FORECASTED MODELING FOR AIR QUALITY INDEX IN VIETNAMESE TOURIST DESTINATIONS: LEVERAGING DEEP LEARNING APPROACHES

Tu Anh Hoang Nguyen[1], Quang-Dieu Nguyen[1], Cong-Bang Luan Nguyen[1], Nguyen Trung Ky[1,2]*

[1]*International University*
[2]*Vietnam National University*
*Corresponding author: *Nguyen Trung Ky, ntky@hcmiu.edu.vn*

GENERAL INFORMATION

Received date: 30/03/2024

Revised date: 20/05/2024

Accepted date: 11/07/2024

KEYWORD

Air Pollution;
Air Quality Index;
Deep Learning;
Low-cost sensors;
Smart Cities;

ABSTRACT

Severe air pollution in Vietnam's tourism areas has become a significant economic issue in recent years. While many studies have found a link between population exposure to air pollution and poor health outcomes, short-term exposure to air pollutants in high-pollution zones can result in acute health consequences; thus, poor air quality jeopardizes visitors' health and well-being and threatens the tourism industry's sustainability. As a result, attempts to correctly estimate the air quality index (AQI) are crucial for effective air quality management, a challenge that smart cities must address as they become more developed soon. However, there are some challenges to predicting AQI. First, the results are influenced by various factors that low-cost sensors frequently skip due to the nonlinear and dynamic nature of multivariate air quality time-series data, leaving a gap for enhancements. Second, standard prediction algorithms often use the training data at fixed intervals and require as many available attributes as possible. This work reviews these issues by applying many Recurrent Neural Network (RNN) deep-learning models for the AQI dataset from PAM AIR stations in 10 Vietnamese tourism areas. Then, it compares each model's impact on the data set by leveraging deep learning models for early predictions based on limited but crucial parameters such as particulate matter 2.5 microns (PM2.5) levels, humidity, and temperature. It presents an appealing method for tackling air pollution problems while dataset quality is uncertain. These findings will result in a fast, efficient, cost-effective, and reliable model that would help reduce the impact on health and add to the literature on meteorology and air pollution while giving theoretical insights and practical guidance in assessing AQI and its dangers. It would support the government in adopting efficient pollution control measures to minimize emissions from various sources by making informed decisions proactively to address air pollution challenges before they increase.

1. INTRODUCTION

According to the World Health Organization (WHO), air pollution is a significant global environmental threat that affects all continents (World Health Organization, n.d.). Southeast Asia and the Western Pacific are the most impacted regions, with Vietnam among the most highly affected. On December 4, 2023, Hanoi recorded an air quality index (AQI) 182 (SGGPO, 2023), placing it third among the world's most polluted cities. This problem necessitates a dedicated study on accurately evaluating the AQI, critical for effective air quality management—a key component in developing smart cities.

Moreover, climate change and environmental pollution in this modern world significantly affect human health. One factor that causes climate change and environmental pollution is air quality. Therefore, many studies (Doreswamy et al., 2020; Minh et al., 2021; Pant et al., 2018; Pruthi & Liu, 2022) have focused on accurately predicting air quality, providing recommendations for local authorities to improve air quality and helping to improve the quality of life.

This research investigates the feasibility of using Recurrent Neural Network (RNN) variants from deep learning models on an AQI dataset gathered from PAM Air stations across ten Vietnamese sites representing diverse locales (*PAM Air*, n.d.). By employing deep learning algorithms to forecast AQI, primarily focusing on critical features such as PM2.5 levels, humidity, and temperature, the study aims to develop a comprehensive AQI forecasting model. The goal is to create a speedy, efficient, cost-effective, and reliable model, providing practical recommendations for AQI estimation and risk assessment.

The study's significance lies in its dual contributions: (i) it offers a comparative and robust experiment for evaluating deep learning models in AQI forecasting using the cross-validation time series split technique, and (ii) it determines which deep learning technique outperforms others within the PAM Air dataset context. The accurate prediction of AQI might be used to evaluate the degree of air pollution and its influence on health. Therefore, our contributions are crucial because they provide insights into improving air quality forecasting, which can help local authorities develop more effective strategies for reducing air pollution based on the early forecasting result, thus enhancing public health and environmental quality.

2. RELATED WORK

Previous research on air pollution in Vietnam has primarily concentrated on large provinces or specialized occupational cohorts (Pant et al., 2018), leaving a significant gap in understanding visitors' brief exposure to air pollutants in less-studied, infrastructure-dependent areas. Moreover, while existing research on using deep learning to predict air quality has produced models tailored to specific locales, such as the Weather Research and Forecasting (WRF) model for PM2.5 prediction in Ho Chi Minh City (HCMC) (Minh et al., 2021) and various combined approaches, including nature-inspired deep learning in India (Pruthi & Liu, 2022) and regression-based modeling in Taiwan (Doreswamy et al., 2020); these studies are limited in scope. They either focus on specific models or prediction zones, thereby underscoring the potential for developing a more holistic and adaptive air quality forecasting model by incorporating data from multiple methodologies across different locations.

