Combining Deep learning & Raman Spectroscopy for Rapid pesticide screening

Esther Kiguru
The University of Nairobi
ASP2024 Alumni Online Presentation

How safe is the food we consume?

- Fruits & Vegetables accounts for 25% and > 80% of diets in rural & urban areas in Kenya.

- The large population creates high demand for fresh produce daily.

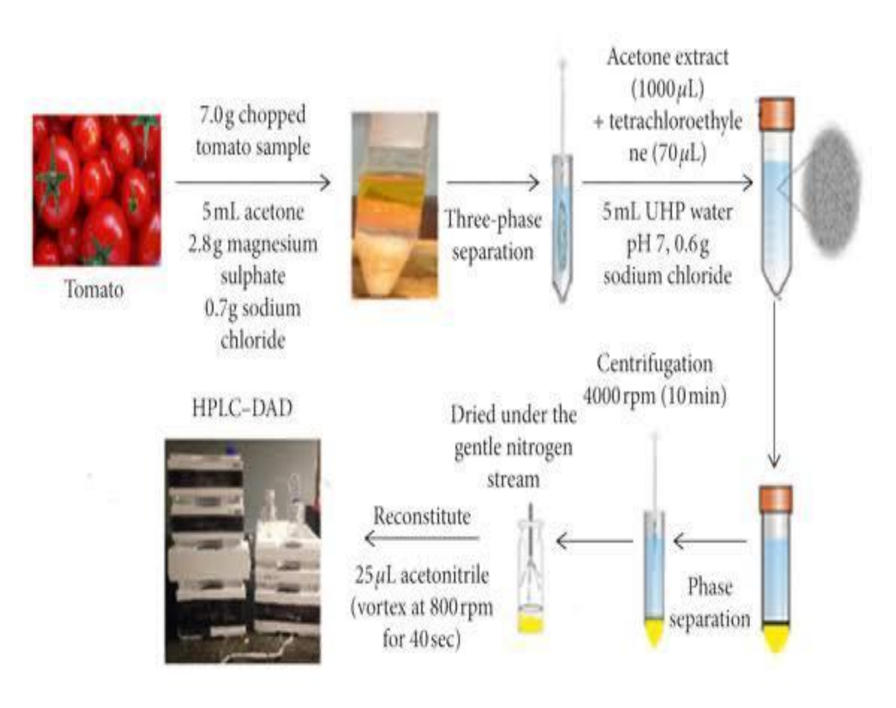
- To satiate the high demand,farmers use pesticides to protect crops from pest & diseases to increase yields;This contaminates the foods.

- Repeated exposure to pesticide residue can result in a cumulative effect on the body overtime

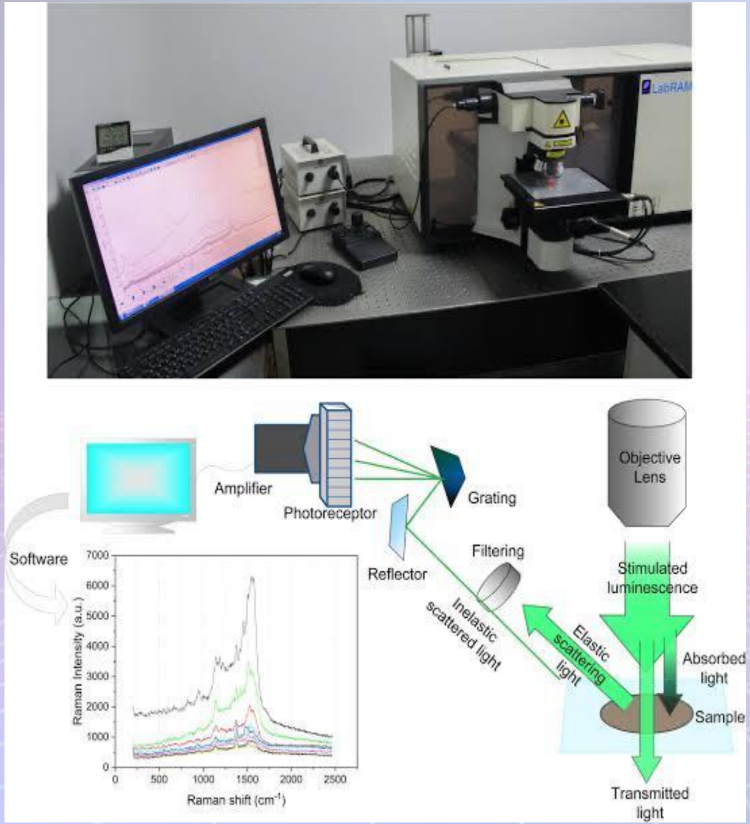
Should we be concerned?

- There is a need for assessment to ensure Maximum Residue Limits are not exceeded

Residue Detection Methods



Standard detection methods e.g GC-MS,LC



Spectroscopy based techniques

Methodology

Pesticide Concentration Preparation

- Pesticide concentrations of Chlorothalonil (CTL) were prepared using serial dilution techniques.

