Microenvironment Pollution Detection Using Plant Leaves

Yuvvan Talreja
The Cathedral & John Connon School
Mumbai, India
talrejayuvvan@gmail.com

Reetu Jain
On My Own Technology Pvt Ltd
Mumbai, India
reetu.jain@onmyowntechnology.com

***Abstract*— Pollution is the introduction or presence of harmful substances or contaminants into the natural environment, including air, water and soil. In the past decade there has been a rapid rise in pollution because of an increased number of motor vehicles, industrial activities and hazardous waste disposal. Air pollution has had major adverse effects on plant growth and development primarily leading to damage to leaves. This study aims to use deciduous plant leaves to predict the air pollution present in the surrounding environment. Using machine learning, the morphology, anatomy and physiology of the leaf is analysed and classified to identify the main chemical in the surrounding environment. The chemicals considered in this experiment include Sulphur Dioxide, Fluorine, Nitrogen Oxides and Particulate Matter. To employ this approach, a dataset of over 5000 images was collected from different places around Mumbai.**

Keywords*— *Air Pollution, CNN, Machine Learning, Plant, Image Processing.

I. Introduction

In 2020, the World Air Quality Report stated that the pollution level of Mumbai was 41.3micrograms/m3, while in March 2022, it was reported to be 45 micrograms/m3. Besides Mumbai, Delhi is also one of the most polluted cities in the world on par with Beijing, Hong Kong and Dhaka. Delhi has an average pollution level of 140 micrograms/m3. This is almost 10 times greater than that of Mumbai. This pollution has been extremely detrimental to the health of humans in

these cities, but also to the flora and fauna that surrounds these cities. In Mumbai especially, the mangroves on the outskirts of the city are severely affected by the pollution. These mangroves have a very thin cuticle that is highly penetrable. Harmful chemicals such as O3 and NOx make their way through the cuticle and directly affect the surfaces of the leaves. The chemicals also diffuse into the stomata while the plants respire and photosynthesis through diffusion. The chemicals interfere with the metabolic processes of the leaf, thus making them far less efficient. Furthermore, the chemicals affect the structural characteristics of the leaves such as the stomatal density, vein diameter and chloroplastic contents of the leaf.

The Nox and SO2 specifically, make the chlorophyll present in the chloroplasts much thinner thus exposing the other pigments such as Xanthophyll and Carotene.

This pollution has several causes: [3].

A. Vehicle exhausts

Vehicle exhausts contribute to air pollution by emitting particulate matter (PM), nitrogen oxides (NOx), carbon monoxide (CO), volatile organic compounds (VOCs), sulfur dioxide (SO2), and greenhouse gases like carbon dioxide (CO2). These pollutants can cause respiratory problems, cardiovascular issues, smog formation, acid rain, and contribute to climate change. Implementing stricter emission standards, promoting cleaner fuels, and encouraging electric vehicle usage are important measures to mitigate the negative effects of vehicle exhausts on air quality.

B. Industries/Factories

Due to stricter emission limits, dense smog and smoke from industrial pollution are now less frequent in wealthy nations. Sulfur dioxide, volatile organic solvents, and particulate particulates, such as metal dust, are the main industrial pollutants that have an impact on air quality. Dioxins and other dangerous chlorinated chemicals can be produced when garbage, especially plastics, are burned. Acid rain is created when sulphur emissions combine with water to form sulfuric acid (H2SO4) particles. Acid rain can harm trees and surface water miles from its source.

II. PLANT RESPONSE TO AIR POLLUTION

The buildup of resources is hampered by air pollution, which is bad for plant development. Many air contaminants, such as O3 and NOx, prevent the plant canopy from fixing carbon dioxide and prevent leaves from functioning metabolically once they are exposed to the atmosphere. Root development is negatively impacted by heavy metals and other air pollutants that are first deposited on the soil, and plants are unable to efficiently use soil resources as a result. Due to these reductions in resource capture, which include the production of carbohydrates through photosynthesis, nutrient uptake from minerals, and water uptake from the soil, different resources will be allocated to different plant structures.The impact on growth when water stress and air pollution stress co-occur depends on a complex interaction of processes within the plant. The competitive dynamics within an ecosystem may change as a result of air pollution, which may also change the composition of the nearby plant community. The effects of these changes in agroecosystems could manifest as a decline in economic yield.

