# Enhanced Deep Maxout Network for Monitoring Particulate Matter 2.5 and 10 Concentration in Air via Interpolated Data Smoothing

Rajendran Thavasimuthu[1], Saranya Arnise[2], Arulkumar Varatharajan*[3], Sridhar Sekar[4], Reshmy A K[5], Ajay Kumar Yadav[6]

[1]Department of Sustainable Engineering, Saveetha School of Engineering, Saveetha Institute of Medical and Technical Sciences, Chennai, Tamilnadu, India. rajendran.thavasimuthusamy@gmail.com

[2]Department of Computer Science and Engineering, Amity School of Engineering and Technology, Amity University, Mumbai, Maharashtra, India. mayflower55@protonmail.com

[3]School of Computer Science and Engineering, Vellore Institute of Technology, Vellore, Tamilnadu, India. arulkumaran.ckpc@gmail.com

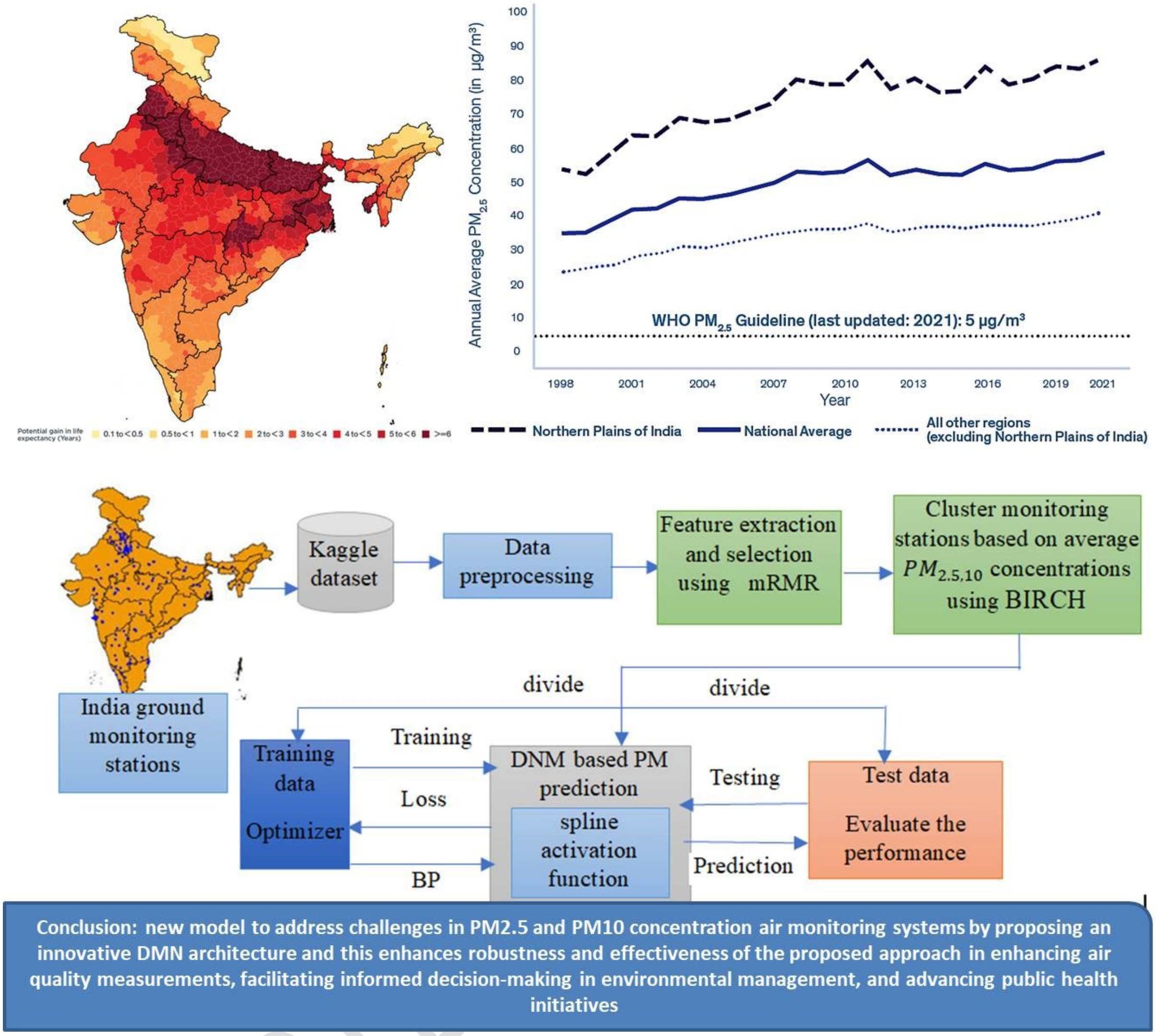
[4]Department of Research, Rajalakshmi Institute of Technology, Chennai, Tamilnadu, India. sridhar.sse@protonmail.com

[5]Department of Computational Intelligence, Faculty of Engineering and Technology, SRM Institute of Science and Technology, Kattankulathur, Chennai, Tamilnadu, India. malu.flower@proton.me

[6]Department of Computer Application, United Institute of Management, Allahabad, Uttar Pradesh, India. ajaykumar.uim@gmail.com

Corresponding Author: arulkumaran.ckpc@gmail.com

## Graphical Abstract




**Abstract:** Currently, most of the global population resides in metropolitan areas, where air quality standards are not properly monitored. As a result, people are constantly exposed to air contaminants that exceed the thresholds set by the World Health Organization (WHO). Air quality monitoring system is often encountering challenges such as discontinuities and missing data in time sequences, affecting the accuracy of measurements. This paper presents an innovative approach to address these issues in PM 2.5 and 10 concentration air monitoring systems proposes a novel Deep Maxout Network (DMN) architecture enhanced with Polynomial and Spline Interpolation methods to effectively handle the discontinuities in data sequences. By smoothing transition fitting curves at interval connections, the proposed model generates an optimal dataset, improving the robustness and accuracy of air quality measurements. First, the data is collected and pre-processed. Then, the features are extracted and selected by using minimum redundancy maximum relevance (mRMR). Then similar features are clustered by using Balanced Iterative Reducing and Clustering using Hierarchies (BIRCH) scheme. Finally, the PM concentration is predicted by using DMN. Experimental results demonstrate the effectiveness of the proposed approach in enhancing the reliability of Matter2.5 and 10 concentration monitoring systems using Air Quality Data in India from kaggle, providing a promising scope for precise $PM_{2.5\ and\ 10}$ concentration forecasting with practical implications for air quality management and public health initiatives.

**Keywords:** Air quality monitoring, BIRCH, Deep Maxout Network, Minimum Redundancy Maximum Relevance, PM Concentration 2.5 and 10 forecasting.

## 1. Introduction

Air pollution is a prominent global environmental issue, and India is among the nations that experience substantial impacts from elevated concentrations of particulate matter (PM) in the environment. Particulate matter comprises micro particles that are spread out in the atmosphere. These particles are classified based on their size, with PM10 (particles measuring 10 micrometres or smaller) and PM2.5 (particles measuring 2.5 micrometres or smaller) being particularly hazardous because they can deeply infiltrate the respiratory system. Monitoring and predicting PM10 and PM2.5 levels in real-time are crucial for assessing air quality and implementing timely interventions to mitigate the adverse health effects associated with air pollution. Therefore, the development of accurate prediction models for PM10 and PM2.5 levels has become a focus of research, especially in regions like India where air pollution poses significant challenges to public health and environmental sustainability.

Recently, the quick growth of industrialization has been followed by a concerning rise in air pollution, drawing global attention due to its severe impacts, resulting in the deaths of approximately 7 million people annually (Jasarevic et al. 2021) (Li et al. 2015). Among the various air contaminants, PM2.5 stands out as a particularly hazardous component, capable of penetrating the nasal passages and reaching the lungs and throat upon inhalation (Di et al. 2017), posing a significant threat to human health. A study by Heft-Neal et al. in 2018 revealed that PM2.5 concentrations exceeding minimum exposure levels contributed to 22% of infant deaths in 30 surveyed countries, resulting in approximately 449,000 additional infant deaths in 2015, a figure more than triple the existing estimates attributing infant mortality to poor air quality (Heft-Neal et al. 2018). Consequently, controlling and preventing air pollution has become an urgent global priority. Real-time monitoring of air pollution levels is essential to achieve this goal (Cheng et al. 2018), and the use of sensors has facilitated the collection of extensive air quality data across various applications (Xue and Chen, 2020) (Xue and Chen, 2019).

