The Growing Threats of Artificial Lights in India: Ecological Implications and Predictive Analysis

A Project

Submitted in partial fulfillment of the requirements for

the award of the Degree of

BACHELOR OF COMPUTER APPLICATION

By

SHANTANU DUTTA

ROLL NO-147 AND REGISTRATION NO-223661010152

SHASHANK SHEKHAR

ROLL NO-148 AND REGISTRATION NO-223661010153

SRIJAN BANERJEE

ROLL NO-177 AND REGISTRATION NO-223661010182

SRIJATA BANERJEE

ROLL NO-179 AND REGISTRATION NO-223661010184

TRISHIKH BHATTACHARYYA

ROLL NO-200 AND REGISTRATION NO-223661010206

DEPARTMENT OF COMPUTER APPLICATION

INSTITUTE OF ENGINEERING & MANAGEMENT

2025

DECLARATION CERTIFICATE

This is to certify that the work presented in the thesis entitled **"The Growing Threats of Artificial Lights in India: Ecological Implications and Predictive Analysis"** in partial fulfillment of the requirement for the award of degree of **Bachelor of Computer Application** of Institute of Engineering & Management is an authentic work carried out under my supervision and guidance.

To the best of my knowledge the content of this thesis does not form a basis for the award of any previous Degree to anyone else.

Date:

Prof. Saikat Mondal
Dept. of Computer Application
Institute of Engineering & Management

Prof. Dr. Rupam Bhattacharya
Head of the Department
Dept. of Computer Application
Institute of Engineering & Management

CERTIFICATE OF APPROVAL

The foregoing thesis entitled **“The Growing Threats of Artificial Lights in India: Ecological Implications and Predictive Analysis”** is hereby approved as a creditable study of research topic and has been presented in satisfactory manner to warrant its acceptance as prerequisite to the degree for which it has been submitted.

It is understood that by this approval, the undersigned do not necessarily endorse any conclusion drawn or opinion expressed therein, but approve the thesis for the purpose for which it is submitted.

(Internal Examiner) **(External Examiner)**

Acknowledgements

We would like to express our special thanks of gratitude to our Guide Prof. Saikat Mondal who helped us a lot in this project, his valuable suggestions helped us to solve tough challenges and without his help this project could not have been completed in time. Secondly, we would like to thank all our respected faculty members and our friends who helped us a lot in finalizing this project within the given time frame.

Name of Student: SHANTANU DUTTA

Enrollment Number: 12022004006152

Name of Student: SHASHANK SHEKHAR

Enrollment Number: 12022004006153

Name of Student: SRIJAN BANERJEE

Enrollment Number: 12022004006182

Name of Student: SRIJATA BANERJEE

Enrollment Number: 12022004006184

Name of Student: TRISHIKH BHATTACHARYYA

Enrollment Number: 12022004006206

Contents

Abstract

Our project comprises three modules aimed at analyzing the growing issue of artificial light pollution and its effects on urban India and the ecosystem. The first module, developed as part of our minor project, extracts night-time light data from Indian cities using Google Earth Engine (GEE) to visualize light intensity and urban expansion over time.

The major project adds two more modules. The second module uses machine learning to forecast light pollution trends and identify key contributing factors, helping to understand how and why certain areas are more affected than others.

The final module includes a visualization dashboard that displays graphs, maps, and patterns in an interactive format. This allows users to explore pollution levels and their ecological impact. Together, these modules provide a practical, data-driven approach to studying and addressing light pollution in India.

Chapter 1

1.1 Introduction

Light pollution, defined as the excessive or misdirected artificial illumination that disrupts natural light-dark cycles, has become a significant environmental issue in recent decades. India has witnessed a substantial increase in artificially lit areas, with an annual growth rate of 1.07–1.09% between 2012 and 2016 and a staggering 26% yearly increase from 1993 to 2013. This rapid rise in light pollution has far-reaching consequences, affecting biodiversity, human health, and ecological stability. Disruptions in natural photoperiods alter the behavior of nocturnal species, interfere with plant growth cycles, and contribute to energy waste and carbon emissions.

This study aims to analyze and map trends in light pollution across India at both national and sub-national levels from 2013 to 2021. By utilizing satellite datasets such as VIIRS-DNB (Visible Infrared Imaging Radiometer Suite Day/Night Band) and MODIS Land Cover Type data, we seek to quantify changes in nighttime light emissions and assess their impact on urban and agricultural landscapes. Additionally, the study examines key drivers of light pollution, including urbanization, industrialization, population growth, and inadequate lighting regulations, while evaluating their ecological and environmental implications.

To achieve these objectives, the research employs geospatial analysis techniques using Google Earth Engine (GEE) and QGIS. This methodology involves delineating regional boundaries, extracting statistical metrics, and conducting time-series analyses to track changes in nighttime light intensity. Machine learning (ML) techniques further enhance the accuracy and depth of the analysis by addressing key challenges in large-scale environmental data processing.

One major challenge in studying light pollution is differentiating between land cover types such as urban areas and agricultural lands. Supervised classification methods, including Random Forest and Support Vector Machines (SVM), categorize satellite images based on spectral signatures, enabling accurate identification of affected regions. To predict future trends, regression-based ML models analyze historical nighttime light data, with time-series forecasting algorithms like Long Short-Term Memory (LSTM) networks capturing temporal patterns to anticipate future hotspots. Clustering techniques such as K-Means categorize regions with similar characteristics to facilitate targeted interventions. Furthermore, anomaly detection methods like Isolation Forests are used to identify areas with unusually high levels of artificial light, ensuring efficient hotspot detection.

By integrating remote sensing techniques with advanced ML tools, this research provides a comprehensive understanding of light pollution trends and their ecological consequences. The findings can inform targeted interventions to mitigate light pollution, support sustainable urban planning, and contribute to biodiversity conservation. Ultimately, this study underscores the need for policies that balance development with environmental sustainability, positioning light pollution management as a critical aspect of global conservation efforts.

