Fall 2023

Estimating Air Pollution Levels Using Machine Learning

Srujay Rao Devaraneni

Estimating Air Pollution Levels Using Machine Learning

A Project Report

Presented To

Prof. Robert Chun

Department of Computer Science

San Jose State University

In Partial Fulfilment

Of the Requirements for the class

Fall-2023: CS 298

By

Srujay Rao Devaraneni

October 2023

The Designated Project Committee Approves the Master's Project Titled

Estimating Air Pollution Levels Using Machine Learning

By

Srujay Rao Devaraneni

APPROVED FOR THE DEPARTMENT OF COMPUTER SCIENCE

SAN JOSE STATE UNIVERSITY

Dr. Robert Chun	Department of Computer Science
Dr. William Andreopoulos	Department of Computer Science
Nikhila Saini	Oracle Cloud Infrastructure

ACKNOWLEDGEMENT

I would like to convey my profound gratitude to Dr. Robert Chun for his priceless guidance, steadfast motivation, and unwavering support throughout the duration of this project. His profound knowledge and insightful advice were instrumental in navigating challenging situations where viable solutions seemed elusive. Dr. Chun's patience and consistent provision of both moral and technical support were pivotal in sustaining my progress on this project.

I wish to express my sincere thanks to Dr. William Andreopoulos, and Miss. Nikhila Saini, who were part of my defense committee, as well as the entire CS faculty. Their collective support has been a cornerstone in my academic journey over the past two years, manifesting in various forms.

Finally, I would like to extend my appreciation to my friends, and family, whose consistent support and motivation have been the driving force behind this undertaking.

ABSTRACT

Air pollution has emerged as a substantial concern, especially in developing countries worldwide. An important aspect of this issue is the presence of PM2.5. Air pollutants with a diameter of 2.5 or less micrometers are known as PM2.5. Due to their size, these particles are a serious health risk and can quickly infiltrate the lungs, leading to a variety of health problems. Due to growing concerns about air pollution, technology like automatic air quality measurement can offer beneficial assistance for both personal and business decisions. This research suggests an ensemble machine learning model that can efficiently replace the standard air quality estimation techniques, which need several instruments and setup and have large financial expenditures for equipment acquisition and maintenance.

Index Terms - ***machine learning, neural network, PM 2.5, prediction model, regression model***

DEFINITION OF TERMS

PM 2.5 – Particulate matter with diameter less than 2.5 micrometers

TABLE OF CONTENTS

I. INTRODUCTION

In recent times, air pollution has emerged as a substantial and pressing global concern. According to the 2020 report from the World Health Organization (WHO) [1], outdoor air pollution has been linked to a staggering global toll of approximately 4.2 million annual fatalities. These statistics underscore the severe and far-reaching impact of air pollution on public health across the world. The primary cause for alarm revolves around suspended particulate matter, with particular emphasis on PM2.5, which denotes exceedingly fine particles measuring 2.5 micrometers or less in diameter. These tiny particles, due to their size, remain suspended in the air for extended periods, unlike larger particles. The small size of these particles allows them to bypass the nose and throat easily and they are found to penetrate deep into the lungs. These pollutants can enter the cardiovascular system and pose a serious threat to human and animal health.

PM2.5 levels are mostly measured by monitoring stations. The instruments used to monitor PM2.5 levels are bulky, expensive and require continual maintenance. This equipment can only measure the PM2.5 levels in a small, local area. These measurements cannot be generalized for other areas. Most developing countries do not have sufficient monitoring stations. This results in people in most parts of the country being unaware of the ambient outdoor pollution levels which lead to exposure to pollutants thus affecting the health of the individual. Figure 1 below depicts various PM 2.5 detection equipment.

Fig. 1 PM 2.5 detection equipment

An ensemble model is developed in this project which can predict PM2.5 concentration in an area using outdoor pictures. The images used were first transformed to accentuate information regarding entropy, contrast, and structural information features of the image. In the next step, three different neural networks based on the Visual Geometry Group 16 (VGG16) [2] neural network architecture were trained using one of these extracted features to develop regression models. The model benefits from a regression meta learner, which is supplied with the outputs generated by the base learners. This incorporation of a meta learner plays a pivotal role in augmenting the overall model's performance, essentially refining its predictive capabilities.

Following this integration, the model's effectiveness is then meticulously assessed. To gauge its performance, it is subjected to a rigorous comparison with state-of-the-art models detailed in the literature. Notably, all these models have one crucial commonality: they were formulated and rigorously evaluated using the very same dataset. This ensures a robust and meaningful comparison, underscoring the model's competitiveness in relation to the most advanced solutions available in the field.

II. Literature Review

Airborne particulate has become a matter of serious concern over the past few decades, and multiple research works have been published which have proposed methods to estimate PM2.5 particle levels without the use of traditional equipment, using computer vision and image processing techniques. In their research, Bo et al. [5], proposed and recommended the adoption of a convolutional neural network (CNN) as a powerful tool for the prediction of PM2.5 values. Their study put forth this neural network architecture as a promising and effective means to estimate PM2.5 concentrations. By endorsing the use of CNNs for this particular purpose, the authors aimed to leverage the network's capacity to extract intricate spatial and temporal patterns from the data, contributing to improved accuracy and reliability in forecasting PM2.5 levels.

This recommendation is a significant step in advancing the field of air quality prediction and underscores the potential benefits of CNNs in environmental research and monitoring. The PM2.5 value obtained is calibrated with other key factors such as humidity and wind speed to train and develop an SVR model which gives the final predicted output. Results of the proposed methods were assessed using two datasets collected from Shanghai and Beijing in China and experimental results showcased the effectiveness of the proposed CNN model, which has paved a path for further research and has served as a foundation for many neural network-based models. Figure 2 below depicts the flowchart of this method.

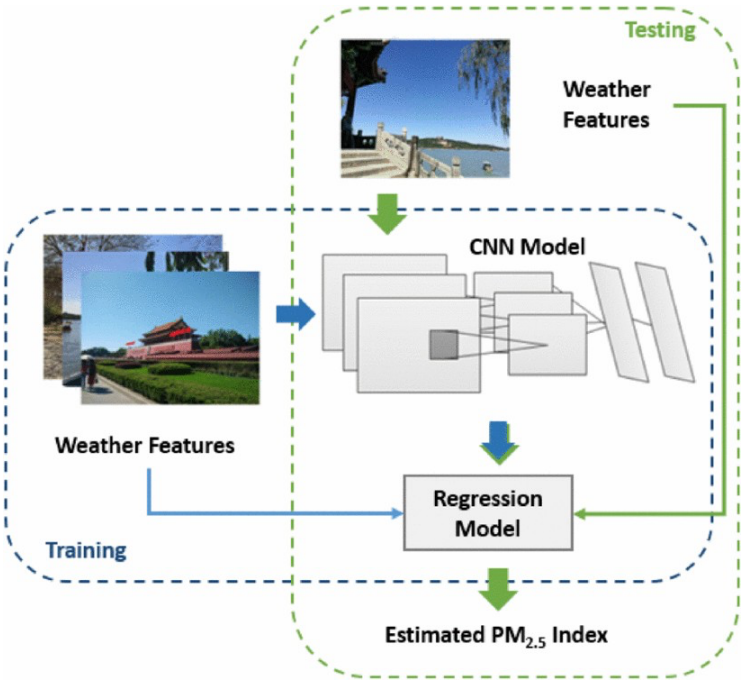

Fig. 2

Zhang et al. [6], developed an efficient convolutional neural network (CNN) with additional ordinal classifiers to evaluate the air quality based on images, influenced by the work of Bo et al. [5]. Three components make up the suggested technique, starting with an ensemble CNN for estimating air quality, set of classifiers namely the complementary log-log ordinal classifier, cauchit ordinal classifier, and negative log-log ordinal classifier, and a modified activation function for the rectifier linear unit (ReLU). The usefulness of this technique on the real-world dataset is demonstrated by experimental findings. This method's rectilinear activation function is challenging to compute, and it doesn't always produce an accurate approximation of the PM 2.5 levels.

Six features were taken from the photos by Liu et al. [7], to devise a method for quantifying PM contamination. In the quest for establishing a relationship between PM levels and image attributes like sky transparency, smoothness, sky color, overall and local image contrast, as well as image entropy, supplementary variables including the

time of day, geographical location, and meteorological conditions were also factored into the analysis. A regression model is built with all these factors to understand and extrapolate the findings to new datasets and thus predict the PM2.5 levels. The six features extracted in this approach were critical and were a basis for many other smartphone image-based models. Figure 3 below depicts the flowchart of this method.

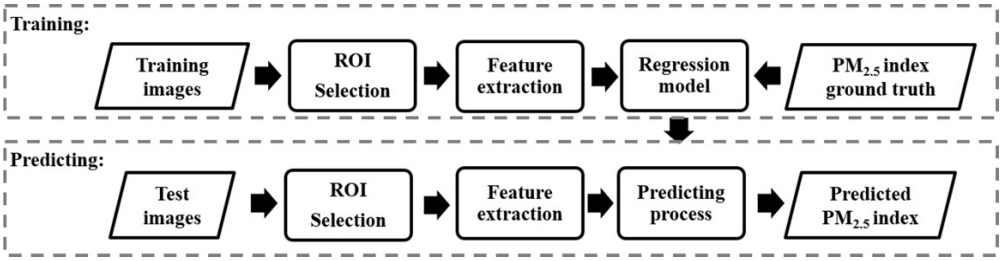

Fig. 3

To measure the concentration of PM2.5, Yue et al. [8], suggested a brand-new photo-based air quality estimator. The proposed method depends on the shape of the distribution of pixel values in a saturation map, gradient similarity, features from photographs taken at various PM2.5 concentrations. The success of proposed methods entirely depends on extracting these two category features. By combining these two features, an additional non-linear function is adapted to map the predicted value of PM2.5 to the actual PM2.5 value to train the model. The usefulness and efficiency of the suggested strategy were sufficiently demonstrated by trials on actual data. The introduction of an extra non-linear function for translating the predicted value into a real-time value lacks consistency, and it amplifies the mathematical intricacy of this model, necessitating the derivation of a function adaptable to all datasets. Figure 4 below depicts the framework of the proposed solution in this method.

