Machine Learning Applied to Material Analysis

Kevin Honeker, Fareen Pourmossavian, Jonathan Alcantara, Benjamin Merberg, David Marrero
(kmh289, fp183, jba50, bhm28, dgm80)@scarletmail.rutgers.edu Advisors: Prof. Turilli, Dr. Hanagandi

RUTGERS THE STATE UNIVERSITY OF NEW JERSEY

Optimal Solutions™

Goal

Determine the best model for wine classification in the pursuit of bringing to market a cutting-edge material analysis application which produces accurate, near real-time readings.

Motivation and Objectives

Motivation

Companies have a pressing need to measure material properties throughout their supply chains.

Objectives

- ❑ Create an automated data preprocessing system that converts text output files from the portable detectors to CSV files for use in training our models.
- ❑ Create models which take the given spectroscopic data and make wine classifications .

Research Challenges

- ❑ **Limited Dataset Size**: practical challenges of experimental setup and data acquisition processes.
- ❑ **Text Data Format**: requires tailored data wrangling.
- ❑ **Abnormalities in the Dataset**: data acquisition process is noisy both at sensor and process levels.

Conclusions

- ❑ Both PLS and Artificial Neural Networks are viable methods for the wine classification. More raw data is needed for either model to to be generalized to new unseen data.
- ❑ The 5 mm path length led to the highest classification accuracies across both detectors, with the Neural Network Detector 2 - 5 mm candidate model classifying both red and white wines with 100% classification accuracy.
- ❑ The PLS Method performed poorly on the red wine data, but showed strong performance on white wine data, as did the Neural Network Method.

Methodology

Data Preprocessing

- ❑ Label data, remove calibration runs and data with input errors.
- ❑ Convert TXT files to usable CSV files.

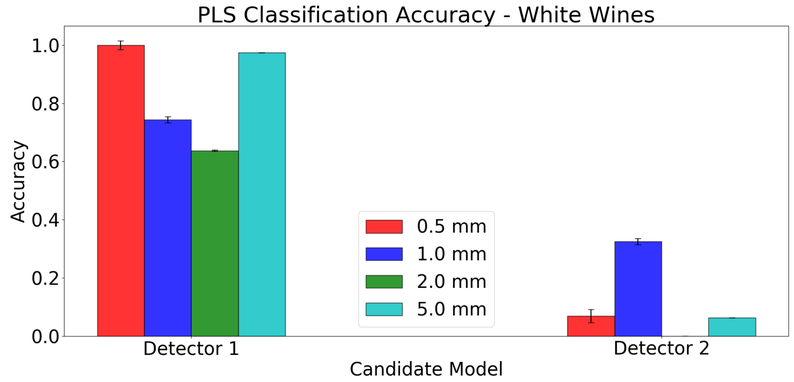
Partial Least Squares

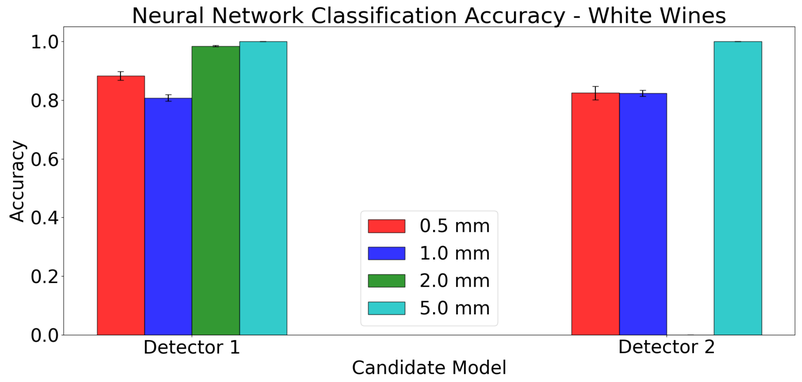
- ❑ The PLS model binarizes the labels, optimizes the number of PLS components between 1 and 40, and utilizes the number of components that yields the lowest mean squared error.

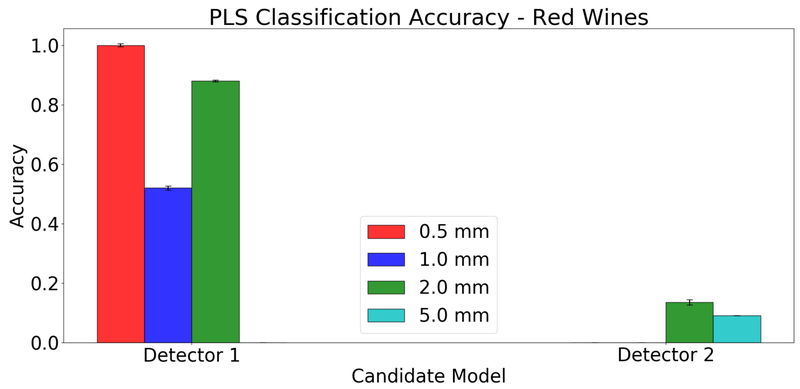
Neural Network

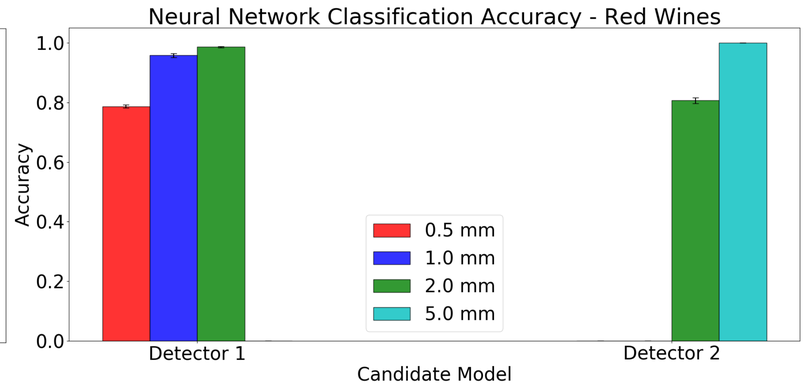
- ❑ Separate data into eight groups, each group representing a candidate model. Two detectors, each with four path lengths at which data was taken.
- ❑ Resample data using the bootstrap technique for each candidate model.
- ❑ Use the Stratified K Fold strategy to determine training and testing set splits.
- ❑ Train a sequential, feed-forward network for each candidate model, using the Keras library with TensorFlow backend.

Results









Average Prediction Time (seconds)

	D1 0.5mm	D1 1mm	D1 2mm	D1 5mm	D2 0.5mm	D2 1mm	D2 2mm	D2 5mm
White	0.0005	0.0001	0.0002	0.0002	0.0001	0.0001	X	0.0001
Red	0.0002	0.0001	0.0004	X	X	X	0.0001	0.0001

Average Prediction Time (seconds)

	D1 0.5mm	D1 1mm	D1 2mm	D1 5mm	D2 0.5mm	D2 1mm	D2 2mm	D2 5mm
White	0.0762	0.0773	0.0749	0.0794	0.0665	0.0685	0.0791	X
Red	0.0661	0.0661	0.1420	0.0692	X	X	0.0645	0.0661