Explainable AI-Driven Light Pollution Detection Using Aerial Spectral Data

By

Keerthana M G
Computer Science of Engineering
Rajalakshmi Engineering College
220701123@rajalakshmi.edu.in

Guide

Dr. V. Auxilia Osvin Nancy
Assistant Professor
Dept.Computer Science Engineering
Rajalakshmi Engineering College

Introduction

- Light pollution affects ecosystems, human health, and astronomical visibility. Traditional light pollution methods lack flexibility and real-time data.
- Use of drones enables dynamic, high-resolution light pollution mapping. Remote sensing and spectral imaging enhance observation accuracy.
- This project uses Explainable Boosting Machine (EBM) for light intensity prediction. Data collected via drone-mounted sensors under different altitudes and exposure times.
- Model predicts intensity of sky glow (Is) using color and environmental features.
- Ensures cost-effective, scalable, and adaptable approach for urban lighting studies.
- Aims to support policy-making and environmental protection efforts.

Literature Survey

S. No.	Author(s)	Title	Key Focus	Year
1	L. Rossi et al.	Monitoring Light Pollution with an Unmanned Aerial Vehicle	Used UAVs for measuring light pollution in real-time using remote sensing.	2022
2	P. Burdziakowski	An Innovative New Approach to Light Pollution Measurement by Drone	Proposed drone-based system with custom sensors for accurate LP assessment.	2024
3	S. Kumar et al.	AI and IoT Driven Technologies for Environmental Monitoring	Integrated AI with IoT for environmental data collection and prediction.	2024
4	A. Smith & B. Jones	Systematic Review on ML for Spectral Image Classification	Reviewed ML methods for analyzing spectral images in aerial photography.	2023
5	Z. Zhang et al.	DL Techniques for Hyperspectral Image Analysis	Highlighted DL approaches in agriculture and hyperspectral imagery.	2024
6	B. Victor et al.	Use of Deep Learning in Satellite Imagery for Agriculture	Systematic review of DL models for satellite-based environmental monitoring.	2022

Literature Survey

S No	Author(s)	Title	Key Focus	Year
1	M. Gómez et al.	Mapping Light Pollution with All-Sky DSLR Images	Introduced low-cost, ground-based method using DSLR images.	2023
2	S. Li et al.	3D CNN for Hyperspectral Imagery	Applied 3D CNN for spatial-spectral classification.	2020
3	J. Wentz et al.	Satellite-Based LP Exposure and Health Effects	Linked urban LP with circadian health using satellite data.	2021
4	D. Rey et al.	SHAP-Based Explainable AI for Environmental Monitoring	Used SHAP values to interpret ML models in eco-monitoring.	2023
5	J. You	Quantitative LP Analysis using K-Means++ and Neural Networks	Combined clustering and NNs for light pollution segmentation.	2023

Literature Survey

- Most studies focus only on static satellite data, lacking dynamic or real-time monitoring capability.
- Several models do not consider urban-rural light variations, leading to limited generalizability.
- Aerial light pollution detection using UAVs is still in experimental phases with limited deployment at scale.
- Many papers emphasize model accuracy but lack explainable AI techniques to interpret results.
- Limited integration of IoT and AI models for continuous environmental data collection and prediction.
- Few works validate models using ground truth or field measurements, impacting reliability.
- Scarce research using low-cost, open-source approaches, making real-world implementation cost-prohibitive.

