

Title

Christian Bager Bach Houmann

Aalborg University
Aalborg, Denmark
chouma19@student.aau.dk

Patrick Frostholt Østergaard

Aalborg University
Aalborg, Denmark
pfas19@student.aau.dk

Ivik Lau Dalgas Hostrup

Aalborg University
Aalborg, Denmark
ihostr16@student.aau.dk

Abstract— Abstract here

ANN – Artificial Neural Network.

BPNN – Back-Propagation Neural Network.

ChemCam – Chemistry and Camera.

CNN – Convolutional Neural Network.

DF – Dominant Factor.

ENet – Elastic Net.

FCNN – Fully Connected Neural Network.

GBR – Gradient Boosting Regression.

K-ELM – Kernel Extreme Learning Machine.

KNN – K-Nearest Neighbors.

KSVR – Kernelized Support Vector Regression.

LASSO – Least Absolute Selection and Shrinkage Operator.

LDA – Linear Discriminant Analysis.

LIBS – Laser-Induced Breakdown Spectroscopy.

LR – Logistic Regression.

MAE – Mean Absolute Error.

MarSCoDe – Mars Surface Composition Detector.

MDS-BPNN – Multidimensional Scaling-back Propagation Neural Network.

MER – Mars Exploration Rover.

MOC – Multivariate Oxide Composition.

MSE – Mean Squared Error.

MSL – Mars Science Laboratory.

MLR – Multiple Linear Regression.

MLR – Multivariate Linear Regression.

OLS – Ordinary Least Squares.

OMP – Orthogonal Matching Pursuit.

PCA – Principal Component Analysis.

PCR – Principal Component Regression.

PLS – Partial Least Squares.

PLSR – Partial Least Squares Regression.

RF – Random Forest.

RMSE – Root Mean Squared Error.

RMSE-CV – Root Mean Squared Error of Cross-Validation.

RMSEP – Root Mean Squared Error of Prediction.

SVM – Support Vector Machine.

SVR – Support Vector Regression.

TAS – Total Alkali-Silica.

ULR – Univariate Linear Regression.

UMAP – Uniform Manifold Approximation and Projection.

I. INTRODUCTION

NASA has been studying the Martian environment for decades through a series of missions, including the Viking missions[1], the MER (*Mars Exploration Rover*) mission[2], [3], and the MSL (*Mars Science Laboratory*) mission[4], each building on the knowledge gained from the previous ones. Today, the rovers exploring Mars are equipped with sophisticated instruments for analyzing the chemical composition of Martian soil in search of past life and habitable environments.

Part of this research is facilitated through interpretation of spectral data gathered by LIBS (*Laser-Induced Breakdown Spectroscopy*) instruments, which fire a high-powered laser at soil samples to create a plasma. The emitted light is captured by spectrometers and analyzed using machine learning models to assess the presence and concentration of certain major oxides, informing NASA's understanding of Mars' geology.

However, predicting major oxide compositions from LIBS data presents significant computational challenges, including the high dimensionality and non-linearity of the data, compounded by multicollinearity and matrix effects[5]. These effects can cause the intensity of emission lines from an element to vary independently of that element's concentration, introducing unknown variables that complicate the analysis. Furthermore, due to the high cost of data collection, datasets are often small, which further complicates the task of building accurate and robust models.

Various machine learning models have been used to predict the composition of major oxides in the sample, including CNNs (*Convolutional Neural Network*) [6], [7], SVR (*Support Vector Regression*) [8], and hybrid models like DF (*Dominant Factor*)-K-ELM (*Kernel Extreme Learning Machine*) [9] that incorporate domain knowledge to enhance model interpretability and performance. However, the high dimensionality and multicollinearity of the spectral data remains a significant challenge for these models.

Building upon the baseline established in C. B. B. Houmann, P. F. Østergaard, and I. L. D. Hostrup [10], this thesis aims to explore approaches for tackling the challenges in predict-

ing major oxide compositions from LIBS data. We develop machine learning models that seek to enhance the accuracy and robustness of these predictions. We define accuracy as the ability of a model to predict the composition of major oxides in Martian geological samples, while robustness refers to the stability of these predictions across different samples and oxides.