III. EFFECTS

These days, there is a big problem with how air pollution affects vegetation. The effects of air pollution on plants and air pollutants have a detrimental impact on plant growth, primarily through interfering with resource gathering. Particularly O3 and NOx, air pollutants can cause harm to leaf structure. Impacts of Air Pollution on Plants result in the deposition of contaminants in the soil, such as heavy metals, which first affect the roots and interfere with the plant's ability to capture soil resources. Via adjustments to the asset parts of the various plant structures, these decreases in asset catch will have an impact on the development of the plant.Several pressures, such as water pressure, are brought on by the effects of air pollution on plants. The impact of air pollution on vegetation has the potential to harm various plant species both now and in the future. brief.

IV. LITERATURE REVIEW

Biological indicators for pollution detection in terrestrial and aquatic ecosystems

The use of biotas, or plants, microbes, and animals, to detect environmental pollutants is discussed in this paper by Swakzi Kooren. Here are the most important biological signs. The mentioned plant indicators include higher plants, while the microbiological indicators include bacteria, fungi, algae, planktons, lichens, helminths eggs, and enzymes, while the animal indicators are earthworms, macro-invertebrates, frogs & toads, insects, and animal poisons.

Certainly, the quantitative response in plant growth is easier to recognize and express than the qualitative influences using the tolerance index (TI) which is calculated as follows:

Tolerance index = growth in polluted soil/growth in unpolluted soil.

A. Advances In Image Processing For Detection Of Plant Diseases

A software solution for the automatic identification and classification of plant leaf diseases in Allen Watkins' book Advances In Image Processing For Detection Of Plant Diseases is tested. Studies of a plant's visually discernible patterns are referred to as studies of a plant characteristic or disease. Many characteristics and diseases affect crops today. One of the main characteristics/diseases is damage to the insect. Because some types of birds may be hazardous to insecticides, their efficacy is not always shown. Moreover, natural animal food systems are harmed. After the segmentation phase, the next two phases are added one after the other. Initially pixels that were primarily green in colour were identified. Following that, these pixels are masked according to predetermined threshold values that are determined by Then, using Otsu's technique, those primarily green pixels are covered. The pixels on the edges of the infected cluster (object) and those with red, green, and blue values of zero were completely deleted as an additional step. The results of the experiments show that the suggested strategy is a reliable method for identifying illnesses of plant leaves. With a precision of between 83% and 94%, the created algorithm can correctly identify and categorise the studied diseases, and it can run 20% faster than the method suggested in Using image processing methodology to detect plant diseases:

Using a variety of image processing techniques and an artificial neural network, Micheal Lee presents a methodology for early and accurate detection of plant illnesses in his paper "Applying image processing approach to identify plant diseases" (ANN). It is extremely difficult for farmers to switch disease control strategies. Many plant diseases pose a serious threat to the agricultural industry by shortening the life of the plants, making it expensive to rely solely on naked eye inspection to detect and diagnose diseases. The goal of the current work is to create an easy-to-use disease detection system for plant diseases. The first step in the process is image capture. Gabor filter used for filtering and segmentation. Following the segmentation process, texture and colour features are extracted, and an artificial neural network (ANN) is trained by selecting feature values that can accurately distinguish between healthy and diseased samples. According to experimental findings, classification performance using an ANN with a feature set has a 91% accuracy rate. Satellite image processing and air pollution detection.

Air pollution detection and satellite image processing are discussed in the article. Digital processing of observed signals and images is strongly related to environmental sensing. The focus of the work is the analysis of mathematical techniques for detecting aerosol particle concentrations seen at ground measurement stations and by satellites. The basic techniques for two-dimensional interpolation are presented in the contribution's first section, enabling the estimation of the observed variables across the entire region of interest.

In the past few years, a number of sensors for detecting environmental pollution have been developed. By offering information about the chemical breakdown of the environment, these are utilised to assist governments and organization. the gas analyzer.

