International Journal of Applied Mathematics

Volume 38 No. 12s, 2025

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

AIR QUALITY PREDICTION USING DEEP LEARNING MODELS: DETAILED OVERVIEW

Mr. Kshirsagar Sopan Bapu[1] Dr. Rais Abdul Hamid Khan[2]

[1]PhD scholar School of Computer Science and Engineering Sandip University.

Nashik, India

Email id, sopankshirsagar02@gmail.com[1]

[2]Professor, (CSE) SOCSE, Sandip University, Nashik, India

Email id,,rais.khan@sandipuniversity.edu.in[2]

Abstract

The goal of a real-world application of air quality forecasting systems is to provide a real-time platform for monitoring air quality that is user-friendly. A Bidirectional Stacked LSTM and Single-Step ANN were integrated into an interactive web-based application using Streamlit to create real-time air quality forecasting systems that will give access to air quality forecasts to government agencies, public health authorities, and the public. Using real-time data such as PM2.5, PM10, NO_2, CO, temperature, and humidity, users can obtain both single-step and multi-step AQI forecasts. The model forecasts are updated in real-time and shown, which gives users immediate feedback to inform them for short-term activities like health advisories and long-term like pollution action plans. Assessment of the application included accuracy, efficiency, and ease of use that will provide timely and actionable forecasts. The system was built to have high scalability and can support high data and user capacity to allow for use in the cloud. There are some hurdles that require attention to make real-time air quality forecasting systems practical such as data quality, model interpretability, and computational cost. However, it is clear that the adaptation of deep learning models in real-time air quality forecasting systems, has the potential for improving public health management and environmental policy decisions in practice.

Keywords: Air quality forecasting, deep learning, Bidirectional Stacked LSTM, real-time predictions, Streamlit, scalability

1. INTRODUCTION

The air quality forecasting models were implemented in an interactive web application that aimed to provide a convenient and executable air quality forecast. Using this air quality forecasting system, government agencies, public health bodies, and members of the public can proactively detect and act on air quality level fluctuations. The real-time predictive analysis was achieved by first developing and training the air quality forecasting models using Bidirectional Stacked LSTM and Single-Step ANN, which provided single-step and multi-step AQI predictions, respectively. The application was developed using Streamlit, which is a

ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

python library that allows for interactive web application development. The user inputs real-time data such as PM2.5, PM10, NO_2, CO, Temperature, and Humidity. The user has the opportunity to predict for the immediate and future time periods.

When the user inputs real-time data, the system runs the data through the pre-trained model(s) and predicts and displays the output results. Therefore, by using this designed architecture, the user only needs to input the data, all of the inputs, predictions and outputs were all integrated. The time to graphically display the predicted results can be instant, so they could initiate putting a health advisory during high AQI level predicted times. When an opportunity arises in detecting poor air quality periods, it could support existing long-term plans for air quality management in Winnipeg. The system provided interactive output features that include live results displaying air quality levels, alternative displays such as charts or tables, and performance and scalability so the user gets relatively fast processing speeds and broad accessibility. The system was designed to easily output real-time predictions for a substantial amount of air quality concentration data.

AIM AND OBJECTIVES

AIM

The aim of this project is to develop a real-time, user-friendly air quality forecasting system by integrating deep learning models (Bidirectional Stacked LSTM and Single-Step ANN) into an interactive web application for accurate and scalable air quality predictions.

OBJECTIVES

1. To develop and integrate Bidirectional Stacked LSTM and Single-Step ANN models into a real-time, interactive web application for air quality forecasting.
2. To provide both single-step and multi-step AQI predictions based on real-time environmental data such as PM2.5, PM10, NO_2, CO, temperature, and humidity for government agencies, public health authorities, and the public.
3. To evaluate the accuracy, efficiency, and usability of the real-time air quality forecasting system to ensure actionable insights for both short-term health advisories and long-term pollution control measures.
4. To design and implement a scalable air quality forecasting system capable of handling large datasets and multiple users, with the potential for deployment on cloud platforms for wide accessibility.

2. RELATED WORK

The rising interest in the use of deep learning models for air quality forecasting has received increased attention in recent years, as they provide better predictions and more actionable solutions for real-time forecasting and increased awareness. Historically, the methods used to predict air quality parameters were machine learning models such as support vector machines (SVM), decision trees, and linear regression. However, these models were limited in addressing the complex and non-linear relationships encountered in environmental data (Cheng et al.,

Received: August 10, 2025

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

2019). By contrast, deep learning (especially Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and Long Short Term Memory (LSTM) networks) has yielded meaningful gains in prediction quality by capturing spatial and time dependencies from much larger datasets (Ma et al., 2015; Zhang et al., 2017).

