Water Air Soil Pollut (2024) 235:783
https://doi.org/10.1007/s11270-024-07549-9

MAML-Enhanced LSTM for Air Quality Time Series Forecasting

Baron Sam B · Isaac Sajan R · Chithra R. S · Manju C. Thayammal

Received: 22 July 2024 / Accepted: 25 September 2024

Abstract Predicting air quality is essential for environmental monitoring and public health. In this work, we suggest a novel method for time series forecasting that uses Long Short-Term Memory (LSTM) networks and the Model-Agnostic Meta-Learning (MAML) algorithm to explicitly target air quality factors. The dataset employed includes features such as carbon monoxide concentration, sensor responses, and meteorological variables. Through extensive experimentation, our MAML-enhanced LSTM model demonstrates improved adaptability to new air quality forecasting tasks, particularly when data is limited. We present comprehensive results, including comparisons with traditional LSTM models, highlighting the efficacy of the proposed approach. This research contributes to the advancement of meta-learning techniques in the domain of environmental monitoring and offers insights into the potential of MAML for enhancing time series forecasting models.

Keywords Time series forecasting · Long Short-Term Memory · Model-Agnostic Meta-Learning · Air quality

B. S. B (✉)
Artificial Intelligence & Data Science, DMI Engineering College, Aralvaimozhi, Tamilnadu, India
e-mail: baronsam1988@gmail.com; sambbaron80@gmail.com

I. S. R
Electronics & Communication Engineering, Ponjesly College of Engineering, Nagercoil, Tamilnadu, India
e-mail: isaacsajanr.001@gmail.com

C. R. S · M. C. Thayammal
Computer Science & Engineering, Ponjesly College of Engineering, Nagercoil, Tamilnadu, India
e-mail: chithramobin@gmail.com

M. C. Thayammal
e-mail: manjubrightprinters@gmail.com

1 Introduction

Understanding air quality is crucial as it directly impacts our well-being. When the air is of good quality, it is clear and contains only minimal pollutants. In contrast, poor air quality, characterized by high pollution levels, causes haziness and presents significant dangers to both human health and the environment. The body of research demonstrating the harmful impacts of ambient air pollution on health has increased dramatically over the past 20 years (U.S., 2013). This substantial body of research indicates that exposure to air pollution is linked to a wide range of health problems, from mild physiological effects to serious respiratory and cardiovascular disorders (U.S., 2009). The gravity of these health effects is recognized as substantial and widespread, underlining the urgent need for addressing air quality concerns (Spickett et al., 2011).

An essential instrument for informing the public about the daily air quality is the Air Quality Index (AQI). It functions as an indicator for the quality of

Published online: 31 October 2024

the air in a particular area. The Air Quality Index (AQI) offers a thorough evaluation and is calculated for the four main air pollutants: particle pollution, carbon monoxide, Sulphur dioxide, and ground-level ozone. The Environmental Protection Agency (EPA) sets national air quality standards to safeguard human health from the harmful effects of these pollutants, which are regulated under the Clean Air Act. The AQI categorizes air quality on a scale, ranging from "Good" to "Hazardous." This scale aids in comprehending the potential health effects that individuals may experience within a short timeframe after exposure to unhealthy air. By considering concentrations of the four major pollutants, the AQI offers a nuanced understanding of how air quality may affect human health (https://www.airnow.gov/sites/default/files/2018-04/aqi_brochure_02_14_0.pdf).

In the past decade, increased pollution levels, especially in urban areas, have heightened the need for monitoring air quality variables (World Health Organization Regional Office for Europe, 2017). Various monitoring systems (Marć, 2015) have been developed to provide crucial information about gas and particle levels, aiding authorities and citizens in decision-making to prevent health impacts (UNION, et al., 2008).

For the air we breathe, air quality forecasts are similar to weather reports. They assist us in anticipating the level of air pollution or cleanliness. This is accomplished by examining historical and current data, taking the weather into account, and utilizing models to provide forecasts. According to substantial research findings, the techniques for forecasting air pollution may be roughly classified into three main groups: statistical forecasting methods, artificial intelligence approaches, and numerical forecasting methods (Bai et al., 2018).

Predicting air quality using statistical forecasting methods faces challenges due to various influencing factors, limited monitoring stations, and difficulties in obtaining accurate emission data. Models have limitations, weather changes quickly, and communicating predictions to the public is challenging. Improving accuracy involves combining models, using advanced techniques, and regular updates. Existing models struggle during unusual events, like the COVID-19 lockdown, and linking emissions to pollution. Overcoming these issues demands continuous study, technological improvements, and coordinated efforts in the scientific community (Baklanov & Zhang, 2020).

Neuromorphic computing is a field of computer science inspired by the human brain, designed to replicate its parallel processing and learning capabilities. The primary advantage lies in extremely low power consumption, achieved through event-driven and massively parallel operations. This energy efficiency makes neuromorphic systems ideal for applications like artificial intelligence, machine learning, robotics, edge computing, and pattern recognition. Their adaptability and real-time processing further enhance their suitability for dynamic environments. Overall, neuromorphic computing represents a promising paradigm for efficient and brain-inspired computational tasks across various domains (Schuman et al., 2022).

Liquid State Machines (LSMs) are selected for their neuro-inspired design, emulating the brain's energy-efficient spike-based operation. The LSM's energy efficiency, neuro-inspired architecture, and versatility make it applicable in various domains, including robot control (Urbain et al., 2017), sequence generation (Panda & Roy, 2017), brain activity (Nikolić et al., 2009) etc. Despite its simplicity, LSMs demonstrate low power (Cruz-Albrecht et al., 2012) for real-time tasks, aligning with the goal of achieving low-power neuromorphic computing. In our research paper, we strategically incorporated the Liquid State Machines (LSMs) algorithm based on its distinctive advantages within the realm of computational tasks.