### *1.1. Air Pollution in Indian Scenario*

India ranks as the second most polluted nation globally. The average life expectancy of an Indian is reduced by 5.3 years due to fine particle air pollution (PM2.5), compared to the life expectancy if the WHO recommendation of 5 $^{1}$g/m3 were observed. Certain regions in India experience much higher levels of air pollution, resulting in a reduction in life expectancy by 11.9 years in the National Capital Territory of Delhi, which is recognized as the most polluted city globally. All 1.3 billion individuals in India reside in regions where the yearly mean level of particle pollution above the guideline set by the WHO. Specifically, 67.4 percent of the population resides in places that surpass the country's own national air quality threshold of 40 µg/m3 (Fiordelisi et al. 2017).

Particulate pollution poses the most significant risk to human health in India, reducing the typical Indian's life expectancy by 5.3 years. Cardiovascular disorders have a negative impact on the average life expectancy of Indians, reducing it by around 4.5 years. Similarly, infant, and maternal malnutrition decrease life expectancy by 1.8 years. The level of particulate pollution has risen throughout the course of time. Between 1998 and 2021, there was a significant rise of 67.7 percent in the average annual particle pollution, resulting in a further decrease of 2.3 years in the average life expectancy. India has accounted for 59.1 percent of the global pollution rise between 2013 and 2021. If present pollution levels remain, 521.2

million individuals, which accounts for 38.9 percent of India's population, in the most polluted region of the country, are projected to have an average loss of 8 years in life expectancy compared to the WHO recommendation and 4.5 years compared to the national norm. If India were to decrease particle pollution to comply with the WHO recommendation, the inhabitants of Delhi, the capital and most populous city of India, would see an increase in life expectancy by 11.9 years. The population of North 24 Parganas, which is the second most populated area in the country, will experience an increase of 5.6 years in their life expectancy.

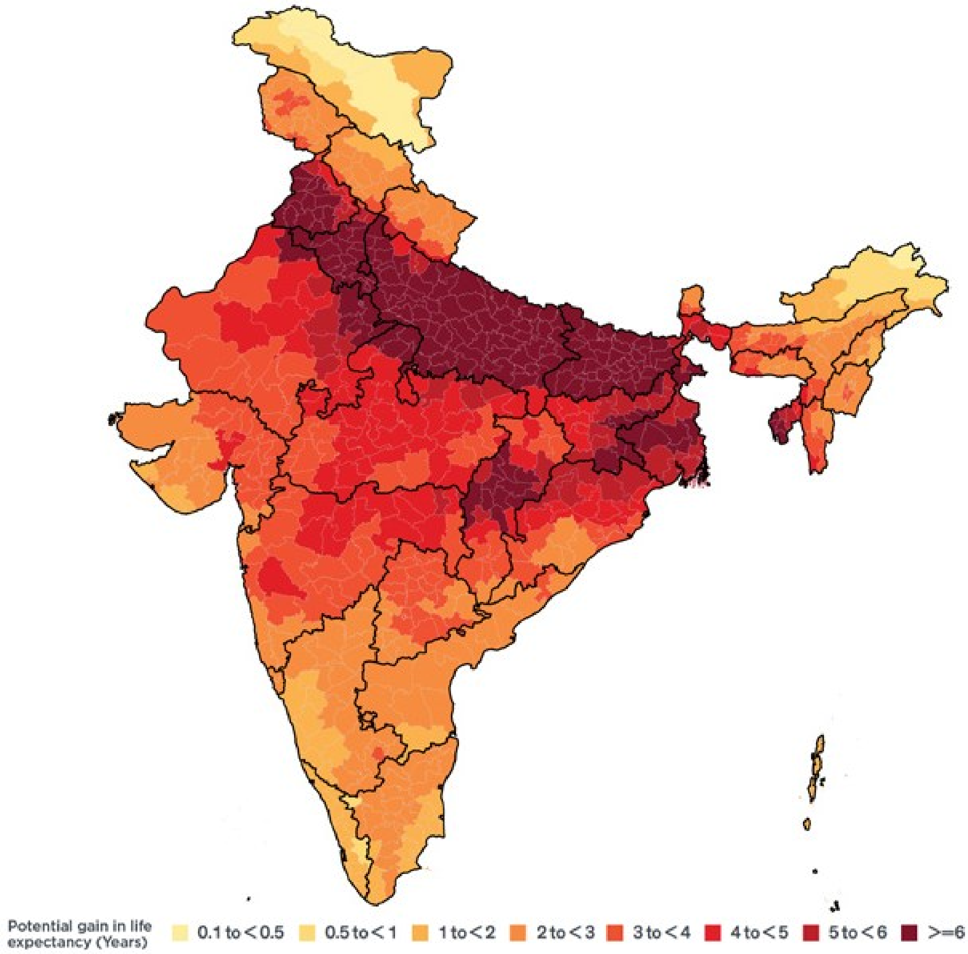



**Figure 1.** Potential Gain in Life Expectancy Reduction on PM2.5 Effect from 2021 in India

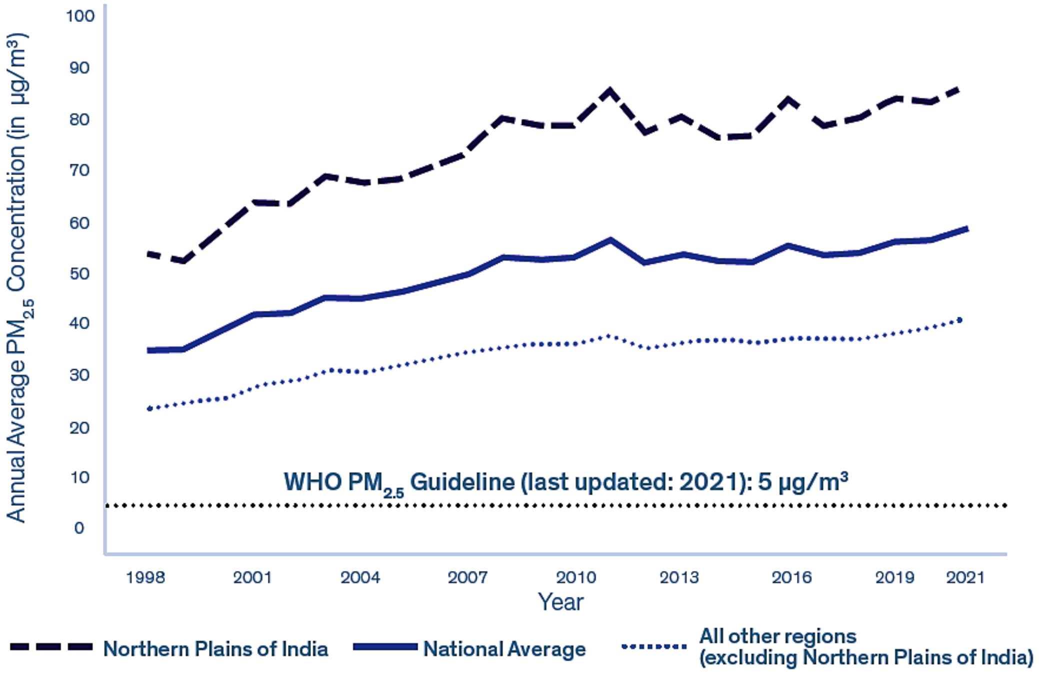



**Figure 2.** Average PM2.5 Concentrations in India

Given the heightened focus on air pollution, numerous researchers have dedicated significant efforts to studying this issue, resulting in a plethora of relevant research studies. Primary among the machine learning (ML) approaches implemented to air pollution prediction are artificial neural networks (ANNs), ensembled learning techniques, support vector machines (SVMs), and various hybrid methodologies (Mendez et al. 2023). Eventually, many current air quality prediction methodologies primarily emphasize model selection and overlook the analysis of factors driving changes in air pollution concentration. Furthermore, the recent surge in deep learning frameworks offers flexibility but can lead to the development of deep and complex models to fit datasets. Consequently, overfitting issues may arise, especially with large neural network models containing numerous parameters.