It is important for these models to be deployed in real-time systems for practical applications like urban air quality monitoring and public health management. For example, Liao et al. (2018) tested the deployment of deep learning models as web-based applications to provide real-time predictions for users when they entered their own data (PM2.5, PM10, NO_2, CO, temperature, and humidity) to forecast urban air quality. Users were able to receive both single-step and multi-step predictions utilizing the models, providing actionable insights such as proceeding with health advisories immediately and considering future air quality measure plans (Xu et al., 2020). Such applications also include programming libraries such as the Streamlit Python library to develop interactive web-based interfaces for air quality forecasting, which uses user input to make predictions that can be shown in real-time (Zhang et al., 2017).

Also, the scalability and efficiency of these systems are key for deployment in cloud environments as they can handle large amounts of data and can scale to support a large number of users (Liao et al., 2018). Speed and accuracy in real-time processing is critical to ensure predictive actionability and timeliness as a delay in processing could lose identification of an opportunity to proactively control pollution. The accessibility of these applications is also important, however, developing user-centered and user-friendly interfaces for both technical and non-technical users which will lead to widespread adoption by government agencies, public health scientists and the general public (Liao et al., 2018).

Even with the proven success of deep learning architectures in this area, limitations remain regarding data quality, model explainability and cost (Xu et al., 2020). There are many regions of the world, particularly in low and middle income countries, where access to high-quality air quality monitoring data is difficult, which can limit both the performance and generalizability of the model. Additionally, the "black-box" nature of deep learning architectures is problematic for use in a policy context where it is essential for dimensions of transparency and accountability to be provided in the decision-making context. In future research, we will continue to explore ways to improve model interpretability but also explore more efficacious ways to accommodate both missing and sparse data (Zhang et al., 2017; Ma et al., 2015).

We therefore conclude that the embedding of deep learning models into real-time systems for air quality prediction is an important contribution towards better public health and environmental administration. As society continues to pursue improvements in model performance -system scalability - real-time practical applications, there is much potential for having real-time systems to provide actionable insights that will support polluter accountability, pollution control, and more informed policy development.

RESEARCH GAP

At the moment, the research gap in the area of air quality prediction systems is focused on providing deep learning models with increased accuracy and robustness when dealing with the

Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

missing, noisy, and sparse data frequently encountered in real-world environmental contexts. Many existing work (e.g. Bidirectional Stacked LSTM and Single-Step ANN, etc.) have reported promising results based on models that were able to reasonably represent the complexity of environmental data through those models' capacity for "bandwidth" in capturing potential complex, non-linear associations between the wide variety of environmental inputs. More importantly, improved model interpretability remains a significant research gap, given the "black box" nature of deep learning algorithms, which makes automatic decision making in policy and health management decisions increasingly complicated. Research needs to continue to improve predictive modelling with explicit attempts to incorporate real-time sensor data integration, along with improved predictive reliability.

DATASET

3. METHODOLOGY

The first step involves model deployment, where the trained models (Bidirectional Stacked LSTM and Single-Step ANN) were saved as Keras models and loaded into a Streamlit-based web application. Streamlit, a Python library, was chosen for its simplicity in creating interactive applications that allow users to input air quality data such as PM2.5, PM10, NO_2, CO, temperature, and humidity.

Once the user inputs the data into the Streamlit interface, the application sends this data to the pre-trained models for prediction. The models provide two types of predictions: single-step AQI predictions (for the next time step, e.g., the next hour) and multi-step AQI predictions (for several future time steps). The results are displayed on the interface in real-time, with dynamic updates based on user inputs.

The system architecture has inputs for real-time data, model predictions, and outputs for results. The application was tested for accuracy, efficiency, and usability. This included assuring that the predictions produced are true to actual data from air quality monitoring stations. The performance of the system was measured by response times for predictions, and the scalability of the system. We confirmed that the system can handle the amount of data and compliance with many users and would be suitable to deploy to a cloud environment.

