Research article

Visually Estimating and Forecasting PM2.5 Levels Using Hybrid Architecture Deep Neural Network

Sirisup Laohakiat, Songpon Klerkkidakan and Nuwee Wiwatwattana

Department of Computer Science, Faculty of Science, Srinakharinwirot University, Bangkok, Thailand

Curr. Appl. Sci. Technol. 2024, Vol. 24 (No. 3), e0258074; https://doi.org/10.55003/cast.2023.258074

Received: 21 March 2023, Revised: 2 May 2023, Accepted: 28 September 2023, Published: 14 December 2023

Keywords

particulate matter;

air pollution;

long short-term memory;

convolutional neural network;

PM2.5

Abstract

Air pollutants, particularly particulate matter with a diameter of less than 2.5 microns (PM2.5), have been linked to a number of negative climatic and health effects, including irritation and damage to internal organs in the short and long term. Knowing current and future levels of PM2.5 is therefore crucial for both public health and environmental management. Photos of the sky and buildings can aid in assessing visibility and air quality because PM2.5 alters the appearance of the sky and its surroundings. This research proposes a method that employs digital images to not only estimate but also forecast PM2.5 concentration levels using deep learning technologies. A convolutional neural network (CNN) was used as the base layer to automatically extract image features, while a long short-term memory (LSTM) network was used as the output layer to analyze the sequence of features. The sky and its surroundings were captured hourly from a high-rise building in Bangkok, Thailand, and the ground-truth PM2.5 data were obtained from the nearest monitoring station facing the same direction. The experimental results show that the hybrid CNN and LSTM model outperformed conventional CNN models in estimating and forecasting PM2.5 levels by 5% and 30%, respectively.

1. Introduction

Air pollutants, particularly PM2.5, have been demonstrated to be detrimental to human health, causing short- and long-term irritation and damage to internal organs such as the heart and lungs, as well as being linked to neurological disorders [1-3]. Health consequences can be fatal, while environmental consequences can both sustain and contribute to climate change [4, 5]. These negative

*Corresponding author: E-mail: nuwee@g.swu.ac.th

consequences have sparked particular concern on a global, regional, and local scale. As a result, it is crucial to take public action and raise public awareness because the PM2.5 problem is worsening by the day in many rapidly expanding cities. PM2.5 concentrations must be measured for people to take the necessary precautions and make preparations. When PM2.5 levels are high, it is advisable to stay indoors to reduce exposure. However, because monitoring stations that regularly report PM2.5 levels are expensive to construct, they are few and far between [6].

For decades, considerable research has been undertaken on developing scientific models to estimate and forecast PM2.5 concentration levels using data from a variety of sources, including air-quality monitoring stations, meteorological stations, and satellite data. Chemical transport-based atmospheric dispersion models, which use equations to simulate the movement and diffusion of particulate matter from emission sources, have long been used by government agencies around the world [7]. In contrast, machine learning models trained on meteorological conditions and past observation data can provide high-accuracy real-time predictions [8, 9]. Machine learning and deep learning can now effectively discover complex patterns and relationships in order to support precise predictions because of increased data availability and computing power.

PM2.5 has long been recognized as one of the causes of smog, particularly in urban areas [10]. Because particles scatter and absorb light, a high concentration of particulate matter in the air can limit visibility. During the most recent COVID-19 lockdown episode, lower PM2.5 levels were reported as visibility increased in Wuhan, China [11]. Photos of the sky and buildings can aid in assessing visibility and air quality because PM2.5 alters the appearance of the sky and its surroundings. When the sky is clear and blue and the buildings are clearly visible, this generally indicates good visibility and air quality. If the sky appears hazy or orange and the buildings appear dusty, this may indicate the presence of pollutants or particles in the air. Figures 1 and 2 highlight our point by comparing images taken at the same time on different days with varying levels of PM2.5. Because IP cameras are omnipresent in metropolitan areas and regularly record digital images of the sky and nearby structures, the cost of digital image acquisition is low. As a matter of fact, digital photographs may be a low-cost solution to represent PM2.5 concentration levels. To effectively use these digital images, efficient image analysis and processing algorithms to interpret sky photos are required.

Image-based machine learning algorithms, which employ satellite or ground-based imagery to estimate particle levels, have shown good potential despite not being as accurate as monitoring-based models. Image-based machine learning may be used to identify patterns and connections in data that humans cannot see, and it can be trained to make predictions based on these patterns. It is important to note, however, that these models still need to be validated or verified against ground measurement data in order to ensure their correctness and trustworthiness. Prior machine learning work has taken two approaches: traditional computer vision algorithms that require explicit feature extraction and end-to-end deep learning algorithms, in which neural networks are used to automatically extract features.

In the traditional approach, a number of key features representing visual cues for PM2.5 are extracted from images. Liu *et al*. [12] used a variety of image-processing algorithms to determine the transmission index, image contrast, image entropy, sky smoothness, sky color, and sun angle of images of city photos. Support vector regressors were trained on these features, and the results confirmed that the approach could be used to predict PM2.5 based on the images. He *et al*. [13] analyzed six features of brightness and colorfulness from images taken with a camera and a smartphone, and ran those features through a recurrent fuzzy neural network. Using artificial neural networks, Chen *et al*. [14] and Wang *et al*. [15] developed models that were trained on spatial and transform-based entropy features derived from a saturation map extracted in hue-saturation-value color space.

Figure 1. At 2 p.m. on December 16, 2019, the PM2.5 level was 14 µg/m^3.

Figure 2. At 2 p.m. on December 11, 2019, the PM2.5 level was 49 µg/m^3.

Yao *et al*. [16] presented techniques to analyze a number of smartphone images from random places. Atmospheric measures such as local transmission index and local extinction index were paired with weak cues such as sky discoloration, luminance gradients, and structural information loss. Support vector regression and random forests were also used in addition to neural networks. These investigations pointed out that the retrieved features must be based on certain hypotheses, such as those relating to colors and light transmission. However, due to regional and temporal changes, they might not always be applicable.

