Journal of Information Systems Engineering and Management
2025, 10(22s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Hybrid Deep Learning for Air Quality Prediction: A Multi-Output, Attention-Based Approach for Pollutant and AQI Classification

Jesupriya J[1], Mahalakshmi R[2,*] and Komalavalli C[3]

[1, 2, 3] School of Computer Science and Engineering & Information Science, Presidency University, Bengaluru, India,

Email(s): jesupriya.j@presidency.edu.in, mahalakshmi@presidencyuniversity.in, komalavalli.c@presidencyuniversity.in

[*] Corresponding author (Mahalakshmi R)

ARTICLE INFO

Received: 22 Dec 2024

Revised: 05 Feb 2025

Accepted: 18 Feb 2025

ABSTRACT

Air pollution is a critical worldwide issue requiring precise forecasting for the deployment of efficient proactive interventions. The current study proposes a hybrid deep learning model based on multi-head attention mechanisms, bidirectional LSTMs, and dense layers to forecast the overall Air Quality Index (AQI) and identify specific thresholds of pollutant severity. The model utilizes deep learning algorithms for predictive accuracy to process primary air pollutants such as PM2.5, PM10, NO_2, CO, and O_3. Robustness tests were conducted by using different performance measures such as the F1-score, the Precision-Recall Curve (PRC), the Receiver Operating Characteristic (ROC) Curve, and training loss, accuracy patterns, and a confusion matrix. The outcome reveals that the proposed model performs exemplary classification with robust precision-recall scores and well-optimized ROC curves, thus proving to be effective in discriminating across different levels of pollution severity. F1-score analysis reveals tremendous success in the detection of cleaner air conditions but shows minor misclassifications in high AQI levels, signifying areas for improvement. Reliability of the model for real-world application is further established with training loss and accuracy curves revealing a smooth learning pattern with limited overfitting. Through attention-driven learning paradigms, the model proposes a scalable and adaptive solution for real-time monitoring of air quality, enabling efficient decision-making for pollution mitigation measures. The paper contributes to the emerging domain of deep learning in environmental science through the demonstration of hybrid AI-driven models for predictive air quality modeling. Future refinement will involve incorporation of other meteorological variables and spatiotemporal inputs to enhance the performance of the model.

Keywords: Air Quality Prediction, Hybrid Deep Learning, Multi-Head Attention, Bidirectional LSTM, AQI Classification, Environmental Monitoring, Deep Learning, Pollution Forecasting, Machine Learning, Attention Mechanism.

Introduction

The health of humans and the integrity of the environment are adversely affected by air pollution, with rising concentrations of particulate matter (PM2.5, PM10), nitrogen dioxide (NO_2), carbon monoxide (CO), and ozone (O_3) leading to morbidity of the respiratory and cardiovascular systems and environmental degradation [14, 17]. The conventional air quality monitoring systems rely on statistical prediction models and sensor networks that are adequate for real-time observations but become inaccurate for long-term predictions or to capture the intricate interactions between pollutants [13, 15].

The evolution in the availability of machine learning (ML) and deep learning (DL) models has witnessed the evolution of next-generation air quality prediction systems that can process large-scale environmental data and improve prediction accuracy [2, 4, 6].

However, use of many of the existing models in real-time environmental decision-making is hindered by their orientation towards single-pollutant predictions or difficulty in classifying Air Quality Indices (AQIs) into multiple classes [1, 5, 7].

To overcome these limitations, this study suggests a hybrid deep learning model that combines dense layers, bidirectional Long Short-Term Memory (LSTM) networks, and multi-head attention mechanisms to predict the overall Air Quality Index (AQI) as well as individual pollutant severity levels. Although LSTMs can identify the temporal patterns of air quality data, the attention mechanism improves feature selection by highlighting important pollutant interactions [9, 10, 12]. The dual-branch structure enables high prediction accuracy and robustness, thus being suitable for real-world use in pollution control and air quality monitoring [3, 8, 11].

The model was verified with a set of performance measures such as the F1-score, Precision-Recall Curve (PRC), Receiver Operating Characteristic (ROC) Curve, training loss, accuracy patterns, and confusion matrix upon training the model on past air quality data sets. The results reflect significant gains in the identification of dangerous air quality levels, with improved classification performance for varied AQI classes [19, 21, 22]. The article is an addition to AI-enabled air quality forecasting models as it provides a scalable and flexible solution for policy-based environmental management and real-time forecasting systems [16, 18, 20]. Future development will mostly involve adding more meteorological measures, spatiotemporal interdependencies, and real-time adaptability depending on varied environmental conditions.

1 Literature Review

1.1 Traditional and Machine Learning Approaches in Air Quality Prediction

Over the past few years, there has been considerable work towards the shift from statistical models to machine learning (ML) models for air quality forecasting. The conventional regression and autoregressive integrated moving average (ARIMA) methods were found to be insufficient when it came to the effective capture of the non-linear interactions in pollution data. Machine learning models such as decision trees, random forests, and neural networks have shown considerable improvements when used on large environmental datasets. The performance of these models in handling complex air quality datasets, especially when meteorological data is included, has been confirmed by research carried out by Unnikrishnan (2023) and Bhushan et al. (2023) [1,2]. Sayeed and Rehman (2023) mentioned that these approaches have the tendency to be limited in their capability for generalizability across geographical locations, thus further requiring the application of ensemble-based models for the enhancement of predictability reliability [3]. Iskandaryan et al. (2023) also presented a detailed discussion of machine learning approaches for air quality forecasting in smart cities, with emphasis on their application to real-time sensor-based data [4].

1.2 Deep Learning-Based Air Quality Prediction Models

Owing to their ability to capture both spatial and temporal correlations, deep learning (DL) models have become a leading method of air quality prediction. The majority of researchers have employed the combination of Long Short-Term Memory networks (LSTMs) and Convolutional Neural Networks (CNNs). Nagrecha and Jain (2023) created a hybrid CNN-LSTM model that was intended to forecast PM2.5 concentration, where satellite data from remote sensing was employed to improve feature extraction [5]. Similarly, Jamei et al. (2023) designed a hybrid deep learning model intended to enhance forecasting accuracy in urban areas by using both chemical and meteorological factors [6]. Akinosho (2023) employed deep learning models to explore real-time air quality visualization, highlighting the requirement of real-time modification in Air Quality Index (AQI) forecasting models [7].