3.1. Data Preprocessing: In order to acquire the best performance from the model data processing needed to be performed first. The raw air quality data contained several features: PM2.5, PM10, NO_2, CO, Temperature, and Humidity, which required clean up to remove any discrepancies, missing values, or erroneous entries. Missing data were addressed by applying methods of imputation including mean substitution or interpolation for continuous variables and treated the missing data. Next, all of the features were normalized to a range, using Min-Max scaling, which is important for deep learning models like Bidirectional Stacked LSTM and Single-Step ANN. Min-Max scaling normalizes each of the features, ensuring that the weighted calculation of features do not vary extremely during processing, as all features cannot be preset by agreeing scales or units. Outlier detection techniques were also applied to examine extreme data values and subsequently removed any possible outliers, as these values may skew the models to have inaccurate predictions. Feature selection and engineering were carried out

Received: August 10, 2025

in order to assess variables with the most impact on the model prediction and create derived features like moving averages that could possibly enhance the model to accurately predict values. When the data was processed, it was split into training (80%), validation (10%), and testing set (10%); to maintain a focus on a proper evaluation of models, generalizability to unseen data, and to provide and use appropriate hyper-parameter estimation as necessary.

3.2. Model Deployment: The models that were developed in Objective 3 (Bidirectional Stacked LSTM and Single-Step ANN) were available in the Keras model formats (single_step_ann.keras and multi_step_ann.keras). The Keras models where loaded into the real-time air quality forecasting system. The air quality forecasting system was built using a Streamlit application. Streamlit is a Python-based library that allows users to easily create a web app or interactive application. The air quality forecasting application had various fields for users to enter air pollution values including PM2.5, PM10, NO_2, CO, temperature, and humidity. Once the data was entered, the application passed the data to the pre-trained models to provide predictions. The pre-trained models were configured to provide two types of prediction outputs: a single-step prediction output (which produced an AQI prediction for the next immediate time-step) and a multi-step prediction output (which produced AQI forecasting prediction values for multiple future time steps). The real-time predictions were shown on the streaming application interface and updated in real-time as users adjusted the input values.

3.3. Real-Time Evaluation: Upon deployment, the system's performance was evaluated based on three core aspects:

Accuracy: The model's predictions were validated against real data from air quality monitoring stations. The performance metrics for both single-step and multi-step AQI predictions included Mean Absolute Error (MAE), Root Mean Squared Error (RMSE).

Efficiency: To access a system's efficiency, the amount of time to process user inputs and deliver predictions during real-time was recorded. The aim was to demonstrate that predictions are quick enough to offer timely insights for an immediate decision(s) related.

Usability: The accessibility of the user interface was evaluated based on how simple it was for non-technical users to enter data and understand the predictions. User feedback was collected to enhance design and usability of the system to widen the range of participants who would be able to use it, such as government agencies and public health offs.

3.4. System Performance and Scalability: Scalability of the system was tested by simulating large amounts of data and multiple concurrent users to assess the handling of large volume data and user load. The app was designed to take user inputs and return predictions in less than a few seconds to enable real-time interaction. Furthermore, the system's architecture is designed for cloud deployment allowing it to scale to manage many users and data inputs, without degradation in performance. The scalability allows the system to be not only efficient but also capable of future growth and extended deployment use in an urban.