This research confronts the critical issue of unreliable air quality monitoring in major cities, where people are constantly exposed to pollutants exceeding WHO safety limits. Existing monitoring systems struggle with gaps and missing data, hindering measurement accuracy. To address this challenge, this research proposes a novel solution: an Enhanced Deep Maxout Network (DMN) architecture, empowered by Polynomial and Spline Interpolation methods. This innovative approach strives to create seamless transition curves at data intervals, generating a more complete dataset and significantly improving the robustness and accuracy of air quality measurements, especially for PM2.5 and PM10 concentrations. The research aims to not only develop but also validate this novel methodology, which is pivotal for advancing air quality monitoring capabilities and facilitating informed decision-making in environmental management. Ultimately, this approach aids public health from the detrimental effects of air pollution. Furthermore, the research carries global potential, offering a solution for improving air quality monitoring systems worldwide and contributing to broader sustainable development initiatives. The main contributions of the work are:

- Cleanse and preprocess the data to handle missing values, outliers, and inconsistencies. Normalize or scale the features to ensure uniformity and facilitate clustering analysis. Extract relevant features from the air quality data, including pollutant concentrations such as PM2.5, PM10, and meteorological variables like temperature, humidity, wind speed, and temporal data like time of day, day of week.

- Utilize techniques such as mRMR or feature selection to reduce dimensionality and enhance clustering performance. BIRCH clustering algorithms is applied to group similar air quality patterns and identify distinct clusters representing different pollution profiles. Analyse the characteristics of each cluster to understand the underlying patterns and factors influencing air pollution levels.
- Identify common features and trends within clusters, such as high pollutant concentrations during specific time periods, Incorporate the clustered information as additional features or contextual factors into air quality prediction models.
- The efficiency of the proposed methodology is demonstrated by the experimental findings in enhancing the reliability of air quality monitoring systems, offering practical implications for air quality management and public health initiatives in India.

The organisation of this work is as follows: section 2 describes the research methodology and section 3 evaluates the performance of proposed scheme; section 4 concludes the work.

## 2. Related Works

Prior research works have examined the forecasting of PM2.5 levels, predominantly employing numerical or statistical learning techniques. Significantly, deep learning (DL) approaches have emerged as a prominent and extensively embraced aspect of statistical learning. These strategies have been found to be helpful in overcoming issues that are often encountered by traditional models. The effectiveness of deep learning in predicting PM2.5 levels is ascribed to its ability to effectively process large datasets, a critical factor in this form of prediction [2.5]. Temporal data on PM2.5 exhibits a dynamic functional connection. Deep learning has demonstrated exceptional proficiency in representing complex relationships and has displayed outstanding results in diverse time-series prediction tasks. As a result, it has emerged as the favoured method for addressing the difficulties associated with predicting PM2.5 concentrations. Deep learning is utilized as a fundamental approach in PM2.5 concentration forecasts to enhance the accuracy and efficiency of mathematical simulation approaches. This study offered a thorough examination of deep learning as the fundamental technique for forecasting the concentration of PM2.5 particles. Multiple studies have examined deep learning-based methods for forecasting PM2.5 levels, offering valuable insights from different viewpoints.

The study conducted by (Liu et al. 2021) employed Q-learning for ensuring the Graph reinforcement learning Convolutional Network-Long Short-Term Memory-Gradient Recurrent Unit (GCN-LSTM-GRU) deep learning approach achieved convergence to an optimum policy with specific constraints. Q-learning was a reinforcement learning technique that was particularly useful for handling contexts with extensive or uninterrupted state spaces. Although the integration of ML and deep learning provide robust models for several applications, these models continue to have certain constraints, including limited interpretability and high computational costs. Consequently, it becomes more challenging to track and utilize these models on devices with limited resources for future utilization.

The work by (Wu et al. 2024) developed a novel hybrid model was created to estimate the mass concentration of PM2.5 and PM10 with minimal reliance on on-site data. The PM10 and PM2.5 concentrations in Beijing, China were estimated utilizing the Gaofen-1 satellite and Moderate Resolution Imaging Spectroradiometer (MODIS) data, with a spatial resolution of 100 m. Subsequently, the PM10/2.5 mass concentrations information from 2020 were utilized to do the spatio-temporal study aimed at examining the characteristics of particulate matter in Beijing. The ground stations provided validation for the estimation results of PM2.5, with R2 values varies from 0.91 to 0.98 and root mean squared errors (RMSE) varies from 4.51 to 17.04 μg/m3. Similarly, the ground stations validated the estimation results for PM10, with R-squared values varies from 0.85 to 0.98 and RMSE values varies from 6.98 to 29.00 μg/m3.

Furthermore, the studies by (Zhang et al. 2022) and (Li et al. 2022) introduced hybrid frameworks that included a Convolutional Neural Network (CNN), a Long Short-Term Memory (LSTM), and an attention mechanism. Additionally, (Zhang et al. 2022) focused on the estimation of air pollution at a fine-grained level. Implementing an attention mechanism enabled the model to selectively concentrate on significant features. DL hybrid models has several notable qualities, including their capacity to effectively capture intricate correlations and patterns inherent in data of PM2.5 concentrations, integrate Spatio temporal data, and demonstrate adaptability to diverse environment situations. Nevertheless, deep learning hybrid methodologies may want substantial computer resources, extensive data, and meticulous model optimization to get ideal outcomes.

The study by (Ejurothu et al. 2023) suggested the use of clusters-based Local Hybrid Graph Neural Networks (HGNN) approach as an alternative to employing a singular GNN for the purpose of monitoring stations-wise multi-steps PM2.5 concentrations forecasts throughout the states of India. This technique acknowledged and accommodated abrupt fluctuations in PM2.5 levels caused by local weather variability. The Hybrid GNN models at the local level were composed of two components: a spatio-temporal component that incorporates the GNN and Gated Recurrent Unit layers. This unit was designed to capture the impact of wind velocity and various meteorological factors on PM2.5 concentrations. The next component was a unit for extracting meteorological features at each station to determine their influence on PM2.5 concentrations. It also examines the temporal relationship among historic data. The work by (Mohan and Abraham, 2023) developed a novel Ensemble Deep Particulate Forecaster (EDPF) model was introduced, which integrates a LSTM network, CNN, and random forests model.

The study by (Gunasekar et al. 2022) introduced an optimized and sustainable hybrid model called ARTOCL. This system combined LSTM and CNN to enhance the accuracy of air quality predictions and minimize false alarms. The hybrid deep learning model presented by (Mao et al. 2023) integrated BiGRU, CNN, and fully connected layers. Both techniques have the benefit of being capable of modelling both spatial and temporal patterns. A hybrid deep learning model was developed in the work by [Chiang and Horng, 2021], which combined a stacked autoencoder (SAE), GRU, and CNN. The method underwent training using an extensive data set of air pollution data from China. It demonstrated the capability to forecast hourly concentrations of various air pollutants with a lead time of up to 24 hours. The model has notable features, such as its capacity to effectively address missing data and its exceptional predictive accuracy. Hence, the hybrid model effectively leveraged the advantages of CNN and GRU to acquire knowledge of both global and local trends in the data of time-series, rendering it highly suitable for predicting air pollutants concentrations. The CNN components of the framework acquired knowledge of local patterns within the temporal sequences of values, whilst the GRU captured longer-term relationships and trends.

The study conducted by (Ding and Noh, 2023) introduced a hybrid model called Interpretable Neural Networks and a Graph Neural Network (INNGNN), which combined an interpretable neural network with a graph neural network. This model effectively captured the temporal and geographical variations in air quality and demonstrated precise prediction of air quality over many steps. The initial step was the utilization of interpretable neural networks (INN) to analyze a time series dataset, with the aim of identifying and extracting significant components that may have been neglected. Subsequently, a self-attention mechanism was employed to capture both local and global dependencies and linkages within the time series. Finally, a city map was generated utilizing a GNN to ascertain the interconnections among cities with the aim of extracting geographically specific characteristics.