The end-to-end deep learning approach has gained popularity over the past few years. Deep learning models, notably convolutional neural networks (CNNs), have demonstrated impressive performance in image analysis [17]. Furthermore, CNNs have a proven track record of success in weather and climate applications [18]. The strengths of CNNs lie in their automated learning of hierarchical data representations, which capture both local and global patterns in visual data. Spatial variation in the appearance of the sky across different regions of a photograph may reflect the position of the sun and objects, the intensity of sunlight, weather conditions, atmospheric phenomena, and of course, the presence of air pollution.

It is not a surprise that convolutional neural networks and their variants are the best-suited architecture for this type of image-based learning. Zhang *et al*. [19, 20] were among the first to use this approach. The authors specifically designed ensemble CNN classifiers and an activation function that were optimized for air quality estimation work. However, they did not try to predict PM2.5 levels in numbers but instead as ordinal classes. Jiang *et al*. [21] devised a spatiotemporal contrastive learning method to pre-train CNN from unlabeled satellite imagery, then used the CNN in supervised learning to achieve lower spatial errors as the number of stations increased. According

to Zaini *et al*. [22], hybrid deep learning could predict air quality more precisely than individual network models, and the trend is moving in that direction. Chen *et al*. [23] tried to generate propagation heat maps of PM2.5 based on a convolutional recursive neural network that used recursive layers as output layers, similar to ours.

Long short-term memory (LSTM) and its hybrid versions are frequently used in time series predictions due to their ability to interpret sequential data, resulting in their widespread applications in PM2.5 concentration predictions, for example in Jiang *et al*. [24] and Shi *et al*. [25], to name a few. LSTM nodes are utilized to capture relationships between image pixels because of their capacity to do so. By using sequential models such as the LSTM together with the CNN model, we anticipated that the hybrid model would be able to learn more aspects of the relationship among pixels than either model alone. Song *et al*. [26] took a step further by integrating Residual Network (ResNet) with LSTM, extracting spatial-temporal features of sequential daytime and nighttime images taken from smartphones, as opposed to ours, which uses a single image as a predictor. In Won *et al*. [27], CCTV image data was fed into a modified VGG-16 model, while weather data was fed into an LSTM model, and both were combined to predict PM2.5 and PM10, yielding an amazing result.

In light of this, we proposed a hybrid deep neural network model that integrates CNN and LSTM networks into a single model to investigate the effectiveness of using IP camera images to estimate and forecast PM2.5 concentration levels. The regression model can be used to estimate air quality in areas where monitoring stations do not exist or are not dense enough to offer a detailed and immediate assessment of the quality of the air. To keep the system simple in terms of both memory and computation requirements so that we could deploy the model on a simple edge device, we aimed at creating a model that used only a single image, not a sequence of images, to estimate PM2.5 levels. The experimental area was located in one of Bangkok's most congested areas. Furthermore, the hybrid model's performance was compared to that of the conventional CNN model.

2. Materials and Methods

The materials and processes used to estimate and forecast PM2.5 concentration levels from pictures are described in this Section. In comparison to a convolutional neural network, the hybrid convolutional neural network and long short-term memory neural network, and also the evaluation metrics utilized in the experiments, are described.

2.1 Study area and data acquisition

The study location was in the heart of Bangkok, Thailand's capital, and commercial center. This study made use of two types of data: PM2.5 observational data (as PM2.5 labels) and skyline photographs. From August 2019 to December 2019, the Pollution Control Department of Thailand's Ministry of Natural Resources and Environment published hourly observation data on PM2.5 concentrations. The PM2.5 monitoring station was at Chokchai Police Station and was a permanent-type roadside air quality station with a 3-m high PM2.5 collection point. The IP camera was mounted in the direction of the monitoring station from the 18th level of a 19-story building, as shown in Figure 3. The forward fill method was used to fill in the missing PM2.5 values from the station. Images were taken on the hour to match the hourly PM2.5 measurements. As this was a regression task, we used the images as the input of the model to predict the current and future values of PM2.5 which were the target values of the model. Therefore, the current and future values of PM2.5 were

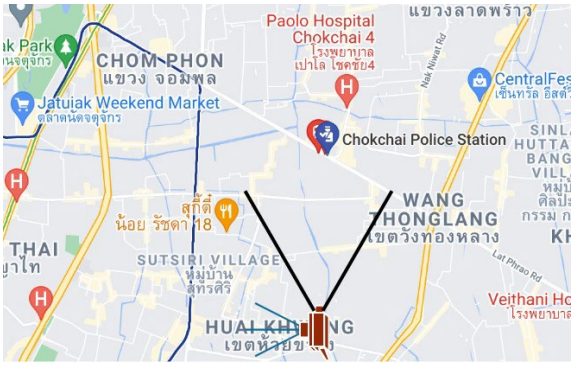

Figure 3. The angle of the camera and the location of the PM2.5 monitoring station

treated as the labels of the corresponding images. However, missing photos, maybe owing to a power outage, could not be restored, and the necessary entries were lost. We only used the data between the hours of 7 a.m. and 6 p.m. because the photographs were too dark otherwise. A total of 1690 hours of observation and photography were collected.

Figure 4 shows the trend of PM2.5 observation data from the monitoring station from August 2, 2019, at 12 p.m., to December 30, 2019, at 6 p.m. During both the rainy and winter seasons, the five-month data set included PM2.5 concentration values ranging from 2 to 99 μg/m^3. This was beneficial to our work because we were able to instruct the models about good and bad air quality instances. The average PM2.5 concentration was 24.04 μg/m^3, with a standard deviation of 14.96 μg/m^3. Figure 5 depicts one-day hourly IP camera photographs, as well as PM2.5 concentration values from the monitoring station on September 27, 2019. Air pollution appeared to turn the sky orange but determining the exact PM2.5 levels using the naked eyes was challenging.

2.2 Convolutional neural network

Convolutional neural networks (CNNs) have been shown to be highly effective in image processing [17]. The goal of CNN is to capture image features through a series of convolutional and pooling operations. The convolution operation can be essentially seen as a low-pass filter process in which a filter with coefficients learned from the data is applied to the image. By applying the filter to the image, the output from the convolution operation, which we call the feature map, can represent some characteristics of the original image.