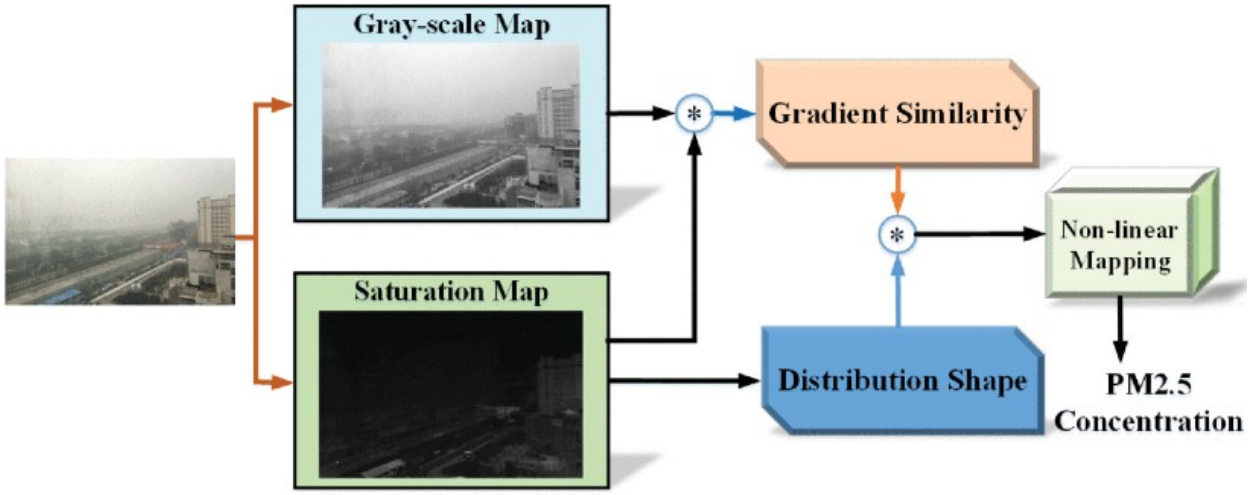

Fig. 4

'Picture-based predictor of PM2.5 concentration' (PPPC) model was suggested by Gu et al. [9]. For this model to work, a sizable collection of images must have been taken in favorable lighting and at low PM 2.5 concentrations. First, the authors develop Naturalness Statistics (NS) based models using pictures captured under low PM2.5 concentrations. The NS models are built on spatial and transform domains. Then, deviation from the above model for a new image is measured. A non-linear function is then introduced to map the deviation of the naturalness statistic of the image to the PM2.5 concentration. Empirical findings confirmed that the proposed PPPC model surpasses the performance of cutting-edge predictors. NS model used in this approach reduces the mathematical complexity in this model compared to the Yue et al. [8] approach. Though this model reduces the mathematical complexity involved, capturing pictures under good weather conditions and low PM 2.5 levels remains a major drawback for this approach. Figure 5 below depicts the flowchart of how this model predicts the PM 2.5 concentration.

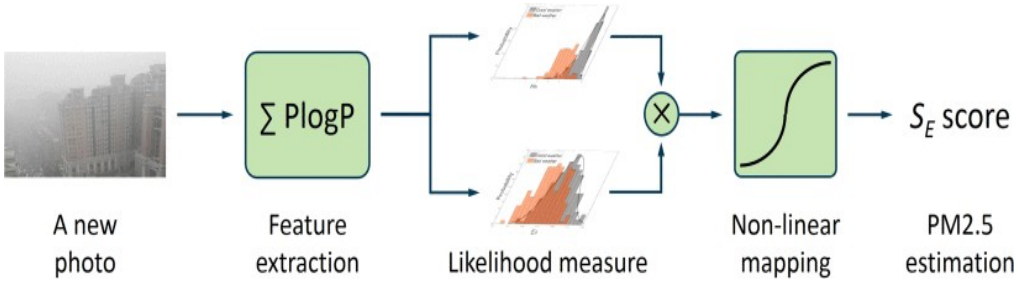

Fig. 5

Li et al.'s [10], introduces a photo based PM2.5 estimator capable of ascertaining an image's depth through the estimated PM2.5 value. This approach employs a hybrid convolutional neural network designed to comprehend and categorize haze-related features. PM2.5 data are added to high-level features retrieved from the convolutional layer in order to train a support vector regression (SVR) model similar to Bo et al.'s [5] method. PM2.5 value obtained from the atmospheric scattering model has been used to deduce the depth map of an image. As this approach involves contributions from more than a model in estimating the final output, error from one model would propagate through each phase and deviate the estimated output to a considerable degree from the actual output. Hence, there is very little space for deviation and each model should make an accurate prediction. Figure 6 below depicts the architecture of the model proposed by this method.

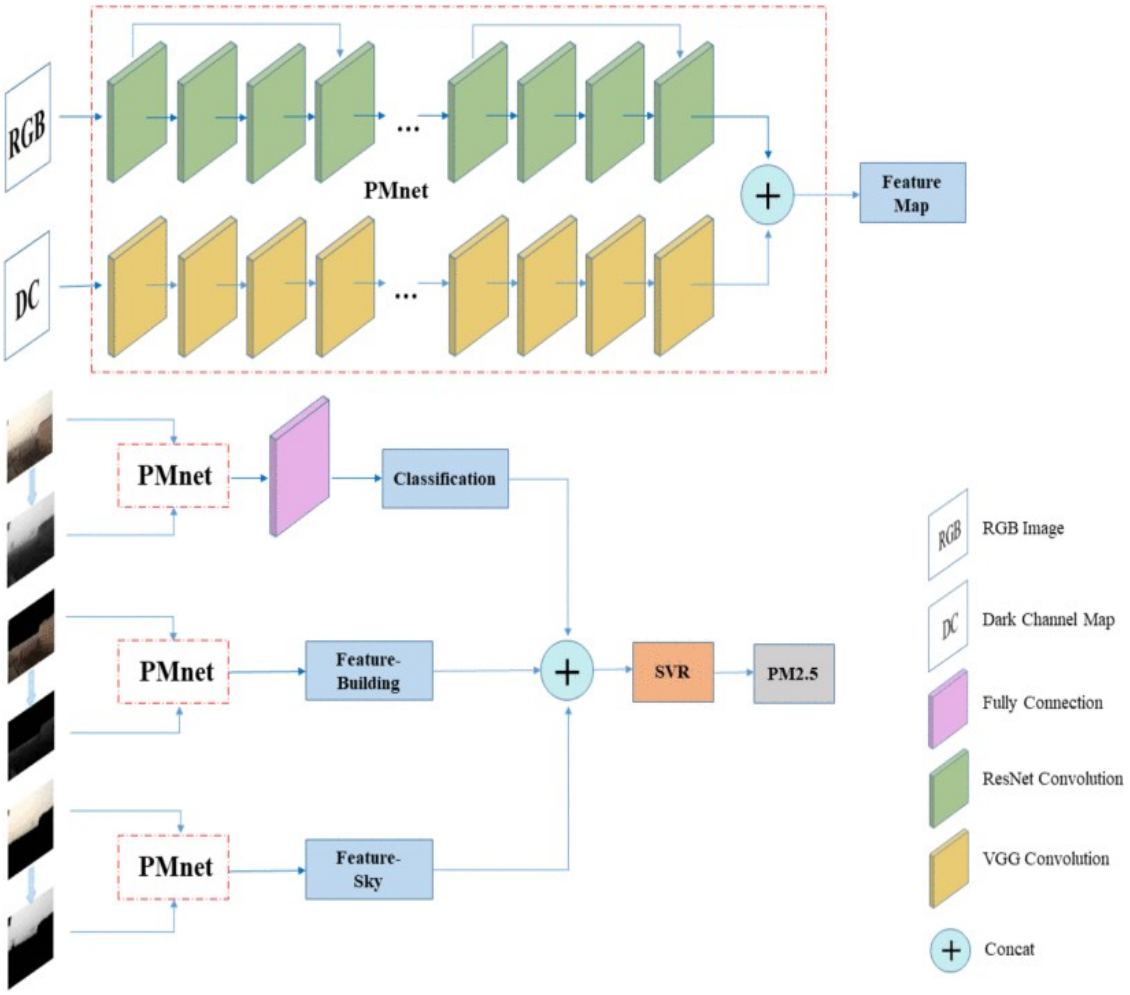

Fig. 6

Ma et al. [11], explores a new approach, proposing a classifier using a hybrid deep neural network to predict air pollution levels from images. This method also employs a dark channel map. Dark Channel Prior is one of the significant dehazing methods based on observations of key features on haze-free images. In order to enhance the features with implicit representation, a secondary neural network is fed the output of the dark channel map. They used a self-made dataset consisting of 1575 images of different scenes and PM2.5 values. As the results were tested on a self-made dataset, further evaluation on compliance of this model with real time dataset is

required. Figure 7 below depicts the architecture of the model proposed by this method.

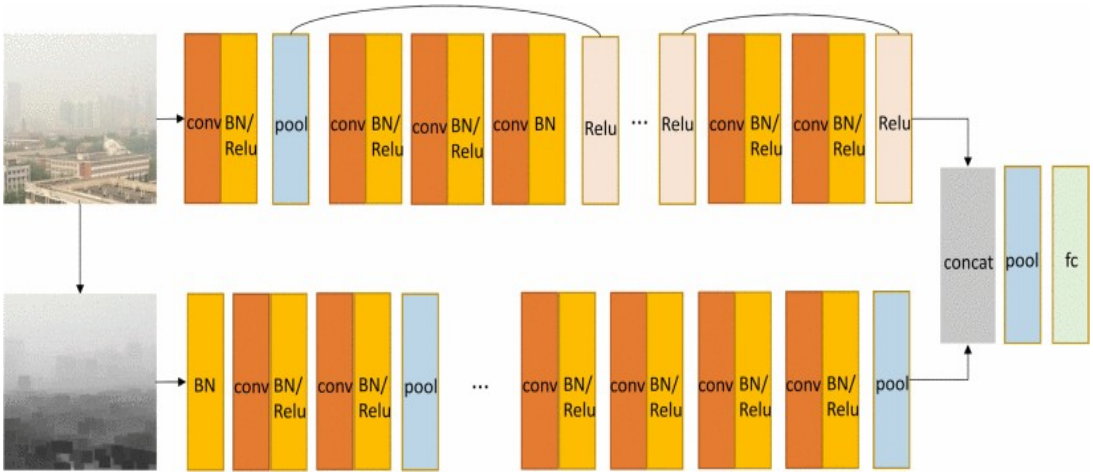

Fig 7.

By harnessing a deep Convolutional Neural Network (CNN), Avijoy Chakma et al. [19], present a method for classifying natural photographs based on their PM2.5 levels. This method just classifies images into different categories based on their PM2.5 concentrations but does not predict the exact value of PM 2.5 levels. Figure 8 below depicts the architecture of the CNN model proposed by this method.

