Uncertainty Quantification for Martian Surface Spectral Analysis using Bayesian Deep Learning

Mark Hinds
mhinds9@gatech.edu
Georgia Institute of Technology,
Los Alamos National Laboratory

Michael Geyer
mgeyer@lanl.gov
Los Alamos National Laboratory

Natalie Klein
neklein@lanl.gov
Los Alamos National Laboratory

Abstract

Laser-induced breakdown spectroscopy (LIBS) is a rapid chemical analysis technique which has many applications in both academia and industry. Traditional techniques for analyzing the spectra to predict elemental composition include partial least squares (PLS) and random forest regression, but these methods are limited in their scalability and performance. Recently, neural networks (NNs) have been applied to this task with the goal of achieving more accurate predictions. However, quantifying the predictive uncertainty of NNs is a challenge. In scientific domains, accurate estimates of predictive uncertainty are critical for evaluating model performance, particularly when trained models are applied to new data. In an effort to solve this problem, Bayesian Neural Networks (BNNs) introduce a probability distribution over model parameters to allow uncertainty to propagate through the network. Predictive queries can then be answered alongside various uncertainty measures. In this paper, we show that BNNs can provide good predictive performance on LIBS data while delivering additional insights on elemental compositions through well-calibrated uncertainty estimates.

1 Introduction

Over the last decade, advances in Bayesian learning in deep NNs have demonstrated improved robustness to overfitting, while remaining competitive in performance in classification and regression tasks [1]. The potential for added uncertainty quantification is highly beneficial in specialized domains such as healthcare, finance, and planetary science. While traditional NNs rely on optimizing point estimates, Bayesian deep learning assumes weights are drawn from a posterior distribution, introducing a measure of uncertainty in the network. The primary challenge of probabilistic methods is accurately estimating the predictive posterior distribution. Common methods like Variational Inference (VI) [2] and the Laplace Approximation (LA)[3] to approximate the posterior distribution of the model given the data. This paper evaluates VI with a Convolutional Neural Network (CNN) utilizing the local reparameterization trick [4] against standard deterministic models in the context of analyzing spectral data.

We compare predictive models on data replicating the atmospheric conditions and material makeup of the martian surface obtained from the ChemCam instrument developed for the Mars rover Curiosity. Curiosity's laser, camera and spectrograph all work together to identify the chemical and mineral composition of rocks and soils. The instrument can rapidly identify the kind of rock being studied (volcanic or sedimentary); determine the composition of soils and pebbles; measure the abundance

Machine Learning and the Physical Sciences Workshop, NeurIPS 2024.

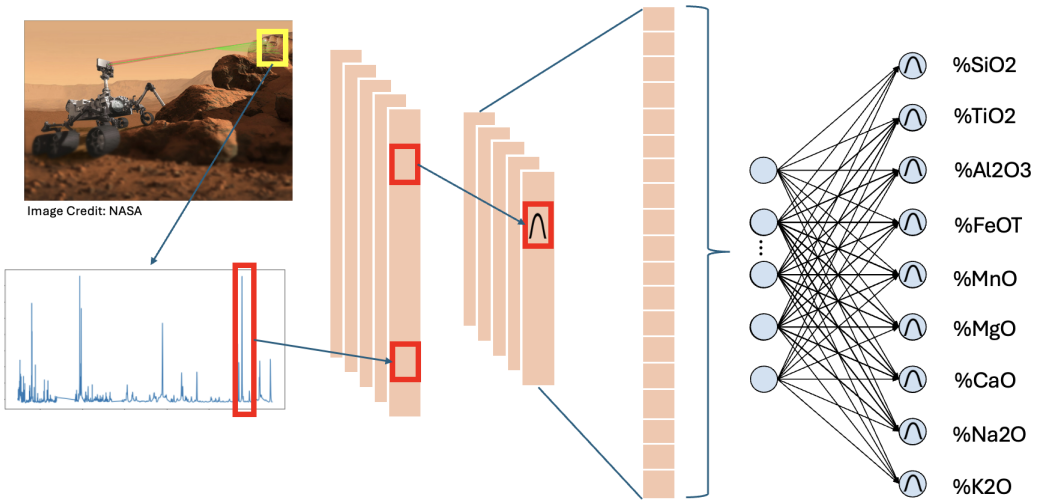

Figure 1: Laser Induced Breakdown Spectroscopy (LIBS) uses a laser to induce plasma on the surface of a target. The spectrum of light produced by this reaction can be used to identify the elemental composition of the target. In our work, data is fed through a Bayesian CNN to identify features which contribute to a certain elemental composition with added (e.g., %SiO2) uncertainty quantification.

of all chemical elements, including trace elements and those that might be hazardous to humans; recognize ice and minerals with water molecules in their crystal structures; measure the depth and composition of weathering rinds on rocks; and provide visual assistance during drilling of rock cores [5, 6].