Chapter 2

2.1 Literature Survey

1. Light Pollution and Ecological Impacts

Light pollution disrupts natural light cycles, with severe consequences for ecosystems and biodiversity:

- Avian Navigation: In *Ecology and Society (2008)* [1] found that nocturnally migrating birds are disoriented by white and red light, which interferes with celestial navigation. Green light, however, caused minimal disruption, suggesting its use in urban lighting to reduce collisions.
- Marine Ecosystems: A study by authors from the University of Seville (2024) in *Marine Environmental Research* [2] revealed that artificial light disrupts vertical migration in coastal species like amphipods and polychaetes, threatening soft-bottom marine biodiversity.
- Terrestrial Fauna: In *Applied Animal Behaviour Science (2024)* [3] demonstrated that light and noise pollution shorten the dusk chorus duration of birds and alter vocal behaviour, exacerbating stress in terrestrial species.

2. Light Pollution and Human Health

- Circadian Rhythm Disruption: *Analysis of Light Pollution in Ticino region during the period 2011-2016* [4] links prolonged artificial light exposure to suppressed melatonin production, increasing risks of sleep disorders, metabolic diseases, and cancers like breast and prostate cancer.

- Public Awareness: A 2022 survey in India *Studying light pollution as an emerging environmental concern in India (2022)* [5] highlighted low awareness of light pollution's consequences, despite rapid urbanization driving a 26% annual increase in lit areas since 2013.

3. Machine Learning in Light Pollution Research

ML techniques advance the analysis and mitigation of light pollution:

- Predictive Modeling: LSTM networks forecast light intensity trends, while Random Forest algorithms identify drivers like urbanization and population density.
- Spatial Clustering: K-Means and DBSCAN algorithms map pollution hotspots, aiding targeted interventions.
- Geospatial Tools: Google Earth Engine (GEE) and MODIS data quantify lit area expansion.

4. Cross-Border and Transboundary Impacts

- Binational Regions: In *Environmental Challenges (2024)* [6] analysed light pollution in the California-Mexico border region, showing urban centres like Tijuana contribute 21% of pollution affecting protected areas. VIIRS data revealed light pollution spreads up to 300 km, necessitating transboundary policies.
- Mitigation Strategies: Legal frameworks in Mexico and the U.S. [6] define "influence zones" around reserves but lack ecological conservation measures for migratory pathways.

5. Research Gap

A synthesis of the key gaps from previous studies highlights the following research voids:

- Cumulative Impact Analysis: Most studies examine isolated species or specific regions, lacking comprehensive longitudinal assessments of light pollution's ecological effects.
- Policy Integration: Current frameworks prioritize economic growth over sustainable lighting policies, neglecting ecological and health costs.
- Machine Learning Model Integration: Few studies incorporate multi-model frameworks that integrate predictive analytics, clustering, and geospatial mapping simultaneously.
- India-Specific Analysis: While global research on light pollution exists, studies focusing on India's long-term trends (2013–2021) and regional disparities remain limited.
- Transboundary Strategies: There is a lack of cross-border collaboration addressing the global ecological footprint of light pollution.

6. Contributions of This Work

Our project addresses these gaps by:

- Multi-Model Integration: Combining LSTM for prediction, Random Forest for driver analysis, and clustering for hotspot identification.
- India-Specific Longitudinal Analysis: Analysing 2013–2021 data highlights regional disparities (e.g., West Bengal's urbanization) and informs policy.
- Policy-Driven Insights: Identifies actionable strategies, such as green-lighting for migratory birds and zoning laws for coastal ecosystems.

Chapter 3

3.1 Experimental Dataset

Dataset Description

The dataset used in this research primarily consists of historical light pollution data collected from 2013 to 2021. It represents the intensity of artificial light emissions across various geographical regions, helping analyse the impact of urbanization and environmental changes on light pollution trends.

Sources of Data

The data has been compiled from multiple authoritative sources, including:

- **Administrative Boundaries**: Acquired from geo boundaries within Google Earth Engine (GEE), specifically the CGAZ ADM1 feature collection, which provides regional administrative boundaries for analysis.
- **Nighttime Light (NTL) Collection**: Extracted from the **NOAA/VIIRS/DNB/ANNUAL V21** image collection in GEE. This dataset includes annual composites of nighttime light data, serving as a proxy for human activity.
- **MODIS Land Cover**: Obtained from the **MODIS/061/MCD12Q1** image collection in GEE, which provides land cover classification data to distinguish between urban and agricultural areas.
- **GSDP Data**: Economic and statistical data related to different regions, obtained from the **e-Sankhyiki Portal**:
    - **Main Website:** https://esankhyiki.mospi.gov.in/catalogue-main/catalogue

- Datasheet Page: https://esankhyiki.mospi.gov.in/catalogue-main/catalogue?page=0&search=tableno&q=SDPAFY24140ANN&product=

- **Area Data**: Geographic area information sourced from the **Know India Portal**:
  - **Know India - States and UTs:** https://knowindia.india.gov.in/states-uts/

Relevance to the Project

This dataset is fundamental for training and testing our machine learning models, allowing us to predict future light pollution trends and assess their environmental impact. By integrating multiple models, we aim to improve forecasting accuracy and detect anomalies in artificial light proliferation.

Data Collection Process

The dataset was compiled using a combination of automated data extraction methods and publicly available repositories.

Data Gathering Methods

- **Google Earth Engine (GEE) Data Extraction**: Administrative boundaries, nighttime light composites, and land cover classifications were retrieved using geospatial processing.
- **Public Datasets**: Official environmental reports and published datasets on light pollution.

- **Data Aggregation from Studies**: Statistical data collected from various research papers and studies.

Preprocessing Steps

Before using the data for training, several preprocessing steps were undertaken:

- **Data Cleaning**: Removal of duplicate and inconsistent entries.
- **Normalization**: Standardization of brightness values for uniform comparison.
- **Handling Missing Values**: Using interpolation and statistical methods to fill gaps in data.

Features and Labels

The dataset contains multiple key variables that contribute to understanding and predicting light pollution trends.

Key Variables (Features)

- **Geographical Location**: Identifies the specific area where data was collected.
- **Brightness Levels**: Measures artificial light intensity.
- **Yearly Trends**: Captures the temporal changes in light pollution from 2013 to 2021.
- **Population Density**: Used to correlate human activity with light pollution.
- **Urbanization Rate**: Determines the expansion of urban areas affecting brightness levels.
- **Land Cover Classification**: Differentiates between urban and agricultural regions.

Target Variable (Label)

- **Light Pollution Severity Level**: Categorized to assess the impact intensity of artificial light emissions.