- According to the Pest Control Products Board,the Maximum Residue Limit(MRL) of Chlorothalonil was 2.0 parts per million (ppm)
- Different concentrations above ,within and below the MRLs were prepared for the pesticide

Selection of suitable vegetables

- Vegetable samples of Tomatoes,Kales,spinach and lettuce were collected from a nearby farm

- The vegetable samples were washed,cut into small pieces and labelled according to the different concentration levels.

- The prepared samples were then sprayed with the different CTL concentrations and left to dry for a few hours before taking spectral measurements.

Data Preprocessing & Augmentation

Data Cleaning

Examining the data for any potential missing values, gaps and
removal of unwanted features

Noise removal

Minimized background noise interference

Applied Outlier Detection to enhance data quality.

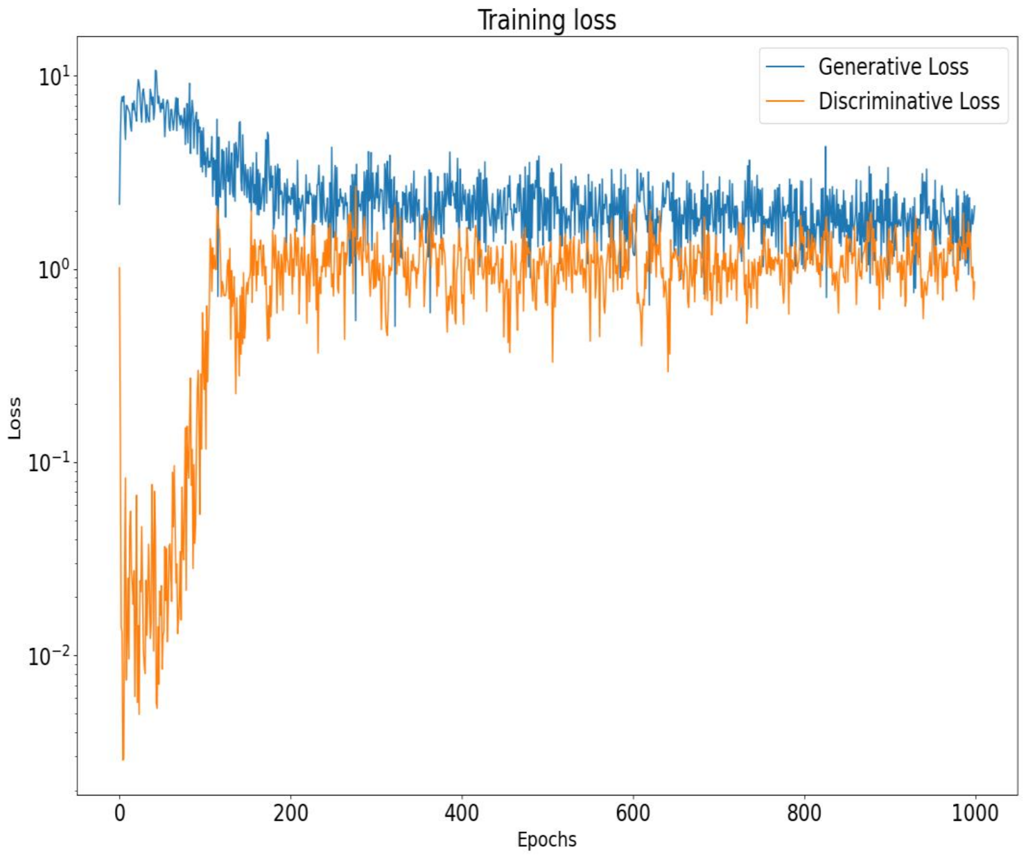
Augmentation

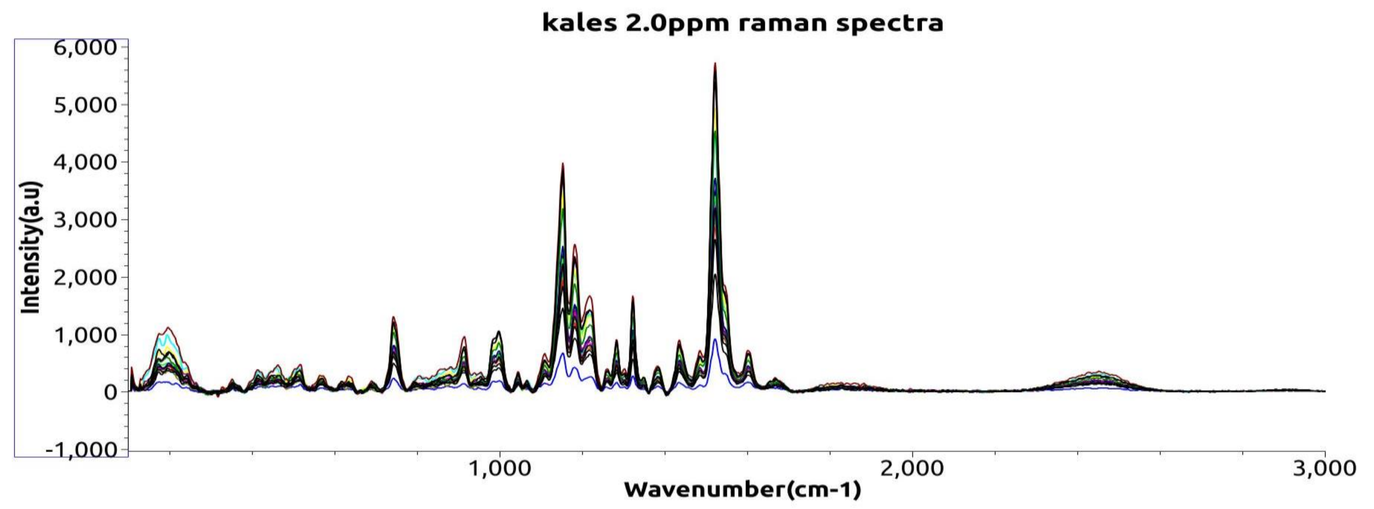
Generative Adversarial Networks(GANs)