In recent studies, one problem is the need to comprehensively evaluate and improve AQI forecasting models by integrating PM2.5 levels with temporal, sensor, location, and meteorological data (Zhu et al., 2021), providing a CNN-LSTM hybrid model in PM2.5 forecasting. In addition, (Nguyen et al., 2023) proposed a CNN-GRU model to predict the hour-average value of pollutants and AQI by learning rich features, where CNN is used as feature extraction, and GRU is used as an auto-regression model to prevent the "gradient-explosion" and RNN's "gradient-disappearance" problems. Zhu and Nguyen used CNN to solve the issue of insufficient features, while GRU and LSTM help avoid the gradient explosion, thereby improving prediction accuracy.

3. APPROACH

3.1. Problem formulation

With the context and the goal set above, this work set a problem to solve the difficulty of building an early forecasting system for diverse sites in Vietnam by using deep learning models to predict the AQI. This problem yields an objective of producing reliable AQI estimations based on the limited data gathered from monitoring stations via sensing devices. The research proposes to capture complicated relationships between four fundamental features through feature extraction and training to ensure accurate forecasts. Furthermore, the study investigates the accuracy and comparison to serve as a baseline for specific locations. This technique seeks to build a dependable and effective forecasting system for various locations in Vietnam, emphasizing the AQI, a crucial indicator of air pollution that has received substantial research in the field (Suman, 2021). Figure 1 shows the details of the steps taken for this approach.

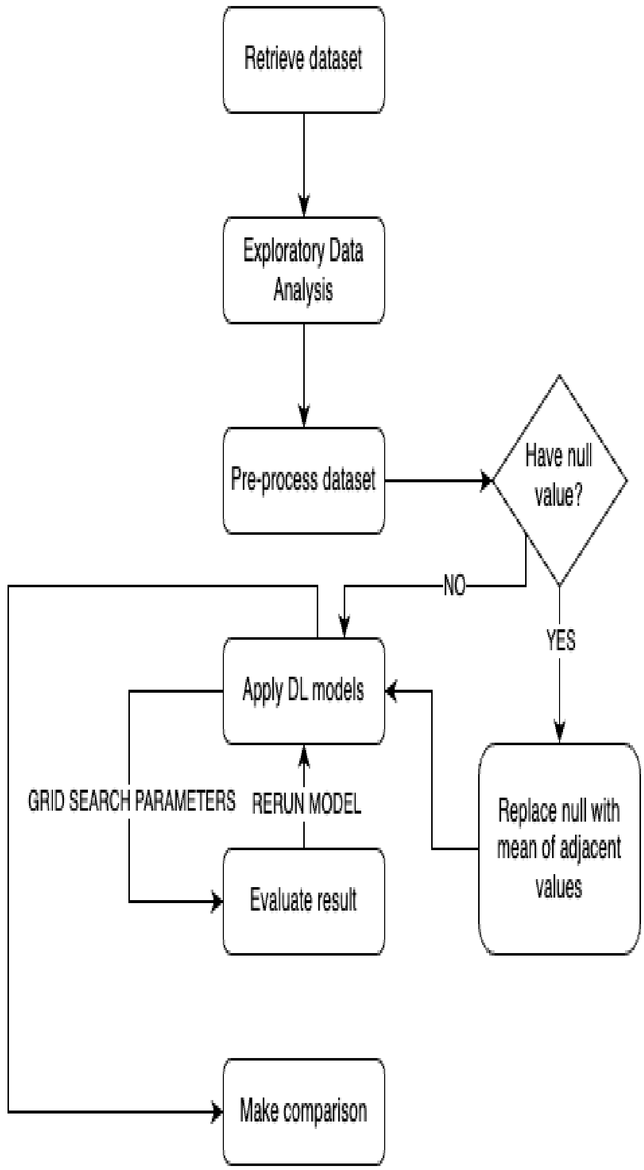

Figure 1. The procedure of the proposed approach.

3.2. Dataset collection

From April 1, 2022, at midnight to February 28, 2023, at 11:00 PM, the PAM Air quality monitoring network gathered data on five essential attributes: observed AQI level, timestamp, PM2.5 value, humidity, and temperature in 10 provinces. These provinces cover Vietnam's key regions, from Ca Mau to Ha Giang, the southern to northernmost point respectively, with some popular tourist destinations and provide a comprehensive picture of Vietnamese terrain and different landscapes, as Figure 2 depicts the distribution of the sensor on the Vietnam terrain.

Figure 2. PAM Air stations' location on the Vietnam map (red points).

