



Predicting the surface exposure time of asteroids using the space weathering features in reflectance spectra: small data machine learning.

Lakshika Palamakumbure¹, David Korda², and Tomas Kohout^{1,3,4}

¹University of Helsinki, Geography and Geosciences, (lakshika.palamakumbure@helsinki.fi)

²Astronomical Institute of the Czech Academy of Sciences, Czech Republic

³Institute of Geology of the Czech Academy of Sciences, Czech Republic

⁴School of Electrical Engineering, Aalto University, Finland

The surfaces of airless planetary bodies exposed to the interplanetary environment experience space weathering (SW). Over time, the SW process distorts certain reflectance spectral features such as spectral slope, albedo, or absorption band depths and widths. Extensive laboratory experiments studied the evolution of these spectral parameters under simulated SW conditions. However, despite these efforts, the precise relationship between SW duration and the alteration of reflectance spectra remains not fully understood. Various experiments give different, often non-linear evolution rates (Loeffler et al. 2009, Vernazza et al. 2009, Hapke et al. 2001). This research aims to develop and apply machine-learning models to estimate the surface exposure time of S-type asteroids as a function of space weathering agents and dose.

The primary source of the training spectral data are published empirical space weathering simulations. We gathered 200 spectra from publications and mineral and meteorite databases: Re-Lab, and C-TAPE. This included materials such as olivine (85), pyroxene (47), mixtures of olivine and pyroxene (9), as well as chondritic meteorites (59). These materials reflect the composition often found in S-type asteroids. Then we calculated the exposure time at 1 AU for the reflectance spectra based on the laboratory simulation conditions (Sasaki et al. 2002, Schwenn 2000). The exposure time of the dataset ranges from fresh material to 10⁹ years. In this study, we developed a Convolutional Neural Network (CNN) model (model 1) and two models using the ensemble learning technique. The ensemble learning technique combines the results of several models to improve the overall model prediction (Phyo et al. 2022). In the first ensemble model (model 2), we combined four decision tree algorithms: Gradient Boosting Tree Regressor (GBT), Decision Tree Regressor (DT), Extra Tree Regressor (ET), and K-neighbors Tree Regressor (KNT). In the second ensemble model (model 3), we combined a CNN with GBT, DT, ET, and KNT. Then we trained and evaluated these models using the stratified K-fold cross-validation method to assess the performance and generalization ability of a predictive model. The models were fed with reflectance spectra and SW conditions as independent variables, while exposure time was predicted as the dependent variable.

Figure 1 illustrates a comparative analysis of actual values versus predicted values from the three models, with the standard deviation across 10 iterations. Notably, as we progress from model 1 to model 3, there is a noticeable enhancement in prediction accuracy, coupled with a reduction in standard deviation. In Figure 2, it is observed that 68% of the predicted values from model 1

estimate time with accuracy better than 2.5 times. However, with the utilization of ensemble model 2 and ensemble model 3, this metric sees a notable improvement, reaching 77% and 80% accuracy respectively. The combination of the CNN model with decision tree algorithms not only reduces absolute error but also improves standard deviation, thus increasing the reliability of predictions. Consequently, the ensembled model, featuring CNN in combination with decision tree algorithms, yields more dependable prediction compared to a standalone CNN model. The constraint of a small dataset size limits the CNN model's capacity to assemble and extrapolate the relationship between SW and surface exposure time from irradiated samples. Conversely, decision tree-based algorithms have shown better performance under such circumstances. It is worth noting that various tree-based algorithms exhibit varying degrees of proficiency across different segments of the dataset. However, their combination notably enhances the overall performance of the model. Furthermore, the combination of the CNN model with the tree-based ensemble model yields further improvements in results. The next step of this work would be to apply this model to asteroid spectra to determine their exposure time.

