Image-based Air Quality Analysis using Deep Convolutional Neural Network Predicting Ambient Air Pollutant Levels

Dr S Ramacharan
Associate Professor, IT Department
GNITS, Hyderabad
s.ramacharan@gnits.ac.in, s.ramacharan@gmail.com

Abstract

The regulation of air pollutant levels is rapidly becoming one of the most important tasks for the governments of developing countries. Among the pollutant index, Fine particulate matter (PM2.5) is a significant one because it is a big concern to people's health when its level in the air is relatively high. However, the relationships between the concentration of these particles and meteorological factors are poorly understood. The project utilizes selected techniques, such as Support Vector Machine(SVM) and Gradient Boost, to predict ambient air pollutant levels based on mostly weather variables. This project attempts to apply some machine learning techniques to predict PM2.5 levels based on a dataset consisting of daily weather and traffic parameters. Due to the uncertainty of the specific number PM2.level, we simplified the problem to be a binary classification one, that is to classify the PM2.5 level into Good, Moderate, Healthy, Unhealthy, Hazardous. The value is chosen based on the Air Quality Level standard which sets 115 ug/m3 to be mild level pollution.

Keywords: **Air Quality Index(AQI), Convolutional neural network, Image Acquisition, particulate matter(PM), Regression Model.**

1. Introduction

This project Air Quality Index Prediction is an online website built on a machine learning model which predicts the air quality index in a region based on the climatic conditions. The air quality index (AQI) is an index for reporting air quality on a daily basis. It is a measure of how air pollution affects one's health within a short time period. The purpose of the AQI is to help people know how the local air quality impacts their health. AQI revolves around the amount of particulate matter (PM) present in the air.

PM is the sum of all solid and liquid particles suspended in air many of which are hazardous. This complex mixture includes both organic and inorganic particles, such as dust, pollen, soot, smoke, and liquid droplets. These particles vary greatly in size, composition, and origin. The AQI is thus calculated by taking the amount of pollutants such as SO2, Nitrate, Ammonia, Ozone etc.

The idea of the project is to take the climatic factors and daily weather forecast into consideration rather than using expensive sensors that can sense the above chemicals. It is reported that weather generally varies within every 5km of radius. The reasons contributing to these changes in weather are humidity, precipitation, barometric pressure, temperature, time of the day, wind-flow and other human factors like pollution from vehicles and factories etc. We know for a fact that Air Quality is at its worst during winter. This drives towards the intuition that correlates the amount of PM present in the air to the climatic factors.

A. Objectives

- Predict AQI in a particular region given their climatic factors
- Achieve high accuracy using efficient Machine Learning Algorithms
- Innovative way to predict AQI
- Cost-Efficient method to calculate particulate matter in the air.
- Build a user friendly website to comprehend the air quality
- Easy and Efficient way that allows people to know the quality of air they breathe
- Sensitive people (who suffer from lung diseases, old and children) can be aware about their surroundings.

B. Scope

The objective of this system is to provide a cost efficient, user friendly website using Python-Django built on a highly efficient Machine Learning model fed with huge training data set and is well tested for high accuracy. This system enables the user to be aware of the surroundings and take precautions if the quality of the air worsens. The correlation between the weather/climatic conditions and the air quality serves as the foundation upon which the system is built.

C. Problem Definition

The AQI website takes in the values of the selected features and runs the highly efficiently developed Machine Learning Model which was fed with accurate, reliable and huge data sets for both testing and training to give the value of PM2.5 present in the air. The PM2.5 value is estimated from the taken climatic factors and generates the hazard level. The hazard levels present are depicted below.