Dataset Size and Structure

The dataset used for this study is structured to facilitate machine learning model training and testing.

Size of Dataset

- **Number of Samples**: The dataset contains a substantial number of records covering multiple geographical regions within India.
- **Time Frame**: Data spans from **2013 to 2021**, ensuring sufficient historical trend analysis.

Data Format

- **File Format**: The dataset is stored in **CSV format** for ease of processing.
- **Structure**:
  - **Rows**: Each row corresponds to a data entry for a specific year and location.
  - **Columns**: Represent features such as brightness levels, urbanization rate, and land cover classification.

Challenges and Limitations

While compiling and using this dataset, several challenges were encountered.

Potential Biases and Missing Data

- **Satellite Data Limitations**: Variations in satellite sensitivity could lead to inconsistencies.

- **Data Gaps**: Some years may have incomplete records due to missing satellite readings.
- **Urbanization Influence**: Light pollution variations in rural versus urban areas may create data imbalances.

Mitigation Strategies

- **Data Augmentation**: Applying statistical techniques to interpolate missing values.
- **Normalization Techniques**: Ensuring uniform data distribution for unbiased model training.
- **Cross-Validation**: Employing multiple sources to verify data accuracy.

By leveraging the dataset and integrating multiple machine learning models, our research provides valuable insights into the long-term effects of light pollution, helping environmental researchers and policymakers take informed actions.

Chapter 4

4.1 Proposed Methodology

1. Data Acquisition and Preprocessing

a) Datasets

The study employs four primary datasets to analyze light pollution trends and their environmental impacts:

1. **Administrative Boundaries:**

   Acquired from the CGAZ ADM1 feature collection within Google Earth Engine (GEE), this dataset provides administrative boundaries for regional analysis at the state and district levels.

2. **Nighttime Light (NTL) Collection:**

   The NOAA/VIIRS/DNB/ANNUAL V21 image collection from GEE includes annual composites of nighttime light data. This dataset is used as a proxy for human activity and urbanization, with a spatial resolution of 500m and global coverage (-65 to 75 degrees latitude).

3. **MODIS Land Cover:**

   The MODIS/061/MCD12Q1 image collection from GEE provides land cover classification data essential for distinguishing between urban and agricultural areas. This dataset has a spatial resolution of 500m and temporal coverage from 2001 to 2021.

4. **Area Dataset:**

   An additional dataset containing the geographic area of each state in square kilometers is integrated into the analysis. This dataset ensures accurate normalization of light emissions by accounting for the size of each region, allowing for meaningful comparisons across states with varying geographical extents.

b) Region Selection

The focus area for this study is India. Administrative boundaries are filtered from the global dataset to extract the specific geometry pertinent to this region for targeted analysis. State-level and district-level boundaries are used to delineate regions of interest.

c) Temporal Range

The analysis covers the period from 2013 to 2021, offering a decade's worth of data to assess trends in light emissions and land cover changes.

d) Data Processing

1. **Nighttime Light Data:**

   The NTL data is filtered according to the specified date range (2013–2021) and the region's geometry. The 'average' band, representing the mean light emission for each year, is selected. This collection is then used to track temporal changes in light intensity.

2. **Land Cover Classification:**

   MODIS Land Cover data is processed annually to distinguish urban from agricultural land classes. A custom function (**extractClasses**) categorizes areas as urban (class value 13) and agricultural (classes 10, 12, and 14). These classifications are used to isolate nighttime light emissions specific to urban and agricultural regions.

3. **Area Dataset Integration:**

   The geographic area (in sq. km) for each state, obtained from the additional dataset, is integrated into the analysis. This allows for normalizing light emissions (e.g., Sum of Lights per unit area) and ensures that comparisons across states account for differences in geographic size.

4. **Sum of Lights (SOL) Calculation:**

The Sum of Lights (SOL) is calculated for urban and agricultural areas using the **reduce Region** method with a summation reducer. SOL values are computed annually for each region, and the time series data generated is used to analyse trends in nighttime light intensity.

2. Environmental Impact Assessment

An urban expansion assessment was conducted to evaluate environmental impacts. Year-over-year changes in urban SOL were analysed:

- A positive change indicated potential negative impacts, such as habitat disruption and increased energy consumption.
- No change or a decrease was labelled as stable/low impact.

3. Integration of Machine Learning

Machine learning algorithms enhanced the analysis:

- **Land Cover Classification:** Supervised models (e.g., Random Forest, SVM) improved the accuracy of urban and agricultural area classification.
- **Predictive Modeling:** Time-series forecasting (e.g., LSTM networks) predicted future light pollution trends.
- **Hotspot Detection:** Clustering (e.g., K-Means) and anomaly detection (e.g., Isolation Forests) identified ecologically vulnerable zones.

4. <u>Visualization</u>

Results were visualized using time-series graphs, geospatial maps, and environmental impact charts. Outputs included:

- Trends in nighttime light intensity across urban and agricultural landscapes.
- Maps highlighting spatial patterns of light pollution.
- Assessments of environmental impacts.