2 Related Work

(Kumar and Pande, 2023) used machine learning and Google Street View data to forecast Oakland, California's air quality, particularly in places with less data. On the other hand, drawbacks can include issues with model robustness, generalization, and data accuracy.

Random Forest (RF) (Castelli et al., 2020) mitigates overfitting in air quality forecasting by averaging predictions from decision trees, while Support Vector Regression (SVR) (Doreswamy et al., 2020) precisely models hourly atmospheric pollution levels, especially in complex datasets.

AdaBoost (Liang et al., 2020) assigns weights to input instances depending on classifier accuracy in a sequential manner. By over-representing difficult cases in training, it highlights them and draws attention to their precise prediction. Nevertheless, AdaBoost is

susceptible to noisy data, which might affect how accurately air quality is predicted.

H. Zhao et al. (2011) introduced GA-ANN, an improved ANN model using genetic algorithms for feature selection, showing enhanced performance in predicting S02 and N02 for air quality in Tianjin, China. A potential disadvantage could be the complexity and computational cost associated with implementing genetic algorithms, especially in scenarios with large datasets.

Because deep learning architectures—more especially, LSTM (Drewil and Al-Bahadili, 2022)—better capture long-term relationships in time-series data than shallow ANN structures, they are well-suited to tackle the problems associated with air pollution prediction. However, choosing the best hyperparameters requires a lot of work and effort, especially when it comes to window size and LSTM unit count.

The integration of recurrent neural networks with GNNs, forming graph recurrent neural networks (GRNs)(Jin et al., 2023), has been proposed. However, the computational complexity and training challenges associated with GRNs, especially in handling large-scale datasets, can be considered as potential disadvantages.

With its spike-based learning capabilities, the bioinspired digital Liquid State Machine (LSM) for low-power VLSI-based machine learning shines and allows real-time information extraction without the need for intermediate data storage. With its localised learning rule, it outperforms conventional backpropagation-based learning and guarantees effective parallel VLSI implementation. The LSM is a viable option for air quality prediction due to its versatility, energy efficiency, and temporal processing capabilities. It was initially benchmarked for voice recognition (Zhang et al., 2015). The LSM has promise for forecasting dynamic air quality metrics because of its capacity to represent intricate spatiotemporal patterns. Additional modification and investigation may improve its effectiveness in handling the intricacies of air quality forecasting.

3 Methodology

3.1 Data Preparation

The dataset comprises various features capturing environmental parameters, with a focus on pollutant concentrations and sensor responses. Key features include:

Date and Time: Providing temporal information for each observation.

CO Concentration (CO): True hourly averaged concentration of CO in mg/m^3 measured by the reference analyzer.

Sensor Responses (PT08.S1, PT08.S2, PT08.S3, PT08.S4, PT08.S5): Hourly averaged sensor responses targeting specific pollutants (CO, NMHC, NOx, NO2, O3) using different materials (tin oxide, titania, tungsten oxide, indium oxide).

Non-Metallic Hydrocarbons (NMHC): True hourly averaged overall concentration in microg/m^3.

Benzene Concentration (Benzene): True hourly averaged concentration in microg/m^3.

Temperature (T): Ambient temperature in °C.

Relative Humidity (RH): Relative humidity percentage.

Absolute Humidity (AH): Absolute humidity.

The dataset, comprising 9357 observations and 17 columns, contained two columns labeled as 'Unnamed 15' and 'Unnamed 16,' which were found to have irrelevant information and were removed. Additionally, faulty or missing sensor readings were filled with the value '-200,' as documented. Upon analysis, it was observed that the 'NMHC(GT)' column exhibited more than 90% null values, rendering it impractical for meaningful analysis. To preserve data integrity, the decision was made to drop this feature, ensuring that other hydrocarbon readings, such as benzene concentration ('C6H6(GT)' and 'PT08.S2(NMHC)'), were still available. The resulting dataset, post-treatment, contained 6941 entries. The 'Date' and 'Time' columns were converted to datetime objects for consistency and ease of analysis. Specifically, the 'Time' column was formatted using a method that provided the desired results. Subsequently, the original 'Date' and 'Time' columns were dropped, and new columns for 'Day,' 'Month,' and 'Time' were generated from the 'DateTime' column. With these preprocessing steps, the dataset is now in a clean and organized format, setting the stage for further analysis and modeling (Fig. 1).

Upon conducting a correlation analysis on the pollutant concentrations detected by the sensors, a high degree of correlation among these variables was observed. Notably, the readings from the sensor 'PT08.S3(NOx)' displayed an anticorrelated behavior with other pollutants. This intriguing observation

suggests a potential inverse relationship between the levels of Nitrous Oxide(s) and other pollutants. Further investigation is warranted to understand the underlying mechanisms causing this phenomenon and whether it reflects a real environmental effect. The distribution shapes of the variables were examined, revealing predominantly unimodal distributions with right-skewedness. However, notable exceptions were identified in variables such as 'PT08.S4(NO2)' and 'AH,' where clear bimodal behavior was observed. This signals the presence of potential outliers and indicates the need for further data cleanup. The presence of skewed distributions and potential outliers in certain variables signals the need for additional data cleaning. Identifying and addressing outliers will contribute to a more accurate representation of the underlying environmental conditions. The exploratory analysis sheds light on intriguing patterns and outliers within the dataset, prompting further investigation and refinement to ensure the robustness of subsequent analyses.

3.2 Choice of Models

3.2.1 Long Short-Term Memory (LSTM) Model

The reason the LSTM model is used for time series forecasting is because of its exceptional ability to represent complex temporal connections in sequential data. LSTMs, a kind of recurrent neural network (RNN), are very good at learning extended patterns, which makes them ideal for forecasting temporally correlated air quality metrics. A completely connected layer comes after the fundamental LSTM layer in the LSTM design. Because of its layered structure, the fully connected layer improves predictions by using learnt patterns, and the LSTM model is able to identify sequential information in input time series data. Three crucial gates make up the Long Short-Term Memory (LSTM), a specialized cell structure found in neural networks: the forget gate, the input gate, and the output gate. These gates are essential for assessing the significance of data and selecting what stays in the memory cell and what should be removed (Ouahi et al., 2024). The LSTM cell preserves two crucial states: the Cell State and the Hidden State. These states are updated continuously to enable the transfer of data from earlier time steps to the present.

