Why Use Non-Linear Chemometrics?

Helix function reconstruction with PCA vs neural network

Reconstruction of Helix from 1D representation

- Helix function
- Conventional PCA reconstruction
- Autoencoder reconstruction
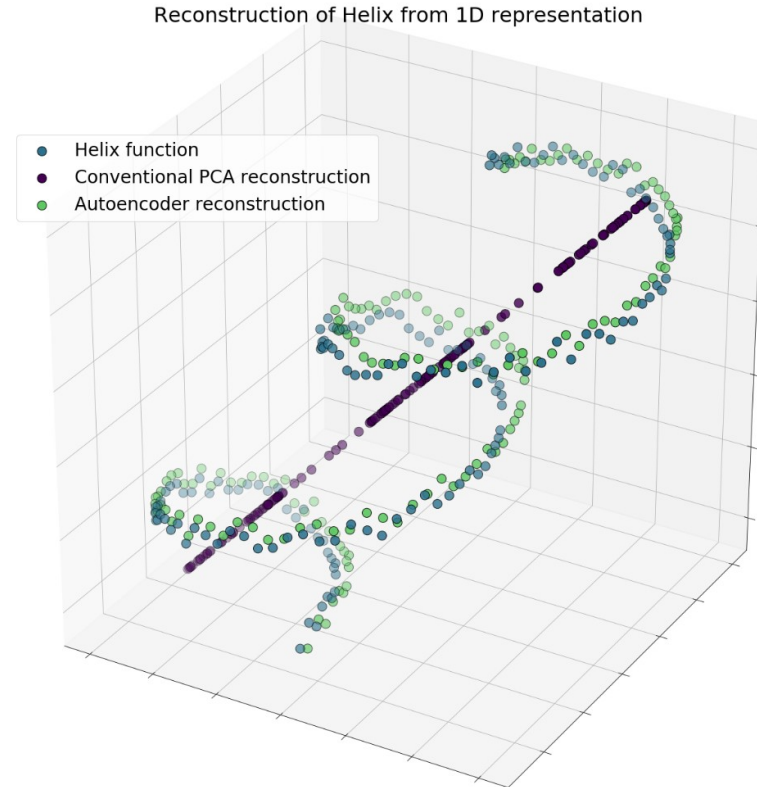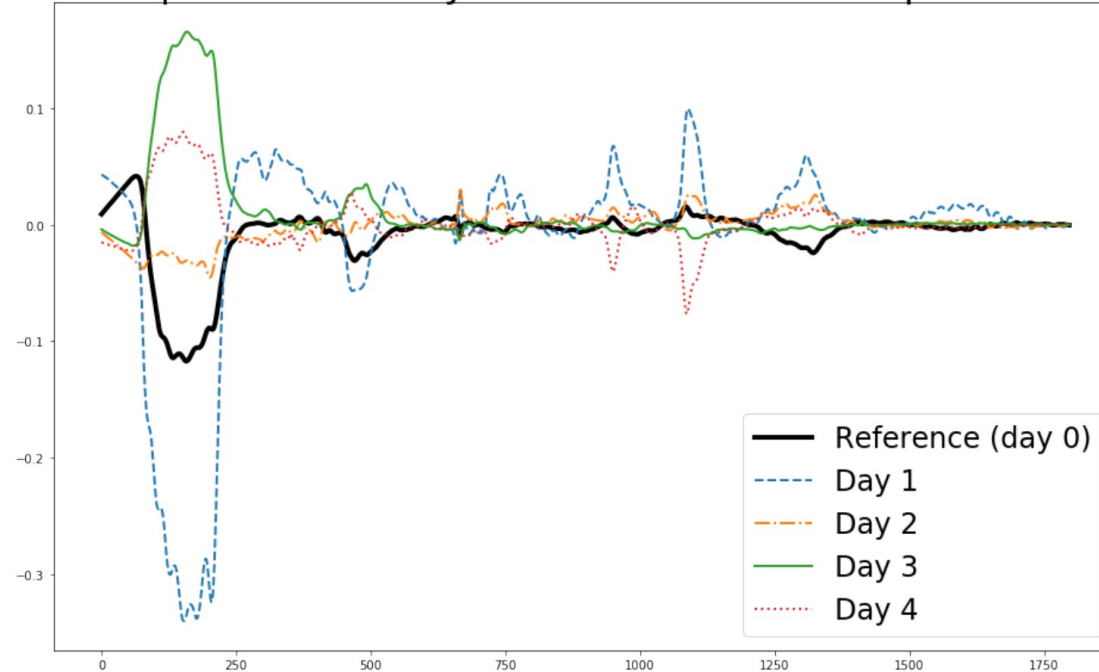
Raman Spectroscopy

Figure 3: **Average Raman spectra of MDCK cells at different days post-exposure to the cytotoxic drug doxorubicin.** Fluorescence background and mean signal has been subtracted from all spectra. While spontaneous Raman is a linear phenomenon, with spectral contributions increasing in direct proportion to the concentration of the corresponding chemical substances, the metabolic and physiological processes of the cell in response to stress or drugs are not necessarily linear.

