

Vision-based Predictive Model on Particulates via Deep Learning

SungHwan Kim[†] and Songi Kim^{*}

Abstract – Over recent years, high-concentration of particulate matters (e.g., a.k.a. fine dust) in South Korea has increasingly evoked considerable concerns about public health. It is intractable to track and report PM₁₀ measurements to the public on a real-time basis. Even worse, such records merely amount to averaged particulate concentration at particular regions. Under this circumstance, people are prone to being at risk at rapidly dispersing air pollution. To address this challenge, we attempt to build a predictive model via deep learning to the concentration of particulates (PM₁₀). The proposed method learns a binary decision rule on the basis of video sequences to predict whether the level of particulates (PM₁₀) in real time is harmful ($>80\mu\text{g}/\text{m}^3$) or not. To our best knowledge, no vision-based PM₁₀ measurement method has been proposed in atmosphere research. In experimental studies, the proposed model is found to outperform other existing algorithms in virtue of convolutional deep learning networks. In this regard, we suppose this vision based-predictive model has lucrative potentials to handle with upcoming challenges related to particulate measurement.

Keywords: Particulate matters, Convolutional neural networks, Transfer learning.

1. Introduction

Particulate matters (PM; a.k.a. atmospheric particulate matter or particulates) are composed of both natural and human-made (i.e., anthropogenic) substances including microscopic solid or liquid matter suspended in atmosphere. Generally they are known to make change to climate and precipitation, and give rise to harmful impacts on human health. Of late, confronting threats of particular matters become no longer negligible as the media and research groups increasingly determine their deleterious effects [1-4, 42]. For instance, a recent study in 2017 reveals that population in old and youth groups are more likely to undergone critical respiratory symptoms [5], and patients of respiratory disorder increasingly rise in number due to particular matters [6]. Under this circumstance, it is imperative to build up a strategic monitoring system and real-time based alarming framework to timely tackle with the effect of particulates. And yet, it is intractable to improve the current particulate reporting system (i.e., county-based whose coverage of particulates remains insufficient to meet civilians' demand.

Korea has more fine-dust pollution than almost any other country in the organization for economic cooperation and development (OECD). The report from the organization, as of 2015, says Korea had an average annual of 32 micrograms per cubic meter: the highest among the 35 OECD member countries. Since 2006, the department of environment in Korea has set the limit, as the warning level, on particulates at $80\mu\text{g}/\text{m}^3$ on the basis of daily average

(<http://www.airkorea.or.kr/>). The fact that a great deal of governmental policies related to particulates have been issued suggests that particulates come into our critical concerns to public health, which appears to be quite immediate in various aspects [7-8].

In light of vision research, atmospheric haze (e.g., mist, cloud, smoke or fog) should be properly removed as blurring out the clarity of the atmosphere. Typically haze undermines the light penetration of dense atmosphere that interferes with detecting distant subjects. Thus, effective controlling for the light effects to remove haze is desired in photography and vision-based applications [9-11]. Cai et al. [12] introduces DehazeNet, an outstanding algorithm to separate unexpected haze to extract clear-cut vision. Related to this, a range of vision-based smoke detection algorithm have been introduced [13]. For example, Toreyin et al. [14] proposed the smoke detection algorithm that builds on temporal variation of wavelet domain. This algorithm utilizes edge and colour to determine the features of smoke. Chen et al. [15] proposed a method for smoke detection algorithm using the total number of smoke pixel and the accumulation of motions that separates smoke and non-smoke moving objects. In short, exploiting temporal variation serves as integral role in learning a vision-based predictive model, and there is much room for extension to a predictive model of particulates that builds on vision-based algorithms such that color, motion, spatial difference, disorder, and image training are mainly taken into account.

When it comes to deep learning, it is well-known that the convolutional neural networks (CNNs) is an efficient image processing algorithm adapted for vision analysis and image recognition, and has proved its superior performance in many applications ([17-19]). The fortes of CNNs are largely two-fold: (1) CNNs locally link to the convolutional

[†] Corresponding Author: Department of Statistics, Keimyung University, Korea. (shkim1213@kmu.ac.kr)

^{*} Department of Statistics, Keimyung University, Korea.
Received: September 19, 2017; Accepted: May 23, 2018

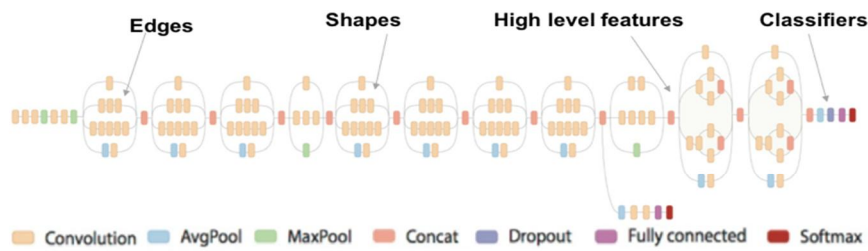


Fig. 1. The model architecture of Inception-v3

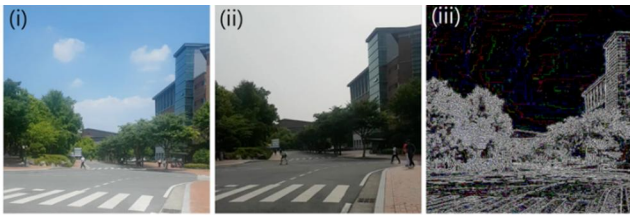


Fig. 2. Image comparisons between safe and harmful atmospheric conditions related to PM_{10} : (i) safe ($80 \mu g/m^3$ below) and (ii) harmful ($80 \mu g/m^3$ above) particulate levels (iii) an example of residual differences derived from two consecutive frames in a video

layers rather than full connections, which indicates that output neurons in the network structure are subject to local input neurons. This allows the locally-connected convolutional layers to readily solve spatially-correlated problems [20,21] (2) CNNs generate new layers (a.k.a. pooling layers) that consists of imagery features relative to receptive region and thereby narrows down the number of parameters to be estimated. For these reasons, the compressed layers, namely pooling layers, in CNNs enable to address large-scale problems [22]. Therefore, in case of prediction models that learn high-dimensional vision data, chances are that CNNs are superior to extant predictive models.

