Developing Data-Driven Artificial Neural Network for a High Throughput Retrieval of Aerosol Optical Depth and Surface Temperature of Mars

Rafael Moreno[1], Michael D. Smith[2], Samuel A. Atwood,[3,4], Emily L. Mason[5], and George Nehmetallah[1]
[1] EECS Department, The Catholic University of America, 620 Michigan Av., N.E., Washington, DC 20064, USA
[2] NASA Goddard Space Flight Center, NASA Goddard Space Flight Center, Greenbelt, MD 20771, USA
[3]Space and Planetary Science Center, and Department of Earth Sciences, Khalifa University, Abu Dhabi, UAE
[4] Laboratory for Atmospheric and Space Physics, University of Colorado, Boulder, CO, USA
[5]Center for Space Sciences & Technology, University of Maryland, Baltimore County, Baltimore, MD 21250, USA

Abstract

In this work, we aim to develop artificial neural network (ANN) techniques to reproduce the retrieval results of physical quantities from spacecraft observations of solar system bodies using radiative transfer methods. The particular application here is the retrieval of dust optical depth, water-ice optical depth, and surface temperature on Mars using daytime observations obtained by the Thermal Emission Spectrometer (TES) onboard the Mars Global Surveyor. Compared against the results obtained from traditional radiative transfer retrieval techniques, our ANN successfully recovered the three quantities using daytime observations. The principal advantage of these machine learning (ML) algorithms is their complete automation and high throughput. Therefore, the algorithms presented here would be useful for very large datasets and would make practical the sampling of many different approximations or boundary conditions related to a given observation dataset and retrieval problem.

I. Introduction

The retrieval of atmospheric properties of solar system bodies from remote sensing observations by spacecraft using radiative transfer techniques has been well established and explored for some time (e.g., the review book by Hanel et al., 2003). The general concept is to use a radiative transfer model (RTM) to simulate the spacecraft observation. The quantities to be retrieved are treated as input parameters to the RTM and are varied until there is a "best fit" between the model output and the observation. The RTM tends to be the most computationally intense part of the retrieval, but it is also generally straightforward with techniques such as discrete ordinates (e.g., Goody and Yung, 1989; Thomas and Stamnes, 2002) or Monte Carlo methods (e.g., Wolff et al., 2006; Vincendon and Langevin, 2010) used to handle the calculation. In general, not all the required inputs to the RTM can be retrieved, and so assumptions or approximations must be made. The fitting portion of the retrieval is generally handled as a type of non-linear minimization problem and often requires constraints, boundary conditions, or other mathematical techniques to choose between different solutions that are often not unique (or essentially non-unique within observational uncertainties).

Because the traditional RTM methods use physics-based calculations to retrieve the underlying physical parameters, the computational time to perform these retrievals can be significant when applied to modern planetary datasets that can have a very large number (e.g., millions) of individual observations. Here we demonstrate that machine learning (ML) predictive modelling approaches based on regression techniques that allow for the prediction of continuous outcomes (Allen 1997) can reproduce certain results retrieved using a simplified radiative transfer model (Smith 2004), which are taken to be “ground truth” for this exercise. The results presented here are intended as a first step that demonstrates the potential of pursuing ML techniques. The ultimate goal for future work would be to develop a ML approach that would significantly reduce the time required for the retrieval, while maintaining fidelity to the results from a full radiative transfer retrieval. We expect that such ML techniques would be most useful for processing very large datasets, or for aiding in repetitive, time-consuming tasks like searches over parameter space for the optimization of different assumptions or approximations.

The ML model on which we center our effort is an artificial neural network (ANN), which is a modeling technique inspired by the human nervous system that has a specific architecture format consisting of neurons connected to each other by weighted links, which allows the establishment of nonlinear correlations and complex interactions between features and outcomes in order to create a model capable of learning by example from representative data that describes a physical phenomenon (e.g., Malekian et al., 2021).

The example we use here is the remote sensing of the atmosphere of Mars using observations taken by the Thermal Emission Spectrometer (TES) onboard the Mars Global Surveyor (MGS) (Christensen et al., 2001) orbiting Mars, which has been used to perform radiative transfer retrievals of dust aerosol optical depth (DOD), water ice aerosol optical depth (WIOD), and surface temperature (ST), among other quantities (e.g., Conrath et al., 2000; Pearl et al., 2001; Smith et al., 2001; Smith 2004)

In Section 2 we discuss a brief theoretical background of the TES thermal infrared spectra of Mars and the typical values of DOD, WIOD, and ST retrieved from those thermal infrared spectra. In Section 3 we discuss the development of different ANN architectures and present our results. The goal is to train a model capable of extracting the very complex and abstract features (i.e., the geophysical quantities) from the given inputs (i.e., the observed thermal-infrared radiance of Mars) that is resilient and robust enough to generate accurate predictions in a short amount of time without relying on RTMs. In Sections 4 and 5, we describe the application of the ML algorithms to the TES observations, and in Section 6 we discuss our conclusions.

II. Theoretical Background of Mars Thermal-Infrared Spectra

Incoming shortwave radiation from the Sun is absorbed by the surface of Mars and warms the planet. Mars then radiates back to space at its surface temperature of 145 – 300 K, which is primarily in the thermal infrared. As this radiation propagates upward, it is absorbed by atmospheric gases (CO_2, H_2O, and others) and aerosols (dust and water ice clouds) at characteristic frequencies (wavenumbers). The radiation that is absorbed in the atmosphere is re-radiated at the temperature of the atmosphere. If the atmosphere and the surface are at different temperatures, this will produce distinctive signatures in the spectrum that can be modeled to estimate the surface temperature and the quantity of the atmospheric absorbers. Figure 1 shows a typical daytime thermal-infrared spectrum of Mars from TES along with blackbody curves at different temperatures. The bottom shows the same spectrum converted to its equivalent “brightness temperature”, which is the temperature of a blackbody with the same radiance as the observation. The large feature centered at 667 cm^{-1} is caused by CO_2 gas absorption and can be used to retrieve atmospheric temperature as a function of height (e.g., Conrath et al., 2000). Dust and water ice aerosols have absorption coefficients (or opacity) with distinctive spectral dependencies. The absorption from dust peaks near 1075 cm^{-1}, while that of water ice aerosols (clouds) peaks near 825 cm^{-1} (Smith 2004). The most transparent portion of the spectrum (which is most representative of surface temperature) is near 1300 cm^{-1}.