Although the feature map from the convolution layer can be representative of the original image, in many cases it still contains too much information. As a result, CNN usually has a pooling layer next to the convolution layer to down-sample the resolution of the feature map so that we can get a more compact representation of the original image.

Figure 6 illustrates the operations of a convolution layer followed by a pooling layer. In the convolution layer, the input image is convolved with the filters, resulting in a filtered feature map. Then, in the pooling layer, the resolution of the feature map is downsampled to remove the redundant information. Note that in this illustration, we show the max-pooling method. Practically, we can add several sets of a convolution layer and a pooling layer to form the deep network operation.

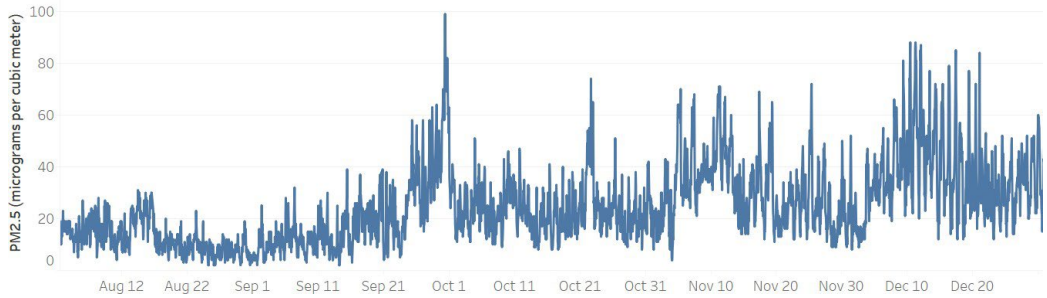

Figure 4. PM2.5 concentration levels data from the monitoring station from August 2, 2019, to December 30, 2019

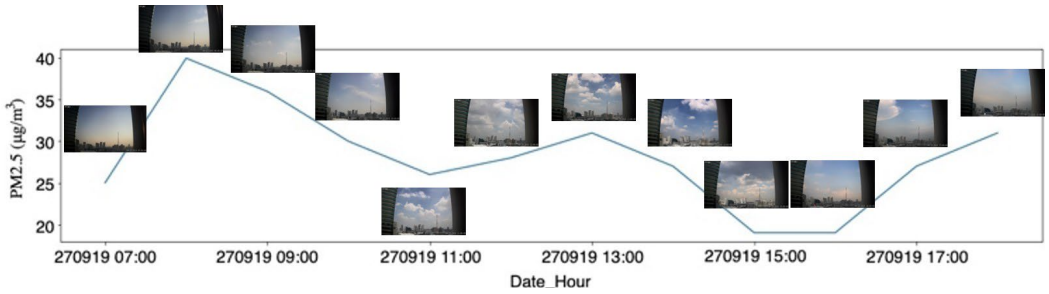

Figure 5. One-day hourly photographs from the IP camera along with PM2.5 readings from the monitoring station on September 27, 2019

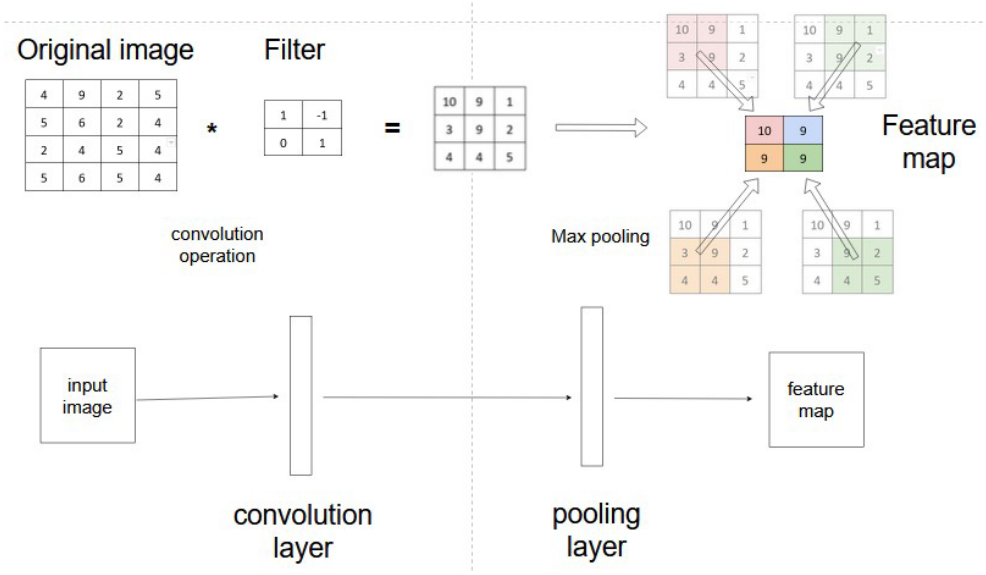

Figure 6. Convolution and pooling processes in a convolutional neural network (CNN)

2.3 Long short-term memory neural network

To analyze sequential data that contains ordinal relationships across samples, we require a particular neural network called recurrent neural networks (RNN). RNN uses hidden states to store the previous data, allowing the network to produce output based on both the current input and previous information. However, RNN cannot effectively deal with long-term dependencies [28], a situation which is the so-called vanishing gradient problem.

To solve this problem, Hochreiter and Schmidhuber [29] proposed the long short-term memory (LSTM) unit, which gave rise to the name of the LSTM neural network, which employs this unit as one of its main components. The LSTM unit generally consists of four basic components, i.e., a memory cell, an input gate, an output gate, and a forget gate, as shown in the diagram in Figure 7. Let X_t, Y_t, and C_t represent the input, output, and state of an LSTM cell at time t, respectively.

