International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Neural Networks for Air Pollution Forecasting

[1]Vidyut Singhai, [2]Dr. Nidhi Sethi

Submitted: 18/06/2024 **Revised**: 22/07/2024 **Accepted**: 21/08/2024

Abstract - This study explores the use of deep learning models for forecasting air pollution, specifically PM2.5 levels, using data from the Central Pollution Control Board (CPCB). The methodology involves extensive data preprocessing, including trend identification, missing value handling, and the extraction of temporal features to capture seasonal variations. The models evaluated include standalone LSTM, Hybrid LSTM-1D CNN, Hybrid LSTM-GRU, and Hybrid GRU-1D CNN. An 80:20 training-testing split was used, with feature extraction methods applied to enhance predictive accuracy. The performance of the models was assessed using common metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), Mean Absolute Error (MAE), and R-squared (R2). The Hybrid GRU-1D CNN model demonstrated the best performance, achieving the lowest MSE (201.2), RMSE (14.1), and MAE (6.7), with a high R2 value of 0.99. These results highlight the potential of hybrid deep learning models for accurate air pollution prediction, offering valuable insights for environmental monitoring and policymaking.

Keywords- *Air Pollution, Forecasting, Deep Learning, Prediction, Hybrid Models*

1. Introduction

Concerns over the damaging effects of deteriorating air quality on ecosystems, human health, and climate change have led to a critical shift in the focus of research and development towards detection of air pollution. Pollutant emissions, including particulates, nitrogen oxides, oxides of sulfur, carbon monoxide, and volatile organic compounds, have increased as industrialization and urbanization continue to grow. necessitating advanced methods for monitoring and managing air quality. Traditional air quality monitoring systems typically involve fixed sensor networks strategically placed in specific locations[1]. While these systems provide valuable data, they often exhibit limitations in terms of coverage, granularity, and real-time responsiveness. Sophisticated technologies have sparked a paradigm change towards more dynamic and all-encompassing air pollution monitoring tactics, especially in the areas for sensor networks, satellite imaging, and data analytics. With pollutants like particulate matter (PM2.5, PM10), nitrogen oxides (NOx), sulphur dioxide (SO2), carbon monoxide (CO), and ozone (O3) profoundly affecting climate patterns, human health, and general ecological balance, air pollution remains one of the most urgent environmental challenges of the twenty-first century. Particularly in heavily inhabited urban regions where industrial emissions, vehicle traffic, and meteorological circumstances contribute to high pollution levels, accurate prediction of air pollution is essential for proactive policy decisions, emergency response preparation, and public health protection[2]–[4].

1Ph.D Research Scholar, Department of Computer Science,
Renaissance University, Indore (M.P.)
2Assistant professor, Faculty of Computer Science, Renaissance College of Commerce & Management, Indore (M.P.)

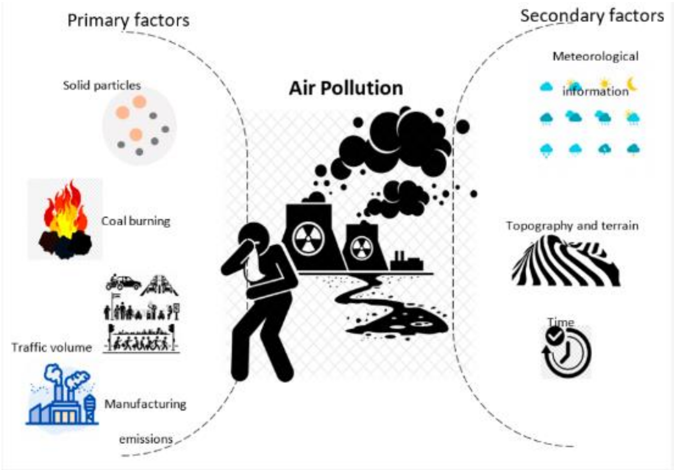

Figure 1 Air pollution Factors[5]

Widely used are traditional air quality prediction models including statistical approaches including autoregressive integrated moving average (ARIMA), multiple linear regression (MLR), and Gaussian dispersion models, together with physics-based numerical simulations including the Weather Research and Forecasting with Chemistry (WRF-Chem) model and Community Multiscale Air Quality (CMAQ) model. These models find it difficult, meantime, to adequately capture the extremely nonlinear, complicated relationships among pollution sources, atmospheric conditions, and geographical factors[6]–[8]. By using hierarchical feature learning, end-to--end modelling, and large-scale data processing, deep neural networks (DNNs) have become a transforming tool for enhancing air quality forecasting accuracy. Convolutional neural networks (CNNs), which are effective in spatial feature extraction from remote sensing images and sensor network grids; recurrent neural networks (RNNs) and their advanced variants, long short-term memory (LSTM) networks and gated recurrent units (GRUs), which are especially suited for modelling temporal dependencies in sequential pollution data; and transformer-based models, which provide enhanced contextual understanding of time-series dependencies through self-attention mechanisms. Improved predictive accuracy follows from hybrid models that combine CNNs with LSTMs or GRUs incorporating both spatial and temporal correlations in air quality data. Furthermore investigated to improve air pollution prediction include attention mechanisms, self-supervised learning, and graph neural networks (GNNs), thereby enabling models to dynamically prioritise pertinent environmental information and link many pollution sources across cities and areas. By allowing constant, real-time monitoring and forecasting at fine geographical resolutions, the combination of deep learning with the Internet of Things (IoT), real-time sensor networks, and edge computing significantly improves predictive capabilities. Combining ground-based sensor readings with satellite imagery from sources such NASA's MODIS, Sentinel-5P, and Copernicus Atmospheric Monitoring Service (CAMS), data fusion methods help to enrich training datasets and increase model generalisability. While hyperparameter optimisation strategies like Bayesian optimisation, genetic algorithms, and reinforcement learning-based tuning improve model performance, preprocessing techniques including normalising, outlier detection, and missing value imputation are absolutely vital in ensuring data quality before feeding it into deep learning models. With methods including Shapley Additive Explanations (SHAP), Layer-wise Relevance Propagation (LRP), and Grad-CAM (gradient-weighted Class Activation Mapping), explainability and interpretability in deep learning-based air pollution forecasting remain vital for ensuring transparency and trust. Transfer learning and domain adaption techniques help models trained on air quality data from one city or region to generalise across many geographical areas, hence reducing the requirement for intensive retraining on new datasets. Promising developments in physically consistent, interpretable air pollution predictions come from physics-informed neural networks (PINNs) and hybrid artificial intelligence-physics models, which combine deep learning with atmospheric chemistry simulations [9]–[11]. Notwithstanding these developments, several problems still exist: data shortage in underdeveloped areas, sensor errors causing noisy measurements, high computational costs related to training deep networks, and adversarial robustness issues whereby small perturbations in input data can cause major predictions errors. To guarantee appropriate AI deployment, ethical issues include data privacy in air quality monitoring systems, algorithmic fairness in exposure assessments, and biases in model projections that disproportionately affect vulnerable populations must also be taken under consideration. New trends like neuromorphic computing, which seeks to create energy-efficient AI hardware, and federated learning, in which several distributed devices cooperatively train models without sharing raw data, present interesting paths for future enhancement of air pollution forecasting capability. These developments are being used by governments, environmental agencies, and researchers to create better, AI-driven environmental policies capable of reducing pollution-related health hazards and thereby supporting sustainable urban design. By enabling more accurate, real-time, scalable predictions, spatial-temporal data sources, and advanced AI technologies including transformers, GNNs, and hybrid models, so paving the path for data-driven air quality management, early warning systems, and global efforts towards cleaner, healthier environments. Deep neural networks have thus transformed air pollution forecasting[12].