The amount of dust and water ice aerosol suspended in the atmosphere is quantified using its "optical depth", which is the integrated value of the absorption coefficient over the path of the observation. Using optical depth is convenient because it does not require knowledge of aerosol composition or other physical properties such as particle size distribution or shape, and it is the natural quantity that is used in the radiative transfer calculation. For Mars at thermal-infrared wavelengths, dust typically has an optical depth of ~0.1 when "clear", ~0.5 when "dusty", and can be greater than unity during large dust storms. Thermal-infrared water ice optical depth can range from 0.1 to 0.5 in thicker clouds but is often less than 0.05 (Smith 2004).

The radiative transfer model of Smith (2004) does not include aerosol scattering, but it still retains all the most important and relevant radiative transfer physics. In fact, for the Mars atmosphere at the wavelengths considered here, the difference between the Smith (2004) radiative transfer and a more complete model that includes aerosol scattering is at least to first order a difference in scaling in the DOD and WIOD values. Therefore, this dataset, while simplified, remains a useful starting point for investigating the potential of ML techniques.

*

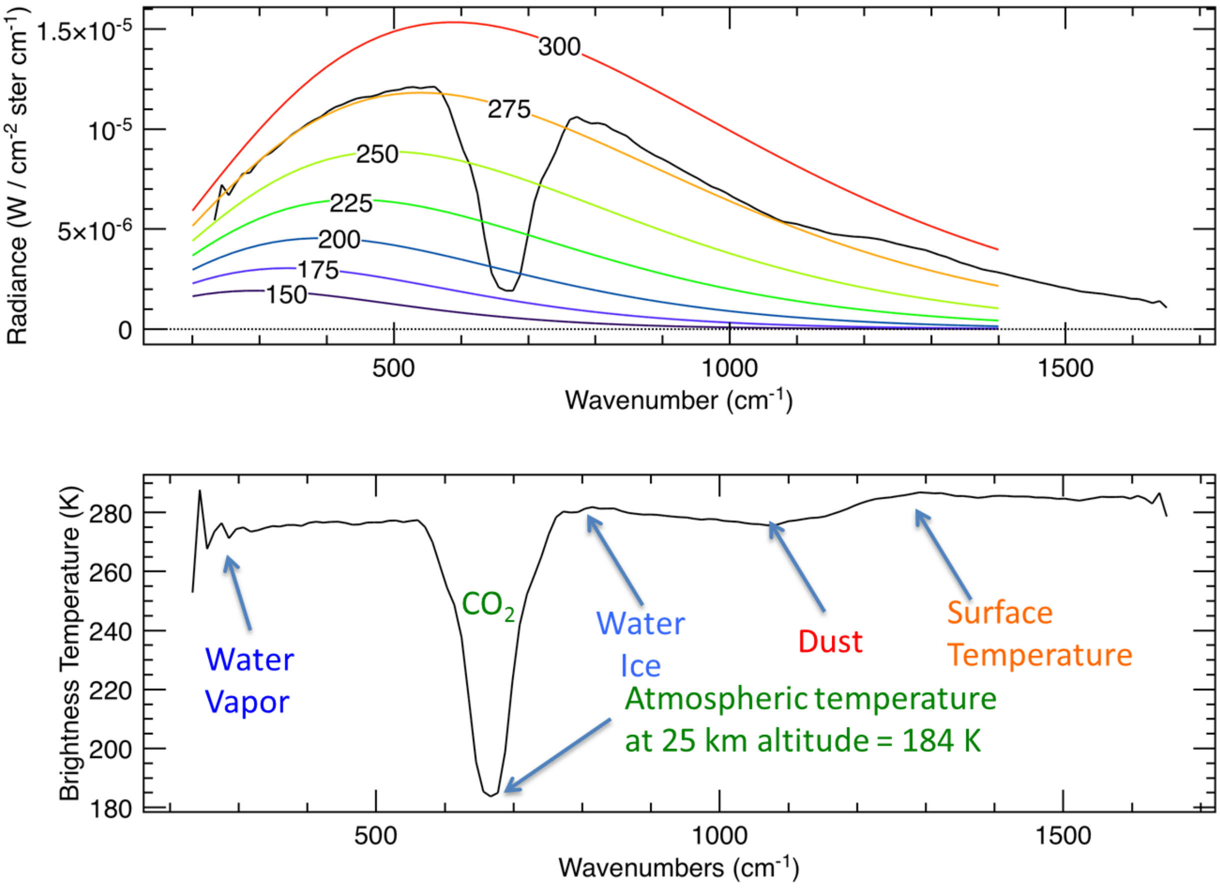

Figure 1: This is a daytime Mars spectrum taken by TES. The atmosphere of Mars is almost transparent at 1300 cm^{-1}, so brightness temperature at that wavenumber is representative of the surface temperature. The CO_2 feature (centered at 667 cm^{-1}) is used to retrieve the atmospheric temperature as a function of height above the surface. The center of the feature is representative of temperature at about 25 km altitude. Dust has a broad absorption feature centered at about 1075 cm^{-1}. Water ice clouds produce an absorption feature centered at about 825 cm^{-1} (not apparent in this spectrum). Water vapor has a distinctive set of narrow absorptions between about 200 and 400 cm^{-1}.

III. Data Description, Metrics, and Machine Learning Algorithms

Because the Mars Global Surveyor spacecraft was in a Sun-synchronous near-polar orbit around Mars, the observations collected from the TES instrument were taken at either 2:00 AM or 2:00 PM local time. Observations of Mars were taken nearly continuously by TES from pole to pole, completing 12 orbits of Mars per day. The complete TES dataset consists of roughly 50 million spectra covering nearly 3 Mars Years (more than 5 Earth years). For the purposes of this study, we have selected representative subsets, each consisting of roughly 50,000 spectra (or roughly five days of data) covering different seasons on Mars.