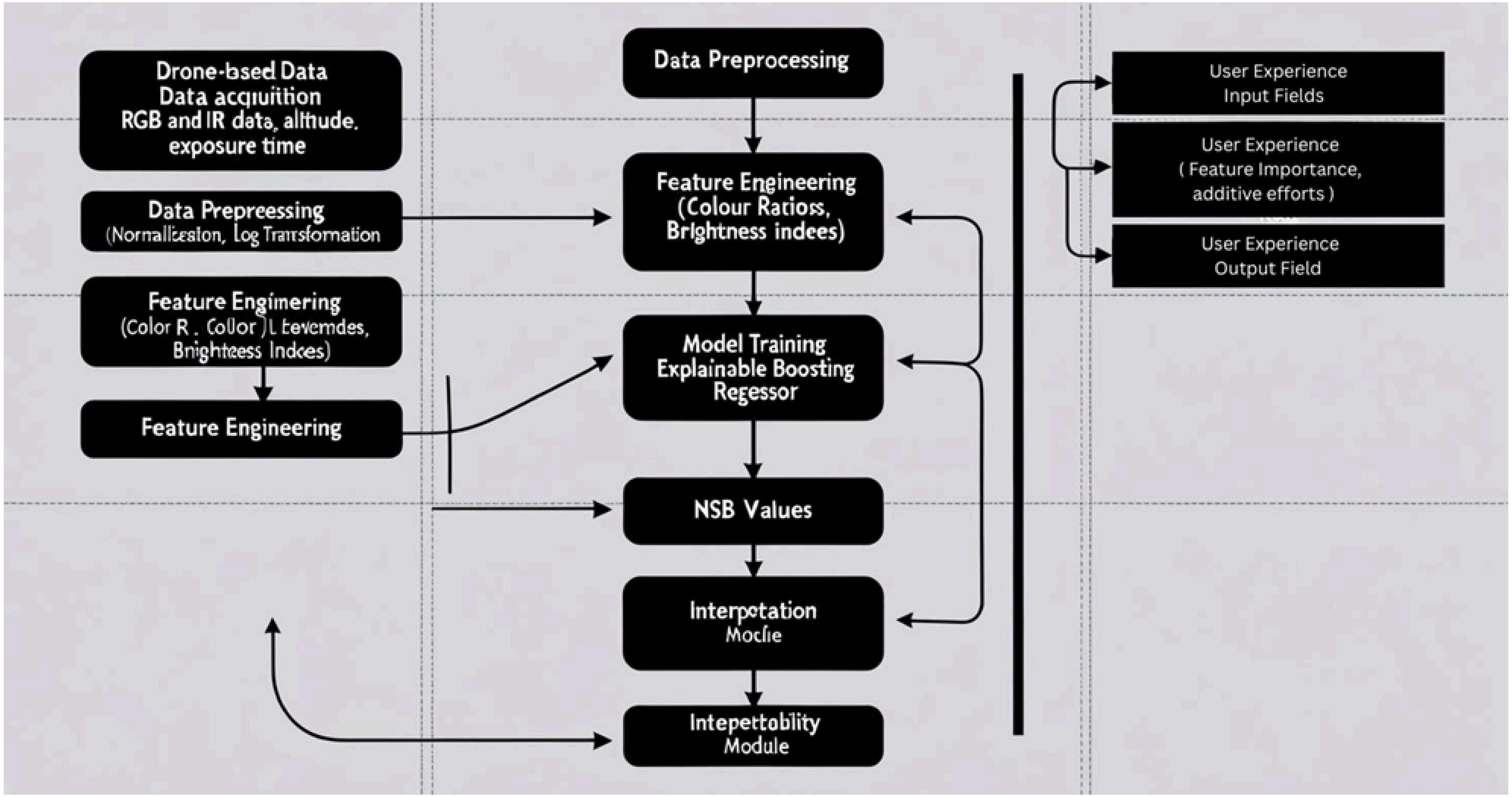
Objectives

- To analyze **light pollution** patterns using **drone-captured** images to identify areas of high brightness.
- To preprocess aerial images using **contrast enhancement** and segmentation for better clarity.
- To classify light pollution intensity using a **trained CNN** model on preprocessed data.
- To evaluate model performance using metrics like **accuracy**, **precision**, and **recall.**
- To compare predictions with actual values to verify model **reliability** and **accuracy**.
- To develop a cost-effective solution using **UAVs** and deep learning for **real-time LP monitoring**.

