University of Peradeniya

Spatial-Temporal Deep Learning Model for Predicting PM2.5 in Sri Lankan Urban Cities

Herath E.M.L.M.B[1], Manodya N.P.K.H.[1], Sellahewa I.B.[1]
Supervised by: W.A.N.I. Harischandra[1], G.Bowatte[2]
[1]DEEE, Faculty of Engineering, University of Peradeniya. [2]Faculty of Allied Health Sciences, University of Peradeniya.

Abstract- This project focuses on predicting PM2.5 air pollution levels in Sri Lankan urban areas using a hybrid CNN-LSTM model. We combined temporal data from LSTM with spatial data from CNN to enhance prediction accuracy. The integration of these models improved performance in estimating air quality for better environmental monitoring.

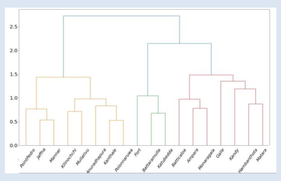
Introduction

This project harnesses the power of deep learning to predict future PM2.5 levels, a critical air quality indicator with significant health implications. By innovatively combining satellite imagery analysis through Convolutional Neural Networks (CNN) with historical data processing via Long Short-Term Memory (LSTM) networks.

Data collection → **Data Preprocessing** → **Clustering** → **Temporal Prediction** → **Spatial Data Integration** → **Building Hybrid Model**

Data collection & Clustering

- Hourly averaged PM2.5, along with spatial data including satellite images, timestamps, were collected from eighteen stations across the island.
- Pearson correlation was used to identify three clusters of monitoring stations with similar PM2.5 behavior, representing regions with same air quality patterns.



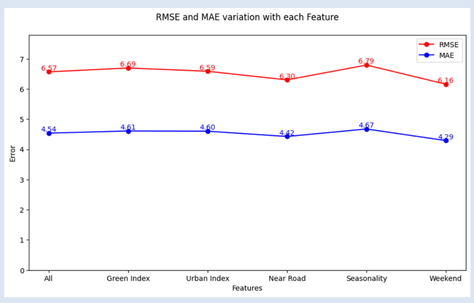
Correlation among the stations

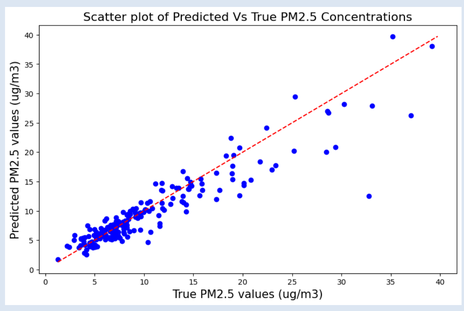
one sample photo among the dataset

18 stations in the map

Spatial Data Integration

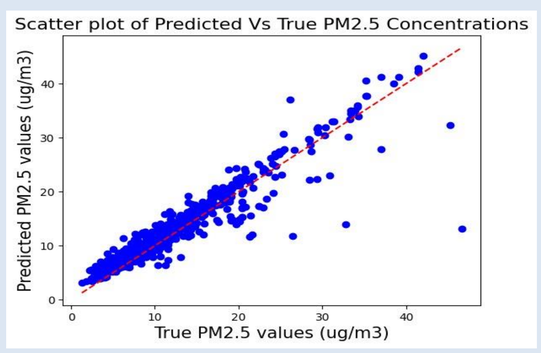
- Utilized a pre-trained VGG16 model to extract deep features from RGB satellite images.
- Manually extracted features like the Green Index, Urban Index and Near Road index are integrated. These features are concatenated and passed through fully connected layers to enhance predictive accuracy.



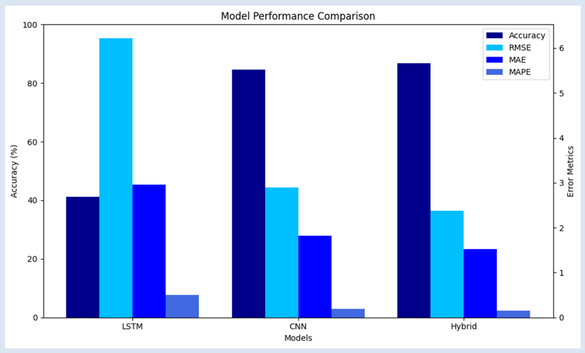


Results

- The hybrid CNN-LSTM model demonstrated superior performance, achieving the lowest errors and it attained an R-squared value of 86.86%, outperforming both the standalone LSTM and CNN models.
- The CNN model also performed well, with an R-squared of 83.84%. The LSTM model, though less accurate, provided insights with an R-squared of 41.13%.