The study by (Zhang et al. 2023) applied enhanced complementary ensemble empirical mode decomposition (CEEMD)-LSTM models that were integrated with Fully Convolutional Network (FCN) and CNN. Convolutional layers were utilized to facilitate feature selection, hence improving the accuracy of predictions. A novel hybrid model was introduced in the work, which integrated various deep learning approaches with statistical techniques. The features were extricated, and the PM2.5 concentration was predicted using LSTM. Nevertheless, a notable limitation of the approach lies in its dependence on the meteorological data, potentially constraining its precision in regions with limited availability of meteorological monitoring stations.

**Table 1.** Review on Analyzed Current Works

| Approach Used | Application | Advantages | Disadvantages |
|---|---|---|---|
| Q-learning, GCN-LSTM-GRU | Air quality forecasting; Handling extensive or uninterrupted state spaces. | Robust models for various applications; Effective convergence to optimal policy with specific constraints. | Limited interpretability; High computational costs. |
| Satellite and MODIS data | Estimation of PM2.5 and PM10 concentrations in Beijing, China; Spatial resolution of 100m. | Minimal reliance on on-site data; Estimation validated against ground stations with high $R^2$ and low RMSE values. | Reliance on satellite data; Limited to specific geographical area. |
| Hybrid CNN-LSTM-Attention | Estimation of air pollution at a fine-grained level; Selective concentration on significant features. | Effective capture of intricate correlations and patterns in PM2.5 concentrations; Integration of spatiotemporal data. | Substantial computational resources; Extensive data requirements; Model optimization challenges. |
| Hybrid GNN | Monitoring stations-wise multi-steps PM2.5 concentrations forecasts in India; Handling abrupt fluctuations due to local weather variability. | Accommodation of local weather variability; Capturing impact of meteorological factors on PM2.5 concentrations. | Limited geographical scalability; Complex architecture. |
| EDPF Model | PM2.5 concentration forecasting; Integration of LSTM, CNN, and random forests model. | Combines strengths of LSTM, CNN, and random forests; Potential for improved predictive accuracy. | Computational complexity. |
| LSTM- | Air quality prediction; | Enhanced accuracy of air | Computational |

| | | | |
|---|---|---|---|
| CNN Model | Combining LSTM and CNN. | quality predictions; Minimization of false alarms. | complexity. |
| Hybrid BiGRU-CNN | Spatial and temporal pattern modeling; Integration of BiGRU, CNN, and fully connected layers. | Modeling of both spatial and temporal patterns; Integration of global and local trends in data. | Complexity in architecture; Computational resources requirement. |
| Hybrid SAE-GRU-CNN | Hourly air pollutants concentration forecasting; Combination of stacked autoencoder (SAE), GRU, and CNN. | Effective addressing of missing data; Exceptional predictive accuracy; Knowledge acquisition of global and local trends in time-series data. | Potential overfitting; Complex architecture. |
| INNGNN Model | Temporal and geographical variations in air quality prediction; Combination of interpretable neural network and graph neural network. | Precise prediction of air quality over many steps; Identification of significant components in time series data. | Complexity in architecture; Reliance on interpretability of neural networks. |
| CEEMD-LSTM-FCN-CNN | PM2.5 concentration prediction; Integration of CEEMD-LSTM models with FCN and CNN. | Improved accuracy of predictions; Feature selection facilitated by convolutional layers. | Dependence on meteorological data; Potential limitations in regions with limited meteorological monitoring stations. |
| Hybrid Model | PM2.5 concentration forecasting. | Effective utilization of various deep learning and statistical techniques; Potential for improved predictive accuracy. | Dependence on meteorological data; Potential constraints in regions with limited meteorological monitoring stations; Complexity in model integration. |

## 3. Data and Methods

This section introduces an innovative approach to enhance the reliability of PM2.5 and PM10 concentrations monitoring systems. It presents a novel Deep Maxout Network (DMN) architecture shown in fig 3, enhanced with Polynomial and Spline Interpolation methods, to effectively handle data discontinuities, thereby generating an optimized dataset for more robust and accurate air quality measurements. The system begins with data collection and preprocessing, followed by feature extraction and selection using mRMR criteria. Similar features are then clustered using the BIRCH scheme. Finally, PM concentration is predicted using the DMN architecture. Experimental results using Air Quality Data from India sourced from Kaggle repository demonstrate the effectiveness of the proposed approach, providing a promising solution for precise forecasting of PM2.5 and

PM10 concentration with practical implications for air quality management and public health initiatives.

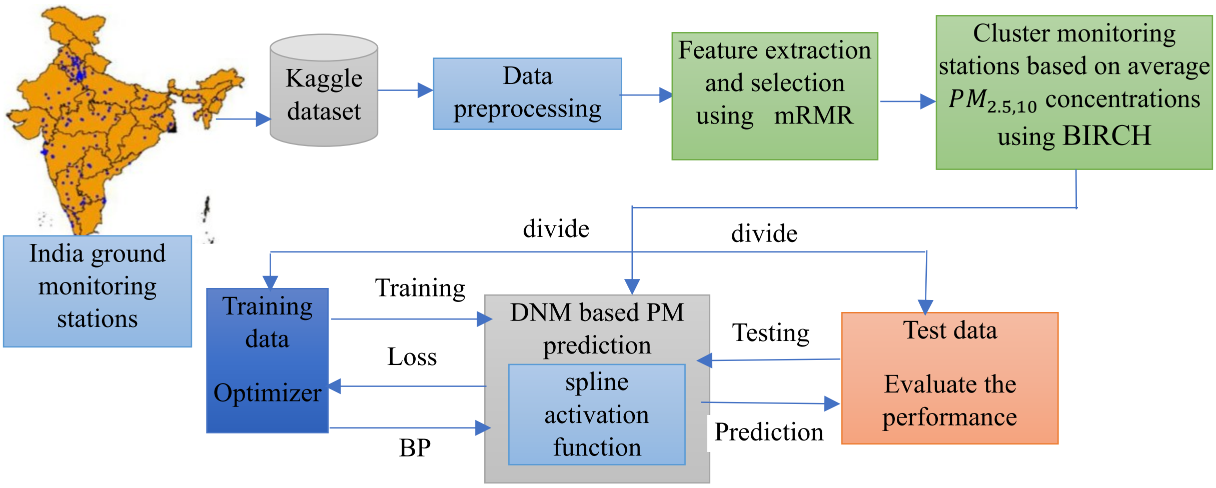



**Figure 3.** DMN based PM Concentrations Prediction Overall Process

*3.1. Dataset Description*

The dataset was collected from the website https://www.kaggle.com/datasets/fedesoriano/air-quality-data-in-india by Fedesoriano in 2022. Providing data on significant air pollutants like particle matter (PM2.5 and PM10), carbon monoxide (CO), sulfur dioxide ($SO_2$), nitrogen dioxide ($NO_2$), and ozone (O3), across several cities in India. The dataset often contains timestamps that correlate to the day and time of measurement, as well as pollutant concentrations measured in quantities such as micrograms per cubic meter (μg/m³). This dataset is an essential source for analysts and researchers to analyze air quality trends, investigate the impact of pollution on public health, develop predictive models for predicting air quality, and evaluate the effectiveness of air quality management strategies and policies. The metadata includes information on monitoring locations and quality control.

*3.2. Preprocessing using Inverse Distance Weighting (IDW)*

IDW interpolation can be used for data preprocessing and normalization in air quality datasets to estimate missing values or to create a continuous surface from sparse monitoring points (De Mesnard, 2013). The IDW formula for air quality datasets can be adapted as follows:

$$AQ(p) = \frac{\sum_{i-1}^{n} AQ_i/d^x}{\sum_{i-1}^{n} 1/d_i^x} \quad (1)$$

Where $AQ(p)$ as the estimated air quality value at location p, $AQ_i$ is the measured air quality value at known location $i$ and $d_i$ as the Euclidean distance among the unknown location $p$ and the location known $i$, $x$ as the power parameter controlling the rate of distance decay and $n$ as the number of known monitoring points used in the interpolation.

From eqn. (1), the numerator represents the weighted sum of measured air quality values at known locations, where the weight assigned to each measurement is inversely proportional to the distance between the known location and the estimation location raised to the power $x$.