To date, measuring particulates mostly relies on the sensor based approaches. The sensor-based measurement has an unparalleled advantage of accuracy, low cost and handy configurations. However, the capturing coverage is hardly to exceed its vicinity region and is not suited to accommodate the degree of particulates' movement. To the contrary, a vision-based approach is capable of scanning wide areas in that the video device of any sort collects spatial information (e.g., accumulated dust in the air) but not limited to picturing terrestrial objects. Suppose that we obtain two video clips. In Fig. 2 (i - ii), two images are drawn the two video sequences recoded at the same spot in a campus region, distinguishable to the eye (i.e., between bright and hazy). Given two images, we can postulate that there are image signals that characterize binary levels sufficient to build a binary predictive model. And yet we should remember that a cloudy sky does not necessarily indicates high-particulate level, and then we should address this challenge and come up with how to adjust varying

atmospheric to uncover the potential features based on particulates.

Combined together, this paper introduces an image-based predictive model of particulate matters, and applies the deep learning architecture of a convolutional neural networks to features extracted from video sequences. To our best knowledge, no vision-based method has been proposed to directly predict particulate matters. Interestingly, the previous works [23-25] consume a great deal of effort to addressing the particulate problems. Yet, most of the existing models are just involved in temporal PM_{10} prediction [26], predominantly designed to forecast its concentration the time ahead, not measuring the level of concentration at present [27]. In theory, this temporal model is largely linked to time dependent methods. Narrowing down the scope of functions, we particularly highlight on a binary prediction model whether PM_{10} is greater than $80 \mu g/m^3$ or not. Importantly, since the convolutional neural network is shown to achieve many successes in diverse research fields (e.g., bio-medicine, vision recognition, robotics [28-32], CNNs are the sensible choice to develop an outstanding predictive model.

The major contributions to this paper include as below: (1) the temporal variation in video sequences and spatial dependencies are considered in the model that applies deep learning architecture of CNNs (2) the binary prediction model of PM_{10} derived from CNNs outperforms, with a high accuracy, other existing models. (3) the proposed predictive model on PM_{10} is widely applicable to temporal monitoring of particulate levels and can result in wide use in the form of electrical gadgets.

The paper is outlined as follows. In Chapter 2, we discuss the previous works and proposed method that leverages the temporal domain of video sequences together with CNNs. In Chapter 3, in an effort to verifying performance, we carry out diverse experiments (e.g., street regions, construction site and forest areas) to test if the proposed algorithm effectively implements the risk prediction. Finally, we discuss future works and leave concluding remarks in Chapter 4.

2. Datasets and Related Works

2.1 Datasets

Below we demonstrate how to collect particulate data

that learn a predictive model. Daegu is traditionally known as one of major cities in South Korea, where a large-scale industrial complex settles in together with high population density. Due to these social and environmental circumstances, gas emission has been a years-long environmental challenge, and high-concentration dusts constantly blew and cover all over the year. This is the reason behind why we choose Daegu as suitable for collecting air pollution information. To measure particulates, we utilize a high-performance device (Aerosol Mass Monitor (AEROCET-831) manufactured by Met One Instruments; <http://metone.com/>), whose capability on measurable particulate size ranges from $PM_{2.5}$ to PM_{10} . This device is particularly featured with high accuracy and robustness as compared with extant others. In this study, we focus solely on the level of PM_{10} for simplicity. The interested regions largely include three categories: (1) street regions by exhausts emission (2) vicinity of construction sites (3) residential areas. In brief, video sequences are recorded with 25 frames per second in 98 street regions, 60 areas nearby construction sites and 60 forest regions for a total of 218 sequences. In an effort to balancing between atmospheric and territory influence, a majority of sequences are recoded in a way that approximately half of screen is occupied by sky and the second half by grounds. Besides, we also take into account residential areas, a group of trees and building complexes featured with only non-atmospheric information (i.e., absence of sky) in order to enhance applicability and practicability of the model. Data are collected three to four times a day over the spring and summer seasons in 2017. To take a glimpse, thumb nails of each video sequence are presented in Table 2. The video sequences are taken by Samsung phone cameras, the most popularly used cellular phones, with the best resolution.