GANs
offer a powerful means to augment limited datasets, reducing the dependency on extensive
and time-consuming laboratory efforts

GANs model

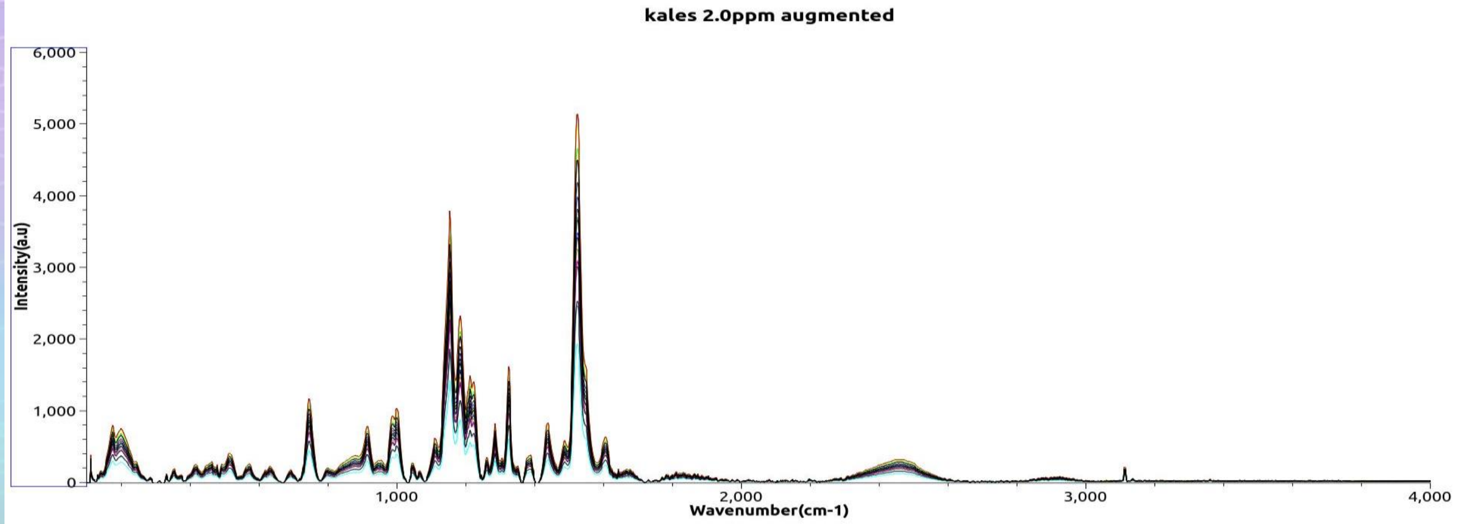
- As training progresses, the generator loss should decrease implicating that the generator is getting better at generating data that is more difficult for the discriminator to distinguish from real data.
- Convergence is reached when the generator loss stabilizes not necesarily reaching zero, but it should remain relatively constant.
- A decreasing discriminator loss means that the discriminator is getting better at distinguishing between real and generated data.
- when the discriminator loss stabilizes at this point, the discriminator is no longer able to easily differentiate between real and generated data.