1.1 Establishing Context and Significance

Air pollution is a global challenge affecting public health and the environment, necessitating accurate forecasting for mitigation. Traditional models struggle with the nonlinear and dynamic nature of pollution, making deep learning techniques such as long short-term memory (LSTM), 1D convolutional neural networks (1D-CNN), and hybrid LSTM-GRU models more effective. LSTMs capture long-term dependencies in pollutant time series, while 1D-CNNs extract local temporal patterns efficiently. Hybrid LSTM-GRU models combine LSTM's sequential learning capabilities with GRU's computational efficiency, improving predictive accuracy[13], [14]. These models benefit from incorporating meteorological data, traffic patterns, and industrial activity to enhance their contextual understanding of pollution trends. Integrating deep learning with IoT-based sensor networks enables real-time air quality monitoring, while edge computing ensures low-latency predictions. Transfer learning helps adapt models to data-scarce regions, improving forecasting reliability. Despite challenges such as missing data and sensor noise, robust architectures and hyperparameter optimization techniques enhance model stability. Explainability methods like SHAP and attention mechanisms make predictions interpretable for policymakers. As AI-driven air quality management gains traction, LSTM-based architectures are emerging as essential tools for high-accuracy pollution forecasting, enabling proactive interventions and data-driven urban planning for healthier environments[15].

2. Literature Review

Kecorius 2024 et al. Urban air pollution, driven by pollutants like nitrogen oxides, ozone, and particulate matter, poses significant health risks. Machine learning, particularly deep learning, has shown potential in predicting pollutant concentrations but suffers from limited interpretability. the Temporal Selection Layer (TSL) technique within deep learning models for time series forecasting, improving both prediction accuracy and interpretability. Applied to hourly pollution data from Graz, Austria, the method enhances model transparency by embedding feature selection directly into the neural network. The results demonstrate that TSL improves model effectiveness, reducing computational costs, and leads to better air pollution management strategies[16].

Rautela 2024 et al. India's fight against air pollution requires a comprehensive approach combining advanced technology, robust regulations, and societal engagement. This study evaluates PM2.5 levels using key aerosols like black carbon, dust, and sulphates, identifying vulnerable regions such as the Indo-Gangetic Plains and Western India. An AI&ML-based convolutional autoencoder model achieved high accuracy in forecasting PM2.5 concentrations, with metrics like Structural Similarity Index over 0.60 and Mean Square Error below 10 μg/m³. Despite technological advancements, regulatory challenges persist. Addressing these demands tailored regional strategies, AI&ML integration, strengthened frameworks, sustainable practices, and global cooperation to mitigate air pollution effectively across India[17].

Fareena 2023 et al. Rapid urbanization and industrialization lead to increased air pollution, affecting health and the environment, making it a public health emergency. Accurate air pollution forecasting is crucial for timely action. Compares deep learning-based models (LSTM, GRU) and statistical models for forecasting five pollutants: NO2, O3, SO2, PM2.5, and PM10. Using a dataset from an air quality monitoring station in Belfast, Northern Ireland, the models are evaluated with RMSE, MAE, and R^2. Results show that deep learning models outperform statistical models, achieving the lowest RMSE of 0.59 and the highest R^2 of 0.856[18].

Zhu 2023 et al. Rapid urbanization worsens water and air quality through contaminants and competition for resources. This paper applies deep learning methods, specifically Convolutional Neural Networks (CNN) and Long-Short Term Memory (LSTM), to classify water quality and assess air quality in urban development. The proposed Conv.LSTM model captures both spatial and temporal dependencies, outperforming traditional models. Performance metrics such as accuracy, recall, precision, and F1-score show that Conv.LSTM achieved 92% accuracy for water pollution classification and 91% for air pollution, surpassing RNN (65%), DBN (78%), and LSTM (82%) in both datasets. Reducing pollution is essential for sustainable urban growth[19].

Guv 2022 et al. Air pollution prediction is crucial for public health, but traditional models often overlook peak value accuracy and lack interpretability. Presents a new Hybrid Interpretable Predictive Machine Learning model for Particulate Matter 2.5 (PM2.5) prediction, combining a deep neural network with a Nonlinear Auto Regressive Moving Average with Exogenous Input model. The model integrates automatic feature generation and selection. Experimental results show the model outperforms others in peak value prediction accuracy and interpretability, with correlation coefficients of 0.9870, 0.9332, and 0.8587 for 1, 3, and 6-hour predictions, respectively, offering an interpretable framework for time-series data[20].

Table.1 Literature Summary

Authors/year	Model/method	Research gap	Findings	References
Abdelkader 2021	IMDA-VAE model enhances pollution forecasting accuracy.	Lack of effective, interpretable, and efficient air pollution forecasting models.	The IMDA-VAE model outperforms traditional models in air pollution forecasting.	[21]
Hähnel 2020	Deep-learning framework for air-pollution forecasting.	Limited deep learning application beyond traditional PDE solver domains.	Deep learning improves air-pollution forecasting, reducing run-time significantly.	[22]
Mohammad 2020	Air quality prediction model.	Need for optimized, multi-stage deep learning models for air prediction.	SAQPM successfully predicts air pollutants using LSTM and PSO optimization.	[23]
Ameer 2019	Comparative regression models for air quality prediction.	Lack of comparative analysis of regression techniques for air quality.	Regression models' performance varies based on error rates and processing time.	[24]

Serrano 2024	Forecasting air pollutant concentrations using Convolutional Long Short-Term Memory networks.	Lack of comprehensive air quality forecasting frameworks for multi-pollutant analysis.	Comprehensive AI framework for multi-pollutant analysis versus existing methods	[25]
Tao 2023	Machine learning models predict PM2.5 concentration using meteorological and soil data.	Assessing machine learning models for high-resolution PM2.5 prediction.	Evaluation of ML models' effectiveness in PM2.5 prediction debated.	[26]

3. Methodology

This work presents the method for deep learning model-based deep learning air pollution detection. First, data on air quality taken from the Central Pollution Control Board (CPCB) guarantees the inclusion of important contaminants including PM2.5, PM10, NOx, SO2, and CO. Trend, missing value, and anomaly identification is accomplished via exploratory data analysis (EDA). Handling missing numbers, normalising pollution concentrations, and extracting temporal information including day, month, and year to capture seasonal fluctuations constitute aspects of data preparation. Training and testing are then split from an 80:20 ratio in the dataset. Predictive accuracy is raised by means of feature extraction methods. LSTM-1D CNN is used in a hybrid deep learning framework including standalone LSTM for sequential dependencies, GRU for computational efficiency, LSTM-GRU model for enhanced long-term dependency handling, and LSTM-1D CNN for spatial-temporal feature learning. Accurate air pollution forecasting is ensured by training and performance evaluation of these models.

3.1 Data Collection

The dataset for this study comprises 18,776 entries with 9 key air quality variables, collected from the open-source Central Pollution Control Board (CPCB) database, focusing on towns in India's North Central Region, primarily Delhi NCR. The data covers major cities, including Sonipat, Panipat, Rohtak, New Delhi, Ghaziabad, and Gurgaon, ensuring diverse environmental conditions and pollution sources. The dataset includes critical air pollutants: PM2.5, PM10, NO, NO2, Ozone, SO2, CO, and NH3, which significantly impact air quality and public health. Each parameter is systematically recorded, ensuring consistency and reliability for analysis. Data collection spans multiple years to capture seasonal and temporal variations, allowing deep learning models to learn long-term patterns. This rich dataset forms the foundation for accurate air pollution forecasting, facilitating advanced preprocessing, feature extraction, and model training for predictive assessments using deep learning techniques such as LSTM, GRU, and hybrid LSTM-1D CNN architectures.

	date	co	no	no2	o3	so2	pm2_5	pm10	nh3
0	2020-11-25 01:00:00	2616.88	2.18	70.60	13.59	38.62	364.61	411.73	28.63
1	2020-11-25 02:00:00	3631.59	23.25	89.11	0.33	54.36	420.96	486.21	41.04
2	2020-11-25 03:00:00	4539.49	52.75	100.08	1.11	68.67	463.68	541.95	49.14
3	2020-11-25 04:00:00	4539.49	50.96	111.04	6.44	78.20	454.81	534.00	48.13
4	2020-11-25 05:00:00	4379.27	42.92	117.90	17.17	87.74	448.14	529.19	46.61

Figure 2 Pandas Dataframe of the dataset

3.2 Data Preprocessing

Data preprocessing is a crucial step in preparing datasets for analysis or Deep learning modeling. It involves cleaning and transforming raw data into a format that makes it suitable for effective analysis and prediction. The process typically includes handling missing values, scaling features, extracting date-based features, creating lagged features, and removing irrelevant or redundant columns.