The denominator represents the sum of the weights, ensuring that the weighted average is properly normalized. The power parameter $x$ controls the rate at which the influence of distant monitoring points decreases as distance increases. Typical values for $x$ range from 1 to 3, with larger values giving more weight to nearby points. By applying the IDW interpolation method, missing air quality data can be estimated, and the dataset can be normalized to create a continuous surface, facilitating further analysis and visualization of air quality patterns.

### *3.3. Feature Selection using mRMR*

The minimum redundancy maximum relevance (mRMR) algorithm is a heuristic feature selection method that aims to identify the subset of features that maximizes the relevance to the target variable while minimizing redundancy among selected features (Radovic, et al. 2017). It does not involve solving a specific mathematical model but rather relies on ranking features based on relevance and redundancy criteria.

Relevance Measure: The relevance of a feature to the target variable (e.g., PM concentration) can be quantified using a suitable metric, such as mutual information, correlation coefficient, or information gain. Mathematically, this can be represented as:

$$R(P_i, Q) = MI(P_i; Q) \tag{2}$$

Where $P_i$ is the $i$-th feature, Q is target variable, and $MI(P_i; Q)$ denotes the mutual information between $P_i$ and Q.

Redundancy Measure: The redundancy between two features $P_i$ and $P_j$ can be quantified using a measure such as conditional mutual information or inter-feature distance. Mathematically, this can be represented as:

$$Red\left(P_i, P_j\right) = MI(P_i; P_j) \tag{3}$$

Where $MI(P_i; P_j)$ denotes the mutual information between $P_i$ and $P_j$.

mRMR Criterion: The mRMR criterion balances relevance and redundancy to rank features. It aims to maximize the relevance of selected features while minimizing their redundancy. Mathematically, the mRMR criterion can be expressed as:

$$mRMR(P_i) = R(P_i; Q) - \frac{1}{k}\sum_{j=1}^{k} Red\left(P_i, P_j\right) \tag{4}$$

Where $P_i$ is the $i - th$ feature, $Q$ is the target variable, $k$ is the number of previously selected features, and $X_j$ represents the $j - th$ selected feature.

Feature Ranking: Features are ranked based on their mRMR scores, with higher scores indicating greater relevance and lower redundancy. The top-ranked features are selected for further analysis or model training.

Iterative Approach: The mRMR algorithm may involve an iterative approach where features are selected one at a time based on the mRMR criterion. After each feature selection step, the relevance and redundancy measures are recalculated to account for the updated set of selected features.

By applying the mRMR algorithm, a subset of features highly relevant to PM concentration can be selected while minimizing redundancy among selected features, facilitating more efficient and interpretable predictive models.

### *3.4. BIRCH Algorithm for PM Concentration Clustering*

After feature selection, the same profile features are clustered by using BIRCH clustering algorithms (Lorbeer et al. 2018). It is used to group similar air quality patterns and identify distinct clusters representing different pollution profiles. Ands Identify common features and trends within clusters, such as high pollutant concentrations during specific time periods, under particular meteorological conditions. Interpret the clusters in terms of their implications for air quality management and potential interventions.

Assume a cluster consisting of *n* d-dimensional data points or objects. The clustering features (CFs) of the clusters are a three-dimensional vector that effectively summarizes data about the cluster of the objects. The CF includes three components: the centroid, the radius, and the number of points in the cluster.

Centroid: This represents the center point of the cluster and is calculated as the average of the coordinates of all points in the cluster along each dimension.

Radius (R): The radius of the cluster indicates the spread or dispersion of the points around the centroid. It can be computed as the maximum distance between the centroid and any point within the cluster.

Number of Points: This component simply denotes the total number of data points in the cluster.

The clustering feature compactly captures essential characteristics of the cluster, enabling efficient computation and storage.

Additionally, BIRCH employs the Clustering Feature Tree (CF-tree) to represent a hierarchy of clusters. The CF-tree structure facilitates scalability and efficiency in handling large or streaming databases, as well as enables incremental and dynamic clustering of incoming objects.

By utilizing these structures, BIRCH overcomes two significant challenges encountered in agglomerative clustering approaches: the inability and scalability to reverse or undo previous clustering decisions. The CF and CF-tree enable BIRCH to efficiently summarize clusters and organize them into a hierarchical structure, making it suitable for handling large datasets and dynamic clustering scenarios. It is defined as follows.

$$CF =< n, LeS, SqS > \qquad (5)$$

where $LeS$ is the linear sum of the *n* points (i.e., $\sum_{i=1}^{n} p_i$), and $SqS$ is the square sum of the data points (i.e., $\sum_{i=1}^{n} p_i^2$).

The utilization of the clustering function in BIRCH allows for the concise summarization of a cluster, hence circumventing the need to retain intricate details pertaining to specific objects or points. Instead, just a constant amount of space is required to preserve the clustering features. This efficiency in space utilization is a key advantage of BIRCH. Additionally, clustering features are additive, meaning that the clustering features of the combined clusters structured by integrating two disjoint clusters (C1 and C2) can be extracted from the clustering features of the individual clusters. This property simplifies the computation of clustering features during hierarchical clustering and contributes to the scalability and efficiency of the BIRCH algorithm. The clustering among ground monitoring stations is shown in fig 4 and the pseudocode of BIRCH is described given below Table 2.

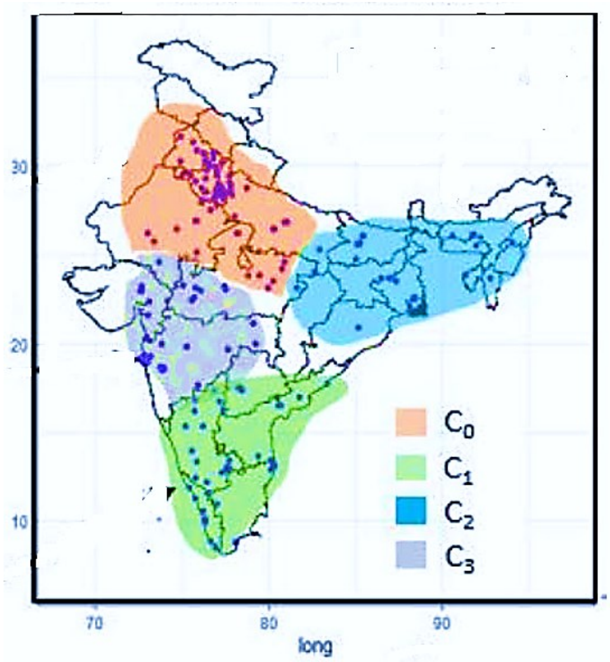



**Figure 4.** Clusters based on BIRCH

**Table 2.** Pseudocode of BIRCH for PM concentration 2.5 and 10 clustering

```
Initialize BIRCH algorithm parameters:
- max nodes per cluster
- max leaf entries
- clustering radius
initialize BIRCH tree structure:
- root = empty node
function insert data point(pm data point):
    current node = root
    while current node is not leaf:
        if pm data point is within current node's clustering radius:
            for each child node in current node:
                if pm data point is within child node's clustering radius:
                    current node = child node
                    break
    if current node is leaf and current node has room for more entries:
        add pm data point to current node
    else:
        split node(current node)
        insert data point(pm data point) # recursively insert pm data point into newly split node
function split node(node):
    if node has reached max leaf entries:
        split node into subclusters using a k-means clustering algorithm
        create new child nodes for each subcluster
        redistribute PM data points among child nodes
        update parent node's summary information
        if parent node has reached max nodes per cluster:
            split node(parent node)  # recursively split parent node if necessary
```

```
function merge clusters():
    recursively merge subclusters within each internal node
    update parent nodes' summary information
    if parent node's children are all leaf nodes:
        merge leaf clusters(parent node)
function merge leaf clusters(node):
    if node has multiple leaf children:
        combine PM data points from all leaf children
        apply k-means clustering algorithm to form new leaf clusters
        update parent node's summary information
        if parent node has reached max nodes per cluster:
            split node(parent node)  # recursively split parent node if necessary
```

### *3.5. PM Prediction using Deep Maxout Networks*

The clustered data is incorporated as additional features or contextual factors into air quality prediction models, then Deep Maxout Networks (Ramkumar et al. 2022) is used for PM prediction, that leverage both historical data and cluster information to predict future air pollution levels. In a typical DMN architecture, each layer consists of multiple units, and each unit computes a linear transformation followed by an activation function. Replace the traditional activation functions (e.g., ReLU, sigmoid) with spline activation functions in one or more layers of the DMN.

