Dust Detection in Satellite Data using Convolutional Neural Networks

CyberTraining: Big Data + High-Performance Computing + Atmospheric Sciences

Changjie Cai[1], Jangho Lee[2], Yingxi Rona Shi[3], Camille Zerfas[4]
Research assistant: Pei Guo[5] Faculty mentor: Zhibo Zhang[6]

[1]Dept. of Occupational and Environmental Health, University of Oklahoma
[2]Dept. of Atmospheric Sciences, Texas A&M University
[3]USRA and NASA GSFC Climate and Radiation Laboratory
[4]School of Mathematics and Statistics, Clemson University
[5]Department of Information Systems, University of Maryland, Baltimore County
[6]Department of Physics, University of Maryland, Baltimore County

Technical Report HPCF–2019–15, hpcf.umbc.edu > Publications

Abstract

Atmospheric dust is known to cause health ailments and impacts earth's climate and weather patterns. Due to the many issues atmospheric dust contributes to, it is important to study dust patterns and how it enters the atmosphere. In the past, many scientists have used satellite data and physical-based algorithms to detect and track dust, but these algorithms have many shortcomings. Herein, we consider Convolutional Neural Networks to classify dust in satellite images to try to improve the accuracy of dust detection. We describe the satellite data used, discuss the model structures, and provide results for the models built. These models show promising preliminary results.

1 Introduction

Atmospheric dust impacts the earth's climate as well as humankind's quality of life. Dust originates from natural sources such as deserts, however dust emissions has been impacted by human activities on land [11]. Tracking atmospheric dust is important because it has been linked to respiratory illnesses, such as asthma, meningitis, and others. Fungi, bacteria, and even some viruses can travel on aerosol particles for miles, causing the spread of diseases and other ailments. Dust storms are also known to destroy crops, reduce visibility, and ruin communication facilities [9]. Furthermore, it has direct and indirect effects on earth's climate and weather. The dust particles can reflect sunlight back into outerspace, or absorb radiation from this light. Dust in the atmosphere also interacts with clouds as ice nuclei, effecting the formation and lifetime of clouds [6]. A major source of atmospheric dust is earth's drylands. If earth's climate warms up, drylands will become more arid and deserts will expand. This will lead to more dust in the atmosphere, and a more substantial need to understand exactly how atmospheric aerosols impacts the earth and quality of life.

Dust detection algorithms have been studied for many years using several different types of satellite data, such as Total Ozone Mapping Spectrometer, Ozone Monitoring Instrument, METEOSAT, AVHRR (Advances Very High Resolution Radiometer), GOES-VISSR (Geostationary Operational Environmental Satellite, Visible Infra-Red Spin-Scan Radiometer), and MODIS (Moderate Resolution Imaging Spectroradiometer) [1, 3–5, 12]. All of these algorithms are based on the dust optical properties because of its distinct characteristics on spectral light absorption. However, these physical-based algorithms do have several shortcomings, such as the empirical threshold depending on the physical and optical properties of the dust. The algorithms may miss dust due to complicated observing conditions or extreme thin or thick dust loading [9]. Many studies have been conducted to understand these dust detection algorithms and their uncertainties.

Because of the shortcomings of physical-based methods, attention has recently turned to using statistical and machine learning techniques to improve dust-detection algorithms [2, 7]. In this article, we apply 1 dimensional (1D) and 2 dimentional (2D) Convolutional Neural Networks (CNN) to detect dust in satellite images, which is organized as follows. The introduction next includes information on the CALIPSO and VIIRS data, then some background on CNN. In sections 2 and 3, we provide the results of the 1D and 2D CNN results, respectively. Lastly, in section 4, we discuss the future directions we intend to take with this project.

1.1 Satellite data used

We now introduce the different types of satellite data used in the models.

1.1.1 CALIPSO

The Cloud-Aerosol Lidar and Infrared Pathfinder Satellite Observation (CALIPSO) satellite was launched on April 28th, 2006, and is part of the A-train constellation. The goals of launching the CALIPSO satellite were to (1) improve the information available on aerosol sources, (2) better understand how these aerosols enter the atmosphere, and (3) study their effect on weather patterns [11]. The Cloud-Aerosol Lidar with Orthogonal Polarization (CALIOP) sensor onboard CALIPSO satellite has 532 nm and 1064 nm two channels.

This satellite can resolve aerosol vertical distribution; most of the currently used passive sensors cannot do this. Its polarization capability also allows it to easily identify particles shapes (spherical vs. non-spherical), which made the sensor ultra sensitive to dust particles, even when the dust signal is low. Combining the backscattering, color ratio, and depolarization ratio, the CALIOP algorithm is able to identify 3 types of clouds (water/ice/mix clouds) and 10 different types of aerosol, including two type of dusts: polluted dust and dust. CALIOP cloud and aerosol layer product is used in this study. The layer product has a horizontal resolution of 1/3km, 1km, and 5 km. Here we uses 5 km product. The CALIOP identified layer feature is used as benchmark for detecting dust in this study.

