Machine Learning for Environmental Health: Optimizing ConcaveLSTM for Air Quality Prediction

Mohammad Diqi*[1], Hamzah[2], I Wayan Ordiyasa[3], Nurhadi Wijaya[4], Benedicto Reynaka Filio Martin[5]
Department of Informatics, Faculty of Science and Technology, Universitas Respati Yogyakarta, Sleman 55281, Indonesia
Email: [1]diqi@respati.ac.id, [2]hamzah@respati.ac.id, [3]wayan@respati.ac.id, [4]nurhadi@respati.ac.id, [5]22220013@respati.ac.id

Abstrak. *Studi ini menyelidiki optimasi model ConcaveLSTM untuk prediksi kualitas udara, berfokus pada interaksi antara panjang urutan masukan dan jumlah unit LSTM untuk meningkatkan akurasi peramalan. Melalui evaluasi berbagai konfigurasi model terhadap metrik kinerja seperti RMSE, MAE, MAPE, dan R-squared, sebuah setup optimal yang menampilkan 50 langkah masukan dan 300 neuron diidentifikasi, menunjukkan kemampuan prediktif yang superior. Temuan menekankan peran kritis penyetelan parameter model dalam menangkap dependensi temporal dalam data lingkungan. Meskipun ada keterbatasan terkait representasi dataset dan variabilitas lingkungan, penelitian ini menyediakan dasar yang solid untuk kemajuan masa depan dalam pemodelan prediktif lingkungan. Rekomendasi termasuk memperluas keragaman dataset, mengeksplorasi model hibrida, dan mengimplementasikan integrasi data waktu nyata untuk meningkatkan generalisasi model dan aplikabilitas dalam skenario dunia nyata.*
Kata Kunci: *prediksi kualitas udara, model ConcaveLSTM, optimasi parameter, dependensi temporal, pemodelan lingkungan*

Abstract. *This study investigates the optimization of the ConcaveLSTM model for air quality prediction, focusing on the interplay between input sequence lengths and the number of LSTM units to enhance forecasting accuracy. Through the evaluation of various model configurations against performance metrics such as RMSE, MAE, MAPE, and R-squared, an optimal setup featuring 50 input steps and 300 neurons was identified, demonstrating superior predictive capabilities. The findings underscore the critical role of model parameter tuning in capturing temporal dependencies within environmental data. Despite limitations related to dataset representativeness and environmental variability, the research provides a solid foundation for future advancements in predictive environmental modeling. Recommendations include expanding dataset diversity, exploring hybrid models, and implementing real-time data integration to improve model generalizability and applicability in real-world scenarios.*
Keywords: *air quality prediction, ConcaveLSTM model, parameter optimization, temporal dependencies, environmental modelling*

1. Introduction

Air quality prediction plays a crucial role in addressing public health challenges, economic impacts, and environmental issues caused by air pollution. Poor air quality poses serious health risks, including respiratory and cardiovascular diseases, and negatively affects economic productivity through increased healthcare burdens and crop yield losses. From an environmental perspective, air pollution damages ecosystems and biodiversity, underscoring the urgency for effective intervention. Using predictive algorithms in air quality forecasting promises significant improvements in accuracy and reliability, enabling more precise and responsive decision-making in response to air quality fluctuations. This directly contributes to enhancing human well-being and environmental sustainability, highlighting the importance of innovation in predictive technology for more effective air quality management.

Air pollution significantly affects both human health and the environment. The inhalation of pollutants like particulate matter, sulfur dioxide, nitrogen oxides, and carbon monoxide is linked to a range of adverse health effects, including issues with respiratory, cardiovascular, and reproductive systems, as well as DNA damage and genetic mutations [1]. The rise in air pollution levels is largely attributed to factors such as urbanization, industrial activities, and increased vehicular traffic [2], [3]. Notably, some urban areas report dangerously high levels of PM2.5 and arsenic, presenting considerable health risks [4]. Annually, air pollution is a factor in millions of premature deaths, contributing to conditions like pulmonary inflammation, exacerbated asthma, and cardiovascular problems [5]. Monitoring and analyzing air quality are essential steps for understanding pollution trends, evaluating its environmental impact, and assessing associated health risks. To mitigate the negative impacts of air pollution, it is vital to implement effective air quality management, adopt renewable energy sources, and enforce policies focused on mitigation strategies.

