Trends in Health Informatics

www.thi.reapress.com

THI. Vol. 1, No. 1 (2024) 63–68.

Paper Type: Case Studies

A Combined CNN and LSTM Model to Predict PM2.5 Concentration in Vietnam

Vo Phuong-Binh*

Faculty of Maths and Computer Science, Dalat University. No. 1, Phu Dong Thien Vuong St., Dalat City, Lam Dong, Vietnam; binhvp@dlu.edu.vn.

Citation:

Received: 26 May 2024 *Revised: 16 July 2024* *Accepted: 01 October 2024*	Phuong-Binh, V. (2024). A combined CNN and LSTM model to predict PM2.5 concentration in Vietnam. *Trends in health informatics, 1*(1), 63-68.

Abstract

In this paper, we propose a combined deep learning model of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to predict PM2.5 concentration in Dalat, a famous tourism city in Vietnam. CNN is utilized to extract spatial features of environmental images. The remaining component, LSTM, is designed for the training of spatial and temporal features to predict PM2.5 over a specific period. In our model, training data consists of environmental images and PM2.5 concentrations collected from lifelog on different roads in Dalat City, Vietnam. Experimental results show that the proposed combined CNN and LSTM model using both spatial and temporal features achieves superior PM2.5 concentration prediction accuracy compared to the traditional LSTM approach.

Keywords: PM2.5 prediction, Deep learning, Convolutional neural networks, Long short−term memory.

1 | Introduction

Air pollution, particularly the concentration of fine particulate matter (PM2.5 and PM10), is a matter of grave concern in the modern world, including Vietnam located in Southeast Asia. Therefore, PM2.5 and PM10 concentration prediction is important for environmental monitoring and public health management.

An association model using a deep learning method called MLP (Multilayer Perception) was proposed to study the association relationship between environmental images and PM2.5 AQI levels of lifelog data collected in Japan [1]. Lifelog or life-log is defined as digital information about personal daily life, which is a set of events (e.g., traffic jam, garbage, …etc.) collected day-by-day [2]. And AQI (Air Quality Index) [3] is an indicator of the rank of air pollution that is usually based on measurements of environmental sensing data, such as PM2.5,

Corresponding Author: binhvp@dlu.edu.vn

https://doi.org/10.22105/thi.v1i1.24

PM10, ozone (O_3), nitrogen dioxide (NO_2), sulfur dioxide (SO_2) and carbon monoxide (CO). According to the aforementioned reference, the AQI rank contains six levels from 1 to 6 for Good, Moderate, Unhealthy for Sensitive Group, Unhealthy, Very Unhealthy, and Hazardous, respectively. Recently, images related to environmental pollution (e.g., traffic congestion) have been used to predict PM2.5 concentration [4].

In this paper, we propose a combined deep learning model of Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) to predict PM2.5 concentration in Dalat, a famous tourism city in Vietnam. CNN is utilized to extract spatial features of environmental images. The remaining component, LSTM, is designed for the training of spatial and temporal features to predict PM2.5 over a specific period. In our model, training data consists of environmental images and PM2.5 concentrations collected from lifelog on different roads in Dalat City, Vietnam. Experimental results show that the proposed combined CNN and LSTM model using both spatial and temporal features achieves superior PM2.5 concentration prediction accuracy compared to the traditional LSTM approach.

2 | Methods

There is many various deep learning models used to predict time-series data in the literature, such as Random Forest (RF), Decision Tree (DT), Auto Regressive Integrated Moving Average (ARIMA), MLP, CNN, and LSTM. In this research, CNN and LSTM models are combined using spatial and temporal features to predict PM2.5 concentration in Vietnam, as illustrated in *Fig. 1*.

