Transforming Landscapes: Potential Impact of Land Use Land Cover (LULC) on Air Quality

LY Major Project-A Report

Submitted in partial fulfillment of the requirements of the Degree of Bachelor of Technology in Computer Engineering

by

Palak Piyush Desai

Kapil Sunil Bhatia

Dakshita Sanjay Kolte

Supervisor
Dr. Jyoti Wadmare

Department of Computer Engineering

K. J. Somaiya Institute of Technology
An Autonomous Institute permanently affiliated to University of Mumbai Ayurvihar, Sion, Mumbai -400022

2024-25

Transforming Landscapes: Potential Impact of Land Use Land Cover (LULC) on Air Quality

LY Major Project-A Report

Submitted in partial fulfilment of the requirements of the Degree of Bachelor of Technology in Computer Engineering

by

Palak Piyush Desai (Roll No. 10)

Kapil Sunil Bhatia (Roll No. 55)

Dakshita Sanjay Kolte (Roll No. 65)

Supervisor
Dr. Jyoti Wadmare

Department of Computer Engineering

K. J. Somaiya Institute of Technology
An Autonomous Institute permanently affiliated to University of Mumbai Ayurvihar, Sion, Mumbai -400022

2024-25

CERTIFICATE

*This is to certify that the project entitled "**Transforming Landscapes: Potential Impact of Land Use Land Cover (LULC) on Air Quality**" is bonafide work **Palak Desai, Kapil Bhatia and Dakshita Kolte** submitted to the University of Mumbai in partial fulfilment of the requirement in Project, for the award of the degree of "Bachelors of Technology" in "Computer Engineering".*

____

Dr. Jyoti Wadmare
Project Guide
Department of Computer Engineering

____

Dr. Sarita Ambadekar
Head of Department
Dept. of Computer Engineering

____

Dr. Vivek Sunnapwar
Principal KJSIT

Place: Sion, Mumbai 400022
Date:

PROJECT APPROVAL FOR L. Y.

This project report entitled **"Transforming Landscapes: Potential Impact of Land Use Land Cover (LULC) on Air Quality"** by

Palak Piyush Desai (Roll No. 10)

Kapil Sunil Bhatia (Roll No. 55)

Dakshita Sanjay Kolte (Roll No. 65)

is an approved Last Year Project **in Computer Engineering**.

Examiners

1.________________

Name and Signature
External Examiner

2.________________

Name and Signature
Internal Examiner

Place: Sion, Mumbai-400022
Date:

DECLARATION

We declare that this written submission represents our ideas in our own words and where other's ideas or words have been included, we have adequately cited and referenced the sources. We also declare that we have adhered to all principles of academic honesty and integrity and have not misrepresented or fabricated or falsified any idea/data/fact/source in our submission. We understand that any violation of the above will be cause for disciplinary action by the Institute and can also evoke penal action from the sources which have thus not been properly cited or from whom proper permission has not been taken when needed.

Palak Piyush Desai________________

Kapil Sunil Bhatia_________________

Dakshita Sanjay Kolte_____________

Date:

ACKNOWLEDGEMENT

Before presenting our LY project work entitled "**Transforming Landscapes: Potential Impact of Land Use Land Cover (LULC) on Air Quality**", we would like to convey our sincere thanks to the people who guided us throughout the course for this project work.

First, we would like to express our sincere thanks to our beloved Principal **Dr. Vivek Sunnapwar** for providing various facilities to carry out this project.

We would like to express our immense gratitude towards our **Project Guide Dr. Jyoti Wadmare** for the constant encouragement, support, guidance, and mentoring at the ongoing stages of the project and report.

We would like to express our sincere thanks to our **H.O.D. Dr. Sarita Ambadekar**, for the encouragement, co-operation, and suggestions progressing stages of the report.

Finally, we would like to thank all the teaching and non-teaching staff of the college, and our friends, for their moral support rendered during the course of the reported work, and for their direct and indirect involvement in the completion of our report work, which made our endeavors fruitful.

Place: Sion, Mumbai-400022
Date:

ABSTRACT

This research examines the integration of LULC classification with air quality analysis using remote sensing data and EO methods. This study therefore evaluates the possible effects of various LULC patterns in urban as well as rural areas. Land use and cover are regarded as the main factors influencing the actual emissions and dispersion of pollutants, which further affect the overall quality of the air in a region. Utilizing the satellite images provided by Google Earth Engine (GEE), we classify high-resolution LULC maps representing urban sprawl, vegetation cover, aquatic bodies, and industrial areas, all with variable impacts on air quality. Concurrent monitoring of LULC data with air quality data, such as SO_2, NO_2, PM_{10} and $PM_{2.5}$, is used to understand how changes in the land-use pattern, especially urbanization and deforestation, affect variances in air pollution level. Through the use of machine learning techniques, work allows the prediction of the spatial distribution of air pollutants using LULC traits to identify high-risk locations that are prone to air quality degradation. The results of this study show that comprehensive sustainable land management practices are required to help alleviate the problem of air pollution. Moreover, the research highlights the need for EO technology to deliver prompt and dependable data to policy-makers and city planners, as well as environmental scientists to set up effective plans to enhance the quality of air. This paper hopes to add to the growing collection of literature on how land use affects the environment and sets urban air quality.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

Sr No.	Abbreviation	Description
1	LULC	Land Use Land Cover Classification
2	GEE	Google Earth Engine
3	AQI	Air Quality Index

I. INTRODUCTION

1.1 Problem Definition

Land use land cover classification is an essential tool to manage the environment and provide information about urban planning, spatial distribution of land cover types and human land use patterns including urban areas, agricultural land, forests, water bodies, and barren land. Examining these classifications allows us to better comprehend the complex interaction between land use, human activity, and environmental activity, particularly air quality. LULC enables the mapping and monitoring of land use and land cover changes across time, highlighting the importance of tracking land cover changes with satellite and remote sensing data.

1.2 Aim and Objective

In Maharashtra, India, two different regions that highlight changing environmental impacts of land use on air quality are Mumbai and Satara. Mumbai, India's financial center and densely populated urban area suffers from significant air pollution because of high vehicular traffic, rapid urban expansion, industrial activities, and construction. The LULC of the city is majorly categorized by built-up land and less green coverage which leads to high pollution levels with pollutants such as Sulphur Dioxide (SO_2), Nitrogen Dioxide (NO_2) and Particulate Matter ($PM_{2.5}$ and PM_{10}). In contrast to this, Satara is a rural region with agricultural cover having less industrial and construction activities, leading to lower pollution and clean air. This striking contrast between these regions make it ideal to study how LULC patterns lead to variations in air quality.

LULC classification helps to categorize and identify land cover types using machine learning and image processing techniques which are used to process satellite images. Different techniques such as Convolutional Neural Network (CNN), Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM) and their hybrid models are used for classification of land cover types by analysing spectral data from the satellite images. For this study, Landsat-8 dataset is utilized for classification and mapping of land cover in Mumbai and Satara. Each LULC type has a unique impact on air quality; for example, densely populated urban areas contribute to increased pollution due to high traffic and industrial activity, whereas forested and agricultural regions typically have better air quality due to vegetation's ability to absorb pollutants and release oxygen.

The core objectives of this study include classification of remote sensing satellite data using LULC, analysing air quality pollutants such as SO_2, NO_2, PM_{10} and $PM_{2.5}$ concentrations in both regions and identify correlation between pollution levels and land cover types. For achieving this, remote-sensing data of both the regions is collected and pre-processed for accurate classification and air quality data is monitored. The LULC classification uses machine and deep learning models to differentiate between different land types and statistical analysis of air quality data is used to assess spatial and temporal distribution of air pollutants. Through the correlation of LULC types with air quality indicators, this study will determine the ways in which different forms of land cover (such as urban, agricultural, and forest) affect air pollution levels.

The findings of this study will provide a relationship between land use and air quality and its valuable insights. The results of Mumbai region can inform techniques to increase areas of green coverage and manage industrial zones to effectively reduce pollution. In Satara, the results can highlight the ways to maintain air quality and sustainable land use practices as development increases. These findings are intended to help policymakers and urban planners make educated decisions about land use management, pollution control measures, and sustainable development practices in Maharashtra, ultimately contributing to public health improvement and environmental conservation.

1.3 Organization of the Report

This project report is intended to be comprehensive in its research of the LULC classification and its consequences for air quality analysis, notably in Mumbai and Satara, Maharashtra. Section 1 highlights essential ideas of LULC classification and emphasizes the necessity of knowing these areas, which show different land use trends in direct association with environmental conditions in most cases (such as air quality). Section 2 is a comprehensive review of literature, methods, and tools that have been developed to date for LULC classification and highlights current practice and future challenges and areas to be filled in the context of conducting the proposed research. Section 3 provides functional and non-functional requirements that describe technical, operational, and performance criteria that define the system, with consideration that it should be designed and developed according to the overall project objectives. Section 4 presents this project development with an emphasis on phases of data acquisition, processing, analysis, and visualization, as well as an iterative process applied for the refinement of the model. Section 5 explains the methods and techniques, tools, as well as the sources of data processed, machine learning models used, and statistical methods considered for classifying LULC and its contribution towards air quality. Section 6 discusses the technical development of the system, detailing the programming languages, libraries, and frameworks applied and, therefore, how the theoretical model was translated into this functioning system. Section 7 communicates the results of the study using data visualization and statistical analysis to interpret the efficiency of the system in the identification of LULC pattern and impacts on air quality. Section 8 summarizes the results and findings of the project, brings out the possible applications of the project, and give future research and system improvement avenues. Every section by itself promotes the understanding of LULC classification and air quality analysis and contributes to a deeper understanding of how these factors affect the environment.

II. LITERATURE SURVEY

The present changes in Land Use Land Cover Classification dramatically influence ecosystems, urban growth, and environmental health. This is thus the basis for some profound insights for intricate relationship between land management and air quality.

K. H. Jodhani et al. pertains to the potential impact of land use/land cover on air quality using synergized Google Earth Engine and Earth observations in the state of Gujarat, India, for the period between 2018 and 2023. It evaluates pollutants such as CO, NO_2, CH_4, SO_2, and HCHO based on six different categories: water bodies, forests, agricultural land, built-up areas, barren land, and scrubland, with the help of Sentinel-5P satellite data. The rising pollutant levels cause the expansion of urban land at the cost of shrinking agricultural and forest areas. With decreased emission, the COVID-19 pandemic has posed a momentary advancement for much better air quality for some time. The significance of LULC monitoring to further support environmental management and public health policy in uncovering significant contributors to pollution in terms of urbanization and industrialization is underlined in the paper [1].

Ajay Kumar Taloor et al. focuses on Tawi Basin land study which uses the CA Markov model in land-use and land-cover change analysis, giving more emphasis on LULC as a marker of environmental and socio-economic changes. It classifies LULC distributions and simulates future scenarios from 2030 to 2100 by using images from 2010 and 2020 from Landsat. Results show that the settlement areas will sharply rise from 5.29% in 2020 to 13.98% by 2100, respectively, due to rapid urbanization and growth in population. The finding is projected to be an important planning tool for future and is thus proposed for land management with sustainable policies towards developmental pressure in agriculture and urbanization. It focuses on the necessity of making informed choices to reduce natural hazards and maintain ecosystem services, hence ensuring sustainable regional development [2].

Chisanga et al. uses Sentinel-2 data to study LULC changes in Zambia during the period 2000 to 2022 and identifies the important trends of a rise in built-up and barren land but, conversely, a decrease in dense forests due to the ongoing increase in urbanization and population. A projection of future LULC change for 2030 was done using an ANN model, with an increase in built-up areas by 71.44% and a decrease in cropland by 0.73%. It is envisaged that Zambia's effort in sustainable land-use planning coupled with natural resource management would contribute significantly to mitigating environmental concerns [3].

Osman MAA et al. analyses the changes in LULC over thirty years in Gedaref state, Sudan using Landsat imagery and a random forest classifier. Scenarios for 2028 and 2048 were predicted with the CA-ANN model, while dynamics in LULC were studied. Results on major shifts between 1988 and 2018 point towards an expansion of 13.92% in cropland, while settlements expanded by 319.61%. Forest and grassland showed a decline of more than 56%. The expansion of cropland mainly usurped forests and grasslands with settlers entering cropland. Mechanized farming and increasing population will slightly increase the cropland lands from 89.59% to 90.43% and the forest area from 0.47% to 0.41% by the year 2048. The research underlines the necessity for sustainable land policies, resource management, and reduction of landscape degradation, thus helping policymakers make regional planning decisions [4].