2.2 Convolutional neural network architectures

In the field of image recognition, the use of convolutional layers in these deep networks [33] has increasingly prevalent, implicating much of its computational and statistical efficiency to a range of applications. Surprisingly, methods employing convolutional neural networks (CNNs) swept major image recognition challenges (e.g., the ILSVRC challenge; <http://www.image-net.org/>) and have recently also been extensively applied to medical image analysis [34]. Convolutional filters in CNNs emulate the pattern recognition of the visual cortex, where so-called receptive fields activate subject to distinct spatial patterns. Simply put, the CNN takes the transformed input image, repetitive convolution and pooling and plug-in derived nodes to fully connected networks. The parameters of the entire network are estimated by forward propagating the input through the network. Inspired by CNNs, Google invests non-trivial efforts to developing high-performing networks, whose architecture builds on CNN, a pre-trained

network on the data [37], in spirit of transfer learning [40]. To this end, Singh et al. [38] proposed the idea of transferring from the image classification adopts a pre-trained model from ImageNet [35] or Coco-DB [36]. It can be understood as a CNN variant aiming to improve computational efficiency. ImageNet is a popularly applied academic data set in machine learning to learn image recognition system. Subsequent to Inception-v1 [39], the algorithm advances towards Inception-v3, currently most widely applied one. Major advantages of Inception-v3 are largely two-fold: (1) the fact that Inception-v3 achieves 21.2% top-1 and 5.6% top-5 error for single frame evaluation suggests its outstanding prediction accuracy on the ILSVRC-2012 classification benchmark (2) affordable computational resources to process the input (e.g., 5 billion multiply-adds per inference with using less than 25 million parameters). More precisely, the form of inductive transfer with an Inception-v3 architecture model displays the consecutive tasks of convolutional layers followed by a pooling layer, convolutional layers, Inception blocks, a dropout layer and a fully connected layer, each of which is learned by training the network on the data. In short, Fig. 1 sketches as an example how TensorBoard deploys the Inception v3 architecture [37]. It cannot be emphasized enough that, unlike a conventional CNN, this inception module-oriented CNN enjoys remarkable saving of computational cost and outstandingly high accuracy. Inception structures are illustrated as in Fig. 3 (a)-(c). Precisely, Inception modules can reduce the dimension of the input representation via 1×1 filters in the beginning stage. The rationale behind this technique is that the strong correlation between neighboring pixels protects loss of information and this dimension reduction even facilitates faster learning. Moreover, one can replace any $n \times n$ convolution by a $1 \times n$ convolution followed by a $n \times 1$ convolution as well as increasing activations per tile. These schemes also promote to learn models faster in effect due to disentangled features.

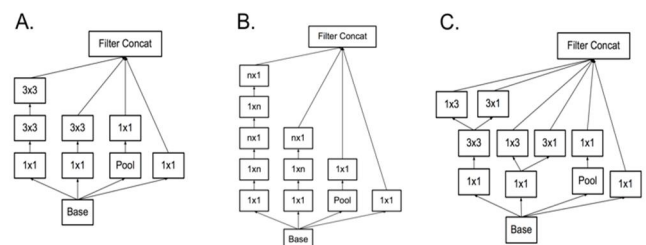


Fig. 3. Three instances of Inception modules proposed by Szegedy et al.[1]. These modules allow to cut down the dimension of the input representation via 1×1 filters. One can replace any $n \times n$ convolution by a $1 \times n$ convolution followed by a $n \times 1$ convolution as well as increasing activations per tile. These techniques promote to learn models faster in effect and produce disentangled features

3. Proposed Method

3.1 Strategic workflows for PM₁₀ predictive model

In this chapter, we introduce a novel predictive model on particulates, named “Deep Haze”, on the basis of temporal video sequences in RGB domain. In what follows, we delineate in order how the predictive model is implemented.

3.1.1 Pre-processing for video sequences

(Step 1) Preparation of image frames: At the initial stage, we layout K temporal sequences of 1080×1920 pixels from a single video for N seconds (i.e., $K = 25$ frames per second on average). Considering three RGB values, a single image frame includes, in turn, a 1080×1920×3 data matrix. Suppose we use video sequences of two seconds, and this produces 50 image frames in total.

(Step 2) RGB residual computation: When shooting image, haze (i.e., particulates) has a tendency to leading to non-smooth patterns, and hence it is vital to eliminate blurring effects at the outset. Inspired by this phenomenon, we may define a benchmark subject to intensity residuals of RGBs between raw and smoothed video sequences:

$$Res^{(k)} = (R^{(k)}, G^{(k)}, B^{(k)}) \\ = (R_o^{(k)} - R_s^{(k)}, G_o^{(k)} - G_s^{(k)}, B_o^{(k)} - B_s^{(k)}),$$

where $R_o^{(k)}$, $G_o^{(k)}$ and $B_o^{(k)}$ are original RGB intensity values and $R_s^{(k)}$, $G_s^{(k)}$ and $B_s^{(k)}$ smoothed RGB values at a pixel, ranging from 0 to 255 for $k = 1, \dots, K$. In regard to smoothing, we apply the Gaussian blurring technique known as one of the most popularly used algorithm for image smoothing. Below the kernel smoothing estimator is denoted as

$$\hat{f}(x, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{x - X_i}{h}\right),$$

where X_i is an observed pixel ($1 \leq i \leq n$), K the Gaussian kernel function and h an arbitrary bandwidth ($h = 3$ in this paper). Consequently, this residual $Res^{(k)}$ still contains 1080×1920 pixels and this seemingly unevenness measures can be a source of haze. In this regard, this residual is expected to capture sheer effects of haze in the sense that particulates are prone to sticking out of neighboring pixels.

(Step 3) Derivation of residual differences: In this stage, we derive residual differences of RGB between two consecutive sequences:

$$diff^{(k)} = Res^{(k+1)} - Res^{(k)} \\ = (R^{(k+1)} - R^{(k)}, G^{(k+1)} - G^{(k)}, B^{(k+1)} - B^{(k)}),$$

for $k = 1, \dots, K - 1$. In theory, the residual differences

account for the degree of haze effect’s variation at a pixel level overtime. Haze effects are typically quite unpredictable and mobile atmospheric conditions (e.g. sunshine, wind, barometric pressure etc). In essence, $diff^{(k)}$ is designed to gauge the total estimation of dust mobility. Importantly, a light effect, a potential distorting factor, commonly interferes with precise image analysis. By means of residual differences, the adjustment of light effects can be made at ease. On account of the form of numeric formulation (i.e., $Res^{(k+1)} - Res^{(k)}$), the common amount of light effects automatically cancels out, so that we hardly need to normalize light (e.g., oversaturation), color, hue variations but rather can directly use $diff^{(k)}$ as a building block to learn a model. In this project, video sequence data are pre-processed via OpenCV (<http://opencv.org/>) implemented by Python, and an example of residual differences is illustrated in Fig. 2 (iii).