We investigate various techniques to handle the high dimensionality, non-linearity, and small dataset size inherent in this problem, and evaluate the performance of these models using appropriate metrics. Through extensive experiments on LIBS data, we demonstrate the superior performance of our approach compared to existing methods in terms of both prediction accuracy and computational efficiency.

Our key contributions are as follows:

- We develop a novel machine learning pipeline that effectively handles the challenges of LIBS data to accurately predict major oxide compositions in Martian soil samples.
- We conduct a comprehensive evaluation of various dimensionality reduction techniques and machine learning models to identify the optimal combination for this task.
- We demonstrate the superior performance of our approach compared to existing methods through extensive experiments on LIBS data.

The remainder of this paper is organized as follows: Section background provides background on the ongoing Mars exploration missions, the LIBS technique, and the baseline MOC (*Multivariate Oxide Composition*) model. Section Section IV formally defines the problem addressed in this work. Section Section V describes our proposed methodology, including data preprocessing, dimensionality reduction, and machine learning models. Section Section VI presents our experimental setup and results. Finally, Section Section VII concludes the paper and discusses future work.

II. RELATED WORK

In addressing the challenge of predicting major oxide compositions from LIBS data, our investigation intersects with a broad spectrum of existing research that tackles similar computational hurdles, such as high dimensionality, multicollinearity, matrix effects, and the challenges of small datasets. This section outlines how prior works relate to our problem, drawing from a range of techniques and methodologies that offer potential pathways for enhancing the accuracy and robustness of our predictions.

R. B. Anderson *et al.* [11] experimented with different machine learning models for quantifying major oxides on Mars using the SuperCam instrument on the Mars 2020 Perseverance rover. They discuss preprocessing, normalization of LIBS spectra, and the development of multivariate regression models to predict major element compositions. For each oxide, they tested different models and selected the best perform-

ing one. In some cases, they used a blend of models to improve the predictions. The models they tested include: OLS (*Ordinary Least Squares*), PLS (*Partial Least Squares*), LASSO (*Least Absolute Selection and Shrinkage Operator*), Ridge, ENet (*Elastic Net*), OMP (*Orthogonal Matching Pursuit*), SVR, RF (*Random Forest*), GBR (*Gradient Boosting Regression*), and local ENet and blended submodels. For SiO₂, they used a blend of GBR and PLS models. Interestingly, they found that PLS performed better at longer distances (4.25m), but GBR was better at 3m. For TiO₂, they selected the RF model for its superior performance at 4.25m and overall lower RMSEP (*Root Mean Squared Error of Prediction*). For Al₂O₃, they used an average of predictions from four models (Local ENet, RF, two variants of PLS) to obtain the lowest RMSEP. For FeOT, they initially selected RF but later replaced it with GBR due to its more realistic stoichiometry predictions for high-Ca pyroxenes and overall performance. For MgO, they selected GBR for having the lowest RMSEP and avoiding negative predictions, despite slightly overpredicting MgO for high concentration samples. For CaO, they used a blend of RF and PLS to address the bimodal distribution of CaO predictions by the RF model alone. For Na₂O, they used a blend of GBR and LASSO models to utilize GBR’s accuracy at low concentrations and LASSO’s superior predictions at higher concentrations. For K₂O, they selected LASSO for its better performance on high K₂O samples, despite the averaged model of five algorithms showing slight improvements at lower concentrations. The findings of this paper are significant to us because they provide a benchmark for the performance of different machine learning models on LIBS spectra. We can use this information to guide our model selection and to compare our results with theirs. Additionally, we might want to try out different models from the ones they tested to see if we can improve the predictions further, or perhaps find a model that is more suitable for our specific use case. Also, SuperCam being the successor to ChemCam (*Chemistry and Camera*) means that the findings of this paper are directly relevant to our work.