ChemCam utilizes Laser-Induced Breakdown Spectroscopy (LIBS) which is a rapid chemical analysis technology that uses a laser pulse to create a micro-plasma on the sample surface called laser ablation as visualized in Figure 1. When a laser pulse ends, the plasma cools, causing electrons in atoms and ions to return to ground states which emit light with distinct spectral peaks. This emitted light is captured and analyzed using an ICCD/spectrograph detector module for LIBS spectral analysis across three spectral ranges: ultra-violet, violet, visible near-infrared. Classical statistical models such as linear models or random forest models have been used to quantify the oxide weight percent in samples [7, 8, 9], but recent works have explored NNs as an alternative [10, 11, 12].

CNNs effectively capture spatially correlated information and local patterns in data, making them well-suited for complex tasks such as spectral data analysis. Another motivation for using NN techniques over traditional linear techniques are matrix effects, in which interactions between distinct elements in the sample produce nonlinear spectral effects, complicating quantification [13]. While previous studies [10, 11, 12] have utilized NNs for ChemCam spectral data analysis, our approach extends this by integrating a Bayesian framework, which retains the strengths of CNNs in capturing spatial correlations and matrix effects, while also providing uncertainty quantification for a more comprehensive analysis.

2 Methodology

For our experiments, we employ a Bayesian Convolutional Neural Network (BCNN), where weights are treated as random variables sampled from a probability distribution. Since exact inference of the posterior distribution of the weights given the data is intractable due to the high dimensionality and complexity of weight space, we approximate the posterior using VI. To formalize this approach, we first define the likelihood function as follows:

$$p(\mathcal{D} \mid \mathbf{w}) = \prod_{i=1}^{N} p(y_i \mid \mathbf{x}_i, \mathbf{w}),$$

where $\mathcal{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{N}$ represents the dataset with inputs $\mathbf{x}_i$ and corresponding outputs y_i, and $\mathbf{w}$ denotes the weights of the NN. The likelihood models the probability of observing the data given the

network parameters $\mathbf{w}$. To incorporate prior beliefs about the model parameters, we define a prior distribution $p(\mathbf{w})$. In this work, we use an independent Gaussian prior. The posterior distribution is then proportional to the product of the likelihood and the prior:

$$p(\mathbf{w} \mid \mathcal{D}) \propto p(\mathcal{D} \mid \mathbf{w})p(\mathbf{w}).$$

The posterior captures our updated beliefs about the weights after observing the data. However, due to the high dimensionality and non-linearity of neural networks, this posterior is generally intractable.

To approximate the posterior, we employ VI by introducing a variational distribution $q(\mathbf{w} \mid \theta)$ parameterized by θ. The goal is to minimize the Kullback-Leibler (KL) divergence between the variational distribution and the true posterior:

$$\mathrm{KL}(q(\mathbf{w} \mid \theta) \parallel p(\mathbf{w} \mid \mathcal{D})).$$

This minimization allows us to approximate the posterior distribution with a tractable distribution $q(\mathbf{w} \mid \theta)$, which we assume to be a independent and identically distributed Gaussian $N(w_j|\mu_j, \sigma_j^2)$. The variance σ_j^2 expresses an uncertainty estimate for one NN parameter w_j. For BCNNs, we sample from the variational posterior using the Local Reparameterization Trick [4], which reparameterizes the Bayes By Backprop equation [1] to sample from activations directly, decreasing the variance of the estimator [14].

3 Experimental Settings

We implemented a Bayesian 5-layer, 1-dimensional CNN with batch normalization and ReLU activation following each convolutional layer. ReLU activations introduce sparsity, and sharper gradients help facilitate more efficient exploration of the posterior weight space. The final layer is a fully connected linear layer which outputs a single oxide prediction. A kernel size of 50 was chosen to capture a broad view of spectral features, which typically span across 5606 or more channels, unlike CNN kernels for images, where smaller kernel sizes are often sufficient. We selected a prior with zero mean and 0.01 variance, reflecting a preference for small weights, which encourages sparsity and acts similarly to ℓ_2 regularization. We also tested wider prior variances (0.1, 2) to see how performance and coverage differs. A smaller prior avoids certain pathologies of wide priors that require arbitrary scaling of the KL divergence term (as in β VAEs). We compare the BCNN to a non-Bayesian CNN with the same architecture, as well as to a deep ensemble of the CNN (eCNN), and to a non-Bayesian fully-connected NN.