Predicting air quality accurately is fraught with challenges due to the spatio-temporal complexity and nonstationarity of air quality data, which complicates forecasting efforts [6]. The accuracy of predictions is further hampered by uncertainties in meteorological condition definitions and the use of inadequate data in model training [7]. Additionally, significant uncertainties in emissions inventories underscore the urgency for enhancing prediction precision [8]. Conventional machine learning techniques often overlook the sequential nature of time series data and long-term dependencies, detracting from the accuracy of predictions [9]. Challenges such as lower classification accuracy and security vulnerabilities in prediction systems further emphasize the necessity for innovative approaches. These should leverage spatial correlations, understand complex interdependencies, and utilize time series data effectively to refine the accuracy of air quality forecasts.

This study addresses the specific problem of inaccuracies and limitations inherent in existing air quality prediction methods, which often struggle to capture complex atmospheric dynamics and environmental interactions, resulting in suboptimal predictive performance. The objective of this research is to develop and validate the ConcaveLSTM algorithm, designed to enhance air quality prediction accuracy by effectively modeling nonlinear temporal relationships and spatial patterns in environmental data. The novelty and contribution of our work lie in the introduction of the ConcaveLSTM model, which incorporates a novel architecture that leverages concave functions within a Long Short-Term Memory framework to better account for the intricacies of air quality fluctuation patterns. This research significantly advances the field of air quality prediction by providing a more reliable and precise tool for forecasting, thus facilitating improved environmental management and policy-making decisions aimed at mitigating the adverse effects of air pollution.

2. Literature Review

Recent studies have underscored the critical impacts of air quality on human health, employing diverse methodologies to elucidate this relationship. Air quality evaluations over determined periods have focused on assessing pollutant concentrations such as arsenic, carbon monoxide, and PM2.5, highlighting their health implications [3]. Additionally, modeling frameworks have been leveraged to explore the intricate connections between air pollution, climate change, and their consequent effects on human health [10]. The advent of Artificial Intelligence (AI)-based predictive models has marked a significant stride toward forecasting air quality impacts on health and quantifying the risks associated with exposure to air pollutants [11]. Moreover, the development of Air Quality Health Index (AQHI) systems aims to articulate health risks more precisely for particular risk groups, facilitating analysis of their spatiotemporal distribution patterns [12]. Such research endeavors have yielded invaluable insights, informing strategies devised by stakeholders and policymakers to mitigate air pollution effectively [13].

The realm of air quality prediction has seen the application of various models, with the long short-term memory network (LSTM) and gated recurrent unit (GRU) being notably prevalent [6]. Other employed models include multi-linear regression (MLR), multi-linear perceptron (MLP), generalized regression neural network (GRNN), and adaptive Neuro-Fuzzy inference system (ANFIS) [8]. The proposition of graph convolutional networks (GCN) for air quality prediction represents an innovative approach to model spatial and temporal correlations, further enriched by the introduction of the CoupledGT model, which assimilates geospatial-temporal couplings for enhanced predictive accuracy [14], [15]. These models aim to refine air quality forecasts by considering various influencing factors, including pollutant sources, meteorological conditions, and spatial-temporal data couplings.

LSTM models, renowned for their efficacy in time series prediction, have been widely adopted for air pollution forecasting. Their capability to capture long-term dependencies and adapt to rapid changes has been demonstrated through their integration with other deep learning methodologies, enhancing predictive precision [16]. The EMD-LSTM algorithm, designed for forecasting short time series characterized by uncertainty and swift changes, has outperformed conventional prediction methods [17]. Moreover, the LSTM-BNN model, a Bayesian neural network built upon LSTM, has significantly reduced forecasting errors for air pollutants [18]. The deployment of multi-layer LSTM artificial neural networks for predicting future air pollutant concentrations has showcased superior performance over single-layer architectures [19].