W. Song, Z. Hou, W. Gu, M. S. Afgan, J. Cui, H. Wang, Y. Wang, and Z. Wang [9] present a novel approach to enhance the performance and interpretability of machine learning models in the context of LIBS quantification. The authors use “knowledge-based spectral lines, related to analyte compositions, to construct a linear physical principle based model and adopts K-ELM to account for the residuals of the linear model.” The method is based on DF and K-ELM and is called DF-gls{k-elm}. This method stands out by offering an intuitive explanation of how knowledge-based spectral lines impact prediction results, thereby enhancing model interpretability without compromising model complexity. DF-K-ELM was tested across 10 regression tasks based on 3 LIBS datasets, comparing its performance against six baseline methods using RMSEP as the evaluation metric. They have 3 coal datasets, and they do regression tasks involving carbon, ash, volatile matter, and heat value analysis. It achieved the best performance in 4 tasks and the second-best in 2 tasks, demonstrating its efficacy. In-

corporation of domain knowledge not only improved the accuracy of the models but also enhanced their generalizability across different tasks. The method's design allows for a more interpretable machine learning model that adheres closer to the physical principles underlying LIBS quantification. The integration of domain knowledge into machine learning models addresses two critical challenges: improving the interpretability of complex models and enhancing their performance by leveraging specific domain insights. The approach demonstrates a practical application of kernel extreme learning machines combined with domain-specific insights. This is particularly valuable in fields like spectroscopy, where understanding the relationship between the spectral data and the analyte concentration is vital. The DF-K-ELM method showcases how hybrid models can outperform traditional machine learning approaches. The approach demonstrates a practical application of kernel extreme learning machines combined with domain-specific insights. This is very relevant to our work considering interpretability is a key requirement for NASA and something they considered when choosing the PLS model for the ChemCam instrument.

M. Rezaei, F. Rezaei, and P. Karimi [8] explore a variety of statistical and machine learning methods, including MLR (*Multiple Linear Regression*), SVR, KSVR (*Kernelized Support Vector Regression*), and ANN (*Artificial Neural Network*), alongside their integrations with PCA (*Principal Component Analysis*) to reduce dimensionality and improve model performance. They use MSE (*Mean Squared Error*) and MAE (*Mean Absolute Error*) as evaluation metrics to compare the performance of the models. This paper clearly demonstrates the effectiveness of dimensionality reduction techniques in improving the performance of machine learning models because it compares the performance of many models with and without PCA: ANN, MLR, SVR, KSVR, PCA-gls{ann}, PCA-MLR, PCA-SVR, and PCA-KSVR. For all elements, a variant of SVR performs the best. For Si, SVR performs the best. For Zn, PCA-SVR performs the best. For the rest of the elements, PCA-KSVR performs the best. The superiority of KSVR is attributed to its ability to handle non-linear relationships in the data effectively, especially when combined with PCA's capability to compress and simplify the input data by focusing on the most relevant variations.

F. Yang, L.-N. Li, W.-M. Xu, X.-F. Liu, Z.-C. Cui, L.-C. Jia, Y. Liu, J.-H. Xu, Y.-W. Chen, X.-S. Xu, J.-Y. Wang, H. Qi, and R. Shu [6] present a study on the application of a deep CNN for classifying geochemical samples using LIBS, with a particular focus on planetary exploration missions such as China's Tianwen-1 Mars mission. The authors demonstrate the effectiveness of a deep CNN in classifying geochemical standard samples using LIBS spectra collected at varying distances. This addresses the challenge of spectral differences induced by distance, showcasing that CNN can learn to classify samples without the need for traditional spectral preprocessing or distance correction. Using a dataset of over 18,000 LIBS spectra from 39 geochemical standard samples, the study compares the CNN model's

performance against four other machine learning models: BPNN (*Back-Propagation Neural Network*) SVM (*Support Vector Machine*), LDA (*Linear Discriminant Analysis*), and LR (*Logistic Regression*). The CNN model exhibits superior classification accuracy, emphasizing its potential for geochemical sample identification/classification in planetary exploration. The paper includes a detailed comparative analysis, proving the CNN model's superior performance. With classification accuracies on the validation set for all models exceeding 95%, the CNN model demonstrated the highest overall accuracy. This was particularly evident as the training set size increased, indicating the model's robustness to varying distances without requiring distance correction. Statistical analysis further confirmed the CNN model's superiority, with higher average Ncorr values compared to other models. The CNN model's ability to accurately classify geochemical samples without preprocessing for distance correction is quite impressive. This is particularly relevant to our work because we are also working with LIBS spectra collected at varying distances. The comparative analysis underscores the CNN model as a best-fit approach for LIBS data analysis, potentially setting a new standard for future research and applications in the field.

