International Research Journal of Multidisciplinary Scope (IRJMS), 2026; 7(2): 1747-1762

Original Article | ISSN (O): 2582-631X DOI: 10.47857/irjms.2026.v07i02.09113

Air Quality Index Prediction Using Hybrid LSTM–GRU and Residual BiLSTM Models

Sushant Suresh Kothari*, Aditi Chhabria, Gargi Phadke

Department of Computer Science and Engineering, Ramrao Adik Institute of Technology, DY Patil Deemed to be University, Nerul, Navi Mumbai, Maharashtra, India. *Corresponding Author's Email: sk619kothari@gmail.com

Abstract

In areas that are urbanizing quickly, air pollution continues to be a major problem because it can lead to enormous dangers to human health and also to environmental sustainability. This paper suggests two strong deep learning models for precise prediction of Air Quality Index (AQI) that makes use of current atmospheric as well as meteorological data so as to raise public awareness, guide policy and also to encourage proactive citizen action. Hybrid Long Short-Term Memory–Gated Recurrent Unit (LSTM–GRU) architecture and a residual bidirectional LSTM (BiLSTM) network with 1×1 convolutional skip connections for enhanced gradient flow and multi-scale representation are the two refined deep learning architectures that are created. For capturing autocorrelations and also short-term dependencies in AQI dynamics both models make use of lagged pollutant data in conjunction along with 27 engineered features, that includes rolling and exponential statistics, indices of heat-humidity, metrics related to rate-of-change, ratios of pollutant concentration as well as cyclical encodings of temporal variables (such as day of week, hour of day). To make sure about steady convergence and robust generalization, the models were trained using the Adam optimizer with Huber loss, adaptive learning-rate decay and early stopping. Inputs are quantile-normalized. By integrating physics-based feature engineering and relevant domain knowledge, the method guarantees organized transparency in the modelling procedure. It also successfully records changes related to time, reduces data interference and adjusts to the changing characteristics of city settings, leading to more trustworthy and understandable forecasts. However, this also helps in creation of intelligent as well as scalable forecasting tools for prediction of AQI for public health dashboards, early warning systems and also for infrastructures of smart city.

Keywords: Air Quality Index, Conv1D, Deep Learning, Forecasting, LSTM–GRU, Residual BiLSTM.

Introduction

Air quality forecasting and monitoring are essential for policy, public education and lowering health risks since urban air pollution is becoming a greater danger to public health. The Air Quality Index (AQI) combines pollutant concentrations into a single, comprehensible figure. The Central Pollution Control Board (CPCB) in India keeps an eye on a number of pollutants, such as particulate matter ($PM_{2.5}$ and PM_{10}), nitrogen dioxide (NO_2), sulphur dioxide (SO_2), carbon monoxide (CO) and ozone (O_3), in order to calculate the AQI. This allows for the evaluation of pollution severity and related health hazards. Nonlinear dynamics, long-range correlations and periodic anomalies like festival spikes and inversion events that characterize urban Indian air quality data are often overlooked by previously used statistical and machine learning approaches. While techniques such as linear regression, ARIMA, ANFIS as well as support vector regression are generally mathematically quite transparent and also computationally less complex but they usually fail to capture the intricate nonlinear relationships between the factors and also long-range temporal dynamics characteristic that constantly keeps changing within urban air pollution. RNNs have significantly enhanced time-series forecasting by identifying sequential patterns in data, especially LSTM and GRU models. Nevertheless, GRU-only networks under represent longer temporal interactions, isolated LSTM stacks have fading gradients and the majority of previous studies lack residual mechanisms and bidirectional context transmission, which are crucial for simulating intricate air quality dynamics. The Hybrid LSTM–GRU model, which balances long-term memory retention with training convergence and the Residual BiLSTM model, which employs 1×1 Conv1D skip connections for deeper recurrent computations with past-future context, are two

(Received 10th November 2025; Accepted 15th April 2026; Published 29th April 2026)

deep learning architectures trained on AQI and meteorological datasets. In order to lessen the influence of outliers, both systems employ a 27-feature input space that includes rolling trends, cyclical encodings, rate-of-change metrics, pollutant ratios and the heat-humidity index. These features are processed via quantile transformation. With a validation R^2 of 0.9475, benchmarking against SVR, KNN and Elastic Net demonstrated the supremacy of the Residual BiLSTM, which makes it perfect for real-time AQI alert systems and smart city dashboards. The literature review depicts how sophisticated deep learning models have outperformed conventional machine learning algorithms. Despite challenges with dynamic temporal correlations, early research utilizing Linear Regression and ANFIS shown that AQI forecasting was viable (1). Even in the absence of long-range sequential context, genetically modified back propagation neural networks showed enhanced convergence (2). Using hybrid CNN architectures with Attention Gate Units (AGU), the model enabled dynamic temporal weighting along with spatial temporal feature extraction, high prediction accuracy was achieved based on Chinese city-level data (3). Deep Air, a multi-component deep sequence model that integrated long-term meteorological, pollutant and socio-geographic data, had limited scalability in regions with sparse data but obtained an overall accuracy of 92.6% (4). For multi-city AQI prediction, Transformer-BiLSTM outperformed CNN-GRU and BiLSTM baselines, although it did not use satellite and meteorological data (5). Although prediction performance was enhanced by XGBoost and dual-LSTM models, careful hyper parameter tweaking was required (6). For PM and AQI prediction in Romania, a hybrid IVS-DT and NARMAX model performed better than standalone ML techniques; however, it was restricted to a single-sensor dataset (7). For multi-city AQI prediction, GLA-Net—which included GAT, LSTM and temporal attention—performed better than GCN-LSTM and CNN-LSTM, but it necessitated accurate spatial graph formation (8). Transformer models with multi-head self-attention improved memory bottlenecks and temporal locality (9, 10). In deep sequence models, gradient flow and network stability were enhanced by residual learning and skip connections (11, 12). The inadequacies of typical machine learning models (Elastic Net, KNN and SVR) in mimicking nonlinear temporal AQI dynamics were validated by comparative research (13, 14). Improved accuracy and control of variability were proven by Bayesian-optimized stacked BiLSTM ensembles and cluster-based random forest predictions (15, 16). The improved flexibility of ensemble techniques was proven by thorough examinations of Neural Networks, Random Forests and Decision Trees (17). Hybrid deep models with CNN-LSTM layers with spatiotemporal clustering were used to exhibit prospective multi-hour and multi-site AQI forecasting (18). Models that include reference stations and series decomposition techniques, such CEEMDAN-LSTM, show that pre-processing procedures like error correction and quantile normalization are essential for precise predictions (19, 20).

Combination of deep learning architectures with attention processes has shown a significant promise for more accurate AQI prediction (21). The usefulness of models such as LSTM and BiLSTM with high R^2values for long-term urban air quality forecasting has also been shown (22, 23). When compared to more conventional data-mining methods like K-Nearest Neighbor (KNN) and ANN–KNN hybrid models, deep learning approaches demonstrated better prediction accuracy and generalization (24, 25). For short-term AQI prediction in Patna, EMD-Transformer-BiLSTM performed better than Wavelet-BiLSTM and solo BiLSTM with an RMSE of 5.6853, but it lacked spatiotemporal modelling (26). Hybrid frameworks, including Prophet-LSTM and GRU-based models, have made it possible to forecast in real time and over several horizons in unpredictable environmental circumstances (27, 28). Attention methods combined with two-stage decomposition have been shown to improve resilience, particularly during irregular environmental occurrences, such as COVID-19 lockdown periods (29). For AQI prediction across Indian cities, a Deep GAN-imputed Stacked Attention GRU with KL divergence performed better than SVR and CNN, obtaining R^2 of 0.9479; nevertheless, it was restricted to particular cities without cross-regional validation (30).