However, LSTM models exhibit strengths and limitations in air pollution forecasting. Their potent architecture for time-series prediction stands out, enabling their application in replacing physical sensors for indoor air pollutants and rendering them highly adaptable to diverse problems [8], [20]. Conversely, their limited generalization capabilities and challenges in long-term forecasting due to inadequate representation of atmospheric processes related to pollution transport mark their weaknesses [21], [22]. Despite these drawbacks, LSTMs offer promising prospects for enhancing air quality forecasting, facilitating real-time monitoring of occupant exposure, and improving building operations [23].

Air quality prediction models crucially depend on external variables like weather and emissions to forecast pollution levels accurately. The significance of these variables is emphasized in the literature, advocating for the exploration of their complex interactions with air pollutant concentrations through various models, including deep learning techniques, neural networks, and fuzzy inference systems [15], [24], [25], [26]. These models aim to establish interpretable relationships, improving prediction accuracy and elucidating meteorological influences on air pollution.

The discussion on model complexity and the accuracy-interpretability trade-off in air quality prediction highlights the widespread use of deep learning models like LSTM and GRU, appreciated for their ability to unravel complex nonlinear relationships and achieve high accuracy [8], [27], [28], [29]. However, their interpretability remains a challenge, prompting some studies to propose models that blend accuracy with interpretability by analyzing variable contributions through techniques like decision plots and Shapley Additive Explanations [23].

The ConcaveLSTM model in this research represents a significant leap in air quality prediction. By ingeniously incorporating concave functions within the robust framework of LSTM, our work not only overcomes the inherent limitations of traditional prediction models but also highlights the ability to model complex temporal and spatial dynamics adeptly. This innovation addresses the pressing need for precise air quality forecasts, thus facilitating proactive and informed decision-making in environmental management and public health policy. The originality of this approach is underscored by its potential to provide more accurate predictions than existing methodologies, contributing a unique perspective to the ongoing discourse on improving air quality monitoring and management. This research advances cutting-edge air quality prediction technology. It sets a new standard for future endeavors in the domain, positioning the ConcaveLSTM model as a crucial tool for researchers, policymakers, and stakeholders committed to combating air pollution and its wide-ranging impacts.

3. Method

3.1. Dataset and Data Preprocessing

The air quality dataset, sourced from Kaggle.com, covers the period from March 2004 to February 2005, comprising 9470 records across 10 attributes: date, time, CO(GT), PT08.S1(CO), NMHC(GT), C6H6(GT), PT08.S2(NMHC), NOx(GT), PT08.S3(NOx), and NO2(GT). Several preprocessing steps are undertaken to ensure the dataset's integrity and readiness for analysis.

Firstly, records displaying zero volume are purged from the dataset to prevent skewed analyses and potential inaccuracies. This elimination is critical as zero-valued records can significantly distort the representation and analysis outcomes.

Subsequently, the dataset undergoes further cleansing to address missing or NaN values by removing such entries, which is vital for preserving the dataset's reliability and ensuring smooth analysis processes. Following the cleaning phase, normalization is applied to standardize the attribute values across the dataset, employing a normalization formula that scales the values to a range between 0 and 1. This normalization process is essential for ensuring uniform scales across different attributes, thereby enabling accurate and meaningful analysis of the air quality data. These preprocessing steps collectively enhance the dataset's reliability and utility for further analyses and modeling efforts.

3.2. Data Splitting

Following the preprocessing steps, the refined dataset includes 9.357 entries. These entries are subsequently segmented into two distinct parts to facilitate the training and testing phases. The primary segment contains 9.317 entries, of which 80% (7.453 entries) are earmarked for model training, and the remaining 20% (1.863 entries) are set aside for testing. The secondary segment, comprising 40 entries, is solely used for testing purposes. This approach to splitting the data guarantees a comprehensive use of the dataset for model training while ensuring a separate portion is available for performance evaluation.

3.3. ConcaveLSTM

The ConcaveLSTM model is a deep learning model designed for air quality prediction, leveraging the Long Short-Term Memory (LSTM) network architecture. This model incorporates multiple LSTM layers and a Bidirectional LSTM layer, aiming to capture both short-term and long-term dependencies in the sequential data of air quality measurements.

The core component of the ConcaveLSTM model is the LSTM layer. An LSTM unit is designed to overcome the vanishing gradient problem common in traditional Recurrent Neural Networks (RNNs) by introducing three gates: the input gate, the forget gate, and the output gate. These gates regulate the flow of information into and out of the cell, allowing it to retain important long-term dependencies and discard irrelevant information. The mathematical equations governing an LSTM cell are shown in Equations 1-4.