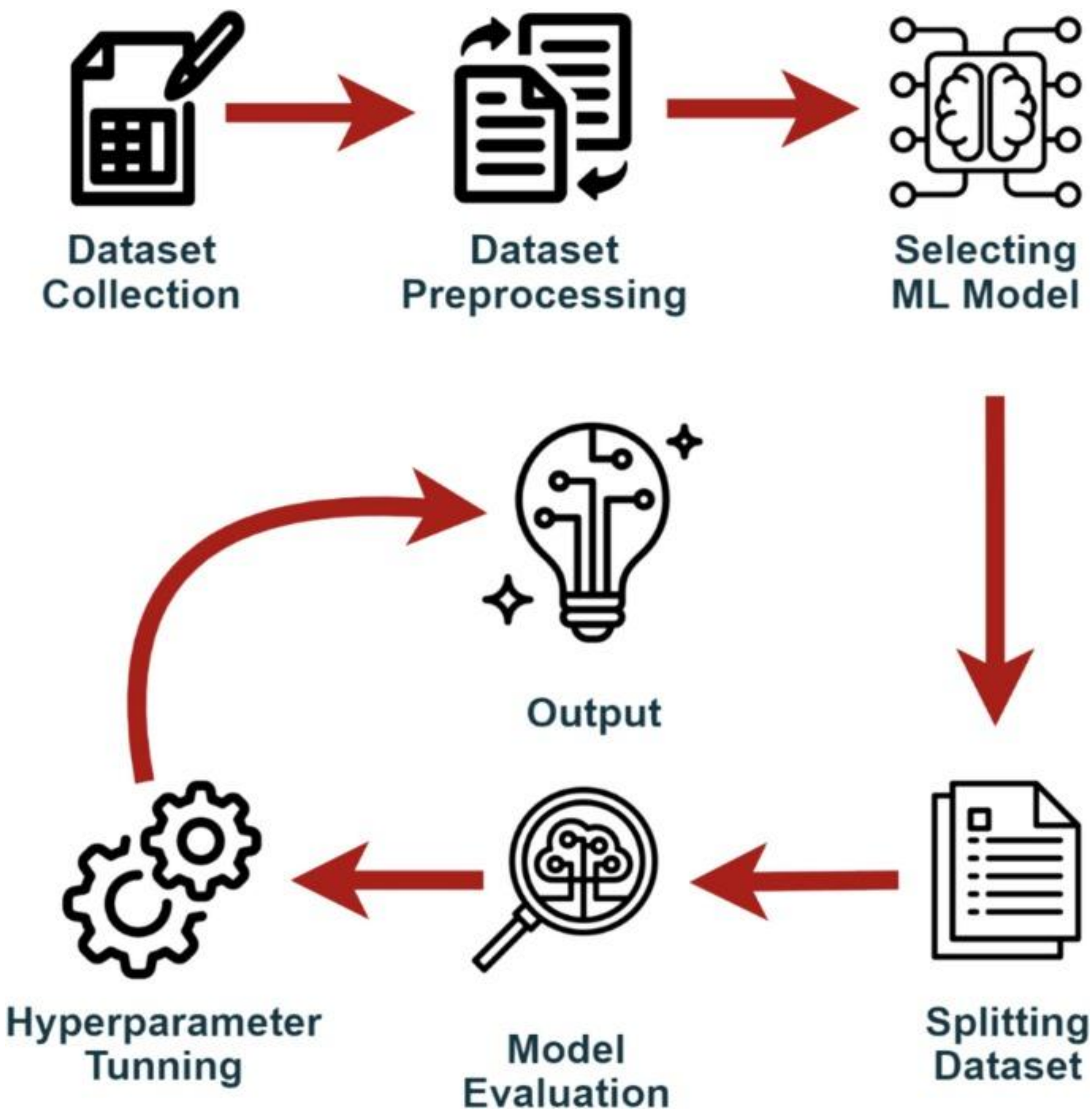


Figure 1.1: Methodology

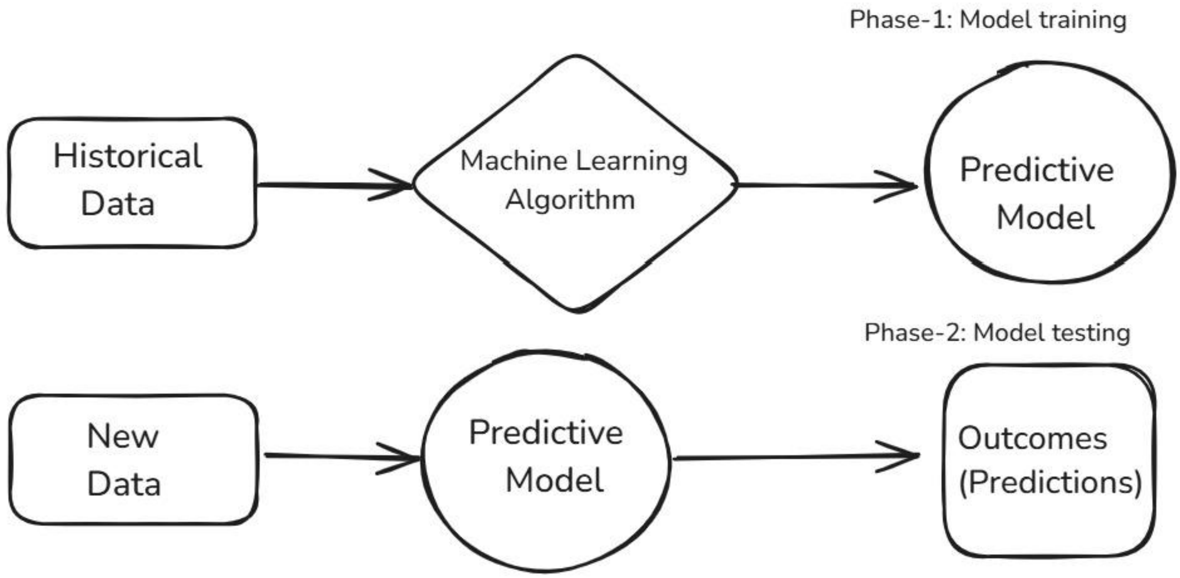


Figure 1.2: Model training & testing

5. <u>Tools and Limitations</u>

The analysis was conducted using cloud-based platforms (GEE and QGIS), ensuring scalability and reproducibility. Limitations include the reliance on satellite data, which may not capture localized factors, and the need for field research to validate findings.

Chapter 5

5.1 Result and Discussions

Results Presentation

The application of multiple machine learning models to analyse light pollution trends has yielded significant insights into the patterns and contributing factors of artificial light emissions. The models were trained on historical data from 2013 to 2021 and evaluated based on their ability to predict light pollution levels, classify regions based on pollution intensity, and identify anomalies in artificial brightness distribution.

Key Findings

- **Long Short-Term Memory (LSTM)**:

  As a recurrent neural network (RNN) model specialized for sequential data, LSTM effectively captured long-term dependencies in light pollution trends. The model demonstrated strong predictive capabilities, with an **R-squared value exceeding 0.85**, indicating a high degree of accuracy in forecasting future light pollution levels. The model's performance was validated against historical data, showing a minimal error rate in predicting brightness intensity levels.

Forecast for Delhi

2025: Urban_SOL = 169331.09, Agriculture_SOL = 26845.75

2026: Urban_SOL = 168902.95, Agriculture_SOL = 26964.40

ℹ️ In nW/cm²/sr, higher values indicate more light pollution.

The image above presents the predicted light pollution levels for Delhi for the years 2025 and 2026 using a Long Short-Term Memory (LSTM) model trained on historical satellite data.