We begin our analysis using daytime spectra (2:00 PM local time) that have sufficient thermal contrast between the surface and atmosphere to produce good spectral signatures. One dataset (the "aphelion" dataset) consists of data taken at Ls=135° during the aphelion season on Mars, which is the time of the Martian year when Mars is furthest from the Sun. This season has relatively cool surface temperatures and tends to have relatively high WIOD and a relatively low optical depth of dust aerosols (e.g., Smith et al., 2004). A second dataset (the "perihelion" dataset) consists of data taken at Ls=235° during the perihelion season on Mars, which has relatively high surface temperature, high dust optical depth, and low water ice cloud optical depth. A third dataset (the "equinox" dataset) consists of data taken at Ls=180°, or equinox, and has surface temperatures and aerosol optical depth values intermediate between the other two datasets.

These three datasets contain five days, or 60 consecutive orbits, of data. These datasets were chosen by visual inspection to be typical for each season and not to include unusual events such as a global dust storm. Because TES observed Mars continuously from a polar orbit, the observations are equally spaced over all latitudes, and the 60 orbits sample longitudes without repeating. While there can be spatial autocorrelations in atmospheric conditions, these are typically regional in scale, a distance that the spacecraft travels in a few minutes. Hence, the observations are largely independent of each other, and these observations do not constitute a time series for conditions at any one location.

The quantities of interest for the ML techniques are the three target variables DOD, WIOD, and ST, which have also been retrieved using standard radiative transfer techniques (e.g., Smith 2004). The quantities used as predictors for the ML algorithms were the 98 spectral channels of the observed radiance spectrum, which extend from roughly 300 to 1350 cm^{-1} (or a wavelength range of 6–50 μm) as shown in Figure 1.

A comprehensive implementation of ML for fast hyperspectral radiative transfer simulations developed by Le, T., et al (2020) uses an ANN model, which they claim capable of alleviating the computational burden by two to three orders of magnitude, and of generating radiances with small relative errors, generally less than 0.5% compared to exact calculations. Another instance on a ML implementation to predict spectral radiance distributions, developed by Del Rocco et al. (2020), uses regression models trained on high dynamic range (HDR) imagery and spectroradiometer measurements. The authors measured the incoming radiance at various points, with those measurements they trained regression models, and used those models to predict atmospheric spectral radiance (350-1780 nm) at 1 nm resolution for the entire sky, given a photo of a clear sky and its capture timestamp.

A very informative exploration of different types of deep neural networks was performed by Kästner et al. (2018) for the processing of mass spectra measured with a miniaturized mass spectrometer to analyze gaseous mixtures with the aim of predicting the concentration of four mixture constituents. They explored using back propagation neural networks (BPNNs) and convolutional neural networks (CNNs) to extract spatial relationships in the dataset. They increased the depth of the ANNs with the help of shortcuts, namely residual blocks and highway blocks. Their best result was achieved with a ResNet configuration containing 120 layers.

In this work, we focused our efforts on feed-forward ANNs, implemented with TensorFlow (Abadi et al., 2015) and Keras. We started with a simple back BPNN before moving to more complex architectures such as a CNN (LeCun et al., 1999) and the state-of-the-art residual neural networks like ResNet (He et al., 2015). We wished to investigate these different ANN architectures with the goal of assessing their performance related to handling noise present in the measurements.

IV. Artificial Neural Networks

IV.1 Back Propagation Neural Network (BPNN)

A multi-layer perceptron (MLP) model is a neural network with at least two hidden layers and qualifies as the simplest DNN (Deng and Yu, 2014). The hidden layers enable composition of features, with the aim of modeling complex non-linear relationships (Bengio, 2016). The BPNN is a fully connected neural network which performs two main actions: propagation (forward and backward) and weight adjustment. When performing forward propagation, an input signal is assigned a weight by the activation function in the hidden layer, and then the weighted signal is passed to the output layer to calculate the output value. Afterwards, the backward propagation will only take place if the difference between the output value and the target output value falls outside the tolerable error range (Hecht-Nielsen, 1992).

In this case, we used the structure shown in Figure 2 which consists of a fully connected BPNN with two hidden layers consisting of 64 neurons and 32 neurons respectively, using the rectified linear unit (ReLU) as the activation function.

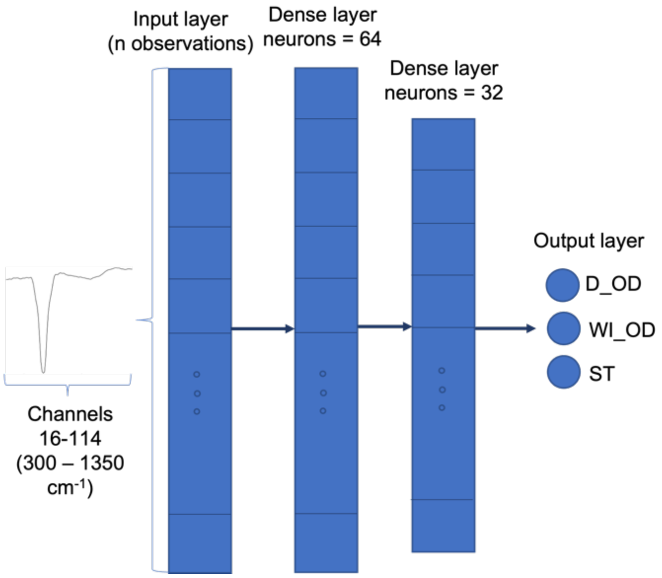

Figure 2: A fully connected BPNN with two hidden layers consisting of 64 and 32 neuron each, using the Rectified Linear Unit (ReLU) as the activation function.