G. Jeon, H. Keum, H. Lee, K. Oh, and J. Choi [12] investigated the effects of feature engineering on the robustness of LIBS for steel classification. They developed a remote LIBS system to classify six steel types, using various feature-engineering and machine learning algorithms, including SVM and FCNN (*Fully Connected Neural Network*), to handle different laser energies in test datasets. They found that using intensity ratios as input data resulted in more robust classification models. It was better than PCA and RF-based wavelength selection. The study highlights the importance of selecting appropriate feature engineering methods to improve model robustness, especially under varying measurement conditions. This is relevant to our project as it demonstrates how feature engineering can enhance the performance and robustness of models for classifying materials based on LIBS data, addressing challenges similar to those we face in predicting major oxide compositions.

The study by F. F. Fontana *et al.* [13]. explores using LIBS for whole-rock geochemical analysis, specifically for major elements like Al, Ca, Fe, K, Mg, Na, Si, and Ti. They averaged LIBS spot analyses over 1-mm spaced transects on drill core intervals, demonstrating strong correlations with lab-based geochemistry for elements like Si, Al, and Na. Different predictive models were used for each element, such as PLS, ENet, LASSO, and PCR (*Principal Component Regression*), showing varied RMSE-CV (*Root Mean Squared Error of Cross-Validation*) values indicating the precision of these models. This method's relevance to our work lies in its potential for rapid, in-situ geochemical analysis, offering a way to overcome challenges related to high dimensionality and non-linearity in LIBS data.

The paper by C. Sun *et al.* [14] introduces transfer learning to LIBS spectral data analysis for rock classification on Mars, significantly improving model performance. Previously, model's

els trained on laboratory standards (pellets) struggled with physical matrix effects when applied to natural rock spectra. Transfer learning, leveraging knowledge from one domain to address related problems in another, was applied to overcome this challenge. The method showed remarkable improvement in TAS (*Total Alkali-Silica*) classification accuracy for both polished and raw rock samples, with rates increasing from 25% and 33.3% to 83.3% respectively using machine learning models to 83.3% with the transfer learning model. This demonstrates the effectiveness of transfer learning in addressing the physical matrix effect and enhancing model robustness for rock classification on Mars.

H. Wang *et al.* [15] discusses an advanced methodology for analyzing stream sediments using remote LIBS combined with a MDS-BPNN (*Multidimensional Scaling-back Propagation Neural Network*) algorithm. This approach yielded highly accurate quantitative analyses of both major and trace elements, with determination coefficients (R^2) for major elements exceeding 0.9996 and for trace elements greater than 0.9837, and RMSE (*Root Mean Squared Error*) less than 0.73 (major elements). The study emphasizes the potential of remote LIBS technology, especially for identifying biominerals in geological samples, highlighting its significance for studying ancient planetary environments.

K. H. Lepore, C. R. Ytsma, and M. D. Dyar [16] provides an in-depth analysis of the effectiveness of using LIBS for geochemical analysis, focusing on the optimization of calibration datasets through the creation of submodels. It outlines the methodology for collecting and processing LIBS spectra, the development of multivariate models for predicting geochemical compositions, and compares the predictive accuracies of different submodel strategies. The study finds that while submodels can improve prediction accuracies under certain conditions, the overall effectiveness is contingent upon having a large and diverse calibration dataset. The research suggests that the optimal use of LIBS for geochemical analysis requires a balance between the specificity of submodels and the breadth of the calibration dataset to ensure accurate and reliable predictions.

