<u>Development of cost-effective sensors and advance AI models for monitoring and forecasting traffic and pollution levels</u>

for
The Big Bang Competition Award

<u>Project Overview (150 words)</u>

Traffic congestion emits pollutants contributing to respiratory illness, cardiovascular diseases and other health problems. A 2018 TomTom study noted 2.2 megatons of Carbon Dioxide traffic emissions annually in London. We developed a cost-effective 6-element sensor and Long Short-Term Memory (LSTM) and You Only Look Once v8 (YOLOv8) models to forecast traffic pollution. By integrating machine learning with real-time sensors, our solution aims to empower communities with adaptive traffic management. Our research demonstrated a time-series LSTM model with low Mean Squared Error (MSE) of 4.12e-4 and a YOLOv8 object detection model with high Mean Average Precision 50 (MAP50) value of 0.934. Our research also discovered a surprisingly weak Pearson correlation between vehicle counts and air pollution, indicating a non-linear relationship between these entities. Integrated into services such as Google/Apple, our solution can enable autonomy to users and cities to plan sustainable traffic pathways in urban areas.

<u>Project Concept (300 words)</u>

Living in London, we have experienced the wrath of road congestion first-hand. According to a 2018 study by TomTom roughly 2.2 megatons of carbon dioxide are emitted yearly in London because of inefficient traffic and congestion. Transport for London (TfL) introduced an Ultra-Low Emission Zone (ULEZ) to curb the severe state of atmospheric pollution every day caused by traffic. Recently, ULEZ was extended to our area, forcing several families, including ours, to drastically change their vehicular use to avoid the ULEZ charge. Curious, we decided to look deeper into the matter. The motivation behind addressing traffic congestion is the notion that by reducing air and noise pollution, we can aim to create more sustainable and livable cities. Was there a way to reduce the pollution in the air and not only reduce the ULEZ charge prices but also improve the daily lives of Londoners? Motivated by this predicament, we decided to research the direct impact of traffic on air pollution. Our research led us to develop cost-effective sensors and use LSTM and YOLOv8 models to monitor and forecast pollution levels based on traffic. By integrating machine learning techniques with real-time data collection, our solution aims to empower communities to mitigate the environmental impacts of congestion, helping integrate sustainability with urban living. As a part of our project, we built a cost-effective sensor to monitor the condition of the atmosphere. we also wanted to forecast traffic along with the air pollution data. Therefore, we trained an LSTM model, with secondary data, to forecast 2 hours into the future, with traffic data of the past. To detect traffic in the local area, we also used a camera to record the traffic in the local area. we then trained and tested a YOLOv8 object detection model to record and forecast traffic.

Project Process (400 words)

Experiment 1: To forecast traffic from historical data, we trained the Autoregressive-Integrated- Moving-Average (ARIMA) model with a dataset containing traffic indices for 840 contiguous quarter-hours for 36 intersections. We also used a Long Short-Term Memory (LSTM) model to capture long-term dependencies. Both models were hyper-parameter-tuned with effectiveness measured. The dataset contained traffic data for 35 days in total, of which the model was trained on 30 days and forecasted on 5 days of data. The effectiveness was essentially the measured error between actual data and forecasted data using the metrics MSE and MAE.

Experiment 2: A literature review revealed a dataset with 4041 aerial roadway images. We used this to train and test an object detection YOLOv8 model to detect the number of vehicles on the roadway. The dataset was split into training/validation/test at (2821 (69.8%)/808 (20.0%)/412 (10.2%)). All results are reported on the test dataset. Next, follow-up experiments were performed with footage from our camera sensor at an intersection, to test the transferability of object detection models from one scene to another when the angle and nature of the background are different.

To collect primary data on air pollution in our local area, we built a sensor monitoring the condition of the atmosphere; we combined a sensor that monitors temperature-humidity, PM2.5, CO_2, volatile gasses, NO_2 and Methane with an ESP-32 board to form one larger sensor. We strategically placed our sensor on a busy road junction and collected data for the state of the atmosphere by coding micro python code onto our ESP-32 controller. At the same time, to obtain aerial images, we also placed a camera next to our sensor to record the vehicles. The timeframes of the camera corresponded to the air pollution data, which allowed us to assess the impact of traffic on greenhouse gas emissions.