1.1.2 VIIRS

The Visible Infrared Imaging Radiometer Suite (VIIRS) on board SuomiNPP mission was launched on October 28, 2011. It has a similar orbit to the A-train constellation. The equator passing time for VIIRS is 1:30 pm, which is the same for CALIPSO. The VIIRS sensor has 16 M bands with 750 meter native resolutions from 412 nm to 12 micron, and 5 I-bands with 375 meter resolution. Compared with CALIOP, VIIRS has a big advantage of spatial coverage. CALIOP observes Earth like a ”curtain” with very limited horizontal data coverage, while VIIRS foot print is 3000 km, which means almost global coverage daily.

For this study, we used collocated VIIRS and CALIOP data produced by Wisconsin SIPS group. The data contains CALIOP layer product information and corresponding VIIRS channel information and viewing/illumination geometries. Due to the narrow swath of CALIOP, the VIIRS data is only one dimension along the CALIOP track. Thus, using the collocated VIIRS & CALIOP data, we trace back to the original level 1B VIIRS file and found the two neighbouring pixels (two on both side) next to the collocated VIIRS pixel. All 5 VIIRS pixel is marked with the CALIOP classification that is corresponding to the center VIIRS pixel.

1.2 Convolutional Neural Networks

Convolutional Neural Networks (CNN) are a type of artificial intelligence that takes an image as input and assigns importance (weights and biases) to objects in the image. It is able to differentiate between various objects in the image. This type of neural network is able to identify the temporal and spatial dependencies in an image, and through the training of the network, it can understand complicated images very well [8].

2 Input data and model structure

In this section we describe the input data used to build the models, as well as the different CNN models considered. We then provide results and discussion for the four CNN models built.

We summarized the CALIPSO aerosol categories in Figure 2.1 showing that the percentage of each category is very similar in each season. Summary of the result is in Figure 2.1. We extracted only semi-clear sky (i.e., excluding cloud mixed with dust) environment. With semi-clear sky filtering, the ration between ”Dust” and ”Clear Sky (non-Dust)” is 1:11. CALIPSO category used is shown in detail in Figure 2.1.

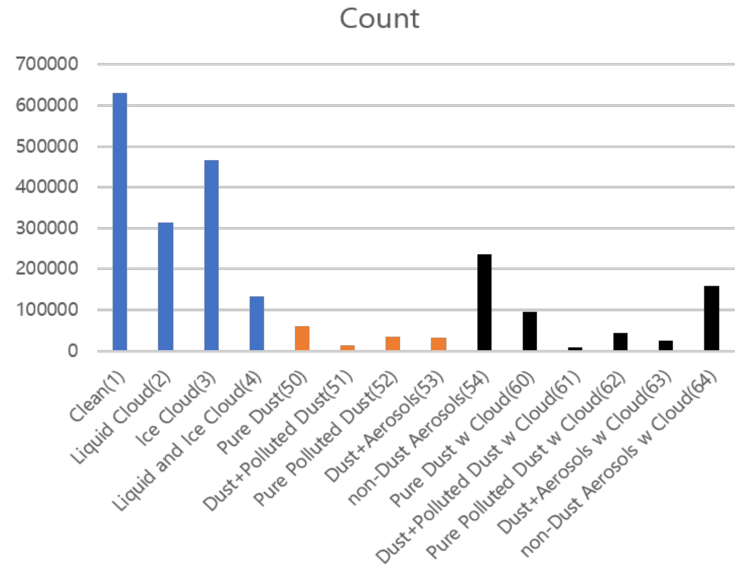

Figure 2.1: CALIPSO aerosol categories. Blue bars denote the classes count as non-dust, while orange bar represents the dust. Black bars are excluded from the study.

Figure 2.2 shows the global spatial distributions of the dust relative frequency in each season, indicating that the dust relative frequency has strong spatial and seasonal pattern. Dust mainly occurs over the major dust source regions including the land of the Northern Africa (Sahara), the Arabian, and the Central Asia (Gobi and Taklamakan). When mobilized dust aerosol are vertically mixed into higher atmospheric levels, they can be carried over long distances by strong wind. For example, in spring, the dust transport was observed from the Asian deserts towards the Pacific Ocean and North/Central America. This transport might be due to the cold fronts in spring (such as 'Kosa' events) that cause strong winds around dust sources. In summer, another clear dust transport was also observed over the subtropical region of Atlantic Ocean, between Northern Africa and Southern America. In winter, relatively high frequency of dust was observed over Northern China and Russia. The observational results from satellite are consistent with previous studies [10]. Due to the strong spatial and temporal variations of the dust relative frequency, including variables related to land types and time might improve the modeling performance.