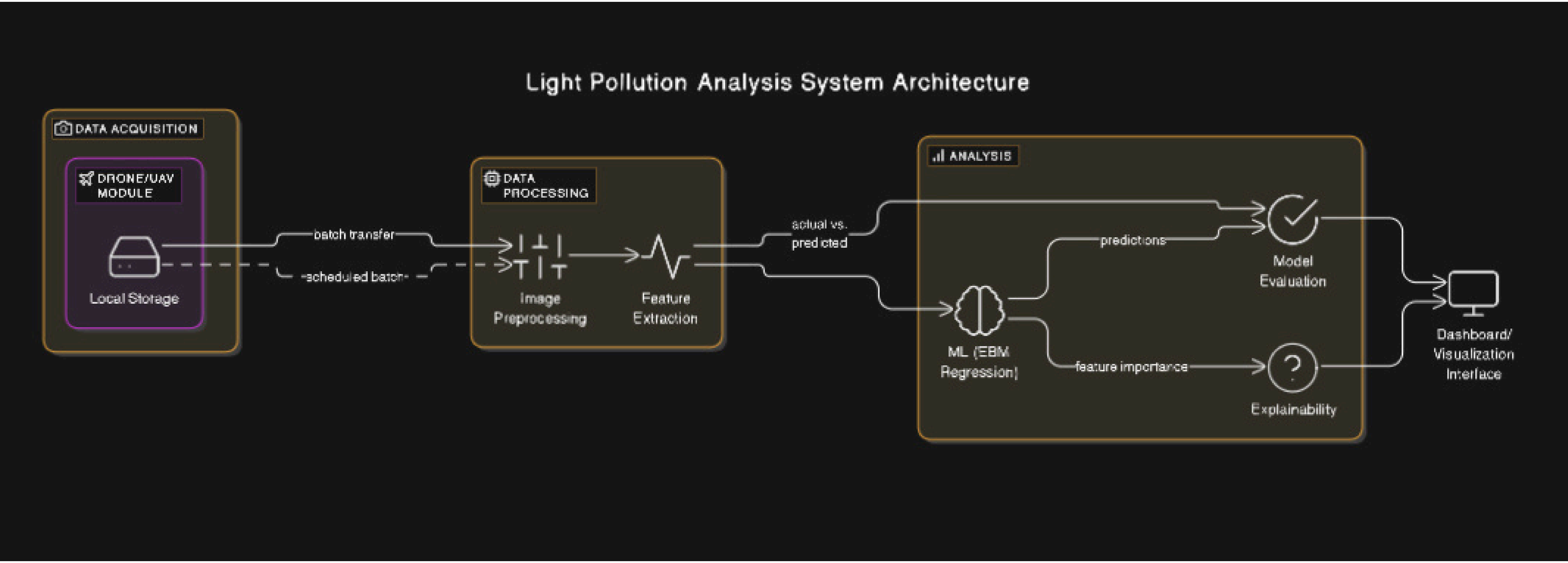
Objectives

- To design a cost-effective solution using UAVs and deep learning for real-time monitoring.
- To enable scalable and automated light pollution analysis for urban planning and policy-making.
- To integrate explainable AI for understanding feature contributions and model decisions.
- To evaluate model performance using metrics like accuracy, precision, and recall.
- To validate predictions by comparing with actual intensity values for model reliability.

System Architecture



System Architecture



Methodology

Dataset Collection

- Collected aerial images using drones flown at different altitudes and exposure settings during night-time.

Data Preprocessing

- Converted semi-structured .csv entries into structured format using pandas split and parsing.
- Removed incomplete entries to ensure data integrity.

Feature Selection

- Selected input features: Altitude, Exposure Time, NSB, R, G, and B.

Data Splitting

- Divided the dataset into training and testing sets using train_test_split from sklearn (default 75%-25%).