ChemCam calibration data is available on the NASA Planetary Data System repository `https://pds.nasa.gov/`. For each target (584), there are different locations (5) where separate laser ablations or shots on the material occur (50), resulting in a set of photon intensities across spectral channels (5606). Targets with missing oxide compositions were removed and the dataset was reduced to train on single shot observations. Spectra were then normalized with respect to each spectral region: ultraviolet (UV, 246.635–338.457 nm), violet (VIO, 382.13–473.184 nm), and visible and near-infrared (VNIR, 492.427–849.0 nm). This results in 401 final targets with 320 targets held out for training and 81 targets held out for testing examples. We use a single shot observation from each location resulting in 2005 total spectra.

4 Results and Discussion

We benchmark the performance of five distinct modeling approaches—Partial Least Squares (PLS), fully-connected Neural Network (NN), Convolutional Neural Network (CNN), Bayesian Convolutional Neural Network (BCNN) and Ensemble CNN (eCNN)—using Root Mean Squared Error (RMSE) as the primary evaluation metric. PLS is considered a benchmark as it is a common approach used on LIBS data [8].

Table 1 shows that the CNN and eCNN generally perform the best in terms of test RMSE, outperforming the benchmark PLS model and under most cases, the NN. We speculate that the NN performs better than the CNN and eCNN for Al, Ca, Na, and K oxides because the spatial correlation assumptions encoded in the receptive field may introduce biases that affect predictive performance. The BCNN generally performs slightly worse in terms of RMSE, though we gain uncertainty quantification capabilities. In contrast, the ensemble CNN also provides uncertainty estimates, but these

Table 1: Comparison of RMSE for oxide composition predictions across different predictive models, with the best result highlighted in bold.

Oxide	PLS	NN	CNN	BCNN	eCNN
SiO2	5.8523	4.5251	4.4327	4.4760	**4.4408**
TiO2	0.6449	0.5123	0.49851	0.6489	**0.4117**
Al2O3	2.93	**2.3593**	3.1097	2.4554	2.5258
FeOT	2.9720	1.8230	**1.6832**	1.7079	1.8906
MnO	0.2769	0.4137	**0.0812**	0.4083	0.7752
MgO	1.2173	2.4293	**0.7340**	1.9163	1.4373
CaO	1.9585	**1.4920**	1.9157	1.5625	1.8153
Na2O	1.5810	**0.7114**	1.6054	1.6391	1.2461
K2O	1.7467	**0.7930**	1.5720	1.3154	1.3924

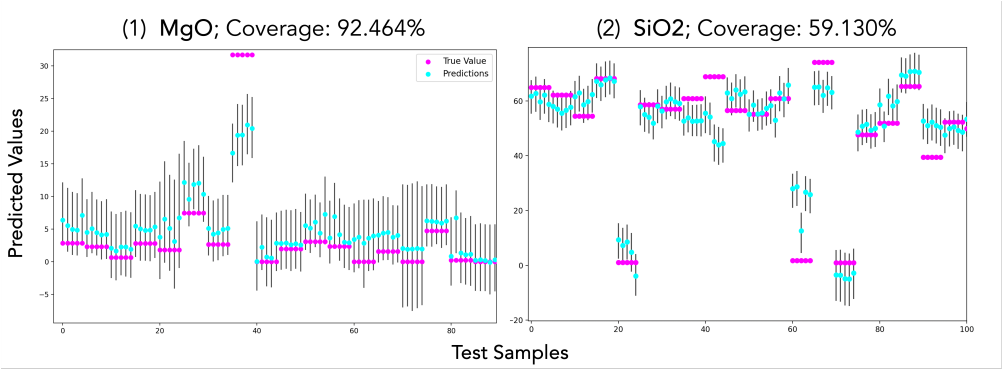

Figure 2: Predictions vs true values for Mg0 and SiO2 with 95% credible intervals (Prior = 0.1)

are not inherently probabilistic, as the method relies on aggregating predictions from multiple models rather than explicitly modeling uncertainty in a probabilistic framework. We note that varied performance across oxides has been noted in previous works (e.g., [8, 9]), influenced both by very different measured ranges for each oxide and different spectral complexity corresponding to different elements.