Decision Trees outperformed Random Forest and Linear Regression in cloud-based SMOTER research for smart-city AQI predictions (31). A NARX-based recurrent model showed resilience with sparse multi-site inputs (32), while an A3T-GCN utilizing attention, GRU and GCN outperformed TGCN, LSTM and GRU in spatiotemporal NO_2 prediction (33). MasterGNN+ enhanced heterogeneous dependency modelling by multi-adversarial learning (34), while a hybrid LSTM-GRU model outperformed standalone ML and DL models for AQI prediction (35). Using CNN–LSTM–ANN and adaptive neighbour matching, ST-DNN improved 48-hour PM2.5 prediction (36). Through evolutionary optimization, GA-KELM increased efficiency and accuracy (37), while an upgraded PSO-BP avoided local minima and obtained 99.03% accuracy (38). For fine-grained multi-city AQI forecasting, Deep Air presented a deep distributed fusion network that combines heterogeneous urban data with spatial transformation (39), while SSH-GNN enabled fine-grained forecasting for unmonitored regions via hierarchical semi-supervision (40).

Methodology

The data was obtained from a stationary high-frequency air quality monitoring station located in Navi Mumbai, Maharashtra, India. Sea-breeze circulation, high humidity and monsoon-driven seasonal variability all have an impact on Navi Mumbai, a coastal urban metropolitan environment with mixed residential, industrial and vehicular emission sources. This results in nonlinear and no stationary pollution dispersion behaviour, which is typical of many quickly developing cities. Between September 2023 and May 2024, measurements were taken every 15 minutes, resulting in 23,040 temporal observations across around 240 days. The following parameters are monitored: ambient temperature (°C), relative humidity (%), dew point (°C), wet-bulb temperature (°C), heat index (°C), particulate matter concentrations (PM_1, $PM_{2.5}$ and PM_{10} in $\mu g/m^3$) and the resulting composite Air Quality Index (AQI). The chosen variables, despite the fact that observations were taken from a single monitoring station, reflect physically governed atmospheric processes that are universal drivers of urban air quality rather than location-specific traits, such as pollutant accumulation and dispersion, boundary-layer stability and humidity-driven aerosol growth. Instead of learning geographic identifiers or site-specific statistical patterns, the suggested Hybrid LSTM–GRU and Residual BiLSTM architectures learn temporal dependencies and pollutant–meteorology interactions, allowing transferability to other urban environments when comparable environmental variables are available. In addition to capturing intra-day emission cycles and quick weather changes, the precise 15-minute temporal resolution enables the framework to describe basic AQI dynamics rather than site-dependent behaviour, promoting application across many cities.

The process was carried out in many steps to ensure that the data is reliable and useful for learning. A unique method was employed to fill in the missing data: linear interpolation was used if the gap exceeded three hours; the next available value was used if the gap was up to three hours and the average of the surrounding data was used if the gap exceeded six hours. A row was removed if more than 30% of the data was missing. The process employed a variety of techniques to identify values: it employed a unique score to identify uncommon values independently and it employed a unique forest strategy to identify odd values while examining several variables simultaneously. Additionally, the procedure ensured that the data was realistic by utilizing our current understanding of the environment and maintaining the numbers for pollutant and meteorological data within appropriate limits. This approach was employed by the process to retain meteorological and pollutant data based on our knowledge of these topics.

Table 1 presents 27 attributes, comprising pollutant ratios, derived metrics, rolling statistics, lag variables, cyclical encodings and rate-of-change indicators, reflecting temporal, meteorological and pollutant trends. Training stability and resilience to outliers were improved by approximating a normal distribution through quantile-based scaling. Following prediction, the target AQI variable was scaled individually and then inverse-transformed to obtain results in real AQI values.

Table 1: Complete List of 27 Engineered Features for AQI Forecasting

Category	Feature(s)	Rationale
Raw Inputs	AQI, PM2.5, PM10, Temperature, Humidity	Direct sensor readings providing baseline pollutant and meteorological state.
Basic Time	Is Weekend, Is Peak Hour	Captures periodic and categorical temporal effects.
Cyclical Encodings	sin / cos (Hour, Month, Day-of-Week)	Models seasonality while avoiding discontinuities.
Pollutant Ratios	PM2.5/PM10	Indicates dominant emission type (dust vs combustion).
Meteorological Index	Heat-Humidity Index	Integrates temperature and humidity influence.
Lag Features	AQI t−1,2,4,8,24, PM2.5 t−1, PM10 t−1	Temporal autocorrelation structure and persistence.
Rolling Stats	AQI EWMAs (span 4, 12, 24)	Reflects short-term smoothed trends across multiple time horizons.
ROC Features	AQI Rate-of-Change (lag 1, lag 4)	Detect abrupt spikes or transitions.

In order to simulate the physical mechanisms controlling pollution generation, transit and accumulation, the 27 designed characteristics were created using accepted atmospheric science concepts and earlier AQI forecasting literature. Sinusoidal sine-cosine encoding was used to maintain circular continuity, a common technique in environmental time-series modeling, as raw integer encoding of cyclic variables like hour and month produces false discontinuities (e.g., hour 23 and 0). Urban AQI fluctuation in Indian cities is primarily driven by industrial and vehicular emissions, which are influenced by socioeconomic activity patterns captured by binary variables such as is_weekend and is_peak_hour. With values close to unity indicating the preponderance of coarse dust and lower values indicating fine combustion-related particles, the $PM_{2.5}/PM_{10}$ ratio functions as a diagnostic indication of emission sources, allowing source distinction without specific apportionment data. Wet-bulb and dew point depression quantify evaporative cooling and atmospheric dryness—important factors influencing aerosol hygroscopic growth, particle size distribution, optical properties and PM-based AQI estimation—while the heat-humidity index represents nonlinear interactions between temperature and moisture affecting boundary layer stability and vertical mixing.

The known temporal autocorrelation of particulate matter is mirrored by lag features at intervals of 15 minutes, 30 minutes, 1 hour, 2 hours and 6 hours. These lag features span the range of pollutant persistence timescales found in urban mixed layers, from boundary layer stabilization and daily repetition cycles to near-instantaneous carryover effects. In addition to this, multi-scale accumulation patterns are captured using exponentially weighted moving averages across 1-, 3- and 6-hour periods, which filter out short-term noise while remaining sensitive to extended pollution events. The onset velocity of rapid pollution transitions, such as abrupt drops associated with rainfall scavenging events, sea-breeze reversals that trap pollutants near the surface and abrupt traffic surges during peak commuting hours, is identified by rate-of-change features computed at 1-step and 4-step differences. These phenomena are all typical of Navi Mumbai's coastal urban meteorology. Together, the 27-feature engineering framework operationalizes domain knowledge from urban climatology and atmospheric chemistry into a structured input representation that is tailored to the research region's environmental and temporal features.

The long-term dependency modelling of LSTMs and the computational efficiency of GRUs are combined in the hybrid LSTM–GRU model. Multivariate shape sequences of batch size 48 (24-time steps, 27 features) comprise the input. Following an LSTM block with 128 units and tanh activation, the architecture includes two GRU blocks with 128 and 64 units, dropout and recurrent dropout. After applying dense layers with batch normalization and ReLU activations, the final result is applied. While Huber loss was used as the objective function, early stopping, learning rate drop on plateau and model checkpoints ensured optimal convergence and minimized over fitting as shown in Figure 1.