1. Handling Missing Data

Missing values can occur in datasets due to various reasons such as errors during data collection or entry. It's important to deal with missing data to prevent it from negatively impacting the performance of machine learning models. In this case, the missing values are identified using data.isnull().sum(), which provides a count of missing values in each column. One common method for handling missing data is interpolation, where the missing values are estimated based on the values of nearby data points. Linear interpolation is a simple and commonly used technique, where missing values are filled based on a linear relationship with adjacent points. By using data.interpolate(method='linear', inplace=True), we apply this technique to fill missing values in the dataset.

2. Converting Date to DateTime Format

Many datasets contain date or time-based information, which is often crucial for time-series analysis or feature engineering. In this case, the dataset includes a 'date' column, which is converted into a proper DateTime format using pd.to_datetime(data['date']). This step ensures that the 'date' column is properly recognized as a date, allowing for easy manipulation and extraction of time-related features such as the hour, day, month, and weekday.

3. Setting the Date as Index

For time-series data, it's beneficial to set the date column as the index of the dataset. This step helps with organizing the data chronologically, making it easier to analyze and perform time-based operations. Using data.set_index('date', inplace=True), we set the 'date' column as the index of the dataset, which facilitates time-based slicing and feature extraction.

4. Scaling Data

Feature scaling is an important preprocessing step, especially when the dataset includes numerical features with different ranges. Machine learning algorithms may perform poorly if features are on different scales, as they can give undue importance to larger numerical values. One common technique for scaling numerical features is MinMax scaling, which normalizes the data to a range between 0 and 1. This transformation is performed using the MinMaxScaler from sklearn.preprocessing. The scaler.fit_transform(data) method scales all the numerical features in the dataset. Scaling helps in improving the performance of models such as neural networks, k-nearest neighbors, and support vector machines.

5. Extracting Time-Based Features

In many time-series problems, it is important to extract useful features from the date and time information. By adding features like the hour, day, month, and weekday, we can provide the model with additional information that may improve predictive accuracy. These features can help the model capture seasonality, trends, and other time-based patterns in the data. The hour, day, month, and weekday are extracted from the index using data.index.hour, data.index.day, data.index.month, and data.index.weekday respectively, and added as new columns in the dataset.

6. Creating Lagged Features

Lagged features are valuable for time-series forecasting tasks, where the value of a variable at a previous time step may help predict its value at a future time step. In this case, three lagged features are created for each of the columns co, pm2_5, and pm10, using the .shift() method. The shift method moves the data by a specified number of periods (in this case, 1, 2, and 3). By creating lagged features, the model can learn the relationships between past and future values, which is especially useful for forecasting tasks. This step is particularly important for models such as LSTM, and other time-series models.

7. Handling Missing Data After Feature Creation

After creating lagged features, it is possible that new missing values are introduced due to shifting operations. For example, the first few rows for the lagged features will have missing values, as there are no previous data points for them. To remove these rows with NaN values, data.dropna(inplace=True) is used. This step ensures that the dataset does not contain any missing values and is ready for analysis or model training.

Pseudocode for Data Preprocessing

```
# Step 1: Load Data
Load data into a DataFrame
# Step 2: Handle Missing Values
If there are missing values:
    Check the sum of missing values in each column
    Apply linear interpolation to fill missing values
# Step 3: Convert 'date' column to DateTime format
Convert 'date' column to DateTime
Set 'date' as the index of the DataFrame
# Step 4: Scale Features
Initialize MinMaxScaler
For each numerical feature in the dataset:
    Scale the feature using MinMaxScaler
    Store scaled features in a new DataFrame
# Step 5: Extract Time-based Features
For each entry in the 'date' index:
    Extract hour, day, month, weekday from the date
    Create new columns for 'hour', 'day', 'month', 'weekday'
# Step 6: Create Lagged Features
For each lag in the range from 1 to 3:
    Create lagged features for 'co', 'pm2_5', and 'pm10'
    Shift the column by lag periods and store in new columns with names 'co_lag1', 'pm2_5_lag1', etc.
# Step 7: Handle Missing Values After Lagging
Drop rows with any remaining missing values (NaN) due to lagging
# Final Output: Cleaned and Preprocessed Data
Return the cleaned and processed DataFrame
```

3.3 Exploratory Data Analysis

Exploratory data analysis (EDA) is an essential first step in grasping the underlying trends and correlations of the dataset. Regarding the detection of air pollution, several visualisations are applied to evaluate patterns and relationships among pollutants. Observing the temporal variations of important pollutants including PM2.5, NO2, CO, and SO2 across time using a line plot helps one to spot seasonal trends and notable changes in air quality. Visualising the interactions between these pollutants depends on a correlation heatmap, which can be helpful for feature selection and knowledge of collinearity. It shows how closely related they are.

Visualising the distribution and spotting outliers in the pollution levels depends especially on box graphs, which also provide understanding of data variability and extreme values. Furthermore, count graphs or histograms can show the frequency distribution of particular pollution levels, therefore helping to identify skewness or imbalanced data. These visualisations taken together offer a complete awareness of the dataset, which guides next preprocessing and model building.

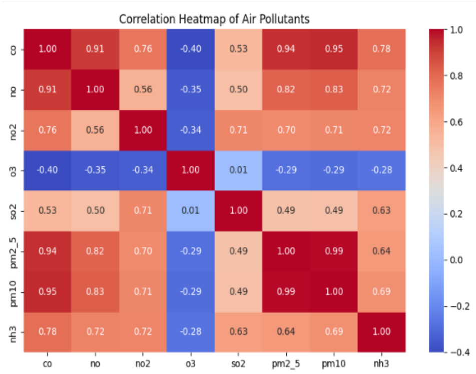

Figure 3 Correlation Heatmap of Air pollutants

This heatmap visually represents the pairwise correlations between various air pollutants such as PM2.5, NO2, CO, and SO2. By using color gradients to depict correlation values, it highlights how strongly different pollutants are related, helping to identify potential multicollinearity and informing feature selection for modeling.

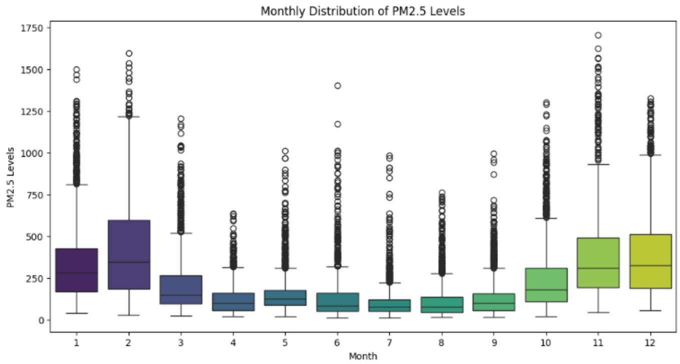

Figure 4 Boxplot of Monthly distribution of PM 2.5 Levels

The boxplot illustrates the spread and distribution of PM2.5 levels across different months. It highlights the median, quartiles, and potential outliers, showing how PM2.5 concentrations vary month-to-month. This helps identify seasonal fluctuations and anomalies in pollution levels over time.