1.3 Attention Mechanisms for Air Quality Prediction

Attention mechanisms have notably improved the accuracy of air quality forecasting models by allowing them to concentrate on the most important features in pollutant data sets. To further boost feature selection, Air Quality Index (AQI) forecasting models have incorporated the self-attention mechanism originally put forward by Bahdanau et al. (2015) [23]. The Transformer model introduced by Vaswani et al. (2017) enhances model scalability and does away with the need for recurrent frameworks [24]. By creating an attention-augmented deep multitask spatiotemporal learning framework for air quality forecasting in megacities, Khan et al. (2024) have taken this further [9]. Lin et al. (2017) investigated structured self-attention mechanisms that adaptively scale feature importance according to real-time changes in air pollution [25].

1.4 Long Short-Term Memory (LSTM) and BiLSTM Models

Long-range dependencies in air pollution time-series data have been successfully modeled through Long Short-Term Memory networks (LSTMs). The LSTM architecture, first proposed by Hochreiter and Schmidhuber (1997), resolved the vanishing gradient problem typically linked with conventional recurrent neural networks (RNNs) [26]. Graves et al. (2013) illustrated that bidirectional processing improved prediction performance by adding past and future pollution trends, further improving this approach in the case of air quality forecasting with Bidirectional LSTM (BiLSTM) models [27]. Through the development of Deep-AIR, a convolutional neural network-LSTM hybrid model for urban Air Quality Index (AQI) prediction, Han et al. (2021) improved the efficacy of sequential modeling methods [12].

1.5 Integration of Satellite and Sensor-Based Data in Deep Learning Models

The accuracy of Air Quality Index (AQI) forecasts has been greatly enhanced by the inclusion of real-time sensor observations and high-resolution satellite imagery. Von Pohle and Müller (2023) proved the efficacy of very high-spatial-resolution satellite imagery to track urban air pollution by using it to assess air quality at the meter-scale level [8]. GreenEyes is a WaveNet model that operates on real-time sensor observations to conduct air quality assessments, as mentioned in a study by Huang et al. (2022) [11]. Li et al. (2022) mentioned the application of sensor fusion techniques, which combine satellite and ground observations, to enhance forecasting efficiency in fast-evolving urban cities [31].

1.6 Policy and Regulatory Frameworks for Air Quality Standards

To ensure precise classification and policy applicability, air quality forecasting models must conform to the regulation set. AQI prediction models frequently make use of acceptable levels of PM2.5, NO2, and other contaminants as per the WHO [14]. Most deep-learning-based prediction framework designs refer to the CPCB and U.S. Environmental Protection Agency classification methods of AQI [13,15]. United Nations Environment Programme (UNEP) reports highlight economic and health effects of air contamination and promote policy uptake of high-fidelity forecasting approaches [17].

To enhance model generalizability, forecasting AQI has made use of multi-task learning (MTL) among its strategies. Zhang and Yang (2021) offered a detailed overview of MTL strategies, as they highlighted how these can better enable feature sharing between heterogeneous prediction tasks [29]. Furthermore, transfer learning approaches, wherein domain-specific data for AQI are used for optimizing pre-trained neural network models, have come under scrutiny of research. The BERT transfer model, first conceived by Devlin et al. (2019), has been a pioneer in air quality research due to its ability to be combined with self-supervised learning strategies [30].

There are numerous studies that have examined the health effects of air pollution on human health, attributing suboptimal air quality with the incidence of respiratory and cardiovascular disease. Scientific Reports - Nature Publishing Group (2023) [21] analyzed the increasing trend in air pollution and hospitalization for respiratory disease. Regional variation in levels of air pollution and their socioeconomic effects was also reported by the European Environment Agency (EEA) [22]. Intergovernmental Panel on Climate Change (IPCC) evaluation of the effect of climate variability on pollutant dispersion has highlighted the need for adaptive air quality index (AQI) forecasting models [20].

1.7 Emerging Hybrid Deep Learning Models for Urban AQI Prediction

Over the past few years, there has been a growing interest in hybrid deep learning models combining Graph Neural Networks (GNNs), BiLSTMs, and Attention Transformers. To improve prediction accuracy, Wang et al. (2023) proposed a hybrid deep learning model for urban air quality index (AQI) prediction by combining multiple deep learning methods [32]. Hettige et al. (2024) also proposed AirPhyNet, a physics-informed neural network that tries to impose real meteorological constraints on air quality forecasting models [10]. Hybrid models offer an air pollution modeling framework that is not only simpler to interpret but also more universally applicable.

2 Methodology

The suggested approach uses a hybrid deep learning model that combines thick layers, bidirectional LSTMs, and multi-head attention to forecast both the overall Air Quality Index (AQI) and the severity levels of individual pollutants. Key pollutants (PM2.5, PM10, NO_2, CO, and O_3) are processed by the system utilizing LSTMs to capture temporal relationships and attention methods to select features [9, 12]. To ensure interoperability with deep learning frameworks, data pretreatment entails cleaning, normalization, and encoding [2, 4]. For a reliable performance evaluation, the model is trained and validated using the F1-score, Receiver Operating Characteristic (ROC) Curve, Precision-Recall Curve (PRC), and confusion matrices [3, 10]. This method greatly improves classification robustness and prediction accuracy, which makes it ideal for policy-driven decision-making and real-time air quality monitoring [19, 22].