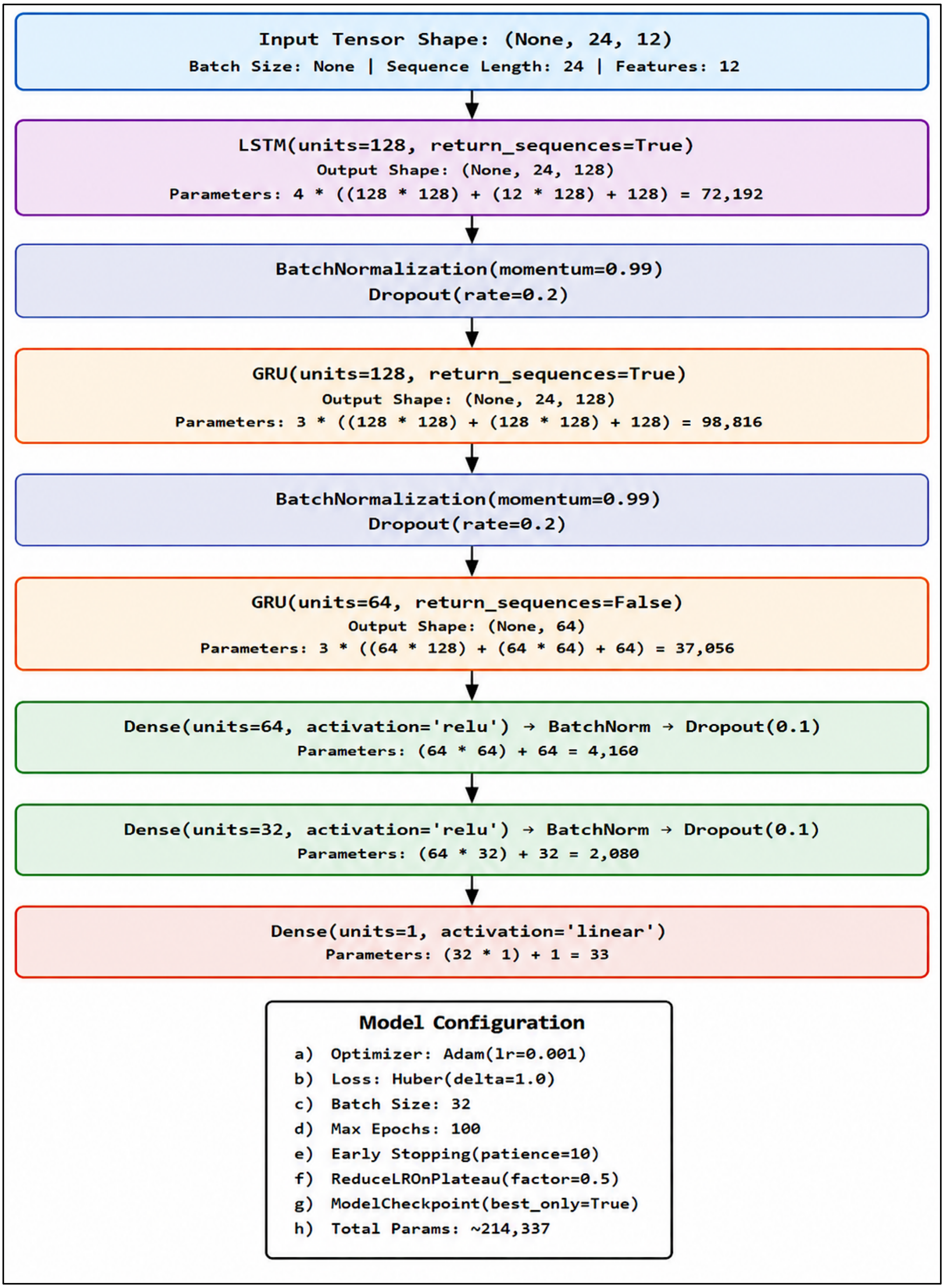

Figure 1: Architecture of the Hybrid LSTM–GRU Model

The Residual BiLSTM incorporates Conv1D skip connections to improve gradient flow and enable deeper temporal modelling. The input format is (24-time steps, 27 features). The model consists of an initial Conv1D layer, dense layers with L1–L2 regularization, a bidirectional LSTM block with 64 units per direction and residual merging using Conv1D and batch normalization. The model is trained using gradient clipping and the Adam optimizer with Huber loss. Early stopping and variable learning rate scheduling were employed to ensure successful convergence as represented in Figure 2.

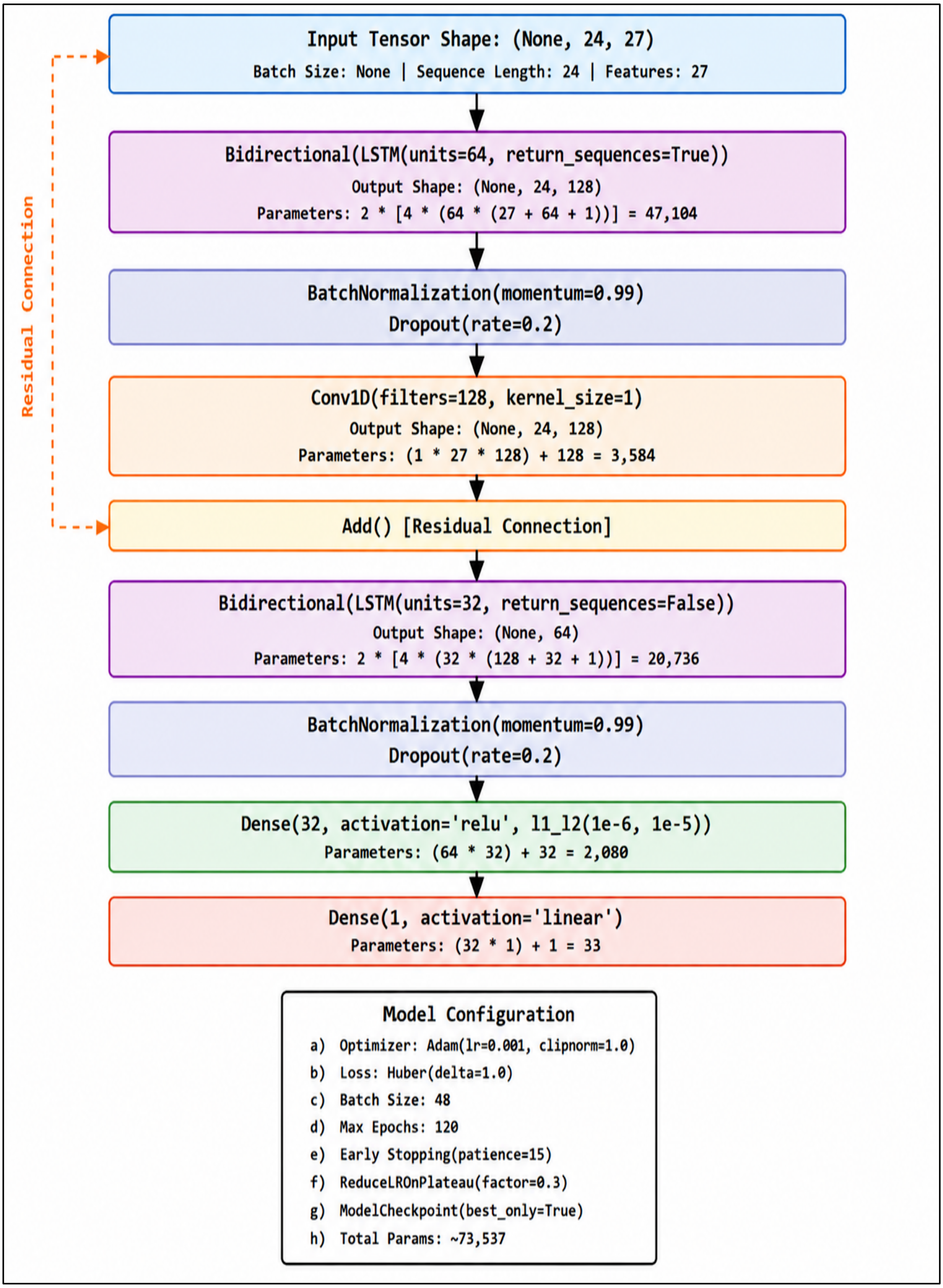

Figure 2: Architecture of the Residual BiLSTM Model

The model used the Adam optimizer with a learning rate of 0.001, default hyper parameters ($\epsilon = 10^{-8}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$) and gradient clipping with a clip norm of 1.0 to guarantee stable training and avoid exploding gradients. Because of its hybrid penalty behaviour, which combines MSE for small residuals and MAE for larger deviations to ensure consistent and dependable convergence and its resilience to outliers, Huber loss was chosen as the objective function. The Huber loss formulation is presented in Equation [1]:

$$L(y,\hat{y}) = \begin{cases} \frac{1}{2}(y-\hat{y})^2 & \text{if } |y-\hat{y}| \leq \delta \\ \delta|y-\hat{y}| - \frac{1}{2}\delta^2 & \text{otherwise} \end{cases} \quad [1]$$

Both models were trained for up to 120 epochs with a batch size of 48. Early stopping was started if the validation loss did not drop by more than 1×10^{-2} for 15 consecutive epochs. After eight plateau epochs, learning rates were reduced by a factor of 0.3 with a minimum limit of 1×10^{-6}. Evaluation metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE) and the Coefficient of Determination (R^2) are used in order to quantify predictive performance of the models that helps to depict accuracy and error sensitivity.