3.1.2 Integrative convolutional deep neural nets

(Step 4) Learning a model on CNNs: Using the pre-processed data (i.e., residual differences), we propose to fuse multiple CNNs (a.k.a. ensemble technique in machine learning) to predict, in a binary fashion, whether a particulate level is harmful or not ($>80 \mu\text{g}/\text{m}^3$). To obtain an individual CNN-based model, as a weak classifier, we apply the Inception-v3 architecture [37]. In theory, Inception-v3 is trained on ImageNet images, and learns on a new top layer that can distinguish other classes of images. To keep generality, we proceed with the typical process of Inception-v3 architecture as stated in Section 2.2 and Table 1. In the pre-processing, a transformed image (i.e., $diff^{(k)}$) is compressed to 2048-dimensional vector and the top layer takes this as input for each transformed image and softmax layer is added on top of this representation. When softmax layer contains N labels, this generally corresponds to learning $N + (2048 \times N)$ model parameters lined up with the learned biases and weights, respectively. In this project, our labels are limited in scope to two (i.e., safe and harmful) and so we fit a model with the total of 4,098 parameters. Resultingly, this architecture provides a two dimensional output for each input sequence (i.e., input size: $1 \times 1 \times 2$), to which we make a main change in the architecture. Each prediction output $f(diff^{(k)})$ can be interpreted as the class label (i.e., safe or harmful) for each frame (i.e., $diff^{(k)}$) of residual differences for $k = 1, \dots, K - 1$.




(Step 5) Integration of weak classifiers: Putting together, we assemble the $K - 1$ weak classifiers of Inception architecture via a major voting to construct an integrative classifier as follows:

$$Pred(f(diff^{(k)}); 1 \leq k \leq K - 1) \\ = \begin{cases} 1 & \text{if } \frac{\sum_{k=1}^{K-1} f(diff^{(k)})}{K - 1} > 0.5, \\ 0 & \text{otherwise} \end{cases}$$

Table 1. The proposed architecture of CNN models in spirit of Inception v3

type	patch size/stride	input size
conv	3×3/2	299×299×3
conv	3×3/1	149×149×32
conv padded	3×3/1	147×147×32
pool	3×3/2	147×147×64
conv	3×3/1	73×73×64
conv	3×3/2	71×71×80
conv	3×3/1	35×35×192
3×Inception	Fig. 3(a)	35×35×288
5×Inception	Fig. 3(b)	17×17×768
2×Inception	Fig. 3(c)	8×8×1280
pool	8×8	8×8×2048
pool	8×8	8×8×2048
linear	logits	1×1×2048
softmax	classifier	1×1×2

Table 2. Thumb nails of target regions where particulate matters are analyzed to construct the predictive model

Category	Thumb nails	# of video sequences
Street regions		98
Construction site		60
Forest regions		60

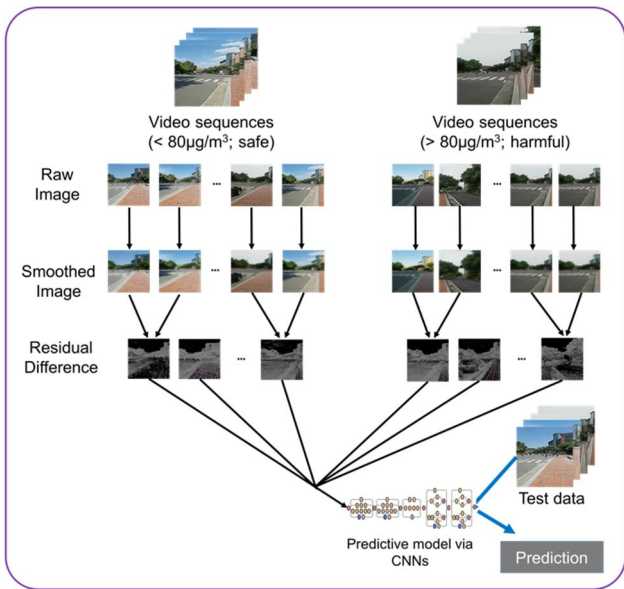


Fig. 4. The workflow of integrative CNNs-based model (Deep Haze) that learns on particulate matters

where $f(\text{diff}^{(k)}) \in \{0, 1\}$ (e.g., 0 denotes “safe” and 1 “harmful”) is a binary Inception model on $\text{diff}^{(k)}$ for $k = 1, \dots, K - 1$. Such integrated weak classifiers are typically found to be statistically superior to weak classifiers [41] in favor of low predictive signals (i.e., ambiguous separation). This is due to the fact that the algorithm, in principle, takes an advantage of data augmentation, say, exploiting multiple frames of difference residuals in the model, which advance single image-based CNNs. In Fig. 4. we outline the sequential flows to construct the integrative Inception architecture of the Deep Haze. For future interested readers, source codes are available online at author’s webpage (<https://sites.google.com/site/sunghwanshome/>).