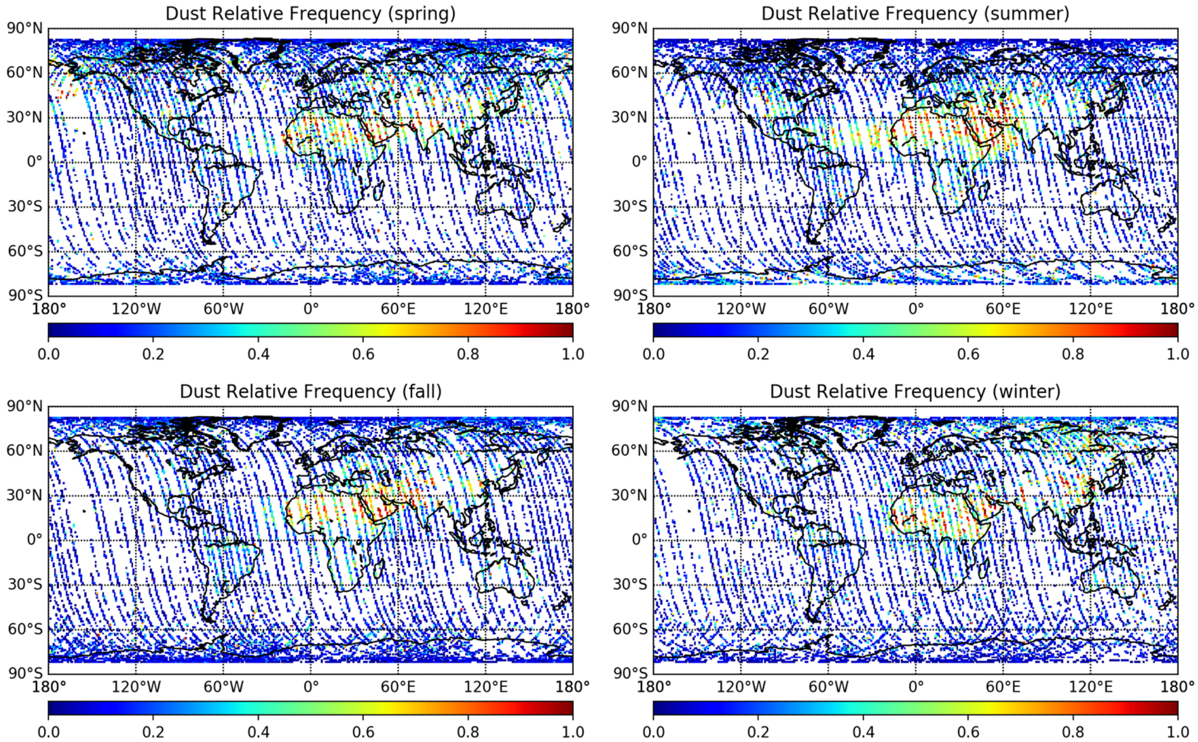

Figure 2.2: Global distributions of dust relative frequency in each season.

Figure 3.6 depicts the dataset we used in the 1D and 2D CNN models. For the 1D CNN model, the collocated CALIPSO and VIIRS satellite data was used. For the 2D CNN model, an additional 4 (2 on left, 2 on right) tracks of VIIRS data was added. The 1D CNN model data was extracted form each observation using a 1×5 moving window, whereas a 5×5 window was used to extract 2D data. The annual data has a total of 6,896,210 elements. With the filtering of the semi-clear sky environment, this reduces to 2,250,339 The dataset was then divided into training and testing sets, respectively containing 1,500,026 (66.66%) and 750,013 (33.33%) data points. Both sets of data were distributed evenly throughout the season, with a three-day interval starting from 01/02/2014 for testing data, and otherwise training. 32 predictor variable for each pixels were selected for both 1D and 2D CNN model, which can be found in Table 2.1. Note that nine combinations of radiative channels are referenced from physical-based models of detecting dust from radiative channels. Using the 32 predictor variables, classification of dust / non-dust in the center pixel ((3,1) for 1D, (3,3) for 2D) was selected as the target variable.