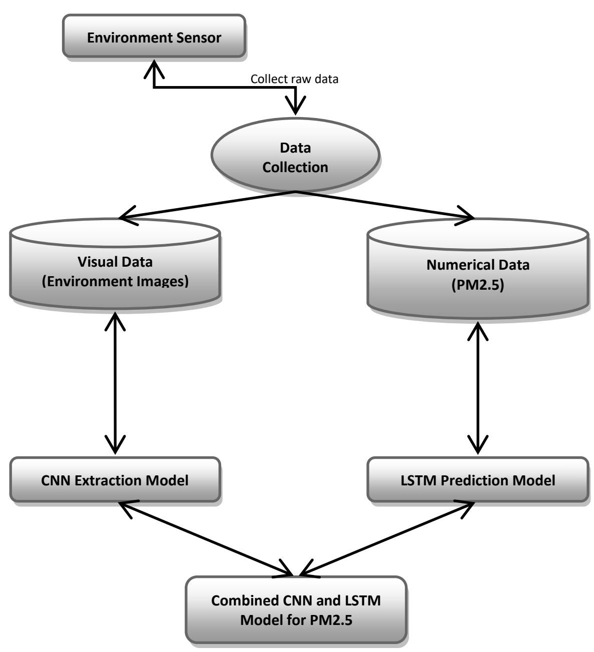


Fig. 1. Overview of proposed combined CNN-LSTM model for PM2.5 prediction.

Data collection

There are eight roads of a route in Dalat city (Vietnam) used to collect lifelog data for our experiment. The length of the route is about 6.6 km. Some different sensors are used to record both environmental images and numerical data. The AQI factor of numerical data, i.e., PM2.5, and images are recorded along the route in the meantime. We invested 33 days for data acquisition with both visual and numerical data from Oct. 15 to Nov. 17, 2019. In addition, we also used 43.824 collected PM2.5 public data days from 2010 to 2014 years for the single LSTM training stage.

CNN extraction model

The lifelog images were processed to extract spatial features, while PM2.5 data was included as an additional input feature to study the association relationship between environmental visual features and AQI levels. CNN is an efficient deep learning method to detect and extract features of visual big data (i.e., images)[5]. Recently, there have been some CNN-based recognition algorithms. Deep Residual Networks (ResNet) [6] was proposed by Microsoft Research as the best model for ImageNet Large Scale Visual Recognition Challenge (ILSVRC) [7] in 2015, and it became possible to construct deep neural networks of up to 1000 CNN layers or more. Each layer is considered to extract more sophisticated and complex features. Place-CNN[8] fine-tuned ResNet for scene classification. Attention-CNN [9] also adopted the ResNet as a CNN seed for image captioning and visual question answering. In our experiment, Attention-CNN is the best feature extractor of the above-mentioned methods that is utilized to extract feature vectors of our lifelog visual data. Then, the combined one-dimensional vectors of extracted features are used as the input data for the proposed CNN-LSTM model training to predict PM2.5 concentrations.

Traditional LSTM prediction model

The traditional LSTM model[10] was trained to capture temporal dependencies in the PM2.5 data. The LSTM model was designed to enable effective prediction of PM2.5 concentrations over a specific period. In our experimental data, PM2.5 values of each route recorded within the same period are strongly different for various days. The PM2.5 concentrations of each route can be fairly predicted using the traditional LSTM model with untrained data of the remaining days.

Proposed combined CNN and LSTM model

The proposed deep learning model combined the CNN Extraction model in *Step 2* and the traditional LSTM model in *Step 3* above to recognize some abnormal changes of PM2.5 data (e.g., traffic congestion, garbage, …etc.). Different from the traditional pure LSTM architecture, the first stage of the proposed combined CNN-LSTM uses CNN-Extractor, and the latter stage is LSTM forecasting, which is used to analyze the features extracted by CNN-Extractor and then to estimate the PM2.5 concentration. The training process involved optimizing the model's parameters to minimize the prediction error. The training was performed until the model reached a satisfactory level of accuracy.

In our combined CNN-LSTM model, the loss function is "categorical_crossentropy" and the optimizer is SGD (Stochastic Gradient Descent) with a learning rate of 0.01.

