IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 June 2023, accepted 3 July 2023, date of publication 7 July 2023, date of current version 13 July 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3293398

RESEARCH ARTICLE

Spatio-Temporal PM2.5 Forecasting in Thailand Using Encoder-Decoder Networks

NATCH SIRISUMPUN[1], KRITCHART WONGWAILIKHIT[2], PISUT PAINMANAKUL[3], AND PEERAPON VATEEKUL[1], (Member, IEEE)

[1]Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Pathum Wan, Bangkok 10330, Thailand
[2]Department of Chemical Engineering, Faculty of Engineering, Chulalongkorn University, Pathum Wan, Bangkok 10330, Thailand
[3]Department of Environmental Engineering, Faculty of Engineering, Chulalongkorn University, Pathum Wan, Bangkok 10330, Thailand

Corresponding author: Peerapon Vateekul (peerapon.v@chula.ac.th)

ABSTRACT PM2.5 is a type of particulate matter that contributes to air pollution in Thailand on a yearly cycle. Exposure to PM2.5 can cause acute health problems, including respiratory and cardiovascular diseases, as well as an increased risk of premature death. In this paper, we present a spatio-temporal model based on a deep learning approach for PM2.5 concentration prediction via an image-like approach at a country-wide level. Our model: SimVP-CFLL-ML is based on a video prediction model, called ''SimVP''. To enhance its performance when attempting to predict high PM2.5 concentration, SimVP includes two major improvements i.e. a cross-feature learning layer (CFLL) using 1×1 convolution layer to learn feature correlation and a masking layer (ML) to calculate loss in specific locations. The experiment is conducted on data collected from the pollution control department (PCD) of Thailand and sensor for all (SFA). Results show that our model outperforms all baselines. Our model's F1 perforance is 3.51% better than the best baseline model for classifying high PM2.5 concentration class.

INDEX TERMS PM2.5 prediction, machine learning, deep learning, video prediction.

I. INTRODUCTION

Air pollution is the problem of unclean air with strange particles in it. When there are more dirty particles in the air, people who live nearby are more likely to get sick. PM2.5 stands for ''particles with a diameter of fewer than 2.5 microns'' that measure how dirty the air is. It is noted that PM2.5 particles are small enough that they can get past the defenses of the human respiratory system, get deep into the lungs, and even get into the bloodstream. Such an outcome can adversely affect people's health, leading to many chronic diseases. Throughout the year in Thailand, particle concentration is exceptionally high. Every year between the months of December and April, farmers in Southeast Asia burn crop residues to clear fields for the following year's planting. Although this method is cost-effective and efficient, it has become a major environmental and health concern as it produces clouds of smoke and air pollutants that can cross international borders. The combustion of crop residues is a primary source of PM2.5 concentration in Thailand. This pollution contributes to the urban industry and vehicle pollution that already exists, making it a significant problem.

The impact of transboundary burning plays a dominant role in determining PM2.5 concentrations in Thailand, accounting for approximately 67% of the influence, while local sources contribute approximately 33% [1]. This emphasizes the critical fact that the primary source of PM2.5 emissions primarily originates from burning activities, occurring outside the country's borders. It is essential, therefore, to improve the accuracy of predictions for high PM2.5 concentrations in Thailand. Preserving public health is vital. Air pollution is a factor that affects the entire population. There is a need for better air quality management strategies, and more direct governmental decisions, regarding pollution control and urban planning.

In Figure 1, the daily average PM2.5 concentration in Thailand in 2022 indicates that the month's PM2.5 concentration is high, which corresponds to the periods when fields are left to burn. Burning can be detected from satellites as hotspots, allowing for the finding of a spatio-temporal correlation of the relationship between PM2.5 and such hotspots.

The associate editor coordinating the review of this manuscript and approving it for publication was Prakasam Periasamy.

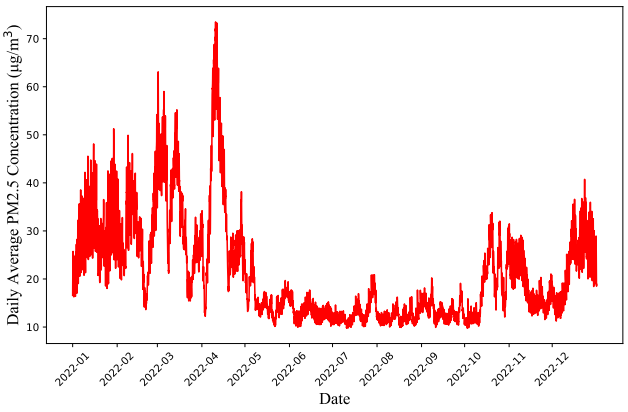


FIGURE 1. The daily average PM2.5 concentration in Thailand in 2022.

Numerous studies have established a correlation between these hotspots and PM2.5 levels [2], [3], [4], [5]. There is also a correlation between PM2.5 concentration and wind features [6], [7], [8], [9]. Wind patterns play a significant role in the dispersion and transport of PM2.5 particles in the atmosphere. Both the direction and wind speed determine the movement of PM2.5 from their source regions to other areas. With the help of fire hotspots and wind characteristics, it is possible to predict where particulate matter travels and when high PM2.5 concentration is likely to occur. According to data obtained, air quality in Thailand has a strong relationship with fire hotspots and wind patterns in Southeast Asia [10], [11], [12], [13].

As a time-series problem, many machine learning models have been used to predict PM2.5's future trend. For air pollution prediction, common traditional time-series machine learning models such as ARIMA [14], [15], [16], and SVM-based regression model [17], [18], [19] are frequently employed. Recently, new models based on deep learning such as recurrent neural networks (RNN) can learn complex non-linearity relations of time-series data, providing greater accuracy [20]. Long short-term memory (LSTM) has been used to predict the next four timesteps of PM2.5 concentrations [21], [22]. Besides, researchers have compared the performance of gated recurrent units (GRU) with RNN and LSTM for air quality prediction using AirNet data [23], [24]. Results show that GRU performed better.

PM2.5 concentration prediction can be differentiated into two distinct categories: temporal PM2.5 concentration prediction and spatio-temporal PM2.5 concentration prediction. For temporal PM2.5 concentration prediction, common traditional time-series machine learning models and common temporal deep learning models such as RNN, LSTM, and GRU are often used. A number of prior works have tried to predict PM2.5 concentration based on temporal prediction techniques [14], [15], [16], [17], [18], [19], [20], [21], [22], [23], [24], [25], [26], [27], [28]. However, such techniques are seen to be limited in their scope for prediction due to their focus on single locations. In contrast, spatio-temporal PM2.5 concentration prediction encompasses predicting PM2.5 at multiple locations at the same time [29], [30], [31]. Moreover, in spatio-temporal prediction, results can be predicted via an image-like approach.

The image-like approach for spatio-temporal PM2.5 concentration means that predicted results can easily be seen viewed on a heatmap by experts at a country-wide level. The image-like approach is easier for ordinary humans to see the picture of PM2.5 concentration distribution at a high level like a heatmap. To the best of our knowledge, only one recent study on how to predict PM2.5 concentration using spatio-temporal data has been shown in an image-like approach. The model was based on CNN-LSTM for estimating PM2.5 concentration inside a city [32]. This model used multiple features, including meteorology, traffic conditions, and urban morphology while learning cross-feature correlations through ResNet and 1×1 convolution layers. However, it is emphasized that this work was conducted at a city-wide level only: fire hotspots were not considered. Using deep learning techniques, video prediction models can be used in an image-like approach to predict PM2.5 concentrations [33], [34], [35], [36], [37], [38]. Video prediction is a rich avenue for the learning of spatio-temporal correlations, providing prediction capabilities. We use each timestep of spatio-temporal PM2.5 concentration as a frame for a video prediction model.

In this paper, the application of a spatio-temporal PM2.5 concentration prediction model having an image-like approach is duly investigated. This approach is based on the video prediction model i.e. "SimVP" that has been specifically designed for spatio-temporal prediction by origin. This inspires us to incorporate SimVP into spatio-temporal PM2.5 prediction. To improve the performance of SimVP when predicting high PM2.5 concentration, SimVP employs a cross-feature learning layer (CFLL) in conjunction with a masking layer (ML). The proposed model is applied country-wide. Herein, our work is new and authentic. Thailand PM2.5 dataset is an hourly prediction for the next 24 timesteps using the previous 24 timesteps. Moreover, fire hotspots and wind patterns ascertained via from different data sources are preprocessed and used in our work.

The contributions of this paper are summarized as follows:

- SimVP, a video prediction deep learning model, is used as a backbone model for spatio-temporal PM2.5 concentration prediction in an image-like approach on a country-wide level.
- The model was further improved with two proposed modules. Firstly, a cross-feature learning layer (CFLL) was added to combine all features (wind and fire hotspots) and enriched the spatio-temporal input before feeding it to the model. Secondly, a masking layer (ML) was added to make the model focus only on regions in Thailand.
- Important features from other sources (wind and fire hotspots) were combined together for Thailand's

PM2.5 prediction. Fire hotspots, especially, due to crop burning, are a major source of PM2.5 pollution.

The rest of this paper is organized as follows: Section II discusses the related work. Section III describes the dataset and proposed method. Section IV describes the experimental settings and the experiment's results. Section V is the discussion and future work. Section VI gives the conclusion of this paper.