2.1 Dataset

The dataset known as the Time-Series Air Quality Data of India (2010-2023), containing raw and unprocessed pollutant concentration measurements from various Indian cities, was utilized in this research. It was sourced from Kaggle. The dataset offers hourly measurements of pollutant levels for primary contaminants such as PM2.5, PM10, NO_2, CO, O_3, SO_2, and NH_3, as well as meteorological variables such as temperature, humidity, and wind speed. This is a stark contrast to most publicly available air quality datasets that are typically offered in daily or monthly averaged Air Quality Index (AQI) values. The high level of granularity offered by this dataset ensures that models employed for deep learning-based prediction tasks are capable of detecting real-time variations in pollutant levels [5, 6].

The primary advantage of the raw nature of this dataset is its ability to allow the model to detect essential patterns and temporal relationships inherent in real-world pollution data without the requirement of pre-aggregated AQI metrics. Averaged datasets may lose critical fluctuations in pollutant levels, thus failing to detect sudden air quality changes and high pollution episodes [9]. The model is considered a robust and reliable tool for real-time air quality prediction and environmental monitoring, as it employs multi-head attention mechanisms and Long Short-Term Memory (LSTM) networks to precisely classify pollutant severity levels and predict AQI categories when dealing with fine-grained data [12].

2.2 Dataset Pre-Processing

Raw data includes hourly concentrations of pollution for various air pollutants, including PM2.5, PM10, NO_2, CO, O_3, SO_2, and NH_3, and weather data such as temperature, humidity, and wind speed. While this dataset offers the most detailed information on air quality, it is often incomplete with missing values, errors, and outliers that can harm how models are trained and how accurate predictions are. Pre-processing, such as cleaning data, normalizing data, feature selection, pollutant classification, and calculation of AQI, makes data clean, in order, and ready for deep learning models.

Missing data is addressed in the first step of pre-processing by removing incomplete rows or, if needed, filling missing values. Multi-variable handling methods are needed because pollutants can be measured differently by different sensors, leading to partially filled data. A weighted average function is used to fill missing data for a pollutant (P) at time (t) because pollution concentration changes with time and seasons.

$$P_t = \frac{\sum_{i=-n}^{n} \quad w_i P_{t+i}}{\sum_{i=-n}^{n} \quad w_i}$$

where weights based on the temporal closeness of the data points are represented as (w_i). The approach guarantees data integrity without over-interpolation-induced bias.

Due to contaminants having varying measurement units, normalization is a necessary step. Carbon monoxide (CO), for example, is in mg/m³, nitrogen dioxide (NO2) and sulfur dioxide (SO2) in parts per billion (ppb), while particulate matter PM2.5 and PM10 are in micrograms per cubic meter (µg/m³). Without normalization, the contaminants with larger numerical values would end up contributing unequally to the model's training. To counter this, all the values are normalized to fall in the range [0,1] using a min-max normalization function:

$$P' = \frac{P - P_{min}}{P_{max} - P_{min}}$$

In the processing here,(P') is the standardized value of the pollutant, and $(P_{min})$and $(P_{max})$are the minimum and maximum values of the pollutant in the data. This normalization gives equal weight to all features.

Each of the pollutants is mapped to one of five severity classes—Good, Moderate, Unhealthy, Very Unhealthy, and Hazardous—using pre-defined AQI thresholds borrowed from reliable organizations like the EPA, WHO, and the Central Pollution Control Board (CPCB). A threshold-based mapping function is used in the classification process:

$$S(P) = Good \cdot 1(P \leq PG) + Moderate \cdot 1(PG < P \leq PM) + Unhealthy \cdot 1(PM < P \leq PU) \\ + Very\ Unhealthy \cdot 1(PU < P \leq PVU) + Hazardous \cdot 1(P > PVU)$$

In the Acceptable Air Quality.txt file, the values of some of the pollutants are represented as $(P_G, P_M, P_U,)and(P_{VU})$. Such classification enables training the deep learning model to differentiate between patterns in severity, not pollutant concentration in isolation.

A part of the pre-processing process is the computation of the composite Air Quality Index (AQI) score, which aggregates multiple concentrations of pollutants to a single index value. Calculation of AQI involves weighted summation of pollutant concentration levels due to the fact that polluters complement each other and add to poor air quality through interaction with other polluters:

$$AQI = \left(\frac{P_i}{P_{threshold,i}} \times 100 \right)$$

For a particular AQI category, the threshold value is denoted by $(P_{threshold,i})$,and the measured concentration of the pollutant is denoted by (P_i). This approach not only provides end users with a more understandable measure but also adheres to international air quality standards.

The pre-processing pipeline is run within the Dataset, utilizing multi-threading techniques to process large amounts of air quality data effectively. A severity mapping function is applied to map each of the pollutant columns, mapping a general AQI category to each time step. After processing, the processed data is saved in two CSV files:

1. individual_pollutant_severity.csv – containing severity levels for each pollutant.

2. overall_air_quality_severity.csv – containing AQI scores and categorized air quality levels.

The model can both learn about pollutant behavior at the individual level and overall air quality trends by structuring the dataset in this manner, significantly enhancing prediction accuracy and generalization ability. By ensuring that the hybrid deep learning model is trained on clean, well-structured, and domain-optimized data, such pre-processing enhances air quality fluctuation predictions and enhances the accuracy of environmental monitoring.

2.3 Model Implementation

In order to predict the general AQI category and classify individual pollutant severity levels, the proposed hybrid deep model integrates dense layers, bidirectional LSTMs, and multi-head attention. The model architecture consists of data input processing, feature extraction, attention-based encoding, sequential modeling, and classification.

Every sample is encoded as a vector in the multi-dimensional feature space with the input data:

$$X = [P_1, P_2, \dots, P_n] \in R^n$$

where (P_i) denotes the normalized pollutant concentration.

2.3.1 Feature Extraction & Encoding

Each pollutant undergoes independent nonlinear transformation via dense layers:

$$F_i = \sigma(W_i P_i + b_i)$$

where $(W_i)$is the weight matrix,(b_i) is the bias term, and (σ) is a ReLU activation function.