Forget Gate

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

Input Gate

$$\begin{aligned} i_t &= \sigma(W_i \cdot [h_{t-1}, x_t] + b_i) \\ \widetilde{C_t} &= \tanh(W_C \cdot [h_{t-1}, x_t] + b_C) \end{aligned} \tag{2}$$

Cell State Update

$$C_t = f_t * C_{t-1} + i_t * \widetilde{C_t} \tag{3}$$

Output Gate

$$\begin{aligned} o_t &= \sigma(W_o[h_{t-1}, x_t] + b_o) \\ h_t &= o_t * \tanh(C_t) \end{aligned} \tag{4}$$

Here, σ denotes the sigmoid activation function, $\tanh$ is the hyperbolic tangent activation function, W and b represent the weights and biases for each gate, respectively, x_t is the input at time step t, h_t is the output vector of the LSTM cell, and C_t is the cell state vector.

The Bidirectional LSTM processes the data in both forward and backward directions (i.e., from past to future and future to past). This allows the model to have both backward and forward information about the sequence at every point. The mathematical representation is similar to the standard LSTM but processed in two directions and then combined.

The Concatenation Layer merges the features learned from the stacked LSTM layers and the Bidirectional LSTM layer. The concatenation does not involve a specific mathematical operation on the data but combines the feature sets for a comprehensive representation.

Through its architecture, the ConcaveLSTM model aims to capture the complex temporal dynamics in air quality data, providing accurate predictions by effectively handling both the short-term and long-term dependencies.

3.4. Parameter Settings

The setup involves defining specific parameters to optimize performance in configuring the model for air quality prediction. The model is structured to forecast 40 steps, utilizing varying input sequences of 30, 50, and 70 time steps to capture different temporal dependencies in the data. It employs three LSTM layers with a progressive increase in neurons: 100 for the first layer, 200 for the second, and 300 for the third, ensuring a comprehensive feature extraction capability. The output layer is configured with 300 neurons to match the complexity of the predictions. The 'adam' optimizer is chosen for its effectiveness in handling sparse gradients and adapting learning rates, combined with the 'mean squared error' (mse) loss function, which is standard for regression problems like air quality forecasting. The training process is set to run for 100 epochs, with a batch size of 32, to balance computational efficiency and the model's ability to converge to a solution, ensuring that the model is adequately trained to predict air quality with high accuracy.

3.5. Model Evaluation

The efficacy of the ConcaveLSTM model in forecasting air quality is assessed through various metrics, providing distinct perspectives on the accuracy and dependability of its predictions. Root Mean Squared Error (RMSE) serves as a conventional metric for quantifying a model's error in numerical predictions, calculated as the square root of the mean squared deviations between forecasted and observed values, as mathematically outlined in Equation 5.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i)^2} \tag{5}$$

where y_i is the actual value, $\hat{y}_i$ is the predicted value, and n is the number of observations. RMSE is particularly useful for highlighting larger errors because it squares the errors before averaging, thus penalizing more significant mistakes more heavily than smaller ones.

Mean Absolute Error (MAE) quantifies the average size of errors in a series of forecasts, disregarding the direction of these errors. It is determined by computing the mean of the absolute discrepancies between forecasted and real values, as specified in Equation 6. MAE offers a direct assessment of the accuracy of predictions, where smaller values denote improved accuracy. Contrary to RMSE, MAE assigns equal weight to all errors, making it a stronger metric for evaluating performance in the presence of significant outliers.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|y_i - \hat{y}_i| \tag{6}$$

Mean Absolute Percentage Error (MAPE) represents the accuracy of predictions by indicating the error as a percentage, offering valuable insights into the relative accuracy of forecasts. This measure is detailed in Equation 7. MAPE is advantageous for conducting comparisons between various datasets or predictive models because it delivers an error metric that is independent of scale, thereby simplifying the understanding of a model's accuracy through percentage-based errors.

$$MAPE = \frac{100\%}{n} \sum_{i=1}^{n} \left| \frac{y_i - \widehat{y_i}}{y_i} \right| \tag{7}$$

R-squared (R2), also known as the coefficient of determination, measures the proportion of the variance in the dependent variable that is predictable from the independent variables. It is calculated in Equation 8.

$$R2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y_i})^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2} \tag{8}$$

where $\bar{y}$ is the mean of the actual values. R2 is a statistical measure that provides insights into the goodness-of-fit of the model, with values closer to 1 indicating a better fit, meaning the model can explain a higher proportion of the variance in the observed data.