II. RELATED WORK

A. TEMPORAL PM2.5 CONCENTRATION PREDICTION

The techniques most frequently employed by researchers for air quality prediction are common deep learning techniques such as RNN, LSTM, and GRU. Liu et al. [25] established an air quality predictor (AAQP), which is an encoder-decoder model employing a fully connected neural network as an encoder and RNN as a decoder, using either LSTM or GRU. Nguyen et al. [26] used a genetic algorithm for selecting features and an encoder-decoder model with long short-term memory (LSTM) for predicting PM2.5. In addition, there are studies that use PM2.5 data from neighboring stations to improve the accuracy of predictions. Cheng et al. [27] proposed the attentional deep air quality inference network (ADAIN), which combined feedforward neural networks, recurrent neural networks, and an attention-based pooling layer that automatically learns the weights of features from various monitoring stations. Liu et al. [28] presented the geographic self-organizing map (GeoSOM) in conjunction with GRU, which clusters groups of stations and predicts PM2.5 concentrations at the target station, using data from all stations within the cluster.

For temporal PM2.5 concentration prediction, the deep learning models often used are LSTM and GRU. These works apply the models to their own techniques. However, prediction remains focused on individual stations, and the ability of the models to predict multiple stations simultaneously is still limited. Moreover, spatial feature factors from other locations that can affect PM2.5 concentrations have not been fully considered. In this work, we use LSTM and GRU as baseline models for temporal PM2.5 concentration prediction.

B. SPATIO-TEMPORAL PM2.5 CONCENTRATION PREDICTION

PM2.5 concentration can be predicted at multiple locations: in a spatial way. This strategy's main goals are to increase efficiency by utilizing a single model to anticipate data from multiple PM2.5 stations at once, and learn a high-level PM2.5 correlation for the interested region. Spatio-temporal PM2.5 prediction can be separated into two categories i.e. predictions based on multiple PM2.5 stations, and image-like data prediction. Xu et al. [29] presented an LSTM autoencoder multitask learning model to predict PM2.5 time-series at a number of different places in the city. Shi et al. [30] introduced a spatial attention-based long short-term memory (SA-LSTM), which combined LSTM and a spatial attention mechanism to utilize the spatio-temporal information of multiple factors in an adaptive manner. Wen et al. [31] set up ST-E-LSTME that used k-nearest stations in combination with CNN and LSTM to predict PM2.5 at multiple stations. Zhang et al. [32] initiated Deep-AIR, a novel hybrid deep learning framework that combined CNN with LSTM. Deep-AIR uses city-wide urban dynamics as an image-like data approach. Deep-AIR incorporates both CNN having 1×1 convolution layers to extract spatial features and LSTM to learn the temporal correlation of the extracted features. The 1×1 convolution layers are adopted to strengthen the learning of cross-feature spatial representation between air pollution and various important urban dynamic features.

As for spatio-temporal PM2.5 concentration prediction, in this work, we focus on an image-like approach. The Deep-AIR model is similar to our approach. However, Deep-AIR is still based on simple deep learning models. Deep-AIR uses CNN for spatial learning and LSTM for temporal learning separately. Herein, the Deep-AIR model has been chosen as a baseline model for spatio-temporal PM2.5 concentration prediction.

C. VIDEO PREDICTION BASED ON DEEP LEARNING

Video prediction is a subfield of computer vision's deep learning that involves predicting upcoming frames based on past frames. This is the problem involving the prediction of spatio-temporal sequences. Shi et al. [33] set up ConvLSTM: a type of recurrent neural network for spatio-temporal prediction that employed convolutional structures in both the input-to-state and state-to-state transitions. Wang et al. [34] proposed PredRNN; the core of this model is a new spatio-temporal long short-term memory (ST-LSTM) unit that extracts and memorizes spatial and temporal representations simultaneously. Wang et al. [35] put forward MIM networks for learning higher-order non-stationarity from spatio-temporal dynamics. Lin et al. [36] introduced a self-attention mechanism into ConvLSTM, which is self-attention memory (SAM), in order to remember features with long-range dependencies in both the spatial and temporal domains. Chang et al. [37] submitted a motion-aware unit (MAU) to capture dependable inter-frame motion information by expanding the temporal receptive field of the predictive units. Zhangyang et al. [38] demonstrated SimVP (Simple yet better Video Prediction), which is fully CNN-based but achieved state-of-the-art performance on five benchmark datasets. SimVP is easy to comprehend and use as a common benchmark due to its simplicity.

Originally designed for predicting spatio-temporal problems, video prediction techniques can be applied to our work. Our work focuses on predicting spatio-temporal PM2.5 contentrations, extending across the whole country. In this paper, ConvLSTM, PredRNN, and SimVP were chosen as baselines for the video prediction approach (spatio-temporal). Moreover, since SimVP showed the best performance in recent video prediction works, it was chosen as our model backbone.

III. METHODOLOGY

This paper addresses spatio-temporal PM2.5 prediction using an image-like approach and a model based on video prediction. From beginning to end, the modeling process consists of two primary parts. First, the data is preprocessed, employing different preprocessing techniques, depending on the data source. The second part is the proposed model: SimVP, which is the backbone model based on video prediction.

A. DATASET AND DATA PREPROCESSING

1) DATASET DESCRIPTION

The PM2.5 data used in this study was collected from the sensor for all (SFA)[1] and pollution control department (PCD)[2] in Thailand. Both centres are air quality monitoring systems in Thailand. These systems deploy sensors in various locations across the nation, enabling real-time monitoring of air quality data. The Faculty of Engineering at Chulalongkorn University, the Ministry of Energy, and the Electricity Generating Authority of Thailand have collaborated to provide access to the SFA data through their website and application. Presently, in Thailand, there are 501 operational PM2.5 measurement stations for SFA sources and 213 operational PM2.5 measurement stations for PCD sources. SFA and PCD are ones of the major organizations responsible for handling PM2.5 concentration and providing the reliable data in Thailand.

ERA5 [39], which stands for ECMWF Reanalysis v5, is the main source of the wind data utilized in this study. The Copernicus Climate Change Service, which is funded by the European Union, created ERA5, the fifth generation of a reanalysis of the global atmosphere. ERA5 provides hourly estimates of a large number of atmospheric, land, and oceanic climate variables, making it essential for understanding and monitoring climate change processes and forecasting future climate change. In this study, we used two wind data variables from ERA5: the u-component of wind, which represents the Easterly wind speed in meters per second, and the v-component of wind, which represents the Northern wind speed in meters per second. High PM2.5 concentrations are correlated with a wind speed of 850 hPa [6]. Such wind data at the 850 hPa pressure level was selected because it is most likely the primary factor influencing the movement of PM2.5 particles from neighboring countries into Thailand, where crop burning is the source of these particles. It is noted that when crop burning occurs in neighboring nations, wind patterns play a crucial role in the transport of these particles into Thailand.

This study utilized fire hotspots data from FIRMS: an acronym for Fire Information for Resource Management System [40], [41], [42]. This system was developed by NASA to provide information on global hotspots. Global hotspots are areas with high heat values on the surface of the earth that can be detected almost in real-time. The primary objective of this system is to support fire management efforts by tracking and notifying potential fires. FIRMS uses information from both VIIRS and MODIS satellite sensor systems to detect these hotspots. VIIRS is available on the Suomi-NPP and NOAA-20 satellites, whereas MODIS is available on the Terra and Aqua satellites. Within three hours of detecting a hotspot, FIRMS can provides both the location and time of detection.

In summary, this study makes use of three distinct collections of data: namely, PM2.5 concentration, wind speed, and fire hotspots. In Table 1, a summary of dataset characteristics are given.

TABLE 1. Summary of dataset characteristics.

Feature Name	Units	Description
PM2.5	$\mu g/m^3$	PM2.5 concentration
Fire Hotspot	-	Hotspot detected location
u-component of wind	m/s	Eastward wind speed
v-component of wind	m/s	Northward wind speed

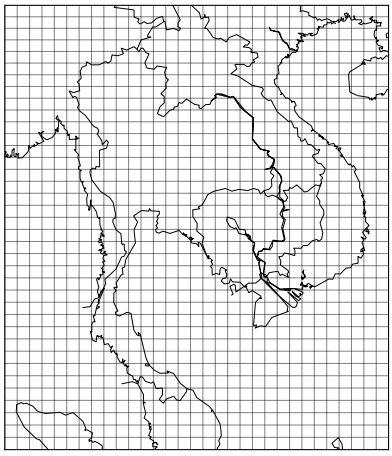


FIGURE 2. The resulting format after preprocessing each feature at each timestep.

2) DATA PREPROCESSING

Data preprocessing can be divided into three parts, based on the type of data from each source. Each source employs different preprocessing techniques, depending on the source of the data. The objective of preprocessing is to create a uniform format that can combine together all the information from each source. The resulting format contains spatio-temporal data with a grid. The size of the grid used in this work was 0.5 degrees (latitude) and 0.5 degrees (longitude). 55.5 km x 55.5 km is the approximate size of the grid. Herein, the area of interest is defined by latitudes ranging from 4.0 degrees North to 22.0 degrees North, and longitudes ranging from 95.0 degrees East to 110.0 degrees East. This area covers the whole of Thailand on the map, resulting in spatial data maps for image-like PM2.5, as displayed in Figure 2. The spatial data maps will be combined into map sequences, yielding spatio-temporal data whereby each map in the sequence represents a one-hour timestep.

