Air Pollution Detection Using CNN

1 Mr.Himambasha Shaik 2. Supraja Bakkamunthala

1 Assistant Professor, Department of Master of Computer Applications,
QIS College ofEngineering & Technology, Ongole, Andhra Pradesh, India
2 PG Scholar, Department of Master of Computer Applications,
QIS College of Engineering & Technology, Ongole, Andhra Pradesh, India

ABSTRACT

The world's population is becoming increasingly urbanised, resulting in degradation of air quality in many of the world's fastest expanding cities. According to a study of air pollution in 20 of the world's 24 megacities, ambient air pollution concentrations are at levels that have been linked to substantial health impacts. While it is well acknowledged that air pollution is a significant issue that impacts public health, agriculture, and urban growth, there are clear limits in current approaches for obtaining accurate air pollution data. The feasibility and accuracy of utilising a convolutional neural network on satellite pictures to forecast the Breezometer Air Quality Index (BAQI) are investigated in this study. The capacity to precisely assess air pollution is one advantage of satellite pictures, which are widely available and inexpensive. We do the study using a collection of satellite pictures and breezometer data records from various cities. On the given satellite imagery, a CNN model with a six-layer convolutional network was trained to predict the BAQI.

1.INTRODUCTION

Air contamination is a combination of regular and counterfeit substances that causes different hurtful impacts on human and the environment. Toxic or harmful pollutants like SO_2, NO_2, CO, PM, and toxic organics are massively released into the atmosphere by the majority of industrial activities. The contamination may likewise Prompt more difficult issues influencing individuals and the entire world, for example, a dangerous atmospheric devation and environmental change. The principal justification for air quality harm is because of the smoke exhaust from businesses, contamination created by power plants, and the smoke exhaust from different vehicles. For the beyond couple of years, numerous techniques and method have been developed and followed to identify air contamination. By picture handling technique, contaminated pictures are gathered from the climate and contrasted and the recordings which are without contamination. The diffusion process has been completed using these images, and the ratio factor is used to determine the pollution level. On the other hand, it works well for images with more noise [1]. Once more, on the off chance that we consider the AI strategy, it identifies the PM (Particulate matter) 2.5 levels in light of air an incentive for a specific day. Auto-regression evaluates the future PM2.5 value based on past PM2.5 values, while logistic regression is used to determine whether a data sample is polluted or not [2]. Finally, about the profound learning approach, which is a sub-group of AI, it utilizes huge informational collection, take care of the issue without separating, utilizing more layers, handling successive layers at the same time [3]. Numerous methods and procedures exist to identify air pollution, which poses a significant threat to nature as well as living things. In the greater part of the cases, the primary focus is generally on a solitary technique and its examination. However, in this paper, the primary spotlight will be on these three strategies advantages and disadvantages, cost and precision of these methods. Comparing and comparing these methods in a single study may make it easier to distinguish between them and choose the best method for a variety of crucial circumstances.

2.LITERATURE SURVEY

2.1 AIR QUALITY INDEX USING MACHINE LEARNING[2]

Publication:An international journal of advanced computer technology,2020

Methodology: The system suggested by this study uses machine learning techniques to calculate the air quality index. It takes into account a readily available Kaggle dataset and provides it as an input to the algorithms to determine the value of the Air Quality Index. compared the effectiveness of the algorithms under consideration. Decision Tree, Random Forest, and Support Vector Machine are three examples of machine learning techniques used in this.

Advantages:

Random forest and support vector machine (SVM), produce promising results for air quality index (AQI) level predictions.

These algorithms can also handle large datasets.

Disadvantages:

Random forest contain large number of trees which can make the algorithm too slow and ineffective for real-time predictions.

2.2 Federated Learning for Air Quality Index Prediction using UAV Swarm Networks[3]

Publication: IEEE Xplore,2021

Methodology: The paper proposes a distributed and decentralized Federated Learning approach within a UAV swarm. Each UAV used its locally gathered data to train a model before transmitting the local model to the central base station. The central base station creates a master model by combining all the UAV's local model weights of the participating UAVs in the FL process and transmits it to all UAV s in the subsequent cycles.