32 Selected Predictor Variables

Radiative Channel (16)		Geometric Information (4)
M01	M09	Solar Azimuth Angle (SAA)
M02	M10	Solar Zenith Angle (SZA)
M03	M11	Viewing Azimuth Angle (VAA)
M04	M12	Viewing Zenith Angle (VZA)
M05	M13	
M06	M14	
M07	M15	
M08	M16	
Observation Data (3)		**Combination of Radiative Channels (9)**
Julian Date (1-365)		M15-M16
Latitude		M15-M12
Longitude		M7-M15
		M4/M3
		M5/M4
		M3/M5
		(M11-M3)/(M11+M3)
		M1/M2
		M1/M11

Table 2.1: Description of predictor variables used in model construction and validation

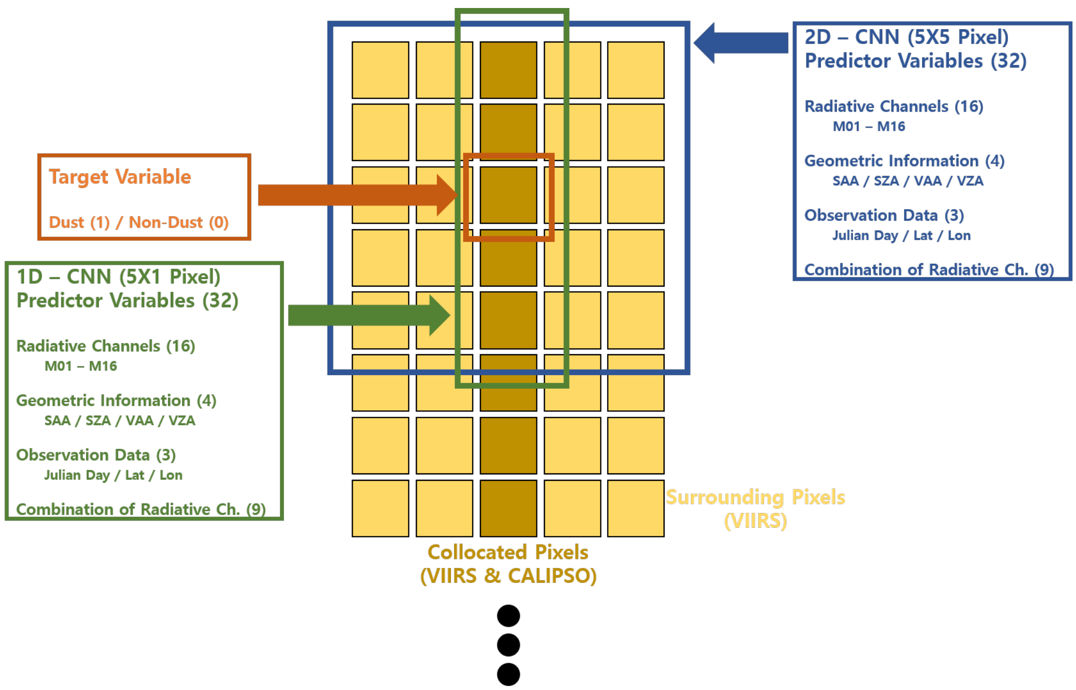

Figure 2.3: Schematic diagram of data used in model construction.

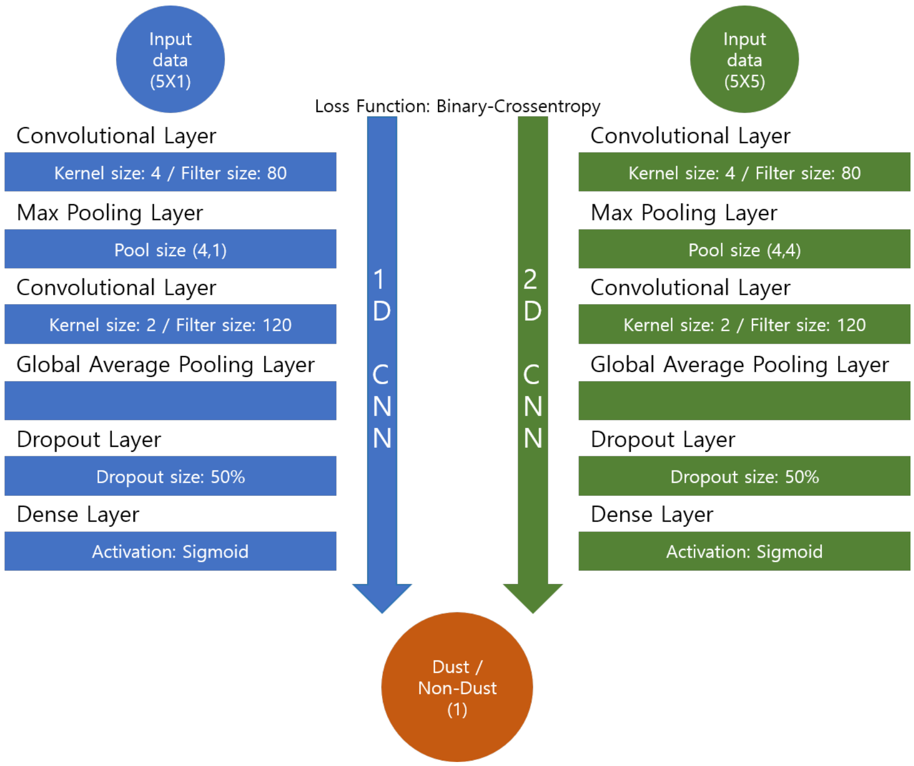

Figure 2.4: Schematic diagram of model structure for 1D CNN (blue) and 2D CNN (green).