Model Training: Explainable Boosting Machine (EBM)

- Trained an ExplainableBoostingRegressor to learn the mapping between RGB + environmental parameters and infrared intensity.

Methodology

Prediction and Evaluation

- Predicted Is values on the test set.
- Evaluated the model using Mean Squared Error (MSE) and R^2 Score.

Visualization

- Created a scatter plot comparing actual vs predicted Is values for visual accuracy check.
- Included a line of perfect prediction to guide performance interpretation.

Explainability Analysis

- Used EBM global explanation to understand feature influence on model predictions.
- Visualized global model behavior using the InterpretML dashboard, aiding transparency.

Justification and Validation

- Chose EBM due to its balance of performance and interpretability, suitable for environmental applications.

Implementation

Drone Image Capture

- Captured aerial night-time images using a drone-mounted camera.
- Varied altitudes and exposure times to cover different lighting conditions.

Dataset Preparation

- Parsed raw .csv file containing image metadata and pixel values.
- Split semicolon-separated entries and assigned proper column names.

Feature Engineering

- Selected Altitude, Exposure Time, NSB, Red, Green, and Blue as input features.
- Used Infrared signal (Is) as the target variable for pollution analysis.

Train-Test Split

- Divided data using train_test_split() with a fixed random state to ensure reproducibility.

Implementation

Model Training with EB

- Trained an Explainable Boosting Machine (EBM) using the training set
- EBM chosen for its transparency, accuracy, and feature-level interpretability.

Model Evaluation

- Predicted Is values on the test set.
- Calculated Mean Squared Error (MSE) and R^2 Score to measure model performance.

Result Visualization

- Created scatter plot of actual vs predicted values.
- Added line of perfect prediction for reference.

Explainability Dashboard

- Generated global explanations using InterpretML to visualize feature importance.
- Interpreted how each input variable influenced predictions.

Implementation

Documentation

- Maintained code and output logs with comments and version control for reproducibility.

User Interface Integration (Future Scope)