Advantages:

measure the quality of the atmosphere; this is a simple technique to predict the area's air quality.

With the help of Unmanned Aerial Vehicle's onboard sensors, we can collect air quality data easily.

It is more efficient than machine learning algorithms.

3.PROPOSED SYSTEM

Air pollution is a problem that affects all of us. People's awareness of pollution levels increases when they are aware of how polluted their surroundings are, allowing them to take appropriate pollution-prevention measures.

3.1 IMPLEMENTATION

3.1.1 ALGORITHM

```
import tensorflow as tf
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Conv2D, MaxPooling2D, Flatten, Dense
from tensorflow.keras.preprocessing.image import ImageDataGenerator
import numpy as np
import matplotlib.pyplot as plt
# Define the paths to your training and validation datasets
train_dir = 'path _to _train _data'
val_dir = 'path_to_val_data'
# Create ImageDataGenerators for data augmentation and normalization
train_datagen = ImageDataGenerator(
    rescale=1./255,
    shear_range=0.2,
    zoom_range=0.2,
     horizontal_flip=True)val_datagen = ImageDataGenerator(rescale=1./255)
# Load the data and apply transformations
train_generator = train_datagen.flow_from_directory(
    train_dir,
    target_size=(64, 64),
    batch_size=32,
     class_mode='categorical'  # Use 'binary' for binary classification)
     val_generator = val_datagen.flow_from_directory(
    val_dir,
    target_size=(64, 64),
    batch_size=32,
    class_mode='categorical')
model = Sequential([
    Conv2D(32, (3, 3), activation='relu', input_shape=(64, 64, 3)),
    MaxPooling2D((2, 2)),
    Conv2D(64, (3, 3), activation='relu'),
    MaxPooling2D((2, 2)),
    Flatten(),
    Dense(128, activation='relu'),
    Dense(3, activation='softmax') # 3 classes: low, moderate, high])
model.compile(optimizer='adam', loss='categorical_crossentropy', metrics=['accuracy'])
history = model.fit( train_generator,
 epochs=10,validation_data=val_generator)
loss,accuracy model.evaluate(val_generator)
print(f'Validation accuracy: {accuracy}')
# Plot the training and validation accuracy and loss
plt.figure(figsize=(12, 4))
plt.subplot(1, 2, 1)
plt.plot(history.history['accuracy'], label='Train Accuracy')
plt.plot(history.history['val_accuracy'], label='Validation Accuracy')
plt.legend()
plt.title1('Accuracy')
plt.stbplot(1, 2, 2)
plt.plot(history.history['loss'], label='Train Loss')
plt.plot(history.history['val_loss'], label='Validation Loss')
plt.legend()
plt.title('Loss')
plt.show()
```

Dataset:

There are mainly 4 modules in this project.

1) Collection of dataset

By photographing a few locations in Autonagar region, the dataset was constructed. 50 photos are included in the dataset.

2) Image pre-processing

Collected dataset is pre-processed using binary segementation to extract the features from images and to prepare them for feeding to the neural network which will be done in further steps.

3) Developing Convolutional Neural Network Model

In this, we build resnet-50 model to predict the percentage of Air Quality Index value.

4) Training and testing the model

The model is trained with images in the collected dataset and is tested by selecting a random image from the dataset.

Fig:1 proposed model

3.2 CNN

Structure of a CNN

1. **Input Layer**: This layer holds the raw pixel values of the input image.
2. **Convolutional Layers**: These layers apply a set of filters (kernels) to the input image. The filters slide over the image to create feature maps that capture the presence of certain features (edges, textures, patterns) at different spatial locations.
3. **Activation Function**: Often, the ReLU (Rectified Linear Unit) function is applied after each convolution to introduce non-linearity into the model.
4. **Pooling Layers**: These layers perform down-sampling operations to reduce the spatial dimensions of the feature maps, retaining the most important information. Common pooling methods include max pooling and average pooling.
5. **Fully Connected Layers**: After several convolutional and pooling layers, the high-level reasoning in the network is done via fully connected layers. Every neuron in a fully connected layer is connected to every neuron in the previous layer.
6. **Output Layer**: This layer produces the final predictions. For classification tasks, it often uses a softmax function to output probabilities for each class.