This selection ensures thorough geographic coverage and a representative sample for the inquiry, including various climatic variables from multiple places. However, the dataset experienced faults owing to device failures and unexpected issues during data collection, resulting in incomplete planned hourly numbers. As a result, considerable pre-processing is necessary before applying the models to remove null values and maintain data integrity. This stage involves thoroughly analyzing the data and assertion of missing values. Table 1 summarizes the null values for each attribute in the dataset.

Table 1. The number of null values in the data set used by each attribute.

Province	PM2.5	Humidity	Temperature
Cao Bang	1	1	0
Lao Cai	1	1	0
Quang Ninh	0	1	4512
Ha Noi	2	0	7
Quang Binh	2	5336	0
Da Nang	2	0	1
Dak Lak	41	0	39
HCMC	3	3221	0
Kien Giang	0	3	2
Ca Mau	2	0	1

3.3. Dataset pre-processing

During the pre-processing procedure for time series data, considerable attention is taken to assure the dataset's integrity and completeness. Initially, the dataset is filtered to include only days with a minimum of 22 hours of recorded data, ensuring that each day contains at least 22 hourly records. This threshold guarantees adequate data coverage while accounting for slight gaps that may occur due to device faults or other unanticipated challenges. From then on, any missing hours within these filtered days, for example, the 23rd and 24th hours, are filled by taking the average of the three closest adjacent data points. For instance, if there is no data for hour 9, the average values from hours 7, 8, and 10 fill the gap. This strategy helps retain the data's temporal continuity and reduces the effect of missing values on future studies. Each trait, such as PM2.5, humidity, and temperature, receives the same treatment for missing data to ensure uniformity across all variables. This technique retains the dataset's general structure and patterns, providing the pre-processed data indicates the original recordings. Following these pre-processing procedures, the transformed data sizes are rigorously recorded and detailed in Table 2. This thorough pre-processing technique pledges that the dataset is clean, complete, and ready for accurate and trustworthy modeling, strengthening the following exploratory data analyses and conclusions.

Table 2. The number of records before and after the pre-processing step.

Province	Before	After
Cao Bang	7865	8016
Lao Cai	7889	8016
Quang Ninh	7855	7990
Ha Noi	4864	8016
Quang Binh	7900	8016
Da Nang	7750	7896
Dak Lak	4728	8016
HCMC	7906	8016
Kien Giang	7846	8016
Ca Mau	6894	7320

This approach is advantageous since PM2.5 is the critical determinant for AQI (Ghobakhloo et al., 2023). It has few null values, ensuring the accuracy of AQI estimates. This rigorous pre-processing technique guarantees that the dataset is complete and consistent, allowing for accurate and trustworthy AQI forecasts using time series data.

3.4. Models

3.4.1. RNN variants and CNN-based models

The experiment examines time-series data using a range of deep learning models of RNN variants such as Long-Short Term Memory (LSTM), Gated Recurrent Units (GRU), and Convolutional Neural Networks (CNN) integrated with LSTM and GRU. Each ensemble was also derived from the basic model like CNN-GRU and CNN-LSTM. These models combine CNN's feature extraction abilities with the temporal processing capabilities of LSTMs and GRUs. GRU and LSTM have also been extensively tested in conventional and bidirectional topologies to provide additional performance comparisons (Méndez et al., 2023). These models use batch size 32 to examine historical patterns and forecast future air quality, including temporal trends from the preceding 23 hours of data. The primary objective is to generate solid and trustworthy forecasts for the following day's air quality, using each model's capabilities to deal with the intricacies of temporal data. By concentrating on the most recent 23 hours, these models are well-equipped to identify and anticipate patterns that affect air quality. Python was the programming language for the whole process, including the TensorFlow library (Abadi et al., 2016) for hyperparameter modification and model selection. The study attempts to identify correlations and trends in the data by optimizing hyperparameters using a grid or random search on cross-validation folds, determining the best configuration for each model to increase prediction accuracy.

3.4.2. Parameters grid search

A parameter grid search was used to determine the ideal parameters for each model to achieve optimum performance. The CNN-GRU, CNN-LSTM, Bidirectional GRU (BiGRU), Bidirectional LSTM (BiLSTM), LSTM, and GRU models were repeatedly trained for 50 epochs and assessed using various parameters such as unit ratio, dropout ratio, pooling, kernel, and filter size. This thorough parameter tuning process ensures that each model performs to its total capacity, producing accurate and dependable air quality projections for each province and model configuration.

3.4.3. Cross-validation chaining

The cross-validation chaining approach analyzes the training dataset by systematically examining each new testing fold, integrating it with the training data, and continuing until all ten folds of the original training set have been used. This comprehensive approach ensures accurate performance assessments and consistency across seasonal air quality variations by eliminating bias when splitting the training and testing sets. As a result, the model's robustness improves, allowing for a more reliable evaluation of its predictive capabilities.