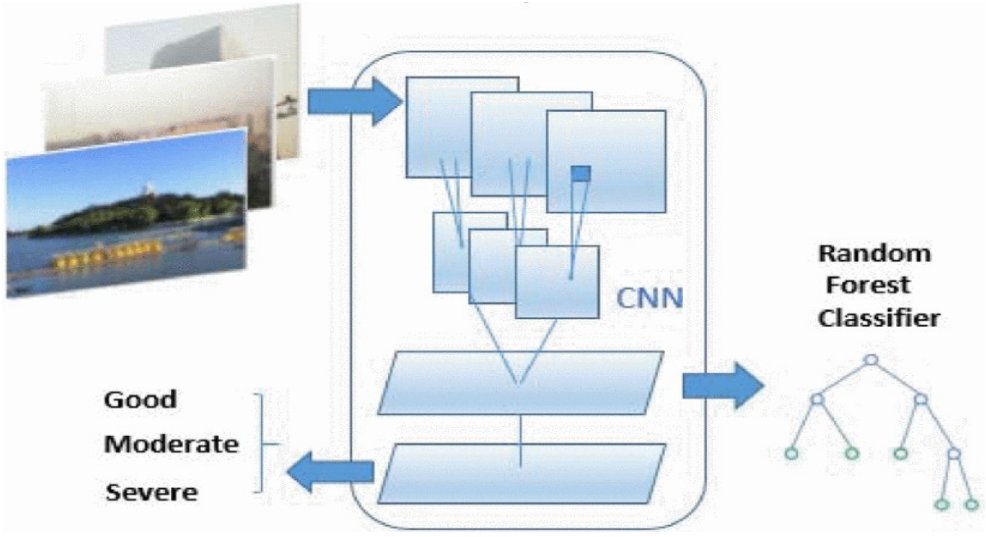

Fig. 8

Through the application of a hybrid deep learning framework, the novel model introduced by Gu et al. [13], for predicting PM2.5 levels effectively captures the spatial-temporal correlation features and dependencies within time series data related to air quality. The core element of this approach includes the utilization of both bidirectional long short-term memory networks (Bi-LSTM) and one-dimensional convolutional neural networks (1D-CNNs). In the first model, spatial-temporal relationships are learned; in the second, relation between characteristics of local trends and features of spatial correlation are learnt. Based on the above base modules, to learn time series data of multivariate air quality, a hybrid deep learning framework has been developed. But the PM2.5 prediction accuracy levels have not been the best and requires adjustments. Predicting the adjustor value is difficult as the results were spread out on a broad spectral range. Figure 9 below depicts the architecture of the proposed framework.

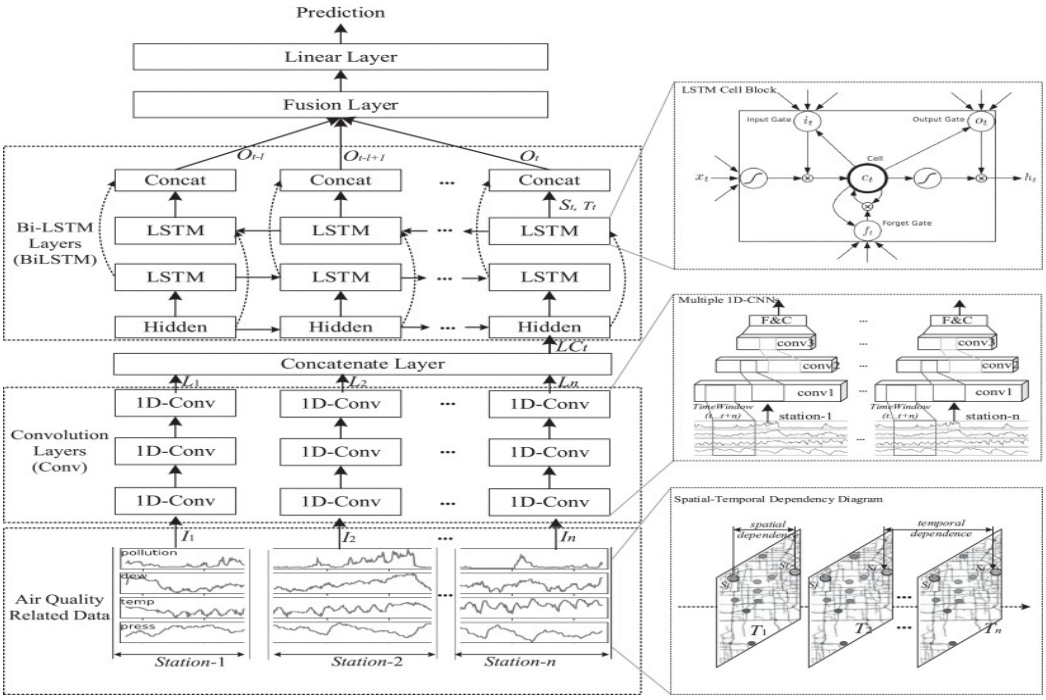

Fig. 9

A spatiotemporal C-LSTME model for forecasting the air quality is put forth by Wen et al. [21]. The model under consideration incorporated historical air pollutant concentrations from the present data collection station, in addition to those obtained from the nearest k neighboring stations. This approach was taken to account for both the temporal and spatial characteristics of the data. Figure 10 below depicts the architecture of the proposed framework.

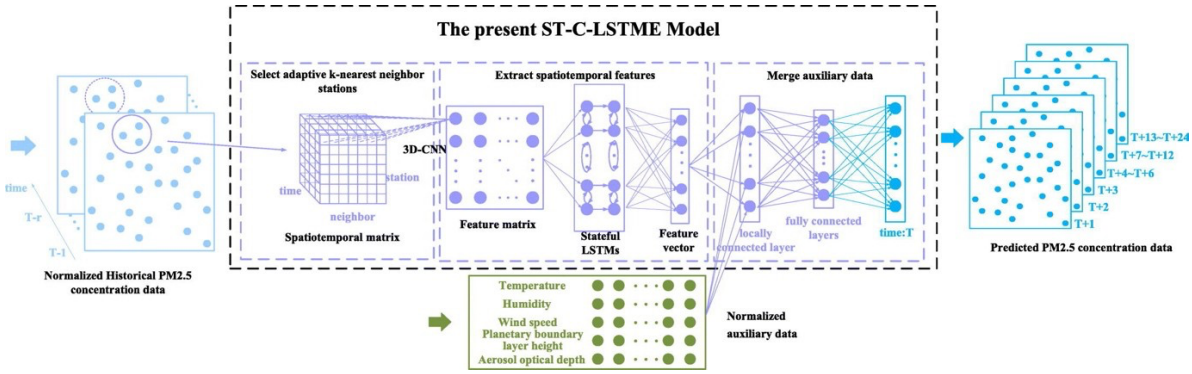

Fig. 10

Utilizing Eeftens et al.'s [22], Land Use Regression (LUR) model, researchers developed spatial variation models for air pollution concentrations at a small scale. Across 20 distinct regions in Europe, 20 sites were carefully chosen, each site comprising 20 specific locations, for the evaluation of PM(2.5) absorbance, PM(coarse), PM(2.5), and PM(10). Within each study area, the geographic variance in annual average concentrations was modeled by examining predictor variables obtained from GIS data, including metrics like traffic intensity, population density, and land usage. Regarding PM(2.5), the median R(2) value of the model accounted for 71% of the variance, whereas PM(coarse) displayed a lower average model R(2), and PM(2.5) absorbance exhibited a higher median R(2). The models employed various traffic-related indicators as predictor variables, typically using between two to five of them.

In order to investigate the connection between aerosol optical depth (AOD), meteorological variables, PM(2.5), and land usage data, Hu et al. [23], formulated a spatially weighted regression model. To evaluate the model's effectiveness, they incorporated data from the North American Land Data Assimilation System and the North American Regional Reanalysis datasets independently. The study primarily focused on the Atlanta Metro area, gathering data from multiple sources for the year 2003.

The study unveiled that the average local R(2) for models trained with the North American Regional Reanalysis was 0.60, while it slightly improved to 0.61 for models trained with the North American Land Data Assimilation System. Regarding forecasting accuracy, the North American Regional Reanalysis achieved 82.7%, while the North American Land Data Assimilation System exhibited a higher accuracy of 83.0%, as confirmed by the root mean squared prediction error.

Park et al. [24], introduced a convolutional neural network (CNN) model designed to forecast the 24-hour average ground-level PM2.5 concentration in the United States for the year 2011, utilizing data from aerosol optical depth (AOD) and land-use information. They gradually combined predictors of neighboring sites to take advantage of the spatial correlation among the predictors, unlike some recent supervised learning-based systems that solely used the predictors from the PM 2.5 prediction area. Using geographically and temporally separated cross-validations (CV), carefully assessed the performance of the approach and demonstrate that this CNN achieves considerable accuracy when compared to baselines developed recently. Furthermore, based on the most recent neural network interpretation approach, they created a novel predictor importance metric for CNN.

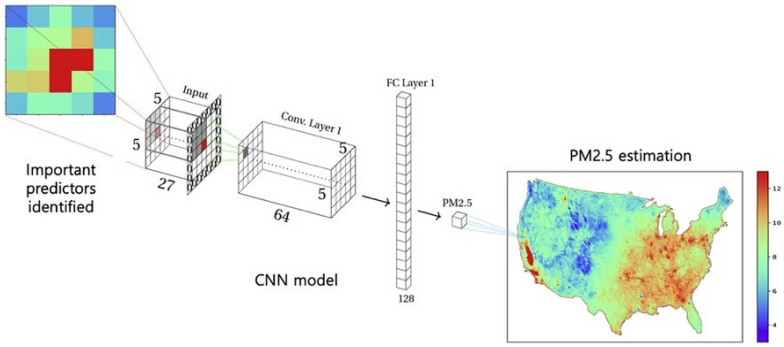

Fig. 11

According to Feng et al. [25], the spatiotemporal fusion method was used to create a spatiotemporal PM2.5 characteristic. An ensemble learning approach was employed, utilizing data that encompassed topological, periodic, spatiotemporal, anthropogenic, vegetation, and meteorological attributes. In this methodology, gradient boosting (XGBoost), back-propagation neural network, and K Nearest Neighbor algorithms were combined at the first level, while linear regression was employed for integration at the second level. The ST-stacking model was the optimal stacking method that took into account the spatiotemporal autocorrelation of PM2.5. Model was trained and tested on China dataset in 2017. The model was used to determine ground-level PM2.5 concentrations for 24 hours in mainland China, on 11th May, 2017.