IV.2 Convolutional Neural Network (CNN)

A CNN is a deep ANN designed for processing structured arrays of data such as spectra, however it is primarily used for image recognition (e.g., Fang et al., 2018) and natural language processing for text classification (e.g., Yin et al., 2017). The key building block in a CNN is the convolutional layer, which contains a convolutional kernel that slides over the arrays of data in order to detect high level patterns (Kuo, 2016).

The kernels or filters are applied to the input data to take into consideration spatial information. In our case, the "spatial" dimension is actually a spectral dimension across the different channels of the TES observations. In addition, a pooling layer was applied to reduce the data dimension by down-sampling. The aim of the CNN is to extract spatial (spectral) characteristics resulting in better predictions by ignoring noise present on the spectra. The schematic of the CNN is shown in Figure 3.

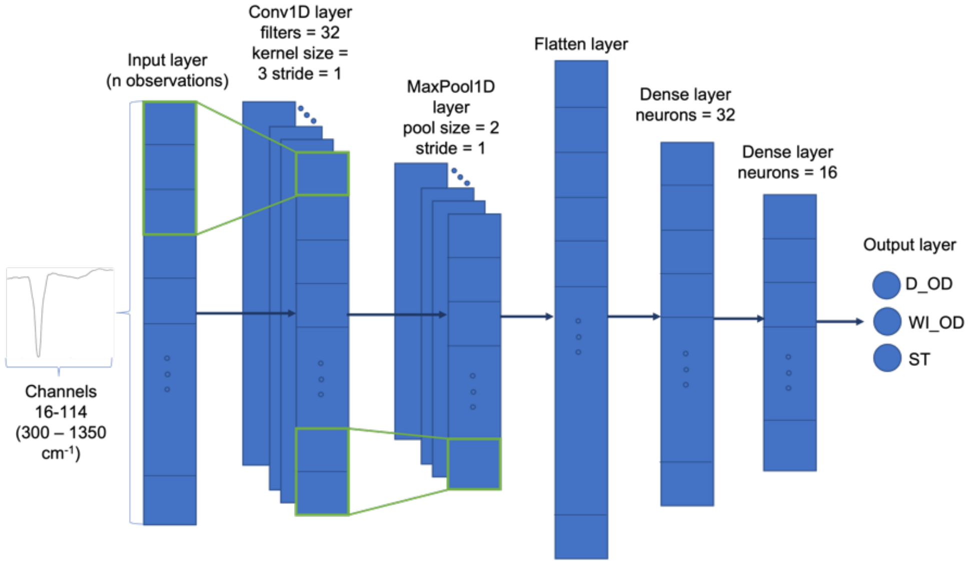

Figure 3: Schematic of the CNN that consists of a Conv1D layer of 32 filters with a kernel size 3, stride of 1 and ReLU as the activation function, followed by a MaxPool1D layer with pool size 2 and stride of 1.

IV.3 Residual Neural Network ResNet 50

ResNets are largely considered to be state-of-the-art networks with a special focus in image recognition tasks (He et al., 2015; Ren et al., 2017), however they have also been successfully adapted in certain instances for regression related tasks (Kästner et al., 2018). ResNet picks up on the idea of a deeper CNN in order to identify low, mid, and high-level features from spatial data by stacking more layers without the drawback of the vanishing gradient. As the gradient is back-propagated to earlier layers, repeated multiplication may make the gradient extremely small. As a result, as the network goes deeper its performance gets saturated or even starts degrading rapidly. This issue is known as the vanishing gradient.

He et al. (2015) addressed this problem by restructuring the layers in blocks, each with a residual function where the output is taken into consideration on the subsequent block of layers through shortcut connections that allow for identity mappings. The overall ResNet50 structure consists of 5 stages each with a convolution and identity block. Each convolution block has 3 convolution layers, and each identity block also has 3 convolution layers. A residual block is a stack of layers set in such a way that the output of a layer is taken and added to another layer deeper in the block. The non-linearity is then applied after adding it together with the output of the corresponding layer in the main path. This by-pass connection is known as the shortcut or the skip-connection. Therefore, the skip connections add the outputs from previous layers to the outputs of stacked layers, making it possible to train much deeper networks than previously possible.

We adapted the architecture proposed by He et al. (2015) for ResNet 50 using 1-deminsional convolutional layers instead of the 2-dimentional ones used for images. The same test and training-validation split as for the two previous ANNs were implemented. Table 1 shows the parameter settings for the three ANN models used in this work.

Table 1: Parameter settings for the three ANN models used in this study.

Model	Parameters							
	Epochs	Filter/Neuron	Layers	Learning rate	Batch size	Kernel size	Patience (early stopping)	Optimizer
BPNN	50	64 neurons	Fully Connected	0.0001	32	-	10	Adam
		32 neurons						
		3 neurons						
CNN	50	32 filters	Convolutional	0.0001	32	3	10	Adam
		32 neurons	Fully Connected			-		
		16 neurons						
		3 neurons						
ResNet 50	50	ResNet50 structure (He et al. 2015)		0.0001	32	-	10	Adam

IV.4 Techniques to prevent overfitting

To avoid overfitting during model training we defined the following Keras callbacks:

- *EarlyStopping:* to stop the progress if the monitored evaluation metric for the validation loss is no longer improved after 10 iterations (Prechelt, 1998).
- *ReduceLROnPlateau:* to decrease the learning rate when the monitored metric has stopped improving (Chang et al., 2010).

IV.5 Evaluation indicators

The following evaluation indicators (Allen 1997) were used to determine the best model:

(i) The root mean square error (*RMSE*), is the square root of the mean of the square of all the errors, where x_i are the target variable observations, $\hat{x}_i = f(x_i)$ are the predicted values of the target variable, and *N* is the number of observations. This is described as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{N}(x_i - \hat{x}_i)^2}{N}}. \tag{1}$$

(ii) The coefficient of determination (R^2), which corresponds to the squared correlation between the observed outcome values and the predicted values by the model, ranges from 0 to 1, with 1 being the best possible result. This is described as:

$$R^2 = 1 - \frac{RSS}{TSS}, \tag{2}$$

$$RSS = \sum_{i=1}^{N}(x_i - \hat{x}_i)^2, \tag{3}$$

$$TSS = \sum_{i=1}^{N}(x_i - \bar{x})^2. \tag{4}$$

where $\bar{x}$ is the mean of the observed data, *RSS* is the residuals sum of squares and *TSS* is the total sum of the squares.