3.5. Metric evaluation

Previous papers often used statistical measures to assess the model's accuracy when evaluating predicted AQI. Root Mean Square Error (RMSE) is a crucial predictor of predictive ability among these measurements. RMSE measures the average size of the errors between projected and observed values, explaining how well the model's predictions match the actual data (Chai & Draxler, 2014) represented by the formula in (1).

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

Conversely, Mean Absolute Error (MAE) is another important parameter for evaluating models. MAE calculates the average absolute difference between expected and observed values, resulting in a consistent assessment of prediction errors. Unlike RMSE, MAE does not square errors and treats all variances equally (Willmott & Matsuura, 2005). This feature makes MAE helpful in estimating the extent of prediction errors in the absence of outliers, with the formula to calculate as in (2). With these two metrics, the efficiency of each model can be analyzed, and from then on, any fine-tuning process would be based on these metrics to make some improvements.

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \hat{y}_i\right| \quad (2)$$

4. EXPERIMENTAL RESULTS AND ANALYSIS

4.1. Exploratory data analysis

The ten provinces' air quality index data show various patterns and seasonal fluctuations. The winter months, namely November, December, January, and February, tend to have the highest AQI readings, suggesting the influence of heating activities, weather patterns, and industrial pollutants, as shown in Table 3.

Table 3. Summary of notable AQI values in each province.

Province	Peak Month		Peak Hour		Min	Max
	Month	Mean Value	Hour	Mean Value		
Cao Bang	February	85.5	11	55.99	0	177
Lao Cai	April	42.3	7	35.65	1	98
Quang Ninh	December	58.0	20	39.69	0	153
Ha Noi	February	39.9	21	24.87	2	129
Quang Binh	February	86.8	8	66.61	1	200
Da Nang	February	26.7	9	25.57	1	90
Dak Lak	April	11.8	17	5.75	0	51
HCMC	February	75.9	7	75.22	4	183
Kien Giang	February	70.7	20	34.90	3	133
Ca Mau	February	42.0	7	37.16	4	139

HCMC had the highest AQI values in February, with a substantial mean of 75.9, suggesting severe pollution events. This result raises significant pollution problems for the winter season. In contrast, Kien Giang elevated AQI trends in February, with a high AQI of 133, but maintained a pretty excellent average AQI throughout the year. Similarly, Cao Bang has high pollution levels in February, with a peak mean value of 85.5 and a maximum of 177, showing severe pollution early in the year.

Other provinces have distinct patterns. Lao Cai has significant swings, with an elevated AQI in April, with a high mean value of 42.3 and a maximum of 98, suggesting pollution at this time. Quang Binh had the highest maximum AQI of 200 in February, indicating severe pollution incidents, with an average score of 86.8. Quang Ninh peaks in December, with a mean of 58.0 and a high of 153, suggesting heavy pollution over the winter.

Da Nang and Ca Mau have seasonal peaks in February, with Da Nang having a mean AQI of 26.7 and a high of 90, indicating moderate air quality with significant increases at the start of the year. Ca Mau also peaks in February, with a mean of 42.0 and a high of 139, suggesting increasing pollution at this season. Ha Noi shows strong AQI oscillations, with high peaks in February (mean value of 39.9 and highest of 129), indicating severe pollution. Dak Lak has the lowest AQI values of any province, with a minor rise in April and a mean score of 11.8, suggesting generally clean air.

4.2. Models' parameters setting

Table 4 displays the optimal parameters for LSTM, GRU, bidirectional versions, and hybrid CNN-based models. After extensive parameter tuning, the best setups for LSTM and GRU models are configured with units of 100 and 150, respectively, and a dropout ratio of 0.1. The optimal parameters for hybrid CNN-based models, combining CNN with LSTM and GRU, include a filter size of 16, a kernel size of 8, and a pooling size of 4, along with an LSTM or GRU unit size of 100 and a dropout ratio of 0.1.

Table 4. Parameters of the model.

Parameters	Models used
unit_1 = 100 unit_2 = 150 dropout_ratio = 0.1	LSTM, BiLSTM, GRU, BiGRU
filter_size = 16 kernel_size = 8 pooling = 4 unit = 100 dropout_ratio = 0.1	CNN + LSTM, CNN + GRU

4.3. Models' result

Before achieving the final performance metrics, the models were rerun on the pre-processed dataset using the epitome parameters determined during the initial model tuning phase, as reported above. Tables 5 and 6 summarize the results for each model throughout the provinces, including the MAE and RMSE. The bolded scores represent the models' maximum accuracy levels, demonstrating the parameter-tuning procedure's efficacy. These findings demonstrate the potential of deep learning models for forecasting AQI levels, providing valuable insights into performance disparities among provinces and serving as a reference point for understanding models' capabilities in various environmental scenarios in a clear and organized format.

Table 5. MAE metric on each model per province.