The model outputs two key metrics:

- **Urban_SOL:** The predicted radiance from urban sources.
- **Agriculture_SOL:** The predicted radiance from agricultural areas.

📊 Model Evaluation

Training MSE: 0.008107 ✅ (Good)

Testing MSE: 0.208933 ✅ (Good)

ℹ️ Lower MSE means better prediction accuracy.

📈 Trend Analysis (Compared to Latest Available Data)

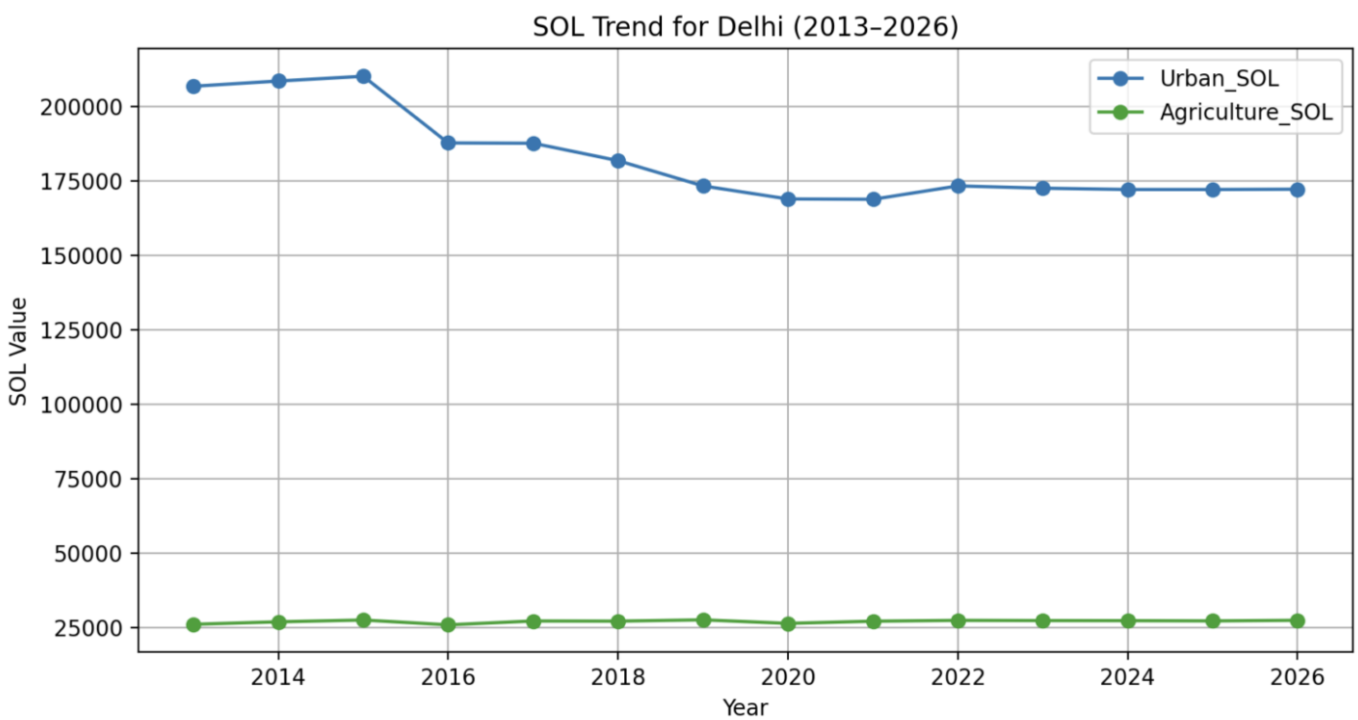
Urban_SOL: 📈 Increasing

Agriculture_SOL: 📈 Increasing

The image summarizes the performance metrics and trend observations of the LSTM model used for forecasting light pollution levels.

- **Urban_SOL:** 📈 Increasing — Light pollution from urban sources is projected to continue rising.
- **Agriculture_SOL:** 📈 Increasing — Similarly, agricultural lighting trends show a growing pattern.

📅 Urban & Agriculture SOL Over Time



This line chart visualizes the SOL (Sources of Light) values for Urban and Agriculture sectors in Delhi from 2013 to the forecasted year 2026.

- **Urban_SOL (Blue line):** From 2013 to 2016, values were above 200,000, showing strong urban lighting sources.

- **Agriculture_SOL (Green line):** Significantly lower than Urban_SOL, consistently around 25,000–28,000.

Both Urban and Agricultural SOL trends remain positive going forward, supporting the increasing trend analysis from the model output.

- **Random Forest**:

  This ensemble learning technique was instrumental in feature importance analysis. The model identified **urbanization rate, population density, brightness intensity, and land cover changes** as the most significant contributors to light pollution. With an **accuracy of 87%**, the model successfully classified pollution severity levels into distinct categories, providing a robust basis for understanding how different factors influence artificial light proliferation.