Matci et al. analyzes the effect of the COVID-19 pandemic on Turkey's air quality through MODIS and Sentinel-5P TROPOMI data from 2019 to 2021, benefiting from the advancement achieved in remote sensing and also the platform provided by Google Earth Engine to perform the analysis efficiently. The findings show a general significant decrease in NO_2 levels, especially during the stringent lockdown of March-April 2020, compared to those in 2019 in urban areas. Moreover, the authors found a link between air quality and temperature; the negative relationship is found to be pronounced in urban areas whereas the other land covers exhibit a positive relationship. The research, while it suggests several positive effects on the environment based on decreased human activities, warns that such benefits might be unsafe. The paper calls for more research to be conducted on strategies that can assist in decreasing urban air pollution [5].

Yang et al. investigates the relationship of $PM_{2.5}$ concentration and LULCC in Shanxi Province, China, using high-accuracy $PM_{2.5}$ data from the ACAG and GWR model. Most studies usually focus on urban areas, but this pays attention to regional impacts of LULCC on the distribution of $PM_{2.5}$, especially in rural areas. GWR analysis revealed a high $PM_{2.5}$-LULCC relation, which is very stable, with an R^2 of 0.94. Higher local R^2 values at sites where significant LULCC changes have occurred corresponded to areas with significant LULCC changes. In summary, above results indicate that the realization of LULCC impact on regional air quality during environmental protection is crucial as well as in future decision-making [6].

Diana B et al. study LULC change impact on the East Baton Rouge, Louisiana air quality from 1991 up to 2021. The study closely looks at satellite imagery obtained through LANDSAT and comparatively, it has data on air quality from the U.S. EPA, for example, CO, NO_2, and $PM_{2.5}$. Despite urbanization and population growth, results reveal that air quality has greatly improved over the last three decades, meaning other factors such as compliance with EPA regulations also contributed to the change. The research draws attention to the importance of resource use efficiency and urban planning. The paper concludes with policy suggestions about renewable energy as a tool to mitigate the adverse impacts of the use of the traditional sources of energy like the coal-fired plants [7].

Gupta et al. assesses the LULCC in Imphal Valley, Manipur, India, during the period 2016-2021 and its implications on air quality variations due to changes in land covers. By mapping LULC using Sentinel-1, Sentinel-2, and ALOS PALSAR data and by the application of the Random Forest algorithm, increased settlements along with horticulture, and a decrease in forests and phumdis formed the trends established. The data of Sentinel-5P shows that CO, HCHO, NO_2, and SO_2 spike in forest fires and shifting agricultural seasons and decline during lockdowns. The authors urge the effective management of land to reduce air pollution in congruence with climate goals by using advanced technologies and data integration for better environmental management [8].

Lu et al. applies the WRF-CMAQ modeling system, coupled with MODIS 2019 land cover data to inquire into the meteorology and air quality related impacts of land use and land cover changes (LULCC) in southwestern China's Sichuan Basin. The accuracy of inputted land cover will affect model performance, and outdated datasets can inhibit actual predictions-especially in complex terrains. Direct comparisons with scenarios of MODIS data to the standard WRF land covers render trends in urbanizations not accounted for in the derived data. The analyses show that LULCC increased 2-m temperature by as high as 2°C, raised the 400-m increase in planetary boundary layer height, and increased basin-wide ozone levels from the weakened NOx titration processes. Real-time acquisition of data would further inform aspects of improving model accuracy and would go ahead to inform pollution control strategies [9].

Bui et al. suggests that gradient boosting algorithms can be integrated with CNNs as applied in the classification of land cover using SPOT7 imagery. Replacing traditional fully connected layers in CNNs with XGBoost, LightGBM, and CatBoost classifiers will improve the performance in classification in this research work. Thus, the methodology involves the image segmentation procedures, feature extraction, and the graph generation stages involved in preparing the data for training. The results reflect marked improvements, wherein it tends to indicate XGBoost at 0.8905 accuracy while LightGBM, as well as CatBoost, reflect 0.8956. This paper highlighted the efficiency of OBIA in proper classification of land covers by considering spatial as well as spectral features. It implies that the fusion of OBIA and CNN-based gradient boosting algorithms is promising for complex urban areas' analysis, thus the research should shift towards feature ordering and data generation techniques for further enhancement [10].

Abdi et al. aims at the performance of four machine learning algorithms: support vector machines (SVM), random forests (RF), extreme gradient boosting (XGBoost), and deep learning (DL), in the context of classification of land cover and land use (LCLU) of a boreal landscape covering a south-central Sweden area, exploiting Sentinel-2 imagery. Diversified types of vegetation were found in the study area; hence, for the algorithms, eight LCLU classes were applied, such as forests, water bodies, and agricultural land, based on multi-temporal imagery. Results: SVM obtained the highest accuracy at 0.758 followed by XGBoost at 0.751 and finally by RF with an accuracy of 0.739 and DL at 0.733. Overall, the study underscores the requirement of using red edge and short-wave infrared bands in classification. Possibly these bands outweigh the conventional vegetation indices applied in boreal landscapes. Such bands are worthy of further research in mono-temporal and multi-temporal scenarios [11].

Beshir et al. analyses the LULCC from 1992 to 2022 in the upper Wabe-Shebele river basin in Ethiopia. The dynamics of population expansion, industrialization, and urbanization on land use and cover changes are presumed as the major focuses. Five main categories of LULC were identified, with vast growths in cultivated lands and settlements; however, they are very depleting in forest lands, pasture areas, and water bodies. The projection would mean further increase in cultivated land, and the techniques of sustainable land management become important to avert environmental as well as community impacts from this change. The research also presents LULC projections up to 2052, which would be helpful in establishing future planning and conservation directions [12].

Faisal et al. analyzes aerosol amounts, using Aerosol Optical Depth (AOD) and $PM_{2.5}$ concentrations, over the Dhaka Metropolitan Area from 1999 to 2019. AOD at spatial resolution of 30 meters can be retrieved using Landsat Imagery and the Simplified Aerosol Retrieval Algorithm (SARA), thereby compensating for the lack of the ground monitoring stations. It fits well with the intensification in $PM_{2.5}$ levels as closely related to an expansion of the urban area (153.47 km²) and a reduction in vegetation cover by 80.8 km². Thus, these findings point towards spatial data of air quality being highly critical for city planning and policies because it tells an assessment both in terms of health and economic impacts. Despite limitations in validating 2019 $PM_{2.5}$ data, the study thereby opens an avenue to show how

2019 has been clogged by serious pollution issues [13].

Mampitiya et al. focuses on projecting PM_{10} levels in Sri Lanka, specifically in metropolitan areas, emphasizing the significance of air quality. It assesses the ability of eight machine learning models, including Artificial Neural Networks (ANN), Bi-directional Long Short-Term Memory (Bi-LSTM), Ensemble, XGBoost, CatBoost, LightGBM, LSTM, and Gated Recurrent Unit (GRU), to forecast PM_{10} concentrations in Battaramulla and Kandy. The study determined that the Ensemble model was the most successful through correlation analysis, hyperparameter tuning, and robust data processing. It accomplished a regression coefficient near 1, demonstrating its excellent accuracy for PM_{10} prediction. These findings help to promote public health and guide air quality policy, addressing Sri Lanka's air pollution challenges [14].

H. Du et al. works on improving LULC classification in Xinjiang using ensemble learning technique that includes incorporating different algorithms like KNN, SVM, RF, ANN, and C4.5, which are tailored specifically for this region's climatic and geographic scenario. By dividing Xinjiang into a more detailed subregional level and supplementing the imagery with Landsat images, especially for urban and forest cover, the study obtained higher classification accuracy. Besides validation of ensemble learning's capability in challenging topographies, the high-precision LULC map obtained has further applicability in areas with similar characteristics and significant benefits in climate simulation and ecological research [15].

J. Yuan et al. addresses most crucial gaps in RS scene datasets, which the WH-MAVS dataset fills, include the lack of usable high-accuracy datasets and georeferenced data for applications such as real estate appraisal and urban planning. It comprises 14 land use classifications based on urban planning, which helps in research into urban land change, slum detection, and housing value calculation. In images captured with Google Earth over Wuhan, this dataset offers 23,567 tagged image patches and helps in scene classification and change detection. Under these models, DenseNet performed better than others; the maximum accuracy was achieved by DenseNet169. WH-MAVS greatly improves urban planning, landscape analysis, and monitoring of land use, and it significantly contributes to researching algorithms as well as practical applications [16].

Rawat et al. compares the effect of three sampling methods on the performance of machine learning algorithms in classifying LULC of Dehradun, Uttarakhand, using Sentinel-2 data. These include stratified random sampling with proportional and equal samples as well as stratified systematic sampling with a minimum distance of 10m. Classification was done using R by the application of 25,000 samples equally split into training and testing datasets. The results showed that SRS proportional had the highest overall accuracy and kappa value of the study, which majorly favoured most classes, such as deciduous and evergreen forests. The study, therefore, emphasizes the significance of sampling strategy in increasing the accuracy of LULC classification, particularly in sensitive terrains having limited reference data [17].

Puttinaovarat et al. utilizes satellite images from both the Landsat 8 and Sentinel-2 machines and uses machine learning techniques towards improving accuracies in LULC classifications. The total correctness value of the proposed method is about 88.8% for classifying four LULC classes that are, the urban, agriculture, forest, and water classes. The study further emphasizes the idea of combining the crowdsourced data, updating information, and building confidence with the LULC data, which makes it more reliable and closer to real conditions. Along with the developed LULC information system, volunteer data input can be taken for real-time updates and comparisons with satellite classifications. Paper further discusses the importance of LULC data in urban planning, disaster risk analysis, and other managerial tasks as a need for correct and time-available information in decision-making processes. In summary, the current study presents a novel approach toward LULC classification in terms of unification of satellite imagery and machine learning for producing high-quality spatial data directly involving a community for different applications [18].

Pan et al. discusses the automation approach to LULC classification using RF and CART within GEE, where the authors revealed an accuracy of 87.24% in Australia and 85.18% in the USA; hence, they concluded that RF performs better than CART, overcoming the typical methods that require large datasets. Unlike CART, the RF outputs were more focused, while CART results tended to be very fragmented. This method enhances the specificity of categorization by using the MCD12Q1 land cover product for appropriate labelling so that it can be used much more precisely as an automated LULC tool for land managers and policymakers in the analysis of satellite imagery [19].

Li et al. investigate the relationship of air quality with socioeconomic status and land use in Phoenix between 2000 and 2010. The research used Landsat imagery to generate fine-resolution aerosol optical depth (AOD) data to discover that pollution levels are higher in lower-income communities, even emphasizing social inequities in exposure to air pollution. Rapid urbanization in Phoenix has compounded such inequalities, and evidence by the study indicates that not all types of vegetation mitigate air pollution alike. This raises a query into how vital it would be to plan green spaces and choose the right types of plants in improving air quality and redressing environmental injustice. The article suggests the use of remote sensing to investigate socio-environmental dynamics, which gives critical knowledge to acquire for development purposes [20].

Thakrar et al. emphasizes the importance of air quality health effects in land-use decisions. This analysis predicts through 2051 implications of a number of land-use strategies in the United States and finds that for most circumstances, air-quality-related social costs and benefits outweigh carbon storage and economic returns from air quality disadvantages frequently. Conclusion This conclusion brings forward the idea of air quality-related factors in land-use planning to improve health outcomes and even more informed decisions [21].

Araki et al. estimated historical $PM_{2.5}$ doses in Japan for the period of 1987 to 2016 using a national-scale model, overcoming the limitation of poor availability of monitoring data before 2014. The model was found to achieve great accuracy with high values of R^2 spatial and R^2 temporal, at 0.76 and 0.73, respectively, using a neural network model that incorporated the non-linear interactions between $PM_{2.5}$ and predictors including air pollution and land use. It realistically replicated monthly fluctuations during 2000 and 2013. The study also pointed out that $PM_{2.5}$ concentrations have declined over the 1990s, further establishing that the model can be a guiding tool in epidemiological research and environmental regulations concerning the health impact of pollution, such as cancer and heart disease [22].

Xiao et al. aims to estimate the concentration of $PM_{2.5}$ - one among important public health estimates and introduces a new model that has been termed Weighted Long Short-Term Memory Neural Network Extended (WLSTME). The model included the two factors most exclusively being ignored in previously proposed methods: site density of monitoring stations, and wind conditions. Weighted historical data are formed using neighbouring stations, which bear wind and distance into consideration, and inserted into an LSTM network to make a spatiotemporal prediction. The results based on the data of the Beijing-Tianjin-Hebei region for the period 2015-2017 show that the WLSTME model is best in fitting other models and possessing a higher correlation and fewer errors. Improvement in wind effects and site density is an accentuated need for improved air quality forecasting by the study [23].