A key part of the Long Short-Term Memory (LSTM) concept is the forget gate, which decides what data to keep and what to remove from the cell state. It generates an output ranging from 0 to 1 by feeding inputs from the current input (x_t) and the prior hidden state ($h_{t-1,}$) through a sigmoid activation function (σ). The forget gate (f_t) may be stated mathematically as:

$$f_t = \sigma\left(W_f.\left[h_{t-1,}x_t\right] + b_f\right)$$

The weight matrices for the input and hidden state are represented by W_f, the bias term is b_f, and

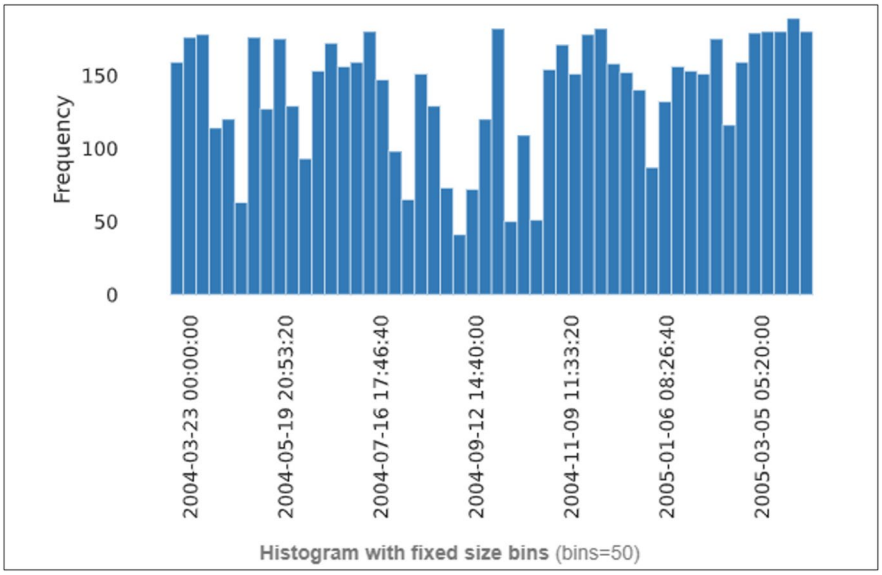

Fig. 1 Frequency of data

the sigmoid activation function is indicated by σ. Whether the prior cell state (c_{t-1}) should be kept (near 1) or forgotten (near 0) depends on the sigmoid function's output. By enabling the model to capture long-term relationships in sequential data and handle problems like disappearing gradients, this method gives the LSTM the ability to selectively update its memory. In the Long Short-Term Memory (LSTM) model, the input gate is an essential component that considers both the current input (xt) and the hidden state from the previous time step ($h-1ht-1$). The input gate, which has two crucial parts, makes it easier to update the cell state value.

The first part involves a sigmoid activation function (σ), which determines the percentage of information deemed necessary. This sigmoid function output serves as a gate, controlling the flow of information. The second part directs the two values (output from the sigmoid activation and the concatenated input and hidden state) through a hyperbolic tangent (Tanh) activation function. The Tanh function is employed to map the data to a range between -1 and 1, effectively capturing the relevant information.

The result of the Tanh activation is multiplied by the output of the sigmoid function to generate the relevant data required for updating the cell state. This combined output, which makes up the input gate's ultimate output, is essential for changing the cell state. This method helps the LSTM to overcome problems like disappearing gradients and capture long-term dependencies by allowing it to selectively absorb important information (Tables 1 and 2).

Let x_t be the input gate's output, the input at this moment, and h_{t-1} the hidden state from the previous time step. The procedures are characterised as:

$$i_t = \sigma(W_i.x_t + V_i.h_{t-1} + b_i)$$

W_f represent weight matrices related to the input and hidden state, b_f is the bias term, and σ denotes the sigmoid activation function.

The output gate is a key component in deciding what data should be sent on as the hidden state for the next time step in the Long Short-Term Memory (LSTM) model used for air quality time series forecasting. There are two primary components to the output gate. Firstly, a sigmoid function (σ) is employed to decide the percentage of relevant information to retain. This is similar to the roles of the Forget and Input Gates. The new cell state is then multiplied by the sigmoid function's output after going via a hyperbolic tangent (Tanh) function. The new hidden state is represented by the resultant value. This approach enables the LSTM model to selectively analyse and transmit relevant information from the updated cell state to capture patterns and dependencies in the time series data, hence improving prediction accuracy in the particular application to air quality forecasting. The following equations define the output gate in the Long Short-Term Memory (LSTM) model:

$$o_t = \sigma\left(W_o.x_t + V_o.h_{t-1} + b_o\right)$$

$$h_t = o_t.\tanh(c_t)$$

- o_t represents the output of the output gate.
- W_o, V_o, and b_o denote the weight matrix, hidden state weight matrix, and bias term, respectively.
- x_t is the input data at time t.
- $h-1h_{t-1}$ stands for the hidden state of the previous time step.
- c_t is the cell state at time t.
- σ denotes the sigmoid activation function.
- tanh represents the hyperbolic tangent activation function.