Together, these metrics offer a comprehensive evaluation of the ConcaveLSTM model's performance, considering both its accuracy and its ability to capture the variance in air quality data effectively.

4. Result and Discussion

4.1. Performance Evaluation

Table 1 presents a comprehensive summary of the performance evaluation for the ConcaveLSTM model applied to air quality prediction based on the testing dataset. The table outlines the results of nine distinct predictive configurations, delineating the interplay between the number of input steps (`n_steps_in`) and the number of neurons (`n_units`) within the LSTM layers. Each configuration has been assessed using four key performance metrics: RMSE, MAE, MAPE, and R2.

Table 1. Summary of the performance evaluation for the ConcaveLSTM model

prediction	n_steps_in	n_units	RMSE	MAE	MAPE	R2
1	30	100	0,02203	0,01549	0,02722	0,8946
2	30	200	0,0201	0,01527	0,02792	0,91232
3	30	300	0,02202	0,01281	0,02162	0,89469
4	50	100	0,02153	0,01484	0,02572	0,89939
5	50	200	0,02388	0,01494	0,02574	0,87615
6	50	300	**0,01413**	**0,00929**	**0,01612**	**0,95668**
7	70	100	0,02466	0,01868	0,03407	0,86799
8	70	200	0,01932	0,01493	0,02686	0,91892
9	70	300	0,02327	0,01555	0,02694	0,88246

Upon examining Table 1, which delineates the performance of various ConcaveLSTM configurations, we observe that the configuration with 50 input steps and 300 LSTM units notably outperforms the others, achieving the lowest RMSE and MAE values and the highest R-squared value of 0.95668, indicating a strong fit between the predicted and actual values. This suggests an optimal balance between the ability to capture temporal dependencies and model complexity, reducing the error and improving predictive accuracy. Conversely, configurations with 70 input steps do not consistently yield better results, especially with 100 units, where the model's performance is the weakest, suggesting potential overfitting or insufficient model complexity for longer sequences. It appears that increasing the model's complexity by adding more units does not linearly enhance performance and may lead to diminishing returns, as seen with the 70 input step configurations. Overall, a mid-range input step size paired with a higher number of units seems to provide the best predictive performance for this dataset, underscoring the necessity of fine-tuning model parameters to achieve optimal results.

Figure 1 provides a graphical illustration of the air quality forecasts over a period of 40 days, placing the model's predictions against the backdrop of actual observed values. The actual air quality measurements are plotted as a blue line, creating a reference point against which the predictions can be assessed. The figure includes nine different predictive scenarios, each represented by a distinct colored line that traces the predictions yielded by various configurations of the ConcaveLSTM model, as delineated in Table 1. This visual format enables a direct and detailed comparison between the predicted and actual air quality levels, offering a clear perspective on the model's precision and the reliability of its predictions over the given period.

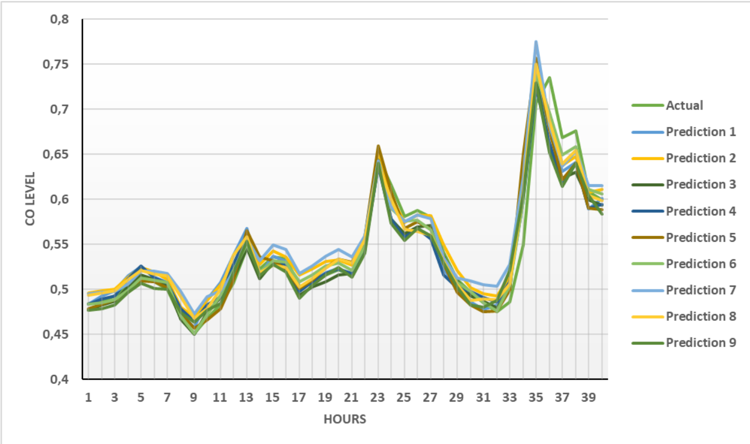

Figure 1. Comparison of Actual and Predicted Values

A closer look at Figure 1 reveals that the blue line, indicative of the actual air quality, is pivotal for gauging the precision of the model's predictions. The spectrum of colored lines, each corresponding to a unique predictive scenario, illustrates the varying degrees of the model's accuracy. Some of the model's forecasts align closely with the actual air quality trajectory, highlighting its efficacy in capturing and reflecting the relevant temporal patterns.

