2023 13th Workshop on Hyperspectral Imaging and Signal Processing: Evolution in Remote Sensing (WHISPERS) | 979-8-3503-9557-0/23/$31.00 ©2023 IEEE | DOI: 10.1109/WHISPERS61460.2023.10431067

# MATERIAL IDENTIFICATION IN COMPLEX ENVIRONMENTS: NEURAL NETWORK APPROACHES TO HYPERSPECTRAL IMAGE ANALYSIS

*Jason Brown*[1,†], *Bohan Chen*[1,†], *Harris Hardiman-Mostow*[1,†], *Adrien Weihs*[1,‡], *Andrea L. Bertozzi*[2,†], *Jocelyn Chanussot*[2,§]

[†]Univ. of California, Los Angeles, Dept. of Mathematics, Los Angeles, CA 90095, USA.
[‡]University of Manchester, School of Mathematics, Manchester, M13 9PL, UK.
[§]University Grenoble Alpes, CNRS, Grenoble INP, GIPSA-Lab, 38000 Grenoble, France.

## ABSTRACT

Hyperspectral imagery is often used in chemometric studies for quality sorting and recycling due to its ability to produce rich spectroscopy data. In this paper, we study plastics detection in a complex environment. In particular, we analyze hyperspectral images of three scenes with spectra in the near-infrared and visible wavelength ranges; our task is to detect plastic within the scenes. The images contain materials with high intraclass variability and significant mixing. Our novel contribution compares various methods for hyperspectral pixel classification in these complicated, real-world environments, specifically deep methods such as contrastive learning and autoencoders, as well as comparing the viability of the hyperspectral cameras' light spectrum for the application of plastic detection.

***Index Terms***— Chemometrics, Spectroscopy, Hyperspectral Imagery, Plastic, Recycling, AutoEncoder, Contrastive Learning, Graph Learning

## 1. INTRODUCTION

Recycling of waste material is an important component of the United Nations' Sustainable Development Goals [1]. Chemometric study of spectroscopy data obtained through Hyperspectral Imaging (HSI) has proven to be a popular framework in quality sorting and recycling tasks, such as classification of polymers [2, 3, 4, 5], classification of food samples [6], contamination identification for plastic recycling [3], and waste identification in copper ore processing [7]. Similarly, we are interested in identifying plastic samples in a tray of waste materials of various chemical compositions. In particular, we want to perform pixel-wise classification on the HSI data of Figure 1d, which contains plastic among other materials.

The novelties of this paper are three-fold. First, while most of the mentioned papers consider spectrum data in the near infrared (NIR) wavelengths (900nm to 1700nm), we aim to compare classification results obtained from the NIR and visible (VIS) wavelengths (400nm to 1000nm). To this end, we present results on two different hyperspectral cameras, which image in the NIR and VIS wavelengths respectively. From an industrial point of view, such results can assist in the choice of the image acquisition setup.

Second, the spectroscopy data in the referenced papers is acquired in a very controlled environment: the samples are well-separated and placed on a uniform white background, allowing for straight-forward analysis of the spectral data. Furthermore, the typical framework assumes expert knowledge of the materials and thus precise definition of the classes in the classification task. However, in our dataset, samples are cluttered on a piece of cardboard with various interfering, partly overlapping elements and more realistic, varied lighting. The dataset is also only vaguely labelled, which leads to a broad definition of the plastic class (see Section 2 for details). The resulting intraclass variability presents a challenge to traditional chemometric classification approaches. This also reflects the more realistic industrial setting where recylcing material is placed directly on a conveyer belt without prearrangement, manual or chemical sorting, and tedious pixelwise labelling.

Third, chemometrics in most of the spectroscopy literature are based on linear methods (see Section 3). Recently, there has been more deep machine learning methods used in these applications [4, 5, 8]. We continue this trend by presenting novel results using contrastive learning, as well as autoencoders combined with graph learning. To the best of our knowledge, this has not previously appeared in the literature.

The rest of the paper is structured as follows: in Section 2, we describe our dataset, data calibration and data exploration;

JB, HHM, and ALB are supported by US National Science Foundation grants DMS-2318817, DMS-2152717, DMS-1952339, and DMS-2027277. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. BC is supported by UC-National Lab In-Residence Graduate Fellowship Grant number L21GF3606. J. Chanussot is supported by MIAI@Grenoble Alpes (ANR-19-P3IA-0003). AW acknowledges support from European Union Horizon 2020 research and innovation programmes under the Marie Skłodowska-Curie grant agreement No. 777826 (NoMADS).

in Section 3 we survey our methods for the classification task; in Section 4 we discuss our results; in Section 5 we summarize our findings and offer future avenues of work.