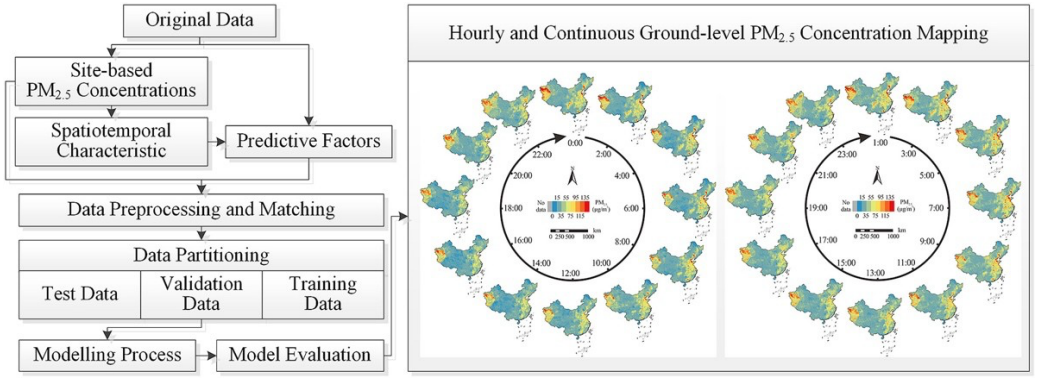

Fig. 12

III. RESEARCH METHODOLOGY

In this study, an ensemble[3] model is created that can estimate the amounts of PM 2.5 in a location from a photograph without any extra information. A machine learning method called ensemble learning combines models to get the best possible predictive model. In ensemble learning, a model that performs much better than the individual models is created by training a meta learner algorithm utilizing the prediction outputs of numerous base learners.

Features like entropy, contrast and structural information from the image were used as these features tend to change with the pollution levels. Following the extraction of these features, a neural network architecture based on the VGG16 architecture was used to create estimator models for each of the features. The relevant photos' PM 2.5 levels were used as the target variable. These neural network models act as the base learners in the ensemble.

Further, a meta learner has been employed by training it on the outputs of the base learners. The performance of the model as a whole was significantly improved by the regression model built on top of base learners. Figure-13 depicts the flowchart of the methodology of this project.

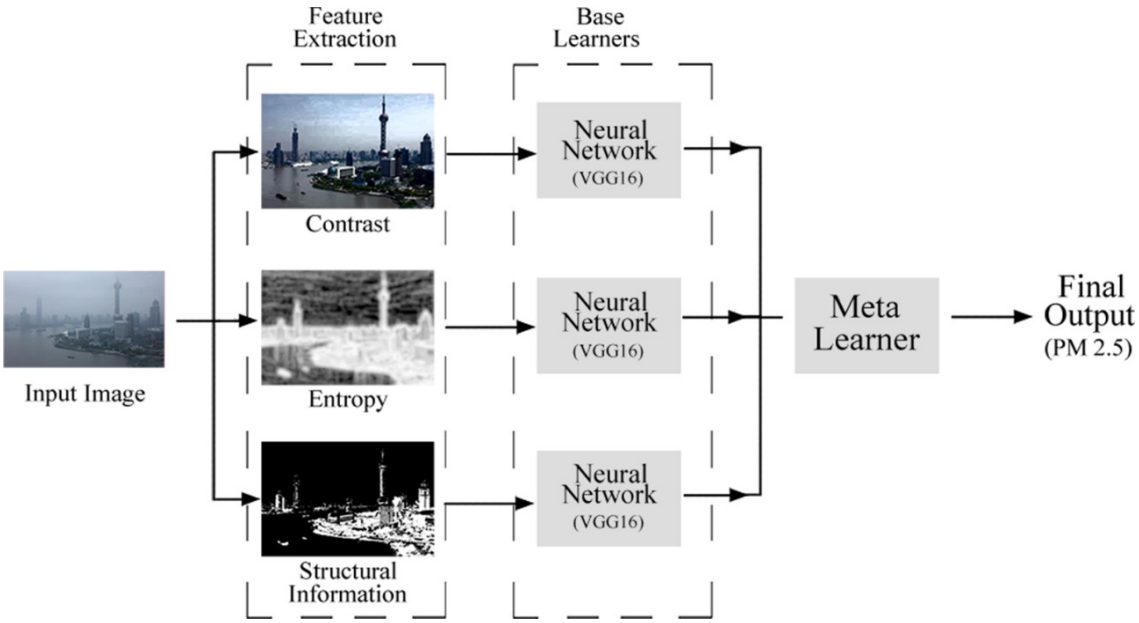

Fig. 13 Flowchart of the proposed method

IV. Data Pre-Processing

In this project, Shanghai dataset has been used to train and develop the models. The dataset contains 1900 photographs taken in Shanghai, China at the Oriental Pearl tower captures between 8:00 and 16:00. The U.S. Consulate in Shanghai, which tracks the city's air quality, provided the PM2.5 figures for Shanghai in public documents.

Figure 14 displays the dataset's distribution of PM2.5 values among the images falling inside a specific range. Figure-15 shows the distribution of images in the dataset based on the time of the day they were captured. The images labeled as ‘Morning’ were captured from 8:00 to 10:45, those labelled as ‘Afternoon’ were captured between 11:00 and 14:30 pm and the ones captured post 14:30 were labelled as captured in ‘Evening’. The images in the dataset had date and location water marks on them. Watermarks were removed to avoid any inaccuracies that might be introduced in the model.

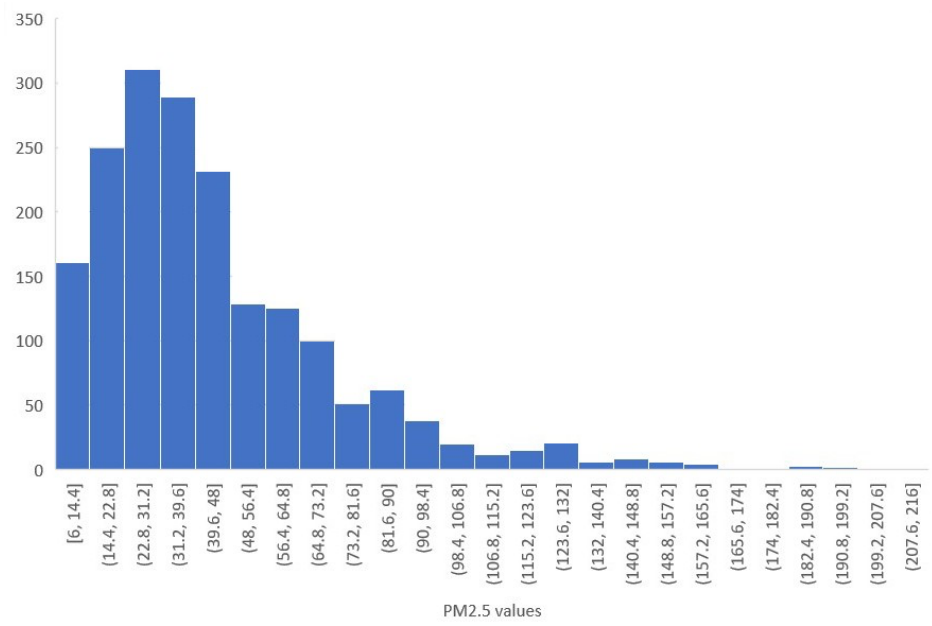

Fig. 14 PM 2.5 distribution in Shanghai dataset

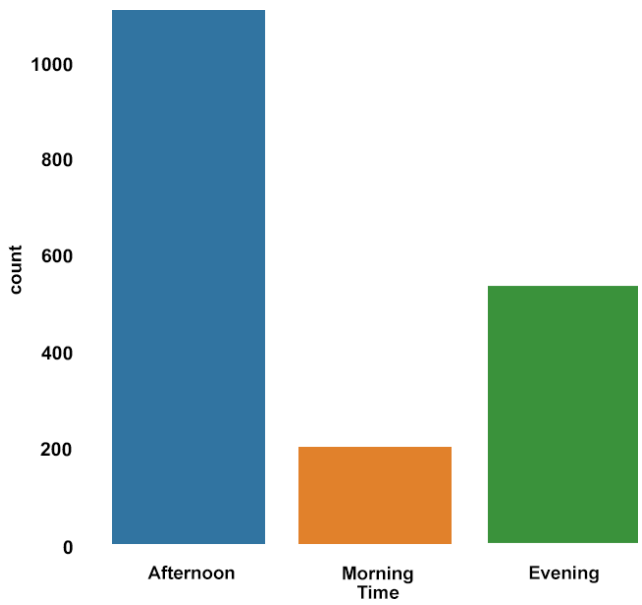

Fig. 15 Distribution of images in the Shanghai dataset based on the time of capture

4.1 ENTROPY

In an image, entropy is connected to the complexity present in a given neighborhood which is usually defined by a structuring element. Subtle variations in the local grey level distribution can be detected by the entropy filter.

The uncertainty of objects can be measured using the term "Entropy" [4]. It can be used to describe the picture's complexity traits and uncertainty distribution. The internal information qualities that are present in the picture can be quantitatively described by the complexity characteristics. Consequently, it is possible to extract a signal's intrinsic properties using the entropy characteristic.

The most commonly used measure of entropy is Shannon entropy [11]. It is defined as

$$H_1(p) = H(p_1, p_2, p_3, \cdots, p_n) = -\sum_{i=1}^{n} p_i \log_2 p_i \qquad (1)$$

Where $p = (p_1, p_2, p_3, \cdots, p_n)$ stands for the existing probability of every event in the collection and it satisfies

$$0 \leq p_i \leq 1, \quad \sum_{i=1}^{n} p_i = 1 \qquad (2)$$

Shannon entropy can take values within the range of $[-\infty, +\infty]$.

Figure-16 shows the variation of Shannon entropy of the images in the Shanghai dataset with the change in PM2.5 levels in the images. Although, the correlation is not very strong, as the PM2.5 levels rise, there is a general tendency for the image's entropy to decrease. This is also in accordance with logic as in case of images with higher values of PM2.5 pollutants, the clarity of the image decreases which leads to a loss of detailed information in these images which in turn causes image entropy to decrease. The fluctuation in entropy values can be because of a wide-ranging weather condition under

which these images were captured.