In light of methodological standpoints, the fortes of the proposed model can be summarized to largely three points. (1) This method conceptually allows to predict particulate

levels only by means of image data. No wonder, this circumvents the traditional method exploiting machine gadgets. (2) The algorithm based on residual difference is robust to light and temperature variations even though particulates are found highly subject to varying weather conditions. (3) Predictive performance is supposed to be stable in image analysis on account of the use of CNN-based algorithms along with Inception schemes and ensemble techniques.

4. Experimental studies

In simulations, we assess some variants of the Deep Haze that utilize different frame number, and compare them to other popularly used classifiers (e.g., random forest (RF), SVM based on linear kernels (SVM-linear) and radial basis function (SVM-rbf) and weighted SVM (wSVM) [43]). Out of 218 video sequences, we used for training 78 video sequences (i.e., 39 safe and 39 harmful labels, respectively), and for testing 140 sequences (i.e., 101 safe and 39 harmful labels, respectively). A range of experimental scenarios are taken into account to evaluate our algorithm’s universal applicability. Table 3 and 4 encapsulates the predictive performance of the Deep Haze. It is evident to say that the proposed algorithm, when using all datasets, outstandingly distinguishes a harmful atmospheric condition with high accuracy (i.e., Deep Haze: 0.8286 ~ 0.8929, Random forest: 0.6500 ~

Table 3. Sensitivity (Sen), Specificity (Spe) and Youden index (=Sensitivity + Specificity - 1) of predictive models with an application to various video sequences

# of frames	5 frames			10 frames			20 frames			40 frames		
Street regions												
	Sen	Spe	Youden	Sen	Spe	Youden	Sen	Spe	Youden	Sen	Spe	Youden
Deep Haze	0.6667	0.8947	0.5614	0.7222	0.9474	0.6696	0.8333	0.9737	0.8070	0.8333	0.9737	0.8070
RF	0.0000	0.8684	-0.1316	0.0000	0.8684	-0.1316	0.0000	0.8684	-0.1316	0.0000	0.8684	-0.1316
SVM-rbf	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
SVM-linear	0.1667	0.6053	-0.2281	0.1111	0.6316	-0.2573	0.0556	0.6316	-0.3129	0.0556	0.6579	-0.2865
wSVM	0.0556	0.6316	-0.3129	0.1111	0.6579	-0.2310	0.1111	0.6579	-0.2310	0.0556	0.6316	-0.3129
Construction sites												
Deep Haze	0.7000	0.6429	0.3429	0.8000	0.7143	0.5143	0.8000	0.7857	0.5857	0.8000	0.7857	0.5857
RF	0.1000	1.0000	0.1000	0.0000	1.0000	0.0000	0.1000	0.9286	0.0286	0.1000	0.8571	-0.0429
SVM-rbf	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
SVM-linear	0.6000	0.4286	0.0286	0.3000	0.7143	0.0143	0.3000	0.7857	0.0857	0.3000	0.7143	0.0143
wSVM	0.6000	0.2857	-0.1143	0.3000	0.7143	0.0143	0.2000	0.7857	-0.0143	0.2000	0.7143	-0.0857
Forest												
Deep Haze	0.8182	0.9184	0.7365	0.8182	0.9388	0.7570	0.8182	0.9592	0.7774	0.8182	0.9184	0.7365
RF	0.2727	0.8163	0.0891	0.2727	0.8776	0.1503	0.2727	0.8776	0.1503	0.2727	0.8367	0.1095
SVM-rbf	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
SVM-linear	0.8182	0.5714	0.3896	0.8182	0.6531	0.4712	0.7273	0.6122	0.3395	0.8182	0.6327	0.4508
wSVM	0.8182	0.5714	0.3896	0.7273	0.6531	0.3803	0.7273	0.6327	0.3599	0.8182	0.6531	0.4712
All datasets												
Deep Haze	0.7179	0.8713	0.5892	0.7692	0.9109	0.6801	0.8205	0.9406	0.7611	0.8205	0.9208	0.7413
RF	0.1026	0.8614	-0.0360	0.1026	0.8812	-0.0162	0.1026	0.8911	-0.0063	0.1026	0.8515	-0.0460
SVM-rbf	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000	0.0000	1.0000	0.0000
SVM-linear	0.4615	0.5644	0.0259	0.3590	0.6535	0.0124	0.3077	0.6436	-0.0487	0.3333	0.6535	-0.0132
wSVM	0.5641	0.5248	0.0889	0.3333	0.6535	-0.0132	0.3333	0.6535	-0.0132	0.5641	0.5248	0.0889

Table 4. Prediction accuracy of predictive models with an application to various video sequences.

# of frames	5 frames	10 frames	20 frames	40 frames
Street regions				
Deep Haze	0.8214	0.8750	0.9286	0.9286
RF	0.5893	0.5893	0.6071	0.5893
SVM-rbf	0.6786	0.6786	0.6786	0.6786
SVM-linear	0.4643	0.4643	0.4464	0.4643
wSVM	0.5000	0.4643	0.4821	0.4464
Construction sites				
Deep Haze	0.6667	0.7500	0.7917	0.7917
RF	0.6250	0.5833	0.5833	0.5417
SVM-rbf	0.5833	0.5833	0.5833	0.5833
SVM-linear	0.5000	0.5417	0.5833	0.5417
wSVM	0.4167	0.5417	0.5417	0.5000
Forest				
Deep Haze	0.9000	0.9167	0.9333	0.9000
RF	0.7167	0.7667	0.7667	0.7333
SVM-rbf	0.8167	0.8167	0.8167	0.8167
SVM-linear	0.6167	0.6833	0.6333	0.6667
wSVM	0.6167	0.6667	0.6500	0.6833
All datasets				
Deep Haze	0.8286	0.8714	0.9071	0.8929
RF	0.6500	0.6643	0.6714	0.6429
SVM-rbf	0.7214	0.7214	0.7214	0.7214
SVM-linear	0.5357	0.5714	0.5500	0.5643
wSVM	0.5357	0.5643	0.5643	0.5571