When the best model from experiment 1 was used to detect traffic in a different scenario, it could not detect the vehicles in the image. To mitigate this scenario, 263 images were labelled. This data was split into training/validation/test at (185 (70.3%)/52 (19.8%)/26 (9.9%)).

Using a combination of experiment 2 and the values collected from the sensors, we were able to create a collated table that contains the 6 elements' values and the mean number of vehicles detected in the same time frame.

Project Outcome (400 words)

For experiment 1 (time series forecasting), the best-performing ARIMA model had Mean Absolute Error (MAE) of 0.05160054070964814. The best-performing LSTM model obtained a Mean Squared Error (MSE) of 4.12e-4. The low error rates show that time series forecasting is effective and can be applied to this problem. Experiment 2, with a secondary dataset of aerial images, demonstrated that we could detect vehicles with a high Mean Average Precision (mAP) of 0.934. Follow-up experiments, with footage that we collected at a local intersection in our community using a camera, demonstrated a high mAP of 0.865, showing effectiveness in a local context.

Whilst collecting air pollution data from our sensor, we encountered some problems. Firstly, the height of the images was not ideal as it was placed too low, preventing the object detection model from detecting vehicles. To address this issue, we altered the zoom of the image, creating the effect of an increased height and improving the model's detection confidence. Another difficulty was with the sensor. The sound sensor did not work as expected at first: unfortunately, the readings of the sound sensor were too low in decibels to be feasible. To mitigate this problem, we tried to increase the sensitivity of the sound sensor. Whilst this did work to increase sound, the readings from the sound sensor were still too erratic to be analyzed. In the future, we may incorporate a higher-quality sensor to detect noise.

We believe that through this research we have discovered several different elements that can help in sustainable traffic planning in the future. The positive results were that traffic could be forecasted using machine learning and that traffic could be detected using low-cost aerial sensors and object detection. we have also demonstrated that low-cost sensors can be used to effectively detect pollution levels. The surprising negative result was the minimal correlation between traffic and pollution. This surprising result can be put down for a few reasons: perhaps there was a seasonal variation which affected pollution levels. A second possibility is that each additional car does not substantially affect air pollution levels, leading to the notion that there was not enough traffic at our junction. It is also possible that low-emission vehicles may affect pollution levels, rendering more active measurement necessary. This finding suggests that, in the future, more data should be collected, perhaps in different countries, with higher-sensitivity sensors.

Next Steps (350)

Our solution proposes the use of economic sensors to measure the pollution levels as well as the traffic at any given location on the pavement and use a machine learning model to forecast the future levels of traffic and pollution. In addition, using a sensor such as the one we built, it is possible to measure greenhouse gasses economically and calculate the correlations at different points in time to measure the local impact of traffic on greenhouse gas emissions.

Our goal is a future where companies like Google and Apple integrate our invention, to enable user autonomy in deciding not only the fastest route but also the healthier route to their destination. our models can be integrated into a GPS like Google Maps to route traffic to minimize pollution so there is less stress on the environment. Where air pollution correlates with traffic, our time series models can enable such programs to predict greenhouse gas levels ahead of time. In the future, the Ultra-Low Emission Zone charge (ULEZ) can even be minimized as people take different routes alleviating pressure on the environment. Such integrations can enable citizen involvement in healthy traffic, with options for users to choose environmentally friendly paths.

The path to this long-term goal will involve developing an even better sensor to enhance the accuracy of our research. Not only will we collect data from different countries, but the next step will be to approach our local council with our research. Given the traffic cameras that are present at most intersections, we aim to incorporate the object detection model into the existing cameras so that traffic can be accurately forecast.

Furthermore, having access to the information base provided by our sensors and models would enable communities to raise awareness and take steps to mitigate the impact of pollution. The use of cheap sensors makes the solution much more accessible to communities across the world. Overall, this approach can help communities to not only measure and predict the traffic flow at their locations but also understand its impact on the environment.