(iii) Mean absolute error (MAE), which is a measure of errors between the predicted versus observed values. This is described as:

$$MAE = \frac{\sum_{i=1}^{N} |x_i - \hat{x}_i|}{N}. \tag{5}$$

IV.6 Data preprocessing

The three datasets (aphelion, equinox, and perihelion) were partitioned in the following manner: The first three days of each were used for training the model, the fourth day of each was used to validate the performance of model in terms of the MAE of the loss function, and the last day of each was set aside for testing the model in terms of R^2 and RMSE.

Each one of these subsets (training, validation, and testing) were also split into the target and predictor variables, which were all standardized in terms of the estimated mean and standard deviation of the training subset. This procedure is also known as standard scaler and is described as:

$$Z_i = \frac{x_i - \bar{x}}{\sigma}. \tag{6}$$

where $\bar{x}$ is the mean of the training subset, and σ is the standard deviation of the training subset.

IV.7 Parameter setting

A large number of trial-and-error procedures were executed to identify the most suitable parameters for each ANN model based on the training and validation datasets, and the obtained parameter settings of the ANN models are presented in Table 1.

V. Comparison of ANN models

Table 2 shows the loss in terms of the MAE for the training and validation stages for the three ANN models. Figures 4, 5, and 6 show the MAE loss for the training and validation subsets per epoch for each ANN model, respectively.

Table 2: MAE loss values for three ANN models.

Model	Training stage	Validation stage
BPNN	0.0712	0.1153
CNN	0.0618	0.0632
ResNet50	0.0844	0.1243

BPNN

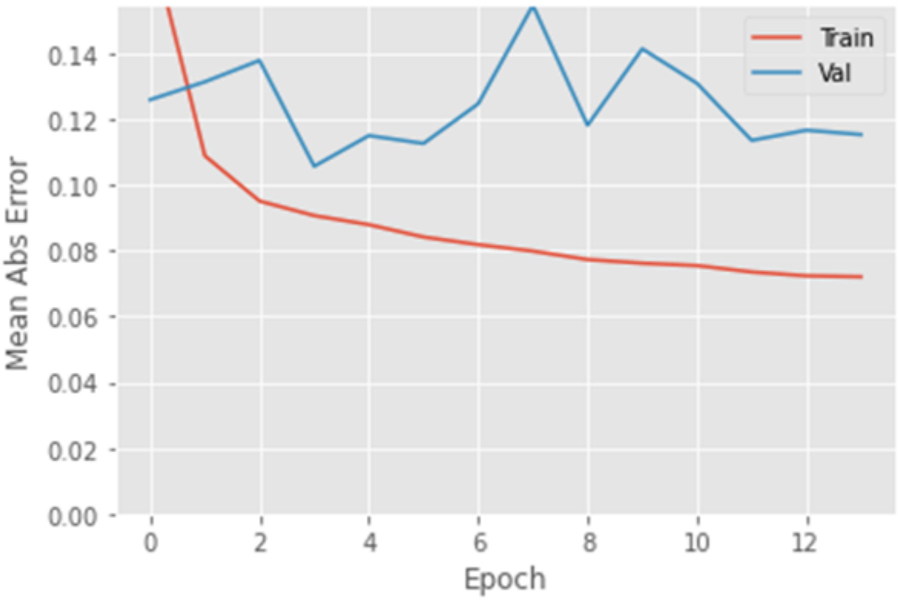

Figure 4: MAE loss per epoch for BPNN model.

CNN

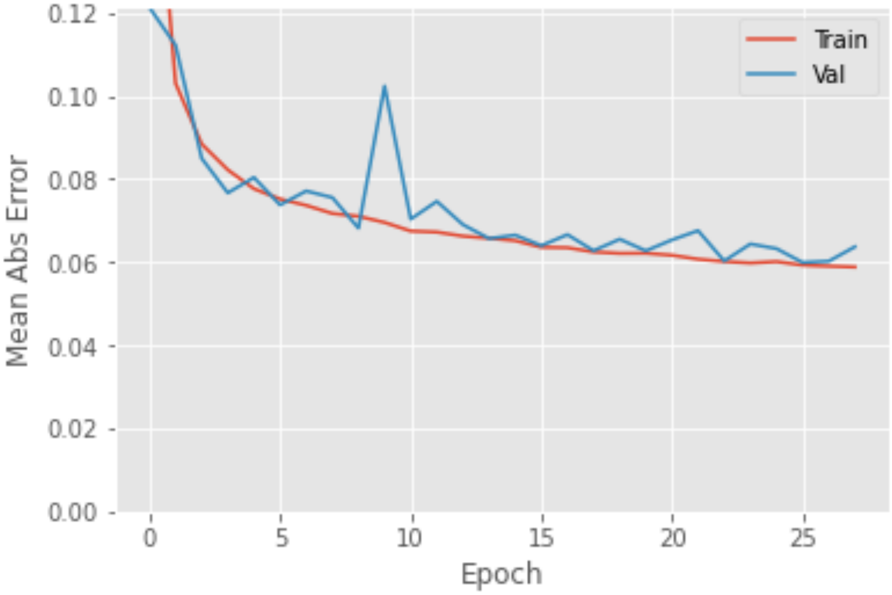

Figure 5: MAE loss per epoch for CNN model.