Branched MLP

Autoencoder functionality replaces principal components analysis.

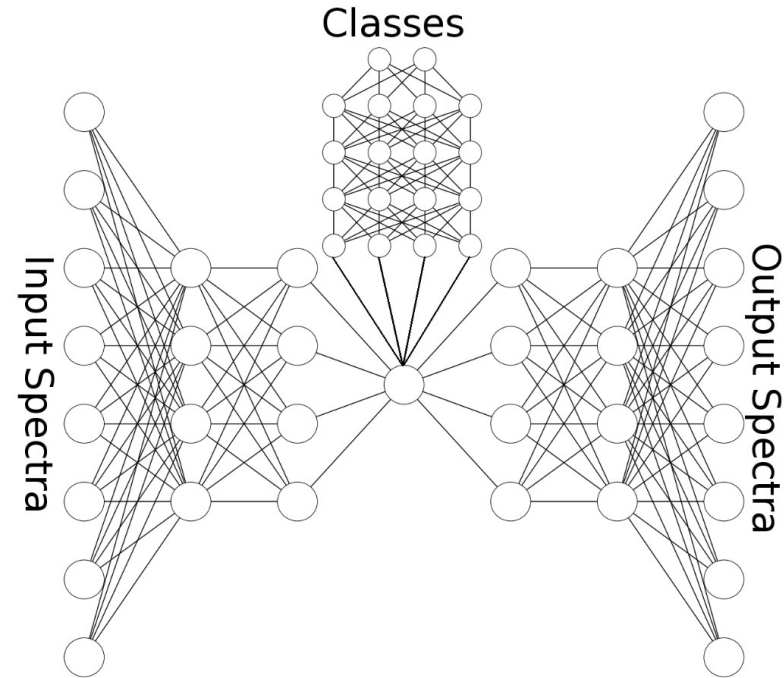Classification arm replaces Linear Discriminant Analysis

Figure 2: **Branched Multilayer Perceptron for autoencoding and classification of Raman spectra** Each node in the AE arm represents 128 nodes, and each small node in the classifier arm represents 64 nodes (except the output classes, which are 5 classes in the case of the doxorubicin experiment, and 2 classes in the prostate cancer cell lines experiment)

PCA/LDA vs. NN Comparison

Approach	Accuracy	Precision	Recall	Specificity	R^2	MSE
PCA/LDA	0.9278	0.9342	0.9235	0.9463	0.7630	0.0221
MLP	0.9833	0.9784	0.9909	0.9784	0.9891	0.0015
Improvement	1.060	1.047	1.073	1.034	1.296	14.67

Table 1: **MLP Outperforms PCA followed by LDA Workflow in Reconstituting and Distinguishing Raman Spectra of DU145 and PNT2 cells.**
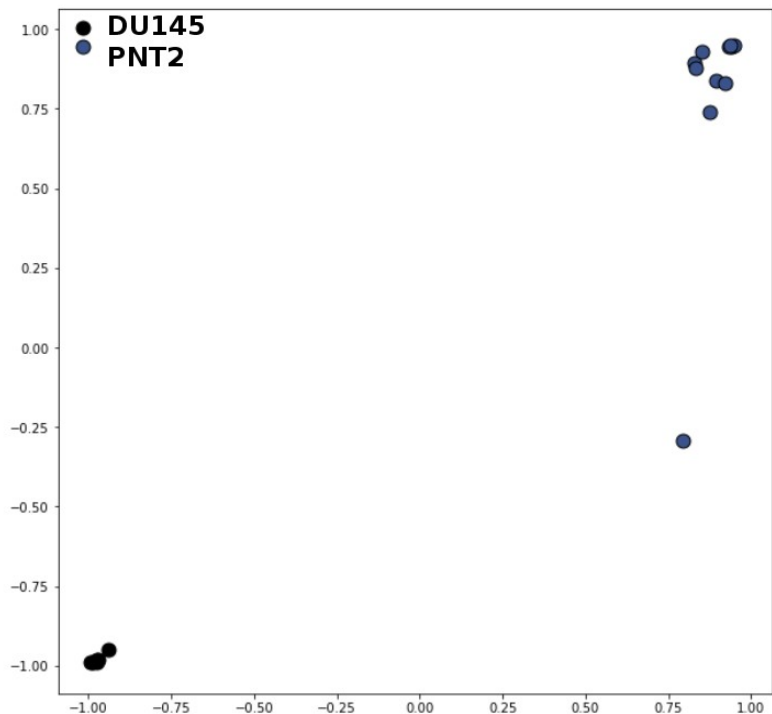
Doxorubicin Exposure Results

| Approach	Accuracy	R^2	MSE
PCA/LDA(39)	0.3355	0.9139	0.002381
PCA/LDA(5)	0.5896	0.9182	0.002263
PCA/LDA(1)	0.8105	0.8935	0.002957
MLP	0.9429	0.9964	0.00009518
Improvement(64)	3.269	1.069	20.56
Improvement(5)	2.235	1.065	19.37
Improvement(1)	1.294	1.102	28.28