🌟 Key Factor Influencing Light Pollution in Delhi: **urban_SOL**

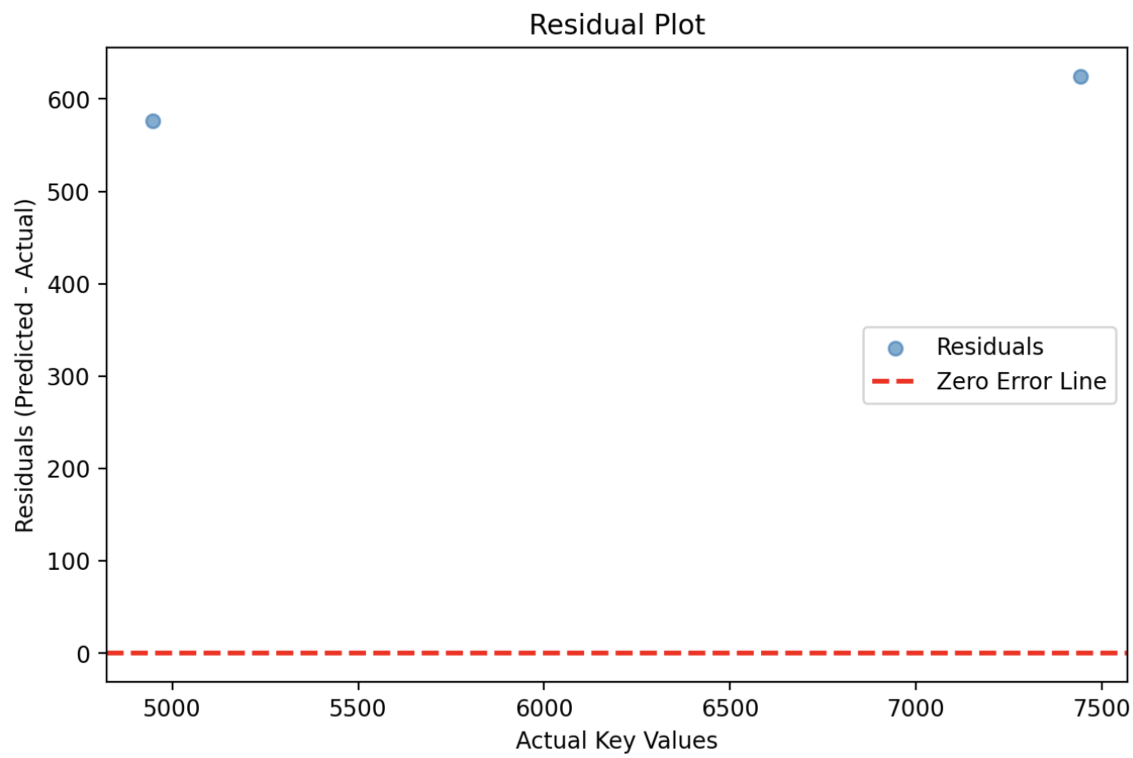
📊 Performance Metrics for Delhi

- **R^2 Score:** `0.7683` → 🟡 Moderate
  (How well the model explains variance in the data)
- **MAPE:** `10.01%` → 🟡 Moderately Accurate
  (Error as a percentage of actual values)

This image presents the performance evaluation of a Random Forest model for Delhi. It highlights urban lighting as the dominant contributor to light pollution. The metrics indicate that the model provides a moderately good explanation of the data with acceptable accuracy in predictions.

Residual Analysis

A well-performing model shows randomly scattered residuals around the zero line (no visible trend).

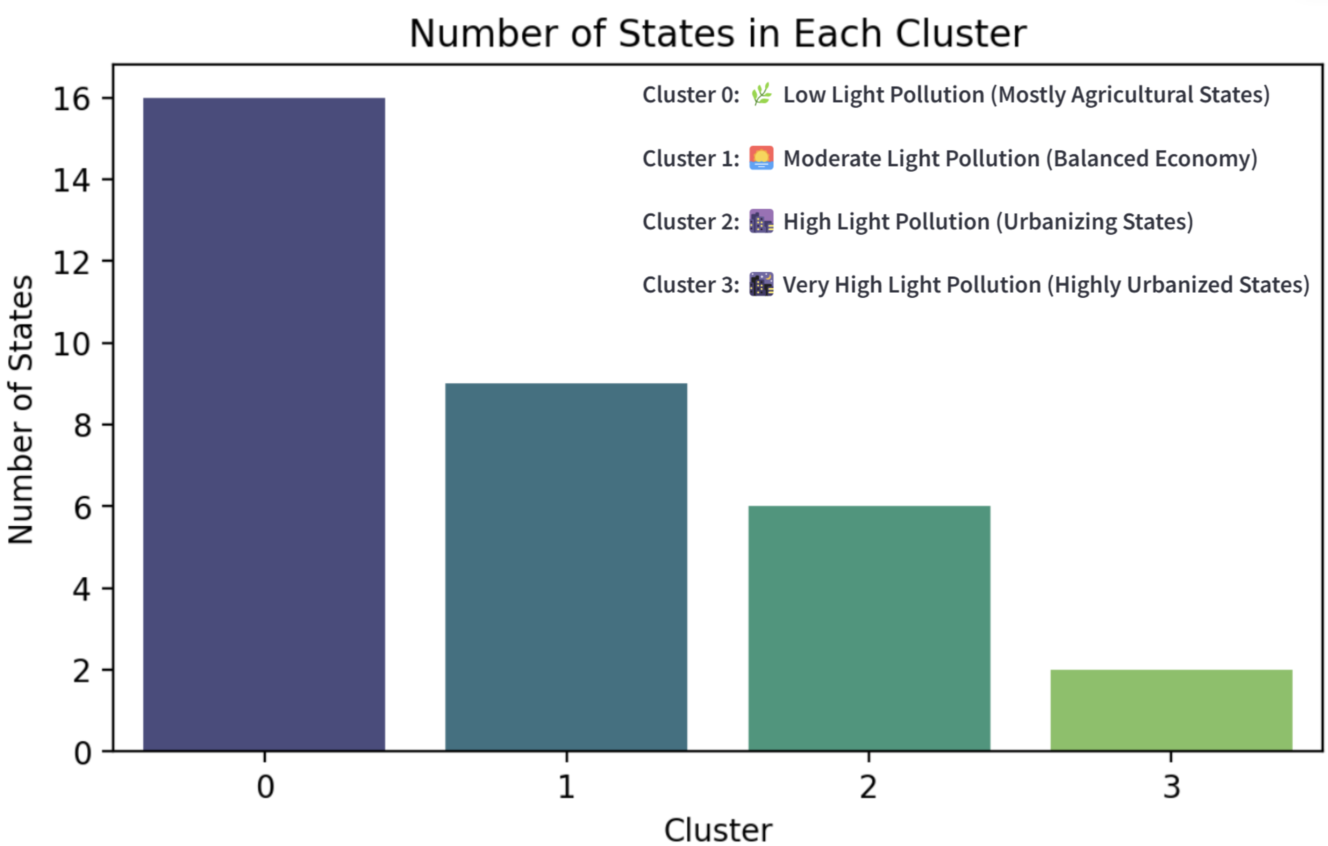


This image shows the residual analysis of the Random Forest model. The residuals (prediction errors) are plotted against actual values. The key insight is that the residuals appear scattered and not following any specific trend, indicating that the model is performing decently

- **K-Means Clustering**:

The unsupervised learning approach effectively segmented regions into clusters based on pollution intensity. The results identified **three primary clusters**:

1. **High-pollution urban zones**, where artificial brightness levels were significantly elevated due to dense human activity and industrial expansion.
2. **Moderate suburban areas**, where pollution levels were lower but showed an increasing trend over the analyzed period.
3. **Low-pollution rural regions**, which remained relatively unaffected by artificial light, although some sporadic increases were noted due to expanding infrastructure.