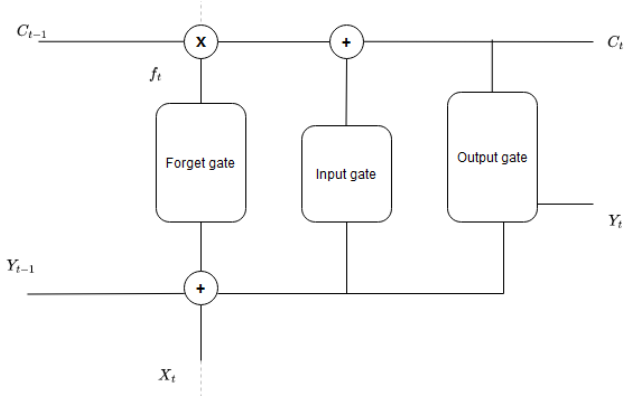

Figure 7. A simplified diagram of a long short-term memory neural network (LSTM) unit

Through the feedback path, the unit receives current input X_t, previous state information C_{t-1}, and previous output Y_{t-1}. The forget gate is used to control the impact of the previous state, C_{t-1}, which is determined by Y_{t-1} and X_t. The output from the forget gate, f_t, is used to modulate the value of C_{t-1}. The value of f_t is obtained from a sigmoid function $f_t = \sigma(Y_{t-1}, X_t)$, and as a result, the value of f_t is between 0 and 1. When f_t is small, the value of C_{t-1} is attenuated after the multiplication. On the contrary, when f_t is large, C_{t-1} will not be attenuated much. As a result, we can see that f_t behaves as a gate controller. When we want to forget the previous state, we attenuate the value of C_{t-1} through a small f_t.

After the forget gate modifies the value of C_{t-1}, we generate the current state C_t from C_{t-1}, Y_{t-1} and X_t. The input gate generates a signal by considering the value of Y_{t-1} and X_t. Then, this value is used to add with the modified C_{t-1} to generate the current state C_t. Finally, the output gate generates the current output Y_t, from C_t, Y_{t-1}, and X_t.

Due to the feedback mechanism of the state signal, the previous state can be used as a representative of the previous information of the unit. In this study, we used LSTM to model the changes in the vertical sky profile to improve the concentration predictions.

2.4 Hybrid convolutional neural network and long short-term memory neural network

Long short-term memory (LSTM) has been used for sequence input/output prediction, but it can also be used to learn spatial feature representation for image learning, as seen in Zhao *et al.* [30]. The

diagram of the proposed model is shown in Figure 8. The input image was fed to CNN to generate the feature map. The feature map was then divided vertically into n sections; in this study, going through a hyperparameter search, $n = 9$ was used. The array of the vertical strips of the feature map was used as a sequential dataset supplied to the LSTM model. By doing so, we enabled the LSTM model to sequentially analyze the vertical profile of the sky. Employing both CNN and LSTM together to capture the features of the image allowed the hybrid model to learn the representation more effectively than a single conventional model. Note that the usage of LSTM here differs from that of the conventional approach in which LSTM is used to process a sequence of data. Here, we use LSTM to process a sequence of vertical strips of a single image. The aim of adding an LSTM layer to the model is to enhance its capability to capture the features of the image.

The architecture of the neural network model is shown in Figure 9. The input image, downsampled from the original image from the camera, was of the size HxWxC = [256x256x3], where H, W, and C are height, width, and the number of channels of the image, respectively. Each CNN unit made use of 3x3 kernel. We used three CNN layers to capture the feature map from the input image. Then, the feature map was sent to a single-layer LSTM. All the activation functions were rectified linear units (ReLU). An Adam optimizer with a learning rate of 0.0001 was used to train the hybrid model.

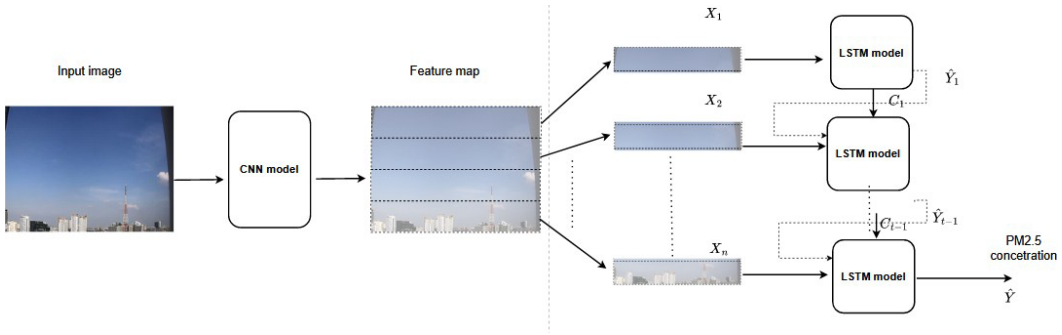

Figure 8. Illustrative diagram of the proposed hybrid CNN-LSTM model

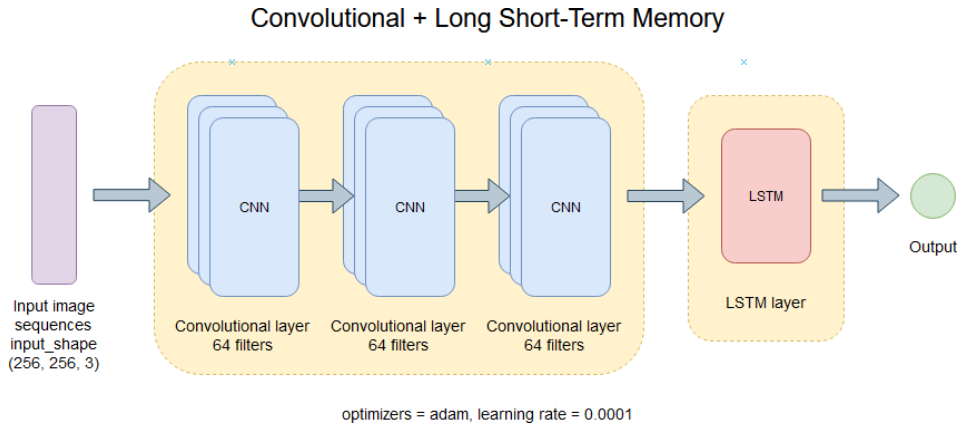

Figure 9. The architecture of the proposed hybrid CNN-LSTM neural network model

2.5 Evaluation metrics

The following measures are performance metrics used to evaluate our methods, where n is the total number of samples.

2.5.1 Root mean square error (RMSE)

In equation (1), RMSE is the square root of the mean squared error (MSE), which is the average squared difference between the predicted PM2.5 values ($\hat{Y}_i$) and the true (observed) PM2.5 values (Y_i). RMSE is excellent at measuring the magnitude of prediction errors because it penalizes large errors.