## 2. DATASET

The dataset consists of six images derived from three scenes imaged by two cameras: Specim FX10 and Specim FX17. The imaging setup is shown in Figure 1a. A rail moves the camera across while two sources of light minimize shadows. The Specim FX10 camera operates in the visible wavelengths (400nm to 1000nm) and the Specim FX17 camera operates in the near infrared wavelengths (900nm to 1700nm). FX10 images 448 evenly spaced wavelengths and FX17 images 224 evenly spaced wavelengths. The two cameras imaging the same scenes allows us to compare the classification results between the different wavelengths.

The three scenes contain two training images and one testing image. The different materials within the image have been very roughly labelled. The first training image (Figure 1b) contains plastic samples on a white, paper background. For FX10, this image is $700 \times 400$; for FX17, this image is $730 \times 320$. The second training image (Figure 1c) contains cluttered scraps of non-plastic material in a cardboard box. The non-plastic materials include copper, fabrics, stones, paper, and metal. For FX10, this image is $1120 \times 570$; for FX17, this image is $1125 \times 480$. The testing image (Figure 1d) contains the same cluttered scraps as the training image, but with the addition of plastics. For FX10, this image is $1135 \times 580$; for FX17, this image is $1130 \times 480$. Despite the images sharing some samples, the different imaging conditions help our methods to not dramatically overfit and our results still reveal meaningful insights about the data and optimal classification methods.

### 2.1. Calibration

To calibrate the Specim cameras, we are provided with a dark image and a reference image. The dark image is taken when the lens is closed and captures the noise inherent the sensors in the camera. The reference image contains a material that has a reflectance of 99%. This is used to normalize the measurements with respect to the lighting source. Using the dark and reference images, we calibrate our images via the industry standard equation, discussed in [2].

### 2.2. Data Preprocessing

Light scattering can cause significant variability in the captured spectra both in diffuse reflectance and transmittance spectroscopy. For our application, this is particularly relevant: diffusively reflected light not only contains information about the chemical content of our samples but also about its micro-structure which causes the scattering (e.g. surface roughness, density fluctuations) [9]. The latter can be modeled using physical models and produces both multiplicative and additive interference in the spectra [10].

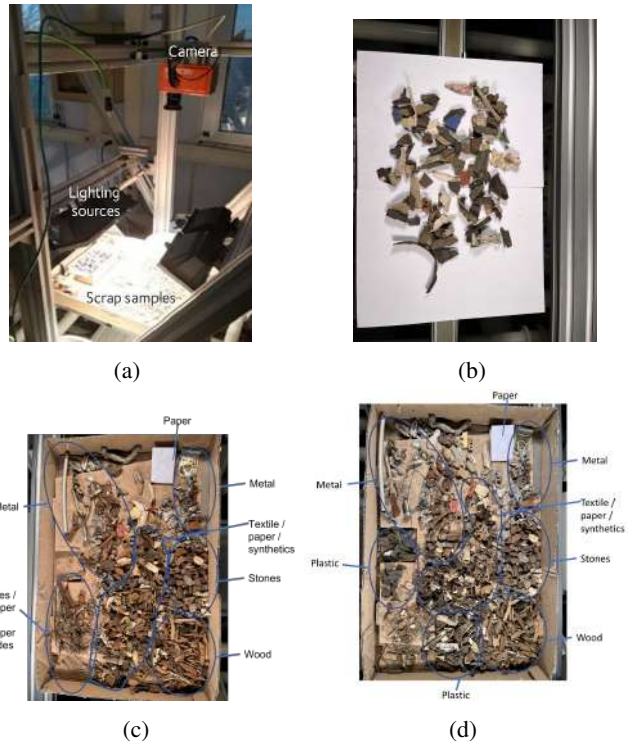


**Fig. 1**: The imaging apparatus and RGB images of different scenes. Panels (a): imaging apparatus; (b): plastics image; (c): mixed image with no plastic; (d): mixed image with plastic. The HSI corresponding to panel (b), (c) are used for training while the HSI of panel (d) is for test.