The mean magnitude of the absolute errors between anticipated and observed AQI values is represented by MAE, which treats all prediction mistakes equally regardless of size, as seen in Equation [2].

MSE is helpful for pointing up significant mistakes since it computes the mean of squared prediction errors and penalizes big deviations more severely; it is expressed in Equation [3].

RMSE, which is the square root of MSE, maintains sensitivity to significant deviations while offering an understandable error figure in the same unit as AQI, as shown in Equation [4].

The R^2coefficient, which indicates quality of fit, quantifies the percentage of variance in real AQI values explained by the model and is represented by the equation [5].

$$\text{MAE} = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \quad [2]$$

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2 \quad [3]$$

$$\text{RMSE} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \quad [4]$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2} \quad [5]$$

The entire AQI prediction process is schematically depicted in Figure 3. Following pre-processing, feature engineering and quantile scaling of the raw data, the Hybrid LSTM-GRU and Residual BiLSTM models are used for sequential modelling. The prediction performance and suitability of the models for real-time AQI application monitoring were assessed using regression metrics and visualizations.

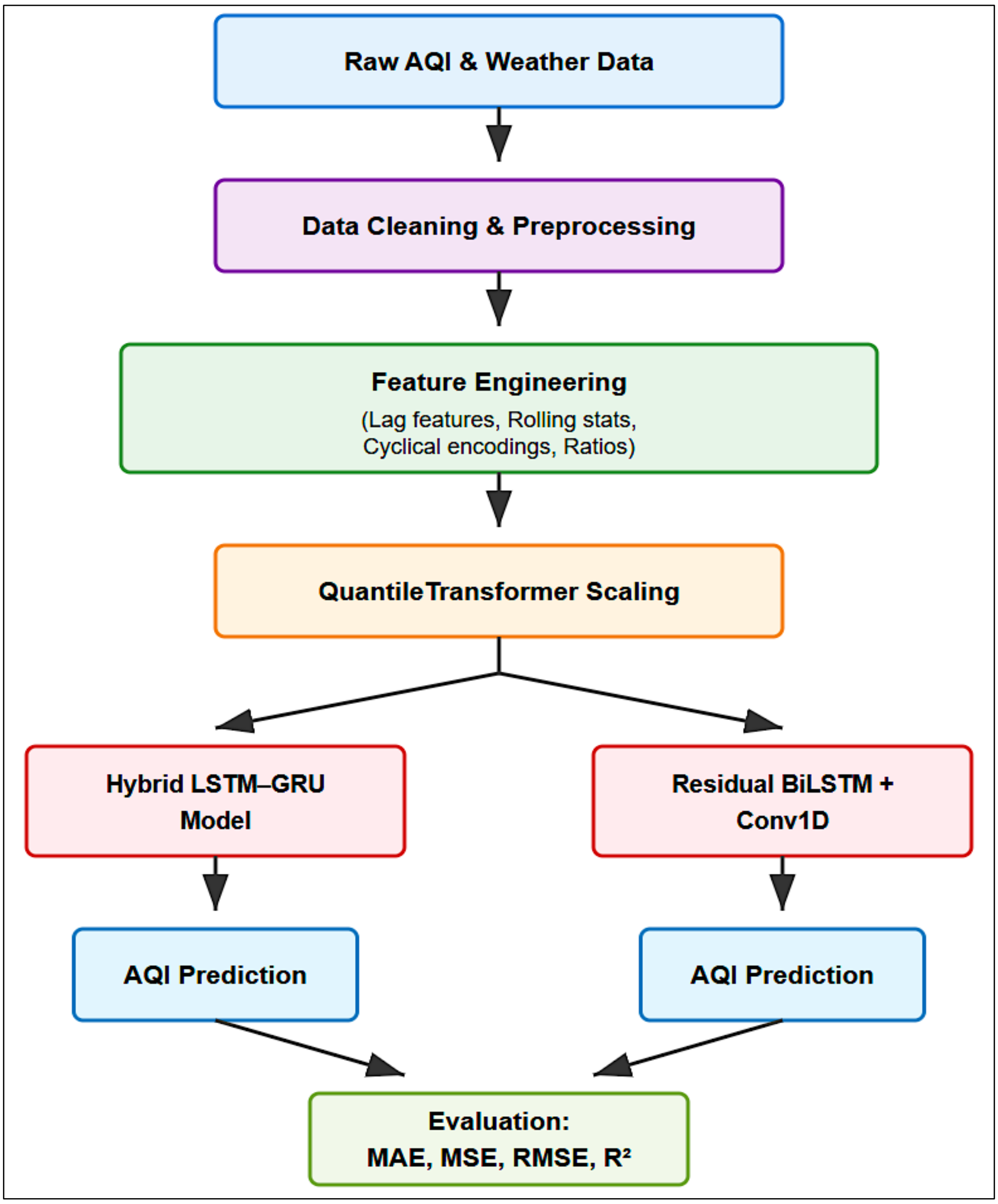

Figure 3: Complete System Flow of AQI Forecasting Process

Results and Discussion

To assess the suggested deep learning models capacity for prediction, the dataset underwent extensive training and validation. Several sophisticated pre-processing methods, such as quantile-based normalization, anomaly correction and imputation of the missing value, were used to ensure data consistency and robustness. The dataset was split into 80:20 ratios for training and validation when the pre-processing phase was finished. The model architectures were trained under controlled settings, such as early stopping and reducing the learning rate on plateau, to improve the model's convergence and prevent over fitting. Standard metrics including Coefficient of Determination (R^2), Mean Absolute Error (MAE), Mean Squared Error (MSE) and Root Mean Squared Error (RMSE) were used to assess the model's correctness and performance.

Three of the most widely used regression models in machine learning were chosen and put into practice. These are Elastic Net, K-Nearest Neighbours (KNN) and Support Vector Regression (SVR). These models were chosen because, after the hyper parameter adjustment procedure, the conventional models were applied and verified on the identical datasets. The SVR model offers kernel-based margin optimization, whereas the KNN model is an instance-based local learner.

Additionally, the Elastic Net model may address multi co linearity as it combines L1 and L2 regularization techniques. KNN was able to give the best validation performance amongst all of the other machine learning models with an R^2 value of 0.81 surpassing the results of SVR and Elastic Net. Table 2 depicts the performance of machine learning (SVR, KNN, Elastic Net) models based on their MAE, RMSE and R^2 scores.

Table 2: Validation Performance of Baseline ML Models

Model	MAE	RMSE	R^2
SVR	29.40	39.91	0.75
KNN	25.23	34.96	0.81
ElasticNet	28.64	35.43	0.75

Forecasting the Air Quality Index is a nonlinear and non-stationary time-series issue that is impacted by a number of interrelated atmospheric factors, such as emission variability, boundary-layer mixing, humidity-dependent aerosol development and particulate matter temporal persistence. In urban pollution environments with rapid concentration variations, daily emission cycles and climatic disturbances, the requirements of linearity and weak stationary that underpin classical time-series techniques like ARIMA and SARIMA are rarely met. Similar to this, single recurrent architectures, such as solo LSTM or GRU models, only work on a small temporal scale and have a tendency to either miss short-term oscillations or forget long-range dependencies. The time-ordered persistence of AQI dynamics, pollutant carry-over effects and temporal autocorrelation are all crucial for precise high-frequency forecasting, but ensemble tree-based techniques like XGBoost and Random Forest are inherently non-sequential and treat each input sample independently, even though they can model nonlinear feature interactions. The Hybrid LSTM–GRU architecture is responsible for addressing this multi-scale temporal behaviour due to the combination of GRU gating for short-term variability and LSTM memory cells for long-term persistence. By using residual skip connections and bidirectional temporal contexts to support deep sequential learning, the Residual BiLSTM avoids needless complexity by basing its design on the physical temporal patterns of air pollution dynamics.