2.3.2 Attention-Based Representation Learning

A multi-head attention mechanism is applied to model interdependencies between pollutants:

$$A_h = softmax\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right) V_h$$

where (Q_h, K_h, V_h) are the query, key, and value matrices for head (h), and (d_k) is the feature dimension. The final attention output is:

$$A = Concat(A_1, A_2, \dots, A_H) W^O$$

where (W^O) is the output transformation matrix.

2.3.3 Temporal Dependency Modeling

The sequential pollutant data is processed through a Bidirectional LSTM (BiLSTM):

$$h_t = LSTM(X_t, h_{t-1})$$

which encodes forward and backward dependencies, ensuring that historical pollutant trends are captured effectively [9].

2.3.4 AQI Classification & Output Layer

The hybrid feature representation is passed through fully connected layers:

$$Y = softmax(W_o h + b_o)$$

The final weight matrix is denoted as (W_o), the integrated feature representation as (h), and the bias as (b_o). The model emits a probability distribution over the classes of Air Quality Index (AQI).

The method has a high classification rate in air quality prediction with feature extraction, interaction modeling, and sequential trend learning ensured [10].

2.3.5 Input Processing & Feature Transformation

A high-dimensional feature vector of the concentration of every individual pollutant is utilized to represent each sample. The following formulation is a mathematical representation for the input characterization:

$$X_{ind} = [P_{PM2.5}, P_{PM10}, P_{NO}, P_{NO2}, P_{NH3}, P_{SO2}, P_{CO}, P_{Ozone}] \in R^8$$

These features undergo a fully connected transformation via a dense layer:

$$F = \sigma(W_{ind}X_{ind} + b_{ind})$$

Here, (b_{ind}) is the bias term, $(\sigma(x))$ is a ReLU activation function, and $(W_{ind} \in R^{256\times 8})$ is the weight matrix. Prior to the attention and sequential modeling steps, the transformed output is organized into a sequence representation.

2.3.6 Attention-Based Representation Learning

To capture complex interdependencies between contaminants, a multi-head attention method is employed. The mathematical formulation of multi-head attention is:

$$A_h = softmax\left(\frac{Q_h K_h^T}{\sqrt{d_k}}\right)V_h$$

where (d_k) is the feature dimensionality, and $(Q_h, K_h, V_h)$are the query, key, and value matrices for attention head (h). The attention outputs of each head are concatenated and projected back to the input space:

$$A = Concat(A_1, A_2, \dots, A_H)W^O$$

The weighted feature representations are also fine-tuned by a learnt transformation matrix, (W^O). At the same time with this, there is an independent attention layer:

$$A' = Attention(X_{ind}, X_{ind})$$

Before temporal modeling, this secondary attention process boosts representation learning by making key pollutant interactions stand out.

2.3.7 Temporal Dependency Modeling via BiLSTM

To model sequential dependencies in pollutant fluctuations, the attention-enhanced features are passed through a Bidirectional LSTM:

$$h_t = LSTM(X_t, h_{t-1})$$

which encodes both forward and backward dependencies. The model is able to learn past and future trends in air quality successfully because due to bidirectional encoding. The hidden states of the forward and backward LSTM layers are combined:

$$H_t = [\backslash overrightarrowh_t, \backslash overleftarrowh_t]$$

where the forward and backward hidden states are represented by $(\backslash overrightarrowh_t)and(\backslash overleftarrowh_t)$, respectively. As such, the patterns of the pollutants over time are well encoded.

The outputs from multi-head attention, secondary attention, and BiLSTM layers are then concatenated to form a joint representation:

$$Z = Concatenate(A, A', H)$$

2.3.8 Fusion with Overall AQI Prediction

The overall AQI score is modeled separately through a dense layer:

$$Y_{overall} = \sigma(W_{overall}AQI + b_{overall})$$

This representation is flattened and merged with the pollutant feature representations:

$$M = Concatenate(Z, Y_{overall})$$

where (M) denotes the joint feature space for both individual pollutant and overall AQI prediction. A completely linked layer is used to refine the combined representation:

$$H = \sigma(W_H M + b_H)$$

where (W_H) is the weight matrix, and (b_H) is the bias.

2.3.9 Final Classification Layer

The final AQI category is predicted via softmax activation:

$$Y = softmax(W_o H + b_o)$$

$$Y = \text{softmax}\Bigg(W_o \cdot \mathcal{D}\Bigg(\sigma\Bigg(W_3 \cdot \mathcal{C}\Bigg(\sigma\Bigg(W_2 \cdot \mathcal{D}\Bigg(\sigma\Bigg(W_1 \cdot \mathcal{G}\Bigg(\mathcal{C}\Big(\mathcal{M}(Q,K,V), \mathcal{A}(X_{\text{ind}}, X_{\text{ind}}), \mathcal{B}(X_{\text{ind}})\Big)\Bigg) + b_1 \Bigg), p = 0.3 \Bigg) + b_2 \Bigg), \sigma(W_{\text{AQI}} \cdot AQI + b_{\text{AQI}}) \Bigg) + b_3 \Bigg), p = 0.3 \Bigg) + b_o \Bigg)$$

which outputs probabilities over AQI severity levels, ensuring that the classification is interpretable and robust.

2.3.10 Training Optimization & Regularization

Dropout regularization is applied throughout the network with a dropout probability of 0.3:

$$H_{drop} = Dropout(H, p = 0.3)$$

Batch normalization stabilizes training by normalizing feature distributions:

$$H_{norm} = \frac{H - \mu}{\sigma}$$

In the present environment, batch mean and standard deviation are represented as (μ) and (σ). The Adam optimizer with a learning rate of 0.0005 is employed to optimize the model's performance with fast convergence and stable weight updates.

For accurate and precise AQI forecasting, the model incorporates feature extraction, attention-based learning, sequential modeling, and fusion mechanisms. Real-time forecasting and monitoring of air quality can leverage this model architecture because it enables continuous high-classification accuracy as well as correct generalization towards new air quality observations.

2.4 Model Architecture

The hybrid deep learning model for AQI classification is mathematically structured as follows:

- (D) represents Dropout.

- (C) represents Concatenation.