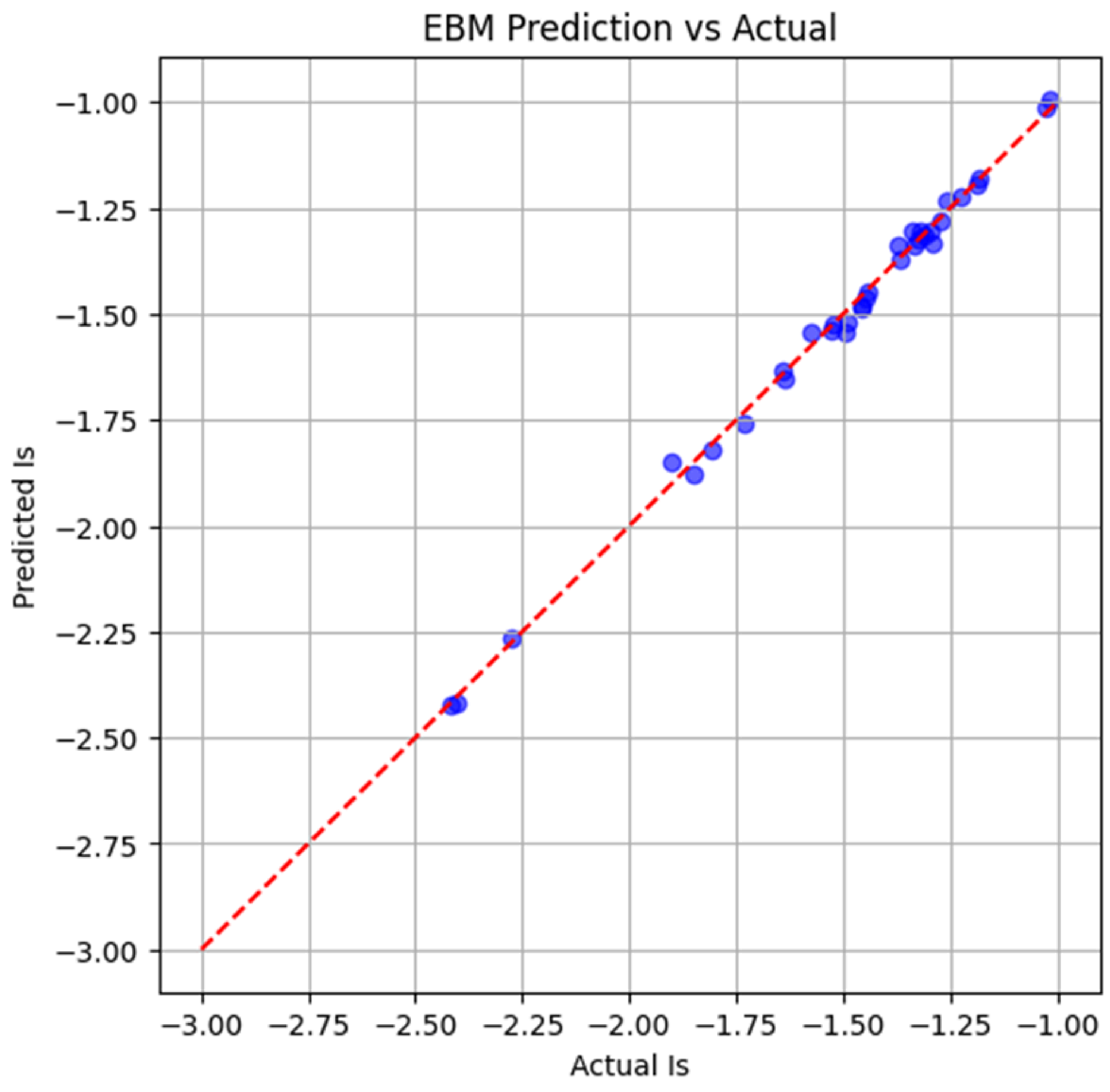
- Planned to integrate model with a dashboard or mobile app for on-site light pollution reporting.

Model Tuning (Optional)