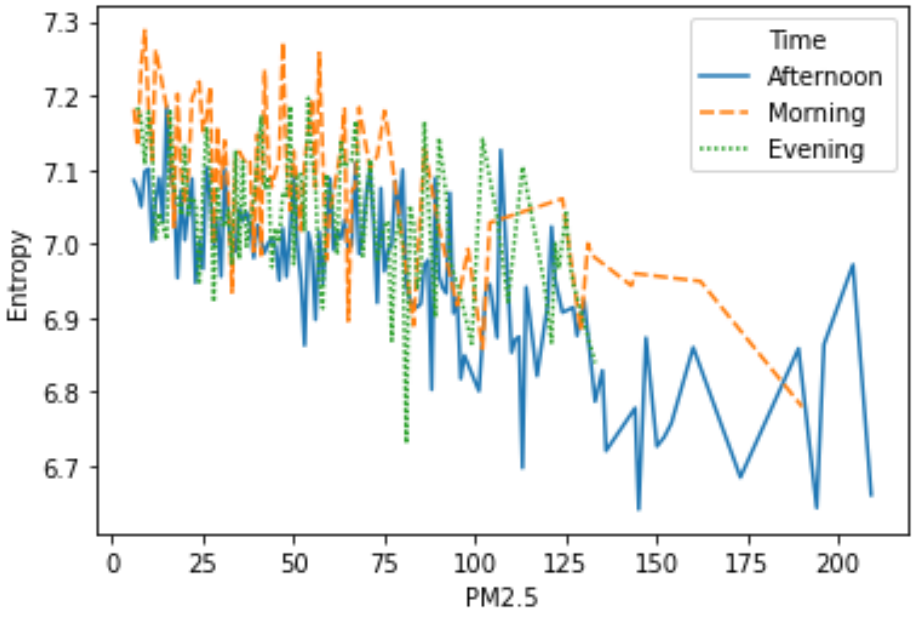

Fig. 16 Line chart showing variation of image entropy with PM2.5 values in images captured during different times of the day

4.2 CONTRAST

In an image, contrast refers to the quantity of differentiation present between the different features of the image. High contrast images usually display a greater degree of color or variation in grey scale when compared with images having a low contrast. For sake of analysis, the contrast of an image is accentuated and meddled with to extract more information and to provide it as an input to a NN(neural network) model.

Calculating an image's contrast value reveals its overall visibility, which is equivalent to the area's air pollution levels. In this research, the histogram equalization technique is used which is the most commonly used technique to alter the contrast of an image. Histogram equalization is a computer image processing method which is used to improve contrast in images. It achieves this by efficiently dispersing the most frequently occurring intensity values, which in turn broadens the overall intensity range of the image. This approach facilitates for regions of lower intensity to acquire a higher intensity. Before being fed into the neural network, the image intensity distribution in all the images in the data set is equalized.

Figure-17 shows the intensity histogram depicting the distribution of the intensities of the pixels in the image which has to be equalized for contrast improvement.

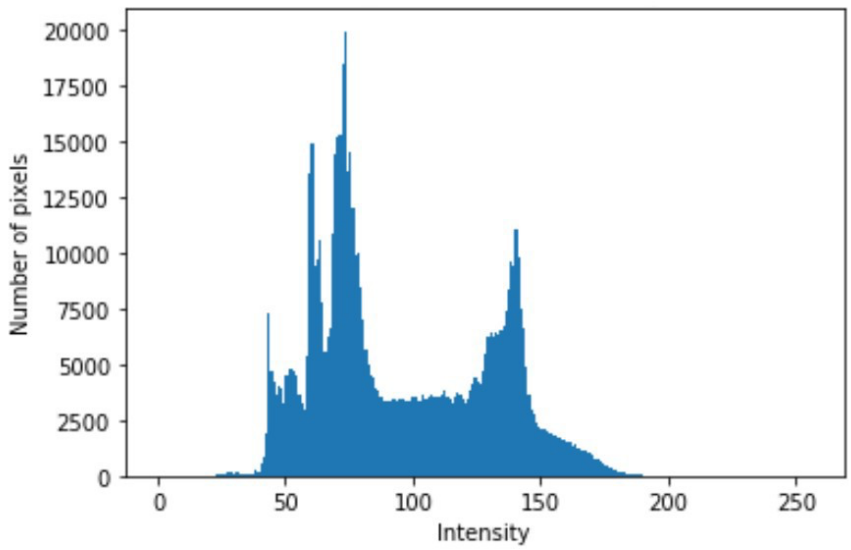

Fig. 17 Intensity distribution histogram of a sample image from the dataset

Figure-18 has RGB channels, and the intensities of all the three channels together depict the overall intensity of the image. Different channels known as YUV are used to separate out intensity and color information from the images.

In case of RGB images, it is a challenge to alter the intensity values of an image without changing its color information. To achieve this, the intensity and color information in the image has to be separated into different channels, for this, the RGB images in the dataset are converted into YUV color space, which is a representation of the same image in three different channels namely Y, U and V. Now, by equalizing the Y channel (Brightness and Intensity values), the intensity range of the image is spread to improve the overall contrast information in the image, it is then converted back to RGB.

The distribution of the intensities of the image pixels after histogram equalization is observed in figure 18.

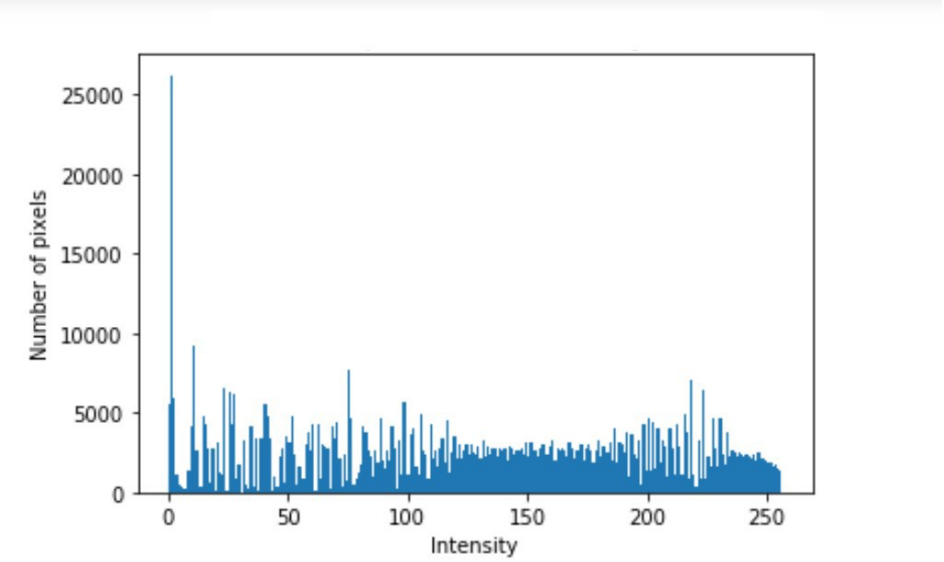

Fig. 18 Intensity distribution histogram of a sample image after Y channel equalization

4.3 STRUCTURAL INFORMATION EXTRACTION

Previous studies [12, 13 and 14] on assessment of image quality have shown that calculating structural similarity between a high-quality image and an unclear or corrupt version of the same image can help to effectively avoid the impact of the contents in the image. Also, it is obvious that the main structural contents of an image like the silhouette of an object or a building will be preserved in situations of both high and low PM2.5 levels. However, the visibility and hence the information about other less prominent features in the image decreases as particulate matter levels increase. This results in a loss of structural information. As clear from image in figure- 19(c), in case of high PM2.5 levels there is a loss in structural information while in the case of low PM2.5 levels, a similar image contains much more structural information. This impact of PM2.5 levels on the structural information of an image was exploited.

Calculating the gradient map of an image is one direct method of measuring structural information loss. This approach has found extensive application in image processing when dealing with reflective structures. The gradient of an image can be used to gauge image sharpness. Any image fusion technique should result in larger gradient values because sharper images are known to have higher gradient values. This is due to the fact that the photos are sharper as a result of this process than photographs with lesser resolutions. The delta of image's clarity and pattern's variance is defined by the mean gradient.

The mean gradient, $\bar{M}$ of an image Z is given by

$$\bar{M} = \frac{1}{N^2}\sum_{i=1}^{N}\sum_{j=1}^{N}\sqrt{\frac{\Delta I_x^2 + \Delta I_y^2}{2}} \quad (3)$$

$$\Delta I_x = Z(i+1, j) - Z(i, j) \quad (4)$$

$$\Delta I_y = Z(i, j+1) - Z(i, j) \quad (5)$$

Where, $\Delta I_!$ and $\Delta I_{\cdot\cdot}$ are the vertical and horizontal gradients per pixel.

In figure 19, we consider two images from the dataset. While the image on the right has a relatively low PM2.5 value, the image on the left is correlated with a high PM2.5 value. Figure-19 (a), 19 (b) and 19 (c) show differences between the images after they undergo the different transformations discussed in the previous sections.

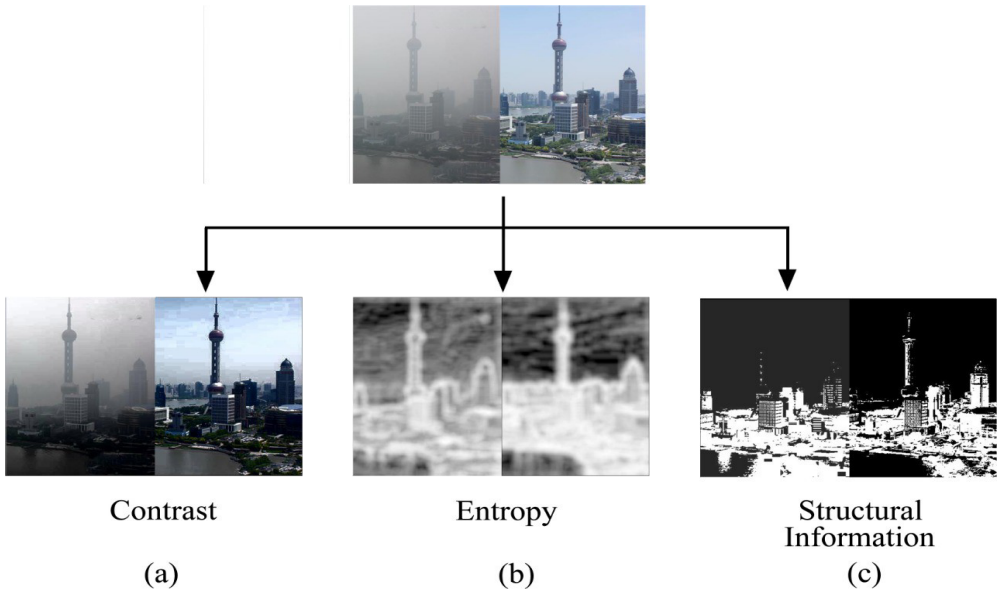

Fig. 19 A comparison between images captured under high (left image) and low (right image) PM2.5 concentrations. (a), (b) and (c) represent the images after they undergo different transformations to accentuate features like contrast, entropy and structural information respectively

V. BASE LEARNERS

In an ensemble, several learners are trained and development to address the same problem. In ensemble modeling, a final output or conclusion is generated by combining output of number of simple models. The model's total power cancels out the biases and variance of each individual model. This provides a composite prediction where the final predictions have a better accuracy than the accuracy of individual models.

