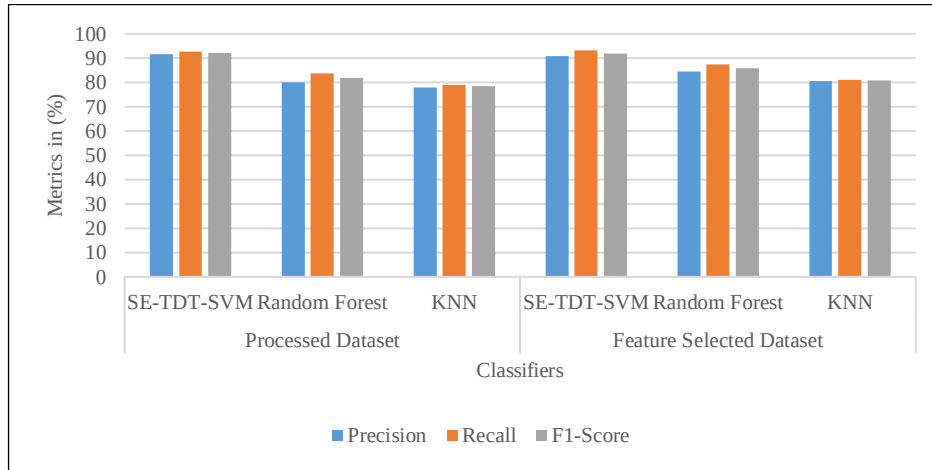


Figure 4: Precision, Recall and F1-score.
Source: Authors, (2026).

SSA-TDT-SVM integration benefits computation in several ways:

1. **Reduced Complexity:** SSA eliminates redundant and irrelevant features, reducing dimensionality of the dataset [19], which reduces SE-TDT-SVM computational load.
2. **Improved Training Efficiency:** SE-TDT-SVM reduces training time by focusing on the most relevant data in the dataset, rather than processing raw, high-dimensional data [20].
3. **Enhanced Predictive Accuracy:** SE-TDT-SVM's ensemble nature ensures robust predictions, while SSA's optimization enhances generalizability.
4. **Scalability:** The algorithms are computationally efficient, enabling real-time processing of large datasets.

AQICN API data was also used to validate the model's predictions. By doing this, the system was tested for its ability to adjust predictions and routing recommendations dynamically. The real-time validation scenarios were:

- Industrial activities or weather changes spike pollution.
- Urban AQI levels vary during the day and night.

Actual AQI values were compared with predictions. Model accuracy was high, with deviations generally below 5%, demonstrating practical applicability. In accordance with the Predicted AQI, the optimal path and shortest distance are recommended to the user. Example: Tirupati is the source city, Bhopal is the destination city. The evaluation of path-1 for the given source and destination city with distance and AQI value is shown in Figure 5.

