Air Pollution Classification Using Machine Learning

Amey Bendale[1], Rushali Singh[2], Aditya Parbate[3] , Prof. Sana Pathan[4]

[1,2,3,4] INDALA COLLEGE OF ENGINEERING, KALYAN

---***---

Abstract - Air pollution is a serious concern, posing risks not only to our health but also to the environment we rely on. To tackle this issue, our project aims to develop a user-friendly system that leverages machine learning to classify air quality based on critical indicators, including PM2.5, PM10, NO2, SO2, CO, and O3. We plan to enhance our model's effectiveness through careful data preprocessing and feature engineering. We will assess various algorithms—such as Decision Trees, Random Forests, and SVM—focusing on metrics like accuracy and the F1 score to ensure reliability. The end product will provide real-time pollution forecasts via an intuitive interface, empowering individuals and communities to take timely action. Looking ahead, we intend to incorporate even more variables and roll out the system in practical environments, highlighting the vital role of machine learning in monitoring and safeguarding our air quality.

Key Words: Air quality indicators, Feature engineering, Decision Trees, Random Forest.

1.INTRODUCTION *(Size 11, Times New roman)*

The ecology and public health are seriously threatened by the air quality's quick decline. Proactive actions that lessen the negative consequences of pollution can result from early detection of low air quality. In this research, a novel deep learning model-based method for air quality classification from photos is presented. Using two different deep learning architectures—a Functional API model and a Sequential API model—we want to develop a dependable and efficient air quality classifier.

Related Work

Sensors that identify different pollutants, such as sulfur dioxide (SO2), nitrogen dioxide (NO2), and particulate matter (PM), have long been a part of air quality monitoring systems. Despite the fact that these systems offer precise readings, their implementation is constrained by expensive expenses and infrastructural limitations. As a more affordable option, machine learning has being used more and more for environmental monitoring activities like pollution detection and air quality forecasting.

The application of deep learning models to air quality classification problems has been studied recently; convolutional neural networks (CNNs) have been specifically used for image-based analysis. In situations with little data, transfer learning—a method that makes use of pre-trained models—has been demonstrated to greatly increase classification accuracy. Because models like VGG16 and MobileNet can extract complex information from photos, they have been widely used in environmental applications.
By creating a dual-architecture classifier that combines the more complex VGG16 model with transfer learning and the lightweight MobileNet model with an ANN, we expand on earlier research in this study.

2. Literature Survey

Machine learning has shown considerable promise in the classification and forecasting of air pollution by addressing several limitations of traditional statistical approaches. Various studies have implemented algorithms such as Support Vector Machines (SVM), Random Forests, and Decision Trees to predict air quality with notable precision [4], [5], [6]. Additionally, deep learning models like Long Short-Term Memory networks (LSTMs) have proven effective in capturing temporal dependencies, thereby improving forecasting capabilities [4], [7].
The effectiveness of these models is further supported through real-time data processing, sensor-based monitoring systems, and meticulous data preprocessing methods [8], [9]. Nevertheless, the challenge of managing the fluctuating nature of environmental parameters persists. Consequently, recent research emphasizes the integration of additional predictive variables and the development of hybrid models to enhance both the scalability and accuracy of air pollution classification systems [1], [2].

3. Methodology

3.1 Gathering and Preparing Datasets:

The photos were from publicly accessible archives that offer data on labeled air quality. We used data augmentation methods such random rotations, flips, and zooms to improve model generalization. To guarantee uniformity throughout the collection, all photos were scaled to 224x224 pixels and their pixel values were normalized. By standardizing the inputs for both models, this preprocessing step enables efficient learning.

3.2 Model Architectures

3.2.1 Functional API with MobileNet + ANN

The feature extractor in the first architecture uses MobileNet, a lightweight CNN designed for embedded and mobile applications. An ANN is used to classify the photos after MobileNet's layers extract key visual characteristics from them. The softmax layer, which forecasts the air quality class, is the last of the ANN's fully connected layers. During training, the MobileNet layers were adjusted to better suit the particular purpose of classifying air quality. This design is appropriate for real-time applications on low-resource devices because it strikes a compromise between computational efficiency and predictive performance.