0.6429, SVM-rbf: 0.7214, SVM-linear: 0.5357~0.5714 and wSVM: 0.5357~0.5643) in Table 3 and low false detection (i.e., Youden Index = Sensitivity + Specificity - 1,

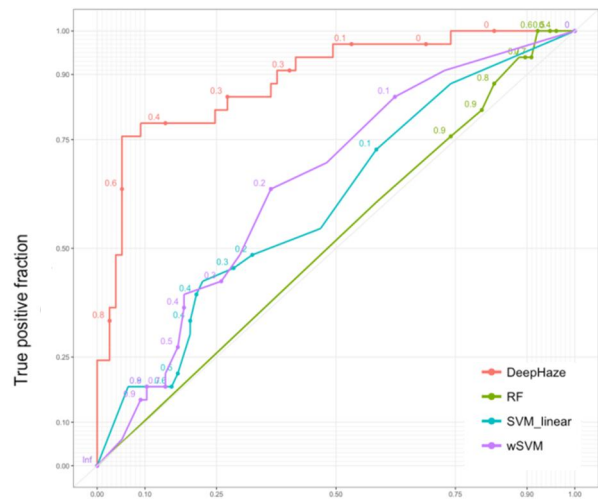


Fig. 5. ROC curves of the deep learning-based predictive model (DeepHaze), random forest (RF), SVM based on linear kernels (SVM-linear) and weighted SVM (wSVM) on two classes (i.e., safe and harmful)

Deep Haze: 0.5892~0.7611, Random forest: -0.0460~0.0063, SVM-rbf: 0, SVM-linear: -0.0487 ~ 0.0259 and wSVM: -0.0132 ~ 0.0889) in Table 4 and Fig. 5.

4.1 Street regions

In this experiment, the proposed model applies to data

captured from street areas influenced exhaust emission. Exhaust emission is thought of as one of major man-made sources of particulates, meaning that we are likely to observe relative high particulate levels. Table 3 and 4 shows the surpassing performance of the Deep Haze as compared to SVM and RFs. Note that the Deep Haze performs with high accuracy (0.8214–0.9286) as opposed to Random forest (0.5893), SVM-rbf (0.6786), SVM-linear (0.4643) and wSVM (0.5000). It is important to note that accuracy, as expected, tends to be increasing as the frames augmented from 5 to 40. Besides, the Deep Haze is superior to others with the low false detection (i.e., high Youden index; Deep Haze: 0.5614–0.8070, Random forest: -0.1316, SVM-rbf: 0, SVM-linear: -0.3129 ~ -0.2281 and wSVM: -0.3129 ~ -0.2310). Putting another way, the low Youden index values imply Random forest and SVM are not suitable to image-based particulate prediction.

4.2 Construction sites

It is commonplace that vicinities of construction sites accompany a bulk of vehicles, facilities and industrial factories and so on. In this sense, chances are that construction areas are more likely to be influential to artificial pollutions even if common atmospheric factors are excluded. In addition, all other experiments (e.g., street regions, construction sites and forest) presents consistent messages of low false detection in Table 3 (i.e., Deep Haze: 0.3429 ~ 0.5857, Random forest: -0.0429 ~ 0.1000, SVM-rbf: 0, SVM-linear: 0.0143 ~ 0.0857 and wSVM: -0.1143 ~ 0.0143). Interestingly, Random forest and SVM is found to be poor to detect harmful air sequences (i.e., low sensitivity) despite rather moderate accuracy.

4.3 Forest areas

We hypothesize whether our predictive models are effectively applied to data from forest areas. It is plausible that residential areas are featured with an intermediate level of particulates as typically located at distant from construction zones and massively congested roads. Coherent to experiments above, Table 3 and 4 shows the the proposes models are superior in accuracy to SVM and RFs (i.e., Deep Haze: 0.8286 ~ 0.8929, Random forest: 0.6500 ~ 0.6429, SVM-rbf: 0.7214, SVM-linear: 0.5357 ~ 0.5643 and wSVM: 0.5357~0.5571 and in low false detection (i.e., Deep Haze: 0.7365 ~ 0.7774, Random forest: 0.0891~0.1503, SVM-rbf: 0, SVM-linear: 0.3395~0.4712 and wSVM: 0.3599 ~ 0.4712).

5. Discussion and Conclusion

Of late, the concerns of high-concentration of particulate matters has been increasingly widespread across in vicinity of East Asia. In spite of low accessibility of particulate

measurements, there is still room that the vision-based predictive model can determine its quantity on the real time basis. With an emphasis of high applicability of vision data, we propose a predictive model that learns on video sequences to predict whether the level of particulates (PM₁₀) in real time is harmful (>80 $\mu\text{g}/\text{m}^3$) or not. To our best knowledge, no vision-based predictive model has been introduced in atmosphere research. In experiments, we confirm the Deep Haze and its variant methods outperform existing popular methods, and hence are shown to be trustworthy as a prediction model over particulate levels. Putting together, this outstanding performance suggests that our algorithm can potentially facilitate particulate monitoring down the road. Nonetheless, it is also found in our internal experiment that the proposed algorithm falls on false detection in case of low particulate signals especially in case of near-sighted scenes. Therefore, it is also interesting to see if haze removal algorithms [12] can handle this obstacle effectively. Importantly, this technique can be favorably extended to other areas, including the detection of chemical leaks, fire and smoke alarm and etc. For maximizing its utility, the Haze is planned embed into a portable go-to electronic gadget and will launch it with the fashion of phone application soon in pursuit of vision-based (or partly vision-guided) monitoring system on particulates. By extension, the model can be extended to multi-class prediction to elaborately categorize concentration of particulates. To this end, another Inception-type architecture can be possibly considered to improve accuracy. We leave these topics for future study.