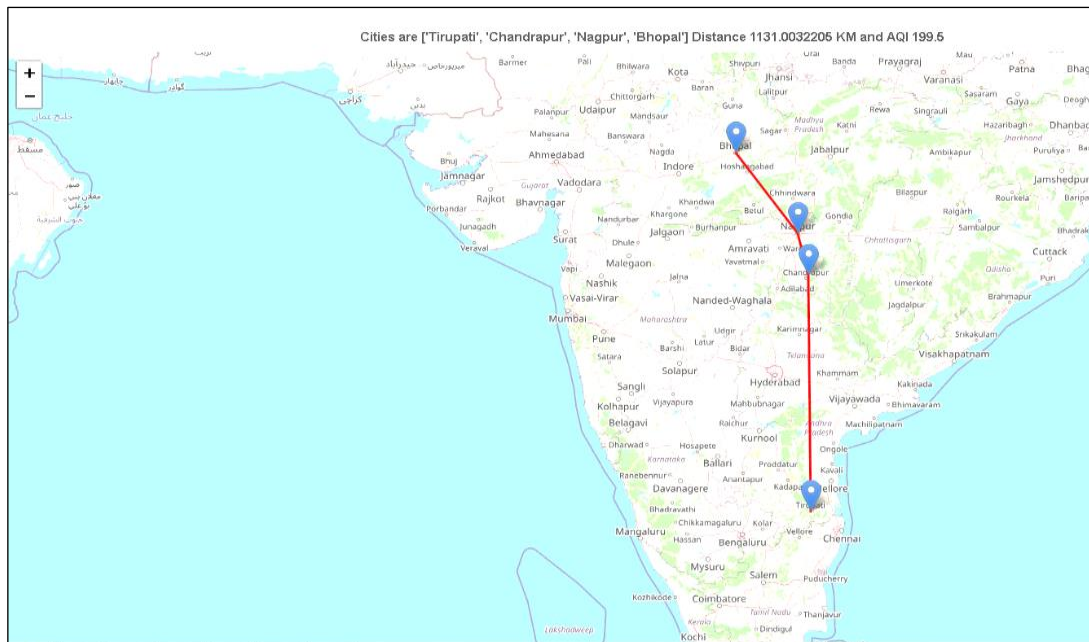


Figure 5: Shortest Path-1.
Source: Authors, (2026).

Figure 6 shows the possible path-2 for the source city (Tirupati) to the destination city (Bhopal) with the evaluated value of distance and AQI.

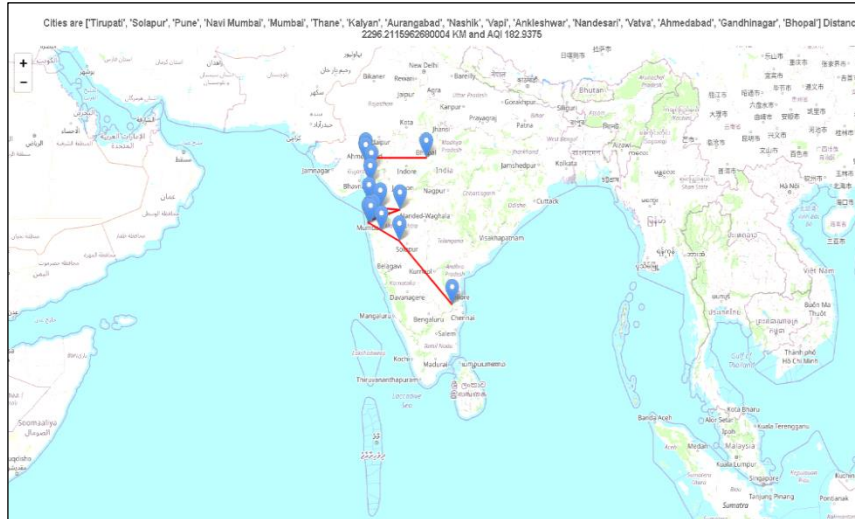


Figure 6: Shortest Path-2.
Source: Authors, (2026).

The possible path-3 for the source and destination city based on distance and AQI is shown in Figure 7. Advantages of this work are, real-time AQI data are used from AQICN.org and integrated with a predictive model (SE-TDT-SVM) for AQI forecasting.

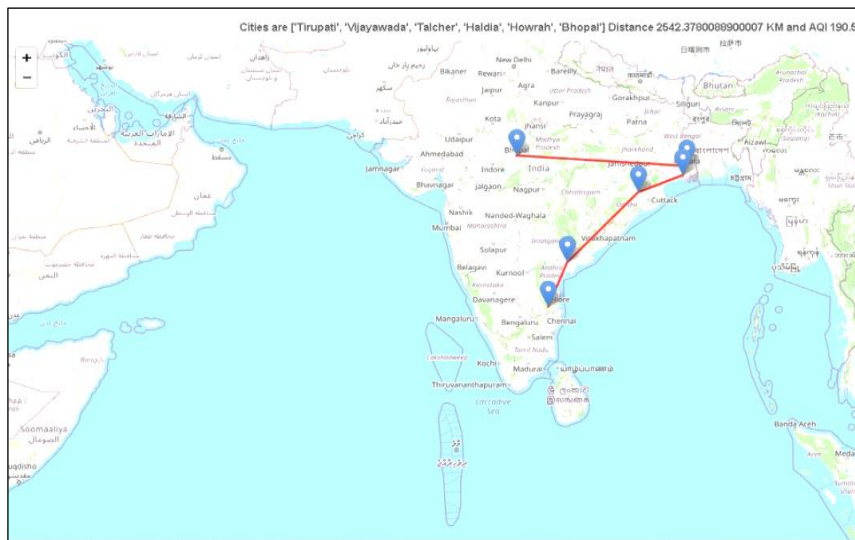


Figure 7: Shortest Path-3.
Source: Authors, (2026).