3 Model performance

3.1 Input dimension sensitivity

To test the input dimension sensitivity of the model, 8 different models were tested as mentioned in Figure 2.3 and 2.4. Each of the 8 models were tested in two different weights, namely 10w and 20w in which will be discussed in more detail in section 3.2. The accuracy of the model is more dependent on weight selection than dimension selection, and did not differ a lot from each other. We calculated error matrix to represent the accuracy of each model set up. The error matrix is calculated in four categories: 1) the dust-dust percentage is the number of cases when CALIOP observed and model predicted are both dust divided by the number of CALIOP observed dust. 2) the dust-nondust percentage is the number of cases when CALIOP observed dust and model predicted nondust divided by the number of CALIOP observed dust. 3) the nondust-nondust percentage is the number of CALIOP observed and model predicted are both nondust divided by the total number of CALIOP identified nondust cases. 4) the nondust-dust percentage is the number of CALIOP observed nondust and model predicted dust divided by the total number of CALIOP identified nondust cases.

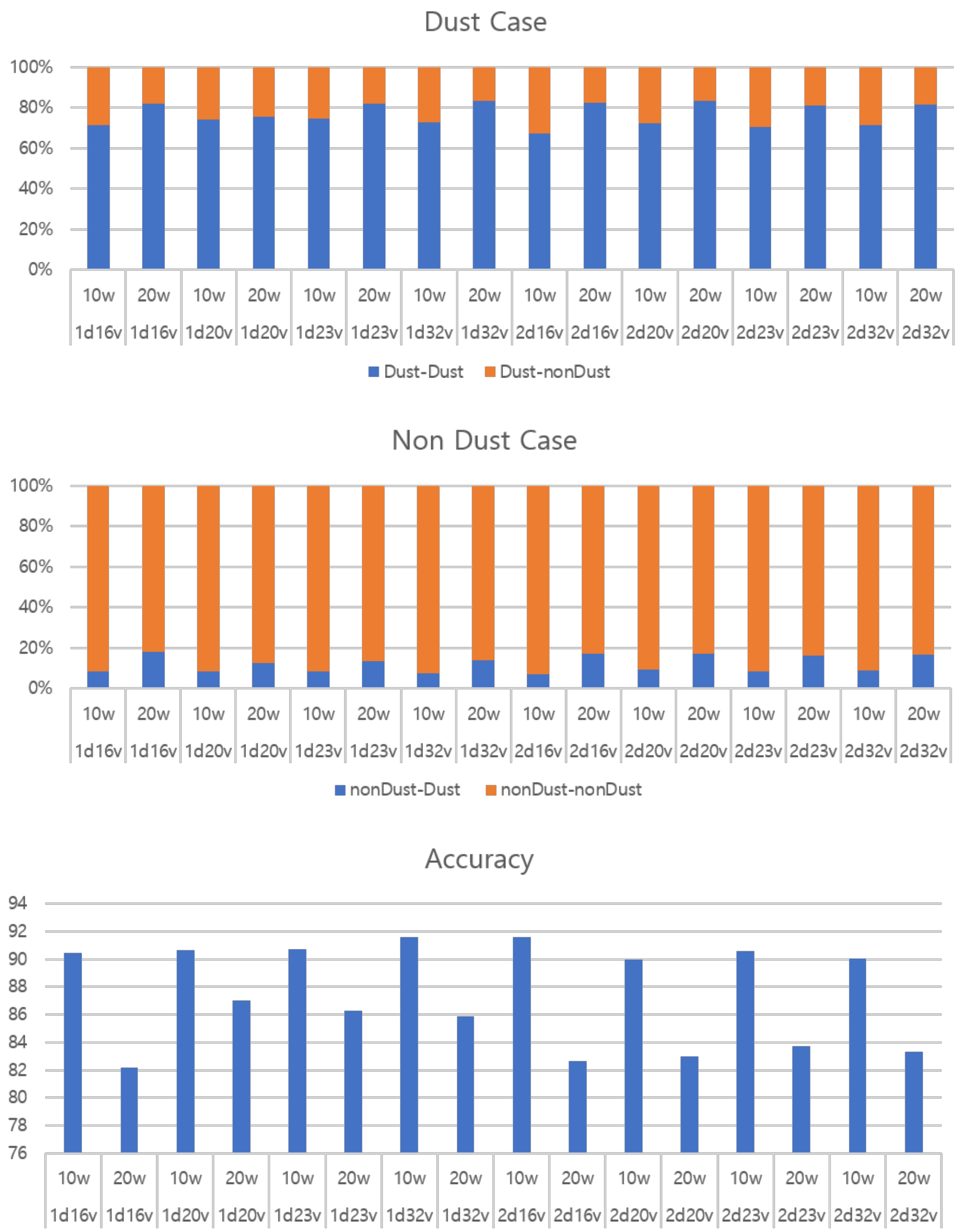

Figure 3.1: Accuracy in percentage of model with different dimension and weight.