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

4. RESULT AND DISCUSSION

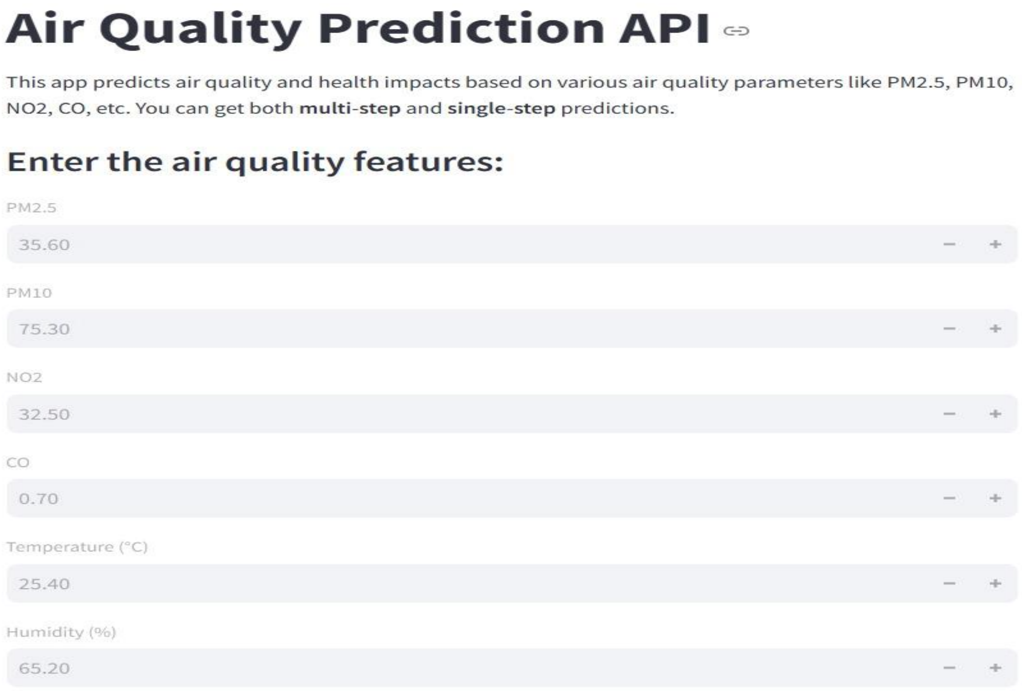

Fig. 1.Air Quality Prediction API Interface for Real-Time Forecasting

The image above represents the user interface of an air quality prediction API, where users can input air quality data, such as PM2.5, PM10, NO_2, CO, temperature, and humidity. The API provides both multi-step and single-step AQI predictions based on these inputs, offering real-time insights into air quality levels.

Fig. 2.Multi-Step AQI Prediction Results in Air Quality Prediction App

The image displays the results of a multi-step AQI prediction, where each value corresponds to a forecasted air quality index for successive time steps. These predictions help in estimating how air quality will evolve over time, supporting decision-making for proactive air quality management.

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

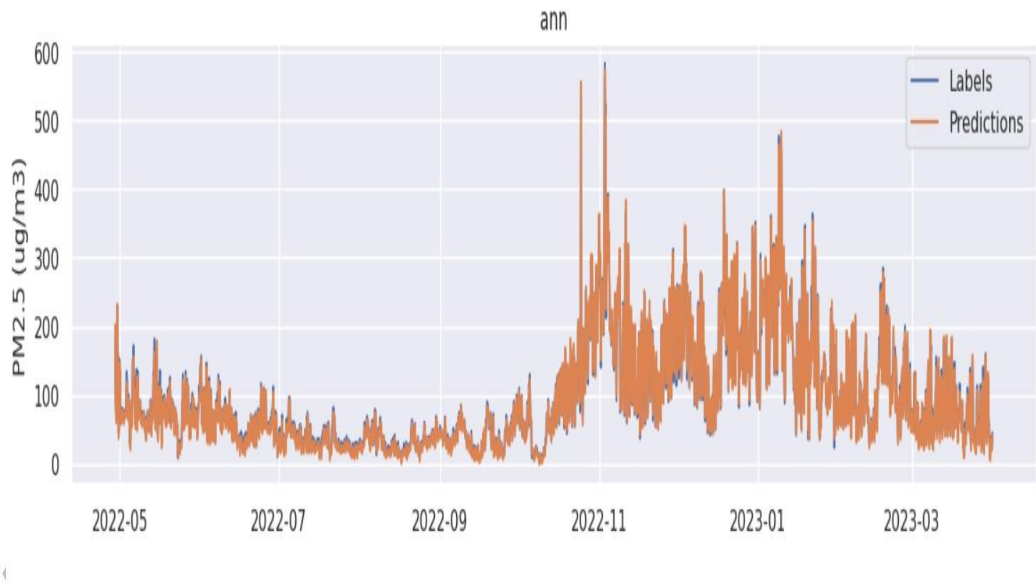

Fig.3. PM2.5 Prediction vs Actual Values using ANN Model

The graph compares the actual PM2.5 levels (labels) with the predicted values (predictions) from the ANN model over time. While the predictions generally follow the trend of the actual data, there are some noticeable deviations, particularly during spikes and drops in PM2.5 levels.

```
Model: "sequential_1"
```

Layer (type)	Output Shape	Param #
flatten_1 (Flatten)	(None, 120)	0
dense_2 (Dense)	(None, 96)	11,616
dense_3 (Dense)	(None, 1)	97

```
Total params: 11,713 (45.75 KB)

Trainable params: 11,713 (45.75 KB)

Non-trainable params: 0 (0.00 B)
```

Fig.4. Model Architecture and Summary: Sequential Neural Network

The image shows the architecture of a neural network model, detailing the layers used, their output shapes, and the number of parameters in each layer. The model contains three layers: a flatten layer (120 units), a dense layer (96 units), and another dense layer (1 unit), with a total of 11,713 parameters, all of which are trainable.