While deep learning methods are capable of dealing with this variability, for classical linear baseline methods and data exploration (see Sections 2.3 and 3), preprocessing of the spectra is necessary. Preprocessing is highly specific to the dataset [3, 11]. After considering Mutiplicative Scattering Correction, Standard Normal Variate and Spectral Derivative methods (see [9] for a review), we decided to use the Savitzky-Golay (SG) filter [12]. Using a moving window, the filter applies smoothing to the spectra by performing polynomial approximation. This approximation allows one to take derivatives of the spectra easily. We will write SG-0 for the smoothed spectra and SG-1 for the first derivative of SG-0. Due to the moving window, both the first and last bands of the spectrum can be discarded.

### 2.3. Data Exploration

In order to use supervision for pixel classification of the testing image (Figure 1d), we need to define Regions Of Interest (ROIs) to train and evaluate our algorithm on. To obtain pure plastic samples, we remove the background of the plastic image (Figure 1b) by thresholding on specific bands [6]. For

non-plastic samples, we use all of the pixels in Figure 1c as negative training samples, as the scene does not contain any plastic. We also note that removing the background in Figure 1d and defining a ROI is significantly harder than in some of the datasets considered in [2, 3, 6] due to its complexity. Using only an unmixing method on materials with high intraclass variability - such as in this recycling setting - may fail to identify ROIs with plastic. To avoid excluding plastic pieces from the ROI, we consider the whole scene as our testing ROI. This also makes our methods applicable to real-world settings where defining an ROI may not be realistic.

By manually sampling the plastic ROI in Figure 1b, we observe that the plastic pieces may be composed of different polymers. This implies large intraclass variabilty in the plastic and reflects recycling in practice, where one might be interested in various plastics types (and potentially in plastics with chemical compositions not apparent in the training dataset). This motivates deep learning methods, whose larger capacity can capture intraclass variability more accurately than classical methods. The large variability in our classification task is in contrast to the usual setting explored in [2, 3, 6] where the target classes are known in advance and well-defined.

## 3. METHODS

In this section, we survey the different methods used for the detection of plastic in the mixed image. The methods vary from linear to highly nonlinear. We first describe Partial Least Squares Discriminant Analysis (PLS-DA) which we use as our baseline. Following the motivations detailed in Sections 2.2 and 2.3 we also explore deep learning methods. In particular, we perform dimensionality reduction through autoencoders (AEs) and constrastive learning (CL); our classifiers include multilayer perceptrons (MLP), $k$-nearest neighbors (k-NN), and graph learning (GL). Unsupervised autoencoders with semi-supervised graph learning classifier were chosen as the low-label rate model while the supervised contrastive learning with k-NN and separate MLP classifier were chosen as the supervised methods.

### 3.1. Partial Least Squares Discriminant Analysis

We refer to [13] for a review of the methods discussed in this section. A standard assumption in multivariate linear regression is that the variables are uncorrelated. This is not satisfied when using spectra and we therefore aim to transform our data into a small number of orthogonal vectors. Unsupervised Principal Component Analysis achieves this by projecting data onto principal components which point in relevant directions based on the variance of the spectra. Performing regression using the principal components as variables is called Principal Component Regression (PCR). The supervised counterpart to PCR is PLS-DA: the directions of the orthogonal vectors used in the regression problem now maximize covariance with the response variable. Regression yields a continuous response prediction and we threshold at 0.5 to obtain labels in $\{0, 1\}$ (for binary classification). PLS-DA is the standard classification method used in spectroscopy [6, 3, 2] and in our experiments, we will use 3 orthogonal vectors for PLS-DA.

### 3.2. Nonlinear dimensionality reduction

As explained in Section 3.1, dimensionality reduction is essential when dealing with spectral data. We now detail two deep learning approaches for this task.

#### *3.2.1. Autoencoder*

Autoencoders are a deep, unsupervised neural network structure that seek to learn a lower-dimensional representation of input data. This can be helpful for denoising, learning latent features, or other downstream tasks. AEs have an *encoder* step and a *decoder* step; the output from the encoder is the input to the decoder. A loss is computed between the original input and the reconstructed output from the decoder. This encourages the network to learn a good encoded (or *latent*) representation of the input data, which can lead to insights about the data. We refer the reader to [14] for a more thorough review of deep learning concepts.