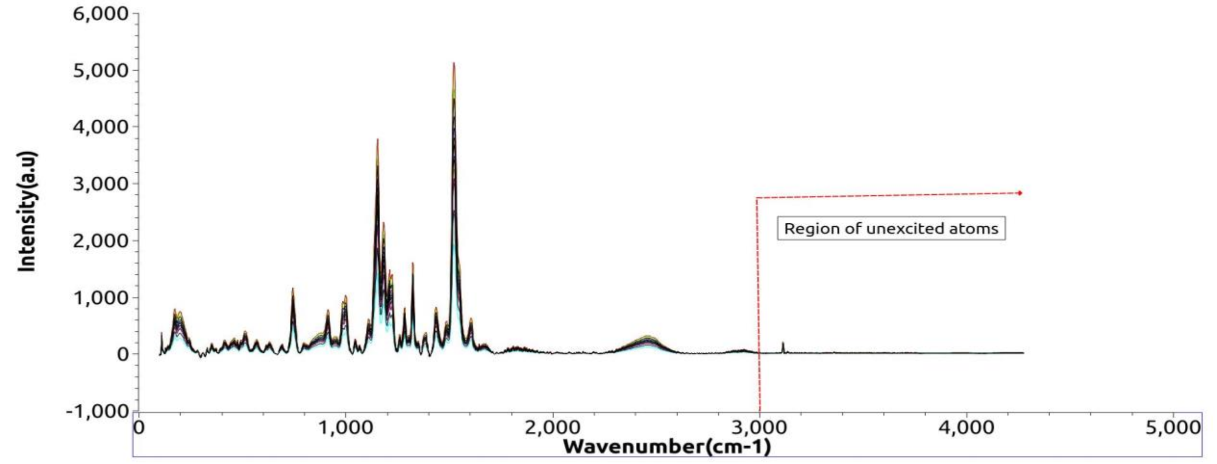




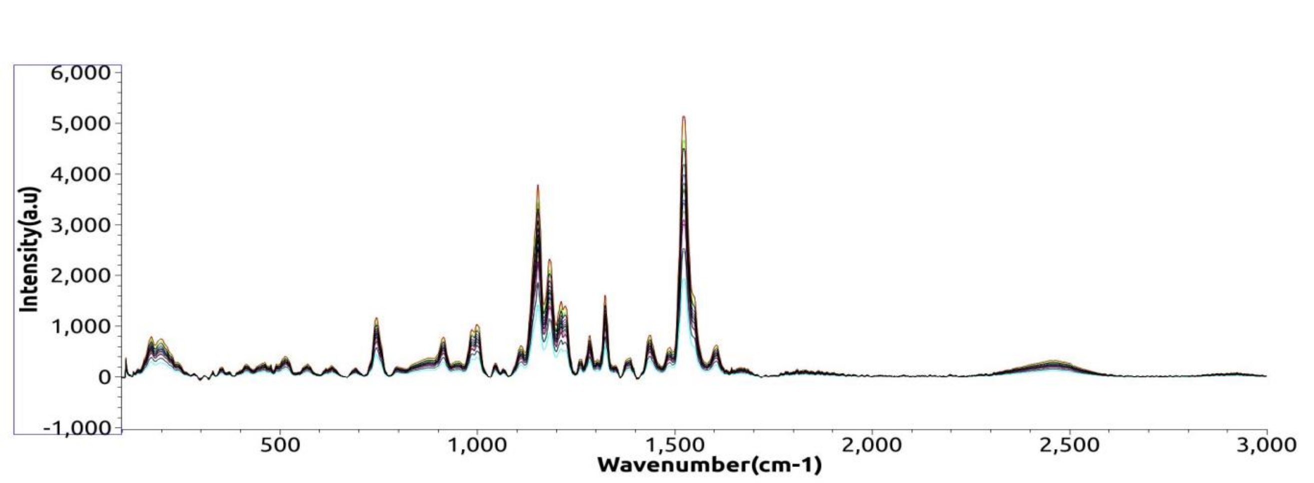
Original spectra

Augmented spectra





Red laser with wavelength of 785 nm is limited to excite atoms at a raman shift of 3000 hence no significant spectra observed pass the region



Model Development

Classification Model

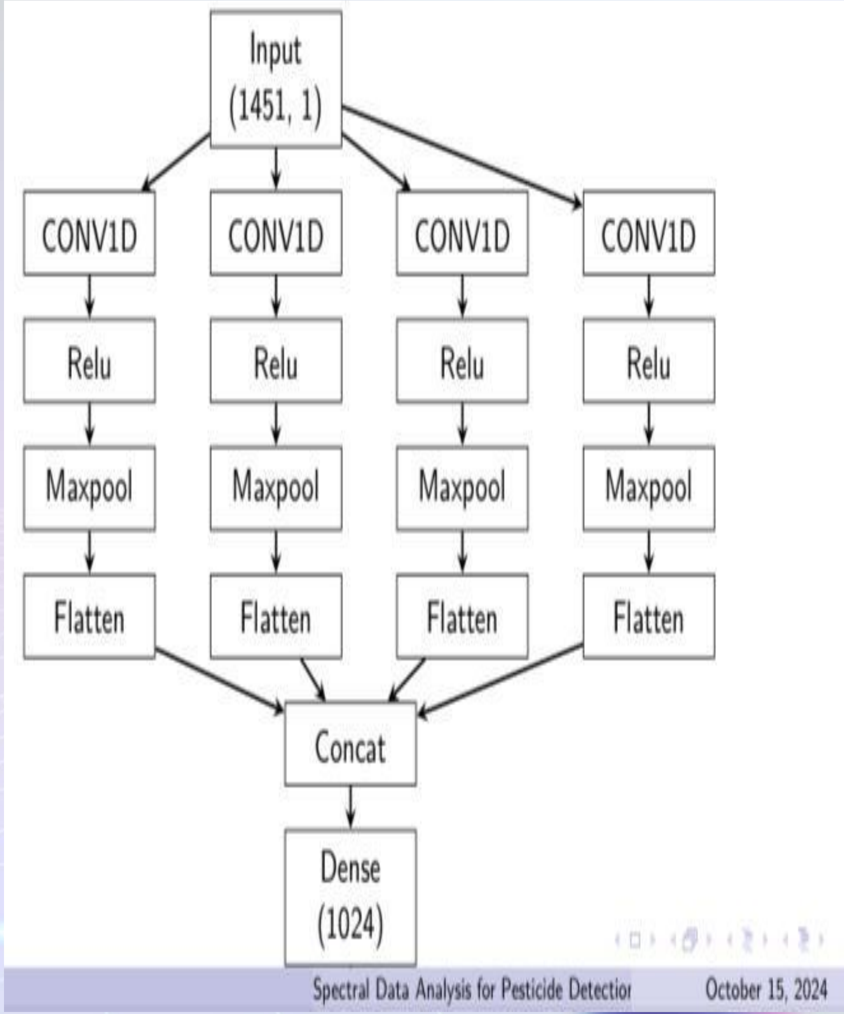
qualitative analysis of pesticide presence in spectral data.