Acknowledgements

This research is supported by the Keimyung University.

References

- [1] R. Bascom, P. Bromberg, D. Costa, R. Devlin, D. Dockery, and M. Frampton, et al, "Health effects of outdoor air pollution," *Am J Respir Crit Care Med*, vol. 153, no. 1, pp. 477-498, Jan. 1996.
- [2] B. Brunekreef, and S. T. Holgate, "Air pollution and health," *The Lancet*, vol. 360, pp. 1233-1242, Oct. 2002.
- [3] M. Kampa, and E. Castanas, "Human health effects of air pollution," *Environmental Pollution*, vol. 151, no. 2, pp. 362-367, Jan. 2008.
- [4] C. M. Wong, N. Vichit-Vadakan, H. Kan, and Z. Qian, "Public Health and Air Pollution in Asia (PAPA): a multicity study of short-term effects of air pollution on mortality," *Environmental Health Perspectives*, vol. 116, no. 9, pp. 1195-1202, sep. 2008.
- [5] X. Xia, A. Zhang, S. Liang, Q. Qi, L. Jiang, and Y. Ye, "The Association between Air Pollution and

- Population Health Risk for Respiratory Infection: A Case Study of Shenzhen, China,” *Int J Environ Res Public Health*, vol. 14, no.9, pp. 950, Aug. 2017.
- [6] H. J. Kim, M. G. Choi, M. K. Park, and Y. R. Seo, “Predictive and Prognostic Biomarkers of Respiratory Diseases due to Particulate Matter Exposure,” *J Cancer Prev*, vol. 22, no. 1, pp. 6-15 Mar. 2017.
- [7] H. C. Kim, E.H. Kim, C. H. Bae, J. H. Cho, B. U. Kim, and S. T. Kim, “Regional contributions to particulate matter concentration in the Seoul metropolitan area, South Korea: seasonal variation and sensitivity to meteorology and emissions inventory,” *Atoms. Chem. Phys.*, vol. 17, no. 17, pp. 10315-10332, Sep. 2017.
- [8] S. H. Hwang, J. Y. Lee, S. M. Yi, and H. Kim, “Associations of particulate matter and its components with emergency room visits for cardiovascular and respiratory diseases,” *PLoS ONE*, vol. 12, no. 8, Aug. 2017.
- [9] J. A. Stark, “Adaptive image contrast enhancement using generalizations of histogram equalization,” *IEEE Transactions on Image Processing*, vol. 9, no. 5, pp. 889-896, May. 2000.
- [10] R. Eschbach and B. W. Kolpatzik, “Image-dependent color saturation correction in a natural scene pictorial image,” US Patent 5,450,217, Sep. 1995.
- [11] Y. Y. Schechner, S. G. Narasimhan, and S. K. Nayar, “Instant dehazing of images using polarization,” *Conference on Computer Vision and Pattern Recognition (CVPR)*, vol. 1, pp. 325-332, June 2001.
- [12] B. Cai, X. Xu, K. Jia, C. Qing, and D. Tao, “Dehaze Net: An End-to-End System for Single Image Haze Removal,” *IEEE Transactions on Image Processing*, vol. 25, no. 11, pp. 5187-5198, Nov. 2016.
- [13] C. Y. Lee, C. T. Lin, C. T. Hong and M. T. Su, “Smoke Detection Using Spatial and Temporal Analyse,” *International Journal of Innovative computing, Information and Control*, vol. 8, no. 6, pp. 4749-4770, June. 2012.
- [14] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, “Contour based smoke detection in video using wavelets,” *European Signal Processing Conference*, pp. 123-128, Sep. 2006.
- [15] T. H. Chen, Y. H. Yin, S. F. Huang, and Y. T. Ye, “The smoke detection for early fire-alarming system based on video processing,” in *Intelligent Information Hiding and Multimedia Signal Processing*, 2006.
- [16] M. S. Singh, V. Pondenkandath, B. Zhou, P. Lukowicz, and M. Liwicki, “Transforming Sensor Data to the Image Domain for Deep Learning - an Application to Footstep Detection,” *International Joint Conference on Neural Networks (IJCNN)*, May 2017.
- [17] X. Ma, Z. Dai, Z. He, J. Ma, Y. Wang and Y. Wang, “Learning Traffic as Images: A Deep Convolutional Neural Network for Large-Scale Transportation Network Speed Prediction,” *Sensors (Basel)*, vol. 17, no. 4, Apr. 2017.
- [18] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet Classification with Deep Convolutional Neural Networks,” *In Proceedings of the 25th International Conference on Neural Information Processing Systems*, pp. 1097-1105, Dec. 2012.
- [19] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, “Learning and Transferring Mid-level Image Representations Using Convolutional Neural Networks,” *Computer Vision and Pattern Recognition*, pp. 1717-1724, June. 2014.
- [20] S. Lawrence, C. L. Giles, A. C. Tsoi, A.D. Back, “Face recognition: A convolutional neural-network approach,” *IEEE Trans. Neural Networks*, vol. 8, no. 1, pp. 98-113, Feb. 1997.
- [21] S. Ji, W. Xu, M. Yang, and K. Yu, “3D Convolutional Neural Networks for Human Action Recognition,” *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 35, no. 1, pp. 221-231, Jan. 2013.
- [22] A. Karpathy, G. Toderici, S. Shetty, T. Leung, R. Sukthankar, and F. F. Li, “Large-Scale Video Classification with Convolutional Neural Networks,” *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1725-1732, June. 2014.
- [23] U. Brunelli, V. Piazza, L. Pignato, F. Sorbello, S. Vitabile, “Two-days ahead prediction of daily maximum concentrations of SO₂, O₃, PM₁₀, NO₂, CO in the urban area of Palermo, Italy,” *Atmospheric Environment*, vol. 41, no. 14, pp. 2967-2995, May. 2007.
- [24] R. Vautard, B. Bessagnet, M. Chin, and L. Menut, “On the contribution of natural Aeolian sources to particulate matter concentrations in Europe: Testing hypotheses with a modelling approach,” *Atmospheric Environment*, vol. 39, no. 18, pp. 3291-3303, June. 2005.
- [25] T. Ni, B. Han, Z. Bai, “Source Apportionment of PM₁₀ in Four Cities of Northeastern China,” *Aerosol Air Qual. Res.*, vol. 12, pp. 571-582, 2012.
- [26] J. Hooyberghs, C. Mensink, G. Dumont, F. Fierens, O. Brasseur, “A neural network forecast for daily average PM₁₀ concentrations in Belgium,” *Atmospheric Environment*, vol. 39, no. 18, pp. 3279-3289, June. 2005.
- [27] E. Stadlober, S. Hörmann, and B. Pfeiler, “Quality and performance of a PM₁₀ daily forecasting model,” *Atmospheric Environment*, vol. 42, no. 6, pp. 1098-1109, Feb. 2008.
- [28] S. Min, B. Lee and S. Yoon, “Deep learning in bioinformatics,” *Briefings in Bioinformatics*, July 2016.
- [29] Q. Li, T. W. Cai, X. Wang, et al, “Medical image classification with convolutional neural network,” *In: 2014 13th International Conference on Control Automation Robotics & Vision (ICARCV)*, pp. 844-848, Mar. 1997.