The ERA5 wind dataset is organized such that each grid is 0.25 degrees in both latitude and longitude. The timestep

[1] https://sensorforall.eng.chula.ac.th/
[2] https://air4thai.pcd.go.th/

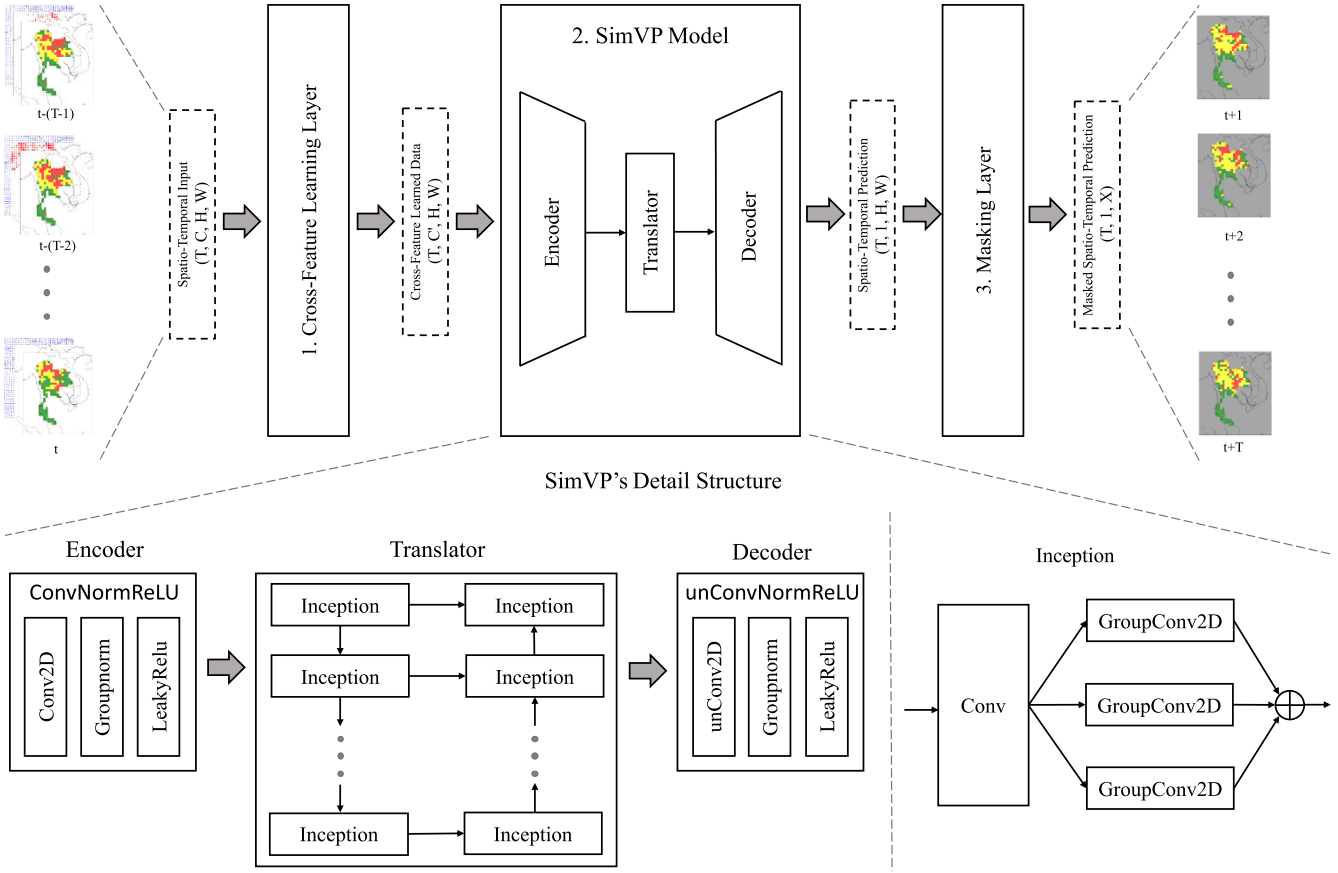


FIGURE 3. The proposed model uses SimVP as a backbone model along with additional layers: Cross-Feature Learning Layer (CFLL) and Masking Layer (ML). *T* denotes the number of timesteps for input and output, *C* denotes the input channel, *C'* denotes the channel after cross-feature learning, *H* denotes height, *W* denotes width, and *X* denotes the number of focus grids after the mask.

is also one hour. As a result, it is sufficient to filter the grid size from 0.25 degrees to 0.5 degrees. In the section quantifying the fire hotspot dataset, the characteristics of the obtained data incorporate the latitude, longitude, and detection time of the heat detected on the Earth's surface. The data, therefore, will be grouped according to the specified grid position and timestep during the preprocessing phase. Then, the number of heat detections for each grid will be counted as a fire hotspot and represented on the grid. Due to the fact that each PM2.5 station stores data every 5 min on average for SFA, it is necessary to calculate the hourly average of the PM2.5 concentration measured at each station for the PM2.5 dataset. For PCD, the timestep is already one hour. Then, stations within the same grid are grouped together, and the average PM2.5 concentration from each station is calculated to represent the hourly PM2.5 concentration of the grid. In some grids and during some hours, some data may be missing due to the fact that all PM2.5 station data was not collected. The missing values in the grid data are estimated using the grid of the nearest valid neighbor. Because of the spatial limitation of PM2.5 data, the values are only filled in for grids located in the Thailand zone. Due to the differences in their measurement units, the value ranges of data obtained from various sources can vary considerably. To eliminate the effects of these varying numerical ranges, the data must be normalized. To normalize the data, we utilized various normalization techniques for the various features. The number of fire hotspots and PM2.5 concentration were normalized in the range of 0 to 1, using min-max normalization. Meanwhile, standard normalization was applied to the u-component and v-component features of the wind dataset.

B. SPATIO-TEMPORAL MODEL BASED ON SimVP

Once data preprocessing is complete, the data become spatio-temporal data consisting of four channels, including fire hotspot data, PM2.5 concentration data, the u-component of wind, and the v-component of wind. In Figure 3, the model architecture composed of 3 modules is illustrated. The first module is the cross-feature learning layer (CFLL). The second module is the backbone video prediction model, which is SimVP. The last module is the masking layer (ML)

The first module is cross-feature learning, which employs a 1×1 convolution layer to understand feature correlation across the four channels. The processing of this convolution layer's kernel in each grid is important, as each grid has four distinct characteristics, and the kernel aids in learning the

relationships between them. The 1 × 1 convolution layer is identical to a linear combination of multiple feature maps in each grid. This is inspired by Deep-AIR [32], which uses the concept of a cross-feature learning layer known as AirRes (ResNet + 1 × 1 convolution layer). Once the channel interrelationship has been learned and encoded into a new number of channels to account for their mutual correlation, the encoded data will be sent to the SimVP video prediction model.

The second module is the core model which is SimVP. SimVP is the Simpler yet Better Video Prediction model, which is entirely built on CNN and trained by mean square error (MSE) loss in an end-to-end fashion. The model achieves state-of-the-art performance on five benchmark video prediction datasets without introducing any additional tricks or complex strategies. For our spatio-temporal PM2.5 concentration prediction, the SimVP model was selected due to its efficiency and simplicity. SimVP consists of the following three primary components: encoder, translater, and decoder. (i) A stack of ConvNormReLU blocks (Conv2d + LayerNorm + LeakyReLU) constitutes the encoder, which is responsible for extracting spatial features. (ii) The translator used to study temporal evolution is made up of Inception modules. These modules contain a Conv2d bottleneck with a 1 × 1 kernel, followed by GroupConv2d parallel operators. (iii) A stack of UnConvNormReLU blocks (ConvTranspose2d + GroupNorm + LeakyReLU) constitute the decoder, which integrates spatio-temporal information to predict future frames. Due to the fact that all of SimVP's components are based on CNN, the model is easy to comprehend, quick to train and predict, and provides exceptional prediction performance. Since we only wish to predict PM2.5 concentration, the final layer of the SimVP model, the convolution layer, is modified to predict only one channel.

For the last module, we use a masking layer to improve the performance of the model. The masking layer helps to restrict loss function by calculating the loss for the grids within a specific location, in this case: Thailand's grids. This ensures that the predicted values for grids located outside the country do not factor into the loss function. In conclusion, for each timestep, the model's final output includes only the grids within Thailand and their PM2.5 concentrations.

IV. EXPERIMENTAL SETTING AND RESULTS

A. EXPERIMENTAL SETTING

1) DATASET

The data utilized in our study were collected in the year 2022. Due to the limited period of PM2.5 data from SFA in Thailand, we chose the last ten days of each month as the test dataset, while the remaining days were split into training and validation datasets. This selection was made to reserve data trends that occur on a monthly basis and to assess the performance of the model in predicting various trends. After separating the data into train, validate, and test sets for each month, we applied a window sliding technique. The window sliding technique generates input and output for model training in order to forecast PM2.5 concentrations for the next 24 hours using data from the previous 24 hours.

2) BASELINE MODEL

This study evaluates and compares six baseline models that can be separated into two distinct categories. The first category is the PM2.5 prediction baseline. This category can be separated into two subcategories: temporal PM2.5 prediction, and spatio-temporal PM2.5 prediction in an image-like approach. For the temporal PM2.5 baseline models, both LSTM and GRU models, based on an encoder-decoder technique, were chosen. These models learned only temporal characteristics within each grid. In this work, Thailand had 198 grids, so it was necessary to train 198 distinct temporal models for each grid. For the spatio-temporal PM2.5 prediction in the image-like approach, the Deep-AIR network (ResNet + 1 × 1 convolution layer + LSTM) was chosen. The second category is the video prediction category, which involves both temporal characteristics and spatial correlation with deep learning. As baselines for this category, the standard video prediction models: ConvLSTM, PredRNN, and the most recent video prediction, SimVP were chosen.