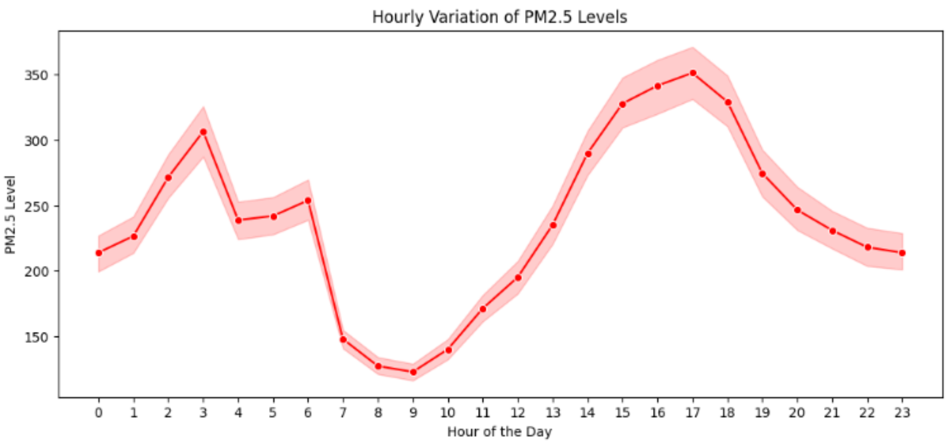

Figure 5 Line plot of hourly variation of PM 2.5 Levels

This line plot tracks the hourly variations in PM2.5 levels. By presenting pollutant data across time intervals, it reveals daily patterns, peak pollution times, and the fluctuation of air quality throughout a typical day, which is important for understanding temporal trends in air pollution.

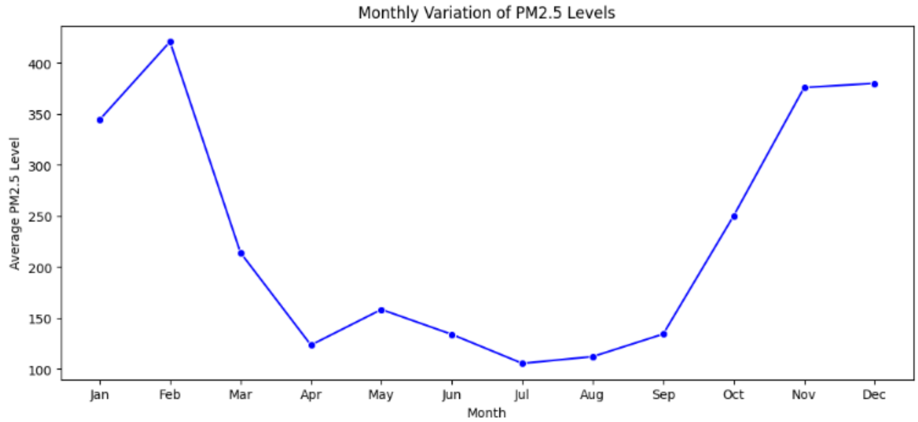

Figure 6 Monthly variation of PM 2.5 Levels

This plot displays the monthly trends in PM2.5 levels. It offers insights into seasonal patterns, identifying months with higher or lower pollution levels, thus providing a clear view of how air quality changes on a longer time scale, often reflecting environmental or climatic factors.

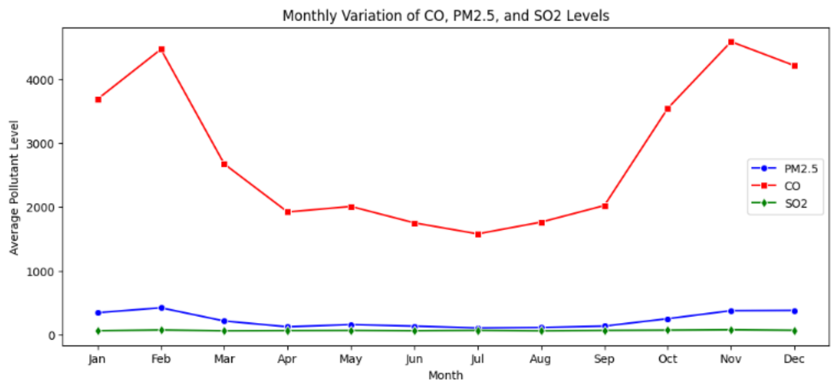

Figure 7 Monthly variation of Co, SO2 and PM 2.5 Levels

This figure compares the monthly trends of CO, SO2, and PM2.5 levels, highlighting how different pollutants behave across the same timeframe. It helps understand how each pollutant fluctuates independently and in relation to others, offering insights into pollution sources and seasonal variations.

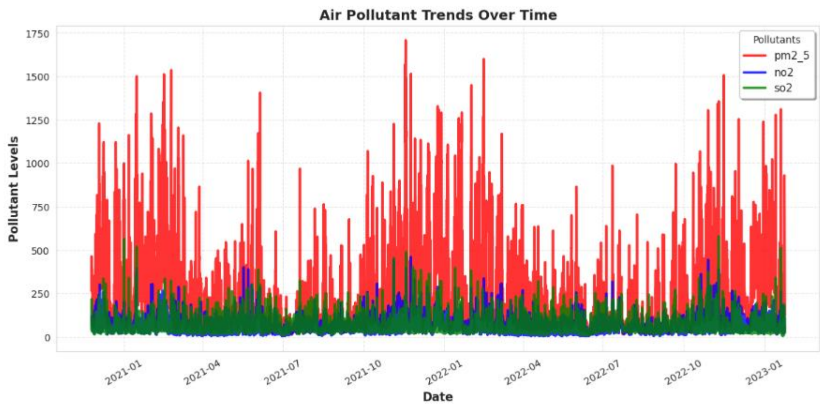

Figure 8 Air pollutants trends over time

This plot visualizes the long-term trends of key air pollutants (PM2.5, NO2, CO, SO2) over time. By tracking these pollutants, the figure reveals how their concentrations have evolved, indicating potential environmental or regulatory changes and providing a clear view of air quality over the study period.

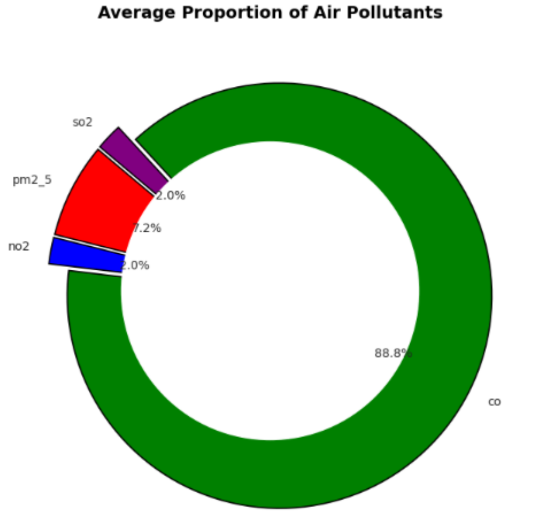

Figure 9 Average Proportion of air pollutants

This figure displays the average proportion of different air pollutants, helping to understand the relative contribution of each pollutant to overall air quality. It highlights the dominant pollutants and their impact on air pollution, providing a basis for focusing on the most critical factors in pollution control efforts.