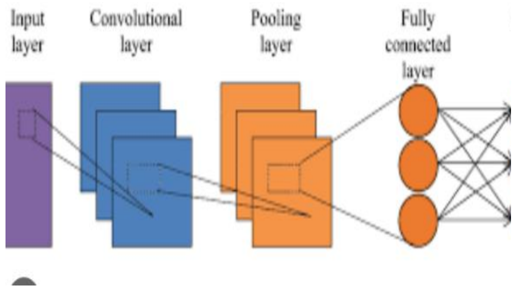

Fig:2 Structure of layers

Key Concepts

- **Convolution Operation**: Involves sliding a filter over the input image and performing element-wise multiplication and summation to produce a feature map.
- **Padding**: Adding zeros around the border of the input image to control the spatial dimensions of the output feature maps.
- **Stride**: The number of pixels by which the filter moves across the input image.
- **ReLU Activation**: Introduces non-linearity by converting all negative values to zero.
- **Pooling**: Reduces the dimensionality of the feature maps while retaining important features.

Training a CNN

1. **Forward Propagation**: Input data is passed through the network, and predictions are made.
2. **Loss Calculation**: The difference between the predicted output and the true output is measured using a loss function (e.g., cross-entropy loss for classification).
3. **Backpropagation**: The network weights are updated to minimize the loss by propagating the error backward through the network and adjusting the weights using gradient descent.
4. **Iteration**: Steps 1-3 are repeated for many epochs (iterations over the entire dataset) until the model converges to a minimum loss.