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}\left(\hat{Y}_i - Y_i\right)^2} \tag{1}$$

2.5.2 Mean absolute error (MAE)

In equation (2), MAE is the average absolute difference between the predicted PM2.5 values ($\hat{Y}_i$) and the true PM2.5 values (Y_i).

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|\hat{Y}_i - Y_i\right| \tag{2}$$

2.5.3 Mean absolute percentage error (MAPE)

In equation (3), MAPE is the average absolute percentage error between the predicted PM2.5 values ($\hat{Y}_i$) and the true (observed) PM2.5 values (Y_i). MAPE is scale independent since it normalizes each error by its true value.

$$MAPE = \frac{100}{n}\sum_{i=1}^{n}\frac{\left|\hat{Y}_i - Y_i\right|}{Y_i} \tag{3}$$

While MAPE is expressed as a percentage, the units of RMSE and MAE are the same as the values we predict or forecast, which in this case are micrograms per cubic meter (μg/m^3). We are now ready to discuss the experimental results.

3. Results and Discussion

We chose four months of data between August 2, 2019, at 12 p.m., and December 1, 2019, at 10 a.m. for training, corresponding to 1352 images, and the remaining month for testing, corresponding to 338 images. As can be seen from Figure 4, the range of PM2.5 concentrations was the same in training and testing data sets; both were positively skewed. However, the test data had higher PM2.5 levels on average. Each model was trained to predict the current day and the next day's PM2.5 levels. The target results of the models were referred to as estimated current values (t+0) and forecast values for the next 24 h or 24-h forecast (t+24). PM2.5 levels are normally expected for the next

several hours ahead for public preventive measures. Forecasting too far into the future may not create enough urgency for public attention, and the outcomes are generally less accurate. The concept of forecasting next-day PM2.5 levels arises from the fact that the PM2.5 cycle normally repeats every day. The levels of PM2.5 in the mornings of today and tomorrow are highly correlated. For example, the correlation coefficient between t+0 and t+24 PM2.5 values in our data was approximately 0.65.

Models were tuned using the training data, and predictions were produced using only the test data. Table 1 shows the results of the hybrid model (CNN+LSTM) compared with those of the CNN model alone. In this study, we evaluated the performance of the models in two respects. First, we assessed the performance of the models in estimating the current level of PM2.5 from the captured image. The results are shown in the columns "Estimated Current (t+0)" in Table 1. We found that the proposed hybrid model outperformed the conventional CNN model by almost 5% in both RMSE and MAE. However, in the case of MAPE, both models yielded similar results. This could result from the fact that the error of the CNN occurred when the actual concentration was relatively large. As MAPE was obtained from the error divided by its actual value, when we divided the error by the large actual value, the result became smaller.

In the second experiment, in which we examined model forecasting performance, the hybrid model clearly outperformed the conventional CNN model. Based on the current image, the models forecast the PM2.5 concentration for the next 24 h. The results are displayed in the columns "24-hour Forecast (t+24)" in Table 1. The hybrid model yielded better results in all three evaluation metrics. In terms of RMSE and MAE, the hybrid model significantly outperformed the CNN model by nearly 30%. Furthermore, the MAPE of the hybrid model was 8% lower than that of the CNN model. The big improvement over the CNN model for the forecasting experiment justified the fact that the LSTM in the hybrid CNN+LSTM model had taken advantage of the vertical sky profile sequence of the photographs.

Forecasting 24 h ahead was less accurate than estimating the present concentration from the image, as expected, but unlike the CNN standalone model, the hybrid model's forecasting performance was not far off from the estimate performance (less than 2 $\mu g/m^3$ in both RMSE and MAE). Previous research done in the Bangkok area using machine learning models to fit MODIS satellite data and other meteorological data [31] achieved RMSE values of 6 to 13 $\mu g/m^3$ depending on the season, with a larger number in the winter when PM2.5 levels were high. Our proposed models, also tested on winter data, produced comparable results while eliminating the collection of costly data.

To visually compare the prediction performance, allowing us to inspect insights from the individual predictions, Figures 10 and 11 depict the scatter plots between the true values (X-axis) and the predicted (forecast) values (Y-axis), as well as the percentages of underestimation, exact estimation, and overestimation. The straight line represents the slope 1 line, with points above the line indicating overestimations (predictions higher than actual values), points below the line indicating underestimations, and points on the line indicating accurate estimations.

Table 1. Performance of the CNN and the CNN+LSTM models on the test data set

Metrics	Estimated Current (t+0)		24-hour Forecast (t+24)	
	CNN	CNN+LSTM	CNN	CNN+LSTM
RMSE ($\mu g/m^3$)	11.13	10.53	16.42	12.39
MAE ($\mu g/m^3$)	8.08	7.57	13.26	9.10
MAPE (%)	24.60	24.39	35.57	27.38

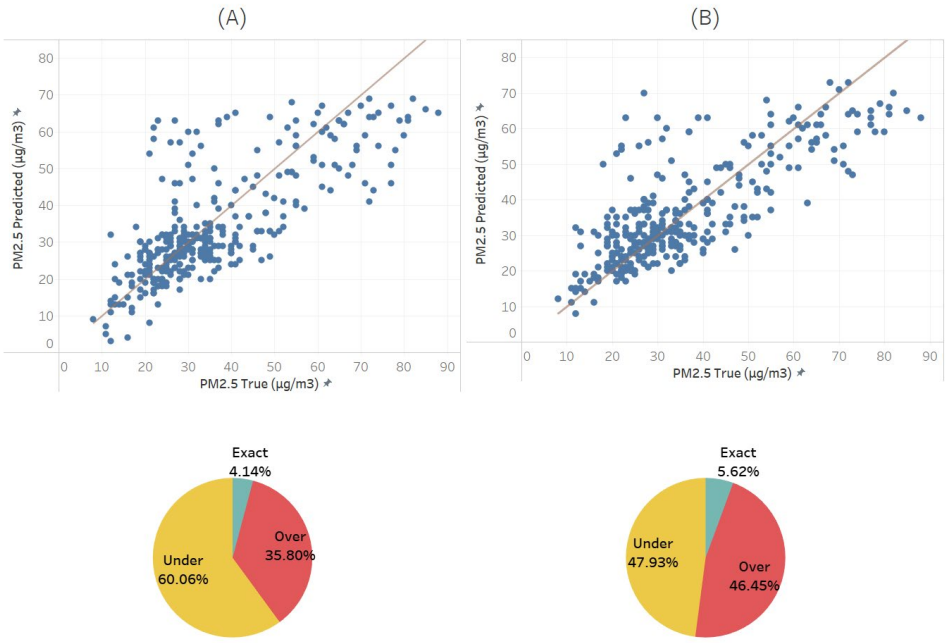

Figure 10. The scatter plot of true values vs. estimation values and the underestimation, exact estimation, and overestimation percentages of: (A) the CNN model; and (B) the CNN+LSTM model