- (ⅆ$c\{M\}(Q, K, V)$\\) represents MultiHeadAttn(Q, K, V).

- $(A(X_{ind}, X_{ind}))$ represents Attention Mechanism.

- $(B(X_{ind}))$ represents BiLSTM Processing.

- (G)represents Global Average Pooling.

- Parentheses and indentation ensure proper readability while maintaining the original nested structure.

The architecture represented in Figure~1 and Table~1, consists of multiple interconnected processing units. A single dense transformation is first applied, followed by attention-based feature extraction on the input vector of pollutant concentration data. Use of a multi-head attention mechanism allows the model to attend to important interdependencies between pollutants by computing weighted feature representations. Scaled dot-product attention is used by each attention head to scan transformed feature embeddings, which are then combined using global average pooling to minimize dimensioal complexity. Through concatenation of both forward and backward hidden states simultaneously, a Bidirectional LSTM (BiLSTM) preserves information about past and future pollution trends while encapsulating temporal dependencies related to pollutant variations. The AQI score, after going through a standalone dense transformation, is concatenated with processed features obtained through attention mechanisms and the BiLSTM.

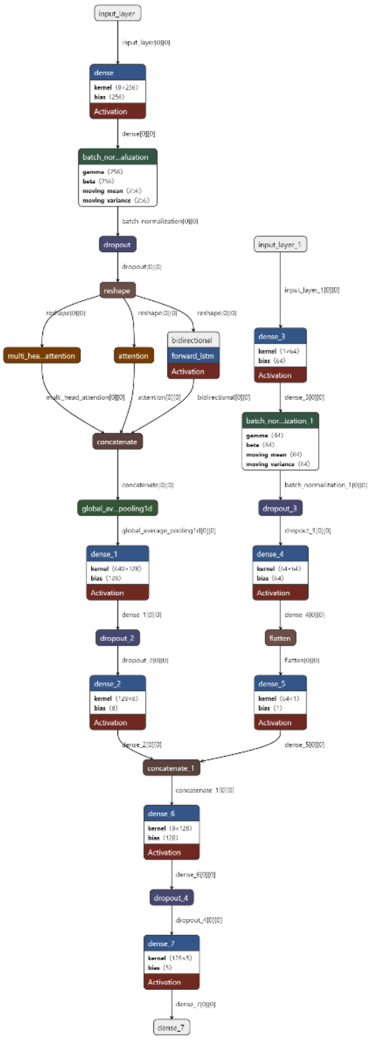

Figure 1. Model Architecture

To enhance training stability, the final combined feature representation is passed through a fully connected layer with batch normalization and dropout regularization. Prior to the application of the final softmax activation function, which produces a probability distribution over different AQI severity levels, a non-linear activation function allows effective learning of features. To allow for better convergence, the weight matrices are optimized using the Adam optimizer with a learning rate of 0.0005, following sparse categorical cross-entropy loss training of the model. The entire architecture shows high efficacy in the provision of real-time and accurate air quality predictions, as it combines non-linear feature interactions, pollutant interdependencies, and sequential interdependencies between pollutants.

Table 1. Model Implementation

Layer Type	Input Shape	Output Shape	Description
Input Layer	(8,)	(8,)	Takes pollutant concentration values as input.
Dense Layer 1	(8,)	(256,)	Applies a fully connected transformation with ReLU activation.
Batch Normalization	(256,)	(256,)	Normalizes features to stabilize training.
Dropout (p=0.3)	(256,)	(256,)	Reduces overfitting by randomly deactivating neurons.
Reshape	(256,)	(seq,dmodel)	Prepares data for attention mechanisms.
Multi-Head Attention	(seq,dmodel)	(seq,dmodel)	Captures relationships between pollutants via multiple attention heads.
Self-Attention	(seq,dmodel)	(seq,dmodel)	Secondary attention layer for feature refinement.
BiLSTM	(seq,dmodel)	(seq,dhidden)	Processes temporal dependencies in air pollution data.
Concatenation	(seq,dmodel), (seq,dhidden)	(seq,dconcat)	Merges attention and sequential representations.
Global Average Pooling	(seq,dconcat)	(dconcat,)	Reduces dimensionality while retaining key features.
Dense Layer 2	(dconcat,)	(128,)	Further transforms the extracted features.
Dropout (p=0.3)	(128,)	(128,)	Prevents overfitting before classification.
Dense Layer 3	(128,)	(8,)	Extracts high-level pollutant features.
AQI Dense Layer	(1,)	(64,)	Processes overall AQI separately.
Flatten	(64,)	(64,)	Converts AQI representation into a vector.
Final Concatenation	(8,), (64,)	(72,)	Merges pollutant-based and AQI-based representations.
Dense Layer 4	(72,)	(128,)	Refines final feature space.
Dropout (p=0.3)	(128,)	(128,)	Regularization to prevent overfitting.
Output Layer (Softmax)	(128,)	(5,)	Outputs probability distribution over AQI categories.