ResNet50

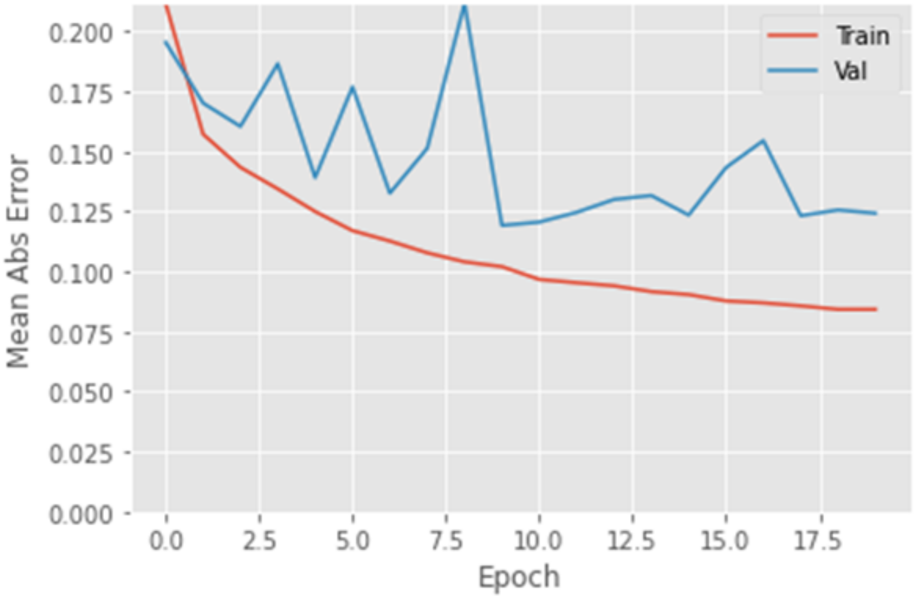

Figure 6: MAE loss per epoch for ResNet50 model.

Figures 7, 8, and 9 show the scatterplot of predictions vs. observed values for DOD, WIOD, and ST for each ANN model, respectively, during the testing stage. Table 3 shows the R^2 and RMSE values between observed and predicted DOD, WIOD, and ST for three ANN models during the testing stage.

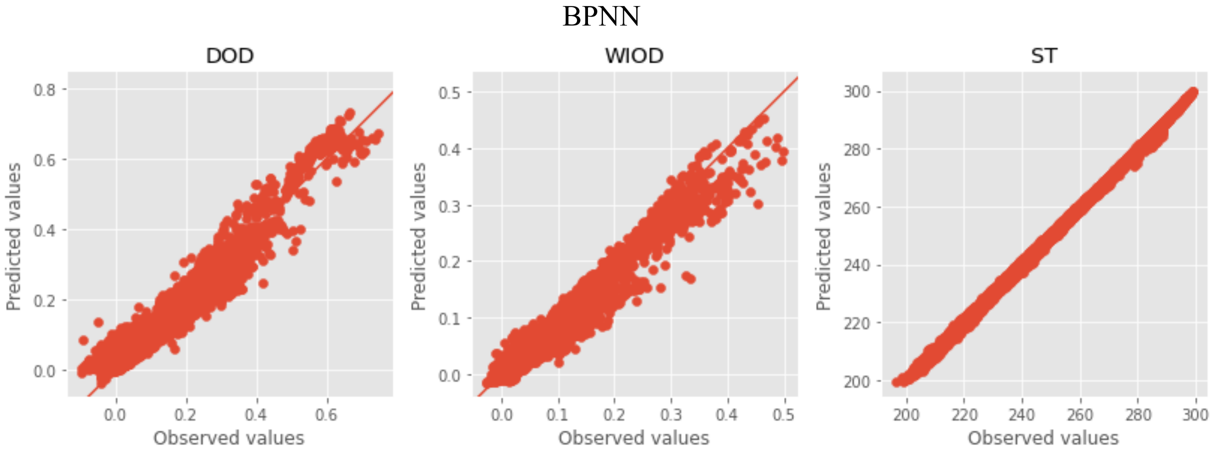

Figure 7: Scatterplot of predictions vs observed values target variables for BPNN