The syntax for the activation function of maxout is as follows: When provided with an input $\mathbf{p} \in \mathbb{R}^d$, where *p* represents the input vector or the state of a hidden layer, let us consider the number of linear sub-units merged by a maxout activation (also known as maxout rank) as *R* and $R \ll d$. In this scenario, a maxout activation first calculates $R$ linear feature mappings $\mathrm{q} \in \mathbb{R}^d$.

$$q_i = wt_i^{\mathrm{T}}\mathrm{p} + b_i, \mathrm{wt}_i \in \mathbb{R}^d \; i \in [R] \quad (6)$$

Where $\mathrm{wt}_i$ is a weight vector associated with the linear transformation. It has the same dimensionality as the input vector p, T is input vector and $b_i$ is the bias term associated with the linear transformation. Subsequently, the resultant value of the maxout hidden unit, denoted as $h_{mt}$, is provided as the highest value among the *R* feature mappings.

$$h_{mt}(\mathrm{p}) = \max\{q_i R\}_{i=1}^{R}. \quad (7)$$

Thus, suppose $wt_i'$ are independent linearly, the activation function of maxout could be viewed as carrying out a pooling operation across an input space of *R* dimensions.

$$\mathcal{A} = b + \mathcal{W}, \quad (8)$$

This describes the computation of the matrix $\mathcal{A}$ by adding the bias vector b to each column of the weight matrix $\mathcal{W}$. The activation function of cross-channel max pooling selects the maximum output across different channels or feature maps, which is then forwarded to the next layer. The maxout activation distinguishes itself from standard activation units by its distinctive structure, which often functions inside a one-dimensional linear space. In the context of a fully-connected deep maxout network with *l* hidden layers, let us consider the $l_{th}$ layer. If the $l_{th}$ layer consists of $N_l$ hidden units with a maxout rank *R*, the output of the $l_{th}$ layer, denoted as $O^l$, can be expressed as follows:

$$O^{l} = \left\{h_{mt}^{i,j}(O^{l-1})\right\}_{j=1}^{N_{l-1}}, \tag{9}$$

where the superscript $l, j$ of $h_{mt}^{i,j}$ represents the $j_{th}$ maxout unit in the $l_{th}$ layer, and $h_{mt}^{i,j}$ has the structure defined in (6) and (7) with (6) adapted to $O^{l-1}$:

$$q_{i}^{l,j} = \left(\mathrm{wt}_{i}^{l,j}\right)B^{T}\mathrm{x} + \mathrm{b}_{i}^{l,j}, \quad \mathrm{wt}_{i}^{l,j} \in \mathbb{R}^{N_{l-1}},\ i \in [R]. \tag{10}$$

Where $B$ is spline activation function.

*3.5.1. Spline Activation Function for DMN*

This process maps the linear transformation output to the desired non-linear response. The spline function can be constructed using piecewise polynomial functions, such as cubic splines, with knots defining the transition points between segments. The parameters of the spline function, including the coefficients of the polynomials and the positions of the knots, can be learned during training using backpropagation. The present study focuses on the treatment of SAFs, specifically examining the simplest scenario with a single neuron possessing a flexible AF. The computation of the outcome of the SAF is performed based on a generic input $p \in X^{D}$ using the following equation.

$$sp = wt^{T}p, \tag{11}$$

$$q = \varphi(sp; \mathrm{R}), \tag{12}$$

The eventual bias term, denoted as $wt \in X^{D}$, is immediately included into the input vector. The $AF\varphi(\cdot)$ is then characterized by a vector $q \in X^{Q}$, which consists of internal parameters known as knots. The knots in the dataset reflect a subset of the $AF$ value over $Q$ points that encompass the whole function. Specifically, this approach assumes that the knots are evenly distributed, with a constant value of $\Delta p \in X$, and symmetrically distributed about the origin. The output is calculated by performing spline interpolation on the nearest knot and its $Pn$ nearest neighbors, given the value of $sp$. The often-employed value of $Pn = 3$, as utilized in this study, aligns with cubic interpolation. This decision is widely regarded as a desirable compromise between the localization of the output and the precision of interpolation. The normalized the margin value among $q_i$ and $q_{i+1}$ could be defined based on the index $i$ of the nearest loop.

$$u = \frac{sp}{\Delta p} - \left\lfloor \frac{sp}{\Delta p} \right\rfloor. \tag{13}$$

The floor operator was denoted as $\left\lfloor \frac{sp}{\Delta p} \right\rfloor$. The normalized reference vector could be computed from $u$, while the necessary control points can be extracted from $i$ and referred to as the $i$th span in the vector $q_i$. The result (12) is then calculated as:

$$q = \varphi(sp) = \mathrm{u}^{T}\mathrm{B}\mathrm{q}_{i}, \tag{14}$$

The spline basis matrix is denoted as $B \in X^{(Pn+1)\times(Pn+1)}$. The Catmull-Rom Spline (CRS) with a value of $Pn = 3$ is employed in this research, as presented below.

$$B = \frac{1}{2}\begin{bmatrix} -1 & 3 & -3 & 1 \\ 2 & -5 & 4 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & 2 & 0 & 0 \end{bmatrix} \tag{15}$$

This study examines the scenario of a single hidden layer neural network, where the input is of size *D*, the hidden layer consists of *H* neurons, and the output neurons are of dimension *O*. Each individual neuron inside the network employs a SAF that include distinct adaptive control points, which are established autonomously throughout the training procedure. To facilitate computational efficiency, let us assume that the sample set of the splines remains consistent across all neurons, and that they possess a singular common basis matrix *B*. DMN architectures, when combined with spline activation functions, can effectively reduce the dimensionality of the input data while preserving important features. This can lead to more efficient processing and training, especially in scenarios with high-dimensional data.

## 4. Results and Discussion

In this section, the proposed DMN is predicted the PM concentration 2.5 and 10 and the performances are evaluated and compared with existing schemes like HGNN, INNGNN, CEEMD-LSTM in terms of RMSE, Coefficient of Determinations ($R^2$), and Mean Absolute Error (MAE). In the experimental setup, the hyperparameters of the model are set as follows: The training procedure consists of 1,000 iterations, each utilizing a batch size of 128 and a rejection rate of 0.1. The architecture incorporates GRU (Gated Recurrent Unit) layers with 64 hidden units, while the input data comprises sequences of eight long-term historical data points. Training is conducted using the Adam optimizer, with the Mean Square Error (MSE) serving as the designated loss function. The primary objective is to minimize the MSE across the training iterations, thereby optimizing the predictive performance of the model.

The overall performance of proposed scheme in terms of RMSE, MAE and $R^2$ are depicted in table 2. It shows the performance numeric evaluation and compared with existing schemes numeric values. Its show the proposed DMN attained better performance results compared than exiting schemes. The proposed DMN models are relatively easier to train and tune compared to complex convolutional architectures like HGNN, INNGNN, CEEMD-LSTM. DMNs may offer advantages in terms of flexibility, efficiency in parameter learning, interpretability, and scalability, especially for tasks like PM concentration prediction. DMNs, with their maxout activation functions, offer flexibility in capturing complex non-linear relationships in the data. This flexibility allows DMNs to adapt well to various types of data and tasks, including PM concentration prediction. Due to this, the proposed DMN attained high results compared than others.