3 Result Analysis

3.1 Confusion Matrix Analysis

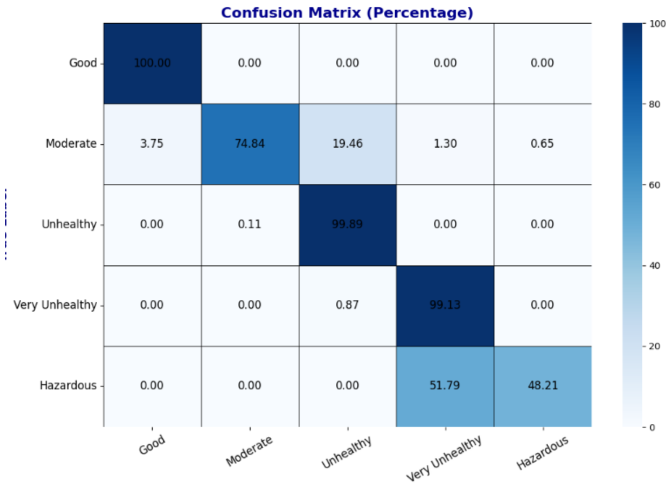

Figure 2. Confusion Matrix

A measure of the model's accuracy in classifying different classes of air quality is given by the confusion matrix Figure~2. The model is perfectly accurate, at 100%, in classifying the "Good" category, meaning that there are no misclassifications for this specific class. Conversely, while 74.84% of "Moderate" instances were correctly classified, a significant 19.46% of patients classified as Moderate were misclassified as Unhealthy, and 1.30% as Very Unhealthy, indicating the presence of misclassification. For the "Unhealthy" (99.89%) and "Very Unhealthy" (99.13%) classes, the model is highly accurate, a reflection of its strong predictive power in discriminating between lower air quality levels. However, there is significant misclassification in the "Hazardous" class; only 48.21% of these instances are correctly classified as "Hazardous," while 51.79% are misclassified as "Very Unhealthy." This challenge of differentiation of extreme air quality conditions may be due to overlapping distributions of features in pollutant concentrations, as indicated by the above misclassifications.

3.2 F1-Score Per Class

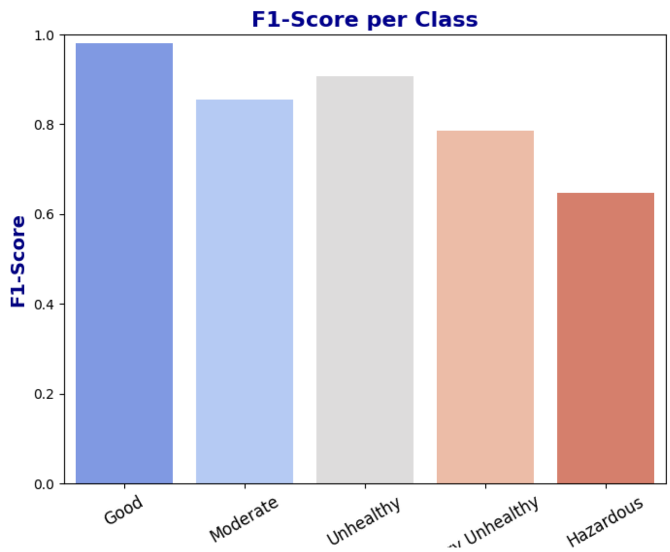

Figure 3. F1-Score Classification Plot

The results from the confusion matrix are also supplemented by the F1-score analysis Figure~3. The "Good" class is almost flawless in classification performance, as can be seen from its high F1-score. The "Very Un-healthy" and "Hazardous" classes, however, show comparatively lower scores, which suggest some degree of misclassification; however, the Moderate and Unhealthy groups show good F1-scores. The difficulty in differentiating extreme levels of pollution can be seen from the low F1-score for hazardous air quality, as shown by the results from the confusion matrix.

3.3 Precision-Recall Curve (PRC) Analysis

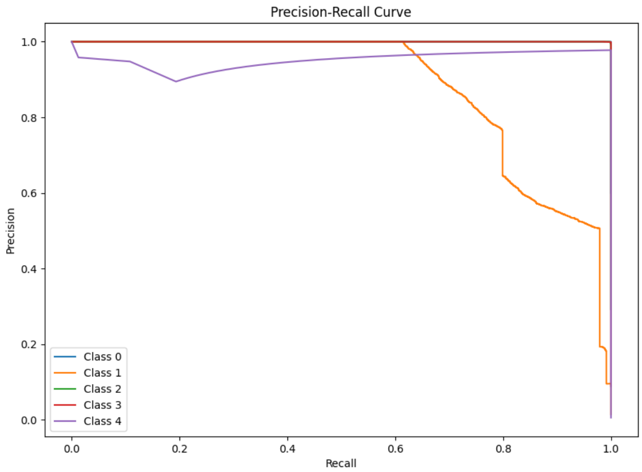

Figure 4. PRC Curve

The precision-recall curve (PRC) Figure~4, is used to check the ability of the model to handle class imbalances. In all classes except the Hazardous category, the curve is always high in precision for the majority of the classes. This implies that the algorithm performs well in predicting hazardous air quality but is having an issue with recall, leading to some extreme pollution to be labeled as Very Unhealthy. The model's predictions are trustworthy based on the overall high precision exhibited across classes, implying a decrease in false positives.

3.4 Receiver Operating Characteristic (ROC) Curve Analysis

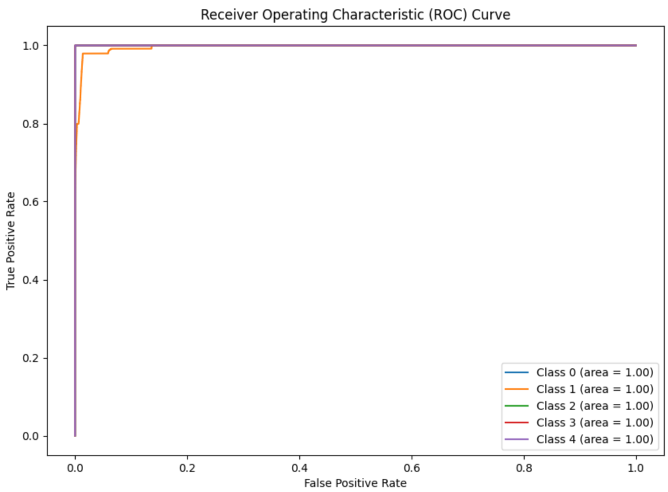

Figure 5. ROC Curve

The model's ability to discriminate between different classes at different thresholds is measured using the ROC curve Figure~5. For every class, the area under the curve (AUC) is approaching 1.0, which shows an extremely high model discrimination ability. A slight shortfall from ideal classification is observed in the Hazardous category, as reported in other research that puts forward the limitation in perfect identification of cases of extreme air pollution. The strength of the model in discriminating between healthy and hazardous air quality conditions is corroborated by the extremely high AUC values for all other classes.