Province	LSTM	BiLSTM	GRU	BiGRU	CNN + LSTM	CNN + GRU
Cao Bang	2.924	**2.752**	2.902	2.793	6.210	5.862
Lao Cai	2.271	2.165	2.191	**2.157**	4.022	3.981
Quang Ninh	1.751	1.687	**1.661**	1.676	3.651	3.844
Ha Noi	1.362	1.337	1.336	**1.252**	3.219	3.197
Quang Binh	2.528	2.295	2.232	**2.144**	5.108	5.100
Da Nang	1.984	1.982	1.930	**1.957**	3.488	3.610
Dak Lak	0.710	0.676	0.685	**0.665**	1.280	1.316
HCMC	3.123	**2.988**	3.075	3.084	6.015	6.399
Kien Giang	2.258	2.256	**2.253**	2.261	4.725	4.418
Ca Mau	2.440	**2.420**	2.507	2.442	4.360	4.287

Table 6. RMSE metric on each model per province.

Province	LSTM	BiLSTM	GRU	BiGRU	CNN + LSTM	CNN + GRU
Cao Bang	5.057	**4.907**	5.104	4.921	9.267	8.778
Lao Cai	4.405	4.268	4.409	**4.252**	6.532	6.408
Quang Ninh	2.715	2.640	2.626	**2.620**	5.151	5.365
Ha Noi	2.161	2.157	2.187	**2.102**	4.937	4.926
Quang Binh	4.704	4.463	4.454	**4.350**	8.382	8.350
Da Nang	4.220	4.157	4.106	**4.074**	6.635	6.708
Dak Lak	1.309	1.275	1.303	**1.270**	2.114	2.154
HCMC	5.598	**5.427**	5.552	5.548	9.696	10.264
Kien Giang	4.473	4.452	**4.451**	4.455	8.291	7.232
Ca Mau	4.958	4.895	5.003	**4.893**	7.595	7.322

The MAE and RMSE metrics reveal that BiLSTM and BiGRU models outperform their regular usage, with fewer errors across most provinces, signaling the likelihood of being used on general prediction across Vietnam. Notably, locations like HCMC and Quang Binh had more significant prediction errors, indicating complicated air quality patterns. In contrast, provinces like Dak Lak and Kien Giang had lower errors, implying more stable conditions for the model to learn. To further examine the result in each fold, Figures 3 and 4 show the MAE and RMSE metrics on each fold during the training process for ten provinces using the optimal parameters, respectively.

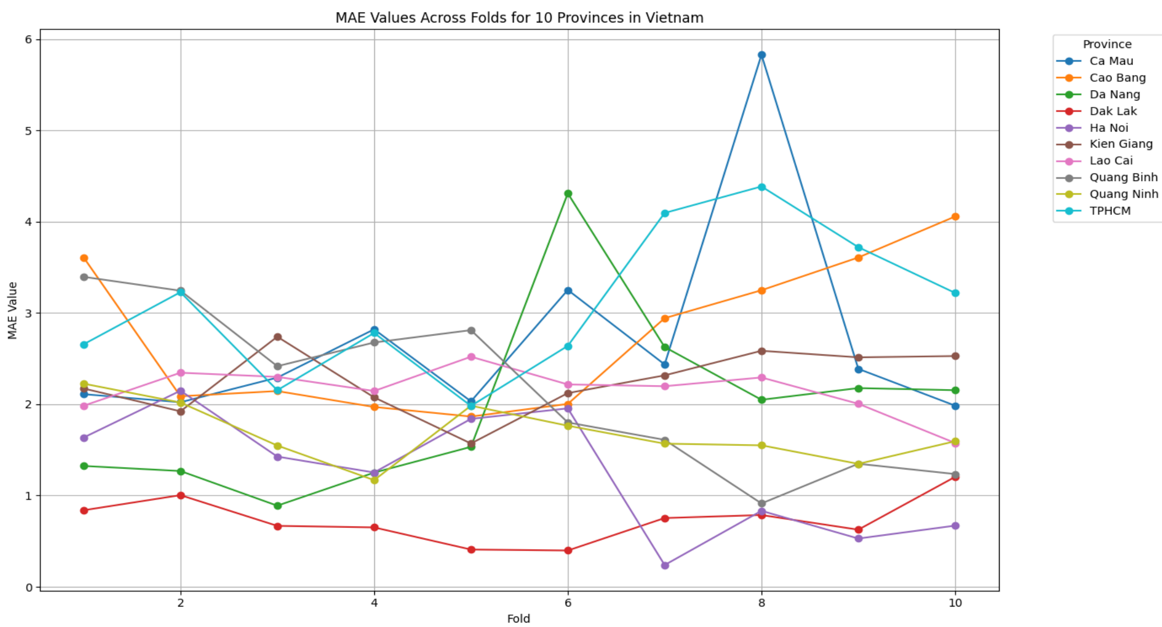

Figure 3. MAE metrics for each fold per province.