**Table 3.** Overall Performance Comparison among PM Concentration Prediction Schemes

| **Methods** | **RMSE ($\mu g/m^3$)** | | **MAE($\mu g/m^3$)** | | **$R^2$** | |
|---|---|---|---|---|---|---|
| | **PM 2.5** | **PM 10** | **PM 2.5** | **PM 10** | **PM 2.5** | **PM 10** |
| Proposed DMN | 10.211 | 10.321 | 5.641 | 5.764 | 0.976 | 0.945 |
| HGNN | 11.971 | 12.232 | 6.938 | 7.24 | 0.828 | 0.834 |
| INNGNN | 12.659 | 12.896 | 7.296 | 7.542 | 0.762 | 0.745 |
| CEEMD-LSTM | 13.536 | 13.876 | 7.781 | 8.122 | 0.668 | 0.675 |

### *4.1. RMSE Performance Comparison*

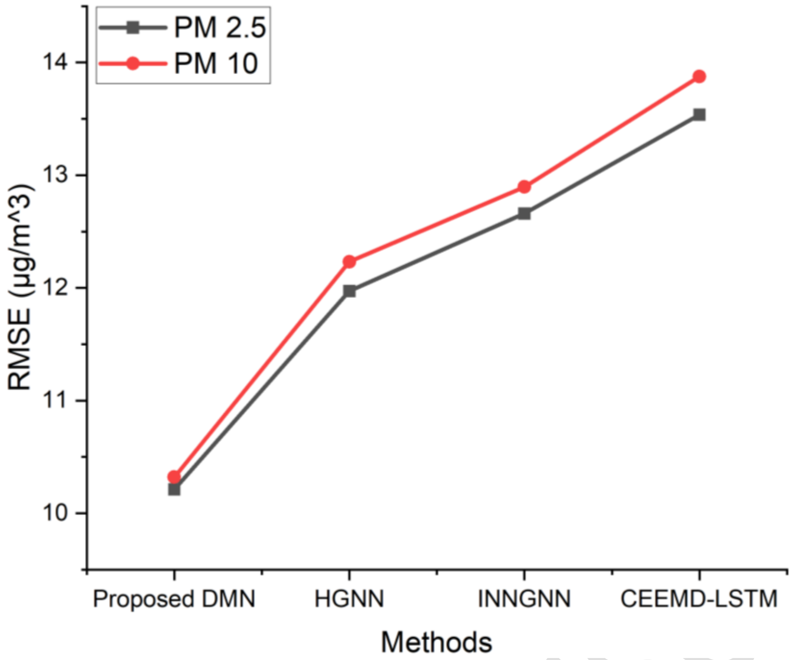



**Figure 5.** RMSE performance comparison among PM concentration schemes

Fig 5 shows the RMSE performance comparison among proposed DMN model and compared with existing PM concentration prediction schemes like CEEMD-LSTM, INNGNN, HGNN and Proposed DMN. Its shows the RMSE of proposed and existing schemes for PM2.5 and PM10, and the outputs show that the proposed scheme attained less RMSE compared than others. The proposed DMN is designed to handle sequential data with varying lengths and time lags. The proposed Deep Maxout Network (DMN) architecture, complemented by Polynomial and Spline Interpolation methods, offers a revolutionary solution for tackling challenges in Matter 2.5 and 10 concentration air monitoring systems. Through meticulous data collection and preprocessing, followed by feature extraction and selection using mRMR and feature clustering via BIRCH, the model optimizes the dataset for robust analysis. Leveraging its innovative design, the DMN adeptly captures complex patterns and relationships within the data, resulting in significantly improved accuracy. Evaluation using RMSE showcases the model's superiority, with PM2.5 and PM10 predictions exhibiting RMSE results of 10.2111 μg/m³ and 10.321 μg/m³, respectively. These impressive results underscore the efficacy of the proposed approach in enhancing air quality measurements and informing environmental management decisions.

*4.2. MAE Performance Comparison*

Fig 6 shows the MAE performance comparison among proposed DMN model and compared with existing PM concentration prediction schemes like CEEMD-LSTM, INNGNN, HGNN and Proposed DMN. Its shows the MAE of proposed and existing schemes for PM2.5 and PM10, and the outputs show that the proposed scheme attained less MAE compared than others. The proposed DMN architecture, augmented by Polynomial and Spline Interpolation methods, presents a transformative solution to address challenges in Matter 2.5 and 10 concentration air monitoring systems. Through meticulous data preprocessing, feature extraction, and clustering using mRMR and BIRCH, the model optimizes the dataset for accurate analysis. Leveraging its innovative design, the DMN effectively captures intricate patterns in the data, resulting in significantly improved predictions. Evaluation using Mean Absolute Error (MAE) reveals remarkable performance, with PM 2.5 and PM 10 predictions exhibiting MAE values of 5.641 and 5.764 μg/m³. These outcomes underscore the robustness and effectiveness of the proposed approach in enhancing air quality measurements and facilitating informed decision-making in environmental management.

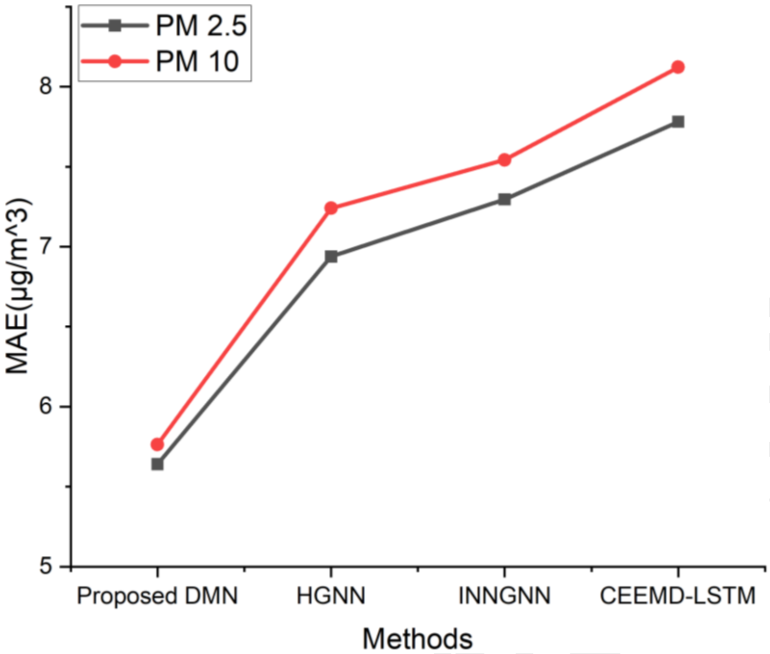



**Figure 6.** MAE performance comparison among PM concentration schemes

### *4.3. $R^2$ Performance Comparison*

Fig 7 shows the $R^2$ performance comparison among proposed DMN model and compared with existing PM concentration prediction schemes like CEEMD-LSTM, INNGNN, HGNN and Proposed DMN. Its shows the MAE of proposed and existing schemes for PM2.5 and PM10, and the outputs show that the proposed scheme attained high $R^2$ compared than others. The proposed DMN architecture, enriched with Polynomial and Spline Interpolation methods, presents a pioneering solution to address the complexities of Matter 2.5 and 10 concentration air monitoring. By meticulously preprocessing data, extracting features, and employing clustering techniques like mRMR and BIRCH, the model optimizes dataset representation for accurate analysis. With its innovative design, the DMN adeptly captures intricate data patterns, resulting in exceptional predictive performance. The high coefficients of determination ($R^2$) of 0.976 for PM 2.5 and 0.945 for PM 10 underscore the model's remarkable ability to explain variability in the data, affirming its effectiveness in enhancing air quality measurements and enabling informed decision-making in environmental management.