In this research, a single type of neural network architecture is used to produce homogenous base learners. The ensemble is developed in two parts. First, three base learners are produced in a parallel style. Next, these base learners are combined with a meta learner which completes the ensemble.

VGG16, introduced by K. Simonyan and A. Zisserman [2], is a convolutional neural network model. VGG16 underwent several weeks of training using the ImageNet [26] dataset. The neural network was trained using NVIDIA Titan Black GPUs. Hence, VGG16 is a pretrained neural network that consists of multiple neural network layers.

VGG16, as outlined in [2], stands as a prominent architecture in the realm of convolutional neural networks (CNNs), acclaimed as one of the foremost designs for tasks related to visual perception. Instead of having many hyper-parameters, VGG16 uses convolution layers with a stride of one and a filter size of 3x3. Additionally, a padding layer and a maxpool layer with a 2x2 filter size and two strides are used. In the entirety of its design, the sequence of convolutional layers and max-pooling layers remains consistent. Towards the final stages of this architectural design, there is a distinct progression. It entails incorporating a set of two fully connected layers, and this is subsequently followed by an additional layer that introduces two more fully connected layers. This layering structure culminates in the integration of a crucial

component, the softmax activation function, which plays an essential role in producing the ultimate output [2]. This architectural sequence is strategically engineered to optimize the neural network's capacity for capturing intricate patterns and relationships within the data, ultimately leading to more accurate and meaningful results. The ‘16’ in VGG16 stands for the sixteen layers of neural network in the architecture that have weights associated with them. This neural network is big and has approximately 138 million parameters. The VGG16 neural network was pretrained on the ImageNet [26] dataset, which comprises approximately 14 million images distributed across 1000 distinct categories.

VI. The Architecture

There are sixteen adjustable parameters in VGG16. It has three completely linked layers, five max pooling layers, and thirteen convolution layers. Sixteen layers overall when three completely connected layers and thirteen convolution layers are added together. The sixteen layers listed above do not include the maximum pooling layers since their weights cannot be adjusted. As illustrated in figure 20, the convolutional layers are separated into five groups with 2, 2, 3, and 3 layers in each group, respectively.

At first, an image goes through the first layer of VGG16. This layer takes in an image of a fixed size of 224 x 224 x 3. The first and second layers of VGG16 are convolution layers. They consist of 64 channels of 3x3 kernel and have a maximum pool stride of two with size 2x2. This means that the max pool windows are non-overlapping windows. We use max pooling to provide robustness towards noise or disturbance in the signal. In the framework of this neural network architecture, the utilization of convolution filters is executed with great precision. These filters, fundamental to the network's feature extraction process, are configured to have a stride value of one. This means that the filters systematically traverse the input data, examining each piece of information in a meticulous manner. Moreover, to ensure the integrity of data processing within every convolutional layer, a technique known as "padding" is employed. Row and column padding are applied strategically. The primary purpose of this padding strategy is to maintain a consistent size for both the input feature map and the output feature map at each convolutional layer.

By introducing padding, the network avoids losing valuable information at the edges of the input data, preserving the spatial relationships and structural nuances

present within the data. This meticulous attention to detail is crucial in the field of image processing, as it allows the neural network to capture and analyze even the most subtle features and patterns in the data, contributing to more accurate and comprehensive results. This careful design choice enhances the network's capacity to recognize essential patterns in the input data, making it a powerful tool for a variety of tasks, including image recognition and analysis.

This means that the resolution of the feature map, after the convolution operation has been performed remains the same. Size reduces proportionally as the image passes through each set of convolution layers and the channel size increases. For instance, the first group layer (layer one and two), has an input dimension as 224 x 224 x 64 with a channel size of 64 and its output dimensions are 112 x 112 x 64. Similarly, the second group layer (comprising of layers three and four), has input dimension of 112 x 112 x 128 with a channel size of 128 and has an output dimension of 56 x 56 x 128. VGG16 is comprised of three fully connected layers. The first two layers have 4096 channels, and the last layer has 1000 channels. Each of these layers have a non-linear ReLU activation function.

The original VGG16 architecture was changed to reflect the fact that predicting PM2.5 concentrations is a regression problem. A single, completely connected neuron which has both, Mean Squared Error (MSE) loss function and linear activation function was used to replace the final, fully connected layer in VGG16, which had 1000 channels. The several layers of neural networks in the VGG16 architecture are illustrated in Figure 20.

The Mean Squared Error (MSE) serves as a pivotal metric in the realm of model evaluation. Its fundamental operation involves computing the square of the difference between each predicted value and the corresponding actual value. This process is

performed for all data points within the dataset under consideration. The squared differences are then averaged, producing the mean squared error.

MSE is utilized as a quantitative measure to assess the performance and quality of a predictive model. Its value indicates the average magnitude of errors between the model's predictions and the actual observed values. In essence, the MSE provides a numerical representation of how closely the model's predictions align with the ground truth. Smaller MSE values denote that the model's predictions are closer to the actual values, implying higher accuracy and reliability. On the other hand, larger MSE values signify greater discrepancies between predictions and observations, highlighting areas where the model may need improvement.

By quantifying the quality of a model, MSE plays a critical role in guiding the development and fine-tuning of machine learning and statistical models. It allows researchers and practitioners to objectively compare different models and select the one that best fits the data and yields the most accurate predictions.. If a vector of n predictions is produced from a sample of n datapoints, where $x_{\#}$ is the vector of observed values for the variable being forecasted and $x^{\$}$ is the vector of predicted values, then the within-sample MSE of the model is calculated using the formula

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(x_i - x_i')^2 \qquad (6)$$

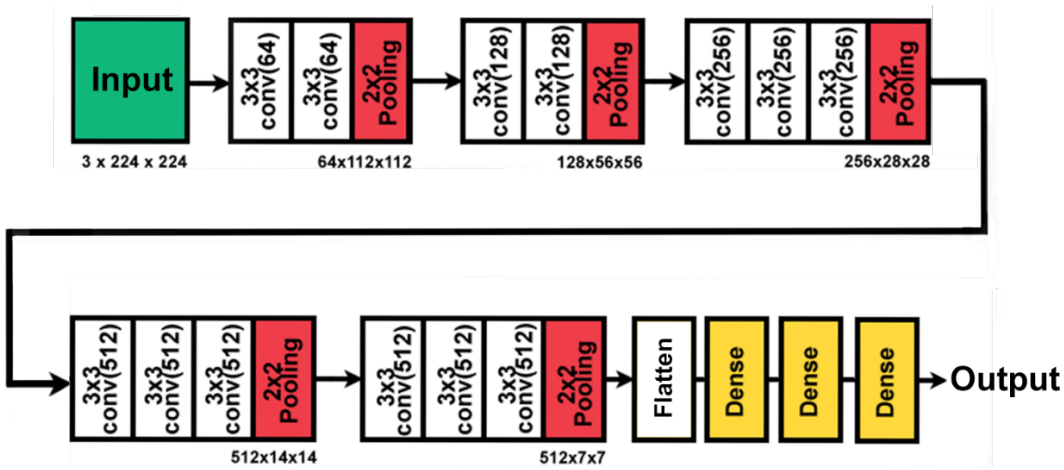

Fig. 20 The VGG16 neural network architecture

VII. EXPERIMENTAL SETTINGS

Given the utilization of the VGG16 architecture, which has been previously pretrained on the extensive ImageNet dataset, for the purpose of transfer learning, a preprocessing step becomes essential. Prior to being employed as inputs to the model, images from the Shanghai dataset undergo this preprocessing procedure. This preprocessing conforms to the methods employed during the initial training of the neural network, guaranteeing that the images are adjusted to be centered around zero for each color channel relative to the ImageNet dataset. After that, the images are resized to 224×224. Since VGG16 neural network was developed to solve classification problems, its final layer consisted of neurons with softmax activation. In response to the regression task at hand, this was substituted with a fully connected dense layer featuring a linear activation function.

The training and test datasets remained uniform for all three base learners. The predicted PM2.5 values from both the training and test datasets were combined and subsequently employed to train the meta-learner, which generates the final output for the proposed model.

Base learners training was carried out in two set of experiments. In the first experiment, all the three neural networks were trained with a batch size of 64 for a maximum of 50 epochs, and an early stopping value of two. Early stopping is used to train the neural networks to the optimal level. It helps to avoid overfitting by terminating the training of the model once the model performance stagnates for a specified number of epochs on the validation dataset.

In the second experiment, a group of 64 were used to train all three neural networks for a maximum of 150 epochs. The value for early stopping was also

increased to five. After every epoch, performance of the neural network was evaluated using the test dataset which was consistent throughout the experiment. It consisted of images which the neural network was not exposed to during training. With MSE as the loss function, Adam optimizer was used. The performance was assessed using RMSE and R-squared (coefficient of determination). The model's performance is considered better when the RMSE value is lower because it signifies that the predicted values are closer to the actual values. R-squared metric value lies between zero and one. It tells us how well the regression model fits the observed data with a value closer to one suggesting a model with a better fit. Table 1 lists the findings from the two experiments. As is clear from table- 1, experiment-2 produces predictors with better performance, in terms of RMSE and R squared, than experiment-1. Figure 21 shows the variation of MSE of the neural networks (y-axis) with the no. of epochs (x-axis) in experiment 1. Figure 22 shows the variation of MSE of the neural networks (y-axis) with the number of epochs (x-axis) in experiment 2.

The application was developed in Python, and the neural networks were trained and evaluated on a Windows 10 64-bit laptop equipped with an Intel i7 (9th generation) processor and 16 GB of RAM. This was accomplished using the Keras (version 2.3.1) library [20] with the TensorFlow (version 2.2.0) backend.