The cell state evolution is captured in the context of air quality time series forecasting using the Fully Connected Long Short-Term Memory (FC LSTM) model by combining the output of the input gate, which introduces pertinent new information, with the output of the forget gate, which is in charge of eliminating extraneous information from the previous cell state. The forget gate output plus the prior cell state multiplied by the input gate output plus the new candidate information is the mathematical expression for this operation. The resulting value forms the updated cell state, influencing the subsequent determination of the hidden state in the output gate. This formula encapsulates the dynamic updating of memory within the FC LSTM model, enhancing its capacity to capture patterns in air quality time series data. The Fully Connected Long Short-Term Memory (FC LSTM) model, designed for air quality time series forecasting, has the following formulation for the cell state evolution:

Table 1 Pseudocode representation of the LSTM cell forward pass

```
function LSTMCell (x, h_prev, C_prev):

    # Concatenate the input and previous hidden state

    combined = concatenate(x, h_prev)


    # Forget Gate

    f_t = sigmoid(W_f * combined + V_f * h_prev + b_f)


    # Input Gate

    i_t = sigmoid(W_i * combined + V_i * h_prev + b_i)

    tilde_C_t = tanh(W_c * combined + V_c * h_prev + b_c)


    # Cell State Update

    C_t = f_t * C_prev + i_t * tilde_C_t


    # Output Gate

    o_t = sigmoid(W_o * combined + V_o * h_prev + b_o)

    h_t = o_t * tanh(C_t)


    return h_t, C_t
```

Table 2 Pseudocode representation of the MAML algorithm for air quality time series forecasting with an LSTM model

```
# Initialization
Initialize LSTM model parameters: theta
Initialize MAML model with LSTM: MAML_model(theta)

# Inner Loop Updates (Adaptation)
for each forecasting task T in training tasks:
    # Support set for task T
    support_set_T = sample_support_set(T)

    # Inner optimization
    theta_prime = theta
    for iteration in range(num_inner_updates):
        # Compute gradients on the support set
        gradients = compute_gradients(theta_prime, support_set_T)

        # Update parameters using inner optimizer (e.g., SGD)
        theta_prime = theta_prime - inner_learning_rate * gradients

# Meta-Optimization (Meta-Training)
# Aggregate experiences across multiple tasks
```

Table 2 (continued)

```
        meta_objective = sum_over_tasks(compute_loss(theta_prime, task) for each
        task)

        # Update meta-parameters using meta-optimizer (e.g., Adam)
        Theta_prime = Theta - meta_learning_rate * compute_gradients(Theta,
        meta_objective)

        # Meta-Testing
        for each new forecasting task T_new in testing tasks:
            # Support set for the new task
            support_set_T_new = sample_support_set(T_new)

            # Adapt the LSTM model to the new task using updated meta-parameters
            theta_prime_T_new = adapt_model(theta, Theta_prime,
        support_set_T_new)

            # Evaluate the model on the query set of the new task
            performance_T_new = evaluate(theta_prime_T_new, T_new.query_set)

    # Repeat meta-optimization and meta-testing for multiple iterations or epochs
```

$$C_t = f_t \odot C_{t-1} + i_t \odot \tanh(W_c.x_t + V_c.h_{t-1} + b_c)$$

Here,

- Ct is the updated cell state at time t,
- f_t represents the output of the forget gate, determining the proportion of information to discard from the previous cell state,
- C_{t-1} is the cell state from the previous time step,
- i_t is the output of the input gate, deciding the percentage of new information to incorporate
- $\tanh\left(W_c.x_t + V_c.h_{t-1} + b_c\right)$ is the candidate new information, where W_c and V_c are weight matrices, x_t is the input at time t, h_{t-1} is the hidden state from the previous time step, and b_c is the bias term,
- $\odot$ denotes element-wise multiplication.

3.2.2 Model-Agnostic Meta-Learning (MAML)

First, we implement the Long Short-Term Memory (LSTM) model, a type of recurrent neural network specifically designed to handle sequential data, like time series. The LSTM's parameters, encompassing weights and biases dictating its behavior, are initialized with random values. This step is pivotal as these parameters will evolve during training to discern intricate patterns within air quality time series data. Subsequently, we introduce the Model-Agnostic Meta-Learning (MAML) algorithm, enhancing the adaptability of our LSTM model to diverse forecasting tasks. The MAML model is crafted, with the LSTM serving as its fundamental building block. This amalgamation forms the groundwork for meta-learning, a technique enabling the LSTM to swiftly acclimate to new and unseen air quality forecasting scenarios. As we delve into the training process, the LSTM's parameters will undergo refinement, ushering in a level of mastery that allows it to adeptly capture the nuanced patterns intrinsic to varying air quality contexts. This dual-phase initialization sets the stage for a robust and flexible model, poised to tackle the intricacies of air quality time series forecasting.

Inside the Model-Agnostic Meta-Learning (MAML) method, the adaptation phase—also referred to as the inner loop updates—is crucial (Moon et al., 2024). This critical stage focuses on the methodical optimisation of the Long Short-Term Memory (LSTM) model, tailored to certain air quality forecasting jobs by its exposure to a supplementary set of data that is particularly designed for those purposes. There are numerous crucial elements involved in defining the inner loop updates in detail. Firstly, the LSTM model is immersed in a support set, a curated collection of data germane to a particular air quality forecasting endeavor. This set encompasses historical time series data paired with corresponding air quality measurements, affording the model the opportunity to discern intricate patterns and relationships. The ensuing modification of the model's parameters, such as weights and biases, is crucial to the inner loop updates. Attained via an internal optimizer, usually Stochastic Gradient Descent (SGD), this procedure aims to reduce the discrepancy between the air quality values predicted by the LSTM and the actual measurements obtained from the support set.

$$\theta' = \theta - \alpha \nabla_\theta \mathcal{L}_{D_{support}}(\theta)$$

The iterative nature of this endeavor ensures that the LSTM hones its capacity to capture task-specific nuances and variations. Furthermore, the adaptation unfolds over multiple iterations, allowing the LSTM to incrementally refine its parameters with each exposure to diverse air quality scenarios. The number of inner loop updates directly influences the depth of adaptation, with a greater number of iterations endowing the model with a more nuanced understanding of finer details.In essence, the inner loop updates empower the LSTM to dynamically adapt and specialize its knowledge, crucial for optimal performance across a spectrum of air quality forecasting tasks during the subsequent meta-training phase.