With this combination, path selection can be made more accurate and up-to-date, bridging the gap between real-time data and predictive modeling. AQI calculations consider multiple pollutants (PM_{2.5}, PM₁₀, NO₂, SO₂, CO, O₃), providing a more comprehensive assessment of air quality than single-pollutant analyses. The AQI-Aware Routing (AAR) algorithm appears to allow for dynamic path selection based on current AQI levels, which is an improvement over static routing. With the SE-TDT-SVM model, decision trees and SVMs are combined in a stacked ensemble that could potentially outperform single algorithms in AQI prediction.

III.2 COMPARATIVE ANALYSIS WITH STATE-OF-THE-ART ROUTING MODELS

Air Quality-Aware Routing (AAR) algorithm is benchmarked against several air quality-integrated routing models, showing distinct advantages in adaptability and comprehensiveness. Mahajan et al.'s and Hameed [21] Clean Air Routing (CAR) algorithm minimizes PM_{2.5} exposure using a weighted function that prioritizes air quality over distance. However, CAR lacks a multi-pollutant approach and dynamic adaptability. AAR incorporates multiple pollutants (PM_{2.5}, PM₁₀, NO₂, SO₂, CO, O₃) and leverages real-time AQI updates to enable dynamic path adjustments. This ensures AAR's responsiveness to sudden environmental changes, a key limitation of CAR. Dijkstra's algorithm and sensor data are used in Zahmatkesh et al., Zhang [22] and Zou [23] urban travel routing system. It focuses primarily on urban environments with dense sensor coverage by factoring AQI as a parameter. The method relies on static data and does not optimize for multi-objective decision-making. AQI trends are forecasted using SE-TDT-SVM, an algorithm that incorporates predictive modeling with a multi-objective optimization framework to balance travel distance and pollution exposure [24]. This IoT-enabled path recommendation system [25] and [26] uses real-time AQI updates, but is tailored to small-scale environments with limited geospatial adaptability. Using advanced spatial indexing techniques, including R-trees, the AAR algorithm exceeds this by computing routes efficiently across diverse regions. Furthermore, its agent-based structure allows for personalized recommendations, such as prioritizing health-conscious routes or optimizing logistics pathways.

Potential Real-World Applications

- i. Smart City Navigation Systems: Promoting healthier routes for pedestrians, cyclists, and drivers by integrating city-wide navigation apps.
- ii. Urban Planning and Policy Making: Assess the impact of urban development projects on air quality and assist city planners in improving air quality.
- iii. Logistics and Delivery Optimization: Air pollution exposure can be minimized by helping delivery companies plan routes.
- iv. Emergency Response Planning: During pollution-related crises, assist emergency services in planning safe routes.

IV. CONCLUSION

By combining the AQI predictions with optimal path selection, the SE-TDT-SVM model achieved an accuracy rate of 94.5% and a precision rate of 5.5% respectively. A feature-selected dataset showed KNN to outperform Random Forest (81.9%). Routes can be optimized based on real-time AQI data to improve urban health. Pollution exposure and travel efficiency are balanced, and hotspots are identified. As a result, pollution exposure is reduced for individuals with respiratory conditions. The algorithm should be improved and its applicability extended in future research. For assessing effectiveness, real-world deployment and user feedback are crucial. As a result of this integrated approach to AQI prediction and path optimization, urban air quality challenges are addressed in a meaningful way. Increasing urbanization and air quality concerns will make smart, health-focused solutions increasingly vital to promoting public health and urban sustainability. In the future, research should explore the incorporation of real-time data feeds and the possibility of deploying the solution as a mobile application for dynamic path recommendations. The results of this work illustrate the critical role that advanced machine learning techniques can play in enhancing urban air quality management and public health protection.

V. AUTHOR CONTRIBUTION

Conceptualization: Alamelu, Christy Eunaicy, Jayapratha.

Methodology: Govindaprabhu, Mahalakshmi Priya.

Investigation: Alamelu, Christy Eunaicy, Jayapratha.

Discussion of results: Alamelu, Christy Eunaicy, Jayapratha.