4.2. Summarization of Key Findings

The research problem at hand focused on addressing the limitations of existing air quality prediction models by implementing and optimizing a ConcaveLSTM model. The major findings from the study reveal that the model's performance is sensitive to the configuration of input steps and the number of LSTM units. Notably, the optimal model configuration with 50 input steps and 300 LSTM units achieved the most accurate predictions, indicated by the lowest RMSE and MAE and the highest R-squared value. This configuration effectively balanced the complexity of the model and its ability to capture temporal dependencies without overfitting. Conversely, configurations with 70 input steps did not consistently enhance performance, suggesting a nuanced relationship between input sequence length and model complexity. These insights pave the way for targeted improvements in air quality forecasting, emphasizing the need for fine-tuning model parameters to the characteristics of the specific dataset in question.

4.3. Result Interpretations

The patterns observed in the data suggest a complex interplay between the length of input sequences and the number of neurons in determining the accuracy of the ConcaveLSTM model's air quality predictions. While an increase in the number of neurons generally correlates with improved accuracy, this trend plateaus and even reverses with longer input sequences, indicating a point at which additional complexity ceases to yield benefits and may lead to overfitting. The results largely met expectations in the mid-range of input steps, where the model captured temporal patterns effectively; however, the diminishing returns at higher complexity levels were somewhat unexpected. This could be attributed to the model's increased difficulty in generalizing from the training data to unseen data or to the inherent noise within the longer input sequences that may have obscured underlying patterns. Alternative explanations might include the need for more nuanced architectures or advanced regularization techniques that can better harness the information within longer sequence lengths without compromising the model's predictive power.

4.4. Research Implications

The research outcomes underscore the critical significance of model parameter optimization in the domain of air quality prediction using machine learning algorithms, aligning with existing literature that highlights the delicate balance between model complexity and overfitting. By demonstrating an optimal configuration for the ConcaveLSTM model, this study contributes new insights into the specific dynamics of sequence length and neuron count, thereby refining our understanding of time-series prediction for environmental data. These findings are particularly relevant for the design of advanced predictive models, providing a concrete benchmark for future research in this field. Furthermore, the research bridges a gap in current knowledge by quantitatively illustrating the non-linear relationship between input sequence length and prediction accuracy, offering a nuanced perspective that could inform the development of more sophisticated models with improved generalizability and reliability in real-world air quality forecasting applications.

4.5. Research Limitations

This study concludes that the careful calibration of input sequence lengths and the number of neurons in LSTM-based models is pivotal for accurate air quality forecasting, which substantiates the hypothesis that model complexity influences prediction performance. Although the research faced limitations, such as the potential non-representativeness of the dataset for broader applications and the unexplored effects of varying environmental conditions on the model's efficacy, the robustness of the results within the studied parameters confirms their validity. The findings remain cogent in answering the research question, affirming the ConcaveLSTM model's capability to predict air quality when optimally tuned, and they provide a foundational methodology for enhancing predictive accuracy in environmental machine learning applications. Despite the limitations, these results contribute valuable knowledge to the field, offering a methodical approach for future studies to refine and apply to a wider array of datasets and conditions.

4.6. Recommendations for Future Research

For practical implementation, future research should focus on expanding the dataset to encompass a wider variety of environmental conditions and geographical areas to enhance the generalizability of the ConcaveLSTM model. Concrete ideas for subsequent studies include exploring hybrid models that combine LSTM with other machine learning techniques to potentially capture more complex patterns and dependencies in air quality data. Additionally, research could investigate the integration of real-time data streams to enable dynamic model updating, improving the responsiveness of air quality predictions. Incorporating explainable AI methodologies would also be valuable, allowing for greater transparency and trust in the model's predictions. Finally, examining the model's performance in a live environment could provide insights into its real-world efficacy and help identify further refinements needed to optimize its predictive capabilities.