3.4 Feature Extraction

Feature extraction is a critical step in preparing data for deep learning-based air pollution detection models. Initially, the dataset includes key air quality variables such as PM2.5, PM10, NOx, SO2, and CO, which are crucial for understanding the impact of pollution on public health. Temporal features, including the day, month, and year, are extracted from the 'date' column to capture seasonal and periodic variations in pollution levels. This is particularly important for predicting air quality changes over time, as these temporal patterns influence pollution concentrations. Additionally, lagged features are created by shifting the values of key variables (e.g., PM2.5, PM10) over previous time steps, allowing the model to learn dependencies between past and future pollution levels. These lagged features provide important context for long-term prediction. The data is also normalized using MinMax scaling to ensure all features are on a comparable scale, improving model performance. By carefully extracting and engineering these features, the model can better capture the complex, temporal relationships inherent in air pollution data, enhancing predictive accuracy.

3.5 Data Splitting

Data splitting and reshaping are essential steps in preparing the dataset for training a hybrid deep learning model. In this process, the dataset is typically divided into training and testing sets, often using an 80:20 or 70:30 ratio, ensuring that the model is trained on a substantial portion of the data while retaining enough unseen data for evaluation. The training data is then reshaped to fit the input requirements of the hybrid model, such as an LSTM-1D CNN architecture. This may involve adjusting the data to a 3D structure, where dimensions represent samples, time steps, and features, especially for time-series data. The reshaped data allows the model to learn both temporal dependencies (via LSTM) and spatial-temporal patterns (via CNN) effectively. Furthermore, scaling of features is typically performed to standardize the data within a consistent range, ensuring optimal performance of the model and preventing issues like vanishing or exploding gradients during training. This structured approach ensures robust model performance.

3.6 Deep Learning Model Implementation

The deep learning models for air pollution forecasting are implemented using sequential architectures in Keras. The LSTM model utilizes stacked LSTM layers to capture temporal dependencies in pollutant data. The Hybrid LSTM-1D CNN model combines a convolutional layer with LSTM layers to capture both spatial features and long-term temporal dependencies. The Hybrid LSTM-GRU model merges LSTM and GRU layers to leverage LSTM's ability to handle long-term dependencies with GRU's efficiency. Finally, the Hybrid GRU-1D CNN model combines CNN for spatial feature extraction and GRU for capturing sequential patterns, improving overall forecasting accuracy.

- LSTM Model

The Long Short-Term Memory (LSTM) model is a powerful recurrent neural network (RNN) architecture designed to handle sequential data, making it well-suited for time-series forecasting such as air pollution prediction. In this model, the LSTM layer is used to capture long-term dependencies in the data, enabling it to learn temporal patterns such as trends and seasonal variations. The model consists of two LSTM layers, one with return_sequences=True to preserve the sequential data for further processing and another to output the final prediction. A dropout layer is included to prevent overfitting by randomly dropping a fraction of neurons during training. The final output layer is a dense layer with a single neuron, predicting the PM2.5 levels. The model is compiled using the Adam optimizer, which adapts the learning rate during training, and mean squared error (MSE) as the loss function to minimize prediction errors. This architecture is effective for modeling sequential dependencies but might struggle with spatial-temporal relationships, which is addressed by hybrid models that combine LSTM with other techniques like CNN or GRU.

- Hybrid LSTM-1D CNN Model

The Hybrid LSTM-1D CNN model combines the strengths of both LSTM and 1D Convolutional Neural Networks (CNN) to handle spatial-temporal data in air pollution forecasting. The CNN layer is used initially to capture local spatial patterns in the input data, such as sudden changes in pollutant levels. The 1D convolutional filter learns relevant features from the data, which are then passed to the LSTM layer. The LSTM layer captures long-term temporal dependencies, learning the sequential trends over time, such as seasonal variations in air quality. The model is structured with multiple LSTM layers, where the first LSTM layer retains sequences for further processing, while the second LSTM layer consolidates the learned information for the final prediction. Dropout layers are included to mitigate overfitting, ensuring the model generalizes well to unseen data. The hybrid nature of this model allows it to effectively handle both short-term and long-term dependencies, making it well-suited for complex air pollution forecasting tasks. Compiled using the Adam optimizer and MSE loss, this model aims to predict PM2.5 levels based on both temporal and spatial features.

- Hybrid LSTM-GRU Model

The Hybrid LSTM-GRU model integrates Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) layers to enhance the model's ability to capture both long-term dependencies and computational efficiency for air pollution forecasting. The LSTM layer, positioned first, is responsible for capturing complex temporal dependencies in the data. The LSTM layer outputs sequences, which are then passed to the GRU layer. The GRU layer, with fewer parameters than LSTM, provides computational efficiency while still capturing significant sequential dependencies in the data. Dropout layers are introduced after each recurrent layer to reduce overfitting and improve generalization. The final fully connected dense layer predicts the PM2.5 levels, which is the target pollutant. By combining LSTM and GRU, this hybrid model leverages the strengths of both architectures—LSTM's ability to learn long-term dependencies and GRU's computational efficiency—making it ideal for time-series forecasting. The model is compiled with the Adam optimizer and MSE loss to minimize prediction errors, allowing it to accurately forecast air pollution levels based on historical data.

- Hybrid GRU-1D CNN Model

The Hybrid GRU-1D CNN model merges the power of Gated Recurrent Units (GRU) and 1D Convolutional Neural Networks (CNN) to effectively model both spatial and temporal dependencies in air pollution forecasting. The model begins with a 1D CNN layer, which is adept at detecting local spatial patterns in pollutant data, such as sharp variations or trends. This convolutional layer applies multiple filters to the input data, extracting key features before passing them to the GRU layer. GRU, known for its computational efficiency and ability to capture sequential dependencies, is employed to learn long-term temporal patterns in the data. The hybrid structure allows the model to first capture spatial features and then model temporal relationships, which is crucial for forecasting pollutants like PM2.5. Dropout layers are used to prevent overfitting, ensuring that the model generalizes well to new data. The final dense layer outputs the predicted PM2.5 levels. This hybrid architecture is highly effective for complex time-series data, where both spatial and temporal dependencies must be captured, leading to accurate and robust air pollution forecasting.

Table 2. Hyperparameter Details of the Models

Model	LSTM	Hybrid LSTM-1D CNN	Hybrid LSTM-GRU	Hybrid GRU-1D CNN
Dropout Layers	N/A	Dropout(0.2)	Dropout(0.2) after LSTM & GRU layers	Dropout(0.2) after GRU layers
	(8,1)	(X_train.shape[1],1)	(X_train.shape[1],1)	(X_train.shape[1],1)
Optimizer	Adam	Adam	Adam	Adam
Loss Function	MSE	MSE	MSE	MSE
Epochs	100	100	100	100
Metrics	MSE, MAE	MAE	MAE	MAE
Total & Trainable Parameters	10,451 10,451	43,507 43,507	26,337 26,337	33,007 33,007
Learning Rate	0.001	0.001	0.001	0.001
Activation Functions	ReLU	ReLU	ReLU for LSTM, Tanh for GRU	ReLU for Conv1D, Tanh for GRU

3.6 Performance Evaluation

Performance evaluation of air pollution forecasting models is conducted using key metrics such as Mean Squared Error (MSE), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) score. MSE quantifies the average squared difference between actual and predicted values, penalizing large deviations. MAE provides a more interpretable measure of absolute differences. RMSE, derived from MSE, retains unit consistency with the predicted variable, offering a more intuitive assessment of model performance. The R^2 score evaluates model explanatory power, with values closer to 1 indicating better fit. Additionally, time-series models are assessed using visual comparisons of actual vs. predicted pollutant levels through line plots. Residual analysis is performed to ensure error distributions follow normality, confirming model reliability.