Dilawar et al. deals with the interaction between land use changes, climate variability, and air pollution in Pakistan for the period 2004 to 2021. Using remote sensing and meteorological data, it finds a very high increase in PM concentrations specifically in urbanized and industrialized areas and considers the impact of precipitation, temperature, and wind speeds over air quality. Results: The paper explained in detail how land use and climate dynamics significantly impact air pollution, bringing the need for strong environmental regulations and effective management techniques to improve air quality and establish sustainable urban settings in Pakistan [24].

Ruidas et al. analyses $PM_{2.5}$ levels in Maharashtra, India, at three different phases of COVID-19, namely pre-lockdown, lockdown, and unlocking. From the results, it is observed that the $PM_{2.5}$ has considerably decreased up to 74% during the lockdown due to minimum industrial and vehicular activities, but the level again increased during the unlocking period. Long-term air quality management, combined with continuous monitoring, would be required and would result in proposals such as relocating industries that cause pollution and stricter emission control, but it highlights that only under less human activity in certain periods can such recovery be brought about and puts even more emphasis on the critical long-term measures in air quality [25].

The reviewed studies show the critical relationship between LULC changes and air quality across various global regions. However, these environmental improvements are not sure to be sustainable especially with continued population growth and industrialization. The studies called for informed land-use management and the implementation of sustainable practices that would mitigate future degradation of the environment. In summary, LULC and air quality monitoring should be done continuously in order to ensure effective management of the environment.

III. REQUIREMENT SPECIFICATION

The requirement specifications of this project are mentioned as given below.

3.1 Hardware Requirements:

a) Laptop with a high-performance GPU

b) 8/16 GB RAM

c) 1TB+ SSD

3.2 Software Requirements:

a) Google Colab

b) QGIS (Quantum Geographic Information System)

c) Google Earth Engine (GEE)

d) ArcMap

3.3 Feasibility Study:

a) Application compatibility: All essential applications, including Google Colab, QGIS, Google Earth Engine, and ArcMap, should be compatible and integrate seamlessly into the project. Every application should be able to incorporate geographic and categorization features.

b) Hardware Capability: Check whether the hardware provided has the capacity to handle such a volume of data in performing the task of classification.

c) Availability of Data: It is essential to look at the availability of credible data on land cover from sources such as Google Earth Engine, which offers several remote sensing datasets. The availability is crucial for an exact classification.

d) Skill Requirements: Team should be equipped with skills in geospatial analysis and machine learning using GIS for LULC classification. If not equipped, a plan for training or hiring would be necessary.

e) Timeline: Prepare a timeline to see whether the project falls within the time frame of collection, pre-processing, model developing, validation, and reporting.

f) Scalability: It must take into account the potential scalability of the project to larger coverage or finer resolution depending on the project's future needs.

3.4 Cost Estimation:

The cost estimation for the project is highlighted in the table given below:

Table 1: Cost Estimation

Category	Item	Cost (in Rupees)
Hardware Requirements	Laptop	₹1,00,000 – ₹2,00,000
	External Storage (optional)	₹8,000 – ₹15,000
Software Requirements	Google Colab Pro (optional)	₹850 per month / ₹10,200 per year
	ArcMap License	₹1,25,000 per year
	QGIS	Free
	Google Earth Engine	Free for academic use
Data Acquisition Costs	Satellite Imagery	₹8,000 – ₹80,000 (if paid)
Miscellaneous Costs	GIS Specialist	₹2,500 – ₹5,000 per hour
	Training and Workshops	₹40,000 – ₹80,000
	Cloud Storage and Compute	₹4,000 – ₹16,000 monthly
Total Estimated Cost		**₹1,89,050**

IV. PROJECT ANALYSIS AND DESIGN

The methodology incorporates both remote sensing and GIS data for examining the Land Use Land Cover classification impact on air quality. Immediate objectives include accurate LULC classification, identification of trends concerning land uses, and comparison of these results with air quality data to offer an insight into urban and environmental planning.

4.1 Functional Requirements:

- **LULC Classification:** Satellite images to classify different types of land coverings which are aquatic bodies, forests, cities and farm lands.
- **Air Quality Analysis:** Analyze by fetching data from trusted sources, such as $PM_{2.5}$ and NO_2 levels and compare it with LULC.
- **Data Preprocessing/Processing:** Filtering, resampling, and scaling of images for the analysis process. The relevant GIS platforms are QGIS and Google Earth Engine, which are to be utilized for purposes of visualization and reporting of the result for the areas of Mumbai and Satara.

4.2 Non- Functional Requirements:

- **Performance:** Ensure the system is able to handle large data and still allow for results that are accurate and efficient.
- **Usability:** The user interface should be accessible for the researchers and other stakeholders.
- **Scalability:** it must be possible to use it on larger areas or data set if needed.
- **Reliability:** Test the reliability of the model by ensuring it is robust in LULC classification.

4.3 Flow Diagram:

Raw satellite images and air quality data are part of the input layer. The data cleaning, LULC classification, air quality interpolation, and correlation analysis form the processing layer. Output layer include GIS visualizations, correlation models analytical reports as shown in Figure 1.

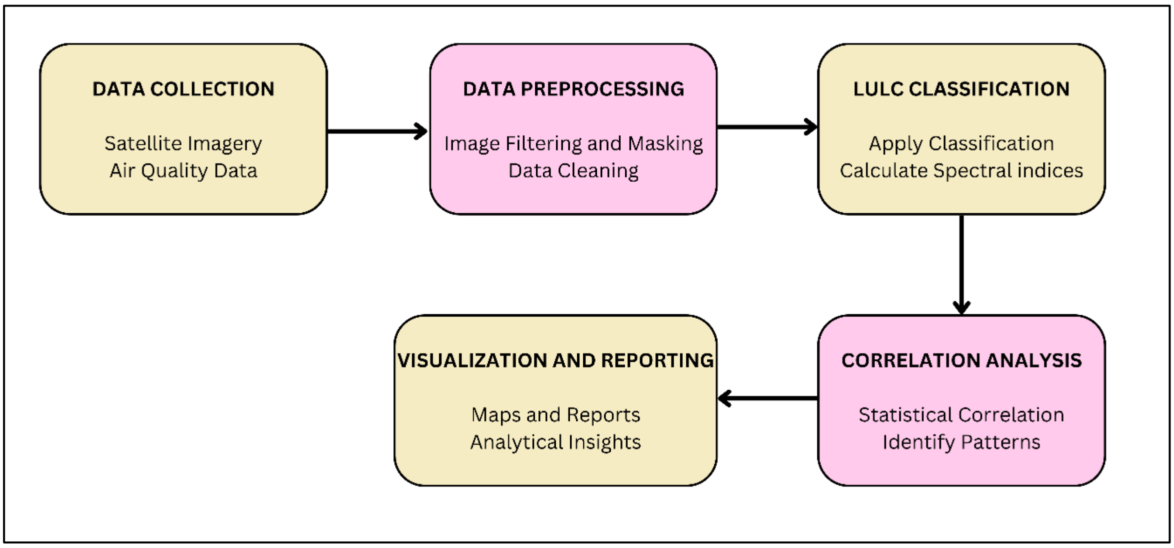


Figure 1: Flow Diagram of Proposed System

V. METHODOLOGY

5.1 Introduction

The study of land use land cover classification and the impact on air quality focuses on two different contrasting regions in Maharashtra, India : Mumbai and Satara. Mumbai located on western coast of Maharashtra at 19.0760° N, 72.8777° E is an urban densely populated region with large-scale industrialization and commercialization across an area of 603 km² and is bordered by Thane in the north and Arabian Sea to the west. Mumbai becomes an ideal region for studying land cover changes and variations in air quality due to rapid urban growth and high pollution levels. Satara on the other hand spans 10,480 km² with coordinates 17.6859° N, 73.9933° E. It is dominated by agricultural and forest cover and has significant low pollution levels which is a contrast to the urban landscape of Mumbai. These regions illustrate various land cover types and the changes that are projected over the year 2025. The geographic locations of the research areas are depicted in Figure 1.

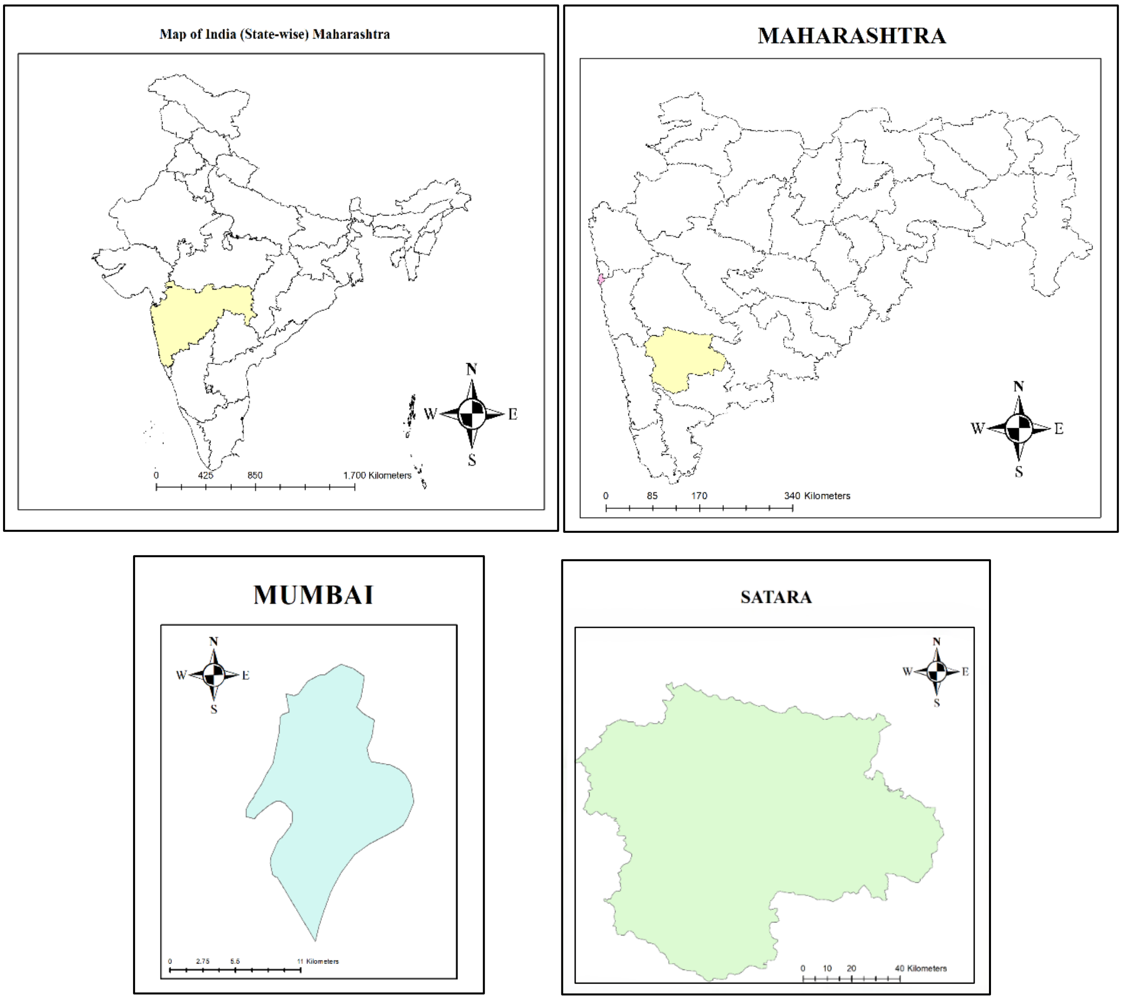


Figure 2: Maps of Mumbai and Satara Districts in Maharashtra, India.

The study utilizes satellite imagery dataset of Landsat 8 and Landsat 9 for LULC classification and future changes as shown in Table 2. It uses Google Earth Engine (GEE) to collect data points of both the regions and for classification, different machine learning and deep learning models were implemented. Future prediction changes are done using the Molusce tool present in open-source software QGIS. The study also employs the use of Sentinel-5p dataset for predicting current air quality of both the regions. The current air quality is predicted using Google Earth Engine and future predictions for air quality AQI indexes, the hybrid model of Gated Recurrent Unit (GRU) and Graph Convolutional Network (GCN) is used. The proposed system is implemented in two parts: 5.1.1 and 5.1.2. The 5.1.1 deals with LULC classification and future predictions and 5.1.2 deals with current air quality AQI index levels and future change predictions in these AQI levels henceforth highlighting contrasting differences among both the regions.