V. METHODOLOGY

The aforementioned diagrams provide a visual representation of the methodological examination of the work that has been given. The first step in the procedure is to take microscopic pictures of the surface of several leaves. These leaves include those of the neem plant and the brinjal plant (Solanum melongena L.) (Azadirachta indica). The leaves themselves are carefully selected from various locations throughout Bombay, where the air is heavily polluted with chemicals and other contaminants. A minuscule camera with a 1000x magnification took these pictures. Afterwards, a dataset of more than 5000 photos is created from these photographs. The information is then divided into leaves influenced by various substances, including particulate matter, nitrogen dioxide, sulphur dioxide, and fluorine. The dataset is then split into a training set and a test set, with 80% of the data earmarked for the machine learning algorithm's training and the remaining 20% being used to evaluate the algorithm's accuracy and other metrics.

A. *Image Processing*

The train dataset was taken with various different distances and angles, therefore the images had to be cropped and rotated through image augmentation and rescaling to make it uniform and making the training process for the model much easier.

1) *Augmentation:* Image augmentation is a method used to artificially expand a dataset. A deep learning algorithm requires a great amount of data to be trained. Hence, image manipulation such as zoom, shear and and rotate are used to create a new set of images that can increase the count of images in the dataset. If a neural network is trained on a limited number of images it tends to over fit. Overfitting is a problem where the evaluation of machine learning algorithms on training data is different from validation data. In this dataset an Image Data generator was used from the Keras API to make 9 images out of a single input image. These images were a combination of rotation by 30, 45 and 90 degrees, a 2 or 3x zoom and shearing in any one axis.

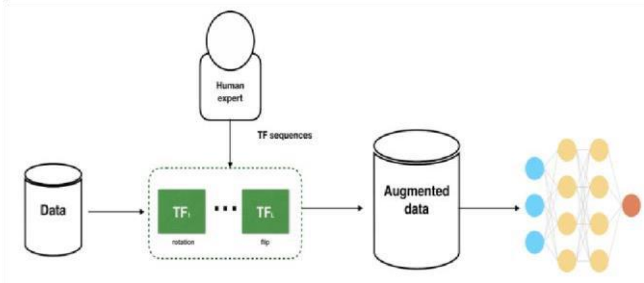


Fig. 1

1) *Rescaling:* To solve the issue of inconsistent images, the process of image rescaling is used. This is when the sizes of the images are downscaled to the smallest image available in the dataset. Most machine learning models train faster on a smaller set of images due to the fact that a smaller number of pixels need to be processed. Large images cause the training time to increase and also complicate the architecture of deep neural networks.

2) *Color conversion:* Deep Neural networks that heavily rely on computer vision and detection and extraction of features require the colours of images to be consistent and uniform throughout the dataset. To maintain this continuity the colours of all the images in the dataset are converted from a RGB to HSV colour space. This is done by projecting the RGB cube shaped colour space to a hexagon. In the diagram above 'H' represents the hue and 'S' represents the saturation of the image.

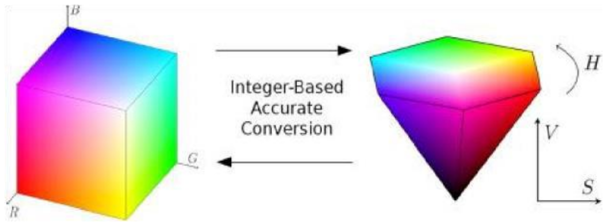


Fig. 2

B. *B. Feature Extraction*

The third step of the proposed approach is to split the Finding and extracting information from photos that can be utilised to interpret a specific sample's meaning is the goal of this approach. Keypoint analysis, masking, and contour analysis were used to achieve this.

1) *Key point anaylsis:* Keypoint detection or analysis is the process of detecting objects by confining their key features in a localized space. Key Points are special locations on a 2D plane that outline the unique factors of an image. The key points of each image in the dataset are matched and images with the

least matches are either discarded or given less weightage while the algorithm processes the dataset.