3) EVALUATION CRITERIA

This paper evaluates the model's performance in three different ways. The first way is the regression aspect, which evaluates the results direct from the model that predicts the PM2.5 concentration values. In this case, this study uses MAE (mean average error) and RMSE (root mean square error) to evaluate the performance. MAE is defined as follows:

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

where n denotes the number of observations in the dataset, y_i denotes the actual value of the target variable for the i-th observation, and $\hat{y}$ denotes the predicted value of the target variable for all observations. RMSE is defined as follows:

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

where n denotes the number of observations in the dataset, y_i denotes the actual value of the target variable for the i-th observation $\hat{y}$ denotes the predicted value of the target variable for all observations.

The second way is the classification aspect, where the model's output is classified into three categories: Low for concentration below 35.5 μg/m^3, Medium for concentration between 35.5 μg/m^3 and 55.5 μg/m^3, and High for concentration above 55.5 μg/m^3. After categorizing the results, we evaluated the classification performance using the F1 score. The F1 score for each class is used to evaluate individual performance. The F1 score is defined as follows:

$$precision = \frac{TP}{TP + FP} \tag{3}$$

$$recall = \frac{TP}{TP + FN} \tag{4}$$

$$F1 = 2 \cdot \frac{precision \cdot recall}{precision + recall} \tag{5}$$

where TP denotes the number of true positives and FP denotes the number of false positives.

The third aspect of our study involves visualizing the performance of the models in terms of spatial information. This is done by applying the Pearson correlation coefficient for each prediction, which is calculated to determine the correlation between the ground truth and the model's prediction in that grid. The Pearson correlation coefficient can be expressed as:

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{6}$$

where n is the number of observations in the dataset, x_i is the i-th observation of variable x, y_i is the i-th observation of variable y, $\bar{x}$ is the mean of variable x, and $\bar{y}$ is the mean of variable y.

B. RESULTS

After training both the proposed SimVP-CFLL-ML model and baseline models on the Thailand dataset, we evaluated the results on the test dataset using four aspects: overall evaluation, the study of feature ablation, evaluation at each timestep, and spatial evaluation.

1) OVERALL EVALUATION

To provide an overall evaluation, we considered all grids in every timestep of the model prediction output. The purpose of this evaluation was to fully compare how well the model worked. The evaluation was conducted to evaluate the regression aspect using RMSE and MAE, as shown in Table 2.

Results demonstrate that SimVP-CFLL-ML was the winner with MAE and RMSE attaining 4.51 μg/m^3 and 7.59 μg/m^3, respectively. The best-performing PM2.5 prediction model baseline was Deep-AIR with MAE and RMSE achieving 4.96 μg/m^3 and 8.27 μg/m^3, respectively. The best-performing video prediction model baseline was SimVP with MAE and RMSE reaching 4.51 μg/m^3 and 7.63 μg/m^3, respectively. As for the PM2.5 prediction model baseline, Deep-AIR, which is a spatio-temporal approach was the winner, showing that the spatial-temporal approach outperformed the temporal approach, As for the video prediction approach, SimVP proved to be the winner.

Another overall evaluation was carried out by categorizing the model prediction output from each grid using pre-defined PM2.5 concentration classes, followed by an evaluation of the classification aspect using F1 per class and Macro F1. In Table 3, the results of this evaluation are shown.

Results reveal that SimVP-CFLL-ML was the winner with F1 scores in the Low, Medium, and High classes corresponding to 0.9613, 0.5419, and 0.4638, respectively: the Macro F1 score was 0.6557. Deep-AIR was the best-performing PM2.5 prediction model baseline with an F1 score of 0.9543 in the Low class, 0.4907 in the Medium class, and a Macro F1 score of 0.5912. However, GRU outperformed Deep-AIR in the High class, with an F1 score of 0.3687. SimVP was the best-performing video prediction model baseline with an F1 score of 0.9607 in the Low class, 0.5282 in the Medium class, and a Macro F1 score of 0.6371. However, it is seen that PredRNN outperformed SimVP in the High class, with an F1 score of 0.4287.

TABLE 2. Overall evaluation in terms of regression: MAE and RMSE. Boldfaces refer to the winners.

Model	MAE ↓	RMSE ↓
LSTM	5.82	9.34
GRU	5.55	8.99
Deep-AIR	4.96	8.27
ConvLSTM	5.02	8.26
PredRNN	4.67	7.87
SimVP	**4.51**	7.63
SimVP-CFLL-ML	**4.51**	**7.59**

TABLE 3. Overall evaluation in terms of classification: F1. Boldfaces refer to the winners.

Model	F1			
	Low ↑	Medium ↑	High ↑	Macro ↑
LSTM	0.9468	0.4388	0.3336	0.5731
GRU	0.9496	0.4551	0.3687	0.5912
Deep-AIR	0.9543	0.4907	0.3347	0.5912
ConvLSTM	0.9544	0.4937	0.3467	0.5982
PredRNN	0.9588	0.5078	0.4287	0.6318
SimVP	0.9607	0.5282	0.4225	0.6371
SimVP-CFLL-ML	**0.9613**	**0.5419**	**0.4638**	**0.6557**

TABLE 4. Features ablation study in terms of regression: MAE and RMSE. Boldfaces refer to the winners.

Features	MAE ↓	RMSE ↓
PM2.5	**4.46**	7.64
PM2.5 + Wind	4.49	7.67
PM2.5 + Fire	4.49	7.63
PM2.5 + Wind + Fire	4.51	**7.59**

TABLE 5. Features ablation study in terms of classification: F1. Boldfaces refer to the winners.

Features	F1			
	Low ↑	Medium ↑	High ↑	Macro ↑
PM2.5	0.9612	0.5336	0.3922	0.6290
PM2.5 + Wind	0.9610	0.5234	0.4038	0.6294
PM2.5 + Fire	0.9609	0.5355	0.4155	0.6373
PM2.5 + Wind + Fire	**0.9613**	**0.5419**	**0.4638**	**0.6557**

2) FEATURES ABLATION STUDY

In the features ablation study, we evaluated the model's performance using the same aspects as for overall evaluation. Thus, we considered all grids at every timestep of the model prediction output. The difference was in the features used for model input, which were the possible combinations of wind and fire hotspots for PM2.5 concentration. This study aimed to demonstrate the importance of feature selection.

TABLE 6. Each predicted timestep evaluation: MAE and Macro F1. Boldfaces refer to the winners. ts denotes the timestep of prediction.

ts	LSTM		GRU		Deep-AIR		ConvLSTM		PredRNN		SimVP		SimVP-CFLL-ML	
	MAE↓	Macro F1↑	MAE↓	Macro F1↑	MAE↓	Macro F1↑	MAE↓	Macro F1↑	MAE↓	Macro F1↑	MAE↓	Macro F1↑	MAE↓	Macro F1↑
1	4.71	0.6308	4.96	0.6435	3.29	0.6690	4.73	0.5263	3.27	0.7719	3.32	0.7557	**3.22**	**0.7833**
2	5.50	0.5659	5.01	0.6404	4.23	0.6311	4.86	0.6102	3.76	0.7289	3.72	0.7260	**3.64**	**0.7486**
3	5.58	0.5951	5.08	0.6342	4.40	0.6380	4.95	0.6073	4.16	0.6907	4.05	0.6892	**3.94**	**0.7224**
4	5.72	0.5828	5.18	0.6253	4.67	0.6175	5.02	0.6044	4.46	0.6598	4.26	0.6731	**4.19**	**0.7014**
5	5.82	0.5798	5.29	0.6156	4.86	0.6073	5.08	0.6022	4.67	0.6383	4.42	0.6637	**4.34**	**0.6868**
6	5.90	0.5751	5.41	0.6060	5.00	0.5978	5.12	0.6013	4.82	0.6249	4.55	0.6511	**4.49**	**0.6770**
7	5.96	0.5730	5.52	0.5963	5.11	0.5919	5.16	0.6001	4.91	0.6173	4.65	0.6458	**4.62**	**0.6632**
8	6.01	0.5713	5.63	0.5881	5.20	0.5862	5.20	0.5997	4.97	0.6131	4.76	0.6319	**4.70**	**0.6535**
9	6.05	0.5700	5.72	0.5806	5.29	0.5809	5.23	0.5994	5.02	0.6114	4.83	0.6284	**4.77**	**0.6476**
10	6.09	0.5690	5.80	0.5756	5.35	0.5783	5.26	0.5998	5.05	0.6113	4.88	0.6224	**4.85**	**0.6448**
11	6.13	0.5690	5.87	0.5725	5.40	0.5761	5.29	0.6003	5.08	0.6109	4.93	0.6241	**4.92**	**0.6361**
12	6.16	0.5686	5.93	0.5721	5.46	0.5746	5.32	0.6001	5.13	0.6111	4.99	0.6249	**4.97**	**0.6318**
13	6.21	0.5678	5.98	0.5720	5.51	0.5738	5.36	0.5996	5.18	0.6108	**5.09**	0.6049	5.10	**0.6262**
14	6.25	0.5668	6.03	0.5733	5.57	0.5733	5.40	0.5990	5.24	0.6099	**5.13**	0.6068	5.14	**0.6221**
15	6.29	0.5660	6.07	0.5747	5.62	0.5733	5.44	0.5980	5.29	0.6092	**5.16**	0.6104	5.18	**0.6191**
16	6.33	0.5655	6.10	0.5767	5.66	0.5740	5.47	0.5971	5.32	0.6092	**5.20**	0.6079	5.21	**0.6163**
17	6.35	0.5650	6.12	0.5799	5.68	0.5751	5.50	0.5971	5.34	0.6096	**5.21**	0.6114	5.23	**0.6209**
18	6.38	0.5652	6.15	0.5822	5.69	0.5769	5.53	0.5969	5.35	0.6103	**5.22**	0.6156	5.24	**0.6284**
19	6.42	0.5653	6.19	0.5837	5.70	0.5792	5.55	0.5958	5.37	0.6107	**5.22**	0.6117	5.28	**0.6213**
20	6.45	0.5649	6.25	0.5828	5.71	0.5810	5.59	0.5938	5.38	0.6105	**5.23**	0.6140	5.29	**0.6234**
21	6.49	0.5632	6.34	0.5789	5.72	0.5817	5.62	0.5909	5.39	0.6096	**5.26**	0.6019	5.31	**0.6244**
22	6.53	0.5607	6.45	0.5714	5.74	0.5813	5.66	0.5880	5.43	0.6065	**5.26**	0.6053	5.35	**0.6167**
23	6.58	0.5574	6.59	0.5616	5.77	0.5791	5.70	0.5848	5.48	0.6015	**5.29**	0.5971	5.38	**0.6122**
24	6.64	0.5533	6.76	0.5504	5.84	0.5744	5.76	0.5807	5.58	0.5927	**5.33**	0.5948	5.42	**0.6054**