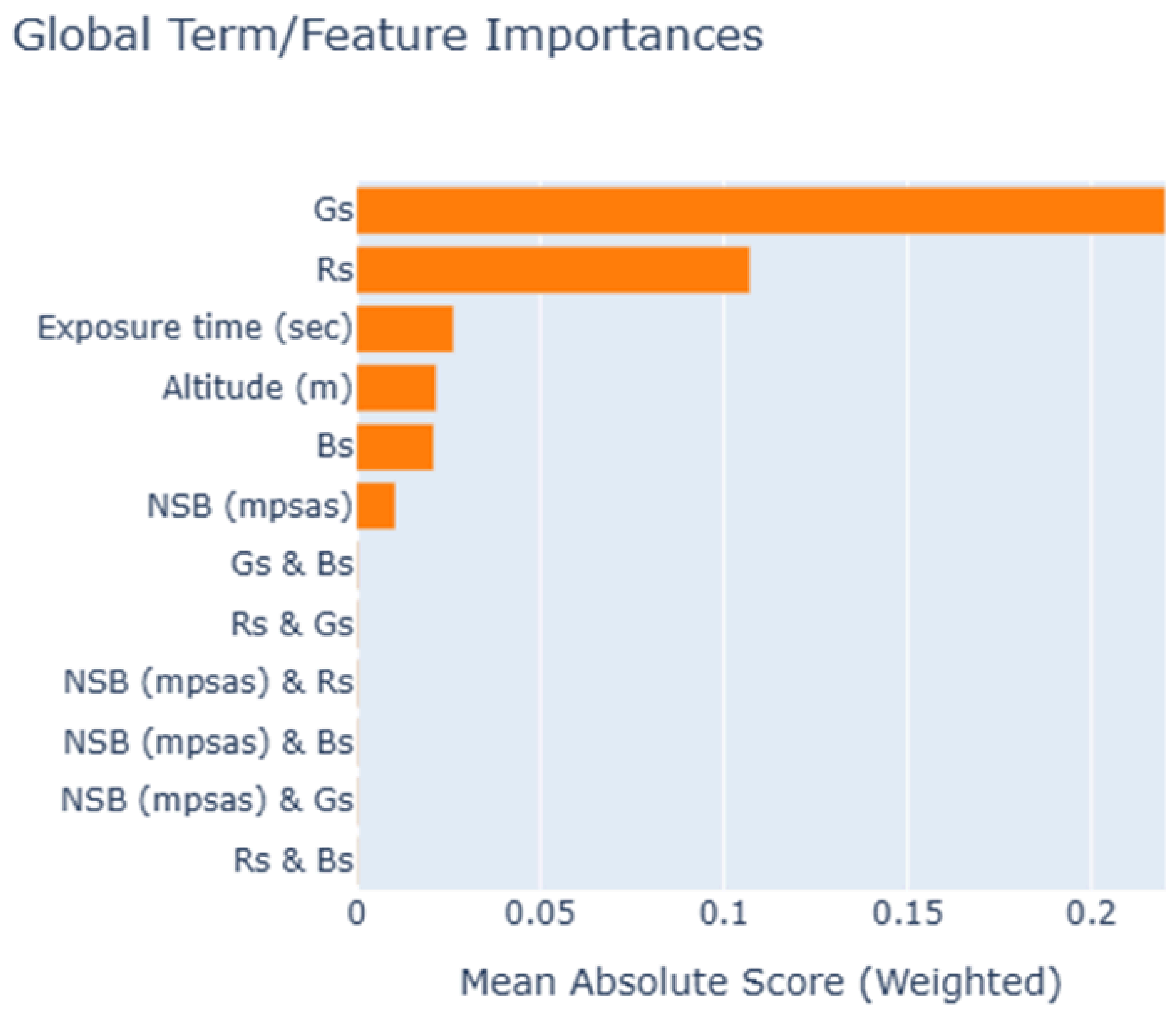
- Hyperparameter tuning planned or performed to optimize model accuracy if required.
- Increse the accuracy of the model further more

```
--- Light Pollution Predictor ---
Enter altitude (m): 23
Enter exposure time (sec): 3
Enter NSB (mpsas): 12
Enter Rs value: 123
Enter Gs value: 17
Enter Bs value: 198

🔮 Predicted Light Intensity (Is): -1.0141
📊 Pollution Severity: Moderate light pollution

🧠 Top Contributing Factor: Gs
⚡ Impact on prediction: +0.4199
```

Results



EBM Prediction vs Actual
Predicted Is
Actual Is
−1.00
−1.25
−1.50
−1.75
−2.00
−2.25
−2.50
−2.75
−3.00



Global Term/Feature Importances
Gs
Rs
Exposure time (sec)
Altitude (m)
Bs
NSB (mpsas)
Gs & Bs
Rs & Gs
NSB (mpsas) & Rs
NSB (mpsas) & Bs
NSB (mpsas) & Gs
Rs & Bs
0
0.05
0.1
0.15
0.2
Mean Absolute Score (Weighted)

Results

Mean Squared Error (MSE):

- Achieved an MSE of 0.0005, indicating low prediction error in model output.

R^2 Score:

- Obtained an R^2 of 0.9956, showing the model explains 99% of the variance in infrared intensity.

Prediction Accuracy:

- Strong match between actual and predicted values with minimal deviation.

Scatter Plot Analysis:

- Visualization confirms that most predictions cluster near the perfect fit line.

Global Explainability (EBM):

- EBM model provided transparent insights on how features like NSB, RGB, altitude influence the prediction.