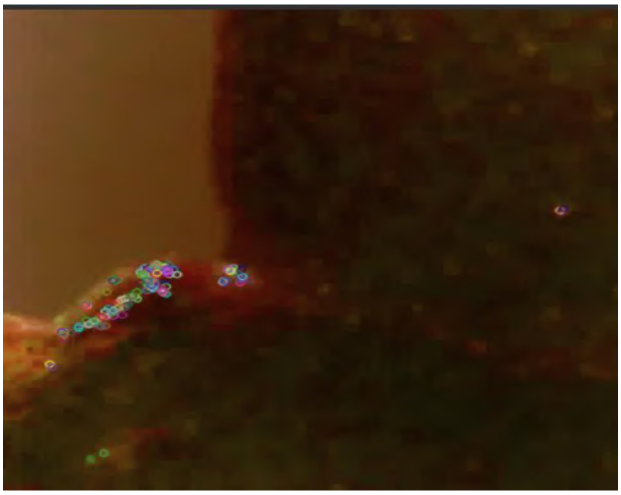


Fig. 3

2) Masking: Masking is the process used in edge detection and motion detection where a temporary overlay known as a 'mask' consisting of zero and non-zero values are used to define or outline a particular portion of an image. This technique is primarily used to find the area of a particular portion of an image by calculating the percentage of zero pixels in the image.

Fig. 4

3) Contour Analysis: Contour analysis is the process used to detect the edges and boundaries of objects by analysing the curvature profile on the image. The profile which is a two dimensional function is projected into the third dimension to create a mesh as shown below. The inconsistencies in this 3 dimensional mesh are then analysed to approximate the position of the border in the three dimensional space. The closed border of the image is then identified using Green's Theorem which relates the boundary line to the area enclosed by that line.

$$\oint_C^{\infty} (L\,dx + M dy) = \iint_D^{\infty} \left(\frac{\partial M}{\partial x} - \frac{\partial L}{\partial y}\right) dx\,dy$$

Green's theorem
'C' represents the closed curve
'D' is the region bounded by the closed curve

Fig. 5

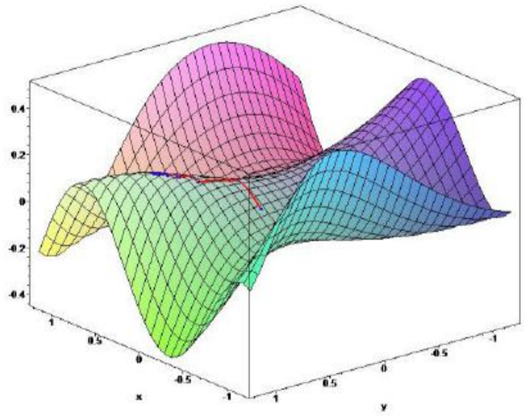


Fig. 6

C. *Recognition and classification*

After the preprocessing, the number of contours plotted on the image are extracted along with the area enclosed by the contours and the percentage area that the contours covered. This information is then saved in a CSV file which is used to train the decision tree classifier machine learning model that classifies the images.

D. *Hypertuning of Decision Tree Classifier*

In figure (9), the emotions angry, disgust, fear, happy, A test dataset and a train dataset were created from the collection of 4200 photos.