3.2 Weight sensitivity

Since there is a bias of numbers in data samples between dust:non-dust (1:11), appropriate weight selection is crucial on setting up the model. For example, when the model weight is set to 10:1, the model will count predicting dust correctly 10 times more valuable then predicting the non-dust correctly. Model weight sensitivity test is done with 2D32V model, which is the most complex model. Accuracy of the model is the best when the weight is set to 1:1 as in Figure 3.2. However, we need to look at the accuracy in more detail, since 95% of the data is non-dust, predicting all data with non-dust will cause 95% accuracy. So the confusion matrix is calculated for each weight model and the results are as in Figure 3.3 and 3.4. Even the accuracy is the lowest in 20w model, the accuracy of predicting dust as dust is the highest in 20w model, with 83%. However, accuracy of predicting non-dust as non-dust is also lowest in 20w model, with 83%. Compared with 1w, which have 32% accuracy of predicting dust and 98% of predicting non-dust, the higher weighted models tend to predict dust far better then lower weighted models (51%), while prediction performance of non-dust case decrease slightly (15%)

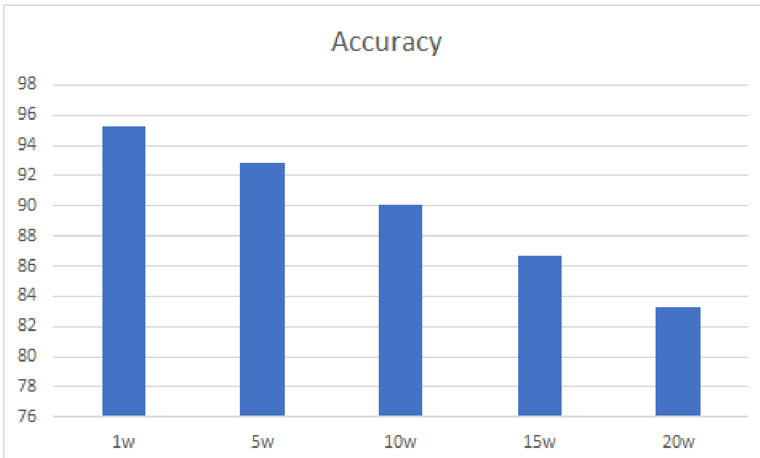

Figure 3.2: Accuracy in percentage of model with different weight selection. 1w, 5w, 10w, etc. stands for dust:non-dust 1:1, 1:5, 1:10 weight sesection.

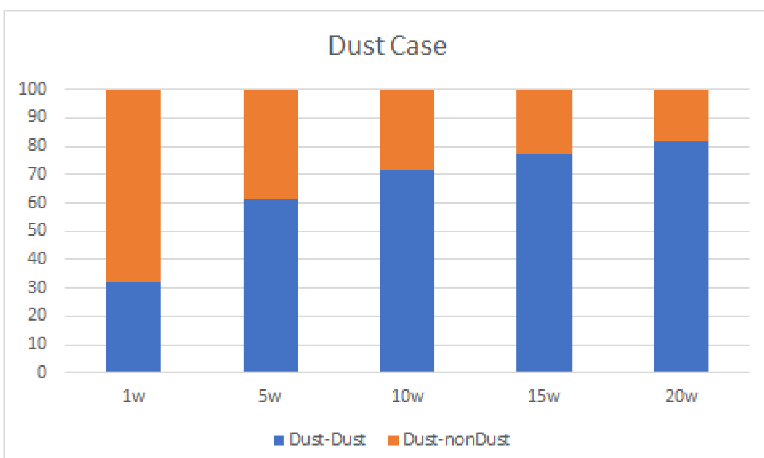

Figure 3.3: Accuracy in percentage of predicting dust samples from observation.

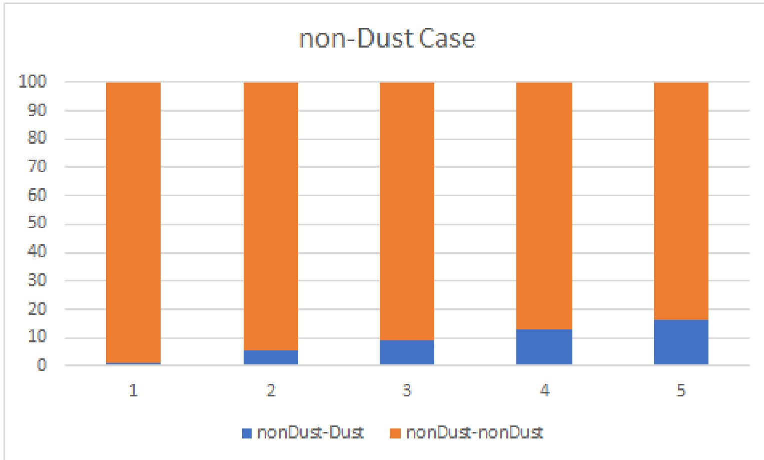

Figure 3.4: Accuracy in percentage of predicting non-dust samples from observation.