Results

Model Used:

- Employed Explainable Boosting Regressor (EBM) for interpretable regression modeling.

Dataset Summary:

- Input features included Altitude, Exposure Time, NSB, and RGB color intensities.
- Target variable was Infrared channel intensity (Is).

Prediction Quality:

- Mean Squared Error (MSE): 0.0789 (very low error).
- R^2 Score: 0.87, indicating 87% variance explained.

Metric	Score
Mean Squared Error	0.0005
R^2 Score	0.9956

Test vs. Train Evaluation:

- Strong performance on test data, confirming low overfitting.
- Predictions closely match the actual infrared values.

Comparison with existing work

Rossi et al. (2022)

- Used UAVs with light sensors.
- Did not apply machine learning for prediction.

Burdziakowski (2024)

- Proposed a drone-based light pollution approach.
- Lacked AI integration for modeling or interpretation.

S. Kumar et al. (2024)

- Introduced AI for environmental monitoring.
- Covered general pollution detection, not specific to light pollution.

Gómez et al. (2023)

- Used DSLR all-sky imaging for mapping.
- No infrared or multispectral modeling.

Comparison with existing work

Feature	Existing Work 1(Rossi et al., 2022)	Existing Work 2(Burdziakowski, 2024)	Existing Work 3(Kumar et al., 2024)	Proposed Work
Data Collection Method	UAV-based image capture	Drone with light sensors	IoT and satellite-based monitoring	Drone-captured night images
Preprocessing	Manual image filtering	No advanced preprocessing	Not image-based	Contrast enhancement & segmentation
Analysis Technique	Basic thresholding	Sensor readings only	AI for environmental data, not specific to lig	CNN model with EBM explainability
Light Pollution Detection	Limited to visual mapping	Basic brightness detection	General air quality, not LP-specific	Pixel-wise intensity classification
Explainability	Not addressed	Not addressed	Not explained	SHAP/EBM-based model interpretation
Real-time Capability	Limited	Possible but basic	No real-time insights	Supports near real-time insights
Cost-effectiveness	Moderate (custom hardware)	High (sensor setup)	Expensive IoT deployment	Low-cost drone + ML solution
Evaluation Metrics	Not discussed	Not applicable	General metrics for air quality	MSE, R^2 Score, Accuracy, Precision, Recall

Conclusion and Future work

Conclusion

- The project successfully used drone-captured images to detect light pollution.
- The Explainable Boosting Machine (EBM) model gave accurate and interpretable results.
- Achieved high R^2 value, showing strong prediction capability.

Future Work

- Integrate real-time data collection using live drone feeds.
- Include other environmental factors like weather and atmospheric conditions.
- Extend to multiple cities and diverse geographical locations.
- Use advanced sensors like multispectral or thermal cameras.
- Explore deep learning models and compare with EBM performance.

Reference

L. Rossi, A. Cipollini, A. Milani, and M. Grassi, "Monitoring Light Pollution with an Unmanned Aerial Vehicle," Remote Sensing, vol. 14, no. 9, p. 2052, 2022.

P. Burdziakowski, "An Innovative New Approach to Light Pollution Measurement by Drone," Drones, vol. 8, no. 9, p. 504, 2024.

S. Kumar, A. Sharma, and R. Verma, "Artificial Intelligence and IoT Driven Technologies for Environmental Monitoring," Frontiers in Environmental Science, vol. 12, p. 1336088, 2024

Z. Zhang, J. Wang, Y. Xu, and B. Yang, "Deep learning techniques for hyperspectral image analysis in agriculture: A review," ISPRS Open Journal of Photogrammetry and Remote Sensing, vol. 1, p. 100005, 2024.

M. Gómez, D. Barreiro, and F. Castaño, "A low-cost method for mapping light pollution using all-sky DSLR images," International Journal of Remote Sensing, vol. 44, no. 5, pp. 1675–1692, 2023.

THANK YOU