The benchmarking was done using data from conventional regressors, which demonstrated the need for the created architectures. The R^2 values were 0.75, 0.81 and 0.75 with greater errors (MAE: 25.23-29.40 and RMSE: 34.96-39.91) for SVR, KNN and Elastic Net regressors and 0.9045 as well as0.9475 for Hybrid LSTM-GRU and Residual BiLSTM models. The inferior performance of traditional regressor models arises because they rely too much on instantaneous feature relationships and also due to their inability to represent temporal autocorrelation, delayed pollutant effects and also meteorology-driven persistence. These results demonstrate that simpler statistical and shallow learning approaches under fit AQI dynamics while on the other hand the proposed recurrent deep learning architectures are able to properly capture the temporal dependencies that are nonlinear and also interactions between pollutant–meteorology, leading to substantially improved generalization of these models while predicting.

Two complex deep learning-based architectures were investigated to evaluate the efficiency and their performance in AQI prediction. One was a hybrid LSTM–GRU model and the second was a residual BiLSTM model with skip-connected Conv1D layers. Both of the models were trained as well as assessed using data normalization under same conditions which helped to ensure a proper as well as a fair comparison. The Hybrid LSTM–GRU model made a remarkable performance based on its short-term prediction capabilities and also offered the lowest MAE score of 0.0288. However, the Residual BiLSTM model performed better when it came to better generalization while obtaining a higher R^2 value of 0.9475 which depicts the model's stability over varied values of the AQI data and abrupt spikes and data fluctuations. The Residual BiLSTM model was ultimately selected for deployment because of its capability of better generalization, its ability to capture complex temporal relations and also by offering better and maintained projected accuracy over extended periods of time. The final evaluation scores of both the proposed deep learning models which are assessed using the MAE, MSE, RMSE and R^2 are showcased in Table 3.

Table 3: Validation Performance of Deep Learning Models (Normalized Scale)

Model	MAE	MSE	RMSE	R^2
Hybrid LSTM–GRU	0.0288	0.0034	0.0582	0.9045
Residual BiLSTM	0.0982	0.0210	0.1449	0.9475

The collected findings were compared with previously published AQI forecasting research using statistical, machine learning and deep learning techniques in order to assess the predictive power of the suggested framework. On an Indian urban dataset (1), Linear Regression and ANFIS reported $R^2 = 0.93$ with RMSE = 18.41, while improved back-propagation neural networks achieved roughly 80.44% accuracy with $R^2 = 0.85$ on Chinese city data (2). The neuro-fuzzy approach and the classic regression method's fixed functional forms limit their ability to capture nonlinear temporal dynamics, long-term relationships and abrupt changes in AQI, particularly under extremely variable air quality circumstances. Nonetheless, the methods show some proficiency, although in rough forecasting.

Modern deep learning models are very accurate, but they frequently depend on geospatial information and dense multi-station infrastructure, which are not accessible in many developing cities (3, 4). For example, Deep Air obtained MAE = 8.23 across 302 cities and a CNN with an Attention Gate Unit reported $R^2 = 0.9872$ (MAE = 6.74), although both rely on substantial geographic and socio-spatial inputs. On the other hand, the suggested framework lacks geographical identifiers and solely employs meteorological and particulate matter variables that are universally quantifiable. The Residual BiLSTM obtained $R^2 = 0.9475$ and the Hybrid LSTM–GRU $R^2 = 0.9045$ using physics-informed temporal feature engineering, learning pollutant–meteorology interactions and temporal carry-over effects without the need for dense sensor networks.

The residual bidirectional architecture captures intricate temporal relationships while being practically deployable in resource-constrained urban environments, in contrast to conventional statistical models that under fit nonlinear AQI dynamics (1, 2) and spatial deep-learning systems that require extensive infrastructure (3, 4). The outcomes show competitive performance with sophisticated spatial models and a definite improvement over traditional approaches with significantly less reliance on data and infrastructure.

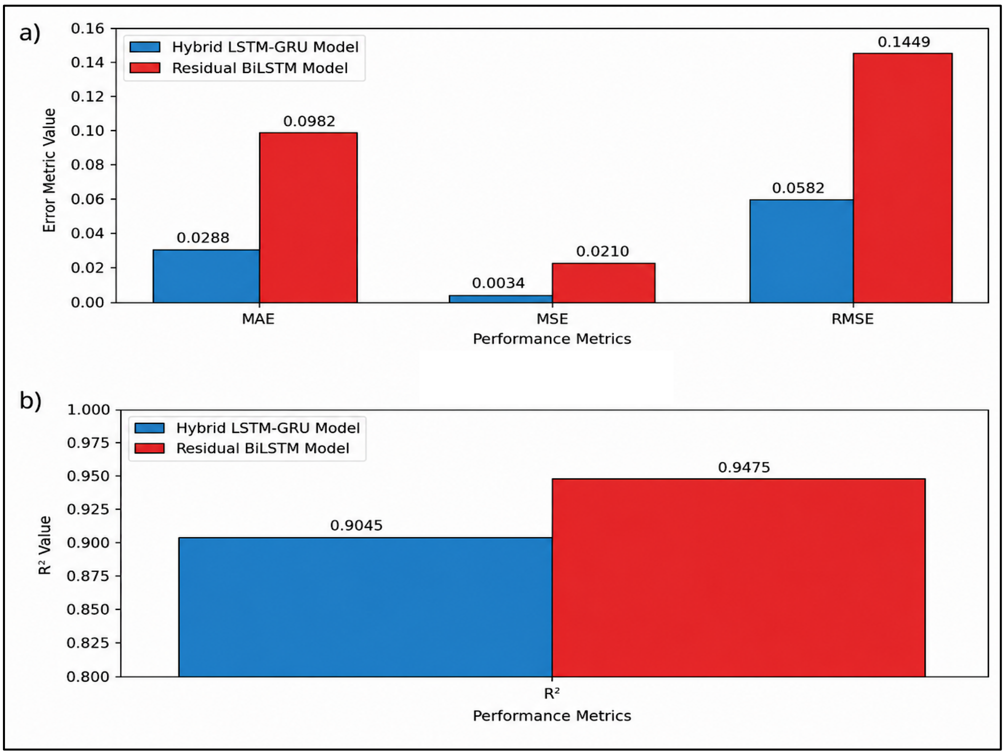

Figure 4: Performance Comparison between Hybrid LSTM–GRU and Residual BiLSTM Models Across: a) MAE, MSE, RMSE and b) R^2

Figure 4 shows a visual comparison of the two deep learning models for every performance criterion. Even though the Hybrid LSTM–GRU model achieved lower short-term errors but the Residual BiLSTM model showcased an improved long-term accuracy as well as generalization which is represented by higher R^2 values and better handling of substantial deviations are represented by using the RMSE and MSE metrics. This illustrates the Residual BiLSTM's suitability for its usage in the dynamic air quality conditions where the interactions between the non-linear pollutant and rapid variations are typical. The model's depth as well as skip linkages improves the adaptability along with robustness in actual forecasting of AQI.

The loss curves in Figures 5 and 6 show that the Hybrid LSTM–GRU model was smoothly converged with very few oscillations whereas the Residual BiLSTM model displayed high volatility during the training but later ultimately achieved better and stronger validation generalization. The Residual BiLSTM model was also very closely aligned with the actual trends within the AQI. This indicates that the model was effectively adapting to the abrupt changes within the environment and it also proves to be better at capturing the complex nonlinear temporal dynamics present in the data. This makes the model more suitable for a real time forecasting related to public health as well as within the alert systems where rapid as well as accurate predictions are very important, whereas the Hybrid LSTM-GRU model responded more slowly to the sudden fluctuations within the data.