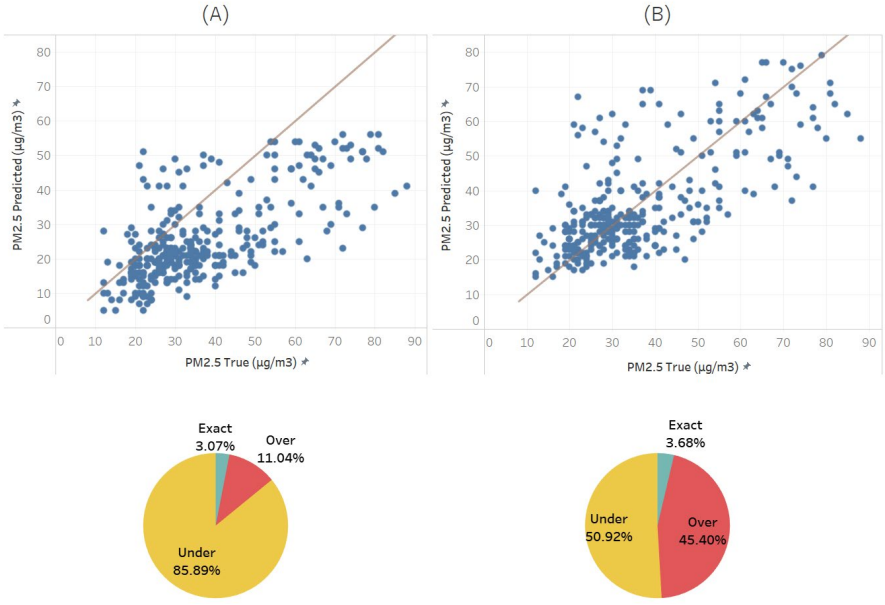

Figure 11. The scatter plot of true values vs. 24-h forecast values and the underestimation, exact estimation, and overestimation percentages of: (A) the CNN model; and (B) the CNN+LSTM model

Comparing the scatter plots in Figure 10, we found that the CNN+LSTM model yielded relatively more accurate predictions, as seen from the fact that there were a larger number of points located around the line with slope 1. The CNN+LSTM model's underestimation and overestimation rates were equivalent, but the CNN model largely generated underestimations, as seen on the pie charts, mostly due to its ability to infer haze features from images.

It is worth noting that both models made more accurate predictions at low values whereas at higher values, most of the predictions were underestimations, particularly the CNN model. When it came to the 24-h forecast, as shown in Figure 11, the CNN model significantly underestimated the values in about 86% of the cases, but the CNN+LSTM model's underestimation and overestimation only differed by about 5%. The CNN model's forecasting performance had dropped in comparison to its estimate performance, resulting in substantial residuals that explained the three-evaluation metrics. The hybrid model, on the other hand, still produced nearly correct forecasts even at high concentrations, as demonstrated by the fact that, at high concentrations, the CNN+LSTM model made predictions that occurred close to slope 1 line. With the help of LSTM, CNN learned abstract features and concepts that indicated the high PM2.5 concentrations in certain picture pixels.

Figures 12 (A), (B), and (C) show examples of underestimation, exact estimation, and overestimation of current PM2.5 concentration values from the CNN+LSTM model. It is possible that the fact that the sky was blue and had lovely fluffy clouds contributed to Figure 12 (A)'s underestimation. Although the 59 μg/m^3 in this case was underestimated by 21 μg/m^3, it is still thought to be significant enough to prompt a healthy alert. The fact that there was no cloud and it appeared gloomy, especially at the skyline close to the structures, may be what caused the small overestimation in Figure 12 (C).

A B C

Figure 12. (A) An underestimation of CNN+LSTM: the true value of 80 μg/m^3 was predicted to be 59 μg/m^3; (B) an exact estimation of CNN+LSTM - 11 μg/m^3; and (C) an overestimation of CNN+LSTM: the true value of 49 μg/m^3 was predicted to be 56 μg/m^3.

In another experiment to analyze the predictions of PM2.5 concentrations according to collection periods and lighting conditions shown in Table 2, the evaluation metrics were measured at three different times: from 7 a.m. to 9 a.m., which represented early morning; from 10 a.m. to 2 p.m., which represented midday; and from 3 p.m. to 6 p.m., which represented late afternoon and evening. The performance results were similar to previous experimental studies, in which the hybrid model outperformed the standalone model by a small margin for estimation but by a larger margin for forecasting. Nevertheless, it is noteworthy to observe a fascinating aspect. For all the metrics, morning estimates and 24-h forecasts were typically the poorest, followed by midday, and evening estimates and forecasts had the best performance. There are a number of factors that could potentially contribute to this phenomenon. Firstly, the orientation of the IP camera was northward, directly facing the rising sun during the morning hours. Thus, on days with fewer clouds, the intense sunlight could potentially mitigate the influence of minute airborne particles on the appearance, color, and texture of the sky and buildings, thereby resulting in underestimations from the models.

Table 2. Performance of the CNN and the CNN+LSTM models on the test data set at three different collection periods: from 7 a.m. to 9 a.m., from 10 a.m. to 2 p.m., and from 3 p.m. to 6 p.m.