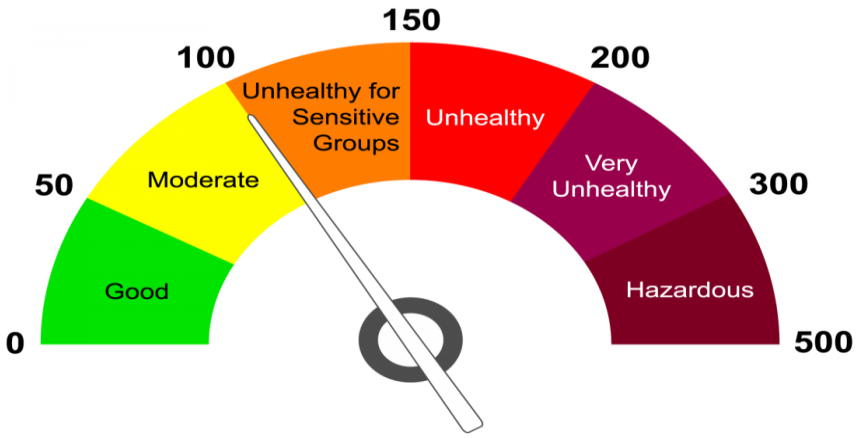

Figure 1.

2. LITERATURE SURVEY

Air pollution and air quality have been buzzwords since two decades. Increase in industrialisation, and concentration of greenhouse gases contributed to rise in global temperatures hence global warming. There has been heavy research going on how to find out how polluted air is and concentration of particulate matter in air using latest technologies like Big Data, Machine Learning, Artificial Intelligence etc.

An ensemble-based model of PM2.5 concentrations across the contiguous United States- an article published in the journal, Environment International, in September 2019 with PM2.5 estimated, the cross-validated R2 was 0.89. The above project used huge datasets with Satellite measurements, land-use terms, and many variables were predictors.

A Machine Learning Project - Image-based air quality analysis using deep convolutional neural network predicts the quality of air by analysing how clear the image is. The cross-validated efficiency came out to be 92%.

2.1 Existing Approaches/ System

There are many existing systems to detect AQI using various algorithms like Support Vector Machine (SVM), CNN, and Gradient Boosting Algorithms. This Project tested all the machine learning algorithms on the data set and the best has been selected with almost 95% accuracy whereas most research papers we came across couldn't reach that accuracy. An excerpt from AIR QUALITY INDEX FORECASTING BASED ON SVM (published in IEEE, Nov, 2019) quotes “The experimental results improve that the proposed method has good AQI forecasting results, and the accuracy of the testing dataset is more than 82.3%”.

A. Drawbacks in Existing System

- Huge amount of data to be fed to the model which comprises chemical composition of air and meteorological data.
- Image based CNN models could get affected due to bad lenses.
- Accuracy less than 90%.

B. Motivation for Proposed System

- This system takes only meteorological data into account for predicting AQI and doesn't need the chemical composition of pollutants which are rather expensive to obtain.
- The accuracy generated is pretty impressive compared to other algorithms priory used. The system generated 92% accuracy on the testing dataset.

3. Overall Description

The Machine Learning concepts have great real-time applications in various fields. Environment field is one such exciting field to predict the situations based on the already available data. It can be of enormous use in Management of natural resources and precautionary steps.

1. **Input:** values for various parameters that are considered

2**. Regression Model**: Input values will be given to the regression model and the predicted value will be given to the output.

3. **Output:** PM2.5 level will be displayed along with the hazard level

A. Product Functions:

The system takes in the values for the selected features and classifies where the AQI falls under. The user should enter valid values (only positive numbers) then the system runs it on the XGBOOST Machine Learning Algorithm to give the PP2.5 value and the hazard level.

B. Design Specification

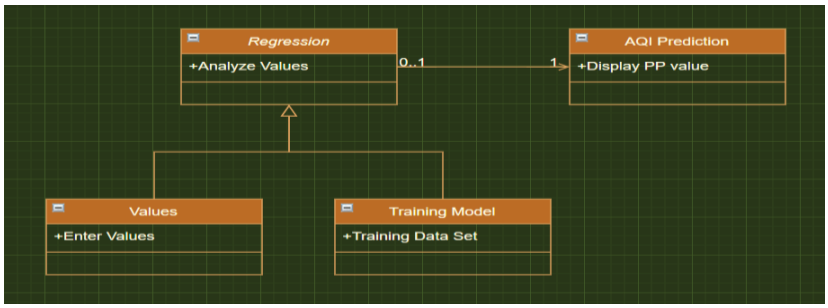

Figure 2.