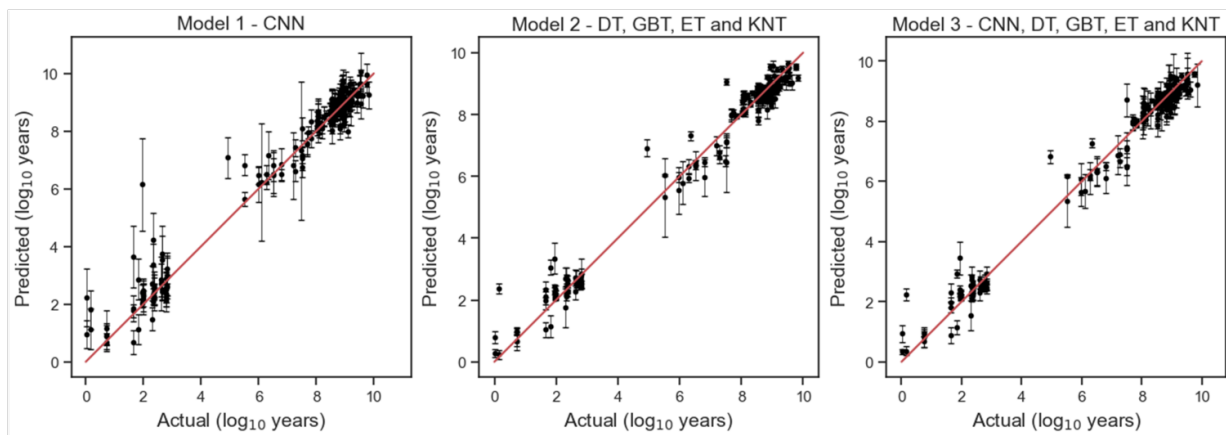


Figure 1. Scatter plots of the true and predicted exposure time of the test data. The red color diagonal line delimit the accuracy of the predictions.

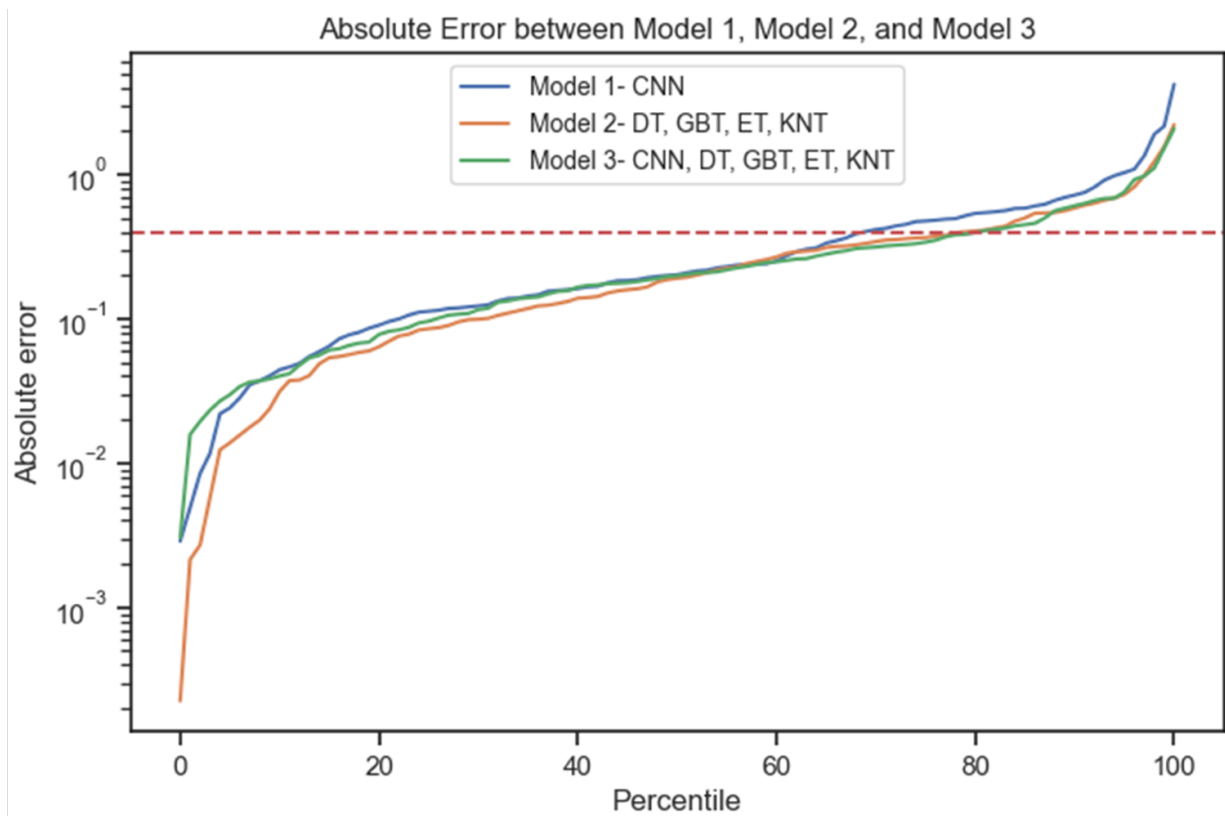


Figure 2. Quantiles of the absolute errors between the actual and predicted values of exposure time in years in log scale. The dashed red line indicate the 2.5 times error.

Hapke, B. 2001 DOI 10.1029/2000JE001338

Loeffler et al. 2009 DOI 10.1029/2008JE003249

Phyo et al. 2022 DOI 10.3390/sym14010160

Sasaki et al. 2002 DOI 10.1016/S0273-1177(02)00012-1

Schwenn, R. 2000 DOI 10.1201/9781003220435

Vernazza et al. 2009 DOI 10.1038/nature07956