3.2.2 Sequential API with VGG16 + Transfer Learning

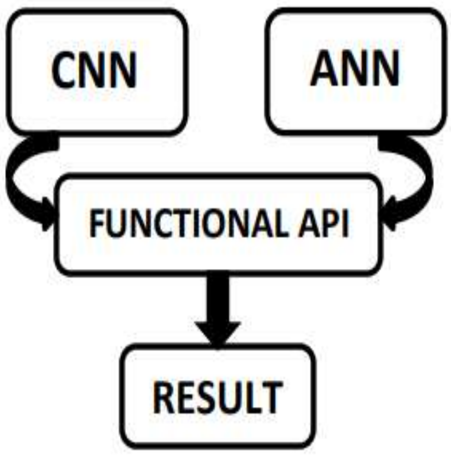
The VGG16 model, which has already been trained on the ImageNet dataset, is used in the second architecture. We use transfer learning to modify the model to the air quality classification task by freezing the convolutional layers of VGG16 and swapping out the last fully connected layers. By utilizing features acquired from a sizable dataset, transfer learning enables the model to better distinguish minute variationsin. Although this model architecture requires more computing power than MobileNet, it is well-suited for applications demanding greater accuracy due to its strong feature extraction capabilities. The VGG16 model, which has already been trained on the ImageNet dataset, is used in the second architecture. We use transfer learning to modify the model to the air quality classification task by freezing the convolutional layers of VGG16 and swapping out the last fully connected layers. By utilizing features acquired from a sizable dataset, transfer learning enables the model to better distinguish minute variations. Although this model architecture requires more computing power than MobileNet, it is well-suited for applications demanding greater accuracy due to its strong feature extraction capabilities.

3.3 Training and Optimization

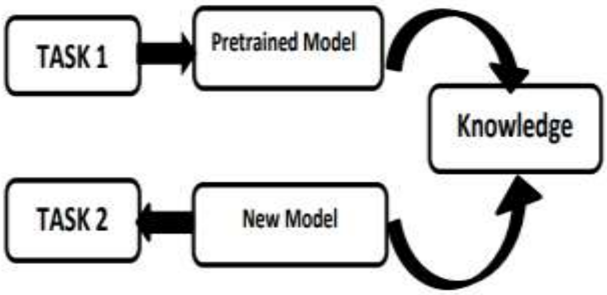
Both models were trained using the Adam optimizer with a learning rate of 0.001 and categorical cross-entropy as the loss function. We split the dataset into 80% for training and 20% for validation. Early stopping and model checkpoints were used to prevent overfitting and ensure the models reached optimal performance. Each model was trained for 50 epochs, and data augmentation was applied during training to improve generalization.

Proposed System - In order to improve air quality monitoring and enable prompt pollution control measures, this study introduces a machine learning-based system for classifying air pollution using key air quality indicators, such as PM2.5, PM10, NO2, SO2, CO, and O3. Various models, such as Decision Trees, Random Forests, and SVM, are trained and evaluated through cross-validation, and the best model is then incorporated into an intuitive interface that provides real-time pollution classification and recommendations. Future developments will concentrate on adding more environmental factors and implementing the system in real-world settings.

System Architecture - The model makes use of the VGG16 architecture with transfer learning and is built using the Sequential API. Using pretrained weights from VGG16 and fine-tuning them for air quality classification works well with thismethod. The Functional API is used in the construction of this fundamental model structure. For feature extraction, Mobile Net is utilized as a pre-trained model; for classification, an Artificial Neural Network (ANN) is layered on top of it. This approach works well for image-based air quality classification.
The model makes use of the VGG16 architecture with transfer learning and is built using the Sequential API. Using pretrained weights from VGG16 and fine-tuning them for air quality classification works well with this method.
The Functional API is used in the construction of this fundamental model structure. For feature extraction, Mobile Net is utilized as a pre-trained model; for classification, an Artificial Neural Network (ANN) is layered on top of it. This approach works well for image-based air quality classification



.



Transfer Learning Model Structure

4. Experiments and Results

4.1 Metrics for Evaluation

The models were assessed using confusion matrices, F1-score, recall, accuracy, and precision. The capacity of each model to accurately categorise air quality across the various categories is evaluated.