Metrics	Collection Periods	Estimated Current (t+0)		24-hour Forecast (t+24)	
		CNN	CNN+LSTM	CNN	CNN+LSTM
RMSE ($\mu g/m^3$)	7 a.m.–9 a.m.	14.83	13.86	19.46	16.26
	10 a.m.–2 p.m.	11.55	11.06	17.88	12.99
	3 p.m.–6 p.m.	6.36	6.03	11.24	7.22
MAE ($\mu g/m^3$)	7 a.m.–9 a.m.	11.18	10.43	16.36	12.52
	10 a.m.–2 p.m.	8.79	8.34	14.23	9.96
	3 p.m.–6 p.m.	4.88	4.50	9.75	5.48
MAPE (%)	7 a.m.–9 a.m.	28.78	30.23	34.53	32.75
	10 a.m.–2 p.m.	26.91	26.90	36.90	30.32
	3 p.m.–6 p.m.	18.62	16.93	34.69	19.71

Secondly, it is worth noting that morning PM2.5 levels tended to be higher compared to the afternoon, leading to greater differences in model projections during the morning as opposed to midday or evening collection periods.

4. Conclusions

In this paper, we investigated the potential of using photos of surrounding buildings and the sky as a low-cost alternative to estimate and forecast the concentration level of particulate matter less than 2.5 diameters (PM2.5). To examine the representation of haze in digital photos captured by an IP camera over a five-month period in Bangkok, a hybrid of a convolutional neural network (CNN) and a long short-term memory neural network (LSTM) architectural model were proposed. Based on experimental results validated using data from a nearby air quality monitoring station, the proposed hybrid architecture outperformed the conventional CNN architecture. Within an order of magnitude error, the hybrid model was able to estimate the current PM2.5 levels and forecast the next day's PM2.5 levels based on the present captured image. Our contribution can assist the public in recognizing the health risks posed by particulate matter over time without relying on official statistics from authorities alone. The image-based approach in this study still has room for improvement. In future work, we can use traditional computer vision techniques to extract features from sky photographs to enhance the abstract features provided by CNN layers. A succession of camera photos may also be utilized to improve prediction accuracy.

Note that the proposed model serves as the prototype for confirming the feasibility of the system. Further considerations and improvements should be made before implementing the suggested model on the actual sites to ensure its applicability in different working conditions. For example, with different levels of lighting exposure on different sites, the model can mistakenly label low-level exposure images as high PM2.5 concentrations images. Some normalization methods, such as batch normalization layers, can be used to allow the model to capture the variance of the sky caused by the varied PM2.5 levels after eliminating the impact of the background exposure.

In future works, we plan to investigate the performance of other recently proposed neural network models for sequential data processing, including Transformer, as discussed in Zeng *et al.* [32]. Furthermore, to reduce the impact of lighting in different areas, image lighting exposure normalization algorithms should be investigated or included into the model to diminish the impact of different lighting exposures in different areas.

5. Acknowledgements

The authors would like to thank the Pollution Control Department, Thailand, for the air quality monitoring station dataset.

References

[1] Bazyar, J., Pourvakhshoori, N., Khankeh, H., Farrokhi, M., Delshad, V. and Rajabi, E., 2019. A comprehensive evaluation of the association between ambient air pollution and adverse health outcomes of major organ systems: a systematic review with a worldwide approach. *Environmental Science and Pollution Research*, 26(13), 12648-12661, https://doi.org/10.1007/s11356-019-04874-z.

[2] Kuylenstierna, J., Malley, C., Büker, P., and Marmon, T., 2020. *Air Pollution and Its Impact on Human Health: An Important Driver for Achieving the 1.5°C Goal of the Intergovernmental Panel on Climate Change*. [online] Available at : https://www.sei.org/publications/air-pollution-and-human-health-impacts/.

[3] Shi, L., Wu, X., Yazdi, M.D., Braun, D., Awad, Y.A., Wei, Y., Liu, P., Di, Q., Wang, Y., Schwartz, J., Dominici, F., Kioumourtzoglou, M.-A. and Zanobetti, A., 2020. Long-term effects of PM2.5 on neurological disorders in the American Medicare population: a longitudinal cohort study. *The Lancet Planetary Health*, 4(12), e557-e565, https://doi.org/10.1016/s2542-5196(20)30227-8.

[4] Tai, A.P.K., Mickley, L.J. and Jacob, D.J., 2010. Correlations between fine particulate matter (PM2.5) and meteorological variables in the United States: Implications for the sensitivity of PM2.5 to climate change. *Atmospheric Environment*, 44(32), 3976-3984, https://doi.org/10.1016/j.atmosenv.2010.06.060.

[5] Ren-Jian, Z., Ho, K.F. and Zhen-Xing, S., 2012. The role of aerosol in climate change, the environment, and human health. *Atmospheric and Oceanic Science Letters*, 5(2), 156-161, https://doi.org/10.1080/16742834.2012.11446983.

[6] Pollution Control Department, 2019. *Air4Thai*. [online] Available at : http://air4thai.pcd.go.th/.

[7] Chen, T.-F., Tsai, C.-Y. and Chang, K.-H., 2013. Performance evaluation of atmospheric particulate matter modeling for East Asia. *Atmospheric Environment*, 77, 365-375, https://doi.org/10.1016/j.atmosenv.2013.05.025.

[8] Mehmood, K., Bao, Y., Saifullah, Cheng, W., Khan, M.A., Siddique, N., Abrar, M.M., Soban, A., Fahad, S. and Naidu, R., 2022. Predicting the quality of air with machine learning approaches: Current research priorities and future perspectives. *Journal of Cleaner Production*, 379 Part 2, https://doi.org/10.1016/j.jclepro.2022.134656.

[9] Zhang, B., Rong, Y., Yong, R., Qin, D., Li, M., Zou, G. and Pan, J., 2022. Deep learning for air pollutant concentration prediction: A review. *Atmospheric Environment*, 290, https://doi.org/10.1016/j.atmosenv.2022.119347.

[10] US EPA, 2022. *Health and environmental effects of particulate matter (PM)*. [online] Available at : https://www.epa.gov/pm-pollution/health-and-environmental-effects-particulate-matter-pm.