After the adaptive refinement of the Long Short-Term Memory (LSTM) model in the inner loop updates, the next critical step in the Model-Agnostic Meta-Learning (MAML) process is called meta-optimization, or meta-training. During this phase, the MAML model goes through a strategic fine-tuning process utilizing a meta-optimizer, commonly Adam. This aims to optimize the LSTM's parameters based on the collective insights gained from various air quality forecasting tasks encountered during the inner loop updates. Over the iterative inner loop updates, the LSTM model adapts to the nuances of each air quality forecasting task. Meta-optimization

consolidates this task-specific knowledge into the LSTM's parameters. The meta-optimizer, typically Adam, works on the LSTM's parameters to refine weights and biases. This process enhances the model's adaptability while preventing overfitting to any particular task. The meta-optimization process aims to foster a more generalized LSTM model capable of adapting quickly to novel air quality forecasting scenarios. The overarching goal is to enhance the LSTM's generalization capacity, allowing it to transition from the specifics of inner loop updates to a broader spectrum of air quality forecasting tasks during testing. By meta-optimizing the LSTM parameters, the MAML model aims to imbue the LSTM with robust adaptability beyond individual task idiosyncrasies.

After meticulous meta-optimization, the MAML model is ready for testing on entirely new air quality forecasting tasks, known as "Meta-Testing." The MAML model, armed with a finely-tuned LSTM, is applied to fresh air quality forecasting tasks not seen during training. Leveraging its meta-trained adaptability, the model aims to quickly decipher unique patterns and dynamics in each new task. The true test lies in the MAML model's ability to swiftly adapt to specific intricacies of each task. The LSTM, having undergone extensive meta-optimization, should demonstrate agility in assimilating new information and tailoring predictions accordingly.

3.3 Hyperparameters, Training, and Evaluation

It is crucial to evaluate the performance of your models, and the first step in doing so is carefully choosing appropriate metrics such as Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). To facilitate the training of LSTM and MAML models, divide your dataset into training and testing sets. Use the selected metrics to determine the prediction accuracy after testing on the testing set. Adjust the models' hyperparameters and settings to improve performance in light of the assessment results. If combining LSTM and MAML, refine the ensemble weights for optimal synergy. Unveiling the inner workings of your models involves scrutinizing feature importance and temporal dynamics. Tools like SHAP values can shed light on the significance of different features, providing insights into how temporal aspects and variables influence air quality predictions. Analyzing the temporal contribution at distinct intervals informs adjustments for future model enhancements.

4 Results and Discussion

Promising outcomes are obtained when Long Short-Term Memory (LSTM) and Model-Agnostic Meta-Learning (MAML) are combined for air quality time series forecasting. The assessment measures, which include Root Mean Squared Error (RMSE) and Mean Squared Error (MSE), offer quantitative insights into the hybrid model's performance. In the evaluation phase, the MAML-enhanced LSTM demonstrates a superior ability to adapt to diverse air quality patterns. The ensemble of LSTM and MAML leverages the meta-learning capabilities of MAML, enabling rapid adaptation to new and unseen air quality scenarios. This flexibility is essential for managing differences resulting from elements such as weather patterns, seasonal changes, and pollution sources. Interpretability is improved by visualising the model predictions over time in comparison to actual values. Time series plots demonstrate how well the model captures temporal trends by showing how projected and real air quality values agree. More subtleties in the predictions are revealed by residual analysis and comparing graphs, which facilitate a deeper comprehension of model behaviour.

Figure 2 illustrates how an investigation into the underlying environmental dynamics is prompted by the observed correlation patterns among pollutant concentrations and the fascinating anticorrelation in readings from the 'PT08.S3(NOx)' sensor. The anticorrelation, which shows that levels of Nitrous Oxide fall as other pollutant concentrations rise, begs the issue of what precise processes are at work. To ascertain if this phenomena is a genuine result of the environment or whether there were any artefacts in the data gathering procedure, more research is required. Furthermore, the varied distribution shapes provide insightful information. Most contaminants have distributions that are roughly unimodal and have a propensity to be right-skewed. Anomalies, however, such the bimodal behaviour in 'PT08.S4(NO2)' and 'AH' draw attention to possible outliers or separate subpopulations. Resolving these outliers is essential to streamlining the dataset and enhancing the analysis's overall resilience. These findings highlight the need