3.3 Temporal sensitivity

Since dust data has temporal characteristics as in Figure 2.2, model's temporal dependency of accuracy has been tested. 2D32V model with 20w was used for this test. The result is in Figure 3.5. The accuracy is fairly constant throughout the season.

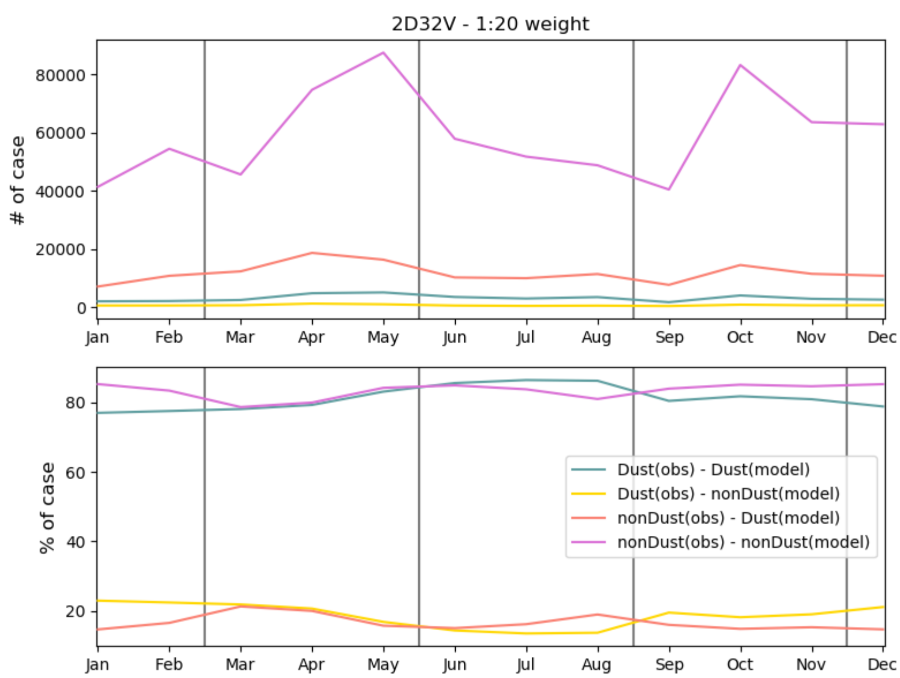

Figure 3.5: monthly averaged number of cases and percentage of accuracy change in 2D32V-20w model for 2014.

3.4 Spatial sensitivity

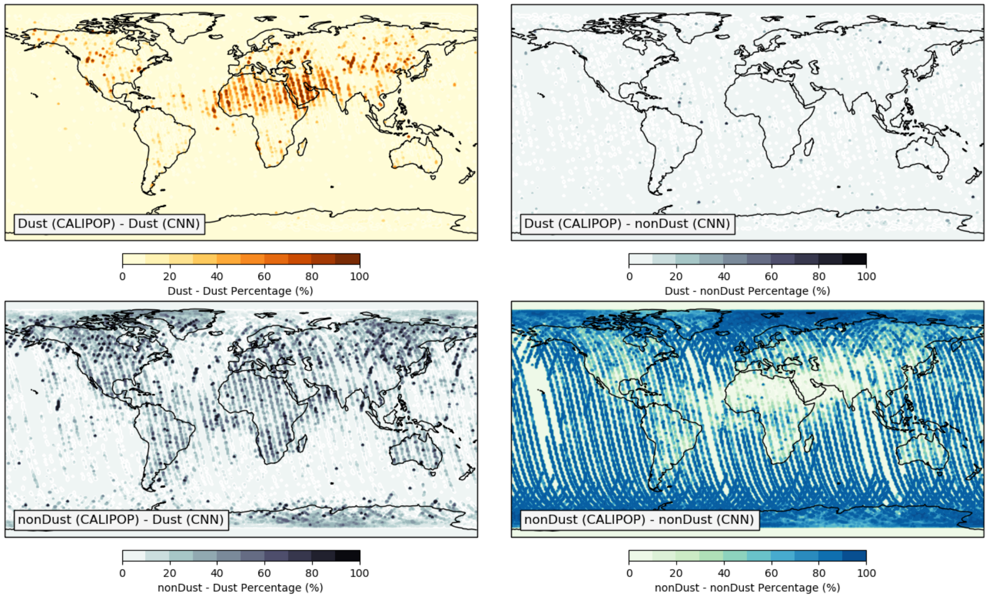

Figure 3.6: Spatial distribution of confusion matrix.