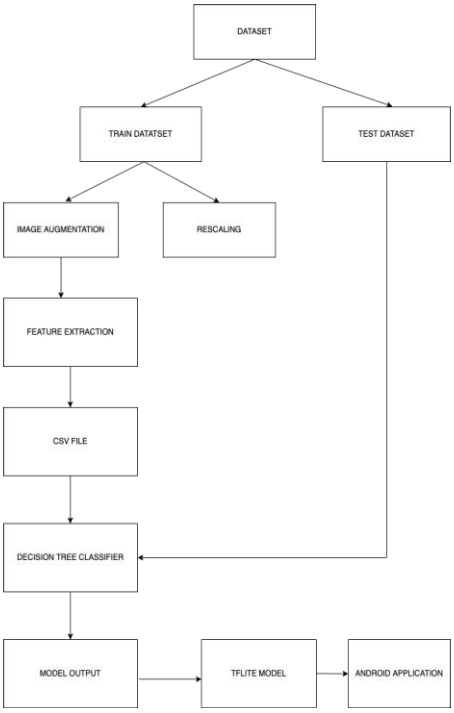


Fig. 6

The model was trained using 80% of the data, and it was tested using 20% of previously unreleased fresh data. The photos needed to be cropped and rotated through image augmentation and rescaling to make them uniform and make the model training process more simpler because the train dataset was shot at many distinct distances and angles. The features of the augmented photos were then extracted. This was accomplished by using key point analysis to pinpoint the important and recognisable points in the image, which were then compared to the other photos in the collection. Following that, the photos with the fewest matches were deemed abnormal and eliminated. The features were retrieved from the remaining photos using contour detection. By comparing the RGB colour values of the photos, contours were formed, which were then painted on the images using the CV2 library. The contours were then extracted from the photos using masking. Utilizing the masks, the contour numbers, contour area, and predicted 'label' of the images—which had four categories—were retrieved. These categories were 0 for healthy, 1 for fluorine-, 2 for sulphur dioxide-, and 3 for dust-related effects.

VI. ALGORIT HM

A. IMAGE PROCESSING

This function is mainly used for image processing and creating the contour lines on the images so that features can be extracted. The function takes an input of the image number and masks it by using the cv2.inRange() function. Cv2 is then used to find the contours and store them in 2 variables. The number of contours are then counted by finding the length of the array of contours and the percentage area is then calculated and stored in another array. These two arrays are then returned to be used later. model.

B. MACHINE LEARNING

Decision Tree Classifier from Sklearn Library. Model splits up data based on critical points known as nodes/. Each area that has been split up is known as a leaf. When a new datapoint is introduced, the model decides which leaf to put it in based on its features, thus classifying it.

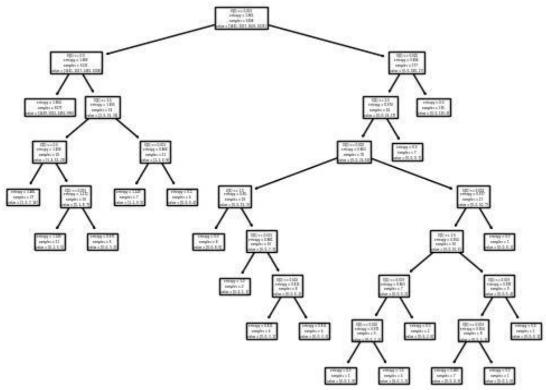


Fig. 6

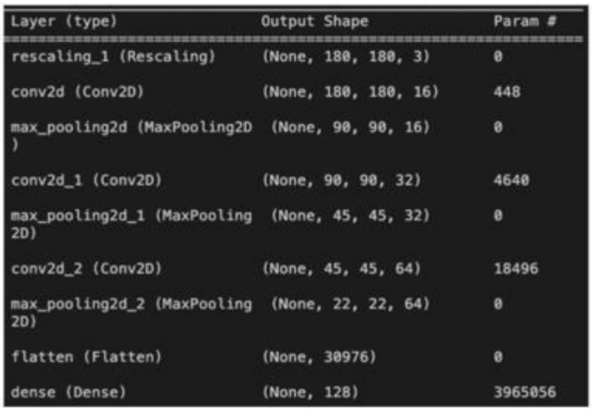
C. CONVOLUTIONAL NEURAL NETWORK

Convolutional neural networks were utilised in this project as an alternative model to the decision tree classifier. The Decision Tree had to be adjusted and changed multiple times until the requisite precision was attained. Hence, a deep learning algorithm like a CNN was needed. The CNN was trained using the same dataset. The dataset was divided into two parts: 20% of the data were used to validate the model and 80% of the data were used to train it. To ensure that all of the photos had the same pixel dimensions, the photographs were first rescaled.

Subsequently, three different convolution layers and a maximum pooling layer were applied to the data. After sets of convolution and pooling layers were best suited for our issue and the CNN's performance. Rectified Linear Unit Function was the activation function employed. This solution was the most methodical for our issue and also gave the neural network access to the large and intricate dataset that was supplied. To achieve the final classification, the output was flattened into a 1D array and placed through two further dense layers. Adam was the optimizer that was used to compile the data.to This made it possible for us to successfully train the model's weights using the squared gradient, a variation on the stochastic gradient descent approach.