Should read "Improvement(39)"

Table 2: **MLP Outperforms PCA followed by LDA Workflow in Reconstituting and Distinguishing Raman Spectra of MDCK cells challenged by doxorubicin.** Restricting the number of principal components actually improves the classification performance of the PCA/LDA approach at the expense of a slight decrease in reconstitution efficacy. The number of principal components in each case is denoted in parentheses.

Feature Space Mapping

DU145 and PNT2 Selected Feature Space

Doxorubicin Exposure Selected Feature Space
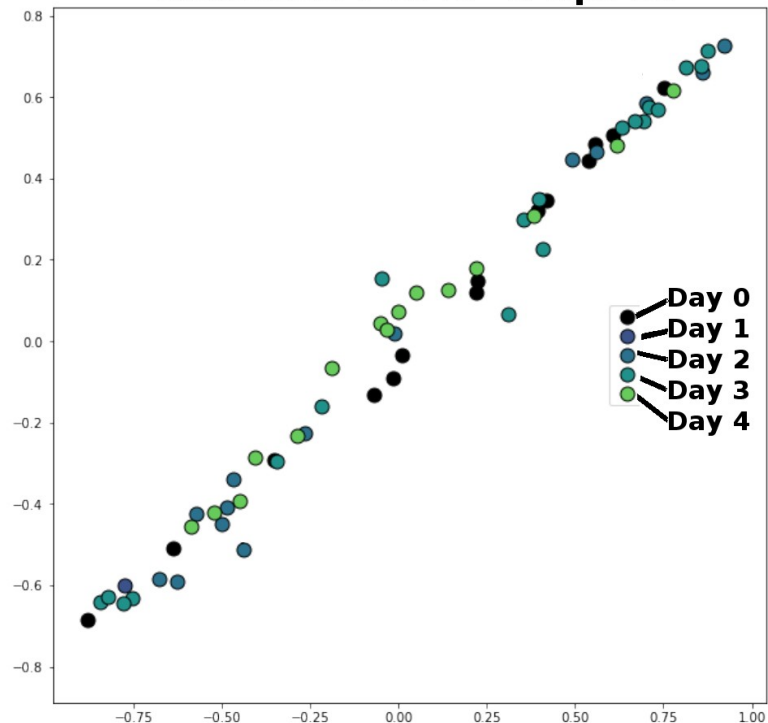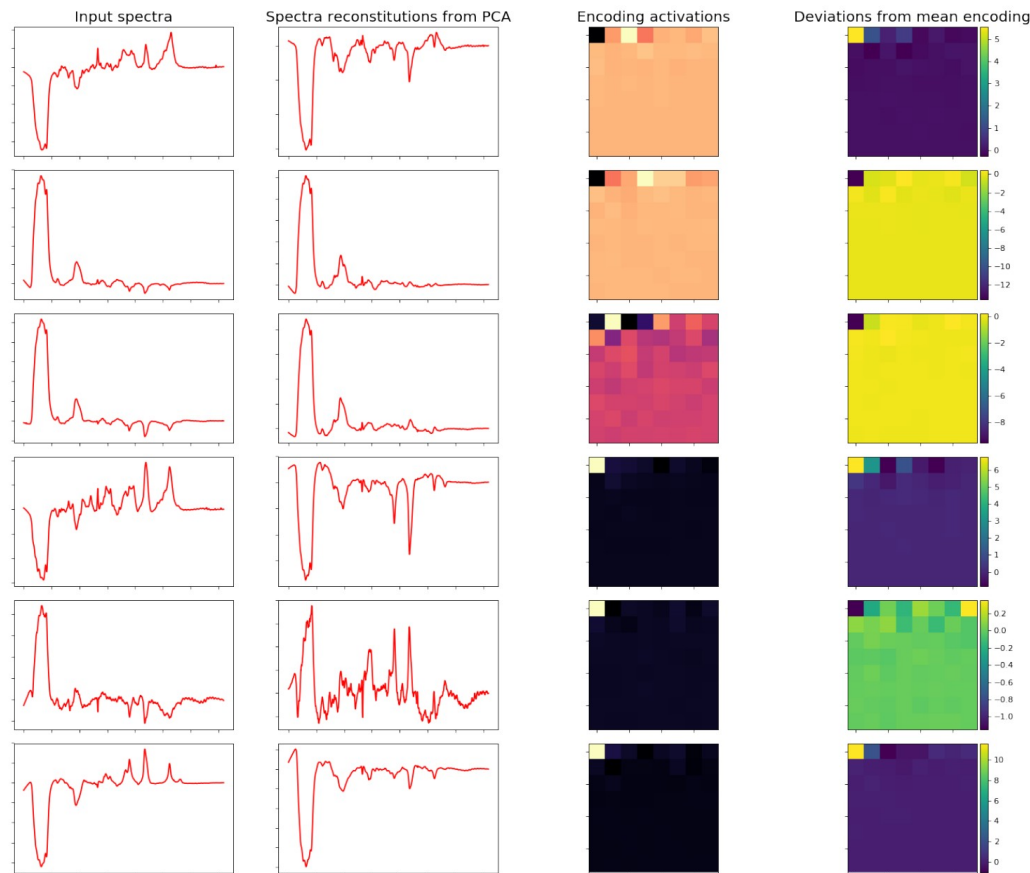
(Features with highest variance shown)