Spatial dependency on performance of the model was also tested as in Figure 3.6. Accuracy of dust-dust is high over Northern Africa (Sahara) and Central China (Gobi). Otherwise, the accuracy is low since the number of observed dust is close to zero. Error of predicting dust as non-dust is fairly low throughout the globe. However, the error of predicting non-dust as dust is relatively higher. The error is high over the land area, especially northern Canada and polar region. Accuracy of predicting non-dust as non-dust is very high throughout the globe, except for regions with lot of dust occurrence (Sahara, Gobi).

3.5 Performance compared to existing models

Performance of the model has been compared to the existing model of VIIRS aerosol detection product (ADP), which is developed by NOAA VIIRS team and is considered as state of art dust/smoke detection product that is based on physics. The product was evaluated using collocated ADP and CALIOP data. The metric used accuracy (Acc), Probability of Correct Detection (POCD), and Probability of false detection (POFD) which is defined as follows:

1)$ACC = (TP + TN)/(TP + TN + FP + FN)$
2)$POCD = (TP)/(TP + FP)$
3)$POFD = (FN)/(TP + FN)$

where TP, FN, FP, FN each represents true positive, true negative, false positive, false

	S-NPP	CNN-1d32v10w
ACC	92.3	91.6
POCD	92.5	72.6
POFD	19.8	27.3

Table 3.1: Comparison of metrics between prediction models

negative. Results are shown in Table 3.1.:

4 Future directions

The authors intend to continue this work in the near future. We will further change the combination of the input data to improve the performance of the model. Furthermore, the comparison between CNN based model and other algorithms (e.g., random forest, physics-based algorithm) will be followed.

Acknowledgement

This study is supported by NSF award (#1730250) for CyberTraining: DSE: Cross-Training of Researchers in Computing, Applied Mathematics and Atmospheric Sciences using Advanced Cyberinfrastructure Resources. We thank the UMBC High Performance Computing Facility (HPCF) for their hardware and maintainance.

Author Contributions

All authors contributed to the write up.

References

[1] P. Alpert, P. Kishcha, A. Shtivelman, O. Krichak, and H. Joseph. Vertical distribution of saharan dust based on 2.5-year model predictions. *Atmos. Res.*, 70:109–130, 2004.

[2] M. I. Chacon-Murguia, Y. Quezada-Holguin, and S. Rivas-Perea, P.and Cabrera. Dust storm detection using a neural network with uncertainty and ambiguity output analysis. *Pattern Recognition, Lecture Notes in Computer Science*, 6718:305–313, 2011.

[3] A. T. Evan, A. K. Heidinger, and M. J. Pavolonis. Development of a new over-water advanced very high resolution radiometer dust detection algorithm. *J. Remote Sens.*, 27:3903–3924, 2006.

[4] D. J. MacKinnon, P. S. Chavez, R. S. Fraser, T. C. Niemeyer, and D. A. Gillette. Calibration of goesvissr, visible-band satellite data and its application to the analysis of a dust storm at owens lake, california. *Geomorphology*, 17:229–250, 1996.

[5] K. K. Moorthy, S. S. Babu, S. K. Satheesh, J. Srinivasan, and C. B. S. Dutt. Dust absorption over the great indian desert inferred using ground-based and satellite remote sensing. *J. Geophys. Res.*, 112, 2007.

[6] J. Prospero. Long-range transport of mineral dust in the global atmosphere: Impact of african dust on the environment of the southeastern united states. *Proc. Natl. Acad. Sci.*

[7] P. Rivas-Perea, J. G. Rosiles, M. I. Chacon Murguia, and J. J. Tilton. Automatic dust storm detection based on super- vised classification of multispectral data. *Studies Comput. Intell.*, 312:443–454, 2010.

[8] Sumit Saha and Sumit Saha. A comprehensive guide to convolutional neural networks- the eli5 way, Dec 2018.

[9] A.H. Souri and S. Vajedian. Dust storm detection using random forests and physical-based approaches over the middle east. *J. Earth Syst. Sci.*

[10] I. Tegen and K. Schepanski. The global distribution of mineral dust. *IOP Conference Series: Earth and Environmental Science*, 7:012001, 2009.

[11] David Winker. Cloud-aerosol lidar and infrared pathfinder satellite observations.

[12] A. Zhu, V. Ramanathan, F. Li, and D. Kim. Dust plumes over the pacific, indian and atlantic oceans: Climatology and radiative impact. *J. Geophys. Res: Atmos.*, 112:3903–3924, 2007.