VII. RESULT AND DISCUSSION

The algorithm used in the application got a 69.2% accuracy and this metric is being improved. The solution was tested near the beach, near mangroves and near buildings and roads. Near the buildings there was SO2 and particulate matter while in the mangrove and beach area there was major Nitrous oxide impurity. The algorithm also had difficulty in identifying leaves which were not green. Furthermore, there were several bugs where all leaves were being detected as healthy.



Layer (type)	Output Shape	Param #
rescaling_1 (Rescaling)	(None, 180, 180, 3)	0
conv2d (Conv2D)	(None, 180, 180, 16)	448
max_pooling2d (MaxPooling2D)	(None, 90, 90, 16)	0
conv2d_1 (Conv2D)	(None, 90, 90, 32)	4640
max_pooling2d_1 (MaxPooling 2D)	(None, 45, 45, 32)	0
conv2d_2 (Conv2D)	(None, 45, 45, 64)	18496
max_pooling2d_2 (MaxPooling 2D)	(None, 22, 22, 64)	0
flatten (Flatten)	(None, 30976)	0
dense (Dense)	(None, 128)	3965056

Fig. 7

Figure 8. Diagram showing accuracy of test set

Figure 9. Diagram showing accuracy of training set

REFERENCES

[1] R. I. Al-Dulaimi, N. Ismail, and M. H. Ibrahim, “Responses of growth of lady’s fingers (Abelmoschus esculentus L.) to different treatments methods of dairy wastewater,” Annals of agricultural and environmental medicine: AAEM, vol. 21, no. 1, pp. 42–48, 2014, Available: https://pubmed.ncbi.nlm.nih.gov/24847548/

[2] S. P. Slavikova, “What Is the Effect of Pollution on Plants? | Greentumble,” Greentumble, Jan. 21, 2022. https://greentumble.com/effect-of-pollution-on-plants#:~:text=Plants%20usually%20show%20damage%20in (accessed Oct. 13, 2023).

[3] T. Valente et al., “Image processing tools in the study of environmental contamination by microplastics: reliability and perspectives,” Environmental Science and Pollution Research, Jul. 2022, doi: https://doi.org/10.1007/s11356-022-22128-3.

[4] A. Chakma, B. Vizena, T. Cao, J. Lin, and J. Zhang, “Image-based air quality analysis using deep convolutional neural network,” IEEE Xplore, Sep. 01, 2017. https://ieeexplore.ieee.org/abstract/document/8297023

[5] X. Jia et al., “Mapping soil pollution by using drone image recognition and machine learning at an arsenic-contaminated agricultural field,” Environmental Pollution, vol. 270, p. 116281, Feb. 2021, doi: https://doi.org/10.1016/j.envpol.2020.116281.

[6] Y. Lin et al., “An optimized machine learning approach to water pollution variation monitoring with time-series Landsat images,” International Journal of Applied Earth Observation and Geoinformation, vol. 102, p. 102370, Oct. 2021, doi: https://doi.org/10.1016/j.jag.2021.102370.

[7] W. Yi, K. Lo, T. Mak, K. Leung, Y. Leung, and M. Meng, “A Survey of Wireless Sensor Network Based Air Pollution Monitoring Systems,” Sensors, vol. 15, no. 12, pp. 31392–31427, Dec. 2015, doi: https://doi.org/10.3390/s151229859.

[8] T. Cheng, F. Harrou, F. Kadri, Y. Sun, and T. Leiknes, “Forecasting of Wastewater Treatment Plant Key Features Using Deep Learning-Based Models: A Case Study,” IEEE Access, vol. 8, pp. 184475–184485, 2020, doi: https://doi.org/10.1109/ACCESS.2020.3030820.

[9] “Effects of Air Pollution on Agricultural Crops,” Gov.on.ca, 2003. http://www.omafra.gov.on.ca/english/crops/facts/01-015.htm