CNN

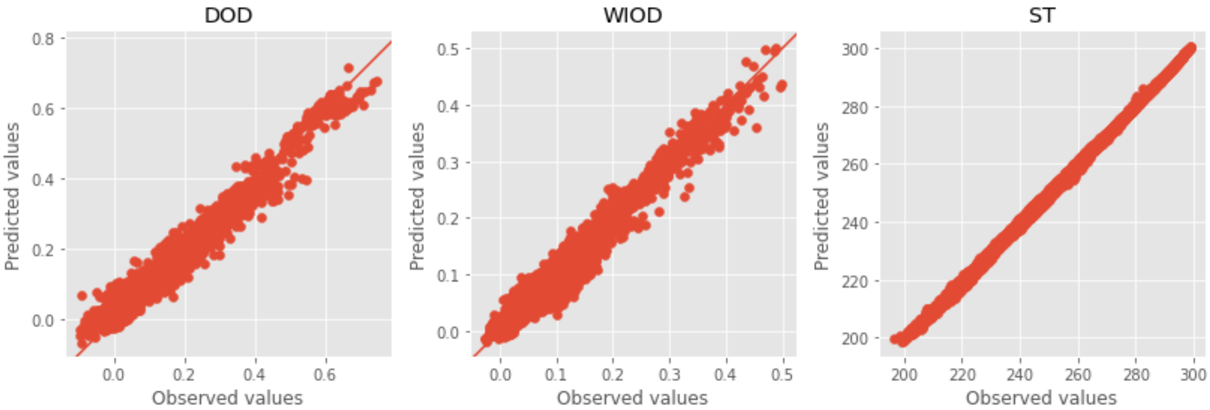

Figure 8: Scatterplot of predictions vs observed values target variables for CNN

ResNet50

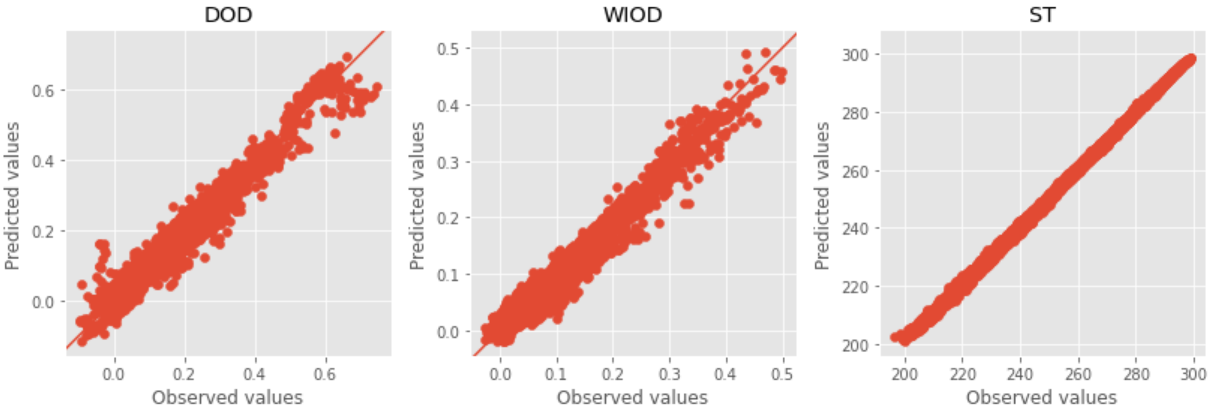

Figure 9: Scatterplot of predictions vs observed values target variables for ResNet50

Table 3: R^2 and RMSE values between observed and predicted target variables for three ANN models.

Model	**Testing stage**					
	DOD		**WIOD**		**ST**	
	R^2	RMSE	R^2	RMSE	R^2	RMSE
BPNN	0.934	0.0004	0.965	0.0001	0.999	0.3027
CNN	0.972	0.0002	0.981	0.0001	0.999	0.2777
ResNet50	0.952	0.0003	0.962	0.0001	0.985	5.0449

Neural networks are trained using a stochastic learning algorithm. Therefore, the same model fit on the same data may result in a different performance. We addressed this in our experiment by repeating the evaluation of each model configuration 30 times and expressing the result in terms of their loss MAE to ensure that the comparison between models is statically significant.

In this case we made use of SciPy's package ttest_ind() function to calculate the T-test for the mean values of MAE from two models. Comparing the means from the BPNN vs. the CNN, we obtained a p-value of 0.00077, and comparing the CNN vs. Resnet we obtained a p-value of 0.00081. Therefore, the means are significantly different with a 95% confidence level. With this analysis we can provide more certainty to the superior performance of the CNN model, which was visually identified from Figures 4,5, and 6 as the best performing model.

Observing the results of the training phase in Table 2 we can appreciate how prevalent the overfitting of the BPNN and ResNet50 models is, whereas it is significantly less in CNN. Taking a closer look at Figures 4, 5, and 6, it becomes evident that there is a higher level of overfitting per epoch present in the BPNN and ResNet50 in comparison to CNN. The training and validation results show significantly less variability for CNN, meaning that the model is more capable of generalizing when confronted with data previously unseen. The testing phase aligns with the results obtained from the training phase, where CNN continues to show better and more consistent predictive performance than the BPNN and ResNet50, as shown in Table 3.

Looking at Figures 7, 8, and 9, we can see in detail how close each one of the observed values align with the predictions from the testing subset. The closer the points in the scatter plot align on the diagonal line the better. With the scatterplots we can confirm in a more nuanced way that the CNN model produces the best fit between prediction and observed values compared with BPNN and ResNet50.

It is worth mentioning that initially we were expecting better results from the ResNet architecture, especially in terms of handling the noise present in the data, based on the findings from the comparison between a somewhat similar CNN architecture and ResNet carried out by Kästner et al. (2018) for a predictive task similar to ours. However, for our case study the high complexity of ResNet made it very difficult to fine tune, whereas the far simpler CNN approach enabled the flexibility to carry out a large number of trial-and-error procedures, which led to an architecture that successfully fits our predictive task.

VI. Conclusions

In this work, we have demonstrated the ability of different ANN models to reproduce the retrieved physical quantities from spacecraft observations of Mars that were obtained using the simplified radiative transfer model of Smith (2004). We have applied several machine learning algorithms to TES daytime thermal infrared spectra of Mars to DOD, WIOD, and ST of Mars. Once trained, our machine learning algorithms successfully estimated these three parameters for daytime observations, with higher accuracy using the CNN above the other two proposed ANN. The main advantages of these data-driven techniques are their complete automation and their very high throughput compared to traditional radiative transfer retrieval algorithms. The main disadvantage to the machine learning algorithms is that reliable results are only possible over the parameter domain where it was trained. As for a recommendation for a specific algorithm, we note that in general the CNN returned the most accurate predictions and the best quality metrics.

Many modern planetary datasets are very large consisting of millions of individual observations. We believe that the speed of machine learning algorithms such as CNN could be especially useful for these types of datasets or when a rapid turnaround of results is desired. The ability to rapidly process a large dataset would also greatly help with sensitivity studies. Radiative transfer models can require many assumptions and approximations, and the associated retrieval algorithms often have free parameters that need to be tuned to optimize the retrieval. With the help of the machine learning techniques described here, one could investigate how changes in the underlying radiative transfer model assumptions or retrieval parameter assumptions affect the retrieval results, something that may well be infeasible for large datasets using traditional radiative transfer algorithms. This work represents a first step towards this goal, demonstrating the possibility that ANNs could be developed to efficiently process large numbers of observations and to optimize retrieval algorithms through sensitivity studies.

Acknowledgments

This research is funded by the NASA Grant and Cooperative Agreement Federal Award Identification No.: 80NSSC21K1085.

References

Abadi, M., et al., 2015. TensorFlow: Large-scale machine learning on heterogeneous systems, Software available from tensorflow.org.

Allen, M.P., 1997. *Understanding Regression Analysis*. ("The coefficient of determination in multiple regression" p. 91–95), Springer, Boston.

Bengio, Y., Learning Deep Architectures for AI, *Foundations and Trends in Machine Learning*, **2 (1)**, 1–127.