The limitations of deterministic models, as discussed in [9], highlight a need for a robust approach for uncertainty estimation in LIBS composition prediction. The composition of a given oxide can vary greatly between targets, and deterministic models are limited in their ability to quantify this difference. Furthermore, accurately identifying the possible range of oxide values feeds into mineralogical interpretation.

The BCNN performed best in terms of RMSE with smaller priors, but resulted in more constrained uncertainty, as shown in Figure 3. Wider priors, on the other hand, improved coverage by allowing for more flexible predictions and accounting for a broader range of possible outcomes, which also lead to lower RMSE. Figure 2 compares BCNN predictions with their true values, along with 95% credible intervals across a subset of test examples (where apparent clusters in the true values correspond to the five measured locations per target). We evaluate the credible intervals with respect to coverage (the percent of intervals that cover the true values). While coverage is near the nominal 95% level for some oxides (e.g., MgO in Figure 2), coverage is not always consistent, especially in oxides with higher intrinsic variability (e.g., SiO2) showing the models limitations to approximate the posterior distribution.

We also observed poor predictive performance and coverage for TiO2 (Figure 3). In this instance, a small number of targets exhibit exceptionally high TiO2 concentrations—over 45%—while most other targets contain less than 1% TiO2. Poorer performance for TiO2 suggests that our model performance is sensitive to high variability in sample distributions. We evaluated an eCNN for TiO2 which improved accuracy and coverage of predictions in our preliminary results which suggest they are reasonable for capturing model uncertainty while retaining high predictive accuracy (Figure 3).

Future work should focus on refining the model or employing additional strategies, such as data augmentation [15] [16] to improve coverage consistency across all oxides.

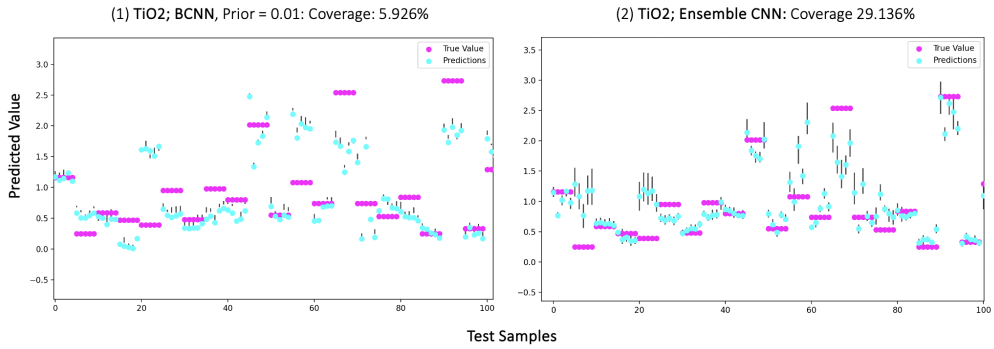

Figure 3: Predictions vs true values for TiO2 with 95% coverage intervals for (1) BCNN and (2) eCNN (Prior = 0.01)

5 Conclusion

We propose that BCNNs are particularly effective for spectral data analysis due to their principled capability for uncertainty quantification. We demonstrate our approach on a regression prediction task: prediction of oxide compositions from LIBS data for the ChemCam instrument on the Mars Curiosity rover. We show that BCNN (i) outperforms a commonly-used PLS model, (ii) remains competitive with the fully-connected NN and deterministic CNNs, and (iii) can effectively quantify predictive uncertainty in oxide compositions, though we leave to future work the task of identifying shortcomings of the model for some of the oxides. This work not only enhances confidence with complex models in analyzing highly variable spectroscopy data but also lays the groundwork for future research into broader applications of laser-induced breakdown spectroscopy.

6 Future Work

As part of our ongoing research, we are currently exploring several avenues to further optimize our performance in uncertainty quantification. We are also exploring the application of VI in modeling the posterior of activation space rather than weight space. Finally, we are actively researching explainability methods for probabilistic deep networks, with the aim of increasing model transparency and interpretability, which will be critical for collaborators wanting to use deep learning in ChemCam research.

References

[1] Charles Blundell, Julien Cornebise, Koray Kavukcuoglu, and Daan Wierstra. Weight uncertainty in neural network. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 1613–1622, Lille, France, 07–09 Jul 2015. PMLR.

[2] David M. Blei, Alp Kucukelbir, and Jon D. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518):859–877, April 2017.