We utilize an AE with an input dimension of 448 or 224 (for FX10 or FX17, respectively), followed by fully connected layers with output dimension 100, 25, and 5, respectively, before fully connected layers with output dimension 25, 100, 448, respectively. Rectified linear unit (ReLU) activation functions are used after each layer except at the bottleneck. We train the network for 50 epochs. We use the Adam optimizer [15] with a learning rate of $0.001$, and $\beta$ values of $0.9$ and $0.999$. The loss is mean squared error. To classify the pixels, we perform graph learning (see Section 3.3.1) on the embeddings learned by the AE.

#### *3.2.2. Contrastive Learning*

Contrastive learning is a successful and recently popularized neural network architecture that, with some supervision, contrasts images within a batch to learn their similarities and differences. In this work, we use SupCon [16], a fully supervised contrastive learning with the goal of learning latent representations for samples so that classes are well clustered.

Typically when using contrastive learning, many data augmentations are used to create more varied data and teach the neural network to be invariant to certain effects. Due to the spectral nature of the data, it is less clear which augmentations are natural to use. Moreover, we have an abundance of pixels, so we instead do not use any augmentations and rely on the noise in the pixels to provide natural variation.

We train a simple multilayer perceptron encoder network with input dimension of 448 or 224 (for FX10 or FX17 re-

spectively) followed by two linear layers with sizes 256 or 128 (for FX10 and FX17 respectively) and 64 with ReLU activation functions. We then have a dropout layer with dropout probability of 50% and then one more linear layer of size 16 with a sigmoid activation function. The projection head is a linear layer of size 4 with a normalization activation function. We train the models for 500 epochs using the Adam optimizer with a learning rate of 0.001, beta values of 0.9 and 0.999, and a gamma value of 0.99 for the scheduler. For the SupCon loss function, we use a temperature of 0.1.

Many of the non-plastic samples have identical latent representations, which makes graph learning (see Section 3.3.1) ill-posed when the number of nearest neighbors is not large enough. For classification, we instead opt to use a k-NN classifier with $k = 5$.

### 3.3. Nonlinear classifiers

In order to go beyond multivariate linear classification (or its variants), we now describe two nonlinear classifiers.

#### *3.3.1. Graph Learning*

Given a set of feature vectors $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N\} \in \mathbb{R}^d$, we construct a graph $G = (\mathcal{X}, W)$, where the angular similarity weight matrix $W$ is computed by k-NN. We define the graph Laplacian $L$ by $L = D - W$, where $D$ is a diagonal matrix with diagonal entries $d_1, d_2, \ldots, d_N$, $d_i = \sum_{j=1}^{N} W_{ij}$. Based on ground truth labels of nodes in the subset $\hat{\mathcal{X}} \subset \mathcal{X}$, Laplace learning [17] predicts labels for $\mathcal{X} \setminus \hat{\mathcal{X}}$ by solving:

$$\begin{aligned} \hat{U} = &\underset{U \in \mathbb{R}^{N \times n_c}}{\arg\min} \frac{1}{2} \langle U, LU \rangle_F, \\ \text{s.t.} \quad &U_i = \mathbf{y}_i^\top, \forall \mathbf{x}_i \in \hat{\mathcal{X}}, \end{aligned} \tag{1}$$

where $n_c$ is the number of different classes, $\mathbf{y}_i$ is the one-hot ground-truth label of $\mathbf{x}_i \in \hat{\mathcal{X}}$, and $U_i$ is the $i^{\text{th}}$ row of the matrix $U$. The predicted class label for $\mathbf{x}_i \in \mathcal{X} \setminus \hat{\mathcal{X}}$ is given by $y_i = \arg\max U_i$. For our experiments, we set $k = 10$, and we use 1000 non-plastic samples and 500 plastic samples as training points to classify the pixels in the testing image according to Equation 1.

#### *3.3.2. MLP Classifier*

We train a 3-layer fully connected neural network for binary classification on the data. The network has hidden layers of size 50 and 10 with ReLU activation functions. We use cross entropy loss. The network uses the same hyperparameters as the AE network. We evaluate the network by performing classification on the mixed scene with plastic (Figure 1d).

## 4. RESULTS

We present classification results on both FX10 and FX17 data in Figure 2. Since we have no pixel-wise ground truth, our evaluation is qualitative. It is clear that the deep methods outperform PLS-DA, the more classical method, in detecting additional plastic, but it is also worth noting that there are notably more false positive pixels in the final result.