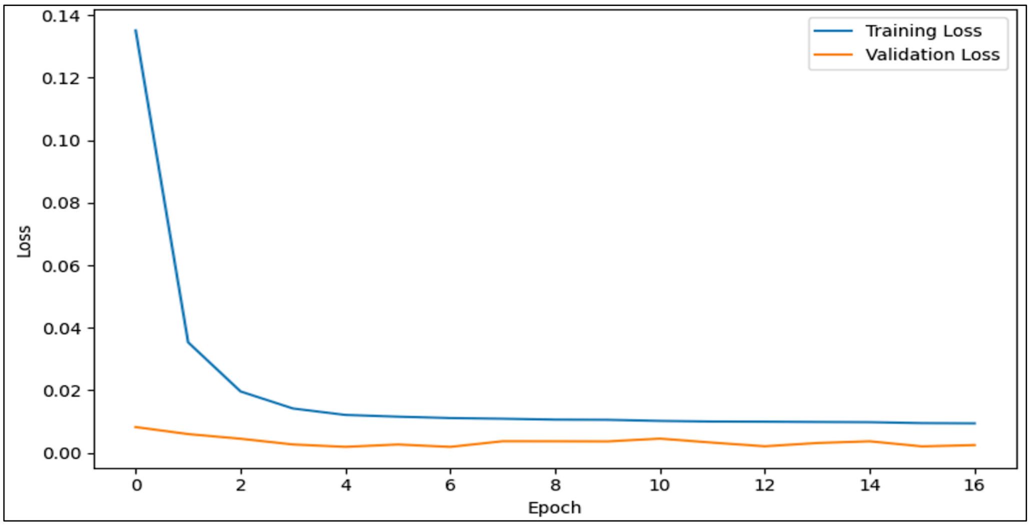

Figure 5: Training and Validation Loss for Hybrid LSTM–GRU

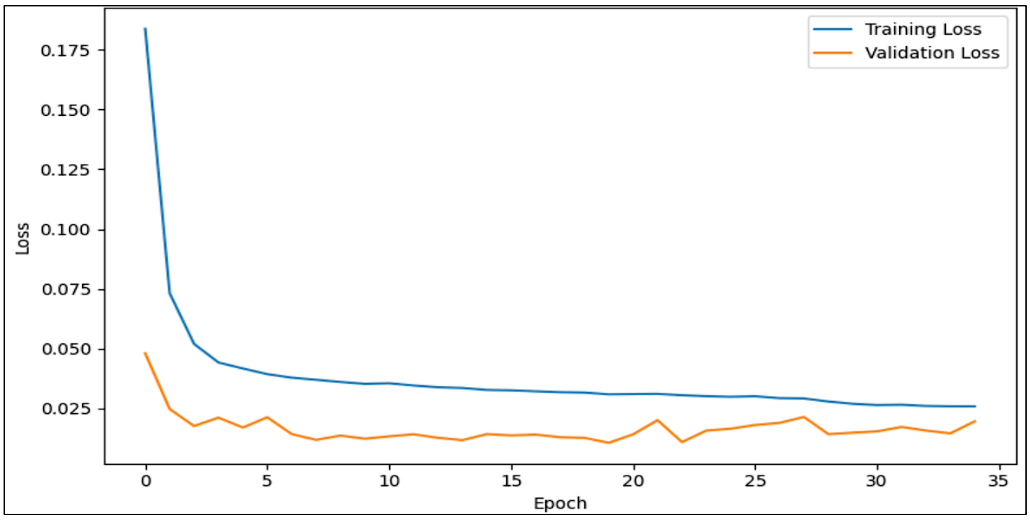

Figure 6: Training and Validation Loss for Residual BiLSTM

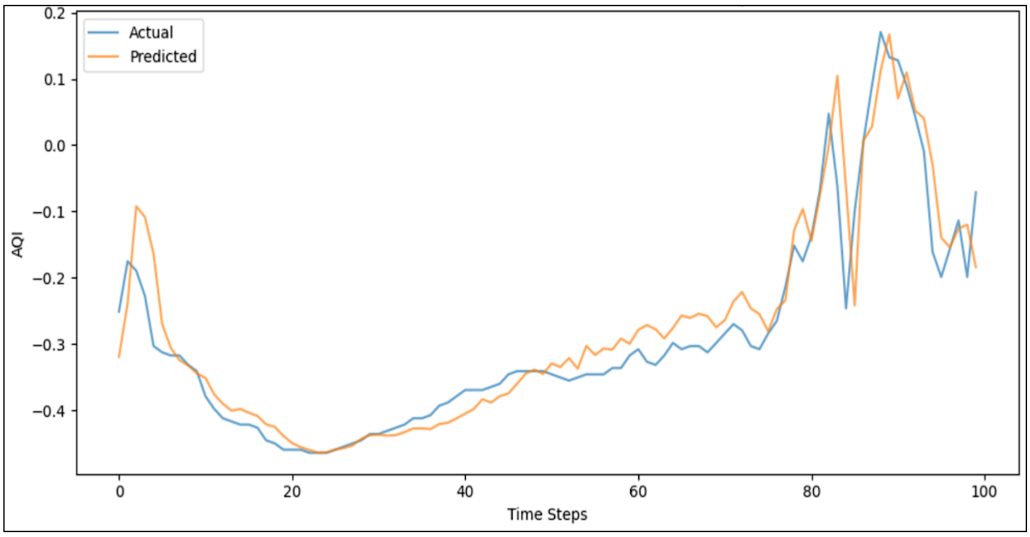

Figure 7: Actual vs Predicted AQI – Hybrid LSTM–GRU

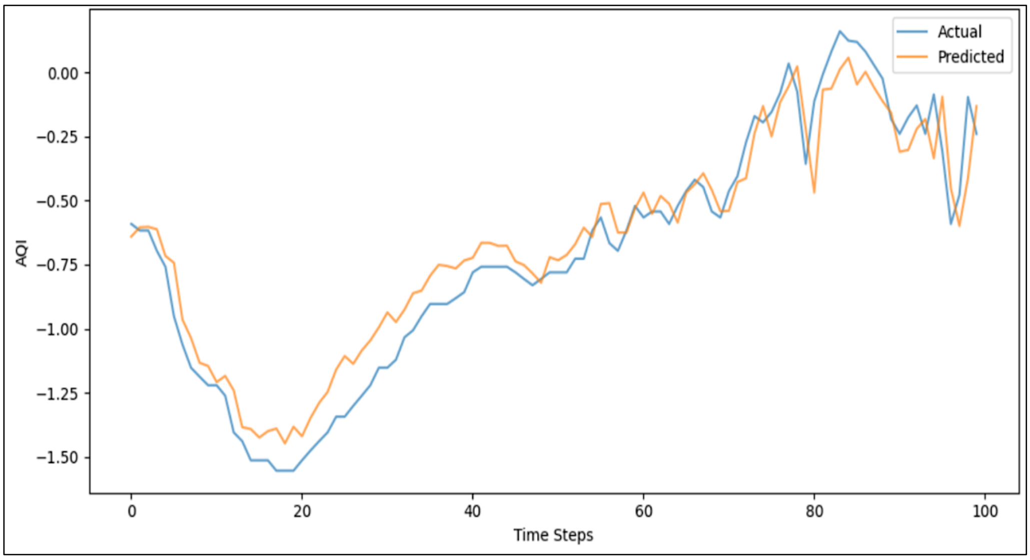

Figure 8: Actual vs Predicted AQI – Residual BiLSTM

The quality of the predictions made by both of the deep learning models were visually evaluated by displaying the real vs predicted AQI trends. The Residual BiLSTM model is very close to the ground truth which can be seen within Figures 7 and 8, specifically when the AQI is changing quickly, especially during the time of spikes as well as sudden drops. This flexibility demonstrates the ability of the model to capture all the non-linear temporal connections as well as react to the abrupt changes within the environment. Conversely, the Hybrid LSTM-GRU model shows somewhat less sensitivity to sudden temporal shifts and occasionally delays or smoothes abrupt transitions. Since timely and also accurate forecasts are very much important for environ-mental warnings as well as public health planning, the Residual model's enhanced temporal precision makes it ideal for real-world monitoring of AQI.