4. Results and Discussion

The evaluation of air pollution forecasting models is conducted using commonly used metrics such as Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and Mean Absolute Error (MAE). These metrics quantify the accuracy of predictions by measuring deviations from actual values. Lower values indicate better model performance, ensuring minimal prediction errors. MSE captures the average squared differences, making it sensitive to larger deviations, whereas RMSE, being the square root of MSE, retains the unit consistency of the predicted variable. MAE provides an intuitive measure of absolute errors. Model performance is compared across different architectures, including LSTM, CNN-LSTM, LSTM-GRU, and GRU-CNN hybrids, assessing their ability to capture spatiotemporal dependencies. A crucial aspect of evaluation is ensuring robust validation techniques; an error in performance assessment occurs when models are trained and tested on a single data split. Cross-validation mitigates this risk, providing a more generalized performance estimate. Additionally, graphical analysis of predicted vs. actual trends enhances interpretability. The discussion also emphasizes hyperparameter tuning and data preprocessing techniques that significantly impact model accuracy. Comparative results indicate that hybrid deep learning architectures generally outperform standalone models, demonstrating superior feature extraction and temporal sequence learning capabilities for air pollution forecasting.

A. MSE (Mean Square Error)

In order to evaluate the quality of images, the MSE has become the most widely used statistic. Since the number serves as an all-inclusive benchmark, the closer it is to 0, the better. The predicted value of this squared error loss can be expressed using a risk function denoted by the acronym MSE. Often, the mean squared error (MSE) is positive because of randomness or because the analyst overlooked information that may have produced a more precise estimate.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 \quad (1)$$

B. RMSE (Root Mean Square Error)

In statistics, the residuals' standard deviation is known as the RMSE. (errors in prediction). The root-mean-square error (RMSE) gauges how erratic the residuals are, whereas the residuals themselves measure the divergence from the regression line. It achieves this by showing how closely the data clusters around the ideal line of fit. The root-mean-square error is a frequently used statistic in climatology, forecasting, and regression analysis to confirm the results of studies.

$$RSME = \sqrt{\frac{\sum_{i=1}^{n}(P_i - O_i)^2}{n}} \quad (2)$$

C. Mean Absolute Error

The mean absolute error provides a quantitative means of comparing mistakes between two measurements that are indicative of the same phenomena. Comparing planned and actual numbers, beginning and finishing times, and one measuring method against another are a few examples of a Y against X comparison.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - x| \quad (3)$$

D. Coefficient of correlation (R^2)

The percentage of the variance of the dependent variable that has previously been linked to its independent variable is known as the R2 coefficient for determination, or R2 or R2. The ratio can be represented by the symbols R2, R2, or "R squared".

$$R^2 = \frac{n(\sum xy - (\sum x)(\sum y))}{\sqrt{[n\sum x^2 - (\sum x)^2]\,[n\sum y^2 - (\sum y)^2]}} \quad (4)$$

Table 3. Performance Evaluation of Proposed Models

Model	MSE	RMSE	MAE	R^2
LSTM	243.5	15.6	10.9	0.99
Hybrid LSTM-1D CNN	330.9	18.9	7.2	0.99
Hybrid LSTM-GRU	697.0	26.4	20.6	0.98
Hybrid GRU-1D CNN	201.2	14.1	6.7	0.99

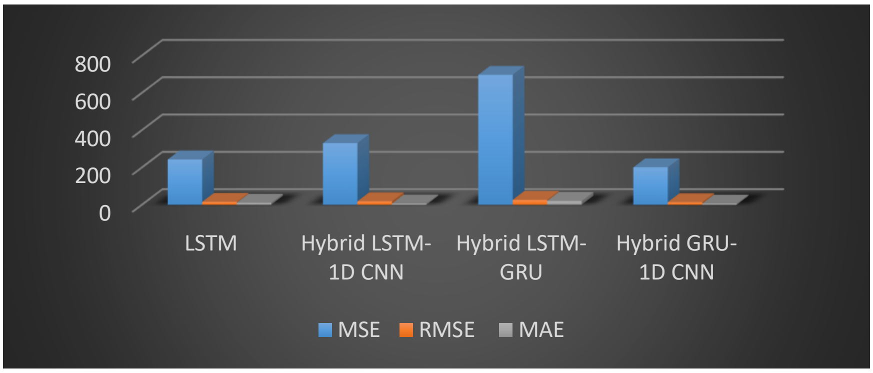

Figure 10 Performance Evaluation of Models

Table 3 presents the performance metrics of different deep learning models used for air pollution forecasting. The LSTM model achieves an MSE of 243.5, RMSE of 15.6, and MAE of 10.9, with an R^2 of 0.99, indicating strong predictive capability. The Hybrid LSTM-1D CNN model, despite having a higher MSE (330.9) and RMSE (18.9), exhibits the lowest MAE (7.2), suggesting better absolute error minimization. The Hybrid LSTM-GRU model shows the highest MSE (697.0) and RMSE (26.4), along with a higher MAE (20.6), which indicates relatively weaker performance compared to other models. The Hybrid GRU-1D CNN model outperforms others with the lowest MSE (201.2), RMSE (14.1), and MAE (6.7), demonstrating superior accuracy and error minimization while maintaining an R^2 of 0.99. Figure 10 visually compares these models, highlighting the effectiveness of hybrid architectures, particularly GRU-CNN, in capturing temporal dependencies and reducing forecasting errors.

4.1 Performance Graphs

Figures 11–14 illustrate the Mean Absolute Error (MAE) progression over training epochs for different air pollution forecasting models. Figure 11 (Hybrid LSTM-1D CNN) exhibits a fluctuating but steady decline in MAE, demonstrating improved learning stability through CNN feature extraction and LSTM sequence modeling. Figure 12 (LSTM) shows a smooth but slower MAE reduction, indicating the model's reliance solely on temporal dependencies without additional feature extraction. Figure 13 (Hybrid LSTM-GRU) presents a less stable convergence pattern, with higher initial errors and slower improvement, suggesting challenges in learning complex pollutant variations. Figure 14 (Hybrid GRU-1D CNN) demonstrates the most rapid and stable decrease in MAE, reinforcing its effectiveness in minimizing prediction errors. Across all models, hybrid architectures integrate complementary strengths, optimizing predictive accuracy. These performance graphs emphasize the necessity of structural modifications to deep learning models to achieve enhanced generalization and lower error rates across varying air pollution datasets.