The SimVP-CFLL-ML model was used for this experiment. As shown in Table 4, evaluation was determined to assess the regression aspect using RMSE and MAE. In Table 5, the aspect of classification was evaluated using the F1 score.

In table 4, it is seen that when input only contained PM2.5 concentration, the model demonstrated the best MAE performance, achieving an MAE of 4.46 μg/m^3. However when the input contained wind, fire hotspots, and PM2.5 concentration features, the model attained the best RMSE performance, demonstrating an RMSE of 7.59 μg/m^3. Moreover, in the classification evaluation aspect, the F1 score was best when all features were taken into consideration, with an F1 score of 0.9613 in the Low class, 0.5419 in the Medium class, 0.4638 in the High class, and 0.6557 in the Macro F1 score. Results indicate that the features did not significantly affect the model's performance in terms of regression, but feature selection was highly effective in terms of classification.

3) EACH TIMESTEP EVALUATION

To evaluate the performance of the model at each timestep, MAE was used to represent the aspect of regression evaluation, and Macro F1 was used to represent the aspect of classification evaluation. The model's performance was calculated by evaluating all grids, corresponding to each timestep. In Table 6, it is noted that as the prediction timestep increased, the performance of all models decreased. In most models, the first timestep revealed the best performance. In the second timestep, however, performance dropped significantly. To illustrate the difference in model performance between short-term and long-term predictions, the first timestep is used to represent the most short-term prediction and the twenty-fourth timestep is used to represent the most long-term prediction. In the first timestep, SimVP-CFLL-ML demonstrated the highest Macro F1 with 0.7833 but had the lowest MAE with 3.22 μg/m^3. In the twenty-fourth timestep, SimVP-CFLL-ML achieved the highest Macro F1 with 0.6054. For MAE, SimVP proved to be the best-performing model for the video prediction baseline and performed better than SimVP-CFLL-ML with MAE reaching 5.33 μg/m^3. Overall, as observed in Macro F1, the SimVP-CFLL-ML model outperformed other models at every timestep. As for MAE, in the first twelve timesteps SimVP-CFLL-ML performed the best. However, in the last twelve timesteps, SimVP recorded the best performance for MAE.

4) SPATIAL EVALUATION

For spatial evaluation, our aim is to see how efficiently the model worked. Using the Pearson correlation coefficient, we examined how well the ground truth and predictions matched up in each grid at every predicted timestep. In Figure 4, the correlation values are given. Besides, green means a high correlation, and red means a low correlation. The Pearson correlation coefficient ranged from -1 to 1, with higher values showing a stronger correlation between what actually happened and what was predicted. Moreover, when the overall performance of temporal PM2.5 prediction model baselines was compared to that of the image-like approach which included spatio-temporal PM2.5 prediction model baseline and video prediction model baselines, it was clearly found that the image-like approach did a much better job of predicting PM2.5 concentration in most parts of Thailand. This outcome demonstrated the fact that its visualization revealed more green grids than the temporal PM2.5 prediction baseline models. For all models, the South of Thailand is seen to be more of a challenge to predict PM2.5 concentration, due to the lower number of green grids in this area. In this area, LSTM and GRU revealed a yellow color, due to the fact that the Pearson correlation coefficient was lower, meaning that the PM2.5 concentration pattern in this area is hard to learn.

In Table 7, the number of grids with Pearson correlation coefficient for each model is shown. Results demonstrate that

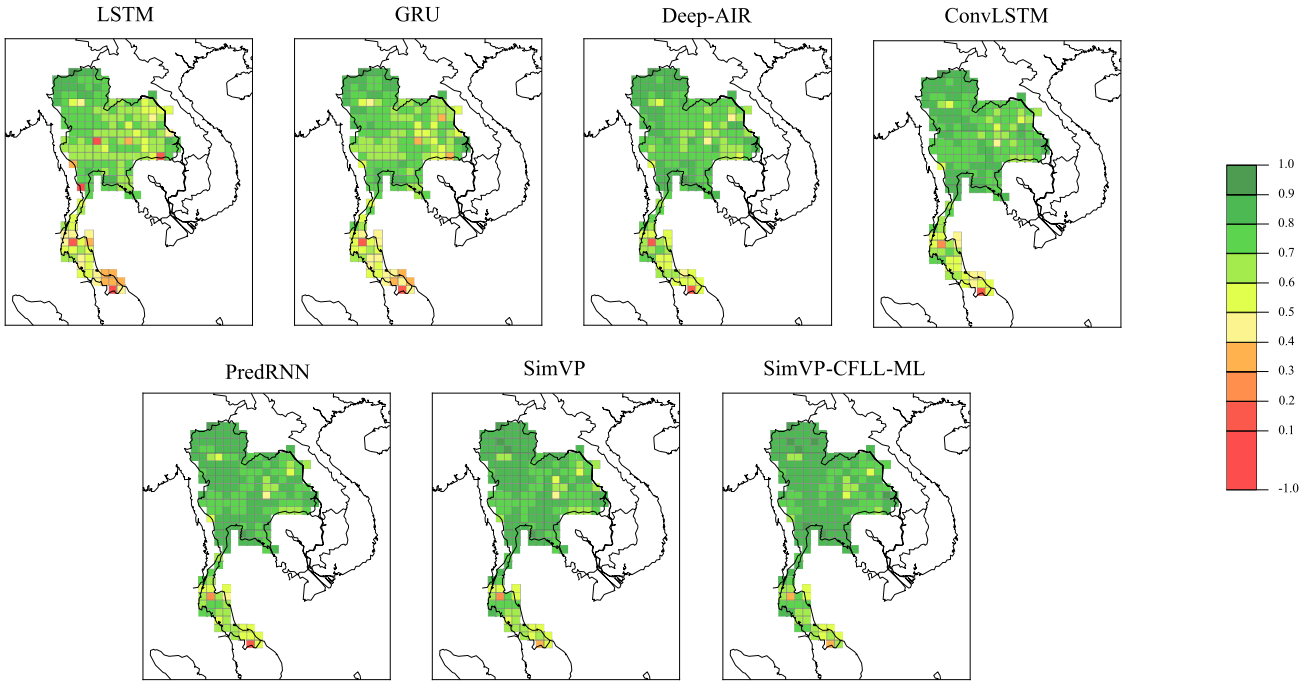


FIGURE 4. Spatial evaluation: using Pearson correlation coefficient in each grid.

TABLE 7. The total number of grids in the spatial evaluation for each Pearson correlation coefficient range.

Model	-1.0-0.2	0.2-0.4	0.4-0.6	0.6-0.8	0.8-1.0
LSTM	5	8	35	129	21
GRU	2	6	35	125	30
Deep-AIR	2	0	21	116	59
ConvLSTM	1	1	22	105	69
PredRNN	1	1	15	83	98
SimVP	0	2	13	78	105
SimVP-CFLL-ML	0	2	8	63	125

SimVP-CFLL-ML performed better than the other baseline models having the highest number of grids within the high Pearson correlation coefficient range of 0.8-1.0. Specifically, the resulting number of grids was 125 being more than half of the total of the predicted grids.

V. DISCUSSION AND FUTURE WORKS

The purpose of this study is to predict PM2.5 concentration in a spatio-temporal manner by combining video prediction deep learning techniques with the Thailand PM2.5 concentration dataset. The study also takes into account both wind and fire hotspot data as additional factors that can affect PM2.5 concentration.