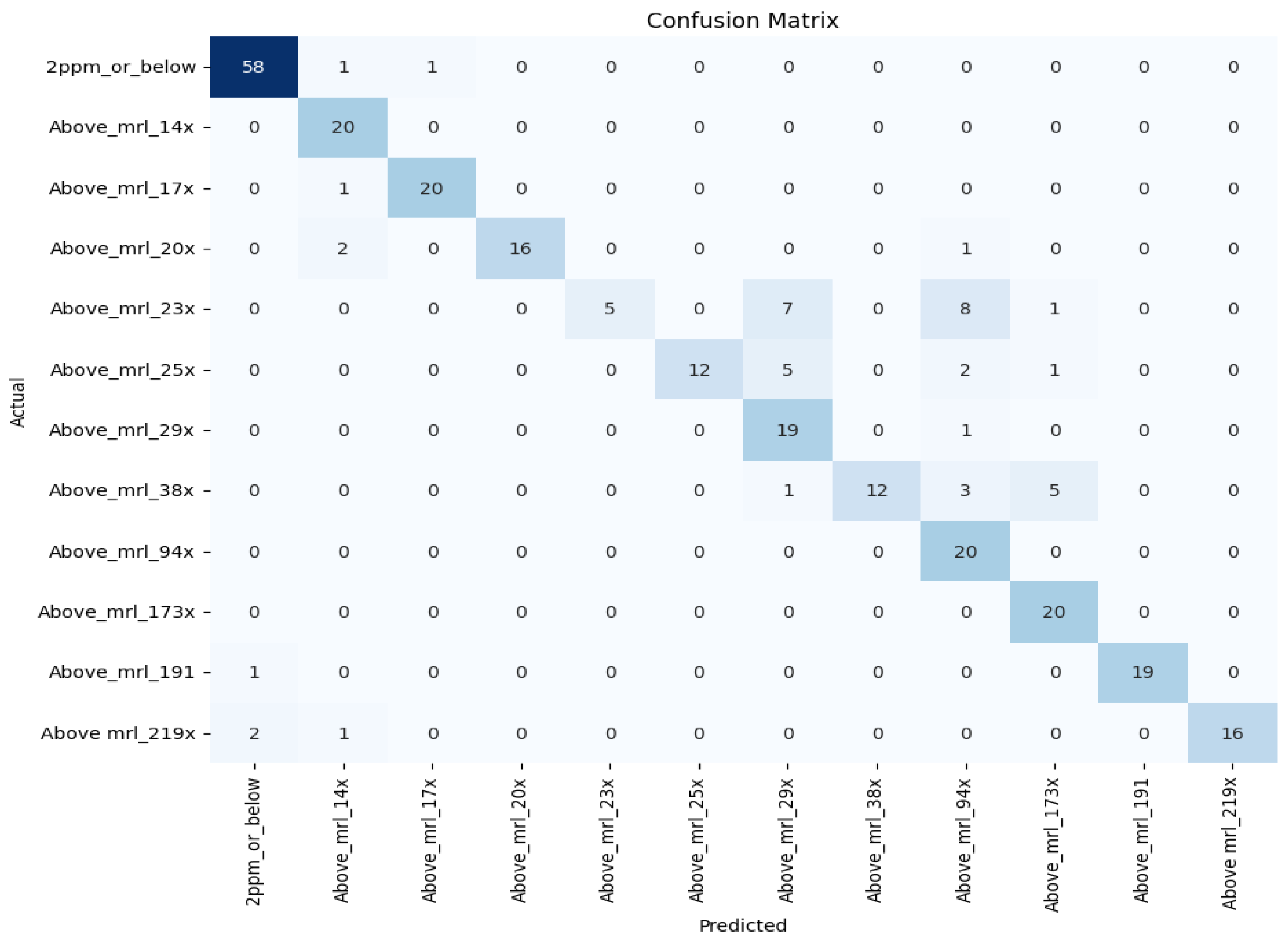
Regression Model

predicts a continuous numerical output based on input features

Classification Model

- The model architecture is a stacked ResNet-type 1D-CNN for spectral data classification
- Layers:
  - **Convolutional** layers for spatial feature extraction.
  - **Relu activation** layer for introducing non-linearity
  - **Flatten layer** for one-dimensional data representation.
  - **Dense layer** establishes the relationship patterns in the features
  - **Max pooling** for downsampling.
  - **Output layer** that represents the predicted classes
- Optimizer: AdamW optimizer for gradient descent.
- Evaluation Metrics:
  - Accuracy: Percentage of correctly classified samples.
  - Loss: Error between predictions and true labels.
  - Precision: Proportion of true positive predictions among all positive predictions.
  - Recall: Proportion of true positive predictions among all actual positive instances.
  - F1 Score: Harmonic mean of precision and recall.
  - Support: The number of actual occurrences of each class in the specified datase