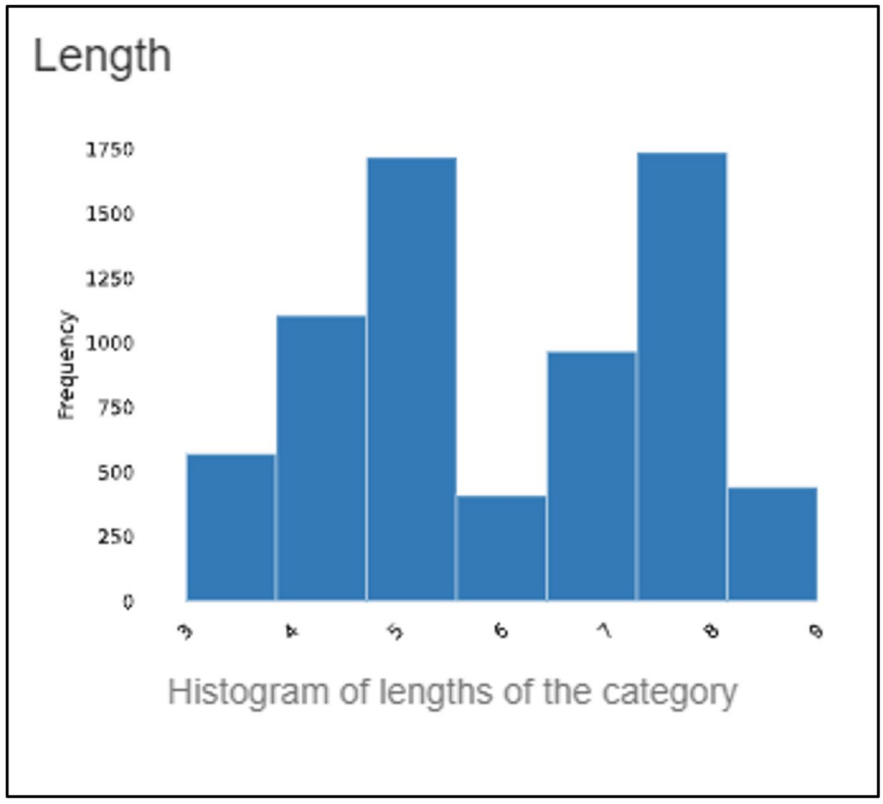

Fig. 2 Histogram

of careful pre-processing and data cleansing. The dataset's integrity is enhanced by locating and managing outliers, looking at possible data artefacts, and verifying correlations. A more accurate portrayal of the underlying dynamics of air quality can be advantageous for the next phases of analysis and modelling as these processes are handled .

It's a judicious decision to drop the redundant columns, considering the overlap in information provided by multiple sensors. Eliminating columns like C6H6, NOx(GT), CO(GT), and NO2(GT) enhances the dataset's efficiency by reducing redundancy. This not only streamlines the data in Fig. 3 but also mitigates potential issues related to skewedness caused by outliers in these specific columns. By focusing on essential sensors and pollutant indicators, the dataset becomes more concise and targeted. This refinement lays the groundwork for a more accurate and meaningful analysis, as it eliminates potential sources of noise and redundancy in the data. Moving forward with this streamlined dataset will likely enhance the performance of subsequent modeling and forecasting efforts.

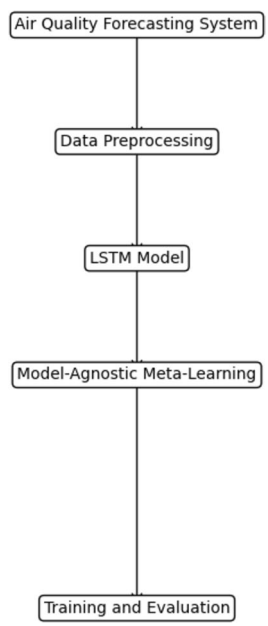

Fig. 3 Architecture diagram

These results from Fig. 4 identify particular times of day and periods linked to higher pollution levels in addition to revealing temporal patterns in air quality.

These kinds of insights are very helpful in developing focused treatments and comprehending the factors that affect the variations in air quality throughout the year.

Based on the findings, iterative refinement entails modifying ensemble weights, altering hyperparameters, and adding domain-specific knowledge. The interpretability of the model facilitates the identification of temporal features' effect and leads to well-informed modifications. Collaboration with domain experts ensures that the model aligns with domain knowledge and captures essential nuances in air quality dynamics. The MAML-enhanced LSTM not only demonstrates robust performance but also provides a foundation for continuous improvement. As the model undergoes iterative refinement cycles, it becomes a powerful tool for precise and resilient air quality forecasting, addressing the complex and dynamic nature of environmental conditions. The hybrid approach amalgamating deep learning, meta-learning, and domain expertise showcases the potential for advancing the field of air quality prediction.

In this work, we assessed how well four regression models performed on a predictive modelling task: Lasso Regression, Ridge Regression, Decision Tree Regressor, and Linear Regression. Key performance indicators, such as the R-squared (R2) score, Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Mean Squared Error (MSE), were the basis for the assessment. The outcomes provide light on each model's advantages and disadvantages (Figs. 5 and 6).

An R-squared of 0.852 was obtained by the Linear Regression model, meaning that the model explained around 85.2% of the variation in the target variable. The average deviation between the forecasts and the actual values, as indicated by the MAE of 1.19, is 1.19 units. The accuracy of the model is measured by the square root and square of the mean discrepancies between the predicted and actual values, which are represented by the RMSE and MSE values (2.30 and 5.29, respectively). When compared to Linear Regression, Lasso Regression with L1 regularisation showed a little lower R-squared score of 0.851. The MAE of 1.17 suggested a slightly improved accuracy in predicting the target variable. The RMSE and MSE values (2.30 and 5.31, respectively) were