For the purpose of deployment, the Residual BiLSTM model was developed and optimized to a Tensor Flow footprint which is around 6 MB, having a more than 0.5 million parameters that are trainable. After ONNX conversion by using 8-bit quantization, the size of the model was reduced down to under 1.5 MB, this makes it more suitable for embedded devices as well as portable AQI sensors. Apart from this it also supports a low latency inference and both streaming as well as batch predictions which makes use of fixed 24-step input tensors. Two deployment options are available for scalable smart city applications: edge inference on hardware using ONNX Runtime or Tensor Flow Lite Micro and containerized backend services via lightweight Docker containers. The

Residual BiLSTM model is ready for practical AQI monitoring in the real world owing to these features, which provide constant inference speed and minimal resource usage.

According to the study, if deep learning architectures are built correctly and trained on relevant environmental features, they can accurately replicate the temporal connections in AQI data. Both the Residual BiLSTM and Hybrid LSTM–GRU models show a notable improvement over traditional machine learning baseline. While the Hybrid LSTM–GRU model specializes in short-term-based predictions along with reduced latency, the Residual BiLSTM model facilitates superior long term-based generalization as well as stability. The ability of the Residual BiLSTM model which captures complex temporal patterns and also the interactions between the non-linear pollutants are improved by making use of residual connections, bi directionality and skip-connected Conv1D layers, within the architecture makes it more reliable and also a dependable tool for predicting air quality within the urban regions and environmental based decision making.

The framework is purposefully designed to model general atmospheric processes rather than location-specific behavior, even though the suggested models were trained and assessed using data from a single monitoring station in Navi Mumbai. The primary physical causes of urban air pollution are represented by all of the input variables, which include temperature, relative humidity, dew point, wet-bulb temperature, PM_1, $PM_{2.5}$, PM_{10} ($\mu g/m^3$) and heat index. These are common environmental factors that are regularly gathered by global monitoring networks. Since the mechanisms of emission generation, atmospheric dispersion, boundary mixing, humidity-driven aerosol growth and short-term particulate persistence are among the universal processes governing the urban AQI, the predictors directly relate to these processes, allowing the models to learn the basic interactions between pollutants and meteorology rather than city-specific patterns. By eliminating spatial coordinates, land use information and geographic information, the Hybrid LSTM-GRU and Residual BiLSTM designs are able to capture transferable temporal patterns such as pollution carry-over effects, diurnal emission patterns and weather-driven dispersion effects. Instead of connecting the forecasts to a specific area, it concentrates on the pollution's temporal patterns. Because of the 15-minute temporal resolution and the quantile normalization method for normalizing the input range, it is also resilient to the variation in baseline pollution levels among cities. Additionally, it can correctly capture patterns such as those brought on by traffic patterns, atmospheric stability at night and the quick impacts of weather in metropolitan settings. As a result, the trained framework operates as a process-driven predictor instead of a location-dependent regressor, proving that it can be applied to other urban areas with comparable environmental variables. Transfer learning techniques can also be used to maximize performance in different emission and climate scenarios.

Conclusion

By making use of deep learning models which are trained on real atmospheric as well as meteorological data, this study showcases a comprehensive and also practical approach in order to estimate the Air Quality Index (AQI) by ensuring accuracy as well as applicability. The creation and comprehensive assessment of two neural network architectures which are a Residual BiLSTM with Conv1D skip connections model and a Hybrid LSTM–GRU model showcases that the deep learning approaches have significantly outperformed traditional machine learning techniques in prediction of the AQI.

In each of the validation metric, the Residual BiLSTM model has consistently performed better than the other models. The best working model which is Residual BiLSTM model gives the results as follows: Root Mean Squared Error (RMSE) was 0.1449, the Mean Absolute Error (MAE) was 0.0982, the Mean Squared Error (MSE) was 0.0210 and the R^2score was 0.9475. These results demonstrate that the model was able to successfully capture both the gradual accumulation of pollutants as well as environmental changes that are abrupt, resulting in better long-term predictions. Because of its bidirectional architecture along with a combination of skip connections, the model was easily able to maintain and store crucial temporal information across various layers, also the model was able to learn long-range correlations and then can also generate reliable results.

Overall, the deep learning technique that is proposed establishes a practical as well as an efficient forecasting system for AQI. It is not only accurate but also contextually sensitive projections help to support deployment within real-time urban settings, this helps to facilitate timely management of environment, decision-based policy formulation and public health alarms.

Abbreviations

AQI: Air Quality Index, BiLSTM: Bidirectional Long Short-Term Memory, Conv1D: One-Dimensional Convolution, CPCB: Central Pollution Control Board, GRU: Gated Recurrent Unit, LSTM: Long Short-Term Memory, ONNX: Open Neural Network Exchange, $PM_{2.5}$: Particulate Matter < 2.5 µm, PM_{10}: Particulate Matter ≤ 10 µm, RNN: Recurrent Neural Network.

Acknowledgment

We thank all our colleagues with full heartfelt gratitude for their support in the completion of this research work.

Author Contributions

All authors have equally contributed.

Conflict of Interest

The authors declare that there is no conflict of interest regarding the content of this article.

Data Availability

The data was collected from an air quality monitoring station located in Navi Mumbai, India, which continuously records both meteorological and air quality parameters. The data is continuously available and accessible.

Declaration of Artificial Intelligence (AI) Assistance

Generative AI tools were used only for language refinement; all scientific content and conclusions are the authors' own.

Ethics Approval

Not Applicable.

Funding

None.

References

1. Sunori SK, Verma D, Negi PB, Juneja P. Air Quality Index prediction using linear regression and ANFIS. In: Proceedings of the 7th International Conference on Inventive Computation Technologies (ICICT); 2024; Nepal. p. 1929-1933. doi:10.1109/ICICT60155.2024.10544842
2. Wang Z, Tian Z. Prediction of air quality index based on improved neural network. In: Proceedings of the 2017 International Conference on Computer Systems, Electronics and Control (ICCSEC); 2017; Dalian, China. p. 200-204. doi:10.1109/ICCSEC.2017.8446883
3. Wang J, Huang M, Jin L, Li X, He S, Wang H. A hybrid air quality index prediction model based on CNN and attention gate unit. IEEE Access. 2022;10: 113343-113354. doi:10.1109/ACCESS.2022.3217242
4. Yi X, Duan Z, Li T, Li R, Zhang J, Zheng Y. Predicting fine-grained air quality based on deep neural networks. IEEE Trans Big Data. 2022;8(5):1326-1339. doi:10.1109/TBDATA.2020.3047078
5. Liu X, Su K, Wang S, Ghazali KH. Intelligent prediction of air quality index based on the transformer-BiLSTM model. Sci Rep. 2025;15: 41838. doi:10.1038/s41598-025-25865-w
6. Chen H, Guan M, Li H. Air quality prediction based on integrated dual LSTM model. IEEE Access. 2021; 9:93285-93297. doi:10.1109/ACCESS.2021.3093430
7. Udristioiu MT, El Mghouchi Y, Yildizhan H. Prediction, modelling and forecasting of PM and AQI using hybrid machine learning. J Clean Prod. 2023;421:138496. doi:10.1016/j.jclepro.2023.138496
8. Li D, Han H, Yu H, Wang J, Liu M, Zou G, Wang L. Spatiotemporal graph convolutional attention network for air quality index prediction of Beijing, Shanghai and Shenzhen. Atmosphere. 2025;16(12): 1314. doi:10.3390/atmos16121314
9. Li S, Jin X, Xuan Y, Zhou X, Chen W, Wang YX, Yan X. Enhancing the locality and breaking the memory bottleneck of Transformer on time series forecasting. Advances in Neural Information Processing Systems; 2019. doi:10.48550/arXiv.1907.00235
10. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, Kaiser L, Polosukhin I. Attention is all you need. Advances in Neural Information Processing Systems; 2017. doi:10.48550/arXiv.1706.03762
11. He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition; 2016. doi:10.48550/arXiv.1512.03385
12. van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A, Kavukcuoglu K. WaveNet: a generative model for raw audio. arXiv. 2016;1609.03499. doi:10.48550/arXiv.1609.03499
13. Lawande S, Gavali VN, Sutar I, Nehe R. Comparative analysis of machine learning models for prediction of Air Quality Index. In: 2024 International Conference on Intelligent Systems and Advanced Applications; 2024. p. 1-5. doi:10.1109/ICISAA62385.2024.10829167