4.RESUTS AND DISCUSSION

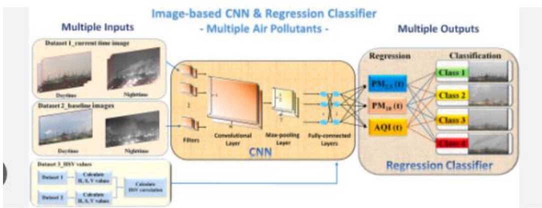

Fig:3Image-basedCNN&Regression clssifier

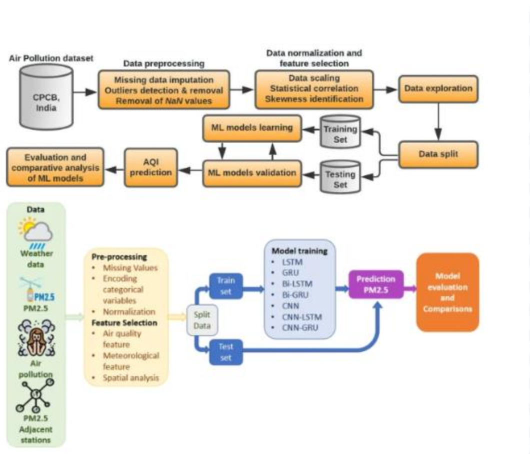

Fig4: Air quality and cuase of degradetion

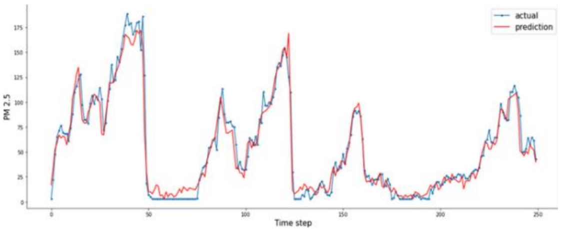

Fig:5 PM2.5 concentration forecasting

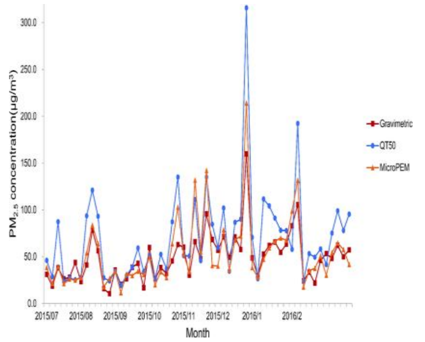

Fig:6 Time trend graph for PM 2.5 level

5.CONCLUSION

Air pollution is one of nature's primary concerns, and it is becoming more severe as cities and industries grow.There are several ways available to detect this. However, as previously said, the focus of this book was on only three easy methods.As a result, the key findings of this research are that image processing can be a good alternative for detection, but additional specific algorithms and sensors are required to detect the impure chemical.However, if more precise pollution detection is necessary, machine learning or deep learning will be a preferable approach.And, in terms of cost, deep learning will be the most expensive due to the vastness of the dataset. In the future, more improved mechanisms in terms of both cost and accuracy may be discovered.

REFERENCES

J. Lee,"Acoustical perceptions of building occupants on indoor environment quality in naturally-ventilated building facades, "Journal of Acoustics, vol.4,no.3,2019.

K. Nahar, A. Jaradat, M. Atoum, and F. Ibrahim, "Sentiment analysis and classification of arab jordanian facebook comments for jordanian telecom companies using lexicon-based approach and machine learning," Jordanian J. Comput. Inf. Technol., vol. 6, no. 03, pp. 247–263, 2020.

Chhikara P., Tekchandani R., Kumar N., Chamola V., and Guizani M., "Dcnn-ga: A deep neural net architecture for navigation of uav in indoor environment, " IEEE Internet of Things Journal, vol. 8, no. 6, pp. 4448–4460, 2021.

X. Lin, H. Wang, J. Guo and G. Mei, "A Deep Learning Approach Using Graph Neural Networks for Anomaly Detection in Air Quality Data Considering Spatiotemporal Correlations," in IEEE Access, vol. 10, pp. 94074-94088, 2022, DOI: 10.1109/ACCESS.2022.3204284.

Fatima Ezzahra Mana, Blaise Kévin Guépié, Raphaèle Deprost, Eric Herber, Igor Nikiforov, "The air pollution monitoring by sequential detection of transient changes", IFAC-papers online, Volume 55, Issue 5,2022, Pages 60-65, ISSN-2405-8963,https://doi.org/10.1016/j.ifacol.2022.07.640.

M. Molinara, M. Ferdinandi, G. Cerro, L. Ferrigno and E. Massera, "An End to End Indoor Air Monitoring System Based on Machine Learning and SENSIPLUS Platform," in IEEE Access, vol. 8, pp. 72204-72215, 2020, doi: 10.1109/ACCESS.2020.2987756

Z. J. Andersen, L. C. Kristiansen, K. K. Andersen, T. S. Olsen, M. Hvid-berg, S. S. Jensen, et al., "Stroke and Long-Term Exposure to Outdoor Air Pollution From Nitrogen Dioxide", Stroke, vol. 43, no. 2, pp. 320-325, 2019.Z. Li et al., "Practical deployment of an in-field soil property wireless sensor network",Comput.Standards Interfaces, vol. 36, no. 2, pp. 278-287, 2019Chuanqi, X. et al. Air pollutant spatiotemporal evolution characteristics and effects on human health in North China. Chemosphere 294, 0045–6535 (2022).

AUTHOR PROFILES

[1].Mr. HIMAMBASHA SHAIK, currently working as an Assistant Professor in the Department of MCA, QIS College of Engineering and Technology, Ongole, Andhra Pradesh. Her area of interest Cloud Computing ,Devops and ,DBMS

[2], Ms, Supraja Bakkamunthala currently pursuing Master of Computer Applications at QIS College of Engineering and Technology (Autonomous), Ongole, Andhra Pradesh. she completed BSC from Sri Harshini Degree College, ongole, Prakasam, Andhra Pradesh. Her areas of interests are Machine Learning.