Table 2: Landsat Satellite Information

Parameter	**Landsat 8 and Landsat 9 Dataset**
Cloud Platform	Google Earth Engine (GEE)
Satellite	Landsat 8, Landsat 9
Sensor	Operational Land Imager (OLI) and Thermal Infrared Sensor (TIRS)
Product Type	Surface Reflectance
Sensor Operational Modes	Multispectral and Thermal
Polarizations	Not applicable (optical sensor)
Orbit Properties Pass	Ascending and Descending
Image Acquisition Date	Based on available dates with <10% cloud cover
Cloud Cover	Less than 10% (filtered for clear-sky conditions)
SENSING_TIME	Specific timestamp for each image acquisition
LANDSAT_PRODUCT_ID	Unique identifier for Landsat product version

5.1.1: Land Use Land Cover Classification and Future Predictions

LULC classification and future prediction involves various steps such as acquiring dataset for land use land cover classification, model training and evaluation and future prediction changes for regions of Mumbai and Satara. The proposed methodology for this study is highlighted in Figure 3 given below.

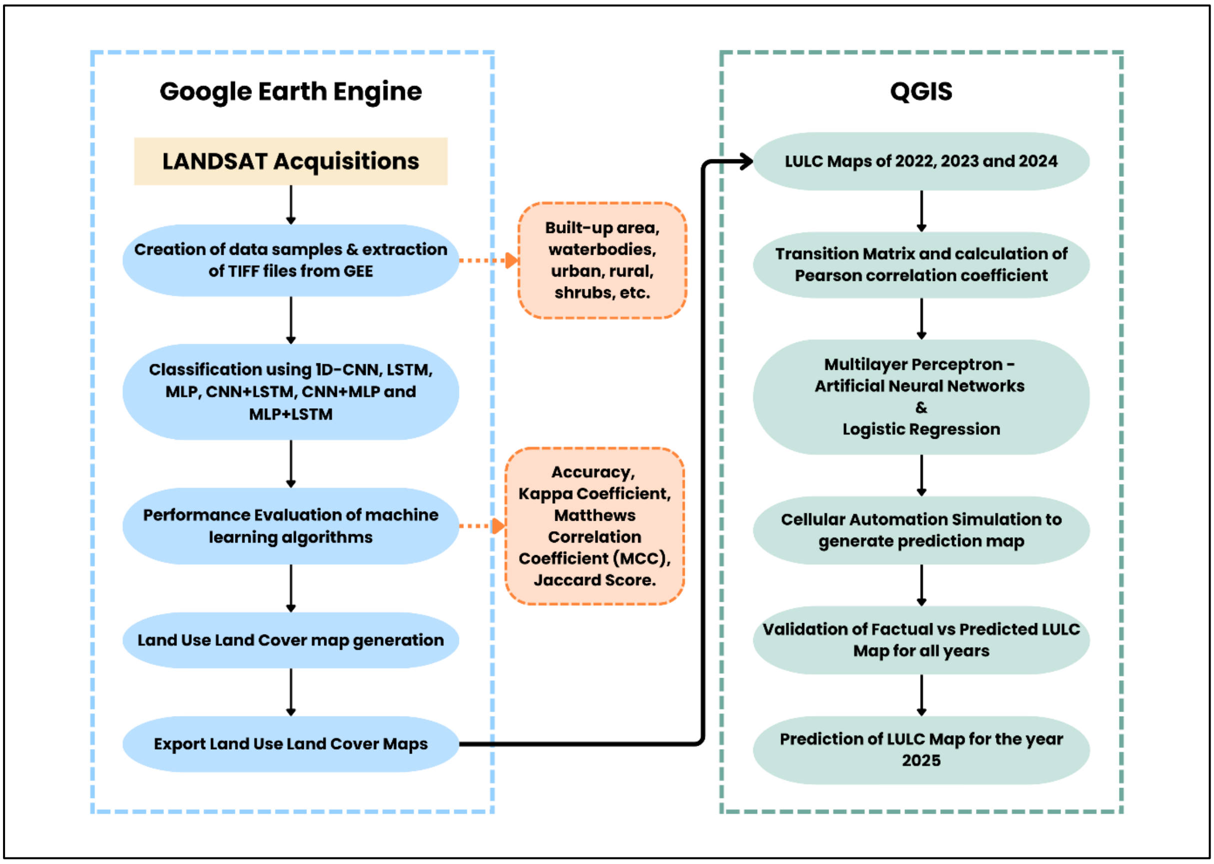


Figure 3: Proposed Methodology of the study

Step 1: Acquiring Dataset for Land Use Land Cover Classification

The study utilized Landsat 8 and Landsat 9 dataset available on Google Earth Engine which provide high-quality multispectral and thermal images. The data consists of spatial resolution of 30 meters and temporal resolution of 16 days. The images are processed to provide cloud-free multi-temporal composite pictures and are later clipped to study Mumbai and Satara regions. The primary land cover types identified in Mumbai and Satara regions are built area, bare land, water, wet land, herbs, dry shrubs, wet shrubs, plantation forests and palm oil. A balanced dataset is essential for ensuring successful machine learning classification. For training the model, different representative samples were marked for regions of Mumbai and Satara on Google Earth Engine which served as training and testing data for the classification model from years 2022 to 2024. For Mumbai, 96320 samples were marked and collected as training dataset and 16898 samples were marked and collected as testing dataset. For Satara, 31372 samples were marked and collected as training dataset and 5945 samples were marked and collected as testing dataset. As a result of this analysis, a CSV file containing data points was obtained along with a TIF file which is a map of Mumbai and Satara regions for further model training and evaluation. Table 3 and Table 4 depicts the details of the dataset samples for Mumbai and Satara regions.

Table 3: Sample Datapoints for Mumbai region

District (Mumbai)	Built	Water	Wetland	Herbaceous	Dry Shrubs	Palm oil	Wet Shrubs	Bare land	Plantation Forest
Geometrical Points	97	30	52	32	28	26	42	34	30

Table 4: Sample Datapoints for Satara region

District (Satara)	Built	Water	Wetland	Herbaceous	Dry Shrubs	Palm oil	Wet Shrubs	Bare land	Plantation Forest
Geometrical Points	67	45	18	15	9	1	19	23	51

Step 2: Model Training and Evaluation

The CSV and TIF file of Mumbai and Satara regions, acquired from the Landsat dataset, is used for training the model. This study employs the use of different machine learning and deep learning models for land use land cover classification such as One-Dimensional Convolutional Neural Network (1D-CNN), Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM) and some hybrid models such as Convolutional Neural Network and Long Short-Term Memory (CNN+LSTM), Convolutional Neural Network and Multi-Layer Perceptron (CNN+MLP) and Multi-Layer Perceptron + Long Short-Term Memory (MLP+LSTM). These models were evaluated on the basis of different evaluation parameters such as Accuracy which measures the correctly classified samples among all the total instances, Kappa Coefficient which evaluates the agreement between observed and expected classifications by chance, taking into consideration both correct classifications and the potential of agreement occurring at random, Matthews Correlation Coefficient (MCC) which measures binary classification by taking true and false positives and negatives in account and Jaccard Score which measures similarity between predicted positive samples and actual positive samples.

Step 3: Future Predictions

After model training and evaluation, the model which performs the best, a TIF file is obtained for that model and this file is then used for detecting future predictions. The future predictions are done using Molusce Tool present in open source QGIS software. In the initial stage of prediction, the initial LULC maps from 2022 to 2024 are fed into the prediction pipeline and Pearson correlation is used to check relations between spatial variables. The change in area is calculated from the 2022 LULC map to 2024 LULC map and a transition matrix is obtained. In the transition potential modeling stage, transition maps are generated using artificial neural network multilayer perceptron and logistic regression which are trained for 300 and 100 epochs respectively. Following this, cellular automata simulation is carried out which generates LULC future prediction maps for year 2025. The generated maps are evaluated using Kappa coefficient values and overall accuracy is calculated.

5.1.2: Predictions of Future Air Quality Changes

Along with changes in land cover types in Mumbai and Satara regions, air quality in both these regions have also changed. The data of AQI index levels for year 2017 to 2024 is obtained from Maharashtra Pollution Control Board (MPCB) for Mumbai and Satara regions. The Sentinel-5p dataset is used to obtain maps for both the regions using GEE processing. The future predictions of air quality changes are done using the hybrid model of Gated Recurrent Unit (GRU) and Graph Convolutional Network (GCN). The architecture of the hybrid GRU and GCN model is depicted in Figure 4 given below.

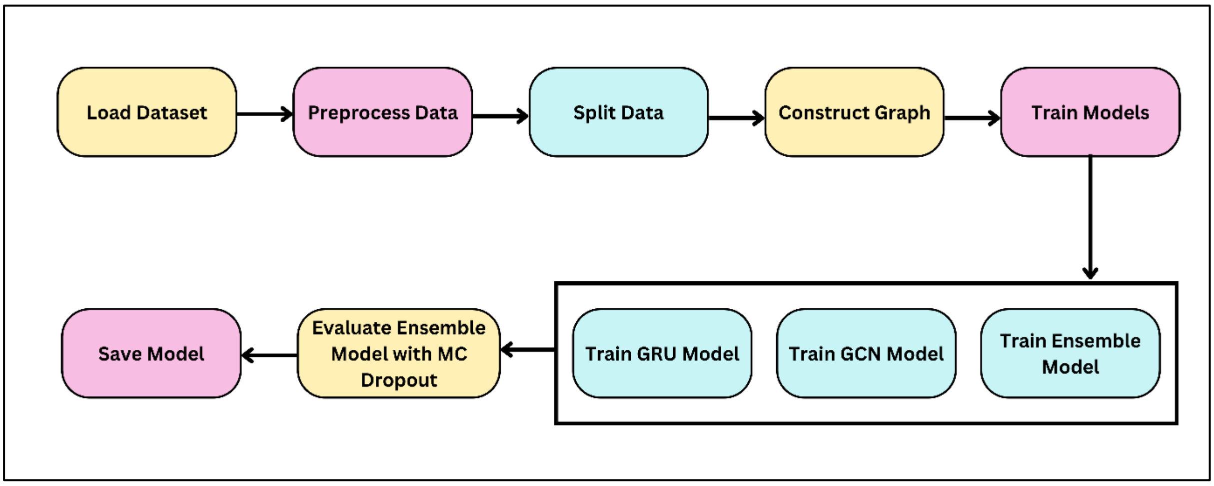


Figure 4: GRU+GCN Architecture

The steps involved in the proposed workflow are as follows:

1. ***Data Preparation***
   - **Compile Pollutant Data:** Past data is compiled for Mumbai and Satara on pollutants

such as SO_2, NO_2, PM_{10} and $PM_{2.5}$. This information can be gathered via Maharashtra Pollution Control Board.

- **Standardize the features:** Pollutant data is scaled to mean 0 and a standard deviation of 1. The model converges more successfully using normalization.

2. Spatial Dependencies

- **Construct Adjacency Matrix (A):** A fully connected graph of each district in the Satara and Mumbai areas is represented as a node. Edges indicate geographical proximity or interaction levels between districts, with $A_{ij} = 1$ showing influence between locations i and j.

3. GRU-Based Temporal Modeling

- **Specify the Time Steps:** Data of pollutants is sorted according to time series data.
- **Implementation of GRU:** A Gated Recurrent Unit (GRU) is used to model time series data. A hidden state h_T is produced at the last step by the GRU, which will record the temporal dynamics of pollutant levels.

4. Graph Convolutional Network based Spatial Modeling

- First GCN Layer transformation is applied to standardize pollutant features and output produced is $H^{(1)}$.
- Output of the first GCN layer is passed to the second GCN layer to capture spatial dependencies across districts producing output $H^{(2)}$ which is used to predict AQI values using weights W_{gcn}.

5. Ensemble Model

- The output of both GRU and GCN models is combined to get final AQI prediction $Y_{ensemble}$ weighted by alpha and beta which leverage both spatial and temporal information.

$$Y_{ensemble} = \alpha . Y_{GRU} + \beta . Y_{GCN} \tag{1}$$

6. Uncertainty Estimation

- **Monte Carlo Dropout** is used to run the ensemble model K times with dropout enabled. This approach gives estimates of prediction uncertainty:

$$Y_{ensemble}(k) = \alpha.Y_{GRU}(k) + \beta.Y_{GCN}(k) \quad (2)$$

- **Mean and Standard Deviation:** Determine the mean prediction and uncertainty using:

$$\bar{Y}_{ensemble} = \frac{1}{K}\sum_{k=1}^{K} Y_{ensemble}(k) \quad (3)$$

$$Uncertainty = \sqrt{\frac{1}{K}\sum_{k=1}^{K} (Y_{ensemble}(k) - \bar{Y}_{ensemble})^{2}} \quad (4)$$

7. ***Application to Localized Models in Mumbai and Satara:***

- **Localized Model:** Adjust the model's settings to reflect the unique peculiarities of Satara and Mumbai, taking into account things like traffic patterns, industrial pollution, and topography.
- **Predictions:** Once trained, the model may generate AQI predictions for both regions at various time scales (e.g., hourly, daily) to monitor short-term and long-term air quality trends.
- **Actionable Insights:** The projections, together with uncertainty measurements, can inform policy decisions, public health advisories, and environmental management strategies in both locations.

