Data Augmentation and Innovative Machine Learning Approaches for Classifying EEL Spectra of Transition Metals Oxides

Daniel Del Pozo Bueno[1], Dr. Demie Kepaptsoglou[2], Prof. Quentin M. Ramasse[3], Prof. Francesca Peiró[1], Prof. Sònia Estradé[1]
[1]LENS-MIND, Dept. d'Enginyeria Electrònica i Biomèdica and Institute of Nanoscience and Nanotechnology (IN2UB), Universitat de Barcelona, Barcelona, [2]School of Physics, Engineering and Technology, University of York & SuperSTEM Laboratory, Heslington - Daresbury, North Yorkshire - Cheshire, [3]Schools of Chemical and Process Engineering & Physics and Astronomy, University of Leeds & SuperSTEM Laboratory, Leeds - Daresbury, West Yorkshire - Cheshire

Background including Aims:
The field of Scanning Transmission Electron Microscopy (STEM), especially through the application of Electron Energy Loss Spectroscopy (EELS), has experienced significant advancements due to technical innovations such as aberration correctors, direct detectors, and increased computing power. These advancements have facilitated the collection of large and complex datasets, highlighting the need for effective data management and analysis solutions in the STEM community. Machine Learning (ML), with its two main categories of supervised and unsupervised learning, has emerged as a crucial tool for addressing challenges in EELS, including classification, clustering of spectrum images, and denoising tasks [1,2].

However, the effective application of supervised ML is often hindered by the requirement for large, labeled datasets, which are difficult to acquire due to the susceptibility of samples to electron beam damage. Addressing this drawback, this study aims to compare the effectiveness of two supervised ML techniques: soft-margin Support Vector Machines (SVM) and Artificial Neural Networks (ANN) in classifying EEL spectra for the determination of oxidation states in transition metal oxides, particularly focusing on iron and manganese oxides. Additionally, we present a novel unsupervised learning approach that employs Generative Adversarial Networks (GANs) for data augmentation to address the problem of labeled data scarcity and improve the precision and effectiveness of oxidation state detection using EELS.

Methods:
This research presents a comparative analysis of soft-margin SVMs and ANNs, adapted to the specific challenges in EELS data classification [3]. We evaluate these classifiers based on their ability to accurately identify features indicative of oxidation states, such as white lines and the oxygen K edge, and the effect of energy shifts or noisy spectra. To enhance the classifiers' robustness and adaptability, we investigate the impact of incorporating energy-shifted

spectra into the training process and explore various normalization methods, including maximum and L2 norms. The latter is analyzed by dimensionality reduction techniques, particularly Uniform Manifold Approximation and Projection. Additionally, we undertake a systematic exploration of ANN architectures through Random Search and Tree-structured Parzen Estimator (TPE) algorithms, aiming to pinpoint the most effective combinations of architecture and parameters for EELS data classification. Finally, we include the innovative use of GANs for data augmentation, which enables the generation of synthetic EEL spectra from a reduced set of experimental data. This approach not only amplifies the diversity and volume of available training data but also reduces the dependency on extensive, experimentally acquired datasets.

Results:
Based on white lines as particularly reliable indicators for oxidation state classification, SVMs are very robust against energy shifts for the EEL features under investigation, and they perform even better when trained on energy-shifted spectra. In comparison, ANNs, especially those employing convolutional layers, demonstrate a superior ability to adapt to the complexities of EEL spectra, achieving a level of precision comparable to the best SVMs. The analysis of normalization techniques and the strategic use of the cosine kernel in SVMs emerge as effective strategies for avoiding normalization while keeping classification accuracy. Finally, the use of GANs for data augmentation marks a pivotal advancement, since this approach generates synthetic data that closely mirrors the variability and complexity of large experimental data collections, facilitating the training of these classifiers to be both more accurate and more generalized, capable of adapting to the diverse spectra encountered in EELS analysis.

Conclusions:
This work not only elucidates the comparative advantages of SVMs and ANNs in the classification of EEL spectra but also introduces a groundbreaking strategy for overcoming the challenges imposed by the limited availability of labeled datasets. SVMs are particularly recommended for simpler classification tasks where data volume is limited, offering an efficient solution that does not compromise performance. On the other hand, ANNs are more suited to tackling complex classification problems that involve larger datasets, benefiting from their enhanced capacity for learning and adaptation. The successful integration of GANs for data augmentation represents a significant advance, substantially reducing the reliance on extensive labeled datasets and paving the way for more efficient and effective classifier training.

Acknowledgments:
This work has been supported by the Spanish projects PDC2021-121366-I00, PID2019-106165GB-C21 and PID2022-138543NB-C21 financed by MCIN/AEI/ 10.13039/501100011033 and by the European Union

NextGenerationEU/PRTR. FP acknowledges the ICREAAcademia 2022 grant from the Government of Catalonia. The authors also acknowledge the support received from the ELECMI - ICTS Electron Microscopy for Materials Science, the funding from Generalitat de Catalunya under project 2021SGR00242, the 2020 FI-SDUR grant from the AGAUR agency, and SuperSTEM acknowledges the Engineering and Physical Sciences Research Council (EP/W021080/1).

Keywords:

ML, EELS, GAN, SVM, ANN

Reference:

[1] del-Pozo-Bueno, D.; Peiró, F.; Estradé, S. Ultramicroscopy 2021, 221, 113190. https://doi.org/10.1016/j.ultramic.2020.113190.
[2] Blanco-Portals, J.; Peiró, F.; Estradé, S. Microscopy and Microanalysis 2022, 28 (1), 109–122. https://doi.org/10.1017/S1431927621013696.
[3] del-Pozo-Bueno, D.; Kepaptsoglou, D.; Peiró, F.; Estradé, S. Ultramicroscopy 2023, 253. https://doi.org/10.1016/j.ultramic.2023.113828.