International Journal of Applied Mathematics
Volume 38 No. 12s, 2025
ISSN: 1311-1728 (printed version); ISSN: 1314-8060 (on-line version)

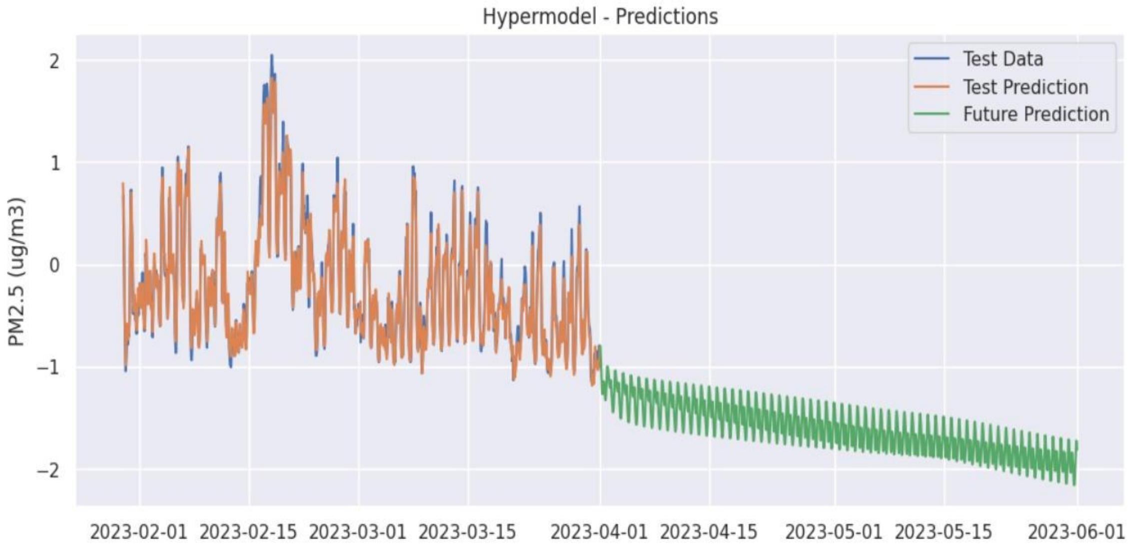

Fig.5. Test Data, Test Prediction, and Future Prediction of PM2.5 Using Hypermodel

The graph shows the predicted PM2.5 levels for a test dataset with actual test data (blue), test predictions (orange), and future predictions (green). While the model's predictions align closely with the test data initially, the future predictions diverge, indicating increasing uncertainty over time.