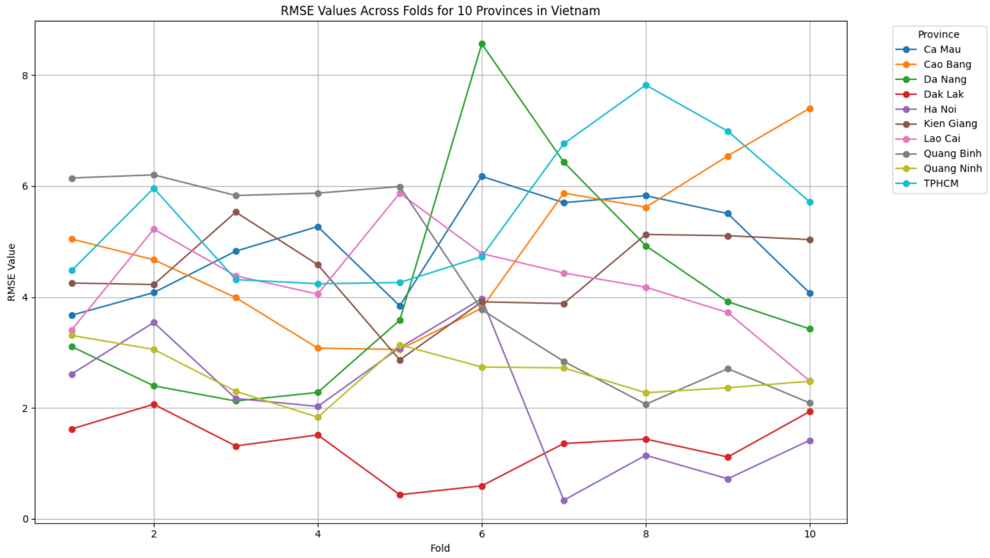

Figure 4. RMSE metrics for each fold per province.

The MAE plot confirms these findings, with Dak Lak continuously exhibiting low error values, suggesting that the models in this area are resilient and reliable. In contrast, HCMC and Cao Bang show significant oscillations and excellent error rates, emphasizing the difficulties in reliably forecasting AQI owing to potentially more complex pollution sources or unpredictability in air quality parameters. Similarly, the RMSE figure demonstrates that provinces such as Dak Lak and Ha Noi consistently maintain low error levels, indicating excellent and reliable forecasting accuracy. However, provinces like HCM and Cao Bang have higher and more variable RMSE values, especially in the last folds of the data. This result indicates that the model's performance in specific locations is less reliable and might benefit from more refinement and development to address the underlying difficulties better. In addition, Figures 5 and 6 present the predicted value versus the actual value of AQI for the best case in Dak Lak using BiGRU and the worst case of HCMC using BiLSTM over the last three days in the dataset.

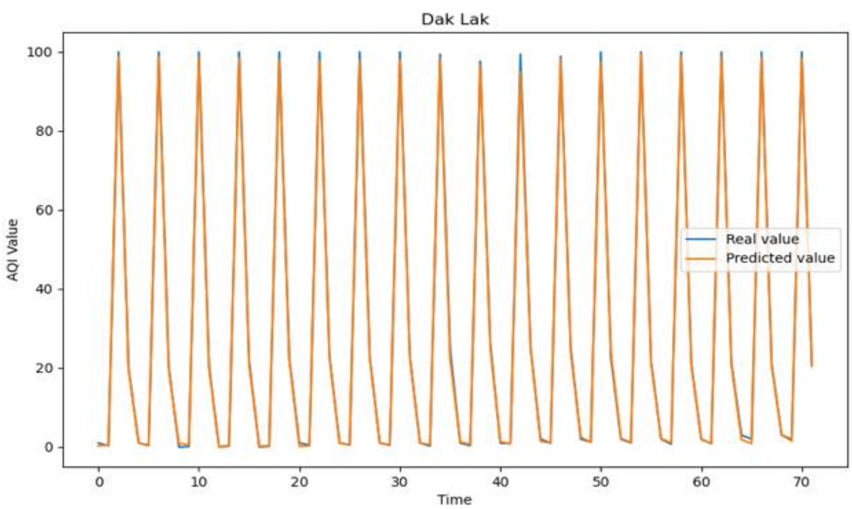

Figure 5. Predicted and actual value plot on Dak Lak's best model configuration.