E. Kepes, J. Vrabel, P. Pořizka, and J. Kaiser [17] discusses enhancing LIBS model accuracy using transfer learning between ChemCam and SuperCam instruments. It proposes a method where ChemCam data transforms to approximate SuperCam spectra, improving CNN regression models' performance. Key methods include data augmentation and fine-tuning of CNNs with pre-processed and normalized spectra. This approach outperforms some existing models for specific oxides, demonstrating transfer learning's potential in LIBS analyses for more accurate quantitative models.

M. F. S. Ferreira *et al.* [18] presents an extensive comparison of various algorithms for quantifying lithium in geological samples, with a focus on both linear and non-linear methods. The study tested algorithms on spectra acquired from a commercial handheld device and a laboratory prototype, highlighting the challenges in quantifying lithium due

to effects like saturation and matrix interference. The results showed that non-linear methodologies, such as KNN (*K-Nearest Neighbors*) regression, SVR, and ANN regression, generally outperformed linear methods by effectively managing saturation and matrix effects, which are common in geological samples. This research provides valuable insights for future applications in geological sample analysis and could potentially be generalized for other elements in similar contexts. The paper's findings are particularly relevant to our project as it demonstrates the effectiveness of non-linear machine learning techniques in handling complex, non-linear relationships in high-dimensional LIBS data, aligning with our research objectives of improving major oxide composition predictions from LIBS data.

The study by X. Liu *et al.* [19] explores the use of MarSCoDe (*Mars Surface Composition Detector*) LIBS for quantitative analysis of olivine in a simulated Martian atmosphere, focusing on multivariate analysis methods to address challenges posed by LIBS data, such as high dimensionality and multicollinearity. The methods evaluated include ULR (*Univariate Linear Regression*), MLR (*Multivariate Linear Regression*), PCR, PLSR (*Partial Least Squares Regression*), ridge regression, LASSO, ENet, and BPNN. The findings demonstrate the effectiveness of dimension reduction techniques, especially PLSR, and nonlinear analysis for improving quantitative analysis accuracy of olivine using LIBS data. This approach is particularly relevant to our work due to the focus on advanced statistical methods and machine learning algorithms for handling complex, high-dimensional LIBS data, aligning with our objectives of improving accuracy and robustness in predicting major oxide compositions.

F. Yang, W. Xu, Z. Cui, X. Liu, X. Xu, L. Jia, Y. Chen, R. Shu, and L. Li [7] is a study where a CNN model is designed to identify twelve types of rocks using LIBS data from the MarSCoDe for the Tianwen-1 Mars exploration mission. The classification performance of the CNN is compared with LR, SVM, and LDA. The CNN model achieved the highest classification accuracy, demonstrating its efficiency in rock identification with LIBS spectra collected in a simulated Martian environment. This indicates that CNN-supported LIBS classification is a promising analytical technique for planetary exploration missions.

N. A. Silva *et al.* [20] introduce clustered regression calibration algorithms for LIBS to address quantification challenges in complex sample matrices or wide concentration ranges, focusing on lithium quantification in geology. They employ unsupervised clustering to group similar samples before applying a tailored linear calibration model to each cluster. This approach, tested on lithium in exploration drills, outperforms standard linear models, especially in lower concentration regions, and demonstrates good generalizability to unseen data from different rock veins. The study highlights the potential of clustered regression methods in improving LIBS quantification accuracy and robustness, particularly valuable in mining environments. Uses Ridge regression, PLS, dimensionality reduction with UMAP (*Uniform Manifold Approximation and*

Projection) and then k-means clustering, followed by a local model for each cluster, univariate calibration curves.

S. Wold, K. Esbensen, and P. Geladi [21] provides an in-depth tutorial on PCA, a fundamental method in multivariate data analysis used for dimensionality reduction, outlier detection, and uncovering the underlying structure in data sets. The paper covers the history and development of PCA, its mathematical foundations, applications across various fields, and detailed instructions for data pre-treatment and PCA implementation. It emphasizes PCA's utility in simplifying complex data sets, enhancing interpretability, and supporting data analysis through the extraction of principal components that capture the most variance in the data. This work is particularly relevant to our project as it underscores the importance of PCA in handling high-dimensional data, which aligns with our objectives of improving the prediction accuracy and interpretability of major oxide compositions from LIBS data by effectively managing multicollinearity and high dimensionality challenges.