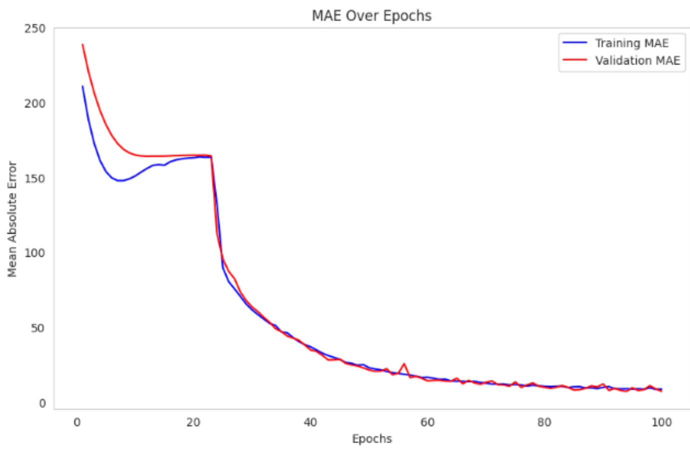

Figure 11 Hybrid LSTM-1D CNN

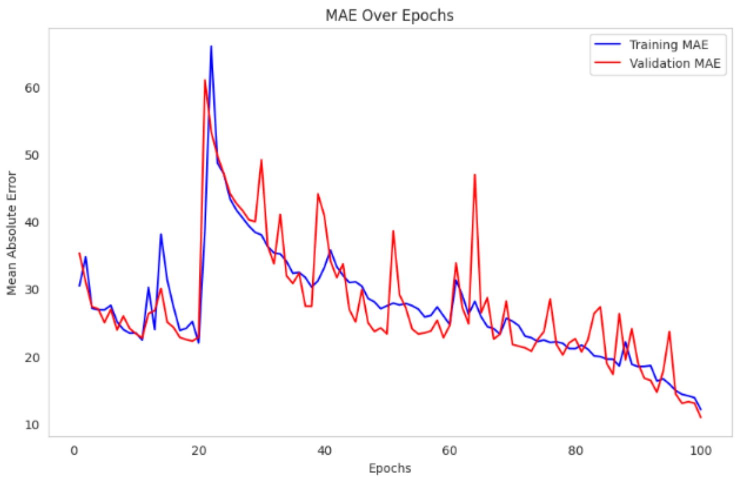

Figure 12 LSTM

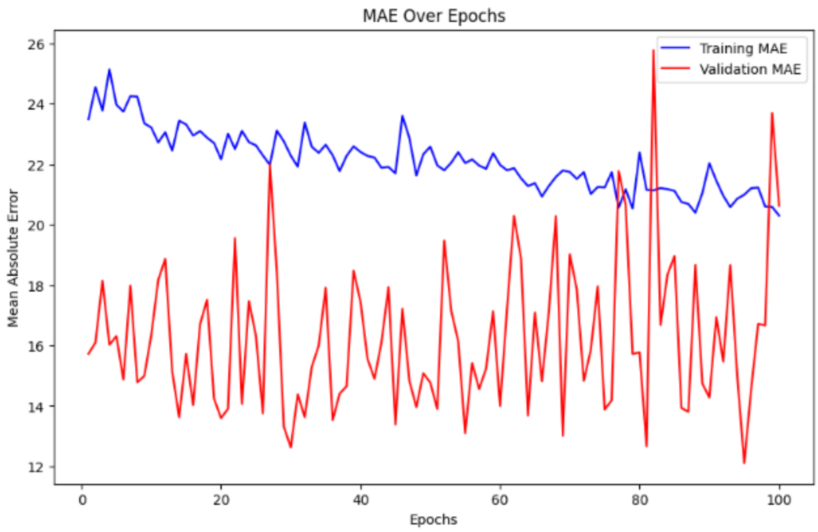

Figure 13 Hybrid LSTM-GRU

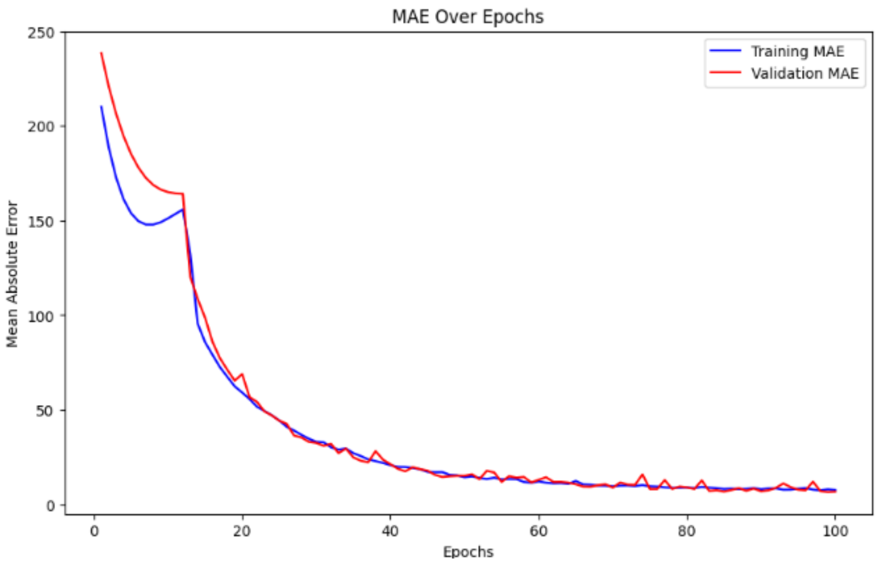

Figure 14 Hybrid GRU-1D CNN

4.2 Prediction Graph of Hybrid GRU-1D CNN

The prediction graph of the Hybrid GRU-1D CNN model demonstrates its superior forecasting performance for air pollution levels. This model effectively captures both spatial and temporal dependencies by leveraging 1D CNN for feature extraction and GRU for sequential learning, ensuring robust generalization. The graph exhibits a strong alignment between predicted and actual pollutant values, highlighting minimal deviations. Compared to other models, this hybrid approach achieves lower MSE, RMSE, and MAE, confirming its predictive accuracy. The smooth convergence and reduced error fluctuations indicate enhanced stability, making it the optimal model for precise air pollution forecasting in dynamic environments.

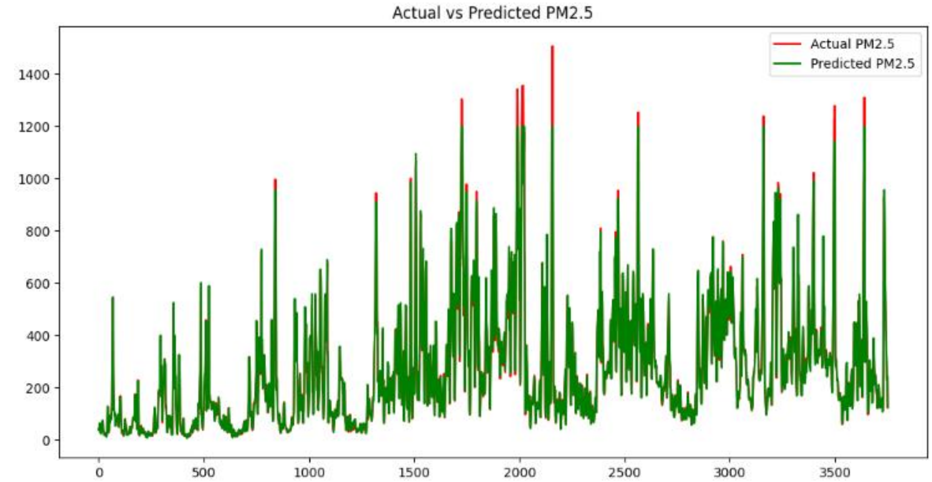

Figure 15 Actual vs. Predicted PM2.5 Levels Using Hybrid GRU-1D CNN Model Forecasting graph

Figure 15 illustrates the comparison between actual and predicted PM2.5 levels using the Hybrid GRU-1D CNN model. The close alignment of the predicted values with actual data indicates the model's strong forecasting capability. Minimal deviations suggest high accuracy, demonstrating the model's effectiveness in air pollution prediction.

5. Conclusion

In Conclusion, this study presents a robust methodology for air pollution forecasting using deep learning models, specifically targeting PM2.5 levels. The dataset, sourced from the Central Pollution Control Board (CPCB), includes critical pollutants such as PM2.5, PM10, NOx, SO2, and CO. Extensive data preprocessing, including exploratory data analysis (EDA), trend identification, and anomaly detection, was performed to prepare the dataset. Key steps like handling missing data, normalizing pollutant concentrations, and extracting temporal features such as day, month, and year were incorporated to address seasonal variations and enhance model accuracy. Four deep learning models were evaluated: standalone LSTM, Hybrid LSTM-1D CNN, Hybrid LSTM-GRU, and Hybrid GRU-1D CNN. The models were trained using an 80:20 train-test split, with feature extraction techniques further improving prediction performance. Among these models, the Hybrid GRU-1D CNN outperformed others, achieving the lowest Mean Squared Error (MSE) of 201.2, Root Mean Squared Error (RMSE) of 14.1, and Mean Absolute Error (MAE) of 6.7, with a high R2 value of 0.99. These results highlight the model's exceptional capability in accurately forecasting air pollution levels. While other models like LSTM and Hybrid LSTM-1D CNN also showed good performance, the Hybrid GRU-1D CNN demonstrated superior predictive accuracy. This research effectively illustrates the potential of deep learning models for environmental forecasting and their crucial role in supporting public health and policy-making.