In experiment-1, the neural network trained on the contrast feature of the image, encountered an early stop at 34 epochs (figure- 21(a)). This occurred because the MSE value did not decrease for two consecutive epochs. Early stopping was also observed for the neural network trained on structural information feature at 34 epochs (figure- 21(b)). In the next experiment (experiment 2), when the patience for early stopping was increased from two to five, the neural network trained on the contrast feature encountered an early stop at 105 epochs (figure- 22(a)). The neural network trained on structural

information also experienced an early stopping at 108 epochs (figure- 22(b)). Table-1 gives the details about the two experiments.

Table-1: Results of experiment-1 and experiment-2

Input Feature	Experiment 1 (50 epochs)		Experiment 2 (150 epochs)	
	RMSE	R2	RMSE	R2
Entropy	17.28	0.64	15.52	0.71
Contrast	17.64 (early stopping at 34 epochs with patience = 2)	0.63	14.86 (early stopping at 105 epochs with patience = 5)	0.74
Structural Information	18.71 (early stopping at 34 epochs with patience = 2)	0.58	16.92 (early stopping at 108 epochs with patience = 5)	0.66

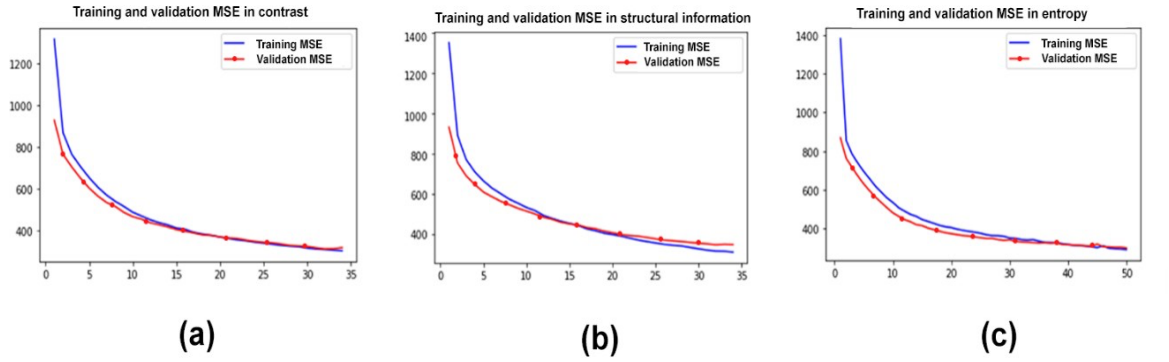

Fig. 21 Curves showing the variation in MSE (on the y-axis) with number of epochs executed (on the x-axis) for experiment 1. 21(a), 21(b) and 21(c) show the variation in MSE for the neural networks trained using the contrast, structural information and entropy features respectively

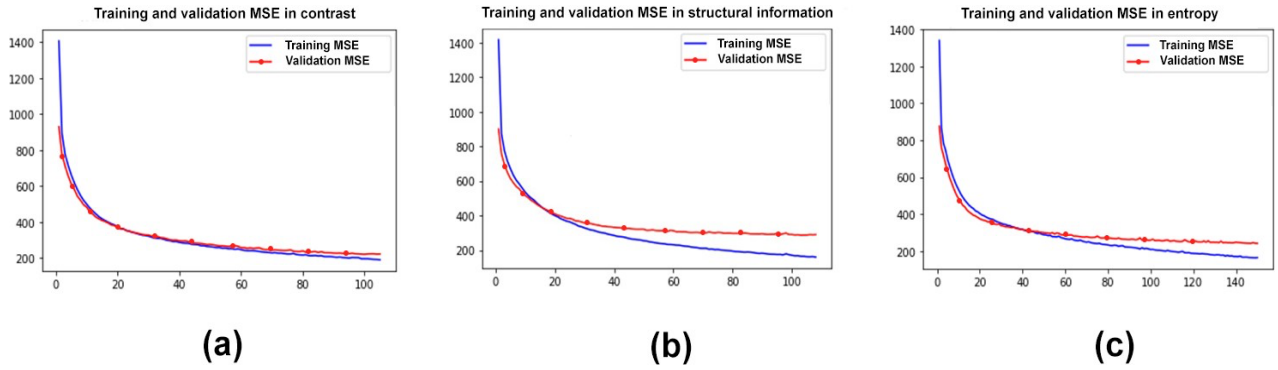

Fig. 22 Curves showing the variation in MSE (on the y-axis) with number of epochs executed (on the x-axis) for experiment-2. 22 (a), 22(b) and 22(c) show the variation in MSE for the neural networks trained using the contrast, structural information and entropy features respectively

VIII. META LEARNER

A meta learner was developed on top of the three homogenous base learners. Multiple machine learning algorithms were tested on top of the base learners. The proposed meta learner was evaluated using an 8-fold cross validation. The complete input data (from the base learners) was divided into eight subsets which were non-overlapping.

The first four subsets were used to train the meta learner and the final one as a test dataset. Above process was repeated eight times and data in every subset were tested once. Out of all the algorithms considered, random forest regressor gave the best performance in terms of predictions.

The random forest algorithm is an additive in nature and performs its final prediction by combining predictions/decisions from a sequence of decision trees. In random forest regressors, the decisions trees are developed independently by using distinct sub-samples from the training data. Rather than relying on individual decision trees, they combine multiple decision trees to make the final prediction.

IX. Evaluation Metrics

In accordance with the recommendation, the model's effectiveness was evaluated by means of R-squared (R2) and root mean square error (RMSE).

The change in a dependent variable that the model can detect is calculated using R squared. It is the correlation coefficient's square (R). R squared substitutes the measured prediction with the mean value and is calculated by dividing the square of the prediction error by the sum of all squares. A greater value of R2 indicates a tighter match between the predicted and actual numbers because its value ranges from 0 to 1. This is a reasonable indication of how well the model fits the dependent variables. That does not, however, account for problems like overfitting.

Mean Square Error (MSE) gauges how effectively the model aligns with the solution, while R2 provides a relative assessment of how closely the model adheres to the dependent variables. It is determined by sum of square of prediction error, which in turn is predicted value minus the actual value by the number of data points. The gap between the results as

predicted and the actual results is expressed as an absolute number. It is not possible to derive other conclusions from a single test, but it provides a specific amount to equate with outcomes from certain projects.

Thus, this aids in selecting the best regression model for a particular issue. MSE's is the square root of Root Mean Square Error (RMSE).

RMSE is defined as

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - y_i')^2} \quad (7)$$

R-squared is mathematically defined as

$$R^2 = 1 - \frac{\sum_{i=1}^{N}(y_i - y_i')^2}{\sum_{i=1}^{N}\left(y_i - avg(y)\right)^2} \quad (8)$$

Where N represents combined total of number of images, y_i and y_i' are the actual (observed) and predicted $PM_{2.5}$ values for the i^{th} image respectively.

X. Results And Discussion

Shanghai dataset was used to test the proposed model. At the conclusion of each stage, RMSE and R squared metrics were generated to assess how well the base and meta learners performed. Table 2 lists performance of the model proposed.

The meta learner model builds on the base learner models and results in significant improvement in predictions as indicated by the RMSE and R squared metrics. RMSE improves by 35.13% and R squared improves by 20%.

Figure 23 shows the ensemble's final output for the Shanghai dataset, with the predicted value on the y-axis and the observed (actual) value of PM2.5 on the x-axis. The red line in the image represents what the output of a model with 100% accuracy would look like. It is observed that as the PM2.5 values increase, The proposed model's forecast precision declines. This is due to the dataset's skewed nature (figure 14) and the fact that it is primarily made up of photos with low PM2.5 values.

Figure 24 depicts some images from the dataset with their observed and predicted PM2.5 values marked below. Due to the multitude of datasets employed for training and testing the different models presented in the literature, making direct comparisons between them is not feasible. But in table 2, a comparison with several cutting-edge models created and trained using the Shanghai dataset has been made. The models have been compared using RMSE and R squared metrics.

As shown in Table 2, the third model integrates weather information, which greatly improves its performance compared to the first model and the second model. But among the three, the proposed model has the best RMSE and R squared values.

Table-2: Comparison of the proposed model with some state-of-the-art models, trained and tested on the Shanghai dataset.

Index	Models	RMSE	R^2
1	**Proposed model**	**9.64**	**0.888**
2	Regression model (without humidity) [5]	19.23	0.570
3	ResNet + Weather features [4]	10.09	0.872

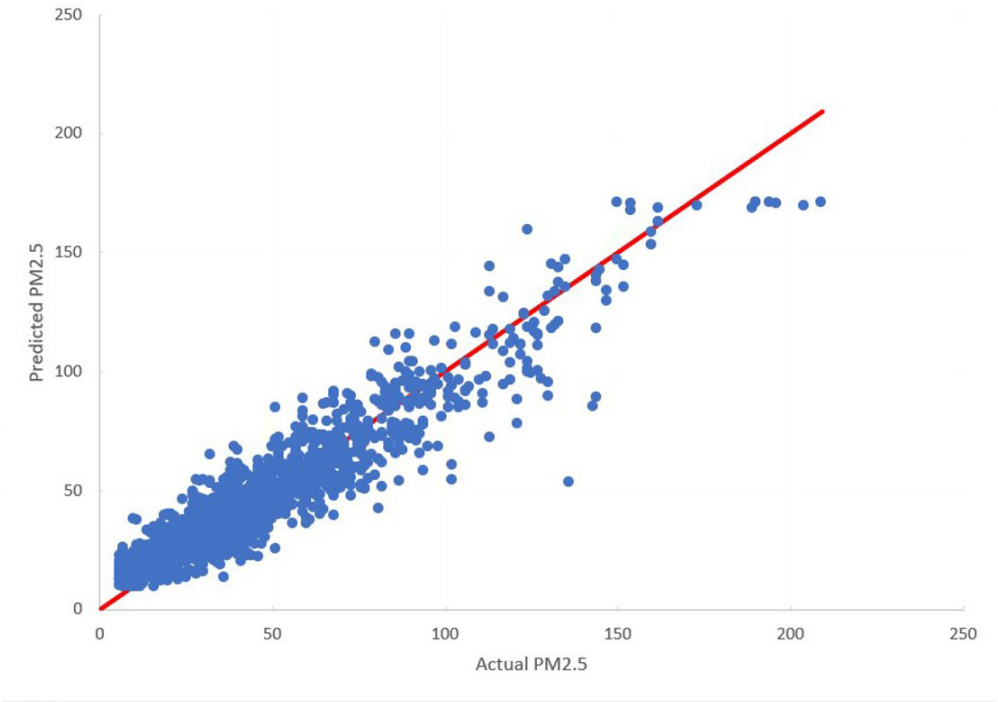

Fig. 23 Graph showing the correlation between actual $PM_{2.5}$ and predicted $PM_{2.5}$ values.