D. Bank, N. Koenigstein, and R. Giryas [22] details various autoencoder architectures and their applications in machine learning, emphasizing their role in dimensionality reduction, feature learning, and generative models. It explores regularized, denoising, and variational autoencoders, highlighting their advantages in compressing data into lower-dimensional spaces while retaining essential information for reconstruction. Autoencoders could significantly enhance our ability to handle the high-dimensional nature of LIBS data. By compressing spectral data into more manageable representations without losing critical information, we can improve model performance, especially in predicting major oxide compositions, by focusing on the most relevant features extracted from the compressed data representation. This aligns with our objectives of efficient dimensionality reduction and robust predictive modeling.

In their work R. Caruana [23], presents a method called *Multitask learning*, which is a method of learning machine models on several related datasets. The motivation for this approach stems from the assumption that utilizing multiple, albeit related, datasets can enhance a model's ability to discern patterns and shapes within the data. R. Caruana [23] suggests that leveraging shared representations for model training can enable the model to identify underlying attributes in other datasets, even when this new data is small. This is relevant for our work as one of the major challenges in analyzing LIBS calibration data for Mars is the scarcity of available data. This scarcity makes it difficult to construct robust models capable of comprehensively understanding the underlying patterns and physical principles within the data. Utilizing related LIBS data could help the models first learn the general outline, shape and patterns in the LIBS data, making it easier for it to grasp the deeper patterns in the Mars related data.

III. BACKGROUND

IV. PROBLEM DEFINITION

V. METHODOLOGY

VI. EXPERIMENTS

VII. CONCLUSION

REFERENCES

- [1] mars.nasa.gov, "Viking 1 & 2 | Missions." Accessed: Jan. 23, 2024. [Online]. Available: <https://mars.nasa.gov/mars-exploration/missions/viking-1-2>
- [2] mars.nasa.gov, "Mars Observer | Missions." Accessed: Jan. 23, 2024. [Online]. Available: <https://mars.nasa.gov/mars-exploration/missions/mars-observer>
- [3] mars.nasa.gov, "Mars Exploration Rovers | Missions." Accessed: Jan. 23, 2024. [Online]. Available: <https://mars.nasa.gov/mars-exploration/missions/mars-exploration-rovers>
- [4] mars.nasa.gov, "Mars Science Laboratory | Missions." Accessed: Mar. 12, 2024. [Online]. Available: <https://mars.nasa.gov/mars-exploration/missions/mars-science-laboratory>
- [5] R. B. Anderson *et al.*, "Improved Accuracy in Quantitative Laser-Induced Breakdown Spectroscopy Using Sub-Models," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 129, pp. 49–57, Mar. 2017, doi: 10.1016/j.sab.2016.12.002.
- [6] F. Yang *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*, vol. 192, p. 106417–106418, Jun. 2022, doi: 10.1016/j.sab.2022.106417.
- [7] F. Yang *et al.*, "Convolutional Neural Network Chemometrics for Rock Identification Based on Laser-Induced Breakdown Spectroscopy Data in Tianwen-1 Pre-Flight Experiments," *Remote Sensing*, vol. 14, no. 21, p. 5343–5344, Jan. 2022, doi: 10.3390/rs14215343.
- [8] M. Rezaei, F. Rezaei, and P. Karimi, "Using various machine learning algorithms for quantitative analysis in LIBS technique."
- [9] W. Song *et al.*, "Incorporating domain knowledge into machine learning for laser-induced breakdown spectroscopy quantification," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 195, p. 106490–106491, Sep. 2022, doi: 10.1016/j.sab.2022.106490.
- [10] C. B. B. Houmann, P. F. Østergaard, and I. L. D. Hostrup, "Identifying Limitations in the ChemCam Multivariate Oxide Composition Model for Elemental Quantification in Martian Geological Samples," Jan. 2024, [Online]. Available: https://vbn-aau-dk.zorac.aub.aau.dk/ws/files/659161040/p9_report_31_01_24.pdf
- [11] R. B. Anderson *et al.*, "Post-Landing Major Element Quantification Using SuperCam Laser Induced Breakdown Spectroscopy," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 188, p. 106347–106348, Feb. 2022, doi: 10.1016/j.sab.2021.106347.
- [12] G. Jeon, H. Keum, H. Lee, K. Oh, and J. Choi, "Effects of Feature Engineering on the Robustness of Laser-Induced Breakdown Spectroscopy for Industrial Steel Classification," *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 212, p. 106857–106858, Feb. 2024, doi: 10.1016/j.sab.2024.106857.