[3] Erik Daxberger, Agustinus Kristiadi, Alexander Immer, Runa Eschenhagen, Matthias Bauer, and Philipp Hennig. Laplace redux – effortless bayesian deep learning, 2022.

[4] Durk P Kingma, Tim Salimans, and Max Welling. Variational dropout and the local reparameterization trick. *Advances in Neural Information Processing Systems*, 2015.

[5] S Maurice, RC Wiens, M Saccoccio, B Barraclough, O Gasnault, O Forni, N Mangold, David Baratoux, S Bender, G Berger, et al. The ChemCam instrument suite on the Mars Science

Laboratory (MSL) rover: Science objectives and mast unit description. *Space Science Reviews*, 170(1-4):95–166, 2012.

[6] Sylvestre Maurice, Samuel M Clegg, Roger C Wiens, O Gasnault, W Rapin, O Forni, Agnes Cousin, V Sautter, Nicolas Mangold, L Le Deit, et al. ChemCam activities and discoveries during the nominal mission of the Mars Science Laboratory in Gale crater, Mars. *Journal of Analytical Atomic Spectrometry*, 31(4):863–889, 2016.

[7] RC Wiens, S Maurice, J Lasue, O Forni, RB Anderson, S Clegg, S Bender, D Blaney, BL Barraclough, A Cousin, et al. Pre-flight calibration and initial data processing for the ChemCam laser-induced breakdown spectroscopy instrument on the Mars Science Laboratory rover. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 82:1–27, 2013.

[8] Samuel M Clegg, Roger C Wiens, Ryan Anderson, Olivier Forni, Jens Frydenvang, Jeremie Lasue, Agnes Cousin, Valerie Payre, Tommy Boucher, M Darby Dyar, et al. Recalibration of the Mars Science Laboratory ChemCam instrument with an expanded geochemical database. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 129:64–85, 2017.

[9] Ryan B Anderson, Olivier Forni, Agnes Cousin, Roger C Wiens, Samuel M Clegg, Jens Frydenvang, Travis SJ Gabriel, Ann Ollila, Susanne Schröder, Olivier Beyssac, et al. Post-landing major element quantification using supercam laser induced breakdown spectroscopy. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 188:106347, 2022.

[10] Junxi Chen, Jorge Pisonero, Sha Chen, Xu Wang, Qingwen Fan, and Yixiang Duan. Convolutional neural network as a novel classification approach for laser-induced breakdown spectroscopy applications in lithological recognition. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 166:105801, 2020.

[11] Fan Yang, Lu-Ning Li, Wei-Ming Xu, Xiang-Feng Liu, Zhi-Cheng Cui, Liang-Chen Jia, Yang Liu, Jun-Hua Xu, Yu-Wei Chen, Xue-Sen Xu, et al. Laser-induced breakdown spectroscopy combined with a convolutional neural network: a promising methodology for geochemical sample identification in Tianwen-1 Mars mission. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 192:106417, 2022.

[12] Lu-Ning Li, Xiang-Feng Liu, Wei-Ming Xu, Jian-Yu Wang, and Rong Shu. A laser-induced breakdown spectroscopy multi-component quantitative analytical method based on a deep convolutional neural network. *Spectrochimica Acta Part B: Atomic Spectroscopy*, 169:105850, 2020.

[13] Kate H Lepore, Caleb I Fassett, Elly A Breves, Sarah Byrne, Stephen Giguere, Thomas Boucher, J Michael Rhodes, Michael Vollinger, Chloe H Anderson, Richard W Murray, et al. Matrix effects in quantitative analysis of laser-induced breakdown spectroscopy (libs) of rock powders doped with cr, mn, ni, zn, and co. *Applied Spectroscopy*, 71(4):600–626, 2017.

[14] Kumar Shridhar, Felix Laumann, and Marcus Liwicki. A comprehensive guide to Bayesian convolutional neural network with variational inference. *Preprint at http://arxiv.org/abs/1901.02731*, 2019.

[15] Justin Salamon and Juan Pablo Bello. Deep convolutional neural networks and data augmentation for environmental sound classification. *IEEE Signal Processing Letters*, 24(3):279–283, March 2017.

[16] Yan Yu and Meibao Yao. When convolutional neural networks meet laser-induced breakdown spectroscopy: End-to-end quantitative analysis modeling of ChemCam spectral data for major elements based on ensemble convolutional neural networks. *Remote Sensing*, 15(13):3422, 2023.