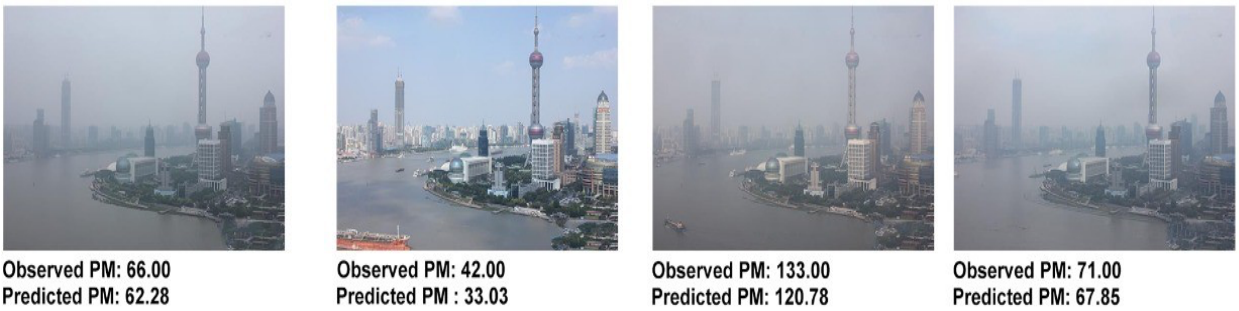

Fig. 24 Images in the dataset with observed and predicted $PM_{2.5}$ level.

XI. Conclusion

Predicting PM2.5 levels, solely, by using images of the outdoor environment, represents a challenging task as it requires an in-depth understanding of different features of an image and the changes in these features with the change in PM2.5 concentrations. This also requires analysis of the visibility and structural information of multiple objects with different depths present in the image.

This study suggests an ensemble method for estimating PM2.5 concentrations from outside photos using deep learning and machine learning techniques. Firstly, the images were pre-processed so as to accentuate information regarding three separate features from them. Following this, three distinct base models were created and trained using the VGG16 architecture in neural networks, each of which was trained using a feature retrieved from the photos. The output of these neural networks was then used to train multiple regressor models. The random forest regressor was picked as the meta learner since it produced the greatest results out of all of them. The model was evaluated using the Shanghai dataset which consists of 1885 images taken from the same location at different times on different days. The meta learner considerably enhanced the model's performance by building on the results of the individual base learners. This led to a final output that was more accurate than any base learner's prediction. Table 2 presents a comparative analysis of the model's performance, specifically focusing on RMSE and R-squared, in relation to the results achieved by other cutting-edge models documented in existing literature. All of these models were assessed using the same Shanghai dataset.

Further, there is a lot of scope for research in this area. Future studies can include testing and using other neural network architectures to develop ensemble methods such

as the one proposed in this paper. A more heterogeneous and larger dataset can be used which contains image samples from different cities and also images of the same location from different angles. Also, a dataset which is more balanced in terms of PM2.5 concentration values can be used. Work can also be done on using other features in an image that might have a higher correlation with the pollution levels. Models which combine other relevant information regarding the environment (such as temperature, humidity, time etc.) along with the features present in the image, can also be pursued. This has the potential to result in more resilient techniques and models for forecasting pollution levels.

REFERENCES

[1] World Health Organization/heath-topic/air-pollution (URL: https://www.who.int/health-topics/air-pollution#tab=tab_1)

[2] Simonyan, Karen & Zisserman, Andrew. (2014). Very Deep Convolutional Networks for Large-Scale Image Recognition. arXiv 1409.1556.

[3] Dietterich, Thomas G. "Ensemble learning." The handbook of brain theory and neural networks 2 (2002): 110-125.

[4] H. L. Cooper and M. I. Miller, "Information measures for object recognition accommodating signature variability," in IEEE Transactions on Information Theory, vol. 46, no. 5, pp. 1896-1907, Aug. 2000, doi: 10.1109/18.857799.

[5] Q. Bo, W. Yang, N. Rijal, Y. Xie, J. Feng, and J. Zhang, "Particle Pollution Estimation from Images Using Convolutional Neural Network and Weather Features," 2018 25th IEEE International Conference on Image Processing (ICIP), 2018, pp. 3433-3437, doi: 10.1109/ICIP.2018.8451306.

[6] C. Zhang, J. Yan, C. Li, X. Rui, L. Liu and R. Bie, “On Estimating Air Pollution from Photos Using Convolutional Neural Network,” 297-301. 10.1145/2964284.2967230.

[7] C.Liu, F.Tsow, Y.Zou, N.Tao (2016) “Particle Pollution Estimation Based on Image Analysis,” PLOS ONE 11(2): e0145955.

[8] G. Yue, K. Gu, and J. Qiao, "Effective and Efficient Photo-Based PM2.5 Concentration Estimation," in IEEE Transactions on Instrumentation and Measurement, vol. 68, no. 10, pp. 3962-3971, Oct. 2019, doi: 10.1109/TIM.2018.2886091.

[9] K. Gu, J. Qiao, and X. Li, "Highly Efficient Picture-Based Prediction of PM2.5 Concentration," in IEEE Transactions on Industrial Electronics, vol. 66, no. 4, pp. 3176-3184, April 2019, doi: 10.1109/TIE.2018.2840515.

[10] K. Li et al., "Discern Depth Under Foul Weather: Estimate PM$_{2.5}$ for Depth Inference," in IEEE Transactions on Industrial Informatics, vol. 16, no. 6, pp. 3918-3927, June 2020, doi: 10.1109/TII.2019.2943631.

[11] J. Ma, K. Li, Y. Han and J. Yang, "Image-based Air Pollution Estimation Using Hybrid Convolutional Neural Network," 2018 24th International Conference on Pattern Recognition (ICPR), 2018, pp. 471-476, doi: 10.1109/ICPR.2018.8546004.

[12] Z. Wang, E. P. Simoncelli and A. C. Bovik, "Multiscale structural similarity for image quality assessment," The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003, Pacific Grove, CA, USA, 2003, pp. 1398-1402 Vol.2, doi: 10.1109/ACSSC.2003.1292216.

[13] K. Gu et al., "Saliency-Guided Quality Assessment of Screen Content Images," in IEEE Transactions on Multimedia, vol. 18, no. 6, pp. 1098-1110, June 2016, doi: 10.1109/TMM.2016.2547343.

[14] K. Gu, L. Li, H. Lu, X. Min and W. Lin, "A Fast Reliable Image Quality Predictor by Fusing Micro- and Macro-Structures," in IEEE Transactions on Industrial Electronics, vol. 64, no. 5, pp. 3903-3912, May 2017, doi: 10.1109/TIE.2017.2652339.

[15] K. Gu, J. Qiao and W. Lin, "Recurrent Air Quality Predictor Based on Meteorology- and Pollution-Related Factors," in IEEE Transactions on Industrial Informatics, vol. 14, no. 9, pp. 3946-3955, Sept. 2018, doi: 10.1109/TII.2018.2793950.

[16] P.Y. Wong, H. Y. Lee, et al., “Using a land use regression model with machine learning to estimate ground level PM.sub.2.5,” vol. 277, 2021, doi: 10.1016/j.envpol.2021.116846.

[17] M. Castelli, F. M. Clemente, A. Popovič, S. Silva, and L. Vanneschi, “A Machine Learning Approach to Predict Air Quality in California,” vol. 2020, pp. 1–23, Aug. 2020, doi: 10.1155/2020/8049504. [Online].

[18] G. Li, C. Fang, S. Wang, and S. Sun, “The effect of economic growth, urbanization, and industrialization on fine particulate matter (PM2.5) concentrations in China” Environmental science & technology, 50(21), pp.11452-11459, 2016

[19] A. Chakma, V. Ben, T. Cao, J. Lin, and J. Zhang, "Image-based air quality analysis using deep convolutional neural network." In Image Processing (ICIP), 2017 IEEE International Conference on, pp. 3949-3952. IEEE, 2017

[20] S. Du, T. Li, Y. Yang and S. -J. Horng, "Deep Air Quality Forecasting Using Hybrid Deep Learning Framework," in IEEE Transactions on Knowledge and Data Engineering, vol. 33, no. 6, pp. 2412-2424, 1 June 2021, doi: 10.1109/TKDE.2019.2954510.

[21] Wen, C., Shufu, L., Yao, X., Peng, L., & Li, X. (11 2018). A novel spatiotemporal convolutional long short-term neural network for air pollution prediction. Science of The Total Environment, 654. doi:10.1016/j.scitotenv.2018.11.086

[22] Eeftens M, Beelen R, de Hoogh K et al. Development of Land Use Regression models for PM(2.5), PM(2.5) absorbance, PM(10) and PM(coarse) in 20 European study areas; results of the ESCAPE project. Environ Sci Technol. 2012 Oct 16;46(20):11195-205. doi: 10.1021/es301948k. Epub 2012 Oct 1. PMID: 22963366.

[23] Hu, X.; Waller, L.A.; Al-Hamdan, M.Z.; Crosson, W.L.; Estes, M.G., Jr.; Estes, S.M.; Quattrochi, D.A.; Sarnat, J.A.; Liu, Y. Estimating ground-level PM2.5 concentrations in the southeastern US using geographically weighted regression. Environ. Res. 2013, 121, 1–10

[24] Park Y, Kwon B, Heo J, Hu X, Liu Y, Moon T. Estimating PM2.5 concentration of the conterminous United States via interpretable convolutional neural networks. Environ Pollut. 2020 Jan;256:113395. doi: 10.1016/j.envpol.2019.113395. Epub 2019 Oct 23. PMID: 31708281.

[25] L. Feng, Y. Li, Y. Wang, and Q. Du, “Estimating hourly and continuous ground-level PM2.5 concentrations using an ensemble learning algorithm: The ST-stacking model,” vol. 223, p. 117242, 2020, doi:10.1016/j.atmosenv.2019.117242. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S1352231019308805

[26] J. Deng, W. Dong, R. Socher, L. -J. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," 2009 IEEE Conference on Computer Vision and Pattern Recognition, Miami, FL, USA, 2009, pp. 248-255, doi: 10.1109/CVPR.2009.5206848.

[27] PM2.5 Estimation Based on Image Analysis. (2020, February 29). KSII Transactions on Internet and Information Systems. Korean Society for Internet Information (KSII). https://doi.org/10.3837/tiis.2020.02.025