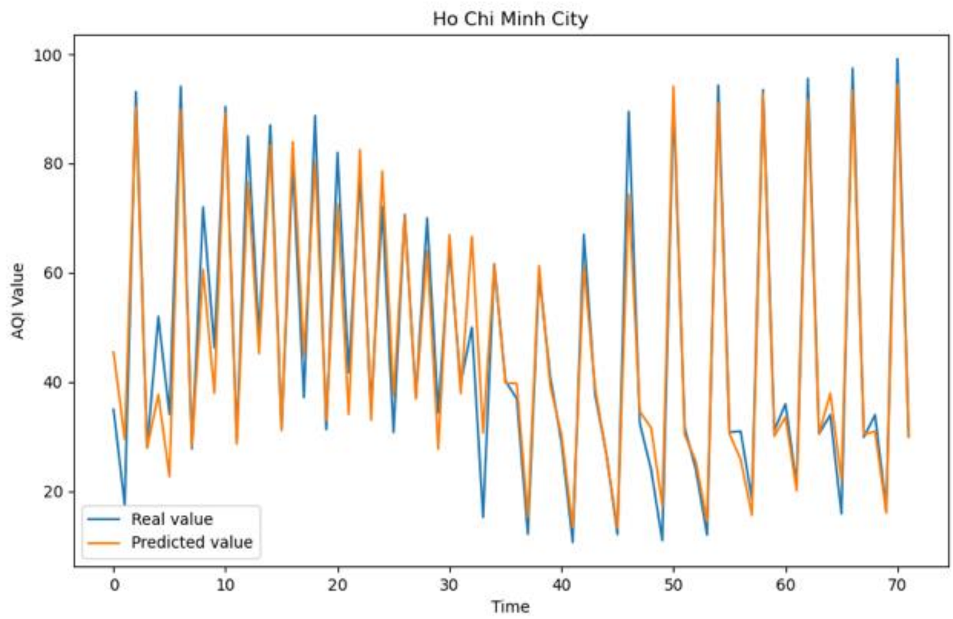

Figure 6. Predicted and actual value plot on HCMC best model configuration.

5. DISCUSSION

The setup for the LSTM and GRU-based models consists of two units of size 100 and 150, respectively, with a dropout ratio of 0.1. These parameters attempt to improve the models' ability to capture complex temporal patterns in data while reducing the danger of overfitting. The dropout ratio guarantees that 10% of neurons are discarded randomly during training, promoting model generalization and resilience.

To accommodate spatial aspects in data before temporal processing, the CNN + LSTM and CNN + GRU models include extra parameters such as filter size, kernel size, and pooling size. The filter size of 16 and kernel size of 8 are intended to successfully discover local patterns within the input data, while the pooling size of 4 aids in data dimensionality reduction, making the computational process more efficient and avoiding overfitting. The processed input is then fed to 100-unit LSTM or GRU units adjusted to capture and learn temporal relationships. The LSTM and GRU unit sizes combine model complexity and the capacity to discover long-term relationships from time-series data. The dropout ratio contributes to model performance by reducing overreliance on specific neurons, which may lead to overfitting. In CNN-based models, the combination of spatial and temporal feature extraction using suitable filter and kernel sizes, followed by pooling, guarantees that the models can manage the intricacies of air quality data. This comprehensive methodology improves the models' capacity to correctly predict air quality measurements, making them valuable tools for forecasting in the research.

Certain provinces have greater error scores than others, indicating disparities in forecast accuracy across areas. For example, Dak Lak and Ha Noi had lower RMSE and MAE values, indicating more accurate forecasts. In contrast, provinces like Ca Mau, Cao Bang, and HCMC have more significant error scores, suggesting more prediction difficulties. These patterns demonstrate the heterogeneity in model performance across regions, indicating that although the model works well in certain areas, particular provinces need focused changes to increase overall accuracy. These observations are critical for improving and calibrating the model to manage Vietnam's geographical peculiarities. They might also be due to the consistency and quality of data available for these places or to the lower fluctuation in AQI levels, making prediction more straightforward. The large error values might be due to more unpredictable AQI patterns, which may be impacted by local environmental conditions, industrial activity, or discrepancies in data collection.

The variation in error levels across the folds emphasizes the difficulties of developing a generally accurate model for AQI prediction. Some provinces' data may have more noise or outliers, affecting the model's performance. Future work should focus on enhancing data quality and consistency by including more advanced models or new features to capture local variances better. Addressing these challenges may result in more dependable and accurate AQI estimates, improving public health monitoring and environmental planning initiatives.

6. CONCLUSION

The paper proposed using several machine learning models and compared how the prediction accuracy of AQI varies between provinces in Vietnam. This heterogeneity emphasizes the difficulty of effectively estimating AQI owing to variables such as data quality, local environmental circumstances, and intrinsic model constraints. Provinces with lower mistake rates may benefit from more consistent data or less ecological variability. In comparison, those with higher error rates may need help owing to noisier data or more chaotic AQI patterns. The research also underlines the need to improve data-processing techniques using low-cost devices while maintaining the minimum set of variables. These improvements may increase model performance, particularly in regions with more significant prediction errors. Future research should create more advanced models capable of handling varied and noisy information and investigating domain-specific factors that may increase prediction accuracy. Furthermore, collaborating with local authorities to standardize the data-collecting techniques and add real-time data might improve model dependability. Future research may enhance AQI estimates by addressing these issues, resulting in improved public health monitoring and environmental management results.