Overall, SimVP-CFLL-ML achieved better performance than the baseline models, including the PM2.5 prediction baseline models and video prediction baseline models. As for the regression aspect, experimental results demonstrated that SimVP-CFLL-ML outperformed the best PM2.5 prediction baseline model (Deep-AIR) by 9.07% and 8.22% for MAE and RMSE, respectively. However, SimVP-CFLL-ML did not surpass the spatio-temporal deep learning model baseline (SimVP) as its performance did not significantly improve (Refer to Table 2). Nonetheless, in the classification aspect, experimental results showed that SimVP-CFLL-ML outperformed both the PM2.5 prediction baseline model and the video prediction baseline model. Besides, SimVP-CFLL-ML performed best in the high PM2.5 concentration class. Moreover, SimVP-CFLL-ML outperformed PredRNN which is the best-performing baseline model by 3.51% in the high PM2.5 concentration class (Refer to Table 3). In all models, the F1 score was the highest for the PM2.5 concentration class: Low. For most of the year, due to the rainy season and dearth of hotspots, PM2.5 concentration is usually Low. Thus, the model's performance in this class did not differ much. In other classes of PM2.5 concentration, especially the high PM2.5 concentration class, the SimVP-CFLL-ML model did much better than the other baseline models.

In the feature ablation study, the effects of wind at 850 hPa and fire hotspots at high PM2.5 concentration in Thailand are shown. The results of the SimVP-CFLL-ML model's feature ablation study show that feature combinations do not have a big effect on how well a model does at regression. When features were added, the model's MAE performance is seen to get a little worse, while its RMSE performance is seen to get a little better (see Table 4). However, in terms of classification, the addition of wind or fire hotspots alongside PM2.5 concentration improved the performance of the model. When using all features along with PM2.5 concentration, the F1 score of the high PM2.5 concentration class increased by 7.16%, and the macro F1 score increased by

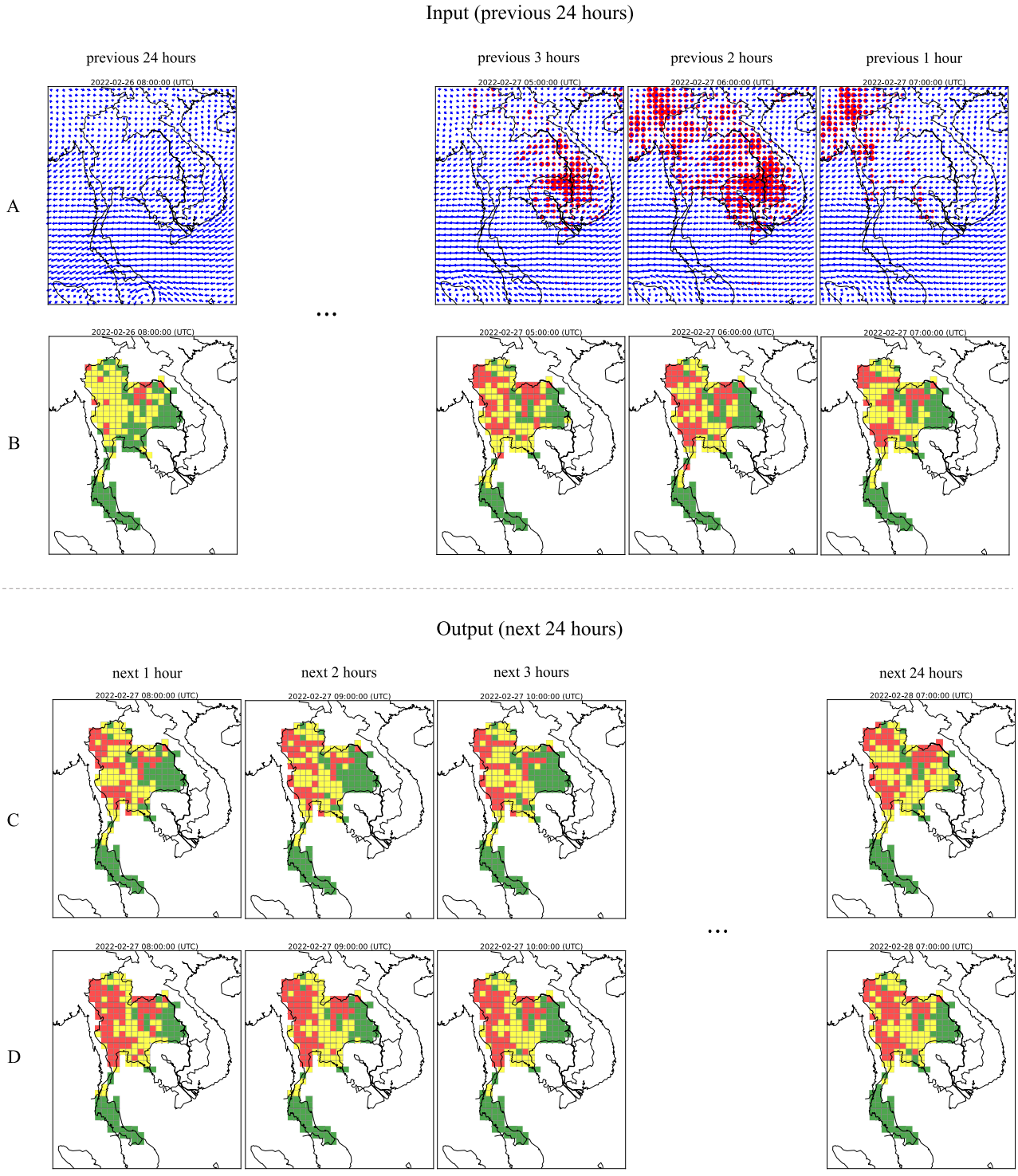


FIGURE 5. SimVP-CFLL-ML model's inputs and outputs for predicting PM2.5 concentration. The images in row A show the wind and fire hotspot data used as input for the model. The wind data is represented by blue arrows, with the direction of the arrow indicating the direction of the wind and the length of the arrow indicating the speed of the wind. The fire hotspot data is represented by red circles, with the size of the circle indicating the number of fire hotspots detected in the grid. The images in row B show the PM2.5 concentration data used as input for the model. The concentration levels are represented by different colors: red indicating high concentration, yellow indicating medium concentration, and green indicating low concentration. The images in row C show the ground truth PM2.5 concentration. The images in row D show the model's predicted PM2.5 concentration.

2.67% (see Table 5). These results suggest that the combination of wind and fire hotspots is the feature that greatly affects PM2.5 concentration. In addition to this, the data herein highlights the fact that the SimVP-CFLL-ML model

is capable of learning the correlation. Morever, the factor of fire, which is the burning of organic materials, is a widespread contributor to PM2.5 pollution worldwide in both developing and developed countries. Fire occurrences, whether through intentional land clearing, agricultural practices, or natural fire events, release significant amounts of particulate matter into the atmosphere.

In this work, the SimVP-CFLL-ML model is able to visualize an example of the inputs and outputs, as shown in Figure 5. The time span of the twenty-four timesteps for the model's input is covered from 2022-02-26 08:00:00 UTC to 2022-02-27 07:00:00 UTC. Moreover, the time span of the twenty-four timesteps for the model's output is covered from 2022-02-27 08:00:00 UTC to 2022-02-28 07:00:00 UTC. This case was specifically selected to demonstrate a scenario in which multiple fire hotspots were discovered in Cambodia; the wind direction flowed through this area, encompassing the central part of Thailand. To be more specific, the fire hotspot data indicated that several fire hotspots were discovered in the previous three hours, which is the same time that the wind data indicated that the wind direction flowed towards the central region of Thailand. As a consequence, there was an increase in the concentration of PM2.5 in this area. A second scenario was also depicted whereby fire hotspots were detected in Laos and the wind direction passed through Thailand, resulting in high PM2.5 concentrations in the north-eastern region of Thailand. A few fire hotspots were also seen in Thailand. In this scenario, a high PM2.5 concentration was recorded in Thailand due to the presence of PM2.5 from both inside and outside the country. When we compared the predictions made by the model with the actual results, we found that the model's accuracy was not completely identical in all grids. However, our model was able to accurately predict the concentration level in each zone at a regional level. This led us to the conclusion that the proposed model was reasonably acceptable in performance.

There are many ways in which this research can be improved. For instance, the low number of air pollution monitoring stations that are currently in operation has hampered the accuracy of the model that we have proposed. Hence, if more monitoring stations were available in Thailand, our model would be more robust and be able to perform better. Further, our experiment focuses only on the wind at 850 hPa, which has been reported to affect high PM2.5 concentration. Wind at different pressures has the potential to also affect PM2.5 concentration. Finally, the cross-feature learning layer would need to be improved if additional features were added, which would also cause the correlation to become more complicated.

VI. CONCLUSION

In this paper, the SimVP-CFLL-ML model was proposed and used for spatio-temporal PM2.5 concentration prediction via an image-like approach. This model based on a video prediction technique utilized SimVP as the backbone model, which included CFLL and ML to learn feature correlation and emphasize specific location predictions, respectively. To the best of our knowledge, this is the first model that applies a video prediction technique to predict PM2.5 concentration via an image-like approach at country-wide level. The features used in this work focus on fire hotspots and wind, playing an important role in PM2.5 concentration patterns. Results demonstrate that the proposed model can outperform all baselines especially when predicting high PM2.5 concentration with 3.51% improvement from the best baseline F1 score. In the features ablation study, results show how important feature correlation is and illustrate that the proposed model can learn the correlation. Although our findings represent a new approach for predicting spatio-temporal PM2.5 concentration, further experiments and screening are needed to validate the effectiveness of the proposed method. In time, with the establishment of more PM2.5 monitoring stations, the image-like data can represent PM2.5 with greater precision, thereby enhancing the proposed model's performance. Finally, the factor of fire, which is the burning of organic materials, is a widespread contributor to PM2.5 pollution worldwide, in both developing and developed countries. We hope that our model can be applied to such scenarios.