This framework's dual focus on temporal and spatial aspects, together with uncertainty quantification, positions it well for making accurate and trustworthy AQI predictions for Mumbai and Satara. The results can offer important information for enhancing air quality control and supporting public health campaigns in these metropolitan settings.

VI. IMPLEMENTATION DETAILS

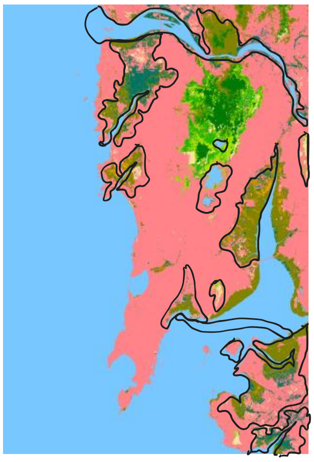
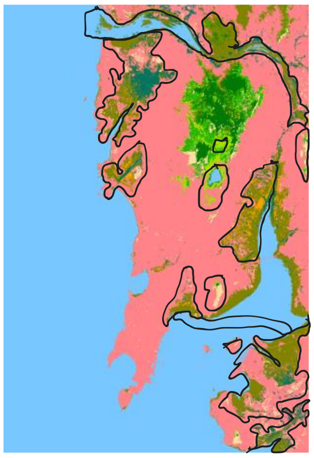
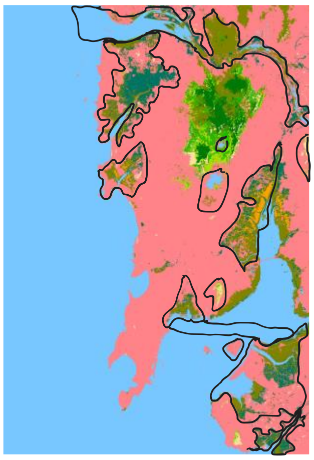
6.1 Introduction

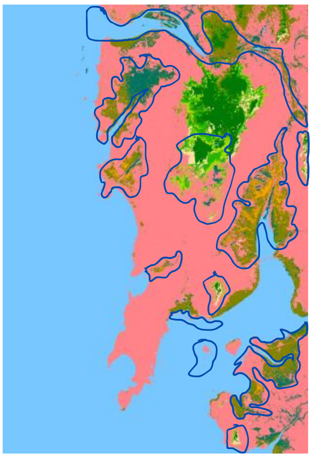
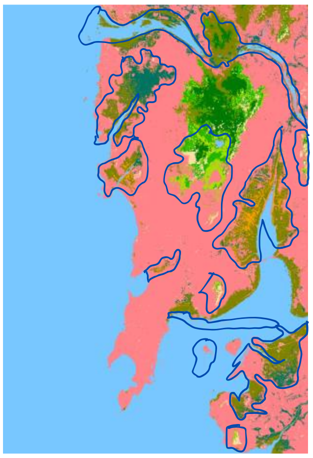
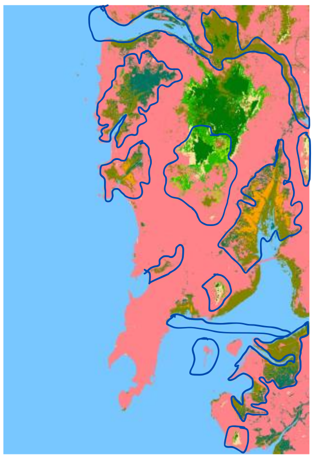
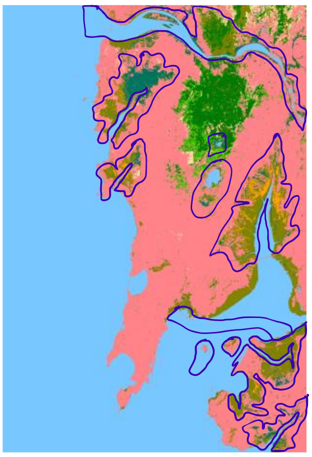
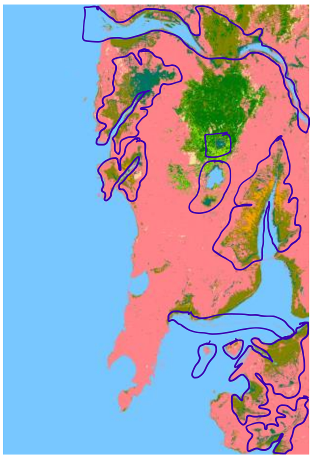
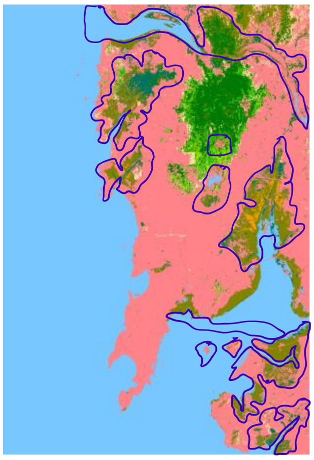
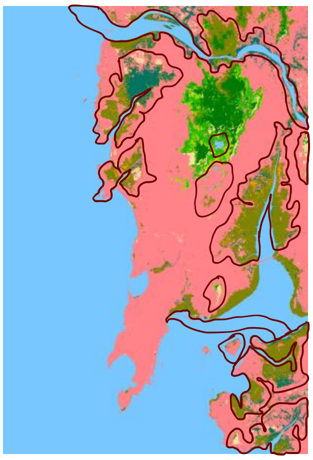
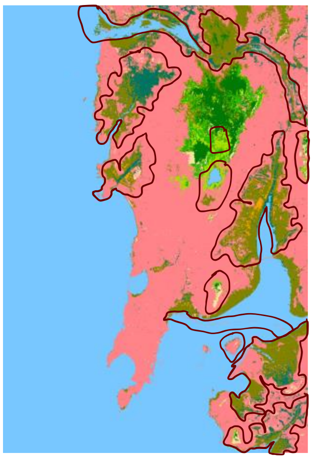
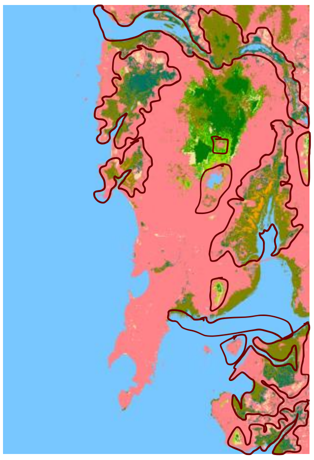
The land use land cover classification is implemented for Mumbai and Satara regions for years 2022, 2023 and 2024. The classification is done using machine and deep learning models such as One-Dimensional Convolutional Neural Network (1D-CNN), Multi-Layer Perceptron (MLP), Long Short-Term Memory (LSTM) and some hybrid models such as Convolutional Neural Network and Long Short-Term Memory (CNN+LSTM), Convolutional Neural Network and Multi-Layer Perceptron (CNN+MLP) and Multi-Layer Perceptron + Long Short-Term Memory (MLP+LSTM). The changes in the classification are marked and analyzed to visualize changes in both the regions happening over time. The future prediction for year 2025 for LULC classification was done using Mollusce Tool in QGIS. The air quality hybrid model of GRU+GCN was implemented to predict future AQI trends for Mumbai and Satara regions highlighting the impact of air quality on LULC changes. Both these models were implemented in Python using various different libraries and maps for both the regions were extracted using GEE processing.

6.2 System Implementation for LULC Classification

Table 5 highlights the LULC changes for Mumbai region implemented using different models.

Table 5: LULC classification changes for Mumbai region

CLASSIFICATION OF CHANGES			
MODEL	YEARS		
	2022	2023	2024
1-D CNN			

LSTM			
LSTM + CNN			
MLP			

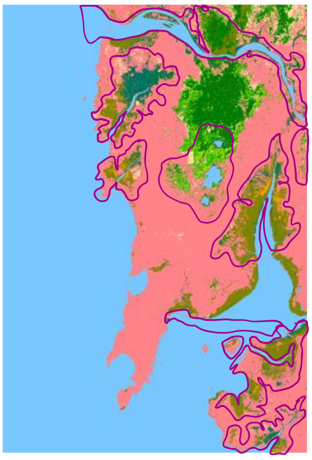
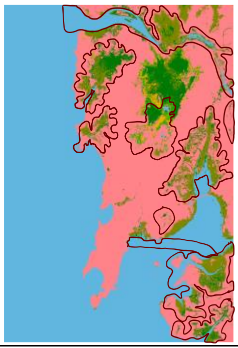
MLP + LSTM			
CNN + MLP			

Table 6 highlights the future predictions for LULC classification of Mumbai for year 2025.

Table 6: Future predictions for year 2025 of Mumbai region

CLASSIFICATION OF PREDICTION CHANGES				
MODEL	YEAR		PREDICTION	
	2022	2023	2024	2025
ANN+MLP+ LSTM				

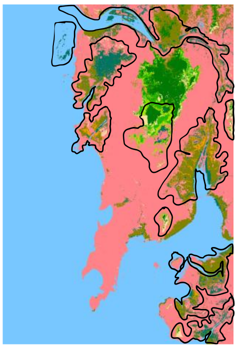
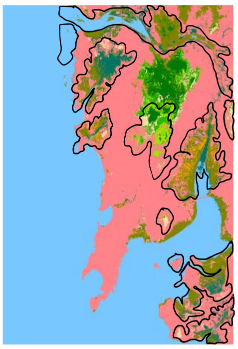
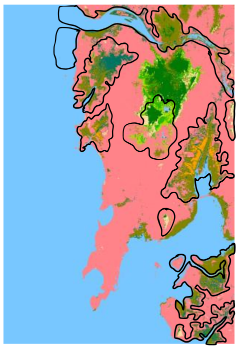
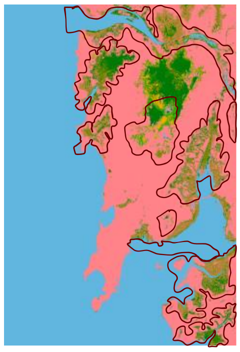
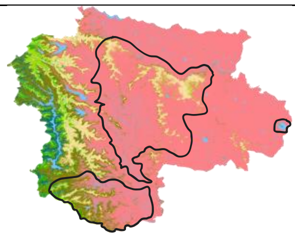
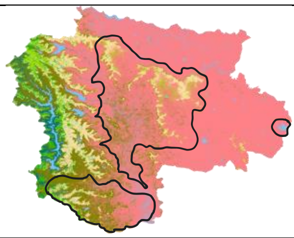
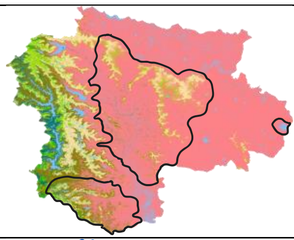
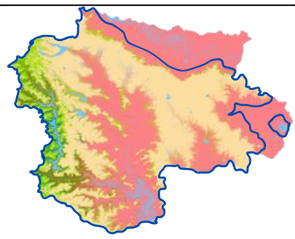
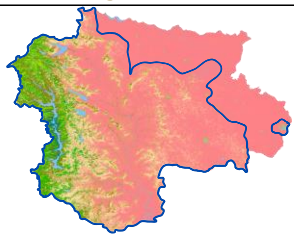
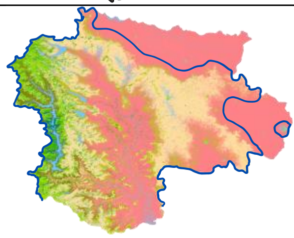
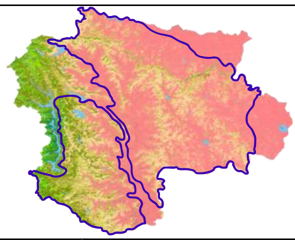
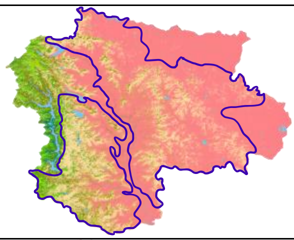
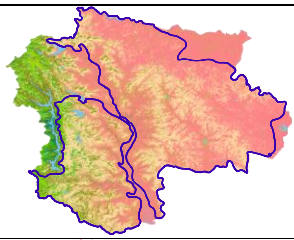
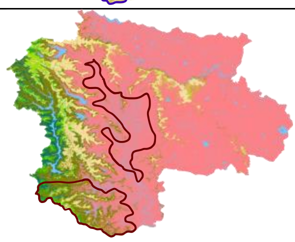
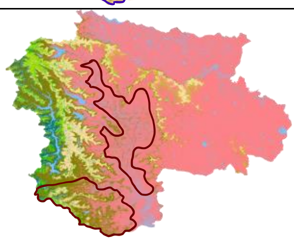
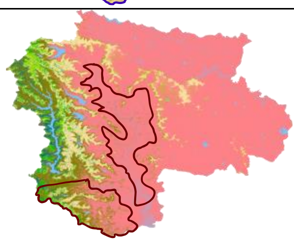
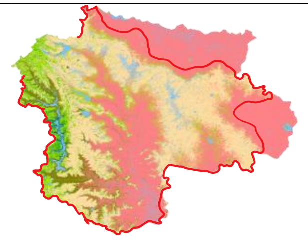
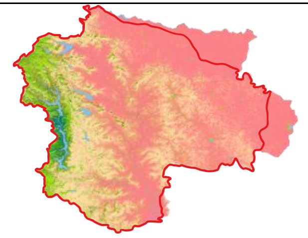
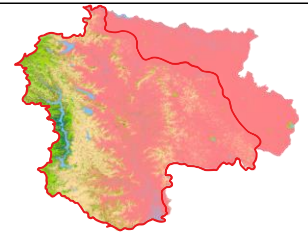
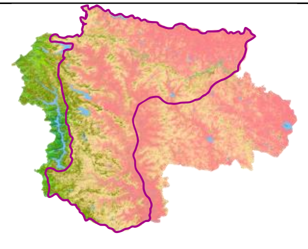
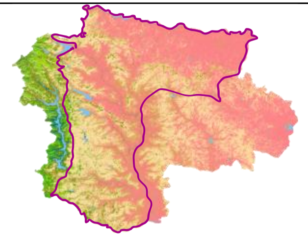
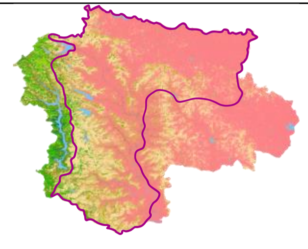
LR+MLP+ LSTM				