ACKNOWLEDGEMENT

The authors express profound thanks to PAM Air for financing this investigation and giving access to their extensive air quality monitoring network, which considerably expanded the breadth and depth of the research. We also thank Dr Tran Thanh Tu, a lecturer at the International University's School of Chemical and Environmental Engineering, for her significant knowledge and advice. Her contributions were critical to the study's success, providing valuable insights and assistance throughout the research process. This research is also supported by the central Interdisciplinary Laboratory in Electronics and Information Technology "AI and Cooperation Robot," International University - Vietnam National University of Ho Chi Minh City.

REFERENCES

Abadi, M., Barham, P., Chen, J., Chen, Z., Davis, A., Dean, J., Devin, M., Ghemawat, S., Irving, G., Isard, M., Kudlur, M., Levenberg, J., Monga, R., Moore, S., Murray, D. G., Steiner, B., Tucker, P., Vasudevan, V., Warden, P., … Zheng, X. (2016). TensorFlow: A system for Large-Scale machine learning. *12th USENIX Symposium on Operating Systems Design and Implementation (OSDI 16)*, 265–283. https://www.usenix.org/conference/osdi16/technical-sessions/presentation/abadi

Chai, T., & Draxler, R. R. (2014). Root mean square error (RMSE) or mean absolute error (MAE)? – Arguments against avoiding RMSE in the literature. *Geoscientific Model Development*, *7*(3), 1247–1250. https://doi.org/10.5194/gmd-7-1247-2014

Doreswamy, K S, H., Km, Y., & Gad, I. (2020). Forecasting Air Pollution Particulate Matter (PM2.5) Using Machine Learning Regression Models. *Procedia Computer Science*, *171*, 2057–2066. https://doi.org/10.1016/j.procs.2020.04.221

Ghobakhloo, S., Khoshakhlagh, A. H., Mostafaii, G. R., Chuang, K.-J., Gruszecka-Kosowska, A., & Hosseinnia, P. (2023). Critical air pollutant assessments and health effects attributed to PM2.5 during and after COVID-19 lockdowns in Iran: Application of AirQ+ models. *Frontiers in Public Health*, *11*, 1120694. https://doi.org/10.3389/fpubh.2023.1120694

Méndez, M., Merayo, M. G., & Núñez, M. (2023). Machine learning algorithms to forecast air quality: A survey. *Artificial Intelligence Review*, *56*(9), 10031–10066. https://doi.org/10.1007/s10462-023-10424-4

Minh, V. T. T., Tin, T. T., & Hien, T. T. (2021). PM2.5 Forecast System by Using Machine Learning and WRF Model, A Case Study: Ho Chi Minh City, Vietnam. *Aerosol and Air Quality Research*, *21*(12), 210108. https://doi.org/10.4209/aaqr.210108

Nguyen, V.-T., Nguyen, H.-D., & Tran, M.-T. (2023). CNN-GRU for Air Quality Index Forecasting. *CEUR Workshop Proceedings*, *3583*.

PAM Air. (n.d.). Retrieved April 17, 2024, from https://maps.pamair.org/

Pant, P., Huynh, W., & Peltier, R. E. (2018). Exposure to air pollutants in Vietnam: Assessing potential risk for tourists. *Journal of Environmental Sciences*, *73*, 147–154. https://doi.org/10.1016/j.jes.2018.01.023

Pruthi, D., & Liu, Y. (2022). Low-cost nature-inspired deep learning system for PM2.5 forecast over Delhi, India. *Environment International*, *166*, 107373. https://doi.org/10.1016/j.envint.2022.107373

SGGPO. (2023, December 4). *Poor air quality recorded in capital city of Hanoi*. SGGP English Edition. https://en.sggp.org.vn/share106725.html

Suman. (2021). Air quality indices: A review of methods to interpret air quality status. *Materials Today: Proceedings*, *34*, 863–868. https://doi.org/10.1016/j.matpr.2020.07.141

Willmott, C., & Matsuura, K. (2005). Advantages of the mean absolute error (MAE) over the root mean square error (RMSE) in assessing average model performance. *Climate Research*, *30*, 79–82. https://doi.org/10.3354/cr030079

World Health Organization. (n.d.). *Ambient (outdoor) air pollution*. Retrieved March 23, 2024, from https://www.who.int/news-room/fact-sheets/detail/ambient-(outdoor)-air-quality-and-health

Zhu, J., Deng, F., Zhao, J., & Zheng, H. (2021). Attention-based parallel networks (APNet) for PM2.5 spatiotemporal prediction. *Science of The Total Environment*, *769*, 145082. https://doi.org/10.1016/j.scitotenv.2021.145082