TABLE 8. Supplementary experiment on the impact of features on the proposed model. Boldfaces refer to the winners.

Features	MAE ↓	RMSE ↓	Macro F1 ↑
(A) Our Work's Features			
PM2.5	4.46	7.64	0.6290
PM2.5 + Wind	4.49	7.67	0.6294
PM2.5 + Fire	4.49	7.63	0.6373
PM2.5 + All Our Work's Features	4.51	7.59	0.6557
(B) Additional Features			
PM2.5 + Temperature	4.48	7.66	0.6278
PM2.5 + Humidity	4.58	7.76	0.6305
PM2.5 + Rainfall	**4.38**	7.48	0.6397
PM2.5 + All Additional Features	4.46	7.57	0.6427
(C) All Features			
PM2.5 + All Features	4.46	**7.47**	**0.6620**

APPENDIX A SUPPLEMENTARY EXPERIMENT

In this appendix, we present an additional experiment that investigates the relationship between additional factors and the prediction of PM2.5 concentrations, using the SimVP-CFLL-ML model. The experiment aims to provide insights into feature selection for future approaches. We specifically focus on the impact of temperature, humidity, and rainfall on PM2.5 concentration. To analyze the influence of these factors, we utilized ERA5 with regards to 2m temperature data, 2m dewpoint temperature data (humidity), and total precipitation data (rainfall). The model's performance was evaluated using metrics such as mean absolute error (MAE), root mean square error (RMSE), and Macro F1.

Table 8 presents a summary of the results obtained from the experiment. When all features were utilized, the SimVP-CFLL-ML model achieved the best RMSE: 7.47 $\mu g/m^3$ and Macro F1 score of 0.6620. However, when considering the best MAE performance, the model only included rainfall as a feature, resulting in a value of 4.38 $\mu g/m^3$. These findings indicate that our model is able to

benefit from the incorporation of additional features, leading to further improvement. Comparing the performance of all features used in our work with the inclusion of additional features in the model, the results suggest that the features employed in our work exhibit superior performance in classifying PM2.5 concentrations. However, there is a slight decrease in MAE and RMSE performance, which is not significant. This can be attributed to the fact that in Thailand, the PM2.5 concentration is generally low throughout most of the year. When additional features are added, the model aims to better predict high PM2.5 concentrations. In conclusion, the features used in our work prove to be quite adequate for predicting PM2.5. When more features are included, the model's performance significantly improves.

ACKNOWLEDGMENT

The authors would like to thank the Sensor for All and Thailand Pollution Control Department for providing the PM2.5 data source used in this work.

REFERENCES

[1] P. Punsompong, S. K. Pani, S.-H. Wang, and T. T. B. Pham, "Assessment of biomass-burning types and transport over Thailand and the associated health risks," *Atmos. Environ.*, vol. 247, Feb. 2021, Art. no. 118176, doi: 10.1016/j.atmosenv.2020.118176.

[2] B. Zhu, J. Su, H. Kang, and Y. Cai, "The impact of crop residue burning on air quality over Yangtze River delta, China: Observation and simulation," in *Proc. IEEE Int. Geosci. Remote Sens. Symp.*, Jul. 2012, pp. 3642–3645, doi: 10.1109/IGARSS.2012.6350627.

[3] Y. Zhuang, D. Chen, R. Li, Z. Chen, J. Cai, B. He, B. Gao, N. Cheng, and Y. Huang, "Understanding the influence of crop residue burning on $PM_{2.5}$ and PM_{10} concentrations in China from 2013 to 2017 using MODIS data," *Int. J. Environ. Res. Public Health*, vol. 15, no. 7, p. 1504, Jul. 2018, doi: 10.3390/ijerph15071504.

[4] R. Lan, S. D. Eastham, T. Liu, L. K. Norford, and S. R. H. Barrett, "Air quality impacts of crop residue burning in India and mitigation alternatives," *Nature Commun.*, vol. 13, no. 1, p. 6537, Nov. 2022, doi: 10.1038/s41467-022-34093-z.

[5] P. Saxena, S. Sonwani, A. Srivastava, M. Jain, A. Srivastava, A. Bharti, D. Rangra, N. Mongia, S. Tejan, and S. Bhardwaj, "Impact of crop residue burning in Haryana on the air quality of Delhi, India," *Heliyon*, vol. 7, no. 5, May 2021, Art. no. e06973, doi: 10.1016/j.heliyon.2021.e06973.

[6] A. P. K. Tai, L. J. Mickley, and D. J. Jacob, "Correlations between fine particulate matter ($PM_{2.5}$) and meteorological variables in the United States: Implications for the sensitivity of $PM_{2.5}$ to climate change," *Atmos. Environ.*, vol. 44, no. 32, pp. 3976–3984, Oct. 2010, doi: 10.1016/j.atmosenv.2010.06.060.

[7] C.-H. Hsu and F.-Y. Cheng, "Synoptic weather patterns and associated air pollution in Taiwan," *Aerosol Air Qual. Res.*, vol. 19, no. 5, pp. 1139–1151, 2019, doi: 10.4209/aaqr.2018.09.0348.

[8] I.-S. Park, M.-S. Park, Y. W. Jang, H.-K. Kim, C.-K. Song, J. S. Owen, S.-H. Kim, C.-R. Cho, and C.-H. Kim, "Impact comparison of synoptic meteorology and nationwide/local emissions on the Seoul Metropolitan Area during high PM multi-event and non-event days," *Asian J. Atmos. Environ.*, vol. 14, no. 3, pp. 263–279, Sep. 2020, doi: 10.5572/ajae.2020.14.3.263.

[9] I.-S. Park, S.-H. Kim, Y. W. Jang, M.-S. Park, J. Lee, J. S. Owen, C.-R. Cho, J.-B. Jee, J.-H. Chae, and M.-S. Kang, "Meteorological characteristics during periods of greatly reduced $PM_{2.5}$ concentrations in March 2020 in Seoul," *Aerosol Air Qual. Res.*, vol. 21, no. 9, 2021, Art. no. 200512, doi: 10.4209/aaqr.200512.

[10] M. Sukitpaneenit and N. T. K. Oanh, "Satellite monitoring for carbon monoxide and particulate matter during forest fire episodes in Northern Thailand," *Environ. Monitor. Assessment*, vol. 186, no. 4, pp. 2495–2504, Apr. 2014, doi: 10.1007/s10661-013-3556-x.

[11] P. Outapa, "The effect of seasonal variation and meteorological data on PM_{10} concentrations in Northern Thailand," *Int. J. Geomate*, vol. 16, no. 56, pp. 46–53, Apr. 2019. [Online]. Available: https://geomatejournal.com/geomate/article/

[12] J. Kayee, P. Sompongchaiyakul, N. Sanwlani, S. Bureekul, X. Wang, and R. Das, "Metal concentrations and source apportionment of $PM_{2.5}$ in Chiang Rai and Bangkok, Thailand during a biomass burning season," *ACS Earth Space Chem.*, vol. 4, no. 7, pp. 1213–1226, Jul. 2020, doi: 10.1021/acsearthspacechem.0c00140.

[13] W. Suriya, P. Chunpang, and T. Laosuwan, "Patterns of relationship between PM_{10} from air monitoring quality station and AOT data from MODIS sensor onboard of Terra satellite," *Sci. Rev. Eng. Environ. Stud. (SREES)*, vol. 30, no. 2, pp. 236–249, Mar. 2021, doi: 10.22630/PNIKS.2021.30.2.20.

[14] W. Wang and Y. Guo, "Air pollution $PM_{2.5}$ data analysis in Los Angeles long beach with seasonal ARIMA model," in *Proc. Int. Conf. Energy Environ. Technol.*, Oct. 2009, pp. 7–10, doi: 10.1109/ICEET.2009.468.

[15] L. Zhang, J. Lin, R. Qiu, X. Hu, H. Zhang, Q. Chen, H. Tan, D. Lin, and J. Wang, "Trend analysis and forecast of $PM_{2.5}$ in Fuzhou, China using the ARIMA model," *Ecol. Indicators*, vol. 95, pp. 702–710, Dec. 2018, doi: 10.1016/j.ecolind.2018.08.032.

[16] X. Yan and X. Enhua, "ARIMA and multiple regression additive models for $PM_{2.5}$ based on linear interpolation," in *Proc. Int. Conf. Big Data Artif. Intell. Softw. Eng. (ICBASE)*, Oct./Nov. 2020, pp. 266–269, doi: 10.1109/ICBASE51474.2020.00062.

[17] A. S. Sánchez, P. J. G. Nieto, P. R. Fernández, J. J. del Coz Díaz, and F. J. Iglesias-Rodríguez, "Application of an SVM-based regression model to the air quality study at local scale in the Avilés urban area (Spain)," *Math. Comput. Model.*, vol. 54, nos. 5–6, pp. 1453–1466, Sep. 2011, doi: 10.1016/j.mcm.2011.04.017.

[18] C.-M. Vong, W.-F. Ip, P.-K. Wong, and J.-Y. Yang, "Short-term prediction of air pollution in Macau using support vector machines," *J. Control Sci. Eng.*, vol. 2012, Jun. 2012, Art. no. 518032, doi: 10.1155/2012/518032.

[19] W. Yang, M. Deng, F. Xu, and H. Wang, "Prediction of hourly $PM_{2.5}$ using a space-time support vector regression model," *Atmos. Environ.*, vol. 181, pp. 12–19, May 2018, doi: 10.1016/j.atmosenv.2018.03.015.