## 5. CONCLUSION AND FUTURE WORK

Using machine learning and artificial intelligence to automate the process of material sorting using hyprspectral cameras shows promise. We have shown, that with relatively limited data, plastic can be detected readily with a variety of methods in both the visible and near infrared spectrums. We find that classification results are stronger for images taken with the FX17 camera, indicating that the near infrared (900nm-1700nm) range is suitable for differentiating plastics from other non-plastic refuse, however the visible spectrum does appear to be viable as well.

It would be valuable to further explore the distinctions between the methods and light spectra with a meticulously crafted, larger dataset containing expert knowledge of the types of plastics.

## 6. REFERENCES


[1] B. X. Lee, F. Kjaerulf, S. Turner, L. Cohen, P. D. Donnelly, R. Muggah, R. Davis, A. Realini, B. Kieselbach, L. S. MacGregor, *et al.*, "Transforming our world: implementing the 2030 agenda through sustainable development goal indicators," *Journal of public health policy*, vol. 37, pp. 13–31, 2016.

[2] T. Arnold, M. De Biasio, R. Kammari, and K. Sayar-Chand, "Development of vis/nir hyperspectral imaging system for industrial sorting applications," in *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXVII*, vol. 11727, pp. 298–304, SPIE, 2021.

[3] G. Bonifazi, L. Fiore, R. Gasbarrone, R. Palmieri, and S. Serranti, "Hyperspectral imaging applied to weee plastic recycling: A methodological approach," *Sustainability*, vol. 15, no. 14, p. 11345, 2023.

[4] A. Sbrana, A. G. de Almeida, A. M. de Oliveira, H. S. Neto, J. P. C. Rimes, and M. C. Belli, "Plastic classification with nir hyperspectral images and deep learning," *IEEE Sensors Letters*, vol. 7, no. 1, pp. 1–4, 2023.

[5] B. Delaporte, T. van Gelder, and K. van der Sluis, "Polymer flake detection through hyperspectral imaging,"

[6] D. Hong, N. Yokoya, J. Chanussot, and X. X. Zhu, "An augmented linear mixing model to address spectral variability for hyperspectral unmixing," *IEEE Transactions on Image Processing*, vol. 28, no. 4, pp. 1923–1938, 2018.

[7] M. Dalm, M. Buxton, and F. van Ruitenbeek, "Discriminating ore and waste in a porphyry copper deposit using short-wavelength infrared (swir) hyperspectral imagery," *Minerals Engineering*, vol. 105, pp. 10–18, 2017.