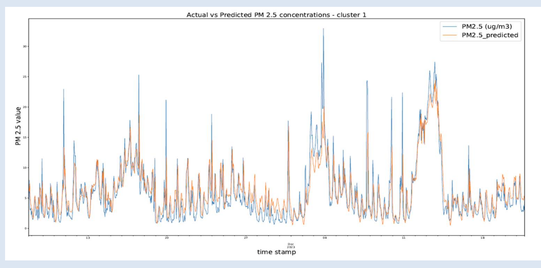
Hybrid model results

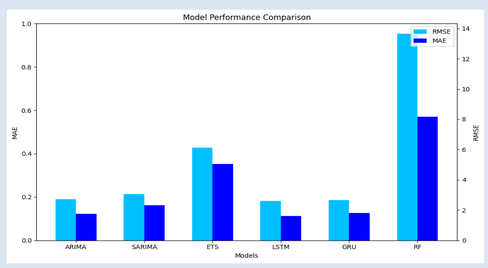


Model Performance Comparison

Temporal Prediction

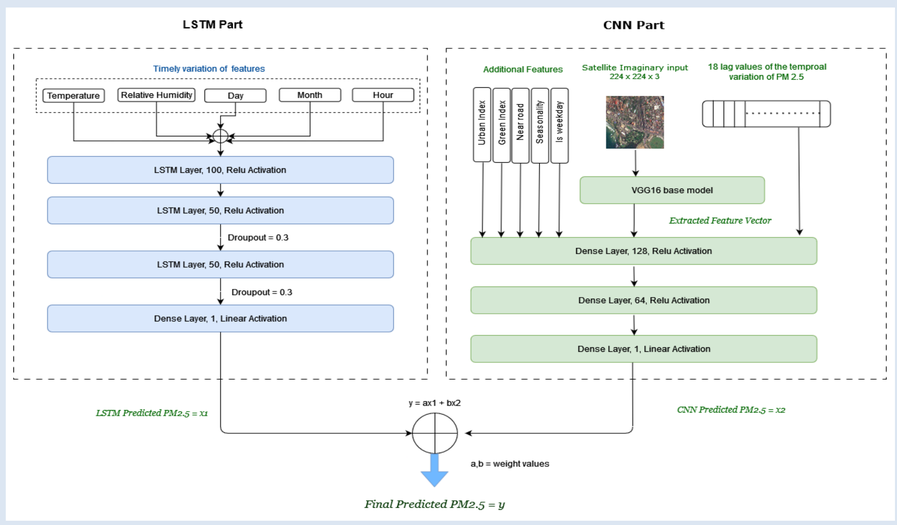
- Compared various statistical forecasting models such as ARIMA, SARIMA, and ETS, alongside deep learning approaches including LSTM, GRU, and machine learning techniques like Random Forest.
- LSTM emerged as the most accurate model for PM2.5 prediction.





Methodology

- The approach involved generating separate prediction lists from the LSTM and CNN models.
- These predictions were then combined using linear regression to calculate optimal weight values, resulting in a more accurate, unified model.



Hybrid Model architecture

Conclusion

The hybrid CNN-LSTM model effectively predicted PM2.5 concentrations in Sri Lankan urban cities, demonstrating high accuracy and reliability compared to standalone models. The model's architecture shows promise for extending predictive capabilities using satellite imagery and advanced deep learning techniques.

Contact details

Name : Dr. W.A.N.I. Harischandra
Tel. No.: +94 81239 3427
Email : nalin@ee.pdn.ac.lk

Dept. of Electrical and Electronic Engineering
Faculty of Engineering
University of Peradeniya
Peradeniya, 20400, Sri Lanka