[20] K. P. Singh, S. Gupta, A. Kumar, and S. P. Shukla, "Linear and nonlinear modeling approaches for urban air quality prediction," *Sci. Total Environ.*, vol. 426, pp. 55–244, Jun. 2012, doi: 10.1016/j.scitotenv.2012.03.076.

[21] Y. Tsai, Y. Zeng, and Y. Chang, "Air pollution forecasting using RNN with LSTM," in *Proc. IEEE 16th Int. Conf. Dependable, Autonomic Secure Comput., 16th Int. Conf. Pervasive Intell. Comput., 4th Int. Conf. Big Data Intell. Comput. Cyber Sci. Technol. Congr. (DASC/PiCom/DataCom/CyberSciTech)*, Aug. 2018, pp. 1074–1079, doi: 10.1109/DASC/PiCom/DataCom/CyberSciTec.2018.00178.

[22] S. Hochreiter and J. Schmidhuber, "Long short-term memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, Nov. 1997, doi: 10.1162/neco.1997.9.8.1735.

[23] V. Athira, P. Geetha, R. Vinayakumar, and K. P. Soman, "DeepAirNet: Applying recurrent networks for air quality prediction," *Proc. Comput. Sci.*, vol. 132, pp. 1394–1403, Jan. 2018, doi: 10.1016/j.procs.2018.05.068.

[24] K. Cho, B. Merrienboer, C. Gulcehre, F. Bougares, H. Schwenk, and Y. Bengio, "Learning phrase representations using RNN encoder–decoder for statistical machine translation," in *Proc. Conf. Empirical Methods Natural Lang. Process. (EMNLP)*, Mar. 2014, pp. 1724–1734, doi: 10.3115/v1/D14-1179.

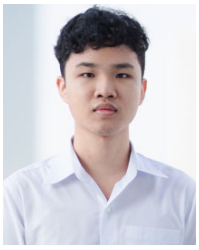
[25] B. Liu, S. Yan, J. Li, G. Qu, Y. Li, J. Lang, and R. Gu, "A sequence-to-sequence air quality predictor based on the n-step recurrent prediction," *IEEE Access*, vol. 7, pp. 43331–43345, 2019, doi: 10.1109/ACCESS.2019.2908081.

[26] M. H. Nguyen, P. Le Nguyen, K. Nguyen, V. A. Le, T. Nguyen, and Y. Ji, "$PM_{2.5}$ prediction using genetic algorithm-based feature selection and encoder–decoder model," *IEEE Access*, vol. 9, pp. 57338–57350, 2021, doi: 10.1109/ACCESS.2021.3072280.

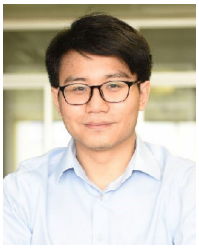
[27] W. Cheng, Y. Shen, Y. Zhu, and L. Huang, "A neural attention model for urban air quality inference: Learning the weights of monitoring stations," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2018, vol. 32, no. 1, pp. 2151–2158, doi: 10.1609/aaai.v32i1.11871.

[28] B. Liu, S. Yan, J. Li, Y. Li, J. Lang, and G. Qu, "A spatiotemporal recurrent neural network for prediction of atmospheric $PM_{2.5}$: A case study of Beijing," *IEEE Trans. Computat. Social Syst.*, vol. 8, no. 3, pp. 578–588, Jun. 2021, doi: 10.1109/TCSS.2021.3056410.

[29] X. Xu and M. Yoneda, "Multitask air-quality prediction based on LSTM-autoencoder model," *IEEE Trans. Cybern.*, vol. 51, no. 5, pp. 2577–2586, May 2021, doi: 10.1109/TCYB.2019.2945999.
[30] G. Shi, Y. Leung, J. S. Zhang, T. Fung, F. Du, and Y. Zhou, "A novel method for identifying hotspots and forecasting air quality through an adaptive utilization of spatio-temporal information of multiple factors," *Sci. Total Environ.*, vol. 759, Mar. 2021, Art. no. 143513, doi: 10.1016/j.scitotenv.2020.143513.
[31] C. Wen, S. Liu, X. Yao, L. Peng, X. Li, Y. Hu, and T. Chi, "A novel spatiotemporal convolutional long short-term neural network for air pollution prediction," *Sci. Total Environ.*, vol. 654, pp. 1091–1099, Mar. 2019, doi: 10.1016/j.scitotenv.2018.11.086.
[32] Q. Zhang, Y. Han, V. O. K. Li, and J. C. K. Lam, "Deep-AIR: A hybrid CNN-LSTM framework for fine-grained air pollution estimation and forecast in metropolitan cities," *IEEE Access*, vol. 10, pp. 55818–55841, 2022, doi: 10.1109/ACCESS.2022.3174853.
[33] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," presented at the 28th Int. Conf. Neural Inf. Process. Syst., Montreal, QC, Canada, 2015, vol. 1.
[34] Y. Wang, M. Long, J. Wang, Z. Gao, and P. S. Yu, "PredRNN: recurrent neural networks for predictive learning using spatiotemporal LSTMs," presented at the 31st Int. Conf. Neural Inf. Process. Syst., Long Beach, CA, USA, 2017.
[35] Y. Wang, J. Zhang, H. Zhu, M. Long, J. Wang, and P. S. Yu, "Memory in memory: A predictive neural network for learning higher-order non-stationarity from spatiotemporal dynamics," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 9146–9154, doi: 10.1109/CVPR.2019.00937.
[36] Z. Lin, M. Li, Z. Zheng, Y. Cheng, and C. Yuan, "Self-attention ConvLSTM for spatiotemporal prediction," in *Proc. AAAI Conf. Artif. Intell.*, Apr. 2020, vol. 34, no. 7, pp. 11531–11538, doi: 10.1609/aaai.v34i07.6819.
[37] Z. Chang, X. Zhang, S. Wang, S. Ma, Y. Ye, X. Xinguang, and W. Gao, "MAU: A motion-aware unit for video prediction and beyond," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 34, 2021, pp. 26950–26962.
[38] Z. Gao, C. Tan, L. Wu, and S. Z. Li, "SimVP: Simpler yet better video prediction," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, New Orleans, LA, USA, Jun. 2022, pp. 3160–3170, doi: 10.1109/CVPR52688.2022.00317.
[39] H. Hersbach, B. Bell, P. Berrisford, G. Biavati, A. Hornyi, J. M. Sabater, J. Nicolas, C. Peubey, R. Radu, I. Rozum, D. Schepers, A. Simmons, C. Soci, D. Dee, and J.-N. Thpaut, "ERA5 hourly data on single levels from 1959 to present," Copernicus, Europe, Tech. Rep., 2023, doi: 10.24381/cds.adbb2d47.
[40] *NRT VIIRS 375 m Active Fire Product VJ114IMGTDL_NRT Distributed From NASA FIRMS*, NASA's Fire Inf. Resource Manag. Syst. (FIRMS), USA, 2023. [Online]. Available: https://earthdata.nasa.gov/firms, doi: 10.5067/FIRMS/VIIRS/VJ114IMGT_NRT.002.
[41] *NRT VIIRS 375 m Active Fire Product VNP14IMGT Distributed From NASA FIRMS*, NASA's Fire Inf. Resource Manag. Syst. (FIRMS), USA, 2023. [Online]. Available: https://earthdata.nasa.gov/firms, doi: 10.5067/FIRMS/VIIRS/VNP14IMGT_NRT.002.
[42] *MODIS Collection 61 NRT Hotspot/Active Fire Detections MCD14DL Distributed From NASA FIRMS*, NASA's Fire Inf. Resource Manag. Syst. (FIRMS), USA, 2023. [Online]. Available: https://earthdata.nasa.gov/firms, doi: 10.5067/FIRMS/MODIS/MCD14DL.NRT.0061.



NATCH SIRISUMPUN received the B.Sc. degree in computer engineering from the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand, in 2020, where he is currently pursuing the M.Sc. degree in computer engineering. His main research interests include machine learning, deep learning, and video prediction.



KRITCHART WONGWAILIKHIT received the dual Ph.D. degree from the Department of Environmental Engineering, Chulalongkorn University, Thailand, and in chemical engineering from INSA Toulouse, France. With a research focus on chemical and environmental engineering, he has extensive experience in various areas, such as chemical process simulation, computational fluid dynamics, carbon capture and utilization, water treatment, and air pollution. His academic pursuits and research activities have provided him with an understanding of the complex interplay between chemical and environmental engineering.

PISUT PAINMANAKUL received the Ph.D. degree from the Department of Environmental Engineering, INSA Toulouse, France, in 2005. He is currently a Professor with the Department of Environmental Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. His research interests include air pollution treatment technology, physico-chemical treatment process, aeration, oxygenation process in water, wine-making process, water, wastewater collection and treatment, environmental management strategy, and innovation.

PEERAPON VATEEKUL (Member, IEEE) received the Ph.D. degree from the Department of Electrical and Computer Engineering, University of Miami (UM), Coral Gables, FL, USA, in 2012. He is currently an Associate Professor with the Department of Computer Engineering, Faculty of Engineering, Chulalongkorn University, Thailand. His research interests include machine learning, data mining, deep learning, text mining, big data analytics, hierarchical multi-label classification, natural language processing, data quality management, and applied deep learning and reinforcement learning techniques in various domains, such as, healthcare, geoinformatics, hydrometeorology, and energy trading.