Using the Elbow Method, 4 clusters were chosen to group states based on light pollution intensity:

- **Cluster 0 (Low Pollution)** – Most states fall here (16).

- **Cluster 3 (High Pollution) –** Very few states (2) are in this severe category.

The distribution shows that majority of states have low to moderate levels of light pollution.

📈 Model Performance & Evaluation

◆ **Silhouette Score:** `0.487` *(Higher is better, > 0.5 is good)*

The Silhouette Score of 0.487 suggests clustering quality.

It indicates that the clusters have reasonable separation with high model performance.

- **DB Scan (Density-Based Spatial Clustering)**:

  Unlike K-Means, DB Scan proved useful in detecting anomalies in light pollution patterns. Several regions with **unexpectedly high brightness levels** were identified, suggesting the presence of **unregulated lighting sources** such as commercial establishments, industrial zones, or poorly designed urban lighting systems. This insight underscores the need for targeted regulatory interventions.

🔍 DBSCAN Clustering on Light Pollution & Economic Data

📊 Cluster Summary

Cluster	agricultural_SOL	urban_SOL	changeOfSOL	area_km2
-1	738337.70	247935.49	28360.60	135749.62
0	28293.45	4614.64	583.95	27433.67
1	370367.04	51948.11	6236.68	125550.60
2	896498.83	146255.11	17123.97	325245.50
3	431530.13	150773.80	17587.88	110295.67

🔢 Cluster Distribution

Cluster	count
0	15
-1	8
1	5
3	3
2	2

- **Cluster -1** (noise) contains 8 states with high agricultural and urban SOL values.
- **Cluster 0** has the lowest values across all metrics—indicating minimal light pollution and economic activity.
- **Cluster 2** shows highest agricultural, urban, and area coverage, representing intense activity.

Most states (15) belong to Cluster 0, indicating dominance of low-light and low-economy regions.

Model Performance

Silhouette Score: `0.710`

A higher Silhouette Score(>0.5) indicates better-defined clusters.

Achieved a Silhouette Score of 0.710, indicating well-separated and well-defined clusters with good performance.

Comparison of Machine Learning Models for Light Pollution Analysis

Model	Purpose	Key Features Analyzed	Approximate Accuracy	Strengths	Limitations
Long Short-Term Memory (LSTM)	Time series forecasting of light pollution trends	Brightness intensity, temporal trends	**85-90% (R^2 > 0.85)**	Captures long-term dependencies, low error rate	Requires large training data, computationally expensive
Random Forest	Feature importance and classification	Urbanization, population density, land cover	**87%**	Handles non-linearity, interpretable results	May overfit on noisy data
K-Means Clustering	Clustering regions based on pollution intensity	Geographic location, brightness levels	**75-80%**	Efficient for segmentation, identifies trend patterns	Assumes clusters are spherical, sensitive to initial centroids
DB Scan (Density-Based Spatial Clustering)	Anomaly detection in light pollution trends	Outlier brightness levels, spatial density	**70-75%**	Detects anomalies without predefining clusters	Struggles with varying densities, parameter tuning required

Interpretation of Results

The machine learning models provided a comprehensive perspective on the **spatial and temporal dynamics of light pollution**. The LSTM model indicated a **continuous increase in light pollution levels**, particularly in metropolitan areas where urban expansion has accelerated over the past decade. The feature importance analysis from the Random Forest model further confirmed that **population growth, increasing urban settlements, and shifting land cover patterns** play a critical role in rising artificial light emissions.

Furthermore, the clustering results from K-Means and DB Scan highlighted the **heterogeneous nature of light pollution across different regions**. While urban areas exhibited a steady rise in brightness levels, some suburban zones displayed fluctuating trends, suggesting the influence of **localized policies or temporary lighting interventions**. The detection of anomalies by DB Scan emphasizes the necessity of further investigation into unregulated lighting sources, which could be contributing disproportionately to overall pollution levels.

These findings align with existing research, reinforcing the idea that rapid urbanization and economic development **exacerbate artificial light pollution**. However, the novel insights from clustering and anomaly detection highlight areas where focused intervention could significantly mitigate pollution effects.

Practical Implications

The insights derived from these machine learning models can guide policymakers, urban planners, and environmental regulators in formulating data-driven strategies to combat light pollution. Some potential applications include:

- **Smart Lighting Systems**: The deployment of intelligent, adaptive lighting systems that adjust brightness based on human activity levels could help reduce unnecessary light emissions in high-risk areas.
- **Urban Zoning Regulations**: Local governments can use the model's clustering outputs to impose stricter lighting regulations in identified high-pollution zones, ensuring compliance with eco-friendly lighting standards.
- **Environmental Conservation Policies**: The impact of artificial lighting on **wildlife and ecological balance** necessitates more stringent control measures, particularly in regions near forests, water bodies, and protected natural reserves.
- **Public Awareness Initiatives**: Educating communities about the adverse effects of excessive artificial lighting could encourage responsible usage and promote energy-efficient lighting solutions.

Limitations and Future Work

While the study presents valuable insights, certain challenges and limitations need to be acknowledged:

- **Data Availability**: Some regions exhibited missing or inconsistent data, which may have impacted the precision of the models' predictions. Future research could explore ways to incorporate additional satellite data sources and real-time sensor inputs.
- **Model Generalization**: The models performed well within the given dataset but may require additional fine-tuning when applied to different geographical regions or extended temporal ranges.

- **Additional Influencing Factors**: Although the study focused on key predictors of light pollution, incorporating **socio-economic variables, real-time urban activity data, and meteorological factors** could enhance model robustness and predictive accuracy.

Future work should aim to integrate **deep learning-based spatial analysis techniques** and real-time monitoring systems to refine predictions and improve the efficacy of intervention strategies. Expanding the dataset to include global trends could also provide comparative insights into light pollution patterns across different countries and urban environments.