14. Gupta S, Gajare S, Moledina A, Kate M, Athavale S. Air quality prediction using machine learning: a comparative study. In: 2023 6th International Conference on Advances in Science and Technology; 2023. p. 485-489. doi:10.1109/ICAST59062.2023.10454930
15. Zhang C, Yuan D. Fast fine-grained air quality index level prediction using random forest algorithm on cluster computing of Spark. In: UIC-ATC-ScalCom-CBDCom-IoP 2015; 2015:929-934. doi:10.1109/UIC-ATC-ScalCom-CBDCom-IoP.2015.177
16. Manikandan K, Rajkumar R, Jayanthi P, Jothimani S, Venkatesh R, Govinda Rao S. An intelligent Bayesian optimization with stacked BiLSTM model for air quality index prediction. In: 5th International Conference on Inventive Research in Computing Applications (ICIRCA 2023); 2023:1699-1704. doi:10.1109/ICIRCA57980.2023.10220650
17. Parameshachari BD, Siddesh GM, Sridhar V, Latha M, Khalid Nazim Abdul Sattar, Manjula G. Prediction and analysis of air quality index using machine learning algorithms. In: 2022 IEEE International Conference on Data Science and Information System (ICDSIS); 2022: 1-6. doi:10.1109/ICDSIS55133.2022.9915802
18. Yan R, Liao J, Yang J, Sun W, Nong M, Li F. Multi-hour and multi-site air quality index forecasting in Beijing using CNN, LSTM, CNN-LSTM and spatiotemporal clustering. Expert Syst Appl. 2021; 169:114513. doi:10.1016/j.eswa.2020.114513
19. Wang SJ, Huang BJ, Hu MH. A deep learning-based air quality index prediction model using LSTM and reference stations: a real application in Taiwan. In: 33rd International Telecommunication Networks and Applications Conference (ITNAC); 2023:1-6. doi:10.1109/ITNAC59571.2023.10368496
20. Cao J, Bhatti UA, Hasnain A, Feng S, Huang M. Air quality index predictions with a hybrid forecasting model: combining series decomposition and deep learning techniques. In: 6th IEEE International Conference on Pattern Recognition and Artificial Intelligence (PRAI); 2023:1-6. doi:10.1109/PRAI59366.2023.10332062
21. Kim D, Han H, Wang W, Kang Y, Lee H, Kim HS. Application of deep learning models and network method for comprehensive air-quality index prediction. Appl Sci. 2022;12(13):6699. doi:10.3390/app12136699
22. Wu Q, Lin H. Daily urban air quality index forecasting based on variational mode decomposition, sample entropy and LSTM neural network. Sustain Cities Soc. 2019;50:101657. doi:10.1016/j.scs.2019.101657
23. Janarthanan R, Partheeban P, Somasundaram K, Elamparithi PN. A deep learning approach for prediction of air quality index in a metropolitan city. Sustain Cities Soc. 2021;67:102720. doi:10.1016/j.scs.2021.102720
24. Sahani M, Singh HK, Kumar V. Air quality index prediction for Gorakhpur city using k-nearest neighbors: model evaluation and analysis. World J Adv Res Rev. 2024;23(2):444-454. doi:10.30574/wjarr.2024.23.2.2373
25. Baran B. Air quality index prediction in Besiktas district by artificial neural networks and k-nearest neighbors. J Eng Sci Des. 2021;9(1):52-63. doi:10.21923/jesd.671836
26. Dong J, Zhang Y, Hu J. Short-term air quality prediction based on EMD-transformer-BiLSTM. Sci Rep. 2024;14(1):20513. doi:10.1038/s41598-024-67626-1
27. Zhou L, Chen M, Ni Q. A hybrid Prophet-LSTM model for prediction of air quality index. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI); 2020. p. 595-602. doi:10.1109/SSCI47803.2020.9308543
28. Furizal F, Ma'arif A, Suwarno I, Masitha A, Aulia L, Sharkawy AN. Real-time mechanism based on deep learning approaches for analyzing the impact of future timestep forecasts on actual air quality index of PM10. Results Eng. 2024;24:103434. doi:10.1016/j.rineng.2024.103434
29. Li Y, Li R. A hybrid model for daily air quality index prediction and its performance in the face of impact effect of COVID-19 lockdown. Process Saf Environ Prot. 2023;176:673-684. doi:10.1016/j.psep.2023.06.021
30. Binbusayyis A, Khan MA, Ahmed MMA, Emmanuel WRS. A deep learning approach for prediction of air quality index in smart city. Discover Sustainability. 2024;5:89. doi:10.1007/s43621-024-00272-9
31. Al-Eidi S, Tashtoush Y, Amsaadi F, Darwish O, Alqahtani A, Niveshitha N. Comparative analysis study for air quality prediction in smart cities using regression techniques. IEEE Access. 2023;11: 115140-115149. doi:10.1109/ACCESS.2023.3323447
32. Zhou Y, Perera C, De S, Ewa G, Moessner K. Data-driven air quality characterization for urban environments: a case study. IEEE Access. 2018;6:77996-78006. doi:10.1109/ACCESS.2018.2884647
33. Iskandaryan D, Ramos F, Trilles S. Graph neural network for air quality prediction: a case study in Madrid. IEEE Access. 2023;11:2729-2742. doi:10.1109/ACCESS.2023.3234214
34. Han J, Liu H, Zhu H, Xiong H. Kill two birds with one stone: a multi-view multi-adversarial learning approach for joint air quality and weather prediction. IEEE Trans Knowl Data Eng. 2023; 35(11):11515-11528. doi:10.1109/TKDE.2023.3236423
35. Sarkar N, Gupta R, Keserwani PK, Govil MC. Air Quality Index prediction using an effective hybrid deep learning model. Environ Pollut. 2022;315:120404. doi:10.1016/j.envpol.2022.120404
36. Soh PW, Chang JW, Huang JW. Adaptive deep learning-based air quality prediction model using the most relevant spatial-temporal relations. IEEE Access. 2018;6:38186-38198. doi:10.1109/ACCESS.2018.2849820
37. Liu C, Pan G, Song D, Wei H. Air quality index forecasting via genetic algorithm-based improved extreme learning machine. IEEE Access. 2023;11: 67086-67097. doi:10.1109/ACCESS.2023.3291146

38. Huang Y, Xiang Y, Zhao R, Cheng Z. Air quality prediction using improved PSO-BP neural network. IEEE Access. 2020;8:99346-99353. doi:10.1109/ACCESS.2020.2998145
39. Yi X, Zhang J, Wang Z, Li T, Zheng Y. Deep distributed fusion network for air quality prediction. Proc ACM SIGKDD Int Conf KnowlDiscov Data Min. 2018. p. 965-973. doi:10.1145/3219819.3219822
40. Han J, Liu H, Xiong H, Yang J. Semi-supervised air quality forecasting via self-supervised hierarchical graph neural network. IEEE Trans Knowl Data Eng. 2023;35(5):5230-5243. doi:10.1109/TKDE.2022.3149815

How to Cite: Kothari SS, Chhabria A, Phadke G. Air Quality Index Prediction Using Hybrid LSTM–GRU and Residual BilSTM Models. Int Res J Multidiscip Scope. 2026; 7(2): 1747-1762.
DOI: 10.47857/irjms. 2026.v07i02.09113