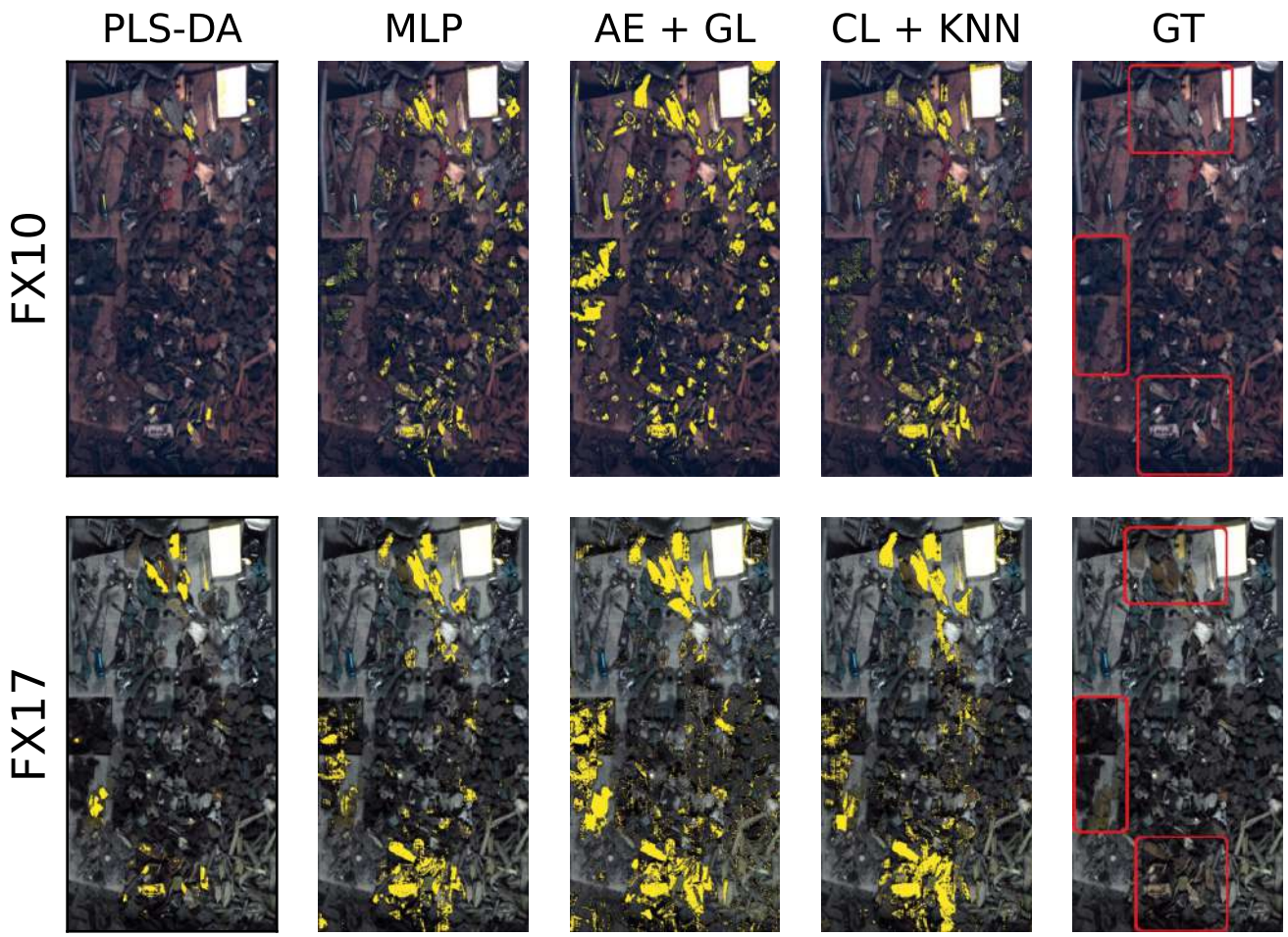


**Fig. 2**: Binary classification on FX10 and FX17 images using four methods: PLS-DA (Partial Least Squares Discriminant Analysis), MLP (Multilayer Perceptron), AE + GL (Autoencoder and Graph Learning), and CL + k-NN (Contrastive Learning and k-Nearest Neighbors). In the first four columns, yellow pixels indicate a plastic classification. The last column shows the approximate "ground truth" areas where plastic was manually placed.

[8] K. Moirogiorgou, F. Raptopoulos, G. Livanos, S. Orfanoudakis, M. Papadogiorgaki, M. Zervakis, and M. Maniadakis, "Intelligent robotic system for urban waste recycling," in *2022 IEEE International Conference on Imaging Systems and Techniques (IST)*, pp. 1–6, IEEE, 2022.

[9] Ãsmund Rinnan, F. van den Berg, and S. ren Balling Engelsen, "Review of the most common pre-processing techniques for near-infrared spectra," *TrAC Trends in Analytical Chemistry*, vol. 28, no. 10, pp. 1201–1222, 2009.

[10] H. Martens, J. P. Nielsen, and S. B. Engelsen, "Light scattering and light absorbance separated by extended multiplicative signal correction. application to near-infrared transmission analysis of powder mixtures," *Analytical Chemistry*, vol. 75, pp. 394–404, 02 2003.

[11] Rinnan, "Pre-processing in vibrational spectroscopy – when, why and how," *Anal. Methods*, vol. 6, pp. 7124–7129, 2014.

[12] A. Savitzky and M. J. E. Golay, "Smoothing and differentiation of data by simplified least squares procedures.," *Analytical Chemistry*, vol. 36, pp. 1627–1639, 07 1964.

[13] A. Biancolillo and F. Marini, "Chemometric methods for spectroscopy-based pharmaceutical analysis," *Frontiers in Chemistry*, vol. 6, 2018.

[14] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[15] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[16] P. Khosla, P. Teterwak, C. Wang, A. Sarna, Y. Tian, P. Isola, A. Maschinot, C. Liu, and D. Krishnan, "Supervised contrastive learning," *CoRR*, vol. abs/2004.11362, 2020.

[17] X. Zhu, Z. Ghahramani, and J. Lafferty, "Semi-supervised learning using gaussian fields and harmonic functions," ICML'03, p. 912–919, AAAI Press, 2003.