Writing – Original Draft: Alamelu, Christy Eunaicy, Jayapratha.

Writing – Review and Editing: Govindaprabhu, Mahalakshmi Priya.

Resources: Alamelu, Christy Eunaicy, Jayapratha.

Supervision: Alamelu, Christy Eunaicy, Jayapratha.

Approval of the final text: Alamelu R M, Christy Eunaicy J I, Jayapratha C, Govindaprabhu G B, Mahalakshmi Priya R.

VI. REFERENCES

- [1] R. Ghorani Azam, P. Riahi-Zanjani, and A. Balali-Mood, "Effects of air pollution on human health and practical measures for prevention in Iran," J. Res. Med. Sci., vol. 21, no. 1, p. 65, 2016, doi: 10.4103/1735-1995.189646.
- [2] D. Tang, Y. Zhan, and F. Yang, "A review of machine learning for modeling air quality: Overlooked but important issues," Atmos. Res., vol. 300, no. 1, pp. 1–10, 2024, doi: 10.1016/j.atmosres.2024.107261.
- [3] L. Zhang, X. Chen, and X. Xue, "Short-term personal exposure to particulate matter and lung function in healthy adults: A panel study in Beijing, China," Sci. Total Environ., vol. 89, no. 1, pp. 1–10, 2024, doi: 10.1016/j.scitotenv.2023.165080.
- [4] M. H. Nguyen, K. Nguyen, and Y. Ji, "PM2.5 prediction using genetic algorithm-based feature selection and encoder-decoder model," IEEE Access, vol. 9, pp. 57338–57350, 2021, doi: 10.1109/ACCESS.2021.3106548.
- [5] S. Suganya and T. Meyyappan, "Forecasting and prediction of air pollution levels to protect human beings from health hazards," Int. J. Sci. Technol. Res., vol. 9, no. 1, pp. 2541–2545, 2020.
- [6] D. Istephan, R. Francisco, and T. Sergio, "Air quality prediction in smart cities using machine learning technologies based on sensor data: A review," Appl. Sci., vol. 10, no. 7, pp. 1–32, 2020, doi: 10.3390/app10072401.
- [7] A. Trenchevski, M. Kalendar, and H. Gjoreski, "Prediction of air pollution concentration using weather data and regression models," in Proc. 8th Int. Conf. Appl. Innov. IT (ICAIIT), 2020, vol. 8, no. 1, pp. 55–61.
- [8] S. Mahajan, Y. S. Tang, D. Y. Wu, and T. C. Tsai, "CAR: The clean air routing algorithm for path navigation with minimal PM2.5 exposure on the move," IEEE Access, vol. 7, pp. 147373–147382, 2019, doi: 10.1109/ACCESS.2019.2946419.
- [9] H. Zahmatkesh, M. Saber, and M. Malekpour, "New method for urban travel route planning based on air pollution sensor data," Curr. World Environ., vol. 83, no. 10, pp. 1–10, 2015, doi: 10.12944/CWE.10.
- [10] C. R. Aditya, R. Chandana, and D. K. Nayana, "Detection and prediction of air pollution using machine learning models," Int. J. Eng. Trends Technol., vol. 59, no. 4, pp. 204–207, 2018, doi: 10.14445/22315381/IJETT-V59P238.
- [11] P. Tripathi, A. Pal, V. Kumar, and N. K. Sharma, "Implementation and study of fuzzy-based IoT-enabled health alert system using AQI," in Proc. Int. Conf. IoT, Commun. Autom. Technol. (ICICAT), Gorakhpur, India, 2023, pp. 1–6, doi: 10.1109/ICICAT57735.2023.10263668.
- [12] B. S. Chandar, P. Rajagopalan, and P. Ranganathan, "Short-term AQI forecasts using machine/deep learning models for San Francisco, CA," in Proc. IEEE 13th Annu. Comput. Commun. Workshop Conf. (CCWC), Las Vegas, NV, USA, 2023, pp. 402–411, doi: 10.1109/CCWC57344.2023.10099064.
- [13] A. Singh, P. Soni, and D. Krishnan, "Ensemble learning with hybrid modeling for multivariate AQI, PM2.5, and PM10 forecasting in Mumbai," in Proc. Int. Conf. Comput. Sci. Emerg. Technol. (CSET), Bangalore, India, 2022, pp. 1–7, doi: 10.1109/CSET58993.2023.10346745.