5. Conclusion

The research conducted provides a significant step forward in the predictive modeling of air quality using the ConcaveLSTM model. Through rigorous testing and evaluation, we've determined an optimal model configuration that offers a balance between input sequence length and neuron density, thereby achieving the highest accuracy in air quality forecasting within the constraints of the dataset used. The study reinforces the importance of parameter tuning in LSTM models and contributes to the body of knowledge by identifying key factors that influence model performance. Despite dataset limitations, the research findings are robust and demonstrate that with the right adjustments, LSTM models are a powerful tool for air quality prediction. The study lays the groundwork for future exploration in the field, suggesting that expanding the diversity of data, integrating real-time monitoring, and developing models with greater interpretability are essential next steps. These findings and recommendations serve as a guide for future research aimed at enhancing environmental monitoring and policy-making, ultimately contributing to better public health and environmental management.

References

[1] "A Study of Environmental Impact of Air Pollution on Human Health: Pm Deposition Modelling," *Global Nest Journal*, 2023, doi: 10.30955/gnj.004613.

[2] S. K. Maji, S. Ahmed, M. Kaur-Sidhu, S. Mor, and K. Ravindra, "Health Risks of Major Air Pollutants, Their Drivers and Mitigation Strategies: A Review," *Air Soil and Water Research*, 2023, doi: 10.1177/11786221231154659.

[3] I. M. Tănasă, M. M. Cazacu, and B. Sluser, "Air Quality Integrated Assessment: Environmental Impacts, Risks and Human Health Hazards," *Applied Sciences*, 2023, doi: 10.3390/app13021222.

[4] L. Šulc, P. Gregor, J. Kalina, O. Mikeš, T. Janoš, and P. Čupr, "City-Scale Assessment of Long-Term Air Quality Impacts on the Respiratory and Cardiovascular Health," *Frontiers in Public Health*, 2022, doi: 10.3389/fpubh.2022.1006536.

[5] R. Singh, R. Jakhar, R. Raj, and P. Sachar, "A Comprehensive Study on Impacts of Air Pollution on Environment and Human Health," *International Journal of Recent Technology and Engineering*, 2022, doi: 10.35940/ijrte.a6976.0511122.

[6] S. Sui and Q. Han, "Multi-View Multi-Task Spatiotemporal Graph Convolutional Network for Air Quality Prediction," *The Science of the Total Environment*, 2023, doi: 10.1016/j.scitotenv.2023.164699.

[7] V. Singh, K. K. Singh, and S. H. Singh, "A Secure and Energy-Efficient Framework for Air Quality Prediction Using Smart Sensors and ISHO-DCNN," *International Journal of Sensors Wireless Communications and Control*, 2023, doi: 10.2174/2210327913666230504122805.

[8] G. Cican, A.-N. Buturache, and R. Mirea, "Applying Machine Learning Techniques in Air Quality Prediction—A Bucharest City Case Study," *Sustainability*, 2023, doi: 10.3390/su15118445.

[9] V. K. Rani and R. Vallikannu, "Air Pollution Prediction Model for Estimating and Forecasting Particulate Matter PM<SUB align="right">2.5 Concentrations Using SOSE-RMLP Approach," *International Journal of Global Warming*, 2023, doi: 10.1504/ijgw.2023.130128.

[10] S. Coelho, J. Ferreira, D. F. Carvalho, and M. Lopes, "Health Impact Assessment of Air Pollution Under a Climate Change Scenario: Methodology and Case Study Application," *Sustainability*, 2022, doi: 10.3390/su142114309.

[11] E. X. Neo *et al.*, "Towards Integrated Air Pollution Monitoring and Health Impact Assessment Using Federated Learning: A Systematic Review," *Frontiers in Public Health*, 2022, doi: 10.3389/fpubh.2022.851553.

[12] X. Deng *et al.*, "Disease Specific Air Quality Health Index (AQHI) for Spatiotemporal Health Risk Assessment of Multi-Air Pollutants," *Environmental Research*, 2023, doi: 10.1016/j.envres.2023.115943.