4. Methodology

The input test image is acquired and pre-processed in the next stage and then it is converted into array form for comparison. The selected database is properly segregated and pre-processed and then renamed into proper folders. The model is properly trained using CNN and then classification takes place. The comparison of the test image and the trained model take place followed by the display of the result. If there is a defect or disease in the plant the software displays the disease along with the remedy.

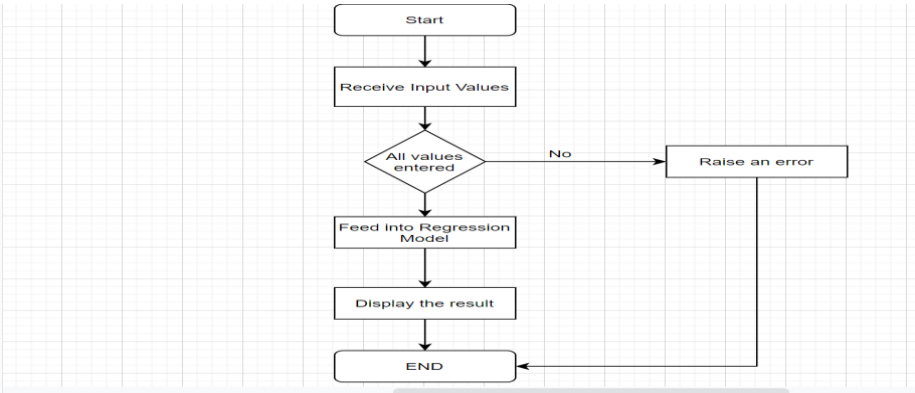

Figure 3.

System Architecture

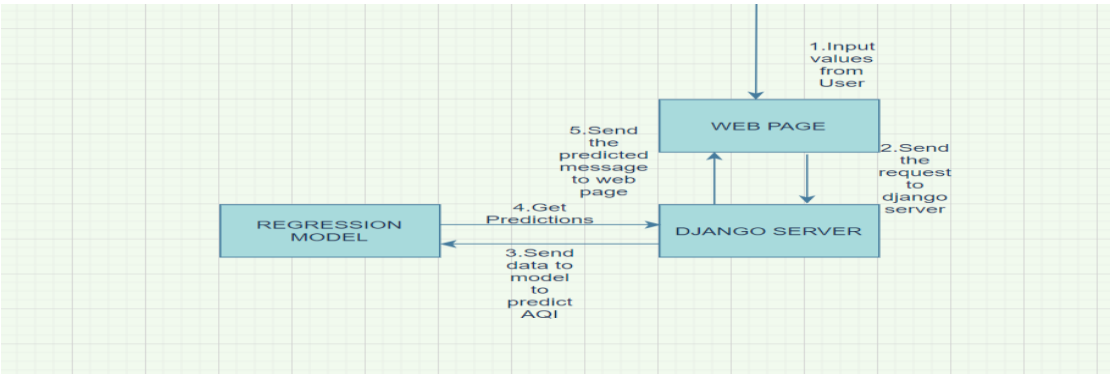

Figure 4.

5. Results and Conclusion

A. Result Analysis / Performance Analysis

Model Accuracy has been increased by increasing the number of epochs. Here we used 500 epochs. Finally got the accuracy of 96.6%.