References

[1] Z. Wu, N. Liu, G. Li, X. Liu, Y. Wang, and L. Zhang, "Meta-Learning-Based Spatial-Temporal Adaption for Coldstart Air Pollution Prediction," *Int. J. Intell. Syst.*, vol. 2023, 2023, doi: 10.1155/2023/3734557.

[2] N. S. Gupta, Y. Mohta, K. Heda, R. Armaan, B. Valarmathi, and G. Arulkumaran, "Prediction of Air Quality Index Using Machine Learning Techniques: A Comparative Analysis," *J. Environ. Public Health*, vol. 2023, pp. 1–26, 2023, doi: 10.1155/2023/4916267.

[3] T. Melese, T. Berhane, A. Mohammed, and A. Walelgn, “Credit-Risk Prediction Model Using Hybrid Deep—Machine-Learning Based Algorithms,” *Sci. Program.*, vol. 2023, pp. 1–13, 2023, doi: 10.1155/2023/6675425.

[4] A. Bikis, “Urban Air Pollution and Greenness in Relation to Public Health,” *J. Environ. Public Health*, vol. 2023, no. X, p. 8516622, 2023, doi: 10.1155/2023/8516622.

[5] Z. Guo, H. Lin, and Z. Qiao, “Air Pollution Detection and Sports Club Environmental Responsibility Based on the RBF Neural Network,” *Comput. Intell. Neurosci.*, vol. 2022, 2022, doi: 10.1155/2022/3693310.

[6] L. Zhang and G. Huang, “A Hybrid Approach for Network Rumor Detection Based on Attention Mechanism and Bidirectional GRU Model in Big Data Environment,” *Mob. Inf. Syst.*, vol. 2022, 2022, doi: 10.1155/2022/4185208.

[7] E. S. Lee *et al.*, “A Machine Learning-Based Study of the Effects of Air Pollution and Weather in Respiratory Disease Patients Visiting Emergency Departments,” *Emerg. Med. Int.*, vol. 2022, pp. 1–20, 2022, doi: 10.1155/2022/4462018.

[8] H. Zheng, M. Duan, X. Sun, G. Han, and Q. Jin, “Landslide Displacement Prediction Based on Transfer Learning and Bi-GRU,” *J. Sensors*, vol. 2022, 2022, doi: 10.1155/2022/4659853.

[9] D. Kothandaraman *et al.*, “Intelligent Forecasting of Air Quality and Pollution Prediction Using Machine Learning,” *Adsorpt. Sci. Technol.*, vol. 2022, 2022, doi: 10.1155/2022/5086622.

[10] S. Chen, M. Hu, Y. Jiang, and Y. Xiong, “Spatial-Temporal Characteristics Analysis of Air Pollution in Hubei Province Based on Functional Data Clustering,” *Wirel. Commun. Mob. Comput.*, vol. 2022, 2022, doi: 10.1155/2022/7023071.

[11] L. Zhiwei, “Bearing Fault Diagnosis of End-to-End Model Design Based on 1DCNN-GRU Network,” *Discret. Dyn. Nat. Soc.*, vol. 2022, 2022, doi: 10.1155/2022/7167821.

[12] Z. Li, X. Bian, J. Yin, X. Zhang, and G. Mu, “The Effect of Air Pollution on the Occurrence of Nonspecific Conjunctivitis,” *J. Ophthalmol.*, vol. 2016, pp. 10–13, 2016, doi: 10.1155/2016/3628762.

[13] C. M. Vong, W. F. Ip, P. K. Wong, and J. Y. Yang, “Short-term prediction of air pollution in macau using support vector machines,” *J. Control Sci. Eng.*, vol. 2012, 2012, doi: 10.1155/2012/518032.

[14] P. V. Moulton and W. Yang, “Air pollution, oxidative stress, and alzheimer’s disease,” *J. Environ. Public Health*, vol. 2012, no. x, 2012, doi: 10.1155/2012/472751.

[15] J. Nikiema, E. M. Nkhalambayausi-Chirwa, and Y. Andrs, “Quality and pollution control technologies for water, air, and soil,” *Int. J. Chem. Eng.*, vol. 2010, 2010, doi: 10.1155/2010/252131.

[16] M. J. Jiménez-Navarro, M. Lovrić, S. Kecorius, E. K. Nyarko, and M. Martínez-Ballesteros, “Explainable deep learning on multi-target time series forecasting: An air pollution use case,” *Results Eng.*, vol. 24, no. October, p. 103290, 2024, doi: 10.1016/j.rineng.2024.103290.

[17] K. S. Rautela and M. K. Goyal, “Transforming air pollution management in India with AI and machine learning technologies,” *Sci. Rep.*, vol. 14, no. 1, pp. 1–19, 2024, doi: 10.1038/s41598-024-71269-7.

[18] F. Naz *et al.*, “Comparative Analysis of Deep Learning and Statistical Models for Air Pollutants Prediction in Urban Areas,” *IEEE Access*, vol. 11, no. June, pp. 64016–64025, 2023, doi: 10.1109/ACCESS.2023.3289153.

[19] L. Zhu, Z. J. B. M. Husny, N. A. Samsudin, H. P. Xu, and C. Han, “Deep learning method for minimizing water pollution and air pollution in urban environment,” *Urban Clim.*, vol. 49, no. February, 2023, doi: 10.1016/j.uclim.2023.101486.

[20] Y. Gu, B. Li, and Q. Meng, “Hybrid interpretable predictive machine learning model for air pollution prediction,” *Neurocomputing*, vol. 468, pp. 123–136, 2022, doi: 10.1016/j.neucom.2021.09.051.

[21] A. Dairi, F. Harrou, S. Khadraoui, and Y. Sun, “Integrated Multiple Directed Attention-Based Deep Learning for Improved Air Pollution Forecasting,” *IEEE Trans. Instrum. Meas.*, vol. 70, 2021, doi: 10.1109/TIM.2021.3091511.

[22] P. Hähnel, J. Mareček, J. Monteil, and F. O’Donncha, “Using deep learning to extend the range of air pollution monitoring and forecasting,” *J. Comput. Phys.*, vol. 408, 2020, doi: 10.1016/j.jcp.2020.109278.

[23] S. Al-Janabi, M. Mohammad, and A. Al-

Sultan, “A new method for prediction of air pollution based on intelligent computation,” *Soft Comput.*, vol. 24, no. 1, pp. 661–680, 2020, doi: 10.1007/s00500-019-04495-1.

[24] S. Ameer *et al.*, “Comparative Analysis of Machine Learning Techniques for Predicting Air Quality in Smart Cities,” *IEEE Access*, vol. 7, pp. 128325–128338, 2019, doi: 10.1109/ACCESS.2019.2925082.

[25] I. I. Prado-Rujas, A. García-Dopico, E. Serrano, M. L. Córdoba, and M. S. Pérez, “A multivariable sensor-agnostic framework for spatio-temporal air quality forecasting based on Deep Learning,” *Eng. Appl. Artif. Intell.*, vol. 127, no. May 2023, 2024, doi: 10.1016/j.engappai.2023.107271.

[26] H. Tao *et al.*, “Machine learning algorithms for high-resolution prediction of spatiotemporal distribution of air pollution from meteorological and soil parameters,” *Environ. Int.*, vol. 175, no. December 2022, p. 107931, 2023, doi: 10.1016/j.envint.2023.107931.