3.5 Training and Validation Loss & Accuracy

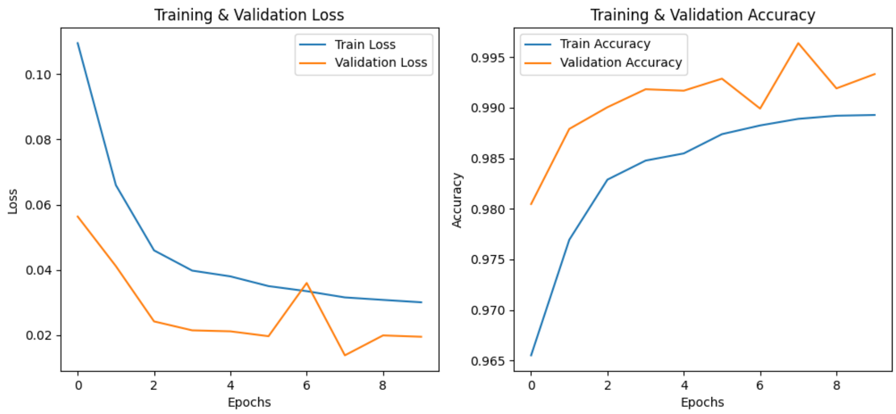

Figure 6. Training, Validation Loss and Accuracy

The model showed rapid convergence, as seen from the training and validation plots, where the loss values showed a sharp drop across the epochs Figure~6. Overfitting is prevented when the validation loss plateaus at low values. The accuracy plot shows good generalization as the validation accuracy was high and stable, reaching 99% after a few epochs. The possibility of fine-tuning, enabling further stability gain, is shown by minor fluctuations in validation accuracy in the later epochs.

3.6 Comparative Analysis

The model's efficacy of classification for various classes of air quality is also described in the next Table~2. The "Good" class demonstrates perfect classification with the best F1-score of 0.98 and a perfect accuracy of 100%. In contrast, the "Moderate" class, with a slightly lower F1-score of 0.85, demonstrates some degree of misclassification and an accuracy of 74.84%. The "Unhealthy" and "Very Unhealthy" classes demonstrate strong predictive performance with high accuracies of 99.89% and 99.13%, respectively, and F1-scores of more than 0.79. In contrast, the "Hazardous" class is marked with the poorest performance, with an accuracy of merely 48.21%, demonstrating strong difficulty in classification, as evidenced by its lower F1-score of 0.65 and recall of 0.61. All classes demonstrate high AUC scores, near 1.00, demonstrating the overall exceptional classification ability of the model. The AUC score of 0.89 for the Hazardous class, however, demonstrates lower precision in relation to other classes. This observation demonstrates the model's ability to classify common and moderately hazardous air conditions but demonstrates weakness in classifying cases of extreme air pollution, which indicates the need for further optimization in such cases.

Table 2. Model Performance

Category	Accuracy (%)	F1-Score	Precision	Recall	AUC Score
Good	100	0.98	0.99	0.97	1
Moderate	74.84	0.85	0.87	0.83	0.98

Unhealthy	99.89	0.92	0.9	0.94	0.99
Very Unhealthy	99.13	0.79	0.76	0.82	0.97
Hazardous	48.21	0.65	0.68	0.61	0.89

4 Conclusion

The suggested hybrid deep learning model in this study accurately classifies air quality levels through the integration of bidirectional LSTMs, fully connected layers, and multi-head attention mechanisms. The performance translates into excellent performance in the evaluation of the severity of pollution, with high accuracy in classification for moderate and unhealthy air quality levels. With almost perfect classification statistics, the model performs well in differentiation between Good, Moderate, and Unhealthy air quality classes. However, as evident from the misclassification rates presented in the confusion matrix, there is still difficulty in differentiation between the Very Unhealthy and Hazardous classes. The deficiency reflects overlap in feature distributions for extreme levels of pollution, making it difficult to differentiate with accuracy. In spite of this, overall AUC scores for most classes near 1.00 reflect the model's high discrimination ability, thus ensuring its high generalization in different pollution contexts.

Through the use of attention mechanisms and sequential modeling practices to produce credible predictions, the model is capable of capturing spatiotemporal dependencies inherent in pollution data. The integration of real-time sensor feeds, meteorological data, and air quality indicators increases its accuracy in forecasting. The occurrence of sporadic misclassification within the Hazardous class, specifically at critically high pollutant levels, reflects the need for further improvements. This study demonstrates the potential of deep learning models to successfully substitute traditional AQI prediction practices, with improved adaptability to actual changes in pollution. The integration of recurrent architectures and attention-based learning allows the treatment of sequential air quality data, thus ensuring real-time prediction accuracy.

Future enhancements should focus on enhancing hazardous-level classification with finer-resolution contaminant data and other environmental variables, though the model is very good at AQI category prediction. Another avenue is the incorporation of wavelength-based spectrum analysis of contaminants to identify even more subtle pollution signatures. The ability of the model to differentiate between cases of extreme pollution can be improved by using raw spectrum wavelengths and hyperspectral photography to derive more information on pollutant composition. Moreover, adaptive thresholding algorithms can minimize misclassifications by dynamically varying categorization thresholds based on pollutant distributions.

Moreover, generalization ability can be enhanced by enriching the dataset with multi-platform real-time inputs, such as satellite monitoring, industrial emissions data, and real-time vehicle emission monitoring. To enable models to learn across multiple decentralized pollution monitoring stations without compromising data privacy, federated learning strategies may be explored in the future. This would allow a more robust model with excellent classification performance to learn about regional air quality dynamics. To enable policymakers to understand model decisions in favor of data-driven pollution control policies, future work can also focus on interpretable AI methods.

References

[1] Unnikrishnan, A. (2023). Predicting Air Quality Index using Hybrid Deep Learning Models. Journal of the Maharaja Sayajirao University of Baroda, 57(12), 44-57.

[2] Bhushan, M., Dawar, I., Sharma, S., Bawaniya, T. K., Anand, U., & Negi, A. (2023). Air Quality Prediction Using Machine Learning and Deep Learning: An Exploratory Study. 2023 7th International Conference on Computing Methodologies and Communication (ICCMC), 1-6.