Table 7 highlights the LULC changes for Satara region implemented using different models.

Table 7: LULC classification changes for Satara region

CLASSIFICATION OF CHANGES			
MODEL	YEARS		
	2022	2023	2024
1-D CNN			
LSTM			
LSTM + CNN			
MLP			

MLP + LSTM			
CNN + MLP			

The changes in the LULC classification marks changes in different natural features such as vegetation, built-up, water bodies and other parameters as mentioned in Table 3 and 4.

6.3 System Implementation for Air quality changes using LULC

Figure 5 represents the NO_2 and SO_2 concentration for Mumbai region for year 2024. Figure 6 represents future predictions for AQI indexes of air quality from year 2017 to 2035.

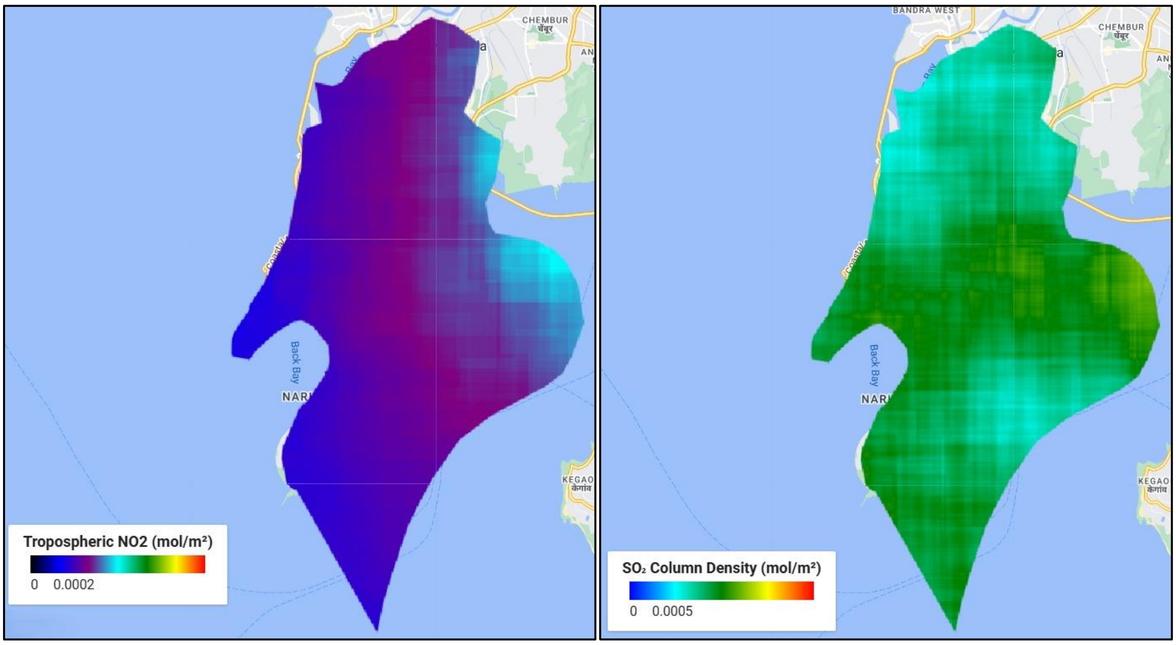


Figure 5: NO_2 and SO_2 concentration for Mumbai region for year 2024

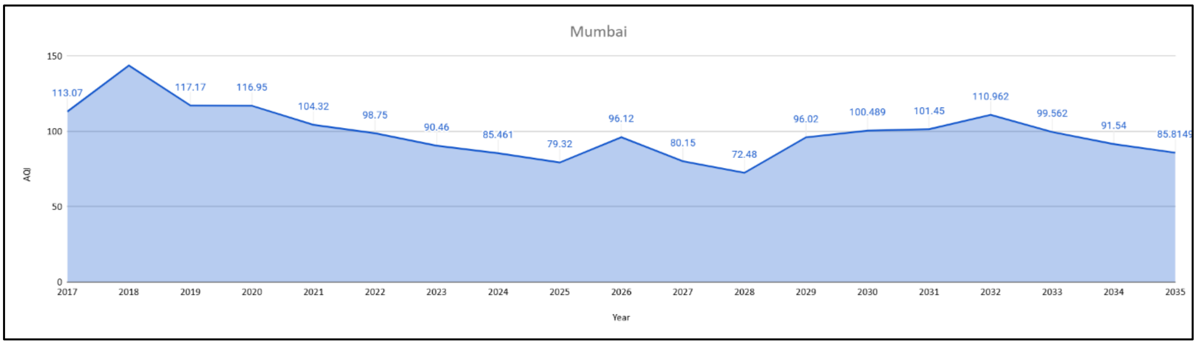


Figure 6: AQI Index values for Mumbai region from year 2017-2035

Figure 7 represents the NO_2 and SO_2 concentration for Satara region for year 2024. Figure 8 represents future predictions for AQI indexes of air quality from year 2017 to 2035.

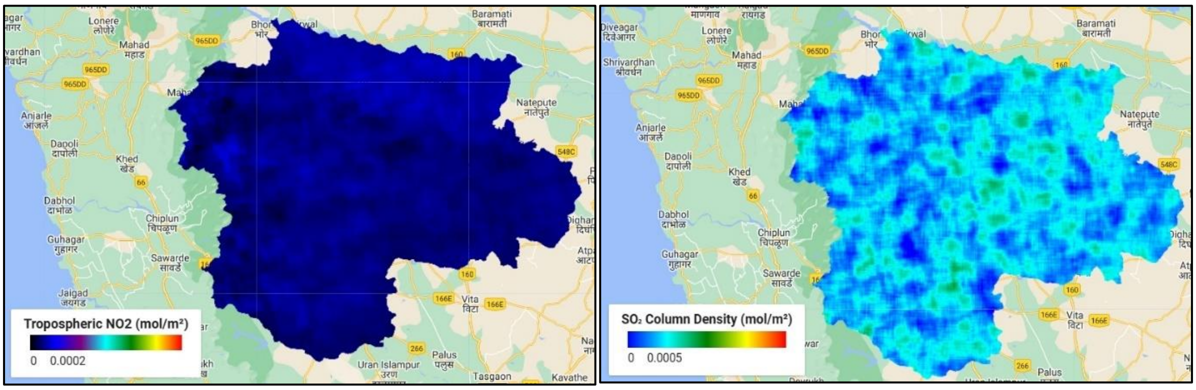


Figure 7: NO_2 and SO_2 concentration for Satara region for year 2024

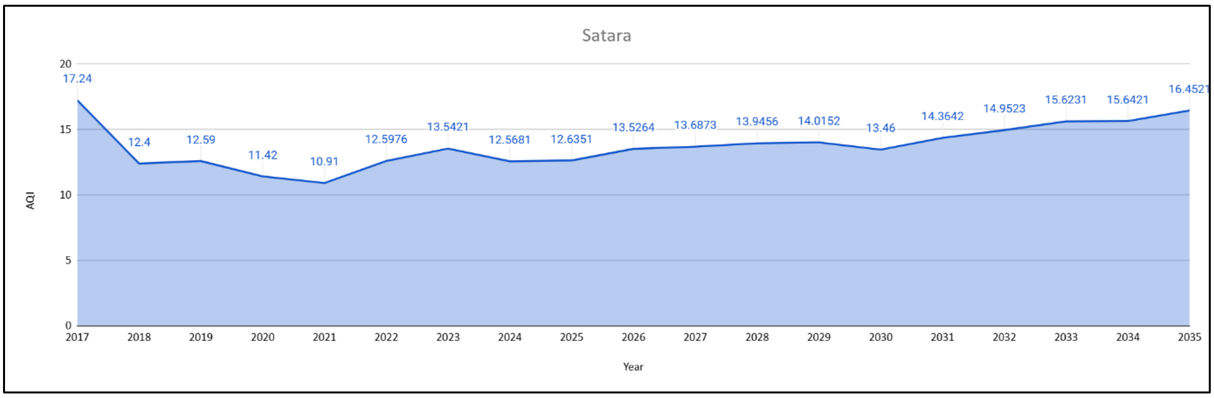


Figure 8: AQI Index values for Satara region from year 2017-2035

VII. RESULT ANALYSIS

7.1 Introduction

The performance of different machine learning models for the classification of Land Use Land Cover are compared over two different districts of Maharashtra, Mumbai and Satara, based on results reported here. For several years between 2022 and 2024, a key performance metric is assessed in the study: accuracy, Kappa Coefficient, Matthews Correlation Coefficient (MCC), and Jaccard Score. In addition, the performance of the models must be assessed in comparison with appropriate classification tasks for both thickly populated urban (Mumbai) and rural (Satara) conditions. Apart from this, the AQI data from 2017 to 2035 helps observe the trends and brings further attention toward how the air quality of Mumbai is worsening while in Satara, it is quite stable.

7.2 Performance Evaluation of LULC classification for Mumbai and Satara regions

Table 8 highlights the performance evaluation metrics for Mumbai District with MLP + LSTM model giving best results among all the years from 2022 to 2024.

Table 8: Performance Evaluation Metrics for Mumbai District

Year	Model	Accuracy (in %)	Kappa Coefficient	Matthews Correlation Coefficient (MCC)	Jaccard Score
2022	1D - CNN	97.10	0.9546	0.9550	0.9537
	MLP	97.20	0.9562	0.9566	0.9539
	LSTM	98.07	0.9698	0.9698	0.9677
	CNN+LSTM	97.47	0.9603	0.9605	0.9569
	CNN+MLP	97.70	0.9641	0.9641	0.9614
	MLP+LSTM	**98.17**	**0.9713**	**0.9714**	**0.9681**
2023	1D - CNN	97.79	0.9655	0.9657	0.9647
	MLP	97.15	0.9555	0.9560	0.9522

Year	Model	Accuracy (in %)	Kappa Coefficient	Matthews Correlation Coefficient (MCC)	Jaccard Score
	LSTM	97.90	0.9672	0.9674	0.9634
	CNN+LSTM	97.23	0.9567	0.9569	0.9529
	CNN+MLP	96.95	0.9525	0.9526	0.9511
	MLP+LSTM	**98.39**	**0.9748**	**0.9749**	**0.9716**
2024 (Till 30th September 2024)	1D - CNN	95.12	0.9237	0.9254	0.9266
	MLP	96.87	0.9511	0.9519	0.9482
	LSTM	97.54	0.9616	0.9619	0.9601
	CNN+LSTM	97.12	0.9550	0.9552	0.9514
	CNN+MLP	96.62	0.9472	0.9473	0.9451
	MLP+LSTM	**97.92**	**0.9674**	**0.9677**	**0.9631**

Table 9 highlights the performance evaluation metrics for Satara District with MLP model giving best results for year 2022 and 2023 and LSTM model giving best results for year 2024.