[13] K. Kamboj and A. Mathur, "Air Quality Assessment With Human Health Effects for Kota Metropolis, Rajasthan India.," *Current World Environment*, 2022, doi: 10.12944/cwe.17.2.10.

[14] S. Ren, B. Guo, K. Li, Q. Wang, and Z. Yu, "CoupledGT: Coupled Geospatial-Temporal Data Modeling for Air Quality Prediction," *Acm Transactions on Knowledge Discovery From Data*, 2023, doi: 10.1145/3604616.

[15] F. Karimi, J. Amanollahi, M. Reisi, and M. Darand, "Prediction of Air Quality Using Vertical Atmospheric Condition and Developing Hybrid Models," *Advances in Space Research*, 2023, doi: 10.1016/j.asr.2023.04.020.

[16] S. Arifin *et al.*, "Long Short-Term Memory (LSTM): Trends and Future Research Potential," *International Journal of Emerging Technology and Advanced Engineering*, 2023, doi: 10.46338/ijetae0523_04.

[17] Y. Liu and D. Wu, "Research on the Prediction of Short Time Series Based on EMD-LSTM," *Journal of Computational Methods in Sciences and Engineering*, 2023, doi: 10.3233/jcm-226860.

[18] X. Xiang and J. Zhou, "An Excess Entropy Approach to Classify Long-Term and Short-Term Memory Stationary Time Series," *Mathematics*, 2023, doi: 10.3390/math11112448.

[19] H. M. Nguyen *et al.*, "Long Short-Term Memory Bayesian Neural Network for Air Pollution Forecast," *Ieee Access*, 2023, doi: 10.1109/access.2023.3265725.

[20] M. Gabriel and T. Auer, "LSTM Deep Learning Models for Virtual Sensing of Indoor Air Pollutants: A Feasible Alternative to Physical Sensors," *Buildings*, 2023, doi: 10.3390/buildings13071684.

[21] M. Teng *et al.*, "72-Hour Real-Time Forecasting of Ambient PM2.5 by Hybrid Graph Deep Neural Network With Aggregated Neighborhood Spatiotemporal Information," *Environment International*, 2023, doi: 10.1016/j.envint.2023.107971.

[22] E. Ramentol, S. Grimm, M. Stinzendörfer, and A. Wagner, "Short-Term Air Pollution Forecasting Using Embeddings in Neural Networks," *Atmosphere*, 2023, doi: 10.3390/atmos14020298.

[23] Z. Zhang, H. Chen, and X. Huang, "Prediction of Air Quality Combining Wavelet Transform, DCCA Correlation Analysis and LSTM Model," *Applied Sciences*, 2023, doi: 10.3390/app13052796.

[24] K. Pan, J. Lü, J. Li, and Z. Xu, "A Hybrid Autoformer Network for Air Pollution Forecasting Based on External Factor Optimization," *Atmosphere*, 2023, doi: 10.3390/atmos14050869.

[25] C. Huang, X. Zhao, W. Cheng, Q. Ji, Q. Duan, and Y. Han, "Statistical Inference of Dynamic Conditional Generalized Pareto Distribution With Weather and Air Quality Factors," *Mathematics*, 2022, doi: 10.3390/math10091433.

[26] Y. Huang, B. Guo, H. Sun, H. Liu, and S. X. Chen, "Relative Importance of Meteorological Variables on Air Quality and Role of Boundary Layer Height," *Atmospheric Environment*, 2021, doi: 10.1016/j.atmosenv.2021.118737.

[27] X.-B. Jin *et al.*, "Variational Bayesian Network With Information Interpretability Filtering for Air Quality Forecasting," *Mathematics*, 2023, doi: 10.3390/math11040837.

[28] T. Li *et al.*, "Contributions of Various Driving Factors to Air Pollution Events: Interpretability Analysis From Machine Learning Perspective," *Environment International*, 2023, doi: 10.1016/j.envint.2023.107861.

[29] W. Huang, Y. Cao, X. Cheng, and Z. Guo, "Research on Air Quality Prediction Based on Improved Long Short-Term Memory Network Algorithm," *Peerj Computer Science*, 2022, doi: 10.7717/peerj-cs.1187.