3 | Results

The results of this research are as follows:

CNN extraction model

In our experiment, a feature extractor called CNN-Extractor[9], which adopted the ResNet as a CNN seed for image captioning and visual question answering, is the best feature extractor of the aforementioned methods that is utilized to extract feature vectors of environmental images. The output of the last fully connected layer of the CNN-Extractor is collected as the input feature vector for the combined CNN-LSTM model training stage.

Proposed combined CNN and LSTM model

The training data of Road 1 are selected randomly with five cases of training data ratios containing 65%, 70%, 75%, 80%, and 85%. The results of data training using the proposed CNN-LSTM model are achieved in terms of mean squared error (MSE) of 5.416, 5.336, 4.927, 4.826, and 4.683, respectively. The decrease in the MSE of the evaluation stage with more trained data shows that the proposed model is probably stable and robust enough to cope with a wide range of different visual data. MSE is formulated as $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^{n}(o_i - p_i)^2$.

where o_i is the ith original value, p_i is the corresponding predicted value for o_i, and n is the number of observations. The Σ indicates that a summation is performed over all values of i.

Table 1 lists experimental results with the same training data ratio of 85% for both traditional LSTM and combined CNN-LSTM models and shows clear evidence of better forecast accuracy of the proposed combined method compared to the original LSTM model. MSE for lifelog data is much smaller with the proposed combined method than with the traditional LSTM.

Table 1. Experimental results.

Roads	Methods	MSE (Mean Squared Error)
Road 1	Traditional LSTM	6.177
	Combined CNN-LSTM	4.683
Road 2	Traditional LSTM	6.129
	Combined CNN-LSTM	4.485
Road 3	Traditional LSTM	6.111
	Combined CNN-LSTM	4.723
Road 4	Traditional LSTM	6.169
	Combined CNN-LSTM	4.776
Road 5	Traditional LSTM	6.127
	Combined CNN-LSTM	3.660
Road 6	Traditional LSTM	6.100
	Combined CNN-LSTM	4.631
Road 7	Traditional LSTM	6.094
	Combined CNN-LSTM	3.599
Road 8	Traditional LSTM	6.079
	Combined CNN-LSTM	4.691

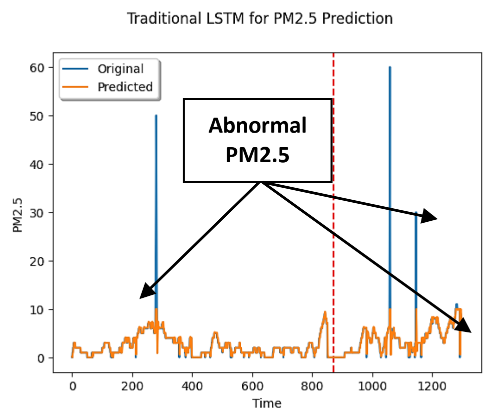
As we know, the LSTM model outperformed traditional statistical models commonly used for time-series prediction. However, the results demonstrated in *Fig. 2a* show that the traditional LSTM model captured only temporal information and was not accurate in some abnormal changes in PM2.5 concentration (e.g., traffic congestion, garbage, …etc.).

As illustrated in *Fig. 2b*, the proposed combined CNN-LSTM model which used both spatial environmental images and temporal PM2.5 information achieved superior PM2.5 concentration prediction accuracy compared to the traditional LSTM approach. In the proposed model, the CNN component effectively extracted spatial features from environmental images, capturing the spatial features associated with PM2.5 concentrations. Then, the LSTM component captured the temporal dependencies and extracted spatial features in the data, allowing accurate predictions over time, especially for abnormal PM2.5 cases. Finally, the experimental results show the effectiveness of the proposed combined CNN-LSTM approach for PM2.5 prediction in Vietnam.