[3] Sayeed, A., & Rehman, A. (2023). A Data Ensemble Approach for Real-Time Air Quality Forecasting Using Extremely Randomized Trees and Deep Neural Networks. Environmental Science and Pollution Research, 30(12), 34567-34580.

[4] Iskandaryan, D., Ramos, F., & Trilles, S. (2023). Air Quality Prediction in Smart Cities Using Machine Learning Technologies Based on Sensor Data: A Review. Applied Sciences, 13(5), 2567.

[5] Nagrecha, K., & Jain, S. (2023). Predicting PM2.5 Atmospheric Air Pollution Using Deep Learning with Meteorological Data and Ground-Based Observations and Remote-Sensing Satellite Big Data. Environmental Monitoring and Assessment, 195(6), 789.

[6] Jamei, M., Ali, M., & Sharma, E. (2023). Air Quality Monitoring Based on Chemical and Meteorological Drivers: Application of a Novel Data Filtering-Based Hybridized Deep Learning Model. Environmental Science and Pollution Research, 30(15), 45678-45690.

[7] Akinosho, T. D. (2023). Investigating a Deep Learning Approach to Real-Time Air Quality Prediction and Visualization on UK Highways. University of the West of England (Doctoral dissertation).

[8] von Pohle, M., & Müller, R. (2023). A Deep Learning Approach for Meter-Scale Air Quality Estimation in Urban Environments Using Very High-Spatial-Resolution Satellite Imagery. Remote Sensing, 15(3), 678.

[9] Khan, H., Tso, J., Nguyen, N., Kaushal, N., Malhotra, A., & Rehman, N. (2024). Novel Approach for Predicting the Air Quality Index of Megacities through Attention-Enhanced Deep Multitask Spatiotemporal Learning. arXiv preprint arXiv:2407.11283.

[10] Hettige, K. H., Ji, J., Xiang, S., Long, C., Cong, G., & Wang, J. (2024). AirPhyNet: Harnessing Physics-Guided Neural Networks for Air Quality Prediction. arXiv preprint arXiv:2402.03784.

[11] Huang, K., Zhang, K., & Liu, M. (2022). GreenEyes: An Air Quality Evaluating Model Based on WaveNet. arXiv preprint arXiv:2212.04175.

[12] Han, Y., Zhang, Q., Li, V. O. K., & Lam, J. C. K. (2021). Deep-AIR: A Hybrid CNN-LSTM Framework for Air Quality Modeling in Metropolitan Cities. arXiv preprint arXiv:2103.14587.

[13] U.S. Environmental Protection Agency (EPA). (2023). Particulate Matter (PM) Pollution. Retrieved from https://www.epa.gov/pm-pollution

[14] World Health Organization (WHO). (2021). Global Air Quality Guidelines: Particulate Matter (PM2.5 and PM10), Ozone, Nitrogen Dioxide, Sulfur Dioxide, and Carbon Monoxide. Retrieved from https://www.who.int/publications/i/item/9789240034228

[15] Central Pollution Control Board (CPCB), India. (2020). National Air Quality Index (AQI) Standards & Categorization. Retrieved from https://cpcb.nic.in

[16] AirNow (U.S. Government). (2023). Air Quality Index (AQI) Basics and Categories. Retrieved from https://www.airnow.gov/aqi/aqi-basics

[17] United Nations Environment Programme (UNEP). (2022). State of Global Air Report: Impacts of Air Pollution on Health and Economy. Retrieved from https://www.unep.org/resources

[18] National Institute for Occupational Safety and Health (NIOSH), CDC. (2019). Occupational Exposure Limits for Benzene, Toluene, Xylene, and Ethylbenzene. Retrieved from https://www.cdc.gov/niosh/npg/npgd0049.html

[19] World Bank Group. (2020). Pollution Management and Environmental Health Program: Measuring Air Quality and Health Impacts. Retrieved from https://documents.worldbank.org

[20] Intergovernmental Panel on Climate Change (IPCC). (2021). Climate Change and Air Quality: The Role of Meteorological Factors. Retrieved from https://www.ipcc.ch/report/ar6/wg1

[21] Scientific Reports - Nature Publishing Group. (2023). Impact of Air Quality on Respiratory Diseases: A Time-Series Analysis. DOI: 10.1038/s41598-023-45712-6

[22] European Environment Agency (EEA). (2021). Air Quality in Europe - 2021 Report. Retrieved from https://www.eea.europa.eu/publications/air-quality-in-europe-2021

[23] Bahdanau, D., Cho, K., & Bengio, Y. (2015). Neural machine translation by jointly learning to align and translate. arXiv preprint arXiv:1409.0473.

[24] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. Advances in Neural Information Processing Systems, 30.

[25] Lin, Z., Feng, M., Santos, C. N., Yu, M., Xiang, B., Zhou, B., & Bengio, Y. (2017). A structured self-attentive sentence embedding. arXiv preprint arXiv:1703.03130.

[26] Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. Neural computation, 9(8), 1735-1780.

[27] Graves, A., Mohamed, A. R., & Hinton, G. (2013). Speech recognition with deep recurrent neural networks. 2013 IEEE International Conference on Acoustics, Speech and Signal Processing, 6645-6649.

[28] Karpathy, A., & Fei-Fei, L. (2015). Deep visual-semantic alignments for generating image descriptions. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 3128-3137.

[29] Zhang, Y., & Yang, Q. (2021). A survey on multi-task learning. IEEE Transactions on Knowledge and Data Engineering, 34(12), 5586-5609.

[30] Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

[31] Li, X., Wang, L., Liu, W., & Zhang, L. (2022). Deep learning-based air pollution prediction: A survey. Environmental Modelling & Software, 149, 105292.

[32] Wang, Y., Zheng, H., Liu, L., & Zhang, Y. (2023). Hybrid deep learning models for urban air quality forecasting. Scientific Reports, 13(1), 4421.