Chang, F.J., Kao, L.S., Kuo, Y.M., Liu, C.W., 2010. Artificial neural networks for esti- mating regional arsenic concentrations in a Blackfoot disease area in taiwan. J. Hydrol. 388, 65e76.

Christensen, P.R., et al., 2001. Mars Global Surveyor Thermal Emission Spectrometer experiment: Investigation description and surface science results. *J. Geophys. Res.*, **106**, 23823–23871.

Clancy, R.T., F. Montmessin, J. Benson, F. Daerden, A. Colaprete, M.J. Wolff, 2017. Mars Clouds, Chapter 5 in *The Atmosphere and Climate of Mars*, Cambridge University Press, Cambridge.

Conrath, B.J., J.C. Pearl, M.D. Smith, W.C. Maguire, P.R. Christensen, S. Dason, M.S. Kaelberer, 2000. Mars Global Surveyor Thermal Emission Spectrometer (TES) observations: Atmospheric temperatures during aerobraking and science phasing. *J. Geophys. Res.*, **105**, 9509–9519.

Del Rocco J., Bourke P.D., Patterson C.B., Kider J.T., 2020. Real-time spectral radiance estimation of hemispherical clear skies with machine learned regression models**.** Sol. Energy, 204, pp. 48-63

Deng, L., and D. Yu, 2014. Deep Learning: Methods and Applications, *Foundations and Trends in Signal Processing*. **7 (3–4)**: 1–199.
Fang, W., Zhang, F., Sheng, V. S., & Ding, Y. 2018. A method for improving CNN-based image recognition using DCGAN. Computers, Materials and Continua, 57(1), 167-178.

Geurts, P., D. Ernst., and L. Wehenkel, 2006. Extremely randomized trees, *Machine Learning*, **63(1)**, 3–42.

Goody, R.M. and Y.L. Yung, 1989. *Atmospheric Radiation: Theoretical Basis*, Oxford University Press, London.

Hanel, R.A., B.J. Conrath, D.E. Jennings, R.E. Samuelson, 2003. *Exploration of the Solar System by Infrared Remote Sensing*, Cambridge University Press, New York.

He, K., X. Zhang, S. Ren, and J. Sun, 2015. Deep residual learning for image recognition, CoRR, vol. abs/1512.03385.

He, K., X. Zhang, S. Ren, and J. Sun, 2016. Identity mappings in deep residual networks, CoRR, vol. abs/1603.05027.

Kästner, F., B. Janßen, F. Kautz, M. Hübner, 2018. Exploring Deep Neural Networks for Regression Analysis, IARIA.

Khare, M.A., J.R. Murphy, C.E. Newman, R.J. Wilson, B.A. Cantor, M.T. Lemmon, M.J. Wolff, 2017. The Mars Dust Cycle, Chapter 10 in *The Atmosphere and Climate of Mars*, Cambridge University Press, Cambridge.

Kulkarni, A., J. Valadi, B.D. Kulkarni, 2009. Review on Lazy Learning Regressors and their Applications in QSAR. *Combinatorial Chemistry & High Throughput Screening*, **12(4)**, 440-50.

Kuo, C.J., 2016. Understanding convolutional neural networks with a mathematical model. ArXiv, abs/1609.04112.

Le, T., Liu, C., Yao, B., Natraj, V., Yung, Y.L., 2020. Application of machine learning to hyperspectral radiative transfer simulations. J Quant Spectrosc Radiat Transf, 246, p. 106928.

LeCun, Y., P. Haffner, L. Bottou, and Y. Bengio, 1999. *Object Recognition with Gradient-Based Learning*, (p. 319–345), Springer, Berlin Heidelberg.

Malekian, A.; Chitsaz, N. Concepts, procedures, and applications of artificial neural network models in streamflow forecasting. In Advances in Streamflow Forecasting; Elsevier: Amsterdam, The Netherlands, 2021; pp. 115–147.

Pearl, J.C., M.D. Smith, B.J. Conrath, J.L. Bandfield, P.R. Christensen, 2001. Observations of Martian ice clouds by the Mars Global Surveyor Thermal Emission Spectrometer: The first Martian year. *J. Geophys. Res.*, **106**, 12,325–12,338.

Prechelt, L., 1998. Early stopping-but when?, 1998. In: Lecture Notes in Computer Science. Springer, Berlin/Heidelberg, Germany, pp. 55e69. 978-3-642-35288-1, 978-3-642-35289-8.

Ren S., He K., Girshick R., Sun J., 2007. Faster R-CNN: Towards real-time object detection with region proposal networks. 39 (6), pp. 1137-1149

Smith, M.D., J.C. Pearl, B.J. Conrath, P.R. Christensen, 2001. Thermal Emission Spectrometer results: Mars atmospheric thermal structure and aerosol distribution. *J. Geophys. Res.*, **106**, 23,929–23,945.

Smith, M.D., 2004. Interannual variability in TES atmospheric observations of Mars during 1999–2003. *Icarus*, **167**, 148–165.

Thomas, G.E. and K. Stamnes, 2002. *Radiative Transfer in the Atmosphere and Ocean*, Cambridge University Press, New York.

Vincendon, M. and Y. Langevin, 2010. A spherical Monte-Carlo model of aerosols: Validation and first applications to Mars and Titan, *Icarus*, **207**, 923–931.

Wolff, M.J., M.D. Smith, R.T. Clancy, N. Spanovich, B.A. Whitney, M.T. Lemmon, J.L. Bandfield, D. Banfield, A. Ghosh, G. Landis, P.R. Chrristensen, J.F. Bell III, and S.W. Squyres, 2006. Constraints on dust aerosols from the Mars Exploration Rovers using MGS overflights and Mini-TES. *J. Geophys. Res. Planets*, **111**, doi:10.1029/2006JE002786.

Yin, W., Kann K., Yu, M. and Schutze H., 2017. Comparative Study of CNN and RNN for Natural Language Processing. eprint arXiv:1702.01923.