Figure 6: **Reconstitution of Raman spectra from doxorubicin experiment using PCA.** Unlike AE reconstitutions, PCA consolidates the majority of spectral variance into the first few components, effecting an efficient decomposition that captures the majority of spectral components in only a few components. However, the PCA is limited to a linear reconstruction and consequently several of the spectra reconstitutions here deviate from the gist of the input spectra.
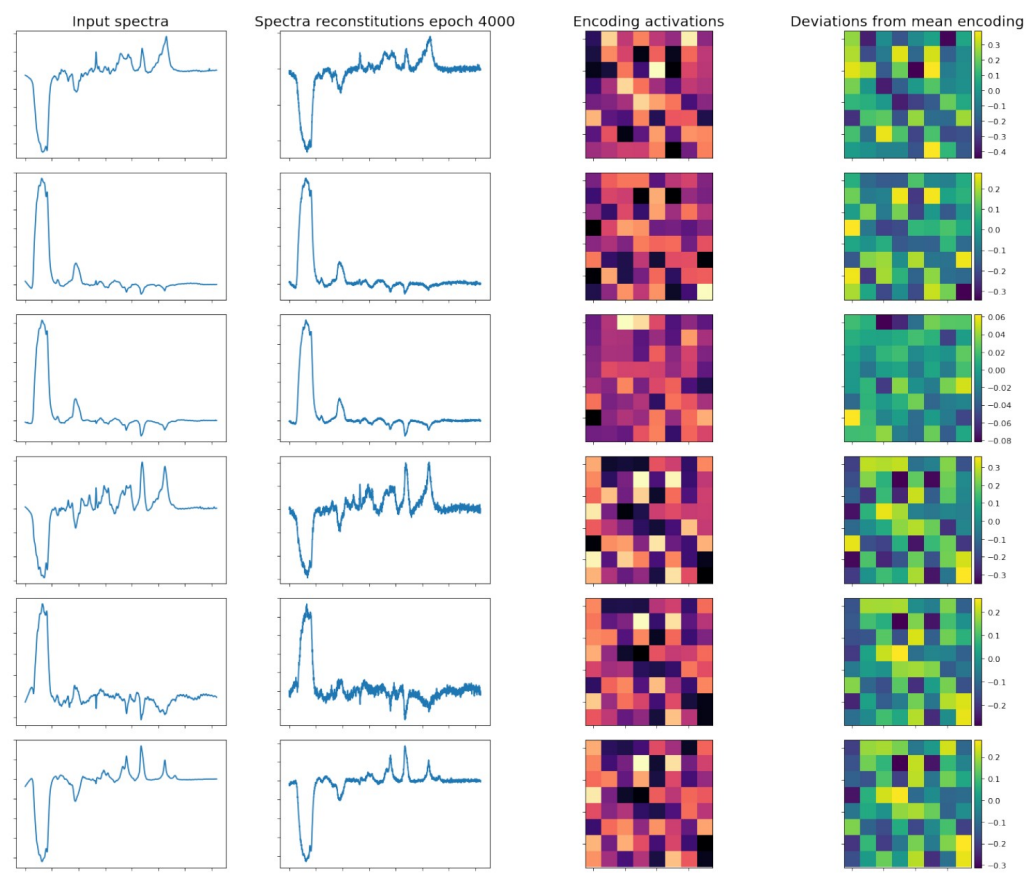
Figure 5: **Reconstitution of Raman spectra from doxorubicin experiment using the MLP autoencoder.** Although some of the reconstitutions contain high-frequency noise, the gist of the reconstitutions always follow closely the overall shape of the input. This is evident in the improved R^2 of the MLP-AE network over that achieved with PCA on test data sets ()

PCA reconstructions

MLP reconstructions