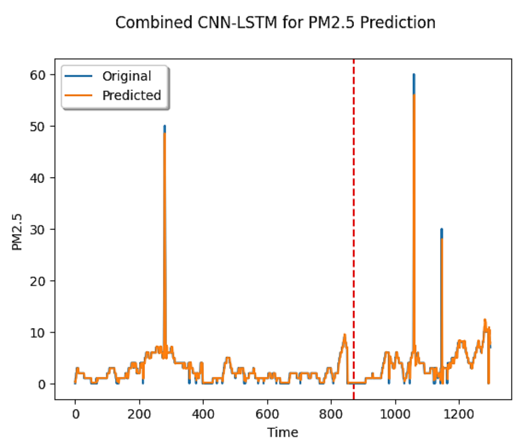
4 | Discussion and Conclusion

In this paper, we proposed a combined CNN-LSTM deep learning model to predict PM2.5 concentration in Vietnam. Experimental results shown that the proposed model using both spatial and temporal features achieved superior PM2.5 concentration prediction accuracy compared to the traditional LSTM model, which has been commonly used for time-series prediction. The proposed combined CNN-LSTM model accurately extracted spatial features related to PM2.5 from environmental images and temporal information from PM2.5 concentrations in the lifelog data, resulting in robustly accurate predictions.

As a result, this research contributes to the prediction of factors causing environmental pollution in the field of public health, offering an overview of air quality control not only in Vietnam but also in other regions (e.g., Asia-Pacific) facing similar environmental pollutants. The proposed accurate PM2.5 prediction model, which employs a combined CNN-LSTM model, assists government leaders in making decisions and strategies for effective air pollution management. In the future, ongoing research and application of this approach may result in better air quality control and promotion of public healthcare in Vietnam and the Asia-Pacific region.



a.



b.

Fig. 2. Experimental results of PM2.5 prediction models; a. traditional LSTM model, b. Combined CNN-LSTM model.

Author Contributions

Vo Phuong-Binh was responsible for the conceptualization of the research, model design, data collection, and analysis. Vo also drafted the manuscript and reviewed the final version for publication.

Funding

The author received financial support from Dalat University (DLU), Vietnam for funding this study, under the DLU Science Research Project 1104/QĐ-ĐHĐL.

Data Availability

The datasets used in this study are available from the corresponding author upon reasonable request.

Declaration of Conflicting Interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

References

[1] Vo, P. B., Phan, T. D., Dao, M. S., & Zettsu, K. (2019). Association model between visual feature and AQI rank using lifelog data. *2019 IEEE international conference on big data (Big Data)*, 4197–4200. IEEE. https://doi.org/10.1109/BigData47090.2019.9005636

[2] Doherty, A. R., & Smeaton, A. F. (2010). Automatically augmenting lifelog events using pervasively generated content from millions of people. *Sensors (Basel)*, *10*(3), 1423–1446. https://doi.org/10.3390/s100301423

[3] AirNow. (2024). *Air quality index (AQI): Basics about the AQI*. https://www.airnow.gov/aqi/aqi-basics

[4] Fang, X., Li, Z., Yuan, B., & Chen, Y. (2024). Image-based PM2.5 estimation from imbalanced data distribution using prior-enhanced neural networks. *IEEE sensors journal*, *24*(5), 4677–4693. https://doi.org/10.1109/JSEN.2023.3343080

[5] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems 25(2).* https://doi.org/10.1145/3065386

[6] He, K., Zhang, X., Ren, S., & Sun, J. (2015). *Deep residual learning for image recognition*. https://arxiv.org/abs/1512.03385

[7] Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., ... & Fei-Fei, L. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, *115*(3), 211-252. https://doi.org/10.1007/s11263-015-0816-y

[8] Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., & Torralba, A. (2017). Places: A 10 million image database for scene recognition. *IEEE transactions on pattern analysis and machine intelligence*, *40*(6), 1452-1464. https://doi.org/10.1109/TPAMI.2017.2723009

[9] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., & Zhang, L. (2018). Bottom-up and top-down attention for image captioning and visual question answering. *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 6077-6086). https://openaccess.thecvf.com/content_cvpr_2018/html/Anderson_Bottom-Up_and_Top-Down_CVPR_2018_paper.html

[10] Khedhiri, S. (2023). A deep learning LSTM approach to predict COVD-19 deaths in North Africa. *Asia pacific journal of public health*, *35*(1), 53-55. https://doi.org/10.1177/10105395221141590