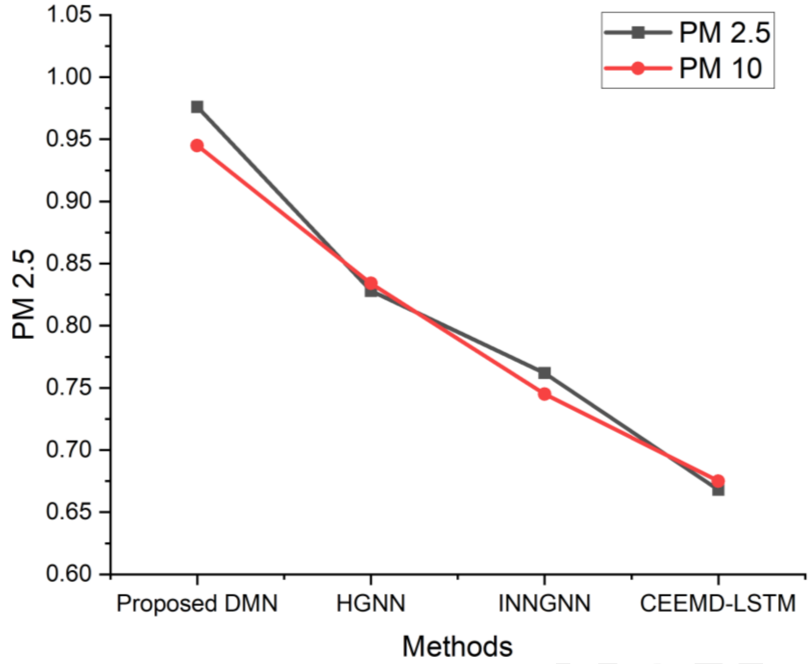



**Figure 7.** $R^2$ Performance comparison among PM concentration schemes

The evaluation using key metrics demonstrates compelling results, with a RMSE of 10.211µg/m³ for PM2.5 and 10.321µg/m³ for PM10, a MAE of 5.641 µg/m³ for PM2.5 and 5.764 µg/m³ for PM10, and high $R^2$ of 0.976 for PM2.5 and 0.945 for PM10. These values underscore the robustness and effectiveness of the proposed approach in enhancing air quality measurements. Furthermore, they emphasize its potential to facilitate informed decision-making in environmental management and advance public health initiatives.

## 5. Conclusion

This research presents a new model to address challenges in PM2.5 and PM10 concentration air monitoring systems by proposing an innovative DMN architecture enhanced with Polynomial and Spline Interpolation methods. By effectively handling discontinuities in data sequences and smoothing transition fitting curves at interval junctions, the proposed model generates an ideal dataset, thereby improving the robustness and accuracy of air quality measurements. Through a systematic process of data collection, preprocessing, feature extraction and selection using minimum redundancy maximum relevance (mRMR), and clustering of similar features using the BIRCH scheme, the paper demonstrates the effectiveness of the proposed approach in enhancing the reliability of PM2.5 and PM10 concentration monitoring systems using Air Quality Data in India from Kaggle. Through meticulous data preprocessing, feature extraction, and clustering techniques such as mRMR and BIRCH, the model optimizes dataset representation for accurate analysis. Leveraging its innovative design, the DMN effectively captures intricate data patterns, resulting in exceptional predictive performance. Evaluation using key metrics reveals compelling results: a Root Mean Square Error (RMSE) of 10.211µg/m³ for PM2.5 and 10.321µg/m³ for PM10, a Mean Absolute Error (MAE) of 5.641 µg/m³ for PM2.5 and 5.764 µg/m³ for PM10, and high coefficients of determination ($R^2$) of 0.976 for PM2.5 and 0.945 for PM10. These values underscore the robustness and effectiveness of the proposed approach in enhancing air quality measurements, facilitating informed decision-making in environmental management, and advancing public health initiatives. Several avenues can be explored in future scope to enhance the proposed methodology further and its practical implications for air quality management and public health initiatives. Firstly, incorporating real-time data streams and sensor fusion techniques could enhance the timeliness and accuracy of air quality measurements. Furthermore, extending the scope of the study to include additional pollutants and considering spatial-temporal variations could provide a more comprehensive understanding of air quality dynamics.

## REFERENCES


Cheng W., Shen Y., Zhu Y. and Huang L. (2018), A neural attention model for urban air quality inference: Learning the weights of monitoring stations, *In Proceedings of the AAAI Conference on Artificial Intelligence*, **32**, 1.

Chiang P.W. and Horng S.J. (2021), Hybrid time-series framework for daily-based PM 2.5 forecasting, *IEEE Access*, **9**.

De Mesnard L. (2013), Pollution models and inverse distance weighting: Some critical remarks, *Computers & Geosciences*, **52**.

Di Q., Wang Y., Zanobetti A., Wang Y., Koutrakis P., Choirat C., Dominici F. and Schwartz J.D. (2017), Air pollution and mortality in the Medicare population, *New England Journal of Medicine*, **376**, 26.

Ding H. and Noh G. (2023), A Hybrid Model for Spatiotemporal Air Quality Prediction Based on Interpretable Neural Networks and a Graph Neural Network, *Atmosphere*, **14**, 1807.

Ejurothu P.S.S., Mandal S. and Thakur M. (2023), Forecasting PM2.5 concentration in India using a cluster-based hybrid graph neural network approach, *Asia-Pacific Journal of Atmospheric Sciences*, **59**, 5.

Fiordelisi A., Piscitelli P., Trimarco B., Coscioni E., Laccarino G. and Sorriento D. (2017), The mechanisms of air pollution and particulate matter in cardiovascular diseases, *Heart Failure Reviews*, 22.

Gunasekar S., Joselin Retna Kumar G. and Dileep Kumar Y. (2022), Sustainable optimized LSTM-based intelligent system for air quality prediction in Chennai, *Acta Geophysica*, **70**, 6.

Heft-Neal S., Burney J., Bendavid E. and Burke M. (2018), Robust relationship between air quality and infant mortality in Africa, *Nature*, **559**, 7713.

Jasarevic T., Thomas, G. and Osseiran, N. (2021), 7 million Deaths Annually Linked to Air Pollution, World Health Organization, *Technical Report*, **7**.

Li F.X., Zhu Y., Hou J., Jin L. and Wang J. (2015), Artificial neural network forecasting of PM2. 5 pollution using air mass trajectory based geographic model and wavelet transformation, *Atmospheric Environment*, **107**, 118e128.

Li D., Liu J. and Zhao Y. (2022), Forecasting of PM2.5 concentration in Beijing using hybrid deep learning framework based on attention mechanism, *Applied Sciences*, **12**, 11155.

Liu X., Qin M., He Y., Mi X. and Yu C. (2021), A new multi-data-driven spatiotemporal PM2.5 forecasting model based on an ensemble graph reinforcement learning convolutional network, *Atmospheric Pollution Research*, **12**, 101197.

Liu H. and Deng D.H. (2022), An enhanced hybrid ensemble deep learning approach for forecasting daily PM2.5, *Journal of Central South University*, **29**, 6.

Lorbeer B., Kosareva A., Deva B., Softić D., Ruppel P. and Küpper A. (2018), Variations on the clustering algorithm BIRCH, *Big Data Research*, **11**.

Mao Y.S., Lee S.J., Wu C.H., Hou C.L., Ouyang C.S. and Liu C.F. (2023), A hybrid deep learning network for forecasting air pollutant concentrations, *Applied Intelligence*, **53**, 10.

Méndez M., Merayo M.G. and Núñez M. (2023), Machine learning algorithms to forecast air quality: a survey, *Artificial Intelligence Review*, **56**, 9.

Mohan A.S. and Abraham L. (2023), An ensemble deep learning model for forecasting hourly PM2. 5 concentrations, *IETE Journal of Research*, **69**, 10.

Radovic M., Ghalwash M., Filipovic N. and Obradovic Z. (2017), Minimum redundancy maximum relevance feature selection approach for temporal gene expression data, *BMC Bioinformatics*, **18**.

Ramkumar M.P., Mano Paul P.D., Maram B. and Ananth J.P. (2022), Deep maxout network for lung cancer detection using optimization algorithm in smart Internet of Things, *Concurrency and Computation: Practice and Experience*, **34**, e7264.

Wu S., Sun Y., Bai R., Jiang X., Jin C. and Xue Y. (2024), Estimation of PM2. 5 and PM10 Mass Concentrations in Beijing Using Gaofen-1 Data at 100 m Resolution, *Remote Sensing*, **16**, 604.

Xue X. and Chen J. (2020), Optimizing sensor ontology alignment through compact co-firefly algorithm, *Sensors*, **20**, 2056.

Xue X. and Chen J. (2019), Using compact evolutionary tabu search algorithm for matching sensor ontologies, *Swarm and Evolutionary Computation*, **48**.

Zhang Q., Han Y., Li V.O. and Lam J.C. (2022), Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities, *IEEE Access*, **10**.

Zhang L., Xu L., Jiang M. and He P. (2023), A novel hybrid ensemble model for hourly PM2. 5 concentration forecasting, *International Journal of Environmental Science and Technology*, **20**, 1.