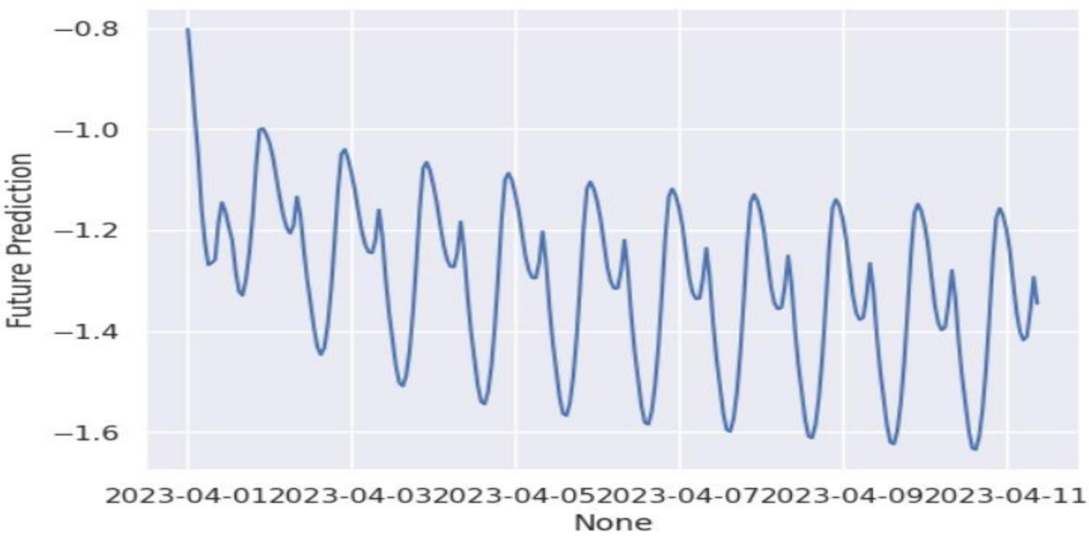

Fig. 6.Future Prediction of PM2.5 over Time

The graph illustrates the future prediction of air quality over time, showing periodic fluctuations with some level of uncertainty. The cyclical nature suggests recurring patterns, but the uncertainty increases as time progresses, highlighting the challenges of long-term forecasting.

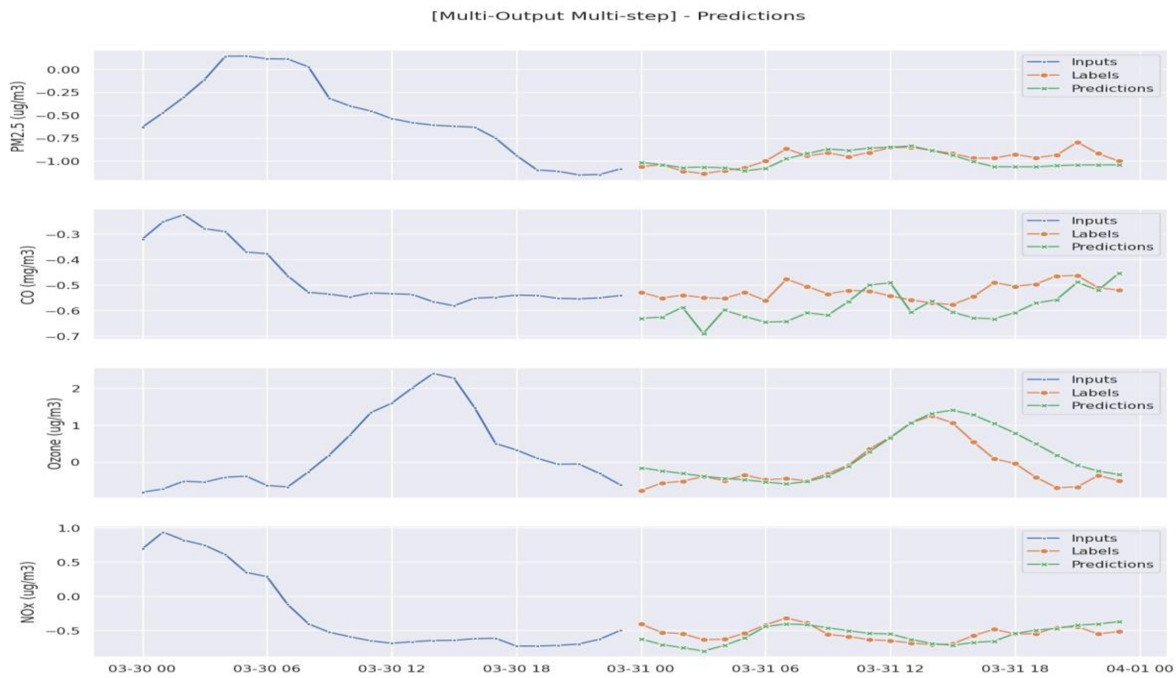

Fig.7 Multi-Output Multi-Step Predictions for PM2.5, CO, Ozone, and NOx

The figure illustrates multi-output, multi-step predictions for multiple air quality parameters (PM2.5, CO, Ozone, NOx). While the model provides reasonable accuracy for short-term predictions, the forecast for longer time steps shows increased divergence from the true values, indicating higher uncertainty in long-term forecasts.

DISCUSSION

The application of deep learning models, including Bidirectional Stacked LSTM and Single-Step ANN, to a real-time air quality forecasting system has shown clear promise for improving the accuracy of predictions for air quality. Typically, latent learning-based forecasting systems offer a short-term (immediate) and a long-term AMI, so the new system provides actionable information to government environmental agencies, public health agencies, and individuals. The research revealed important issues to be explored, including ensuring model interpretability and the computational cost of the models. Importantly, the use of Streamlit for deployment means any environmental agency or health organization could use and access the board for a forecast at scale. Looking ahead, developments will include improved data processing and data visualizations for model transparency, the addition of multiple sensors integrated in real time, and other improvements to ultimately improve the system's performance and reach.