Confusion Matrix
Actual
Predicted
2ppm_or_below 58 1 1 0 0 0 0 0 0 0 0 0
Above_mrl_14x 0 20 0 0 0 0 0 0 0 0 0 0
Above_mrl_17x 0 1 20 0 0 0 0 0 0 0 0 0
Above_mrl_20x 0 2 0 16 0 0 0 0 1 0 0 0
Above_mrl_23x 0 0 0 0 5 0 7 0 8 1 0 0
Above_mrl_25x 0 0 0 0 0 12 5 0 2 1 0 0
Above_mrl_29x 0 0 0 0 0 0 19 0 1 0 0 0
Above_mrl_38x 0 0 0 0 0 0 1 12 3 5 0 0
Above_mrl_94x 0 0 0 0 0 0 0 0 20 0 0 0
Above_mrl_173x 0 0 0 0 0 0 0 0 0 20 0 0
Above_mrl_191 1 0 0 0 0 0 0 0 0 0 19 0
Above mrl_219x 2 1 0 0 0 0 0 0 0 0 0 16
2ppm_or_below
Above_mrl_14x
Above_mrl_17x
Above_mrl_20x
Above_mrl_23x
Above_mrl_25x
Above_mrl_29x
Above_mrl_38x
Above_mrl_94x
Above_mrl_173x
Above_mrl_191
Above mrl_219x

Classification Report

Class	Precision	Recall	F1-Score	Support
2ppm_or_below	0.95	0.97	0.96	60
Above_mrl_14x	0.80	1.00	0.89	20
Above_mrl_17x	0.95	0.95	0.95	21
Above_mrl_20x	1.00	0.84	0.91	19
Above_mrl_23x	1.00	0.24	0.38	21
Above_mrl_25x	1.00	0.60	0.75	20
Above_mrl_29x	0.59	0.95	0.73	20
Above_mrl_30x	1.00	0.57	0.73	21
Above_mrl_94x	0.57	1.00	0.73	20
Above_mrl_173x	0.74	1.00	0.85	20
Above_mrl_191x	1.00	0.95	0.97	20
Above_mrl_219x	1.00	0.84	0.91	19
Accuracy			0.84	281
Macro Avg	0.88	0.83	0.81	281
Weighted Avg	0.89	0.84	0.83	281
cohen kappa coeffient	0.825			

Regression Model

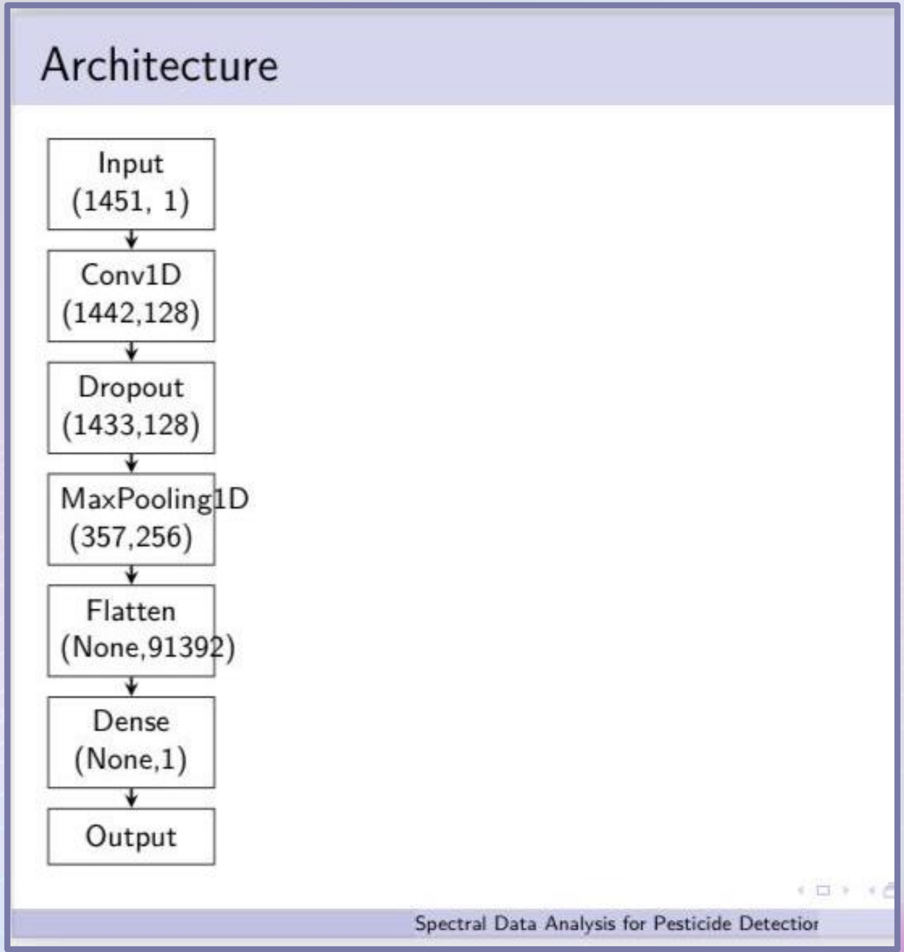
- 1D Convolutional Neural Network (CNN) followed by dense layers.
- Convolutional Layers:

  1D Convolutional layers capture hierarchical features in the input data.