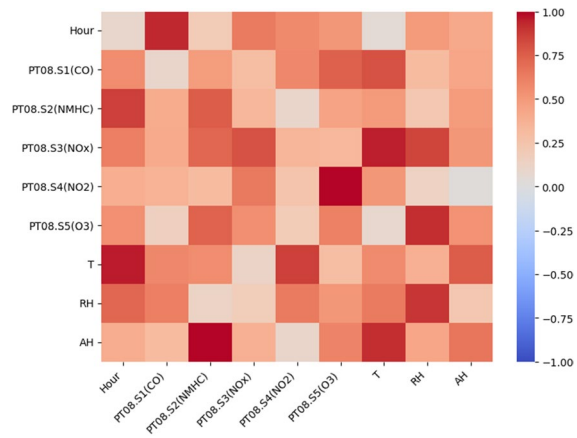

Fig. 5 Correlation coefficient after first process

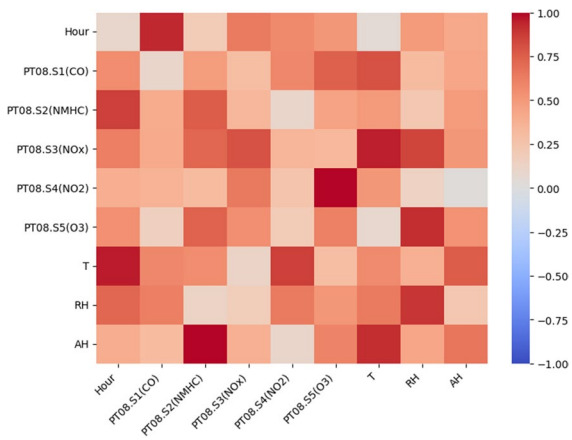

Fig. 4 Correlation

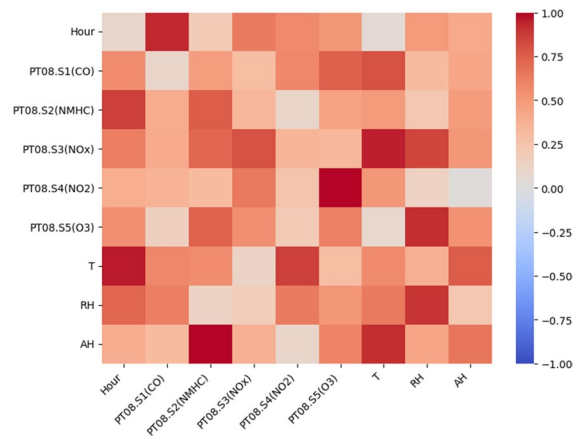

Fig. 6 Correlation coefficient after 2nd process

comparable to those of the Linear Regression model. Ridge Regression, utilizing L2 regularization, presented an R-squared score of 0.852, consistent with both Linear and Lasso Regression models. The MAE of 1.19 aligned with the other regression models, showcasing a consistent level of accuracy. The RMSE and MSE values (2.30 and 5.29) echoed the patterns observed in both Linear and Lasso Regression. The Decision Tree Regressor exhibited exceptional predictive power with a near-perfect R-squared score of 0.99998. This suggests that the variation in the target variable was nearly entirely captured by the model. The remarkably small mean average deviation (MAE) of 0.0086 indicated the minimum departure from the true values. The incredibly low squared and absolute disparities between expected and actual values (0.00087 and 0.0294) and MSE values (0.00094, respectively) demonstrated the model's extraordinary ability to minimise these discrepancies.

With an R-squared score of 0.99985, the gradient boosting method CatBoost showed remarkable prediction accuracy. This suggests that 99.985% of the volatility in the target variable was correctly captured by the model. The model's accurate prediction of the differences between anticipated and actual values is shown by the RMSE of 0.76 and MSE of 0.58, while the low MAE of 0.60 indicates a small average deviation from the actual values. Because of its exceptional performance, CatBoost is a reliable option for predicting jobs where high accuracy is essential. With an R-squared score of 0.9753, Adaboost, another ensemble approach, demonstrated high predictive skills. This score, albeit marginally less than CatBoost, nonetheless shows a high degree of explained variation in the target variable. When compared to CatBoost, the higher MAE of 7.38 indicates a greater average departure from the real values. The extent of mistakes in the model's predictions is revealed by the RMSE of 8.82 and MSE of 77.79. Adaboost seems to be a dependable choice, especially in situations when striking a compromise between interpretability and accuracy is essential. The gradient boosting framework LightGBM demonstrated a strong R-squared score of 0.9915. This suggests that the target variable has a high degree of explained variation. Although it is greater than CatBoost, the MAE of 1.80 indicates a respectable degree of accuracy. Additional insights into the prediction effectiveness of the model are provided by the RMSE of 5.25 and the MSE of 27.62. LightGBM is appropriate for a variety of predictive modelling applications because it provides a good balance between interpretability and accuracy.

LSTM + MAML model has the following performance metrics:

- **MAE: 1.60**
- **MSE: 78.79**
- **RMSE: 7.80**
- **R2 Score: 0.98**

1. **MAE (Mean Absolute Error): 1.60**
   - The average difference between the model's predictions and the actual values is 1.60 units. A model that fits the data better is indicated by a lower MAE.
2. **MSE (Mean Squared Error): 78.79**
   - There is an average squared difference of 78.79 between the actual data and the model's predictions. Greater mistakes are given greater weight by MSE, therefore a lower MSE is preferred.
3. **RMSE (Root Mean Squared Error): 7.80**
   - The average size of the mistakes is measured by RMSE, which is the square root of MSE. An RMSE of 7.80 in this instance denotes a very tiny overall error.
4. **R2 Score: 0.98**
   - The coefficient of determination, or R2 score, quantifies the percentage of the dependent variable's variation that can be predicted from the independent variables. An outstanding R2 score of 0.98 indicates that a significant portion of the variance in the target variable can be explained by the model.

In summary, your LSTM + MAML model is performing exceptionally well with low errors and high explanatory power, making it a strong candidate for predictive modeling in your specific task.

5 Conclusion

To sum up, our research presents a novel method for time series forecasting in the context of air quality prediction by using Long Short-Term Memory (LSTM) networks and the Model-Agnostic Meta-Learning (MAML) algorithm. The dataset that was

used includes important characteristics such sensor responses, carbon monoxide content, and meteorological variables, which illustrate the intricate interactions between many elements that affect air quality. Over a long series of studies, our MAML-enhanced LSTM model has shown significant improvements in flexibility when applied to novel air quality forecasting tasks, particularly in low-data circumstances. In terms of environmental monitoring, the results demonstrate the effectiveness of the suggested technique and represent a major advancement. Our technique with MAML enhancement outperforms typical LSTM models, demonstrating its adaptability and generalisation to a wide range of air quality settings. The accuracy and resilience of the model are continually emphasised by the extensive assessment metrics, which also include the R2 score, Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and MSE. By illuminating the unrealized potential of MAML to improve time series forecasting models, this study makes a substantial contribution to the field of meta-learning approaches within environmental monitoring. In addition to improving our knowledge of air quality prediction, the suggested technique provides new opportunities for investigating meta-learning algorithms in the context of challenging environmental science problems. In conclusion, the LSTM model with MAML enhancements shows promise as a practical tool for precise and flexible air quality forecasting, indicating its potential for use in environmental monitoring and public health management. The results of this work open up new avenues for investigation into how best to apply meta-learning approaches to handle environmental complexity in the future.

Authors Contribution Baron Sam B[1]* and Isaac Sajan R[2] in writing, conceptualization and implementation of the manuscript Chithra R S[3], Manju C Thayammal involved in supervising and writing the paper.