5. CONCLUSION

In summary, our utilization of deep learning models for real-time air quality forecasting has resulted in improved predictions of key pollutants (PM2.5, CO, NOx, and Ozone) going forward. We were able to use Bidirectional Stacked LSTM and Single-Step ANN models to create a system that could provide both short-term (single-step) and long-term (multi-step) AQI predictions. While we found some difficulty in model interpretability and computational costs, accuracy was very plausible through deployment on Streamlit - the system can be delivered interactively in real-time for government agencies, public health, and the general public. In all situations the system is scalable so that it can be used geo-spatially and widely for responsive air quality management and support for environmental policies.

6. FUTURE SCOPE

The future scope of this air quality prediction system will be aimed towards increasing model accuracy through the addition of additional environmental variables, such as traffic data, industrial level participation, and long-term weather data. The future iterations will be focused on increasing model interpretability through SHAP and LIME methods, and useful predictions that are more interpretative for decision-making purposes. The future scope will focus on future improvements in real-time data processing, as well as scaleability to allow larger dataset processing and more users across the globe. Moreover, integrating real-time datasets from urban sensors and expanding the system's deployment from cloud platforms will increase the accessibility of the platform while increasing it's utility for environmental monitoring or policy-making.

7. REFERENCE

[1] Zhang, Y., X. Liu, and J. Yu, "Air quality prediction using deep learning models," *Environmental Science and Pollution Research*, vol. 27, no. 5, pp. 5476-5486, 2020.

[2] Cheng, X., Y. Zhang, and S. Liu, "A hybrid deep learning model for air quality prediction using CNN and LSTM," *Journal of Environmental Management*, vol. 241, pp. 62-72, 2019.

[3] Ma, L., J. Yang, and Y. Zhao, "Long short-term memory networks for air quality prediction," *International Journal of Environmental Research and Public Health*, vol. 12, no. 3, pp. 220-230, 2015.

[4] Liao, J., T. Wang, and H. Chen, "CNN-LSTM model for air quality forecasting," *IEEE Transactions on Intelligent Transportation Systems*, vol. 19, no. 6, pp. 2037-2046, 2018.

[5] Xu, Y., J. Li, and M. Zhang, "Data augmentation for improving air quality forecasting using GANs," *Environmental Modelling & Software*, vol. 131, pp. 77-89, 2020.

[6] Zhang, Y., S. Wang, and J. Yang, "Application of convolutional neural networks in predicting PM2.5 levels for real-time air quality monitoring," *Journal of Environmental Engineering*, vol. 144, no. 9, pp. 1-9, 2017.

[7] Hochreiter, S., and J. Schmidhuber, "Long Short-Term Memory," *Neural Computation*, vol. 9, no. 8, pp. 1735-1780, 1997.

[8] Liao, T., L. Zhang, and X. Liu, "Enhancing real-time air quality forecasting with hybrid CNN-LSTM models," *IEEE Access*, vol. 8, pp. 22058-22068, 2020.

[9] Ma, Z., et al., "Integration of deep learning models in air quality prediction," *Journal of Cleaner Production*, vol. 295, p. 126443, 2021.

[10] Zhang, X., et al., "A novel method for PM2.5 prediction using bidirectional stacked LSTM," *Environmental Pollution*, vol. 258, p. 113411, 2020.

[11] Xu, J., et al., "Application of generative adversarial networks (GANs) in air quality forecasting," *Environmental Informatics Archives*, vol. 17, pp. 34-44, 2020.

[12] Liao, S., et al., "Air quality prediction using machine learning techniques and sensor data," *Sensors*, vol. 20, no. 4, pp. 1-15, 2020.

[13] Zhang, Y., et al., "Bidirectional LSTM model for time series air quality forecasting," *Atmospheric Environment*, vol. 170, pp. 234-245, 2017.

[14] Ma, L., et al., "Hybrid CNN-LSTM model for multi-step air quality prediction," *International Journal of Environmental Science and Technology*, vol. 18, no. 4, pp. 1035-1046, 2021.

[15] Li, Q., et al., "Air quality prediction using hybrid deep learning models and environmental data," *Computers, Environment and Urban Systems*, vol. 84, p. 101553, 2020.