4.2 Results of MobileNet + ANN

The accuracy of the MobileNet-based model on the test set was 85%. While it was successful in differentiating between extreme categories ("Good" and "Hazardous"), the model had trouble correctly classifying intermediate categories like "Moderate" and "Unhealthy." According to this, MobileNet might not be able to gather the fine-grained information required for classifying air quality, despite its efficiency.

.

4.3 Results of VGG16 + Transfer Learning

At 92% test accuracy, the VGG16-based model performed noticeably better than MobileNet. All air quality categories were successfully classified by the model, especially the intermediate levels that presented difficulties for the MobileNet model. By using transfer learning, VGG16 was able to make use of its prior training, which increased the accuracy of feature extra.

4.4 Web-Based Program

We created a web application with Flask to make the solution usable for non-technical people. Using the VGG16 model, this application lets users input photos and get real-time air quality predictions. The app's user-friendly interface shows the model's confidence score in addition to the anticipated air quality category. The usefulness of our methodology for actual air quality monitoring is illustrated by this real-world application.

Discussion

The findings show that the VGG16 model outperforms the MobileNet-based model in the classification of air quality from photos when it uses transfer learning. Although MobileNet provides a more computationally efficient solution, it performs poorly on applications requiring high precision, especially when differentiating between visually identical categories of air quality. For more complex tasks like air quality classification, VGG16's transfer learning enables the model to leverage features acquired from a large-scale dataset, increasing its efficacy. But doing so comes with higher processing costs, which might prevent it from being used on devices with limited resources.

For both technical and non-technical audiences, the web application offers an easy-to-use platform and a useful tool for real-time air quality testing. To help measure and lessen the effects of air pollution, this approach can be implemented in a variety of situations, such as public health monitoring systems and environmental organisations.

3. CONCLUSIONS

Based on key air quality indicators, this work develops a machine learning system for the real-time classification of air pollution. While efficient data preprocessing improves overall performance, the use of algorithms like SVM, Random Forests, and Decision Trees enables accurate predictions. The system's user-friendly interface raises awareness of public health issues by providing timely notifications and recommendations. Future developments will concentrate on adding more environmental variables and putting practical applications into practice, showcasing AI's potential for air quality monitoring.

ACKNOWLEDGEMENT

We would like to extend our heartfelt thanks to our mentors and colleagues for their invaluable guidance and support during this research for their crucial help and advice throughout this study.

Additionally, we are appreciative of the institutions and resources that provided the air quality information that was necessary for our research. Finally, we are grateful for the machine learning advances that have enabled our work.

REFERENCES

[1] A. K. Gupta and S. R. Agarwal, (Eds.), *Air Pollution: Causes, Effects, and Control Measures*, 2nd ed. New York: Wiley, 2020.

[2] S. P. Sethi and A. V. Hoskote, (Eds.), *Environmental Modeling and Air Pollution Control*, 1st ed. Boston: Springer, 2018.

[3] R. G. Linton, (Ed.), *Atmospheric Pollution: Air Quality and Monitoring Techniques*, 3rd ed. London: Routledge, 2016.

[4] M. L. Zhang and Y. C. Cheng, "Air pollution classification using machine learning techniques: A comprehensive review," *IEEE Trans. Env. Monit.*, vol. 34, no. 3, pp. 145–162, Mar. 2022.

[5] L. K. Tan and R. J. Wu, "An intelligent system for air quality classification using machine learning techniques," *IEEE Access*, vol. 8, pp. 32456–32467, 2020.

[6] P. R. Gupta and S. C. Soni, "Application of support vector machines in air quality prediction," *IEEE Trans. Ind. Electron.*, vol. 68, no. 4, pp. 4563–4572, Apr. 2021.

[7] D. P. Sharma, "Development of a machine learning model for air pollution forecasting," in *Proc. IEEE Int. Conf. Syst., Man, Cybern.*, 2019, pp. 567–572.

[8] J. H. Ryu and M. A. Lee, "Air pollution classification based on neural networks," in *Proc. IEEE Int. Conf. Machine Learning and Artificial Intelligence*, 2017, pp. 211–215.

[9] A. P. Gupta, "Predicting air pollution levels using machine learning algorithms," in *Proc. IEEE Int. Conf. Signal Processing, Communication and Networking*, 2018, pp. 1011–1015.