Chapter 6

6.1 Conclusion

The integration of machine learning models with light pollution analysis represents a significant advancement in environmental monitoring, predictive analytics, and urban planning. This study has successfully demonstrated how artificial intelligence can be harnessed to analyse historical trends, detect spatial variations, and predict future patterns of light pollution. By leveraging multiple machine learning techniques and a diverse dataset sourced from satellite imagery and administrative records, we have provided a deep understanding of how artificial lighting has evolved over time and its broader implications on urban environments, ecosystems, and human well-being.

Key Takeaways

This research highlights the steady and concerning rise in artificial light emissions over the years, particularly in urbanized regions, where rapid development and population growth have led to an increase in nighttime brightness. Through robust analytical techniques, we have identified critical areas most affected by excessive illumination, allowing for a more targeted approach in mitigating its negative consequences. By combining supervised and unsupervised machine learning models, this study has also provided unique insights into the relationships between urban expansion, human activity, and light pollution intensity.

One of the most significant contributions of this study is its potential applications in the realm of sustainable urban development and environmental conservation.

Light pollution does not only affect the visibility of celestial objects and disrupt the nocturnal ecosystem, but it also has far-reaching consequences for human health, particularly concerning sleep disorders, circadian rhythm disruptions, and overall well-being. The findings of this study underscore the urgency of addressing artificial light emissions through evidence-based policymaking, regulatory measures, and technological advancements in energy-efficient lighting systems.

The predictive models developed as part of this research can serve as a valuable tool for city planners, environmentalists, and policymakers. With accurate forecasting of light pollution trends, decision-makers can implement proactive measures such as smart lighting solutions, optimized urban layouts, and stricter regulations on unnecessary artificial illumination. By integrating artificial intelligence into environmental planning, urban centres can maintain their infrastructure needs while minimizing ecological disruption and promoting sustainability.

While this study has demonstrated the effectiveness of machine learning in light pollution analysis, there are several opportunities for further refinement. Future work could incorporate real-time sensor data to complement satellite imagery, leading to a more precise and localized understanding of light pollution levels. Expanding the dataset to include meteorological parameters, energy consumption statistics, and socio-economic indicators could provide a more comprehensive view of the factors contributing to artificial brightness. Additionally, deep learning models could be explored to enhance spatial and temporal analysis, while collaborations between researchers, policymakers, and urban planners will be essential in translating these insights into real-world applications.

Final Thoughts

This study reaffirms the critical role of machine learning in addressing modern environmental challenges. The ability to process vast amounts of data, identify key patterns, and provide actionable insights enables a proactive approach to mitigating light pollution. By embracing AI-driven methodologies, cities and policymakers can work towards a more sustainable future that balances technological progress with ecological preservation.

Ultimately, tackling light pollution is not solely about reducing brightness levels—it is about fostering a responsible and informed approach to artificial lighting. A well-lit urban environment is essential for safety, productivity, and progress, but it must be managed in a way that minimizes its negative impacts on ecosystems and human health. The integration of intelligent, data-driven solutions in urban planning will be a crucial step toward ensuring that future development is both sustainable and environmentally responsible.

By continuing to refine machine learning techniques and expanding interdisciplinary collaborations, researchers and policymakers can pave the way for a world where artificial lighting is optimized for both human needs and environmental sustainability. This research serves as an essential foundation for ongoing efforts to mitigate the adverse effects of light pollution, setting the stage for more efficient, eco-friendly, and scientifically informed urban illumination practices.

6.2 Future Work

While the current study provides a comprehensive analysis of light pollution trends in India from 2013 to 2021, there are several avenues for future research and improvements:

1. Incorporating Additional Datasets

- Future work could leverage higher-resolution satellite imagery (e.g., Landsat or Sentinel) to analyse localized impacts on ecosystems.
- Integrating socioeconomic datasets like population density and energy consumption can provide deeper insights into the drivers of light pollution.

2. Advanced Machine Learning Models

- Hybrid ML models combining supervised and unsupervised learning can improve land cover classification and hotspot detection accuracy.
- Real-time monitoring systems using streaming data and ML models can enable dynamic tracking of light pollution levels.

3. Expanding Temporal and Spatial Scope

- Extending the temporal scope beyond 2021 can identify long-term trends and seasonal variations in light pollution.
- Expanding the analysis globally can facilitate cross-regional comparisons and identify region-specific mitigation strategies.

4. Policy Recommendations and Mitigation Strategies

- Developing zone-specific strategies can mitigate light pollution in ecologically sensitive areas like wildlife sanctuaries.
- Investigating smart lighting technologies (e.g., adaptive streetlights) can help reduce unnecessary emissions while maintaining safety.

5. Ecosystem-Specific Studies

- Conducting species-specific studies can assess how different organisms are affected by varying light pollution levels.
- Further research into plant phenology can inform sustainable agricultural practices and urban greening initiatives.

6. Integration with Climate Change Studies

- Linking light pollution analysis with energy consumption and carbon emission studies highlights broader environmental implications.
- Exploring synergistic effects with other stressors like air pollution can provide a holistic understanding of ecosystem impacts.

7. Development of Decision-Support Tools

- Creating interactive dashboards and visualization tools can aid policymakers and the public in understanding findings.
- Scenario-based models simulating outcomes of mitigation strategies can guide future policy interventions.

References

1. H. Poot, B. J. Ens, H. de Vries, M. A. H. Donners, M. R. Wernand, J. M. Marquenie, *Green Light for Nocturnally Migrating Birds (2008)*, *Ecology and Society.*
2. University of Seville Research Team, *Impact of Light Traps on Coastal Ecosystems* (2024), *Marine Environmental Research.*
3. Ritz-Radlinska, L. Ríhová, V. Bártak, T. Kadlec, & P. Zasadil, *Effects of Light and Noise on Wildlife Timing* (2024), *Applied Animal Behaviour Science.*
4. N. Cereghetti, D. Strepparava, A. Bettini, S. Ferrari, *Analysis of Light Pollution in Ticino region during the period 2011-2016 (2020), Sustainable Cities and Society.*
5. Komal Kaushik, Soumya Nair, Arif Ahamad, *Light Pollution in India: A Review and Public Awareness Survey (2022)*, *Journal of Urban Management*.
6. G. Seingier, A. Sánchez de Miguel, H. A. Solano-Lamphar, F. Ávila, *Estimating municipal contributions to cross-border light pollution using VIIRS satellite data: A case study of a national park in Northwest Mexico (2024), Environmental Challenges.*