- [13] F. F. Fontana *et al.*, “Laser Induced Breakdown Spectroscopy (LIBS) for Whole Rock Geochemistry,” *Journal of Geochemical Exploration*, vol. 246, p. 107160–107161, Mar. 2023, doi: 10.1016/j.gexplo.2023.107160.
- [14] C. Sun *et al.*, “From Machine Learning to Transfer Learning in Laser-Induced Breakdown Spectroscopy Analysis of Rocks for Mars Exploration,” *Scientific Reports*, vol. 11, no. 1, p. 21379–21380, Nov. 2021, doi: 10.1038/s41598-021-00647-2.
- [15] H. Wang *et al.*, “Determination of Elemental Composition and Content in Stream Sediments by Laser-Induced Breakdown Spectroscopy,” *Chemosensors*, vol. 11, no. 7, 2023, doi: 10.3390/chemosensors11070377.
- [16] K. H. Lepore, C. R. Ytsma, and M. D. Dyar, “Quantitative Prediction Accuracies Derived from Laser-Induced Breakdown Spectra Using Optimized Multivariate Submodels,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 191, p. 106408–106409, May 2022, doi: 10.1016/j.sab.2022.106408.
- [17] E. Kepes, J. Vrabel, P. Pořízka, and J. Kaiser, “Improving Laser-Induced Breakdown Spectroscopy Regression Models via Transfer Learning,” *Journal of Analytical Atomic Spectrometry*, vol. 37, Jan. 2022, doi: 10.1039/D2JA00180B.
- [18] M. F. S. Ferreira *et al.*, “Comprehensive Comparison of Linear and Non-Linear Methodologies for Lithium Quantification in Geological Samples Using LIBS,” *Spectrochimica Acta Part B: Atomic Spectroscopy*, vol. 195, p. 106504–106505, Sep. 2022, doi: 10.1016/j.sab.2022.106504.
- [19] X. Liu *et al.*, “Comparison on Quantitative Analysis of Olivine Using MarSCoDe Laser-Induced Breakdown Spectroscopy in a Simulated Martian Atmosphere,” *Remote Sensing*, vol. 14, no. 21, p. 5612–5613, Jan. 2022, doi: 10.3390/rs14215612.
- [20] N. A. Silva *et al.*, “Towards Robust Calibration Models for Laser-Induced Breakdown Spectroscopy Using Unsupervised Clustered Regression Techniques,” *Results in Optics*, vol. 9, p. 100245–100246, Dec. 2022, doi: 10.1016/j.rso.2022.100245.
- [21] S. Wold, K. Esbensen, and P. Geladi, “Principal Component Analysis,” *Chemometrics and Intelligent Laboratory Systems*, vol. 2, no. 1, pp. 37–52, Aug. 1987, doi: 10.1016/0169-7439(87)80084-9.
- [22] D. Bank, N. Koenigstein, and R. Giryes, “Autoencoders.” Accessed: Feb. 26, 2024. [Online]. Available: <http://arxiv.org/abs/2003.05991>
- [23] R. Caruana, “Multitask learning,” *Machine Learning*, vol. 28, no. 1, pp. 41–75, 1997, doi: 10.1023/A:1007379606734.