Table 9: Performance Evaluation Metrics for Satara District

Year	Model	Accuracy (in %)	Kappa Coefficient	Matthews Correlation Coefficient (MCC)	Jaccard Score
2022	1D - CNN	78.75	0.6961	0.7014	0.6833
	MLP	**80.08**	**0.7169**	**0.7227**	**0.7033**
	LSTM	77.24	0.6797	0.6844	0.6738
	CNN+LSTM	77.77	0.6875	0.6895	0.6750

Year	Model	Accuracy (in %)	Kappa Coefficient	Matthews Correlation Coefficient (MCC)	Jaccard Score
	CNN+MLP	76.63	0.6700	0.6719	0.6629
	MLP+LSTM	78.79	0.7055	0.7087	0.6919
2023	1D - CNN	78.86	0.7000	0.7061	0.6899
	MLP	**79.89**	**0.7099**	**0.7122**	**0.6880**
	LSTM	79.71	0.7121	0.7153	0.6889
	CNN+LSTM	78.41	0.6936	0.6967	0.6805
	CNN+MLP	77.46	0.6833	0.6852	0.6689
	MLP+LSTM	76.48	0.6676	0.6707	0.6517
2024 (Till 30th September 2024)	1D - CNN	79.16	0.7008	0.7057	0.6833
	MLP	80.17	0.7173	0.7219	0.6970
	LSTM	**81.38**	**0.7373**	**0.7391**	**0.7101**
	CNN+LSTM	77.36	0.6820	0.6839	0.6702
	CNN+MLP	80.67	0.7290	0.7315	0.7026
	MLP+LSTM	77.88	0.6856	0.6899	0.6664

7.3 Air Quality Future AQI Trends

Table 10 and Table 11 highlights the AQI index values for Mumbai and Satara regions from year 2017 – 2035 indicating future predictions of AQI index value highlighting Mumbai's worsening air quality as compared to the stable AQI index values of Satara.

Table 10: Future AQI air quality index values of Mumbai

Mumbai																			
Year	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
AQI	113.07	143.74	117.17	116.95	104.32	98.75	90.46	85.461	79.32	96.12	80.15	72.48	96.02	100.489	101.45	110.962	99.562	91.54	85.8149

Table 11: Future AQI air quality index values of Satara

Satara																			
Year	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035
AQI	17.24	12.4	12.59	11.42	10.91	12.5976	13.5421	12.5681	12.6351	13.5264	13.6873	13.9456	14.0152	13.46	14.3642	14.9523	15.6231	15.6421	16.4521

7.4 Comparative Analysis

The comparative analysis for model performance of Mumbai and Satara districts and their AQI trends for air quality are discussed below.

7.4.1 Model Performance for Mumbai District:

MLP + LSTM models performed the best for LULC classification of Mumbai and Satara regions with an accuracy of 98.17% in year 2022, 98.39% in year 2023 and 97.02% in year 2024 maintaining high Kappa, MCC and Jaccard Score values as indicated in Figure 9, 10 and 11. CNN-based models (CNN+LSTM, CNN+MLP) produced decent results, although they lagged behind MLP and LSTM, implying that hybrid or multi-layer models may be better suited for this dataset.

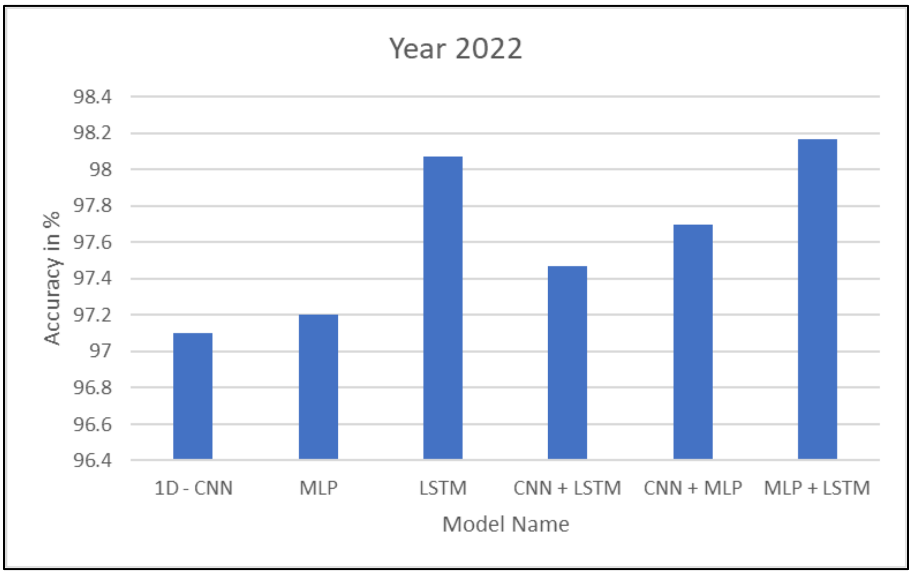


Figure 9: Comparison of accuracies for Mumbai region of year 2022

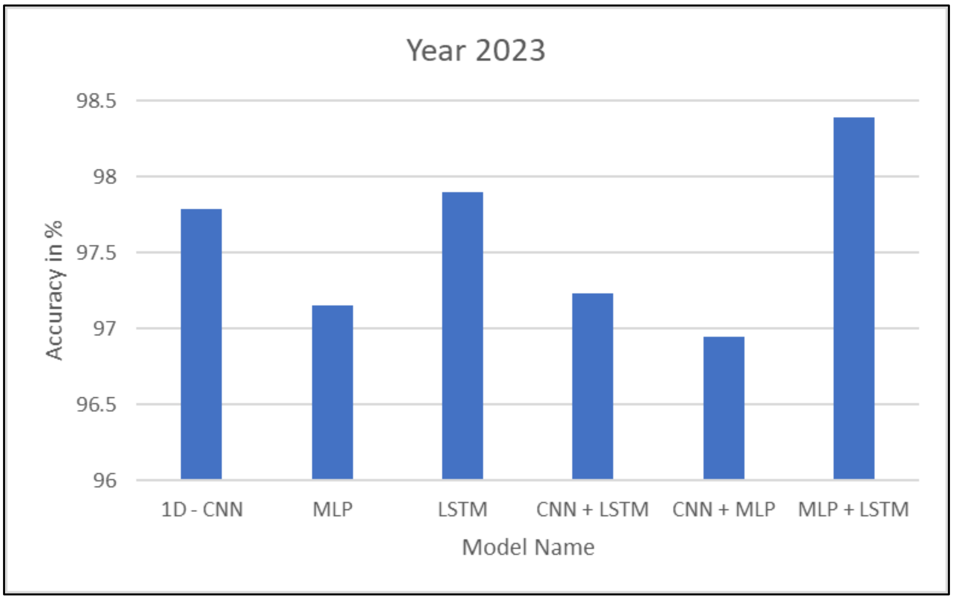


Figure 10: Comparison of accuracies for Mumbai region of year 2023

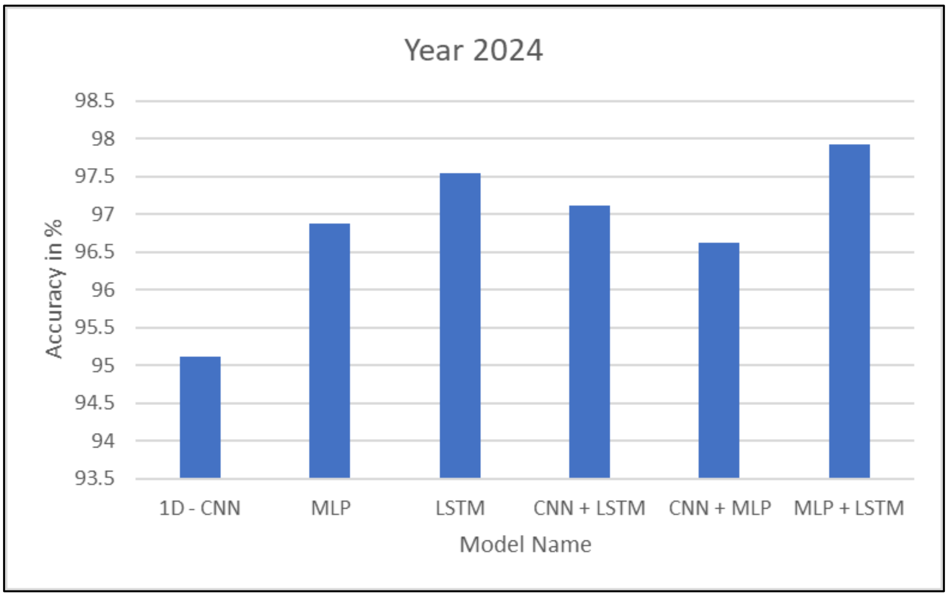


Figure 11: Comparison of accuracies for Mumbai region of year 2024

7.4.2 Model Performance for Satara District:

Accuracies for models in Satara were lower, with MLP model giving highest accuracy of 80.08% in year 2022 and 79.89% in year 2023 as shown in Figure 12, 13 and 14. In year 2024, LSTM model performed the best with the accuracy of 81.38% as shown in Figure 14. The performance of CNN-based hybrid models (CNN+LSTM and CNN+MLP) fluctuated, which could be attributed to poor rural data adaptation.