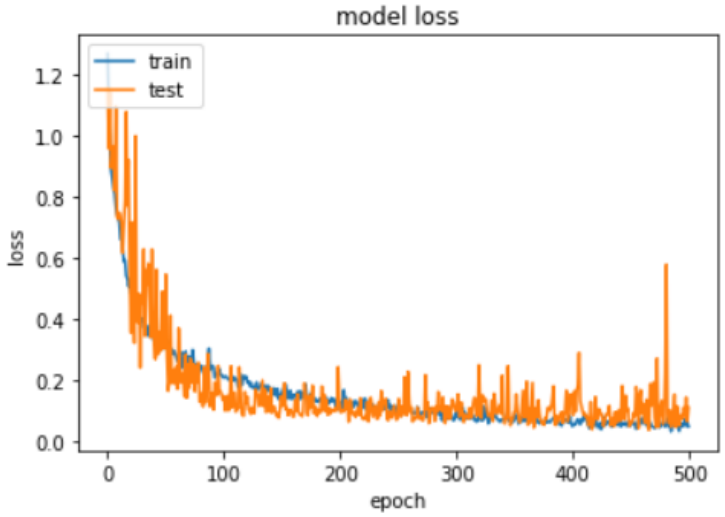

Figure 5.

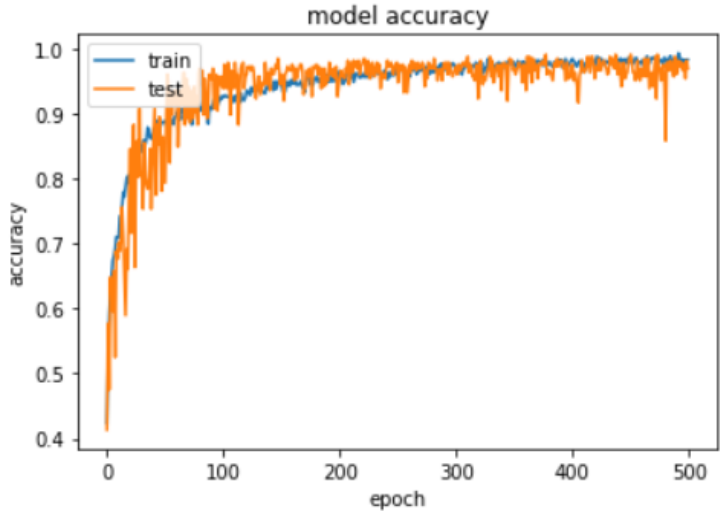

Figure 6.

6. Conclusion & Future Scope

An application of detecting the plant diseases and providing the necessary suggestions for the disease has been implemented. Hence the proposed objective was implemented on the cotton plant. The diseases specific to cotton plant were considered for testing of the algorithm. The experimental results indicate the proposed approach can recognize the diseases with a little computational effort. By this method, the plant diseases can be identified at the initial stage itself and the pest control tools can be used to solve pest problems while minimizing risks to people and the environment.

In order to improve disease identification rate at various stages, the training samples can be increased with the optimal features given as input conditions for disease identification

and fertilization management of the crops. As a part of Future Enhancement the complete process described in this project can be automated so that the result can be delivered in a very short time. Further enhancements can include upgrading user interface and the accuracy to detect specific diseases along with product recommendations.

References

[1] *Pandey, Gaurav, Bin Zhang, and Le Jian. "Predicting submicron air pollution indicators: a machine learning approach." Environmental Science: Processes & Impacts 15.5* ***(2013)****: 996-1005.*

[2] *Athanasiadis, Ioannis N., et al. "Applying machine learning techniques on air quality data for real-time decision support." First international NAISO symposium on information technologies in environmental engineering (ITEE'2003), Gdansk, Poland.* ***(2003)****.*

[3] *Ioannis N. Athanasiadis, Kostas D. Karatzas and Pericles A. Mitkas. "Classification techniques for air quality forecasting." Fifth ECAI Workshop on Binding Environmental Sciences and Artificial Intelligence, 17th European Conference on Artificial Intelligence, Riva del Garda, Italy, August* ***(2006)****.*

[4] *M. Caselli & L. Trizio & G. de Gennaro & P. Ielpo. "A Simple Feedforward Neural Network for the PM10 Forecasting: Comparison with a Radial Basis Function Network and a Multivariate Linear Regression Model." Water Air Soil Pollut* ***(2009)*** *201:365–377.*

[5] *S. Bordignon, C. Gaetan and F. Lisi, "Nonlinear models for ground-level ozone forecasting." Statistical Methods and Applications, 11, 227-246,* ***(2002)****.*