- [14] P. Phadtare, S. K. Singh, S. Karanjkar, S. P. Kulkarni, and S. R. Hiray, "Intelligent navigation system using air quality index," *Int. Res. J. Eng. Technol. (IRJET)*, vol. 5, no. 5, pp. 772–775, 2018.
- [15] Y. Zhou, S. De, E. Gawa, C. Perera, and K. Moessner, "Data-driven air quality characterization for urban environments: A case study," *IEEE Access*, vol. 6, pp. 77996–78006, 2018, doi: 10.1109/ACCESS.2018.2884647.
- [16] D. Khan, A. B. Patankar, R. Motwani, and A. Ailsinghani, "Investigating efficacy of RNN and its variants for AQI forecasting," in *Proc. Int. Conf. Adv. Comput., Commun. Inf. Technol. (ICAICCIT)*, Faridabad, India, 2023, pp. 1303–1307, doi: 10.1109/ICAICCIT60255.2023.10465781.
- [17] G. B. Govindaprabhu and M. Sumathi, "Ethnomedicine of indigenous communities: Tamil traditional medicinal plants leaf detection using deep learning models," *Procedia Comput. Sci.*, vol. 235, no. 1, pp. 1135–1144, 2024, doi: 10.1016/j.procs.2024.04.108.
- [18] G. B. Govindaprabhu and M. Sumathi, "Safeguarding humans from attacks using AI-enabled (DQN) wild animal identification system," *Int. Res. J. Multidiscip. Scope (IRJMS)*, vol. 5, no. 3, pp. 285–302, 2024, doi: 10.47857/irjms.2024.v05i03.0697.
- [19] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of air quality index using machine learning techniques: A comparative analysis," *J. Environ. Public Health*, vol. 2023, pp. 1–26, Jan. 2023, doi: 10.1155/2023/4916267.
- [20] Md. M. Islam, F. A. Jibon, M. M. Tarek, M. H. Kanchan, and S. U. P. Shakil, "A real-time dataset of air quality index monitoring using IoT and machine learning in the perspective of Bangladesh," *Data Brief*, vol. 55, p. 110578, Jun. 2024, doi: 10.1016/j.dib.2024.110578.
- [21] S. Hameed, A. Islam, K. Ahmad, S. B. Belhaouari, J. Qadir, and A. Al-Fuqaha, "Deep learning-based multimodal urban air quality prediction and traffic analytics," *Sci. Rep.*, vol. 13, no. 1, Dec. 2023, doi: 10.1038/s41598-023-49296-7.
- [22] Z. Zhang, S. Zhang, C. Chen, and J. Yuan, "A systematic survey of air quality prediction based on deep learning," *Alexandria Eng. J.*, vol. 93, pp. 128–141, Mar. 2024, doi: 10.1016/j.aej.2024.03.031.
- [23] B. Zou, S. Li, Z. Zheng, B. F. Zhan, Z. Yang, and N. Wan, "Healthier routes planning: A new method and online implementation for minimizing air pollution exposure risk," *Comput. Environ. Urban Syst.*, vol. 80, p. 101456, Jan. 2020, doi: 10.1016/j.compenvurbsys.2019.101456.
- [24] L.-N. Gao, "A short-distance healthy route planning approach," *J. Transp. Health*, vol. 24, p. 101314, Dec. 2021, doi: 10.1016/j.jth.2021.101314.
- [25] B. Sarkar, D. Mukherjee, J. Mukherjee, and S. Sinha, "Traffic optimization and route detection based on air quality and pollution level," in *Proc. IEEE 11th Int. Conf. Commun. Syst. Netw. Technol. (CSNT)*, Apr. 2024, pp. 1203–1208, doi: 10.1109/CSNT60213.2024.10545817.
- [26] Shreyank, S. Kumar V., Shashwath, V. T., and P. T. C., "Safe route prediction based on AQI and machine learning," in *Proc. Int. Conf. Netw., Multimedia Inf. Technol. (NMITCON)*, Bengaluru, India, 2023, pp. 1–7, doi: 10.1109/NMITCON58196.2023.10275929.