[11] Yao, L., Kong, S., Zheng, H., Chen, N., Zhu, B., Xu, K., Cao, W., Zhang, Y. Zheng, M., Cheng, Y., Hu, Y., Zhang, Z., Yan,Y. Liu, D., Zhao, T., Bai, Y. and Qi, S., 2021. Co-benefits of reducing PM2.5 and improving visibility by covid-19 lockdown in Wuhan. *npj Climate and Atmospheric Science*, 4(1), https://doi.org/10.1038/s41612-021-00195-6.

[12] Liu, C., Tsow, F., Zou, Y. and Tao, N., 2016. Particle pollution estimation based on image analysis. *PLoS ONE*, 11(2), 1-14, https://doi.org/10.1371/journal.pone.0145955.

[13] He, Z., Ye, X., Gu, K. and Qiao, J., 2018. Learn to predict pm2.5 concentration with image contrast-sensitive features. *Proceedings of the 2018 37th Chinese Control Conference (CCC)*, Wuhan, China, July 25-27, 2018, pp. 4102-4106.
[14] Chen, Q., Chen, W. and Pan, G., 2021. An improved picture-based prediction method of $PM_{2.5}$ concentration. *IET Image Processing*, 16(11), 2827-2833, https://doi.org/10.1049/ipr2.12204.
[15] Wang, G., Shi, Q. and Jiang, K., 2022. $PM_{2.5}$ concentration measurement based on image perception. *Electronics*, 11(9), https://doi.org/10.3390/electronics11091298.
[16] Yao, S., Wang, F. and Huang, B., 2022. Measuring $PM_{2.5}$ concentrations from a single smartphone photograph. *Remote Sensing*, 14(11), https://doi.org/10.3390/rs14112572.
[17] Al-Saffar, A.A.M., Tao, H., and Talab, M.A., 2017. Review of deep convolution neural network in image classification. *Proceedings of the 2017 International Conference on Radar, Antenna, Microwave, Electronics, and Telecommunications (ICRAMET)*, Jakarta, Indonesia, October 23-24, 2017, pp. 26-31.
[18] Schultz, M.G., Betancourt, C., Gong, B., Kleinert, F., Langguth, M., Leufen, L.H., Mozaffari, A. and Stadtler, S., 2021. Can deep learning beat numerical weather prediction? *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 379(2194), https://doi.org/10.1098/rsta.2020.0097.
[19] Zhang, C., Yan, J., Li, C., Rui, X., Liu, L. and Bie, R., 2016. On estimating air pollution from photos using convolutional neural network. *Proceedings of the 24th ACM International Conference on Multimedia (MM),* Amsterdam, The Netherlands, October 15-19, 2016, pp. 297-301.
[20] Zhang, C., Yan, J., Li, C., Wu, H. and Bie, R., 2018. End-to-end learning for image-based air quality level estimation. *Machine Vision and Applications*, 29(4), 601-615, https://doi.org/10.1007/s00138-018-0919-x.
[21] Jiang, Z., Zheng, T., Bergin, M. and Carlson, D., 2022. Improving spatial variation of ground-level $PM_{2.5}$ prediction with contrastive learning from satellite imagery. *Science of Remote Sensing*, 5, https://doi.org/10.1016/j.srs.2022.100052.
[22] Zaini, N., Ean, L.W., Ahmed, A.N. and Malek, M.A., 2022. A systematic literature review of deep learning neural network for time series air quality forecasting. *Environmental Science and Pollution Research*, 29(13), 4958-4990, https://doi.org/10.1007/s11356-021-17442-1.
[23] Chen, H.-C., Putra, K.T., Weng, C.-E. and Lin, J.C.-W., 2022. A novel predictor for exploring PM2.5 spatiotemporal propagation by using convolutional recursive neural networks. *Journal of Internet Technology*, 23(1), 165-176.
[24] Jiang, X., Luo, Y. and Zhang, B., 2021. Prediction of PM2.5 concentration based on the LSTM-TSLightGBM variable weight combination model. *Atmosphere*, 12(9), https://doi.org/10.3390/atmos12091211.
[25] Shi, L., Zhang, H., Xu, X., Han, M. and Zuo, P., 2022. A balanced social LSTM for PM2.5 concentration prediction based on local spatiotemporal correlation. *Chemosphere*, 291(Part 3), https://doi.org/10.1016/j.chemosphere.2021.133124.
[26] Song, S., Lam, J.C.K., Han, Y. and Li, V.O.K., 2020. ResNet-LSTM for real-time PM2.5 and PM10 estimation using sequential smartphone images. *IEEE Access*, 8(99), 220069-220082, https://doi.org/10.1109/ACCESS.2020.3042278.
[27] Won, T., Eo, Y.D., Sung, H., Chong, K.S., Youn, J. and Lee, G.W., 2022. Particulate matter estimation from public weather data and closed-circuit television images. *KSCE Journal of Civil Engineering*, 26, 865-873, https://doi.org/10.1007/s12205-021-0865-4.
[28] Yu, Y., Si, X., Hu, C. and Zhang, J., 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Computation*, 31(7), 1235-1270, https://doi.org/10.1162/neco_a_01199.
[29] Hochreiter, S. and Schmidhuber, J., 1997. Long short-term memory. *Neural Computation*, 9(8), 1735-1780, https://doi.org/10.1162/neco.1997.9.8.1735.

[30] Zhao, F., Feng, J., Zhao, J., Yang, W. and Yan, S., 2018. Robust LSTM-autoencoders for face de-occlusion in the wild. *IEEE Transactions on Image Processing*, 27(2), 778-790, https://doi.org/10.1109/TIP.2017.2771408.

[31] Peng-in, B., Sanitluea, P., Monjatturat, P., Boonkerd, P. and Phosri, A., 2022. Estimating ground-level $PM_{2.5}$ over Bangkok metropolitan region in Thailand using aerosol optical depth retrieved by MODIS. *Air Quality, Atmosphere and Health*, 15(11), 2091-2102, https://doi.org/10.1007/s11869-022-01238-4.

[32] Zeng, A., Chen, M., Zhang, L. and Xu, Q., 2023. Are transformers effective for time series forecasting? *Proceedings of the 37th AAAI Conference on Artificial Intelligence (AAAI-23)*, Washington DC, USA, February 7-14, 2023, pp. 11121-11128.