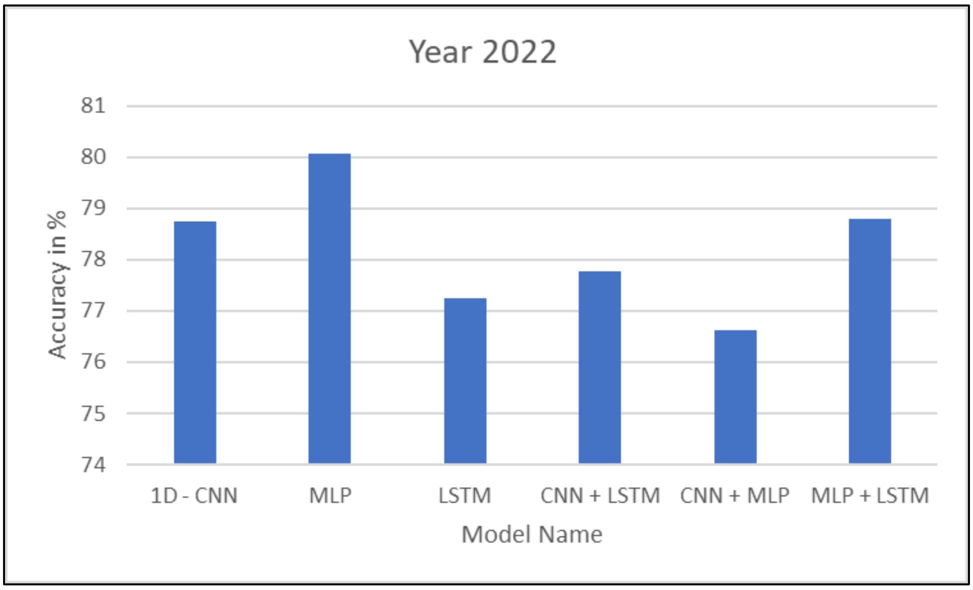


Figure 12: Comparison of accuracies for Satara region of year 2022

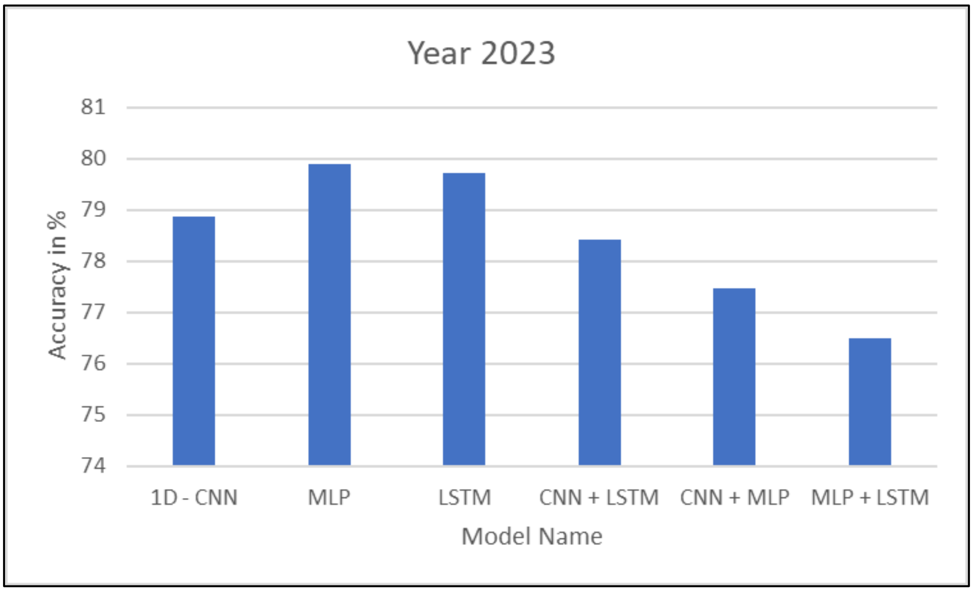


Figure 13: Comparison of accuracies for Satara region of year 2023

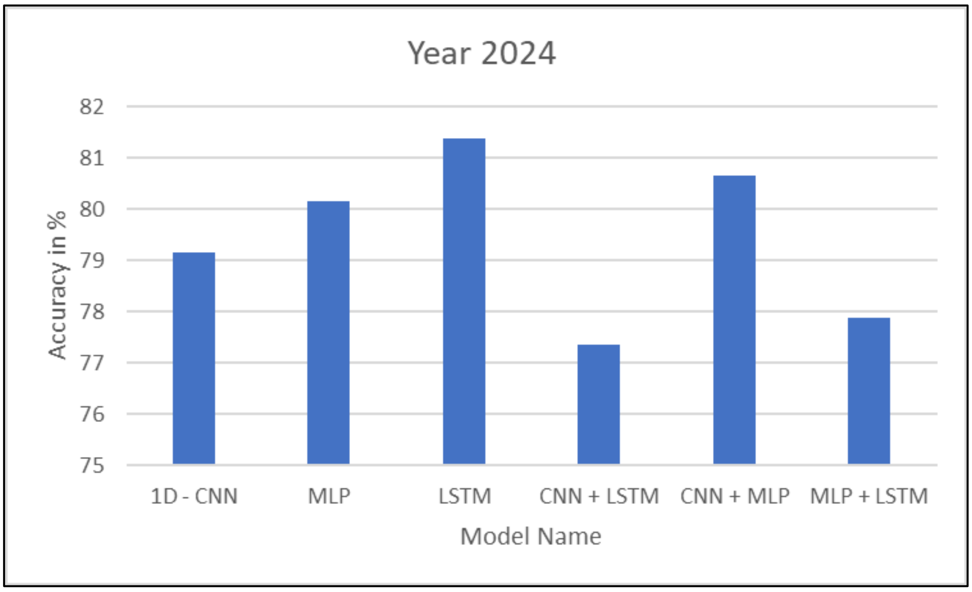


Figure 14: Comparison of accuracies for Satara region of year 2024

7.4.3 AQI trends for Mumbai and Satara

From 2017 to 2035, the AQI index values of Mumbai and Satara, along with the projected values for the future, vary in trends of air quality. The hybrid GRU+GCN model was implemented and the value obtained for MAE is 0.144, MSE is 0.029 and R^2 is 0.964. Mumbai's AQI showed consistent improvement until 2024, apparently because of efficient control of air quality, but fluctuations in the following years prove an impact of continuous urbanization. On the other hand, Satara maintains almost persistently low AQI readings, pointing to negligible industrial pollution, with a minor rising trend beginning from 2030, pointing toward emerging influences on rural air quality.

VIII. CONCLUSION & FUTURE SCOPE

LULC classification and air quality analysis across the Mumbai and Satara districts show how well machine learning models perform with datasets from urban and rural settings. High accuracy and stability of the LULC classification results have been reported for advanced models, such as LSTM and MLP+LSTM suitable for urban environments like the Mumbai district. However, alternative and simpler models like MLP would reveal competing performance for Satara, where landscape variability and resolution of data may challenge performance in complex models. Further, the AQI analysis suggests that although the level of air quality in Mumbai has improved, over time, through such recent management interventions, it is still susceptible to fluctuations based on urban expansion. Satara indicates a relatively stable AQI that tends upward. Such can be an indication of the activities resulting from increased industrial or agricultural practices.

The future scope is as follows:

a) **Model Optimization and Transfer Learning:** The future work may thus focus on optimization of model architectures so that it can better adapt to different geographic settings. Transfer learning techniques may be utilized in order to leverage the model performance on rural datasets, where the availability of data might be scarce.

b) **High Resolution Datasets:** By utilizing the high-resolution data sets, higher accuracy in classification for LULC classification in mostly rural areas will be achieved, where most types of landscapes may become a mix. These combined data sources may also further improve the accuracy of predicting air quality.

c) **Time Series Graphs:** Time series on the changes in LULC and AQI are crucial for identifying long term trends and causal relationships existing between land use and the characteristics of air quality. Therefore, this will be vital while building predictive models so proactive policy measures can be considered in advance.

d) **Integration with policy planning:** These findings can then be incorporated into the policy of local authorities and urban planners regarding planning at the rural and urban development levels. Data-driven recommendations could perhaps reduce the air quality impacts while managing the change in land uses.

e) **Climate and Environmental Variables:** Additional climate-related variables such as temperature and humidity can be included in future studies to see how LULC interact with larger environmental factors. In this way, it will be helpful in giving a more comprehensive appraisal of the effect of the environment.

The summary of general points discussed above highlight the avenues of extension of the considered fields in future studies with the probable contribution to the development of the applicability of LULC classification models and AQI analysis in pursuit of sustainable urban and rural development.

REFERENCES

[1] K. H. Jodhani et al., “Synergizing google earth engine and earth observations for potential impact of land use land cover on air quality,” Results in engineering, vol. 22, pp. 102039–102039, Jun. 2024, doi: https://doi.org/10.1016/j.rineng.2024.102039.

[2] Ajay Kumar Taloor, S. Sharma, Gurnam Parsad, and Rakesh Jasrotia, “Land use land cover simulation using integrated CA-Markov model in the Tawi Basin of Jammu and Kashmir India,” Geosystems and Geoenvironment, vol. 3, no. 2, pp. 100268–100268, May 2024, doi: https://doi.org/10.1016/j.geogeo.2024.100268.

[3] Chisanga, C.B., Phiri, D. & Mubanga, K.H. Multi-decade land cover/land use dynamics and future predictions for Zambia: 2000–2030. Discov Environ 2, 38 (2024). https://doi.org/10.1007/s44274-024-00066-w

[4] Osman MAA, Abdel-Rahman EM, Onono JO, Olaka LA, Elhag MM, Adan M, Tonnang HEZ. Mapping, intensities and future prediction of land use/land cover dynamics using google earth engine and CA-artificial neural network model. PLoS One. 2023 Jul 24;18(7):e0288694. doi: 10.1371/journal.pone.0288694. PMID: 37486922; PMCID: PMC10365312.

[5] Matci, D.K., Kaplan, G. & Avdan, U. Changes in air quality over different land covers associated with COVID-19 in Turkey aided by GEE. Environ Monit Assess 194, 762 (2022). https://doi.org/10.1007/s10661-022-10444-7

[6] Yang, W., Jiang, X. Evaluating the influence of land use and land cover change on fine particulate matter. Sci Rep 11, 17612 (2021). https://doi.org/10.1038/s41598-021-97088-8

[7] "Assessing the Impact of Land Use and Land Cover Change on Air Quality in East Baton Rouge—Louisiana Using Earth Observation Techniques" written by Diana B. Frimpong, Yaw A. Twumasi, Zhu H. Ning, Abena B. Asare-Ansah, Matilda Anokye, Priscilla M. Loh, Faustina Owusu, Caroline Y. Apraku, Recheal N. D. Armah, Judith Oppong, John B. Namwamba, published by Advances in Remote Sensing, Vol.11 No.3, 2022"

[8] Gupta, P., Shukla, A. K., & Shukla, D. P. (2024). ML-Based hybrid SAR and Optical Image LULC mapping and change analysis with variations in the air quality of the Imphal Valley, North-East India. Earth and Space Science, 11(3). https://doi.org/10.1029/2023ea003176

[9] Lu, Y., Yang, X., Wang, H., Jiang, M., Wen, X., Zhang, X., & Meng, L. (2023). Exploring the effects of land use and land cover changes on meteorology and air quality over Sichuan Basin, southwestern China. Frontiers in Ecology and Evolution, 11. https://doi.org/10.3389/fevo.2023.1131389

[10] Bui, Q.-T.; Chou, T.-Y.; Hoang, T.-V.; Fang, Y.-M.; Mu, C.-Y.; Huang, P.-H.; Pham, V.-D.; Nguyen, Q.-H.; Anh, D.T.N.; Pham, V.-M.; et al. Gradient Boosting Machine and Object-Based CNN for Land Cover Classification. Remote Sens. 2021, 13, 2709. https://doi.org/10.3390/rs13142709

[11] Abdi, A. M. (2019). Land cover and land use classification performance of machine learning algorithms in a boreal landscape using Sentinel-2 data. GIScience & Remote Sensing, 57(1), 1–20. https://doi.org/10.1080/15481603.2019.1650447

[12] Beshir, S., Moges, A., & Dananto, M. (2023). Trend analysis, past dynamics and future prediction of land use and land cover change in upper Wabe-Shebele river basin. Heliyon, 9(9), e19128. https://doi.org/10.1016/j.heliyon.2023.e19128

[13] Faisal, A., Rahman, M. M., & Haque, S. (2021). Retrieving spatial variation of aerosol level over urban mixed land surfaces using Landsat imageries: Degree of air pollution in Dhaka Metropolitan Area. Physics and Chemistry of the Earth Parts a/B/C, 126, 103074. https://doi.org/10.1016/j.pce.2021.103074

[14] Mampitiya, L., Rathnayake, N., Hoshino, Y., & Rathnayake, U. (2023). Forecasting PM_{10} levels in Sri Lanka: A comparative analysis of machine learning models PM_{10}. Journal of Hazardous Materials Advances, 13, 100395. https://doi.org/10.1016/j.hazadv.2023.100395

[15] H. Du, M. Li, Y. Xu and C. Zhou, "An Ensemble Learning Approach for Land Use/Land Cover Classification of Arid Regions for Climate Simulation: A Case Study of Xinjiang, Northwest China," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 16, pp. 2413-2426, 2023

[16] J. Yuan, L. Ru, S. Wang and C. Wu, "WH-MAVS: A Novel Dataset and Deep Learning Benchmark for Multiple Land Use and Land Cover Applications," in IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 15, pp. 1575-1590, 2022, doi: 10.1109/JSTARS.2022.3142898.

[17] Rawat, S., & Saini, R. (2023). Evaluating the impact of sampling designs on the performance of machine learning techniques for land use land cover classification using Sentinel-2 data. International Journal of Remote Sensing, 44(24), 7889–7908. https://doi.org/10.1080/01431161.2023.2290994

[18] Puttinaovarat, S., Khaimook, K., & Horkaew, P. (2023). Land use and land cover classification from satellite images based on ensemble machine learning and crowdsourcing data verification. International Journal of Cartography, 1–21. https://doi.org/10.1080/23729333.2023.2166252

[19] Pan, X., Wang, Z., Gao, Y., Dang, X., & Han, Y. (2021). Detailed and automated classification of land use/land cover using machine learning algorithms in Google Earth Engine. Geocarto International, 37(18), 5415–5432. https://doi.org/10.1080/10106049.2021.1917005

[20] Li, Y., & Myint, S. W. (2021). Fine resolution air quality dynamics related to socioeconomic and land use factors in the most polluted desert metropolitan in the American Southwest. The Science of the Total Environment, 788, 147713. https://doi.org/10.1016/j.scitotenv.2021.147713

[21] Thakrar, S. K., Johnson, J. A., & Polasky, S. (2023). Land-Use Decisions Have Substantial Air Quality Health Effects. Environmental Science & Technology, 58(1), 381–390. https://doi.org/10.1021/acs.est.3c02280

[22] Araki, S., Shima, M., & Yamamoto, K. (2020). Estimating historical $PM_{2.5}$ exposures for three decades (1987–2016) in Japan using measurements of associated air pollutants and land use regression. Environmental Pollution, 263, 114476. https://doi.org/10.1016/j.envpol.2020.114476

[23] Xiao, F., Yang, M., Fan, H. et al. An improved deep learning model for predicting daily PM2.5 concentration. Sci Rep 10, 20988 (2020). https://doi.org/10.1038/s41598-020-77757-w

[24] Dilawar, A., Chen, B., Ul-Haq, Z., Ali, S., Sajjad, M. M., Junjun, F., Gemechu, T. M., Guo, M., Dilawar, H., Zhang, H., Zicheng, Z., & Lodhi, E. (2024). Evaluating the potential footprints of land use and land cover and climate dynamics on atmospheric pollution in Pakistan. Frontiers in Environmental Science, 11. https://doi.org/10.3389/fenvs.2023.1272155

[25] Ruidas, D., Pal, S.C. Potential hotspot modeling and monitoring of $PM_{2.5}$ concentration for sustainable environmental health in Maharashtra, India. Sustain. Water Resour. Manag. 8, 98 (2022). https://doi.org/10.1007/s40899-022-00682-5