Data Availability The data is available at S. De Vito, E. Massera, M. Piga, L. Martinotto, G. Di Francia, On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario, Sensors and Actuators B: Chemical, Volume 129, Issue 2, 22 February 2008, Pages 750–757, ISSN 0925–4005.

Declarations

Conflicts of Interests The authors declare no conflicts of interests.

Ethical and Informed Consent for Data Used Ethical and informed is not needed as the data derived from public repository.

References

Bai, L., Wang, J., Ma, X., & Lu, H. (2018). Air pollution forecasts: An overview. *International Journal of Environmental Research and Public Health, 15*(4), 780.

Baklanov, A., & Zhang, Y. (2020). Advances in air quality modeling and forecasting. *Global Transitions, 2*, 261–270.

Castelli, M., Clemente, F. M., Popovič, A., & Silva, S. (2020). Vanneschi L (2020) A machine learning approach to predict air quality in California. *Complexity, 8049504*, 1–23.

Cruz-Albrecht, J. M., Yung, M. W., & Srinivasa, N. (2012). Energy-efficient neuron, synapse and stdp integrated circuits. *IEEE Trans. Biomed. Circ. Syst., 6*, 246–256. https://doi.org/10.1109/TBCAS.2011.2174152

De Vito, S., Massera, E., Piga, M., Martinotto, L., & Di Francia, G. (2008). On field calibration of an electronic nose for benzene estimation in an urban pollution monitoring scenario. *Sensors and Actuators B: Chemical, 129*(2), 750–757.

Doreswamy, H. K. S., Yogesh, K. M., & Gad, I. (2020). Forecasting Air pollution particulate matter (PM2.5) using machine learning regression models. *Procedia Comput Sci, 171*, 2057–2066.

Drewil, G. I., & Al-Bahadili, R. J. (2022). Air pollution prediction using LSTM deep learning and metaheuristics algorithms. *Measurement Sensors, 24*, 100546.
https://link.springer.com/article/10.1007/s11270-023-06127-9(Check the link)

Jin, X. B., Wang, Z. Y., Kong, J. L., Bai, Y. T., Su, T. L., Ma, H. J., & Chakrabarti, P. (2023). Deep spatio-temporal graph network with self-optimization for air quality prediction. *Entropy, 25*(2), 247.

Kumar, K., & Pande, B. P. (2023). Air pollution prediction with machine learning: A case study of Indian cities. *International Journal of Environmental Science and Technology, 20*(5), 5333–5348.

Liang, Y.-C., Maimury, Y., Chen, A.H.-L., & Juarez, J. R. C. (2020). Machine learning-based prediction of air quality. *Applied Sciences, 10*(24), 9151. https://doi.org/10.3390/app10249151

Marć, M., Tobiszewski, M., Zabiegała, B., de la Guardia, M., & Namieśnik, J. (2015). Current air quality analytics and monitoring: A review. *Analytica Chimica Acta, 853*, 116–126. https://doi.org/10.1016/j.aca.2014.10.018

Moon, J., Kim, E., Hwang, J., & Hwang, E. (2024). A Task-Adaptive parameter transformation scheme for Model-Agnostic-Meta-Learning-Based Few-Shot animal sound classification. *Applied Sciences, 14*(3), 1025. https://doi.org/10.3390/app14031025

Nikolić, D., Häusler, S., Singer, W., & Maass, W. (2009). Distributed fading memory for stimulus properties in the primary visual cortex. *PLoS Biology, 7*, e1000260. https://doi.org/10.1371/journal.pbio.1000260

Ouahi, M., Khoulji, S., & Kerkeb, M. L. (2024). Predictive assessment of learners through initial interactions with encoding techniques in deep learning. *Journal of Autonomous Intelligence*, *7*(4). https://doi.org/10.32629/jai.v7i4.1443

Panda, P., & Roy, K. (2017). Learning to generate sequences with combination of hebbian and non-hebbian plasticity in recurrent spiking neural networks. *Frontiers in Neuroscience, 11*, 693. https://doi.org/10.3389/fnins.2017.00693

Schuman, C. D., Kulkarni, S. R., Parsa, M., Mitchell, J. P., Date, P., & Kay, B. (2022). Opportunities for neuromorphic computing algorithms and applications. *Nature Computational Science, 2*(1), 10–19.

Spickett, J. T., Brown, H. L., & Rumchev, K. (2011). Climate Change and Air Quality: The Potential Impact on Health. *Asia Pacific Journal of Public Health., 23*(2_suppl), 37S-45S. https://doi.org/10.1177/1010539511398114

U.S. Environmental Protection Agency. Integrated Science Assessment for Particulate Matter, EPA/600/R-08/139F, 2009.There is no corresponding record for this reference.

U.S. Environmental Protection Agency. Integrated Science Assessment for Ozone and Related Photochemical Oxidants, EPA 600/R-10/076F, 2013.There is no corresponding record for this reference

UNION, E., et al. (2008). Directive 2008/50/EC of the European parliament and of the council of 21 May 2008 on ambient air quality and cleaner air for Europe. Official Journal of the European Union.

Urbain, G., Degrave, J., Carette, B., Dambre, J., & Wyffels, F. (2017). Morphological properties of mass–spring networks for optimal locomotion learning. *Front. Neurorob., 11*, 16. https://doi.org/10.3389/fnbot.2017.00016

World Health Organization Regional Office for Europe. (2017). *Evolution of WHO air quality guidelines: Past present and future*. WHO.

Zhao H, Zhang J, & Wang K, et al. (2011). A GA-ANN model for air quality predicting. IEEE, Taiwan.

Zhang, Y., Li, P., Jin, Y., & Choe, Y. (2015). A digital liquid state machine with biologically inspired learning and its application to speech recognition. *IEEE Transactions on Neural Networks and Learning Systems, 26*(11), 2635–2649. https://doi.org/10.1109/TNNLS.2015.2388544