- [30] N. Liu, J. Han, D. Zhang, et al, "Predicting eye fixations using convolutional neural networks," *In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 362-370, June. 2015.
- [31] G. Hinton, L. Deng, D. Yu, et al, "Deep neural networks for acoustic modeling in speech recognition: the shared views of four research groups," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 82-97, Nov. 2012.
- [32] TN. Sainath, A-R. Mohamed, B. Kingsbury, et al, "Deep convolutional neural networks for LVCSR," *In: 2013 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE, New York*, pp. 8614-8618, 2013.
- [33] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "LeNet-5, convolutional neural networks," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278-2324, Nov. 1998.
- [34] M. Aubreville, C. Knipfer, N. Oetter, C. Jaremenko, E. Rodner, J. Denzler, C. Bohr, H. Neumann, F. Stelzle, and A. Maier, "Automatic Classification of Cancerous Tissue in Laserendomicroscopy Images of the Oral Cavity using Deep Learning," *Computer Vision and Pattern Recognition*, vol. 2, Mar. 2017.
- [35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A. C. Berg, and L. Fei-Fei, "ImageNet Large Scale Visual Recognition Challenge," *International Journal of Computer Vision (IJCV)*, vol. 115, no. 3, pp. 211-252, Dec. 2015.
- [36] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, "Microsoft coco: Common objects in context," *in European Conference on Computer Vision. Springer*, vol. 8693 pp. 740-755, 2014.
- [37] C. Szegedy, V. Vanhoucke, S. Ioffe, and J. Shlens, "Rethinking the Inception Architecture for Computer Vision," *Computer Vision and Pattern Recognition*, vol. 3, Dec. 2015.
- [38] M. S. Singh, V. Pondenkandath, B. Zhou, P. Lukowicz and M. Liwicki, "Transforming Sensor Data to the Image Domain for Deep Learning - an Application to Footstep Detection," *International Joint Conference on Neural Networks (IJCNN)*, May 2017.
- [39] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," *Computer Vision and Pattern Recognition*, June 2015.
- [40] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson, "How transferable are features in deep neural networks?," *Advances in Neural Information Processing Systems 27*, pp. 3320-3328, Dec. 2014.
- [41] L. Rokach, "Ensemble-based classifiers," *Artificial Intelligence Review*, vol. 33, no. 1-2, pp. 1-39, Feb. 2010.
- [42] H. C. Kim, E. Kim, C. Bae, J. H. Cho, B. Kim, and S. Kim, "Regional contributions to particulate matter concentration in the Seoul metropolitan area, South Korea: seasonal variation and sensitivity to meteorology and emissions inventory," *Atmos. Chem. Phys.*, 17, pp. 10315-10332, 2017.
- [43] S. Kim, "Weighted K-means support vector machine for cancer prediction," *SpringerPlus*, 5:1162, 2016.



SungHwan Kim He received the Ph.D. degree in Biostatistics from of University of Pittsburgh in 2015, and his MS and BA from the Department of Statistics at the Korea University. He is working as an assistant professor in the Department of Statistics at the Keimyung University. His research interests include deep learning-based models to address the problems of vision analysis and omic-data integration.



Songi Kim She is a master degree student in the Department of Statistics at the Keimyung University and obtained the BS degree of Statistics at the Keimyung University. Her research topics cover machine learning and reinforcement learning to advance AI theories.