  Dropout layers and **MaxPooling1D layers** are employed for regularization.
- Dense Layers: combine features for final regression prediction.
- Evaluation Metrics:

  Mean Absolute Error (MAE) is used as the loss function.

  R2SCORE, Mean Squared Error (MSE), and Mean Absolute Error (MAE) are used as evaluation metrics.

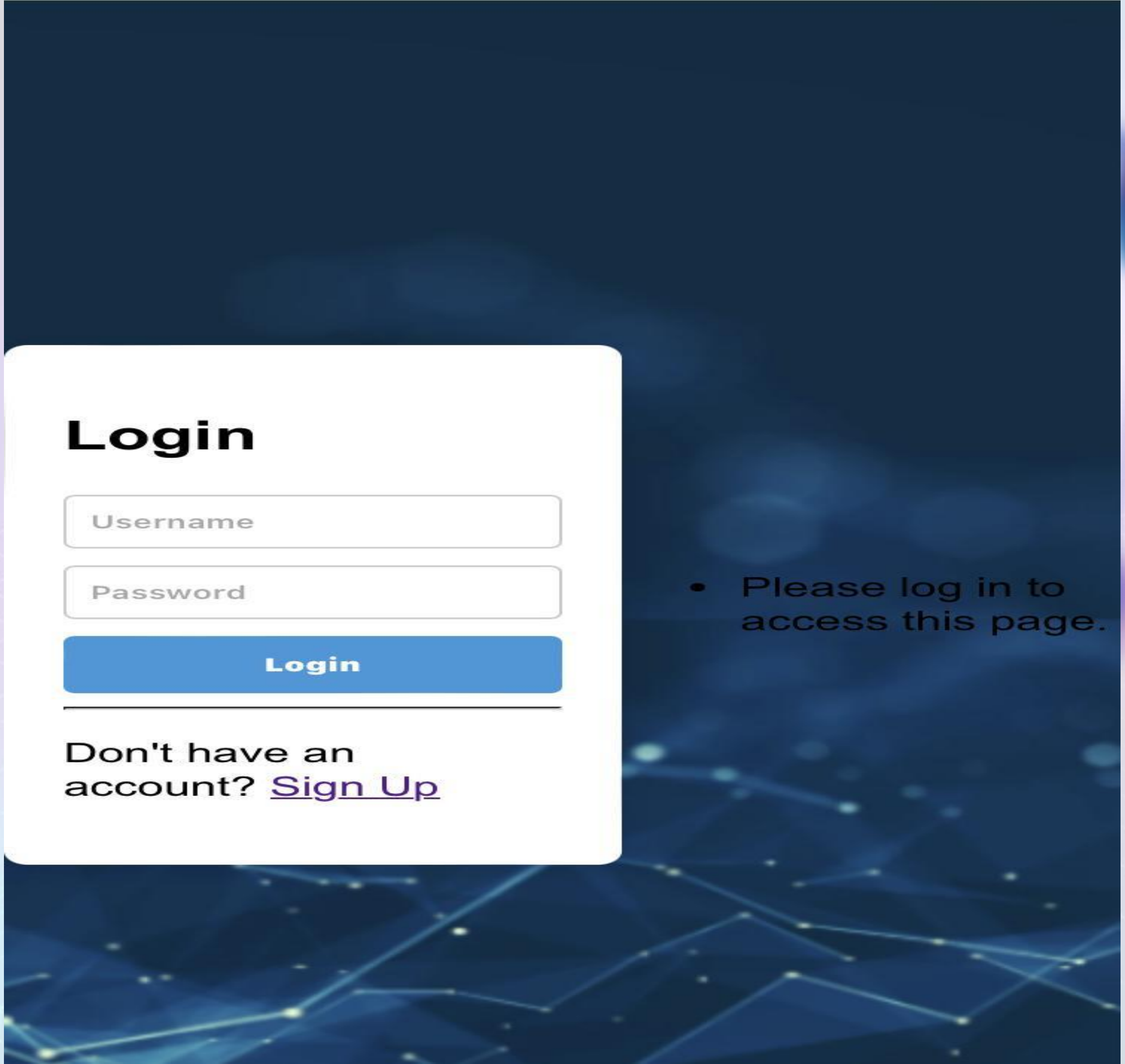


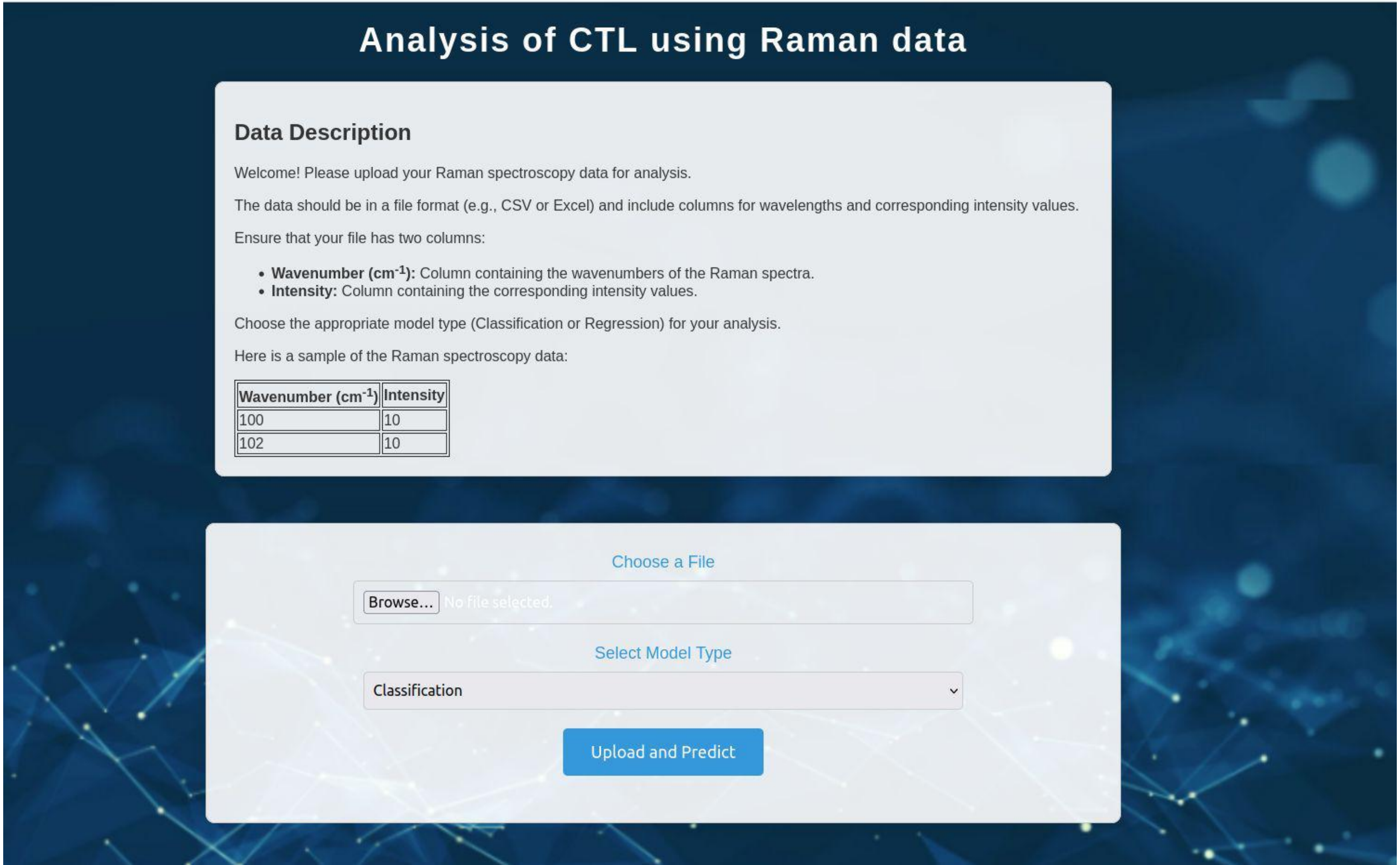
- Indicates the proportion of the variance in the dependent variable predictable from the independent variable(s).
- Ranges from -∞ to 1, with 1 being a perfect fit

R2 Score Trend Over Training Epochs

Application Programming Interface

User authentication Interface




Analysis of CTL using Raman data
Data Description
Welcome! Please upload your Raman spectroscopy data for analysis.
The data should be in a file format (e.g., CSV or Excel) and include columns for wavelengths and corresponding intensity values.
Ensure that your file has two columns:
• Wavenumber (cm^{-1}): Column containing the wavenumbers of the Raman spectra.
• Intensity: Column containing the corresponding intensity values.
Choose the appropriate model type (Classification or Regression) for your analysis.
Here is a sample of the Raman spectroscopy data:
Wavenumber (cm^{-1}) | Intensity
100 | 10
102 | 10
Choose a File
Browse... No file selected.
Select Model Type
Classification
Upload and Predict

Classification prediction

Home

Download CSV

Above_mrl_14x

Above_mrl_17x

2ppm_or_below

Above_mrl_17x

Above_mrl_14x

2ppm_or_below

2ppm_or_below

2ppm_or_below

2ppm_or_below

Above_mrl_17x

Regression Results

Home

Download CSV

Predicted Concentration
26.79236602783203
25.87305450439453
132.3760223388672
49.60273742675781
17.16131019592285
0.7084272503852844
33.79109573364258
3.1136128902435303
1.052881121635437
14.453977584838867
89.1885986328125
7.373283386230469
45.88431167602539
17.620405197143555
18.032611846923828

Thanks!