Overcoming Data Scarcity in Digital Agriculture: A Generative Approach to Hyperspectral Imaging

Nooshin Noshiri[1], Christopher Henry[1], Christopher Bidinosti[2]
Computer Science Department, University of Manitoba, MB, Canada[1] | Physics Department, University of Winnipeg, MB, Canada[2]

Introduction

One of the most valuable tools in modern farming is hyperspectral imaging (HSI) [1], a technology that allows for a detailed analysis of crop health. HSI extends beyond the capabilities of conventional imaging by capturing a wide range of spectral wavelengths, including those outside the visible spectrum. This technology generates a three-dimensional (3D) data structure called a hyperspectral cube (Figure 1) – a stack of images, each corresponding to a specific wavelength – enabling detailed analysis of plant characteristics and early detection of crop issues such as nutrient deficiencies, diseases, and stress markers. However, the application of machine learning (ML) to analyze HSI data is limited by the scarcity of diverse datasets. Challenges include the high costs of HSI equipment, logistical difficulties in data collection, and maintaining standardized acquisition conditions. This research investigates deep generative models as a solution to augment HSI datasets and improve the performance of ML models in agricultural analysis.

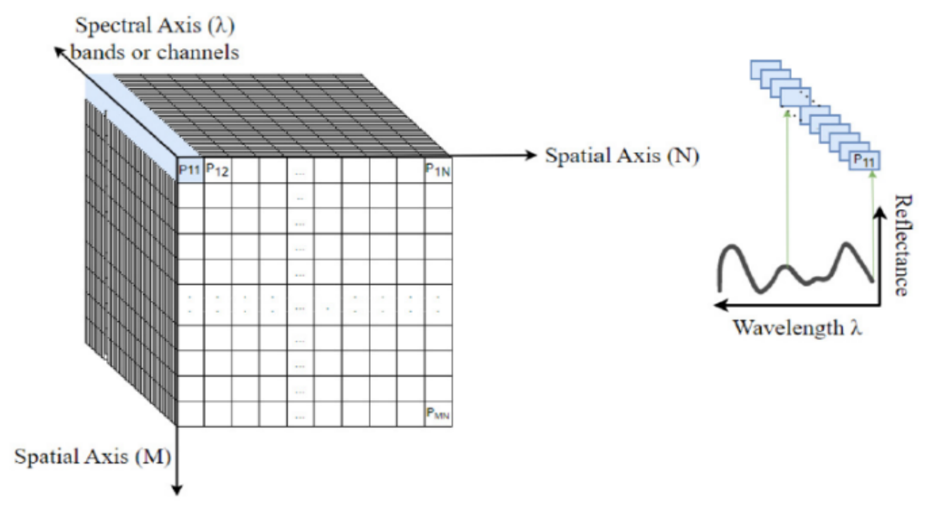


Figure 1. The hyperspectral cube

Objectives and Methodologies

This research aims to address the challenges associated with the scarcity of HSI datasets by exploring the use of deep generative models to augment spectral data. The objectives are the following:

- To evaluate the performance of several generative models, including a customized variational autoencoder (VAE) [2], deep convolutional generative adversarial networks (DCGAN) [3], and Wasserstein GAN with gradient penalty (WGAN-GP) [4], in generating synthetic HSI data.
- To compare the quality, variability, and representativeness of the augmented data from these generative models to a conventional noise addition approach.

In this research, we utilized a hyperspectral dataset of Buttercrunch lettuce under varying nitrogen stress levels, captured using the SPECIM FX10 camera [5]. From the 404 captured images, 104 samples were chosen for data augmentation, representing plants grown with a full standard dose of nitrogen (operates within a spectral range of 400 to 1000 nm). Algorithm 1 was employed for preprocessing, which included data calibration and addressing negative values, while Algorithm 2 (adapted from a method in [6]) was used for segmentation to identify key regions of interest. A sample result of the segmentation process is shown in Figure 2, and the entire normalized dataset is visualized in Figure 3.

Algorithm 1 Preprocess hyperspectral data

```
1: Start
2: Read ".raw" data
3: Convert ".raw" data to 3D array
4: Calibrate the data using black and white calibration (see Equation 1)
5: Replace negative voxel values with the average of sum of neighboring voxels with values $\geq 0$
6: Stop
```

Algorithm 2 Leaf segmentation from hyperspectral data using NDVI and EGI

```
1: Input: Hyperspectral image data
2: Output: Segmented leaf areas along all bands
3: Calculate mean across spectral bands for Blue (bands $1 - 38$)
4: Calculate mean across spectral bands for Green (bands $39 - 76$)
5: Calculate mean across spectral bands for Red (bands $77 - 114$)
6: Calculate mean across NIR spectral bands (bands $115+$)
7: Compute NDVI for each pixel:
       $\mathrm{NDVI} = \frac{\mathrm{NIR} - \mathrm{Red}}{\mathrm{NIR} + \mathrm{Red}}$
8: Compute EGI for each pixel:
       $\mathrm{EGI} = 2 \times \mathrm{Green} - \mathrm{Blue} - \mathrm{Red}$
9: Create binary masks based on thresholding NDVI and EGI
10: Element-wise multiplication of NDVI and EGI masks using thresholding to form the primary mask
11: Morphological Operations:
12: Apply morphological closing using a $3 \times 3$ structuring element
13: Apply morphological opening using a $3 \times 3$ structuring element
14: Extract and output the final segmented leaf areas
```

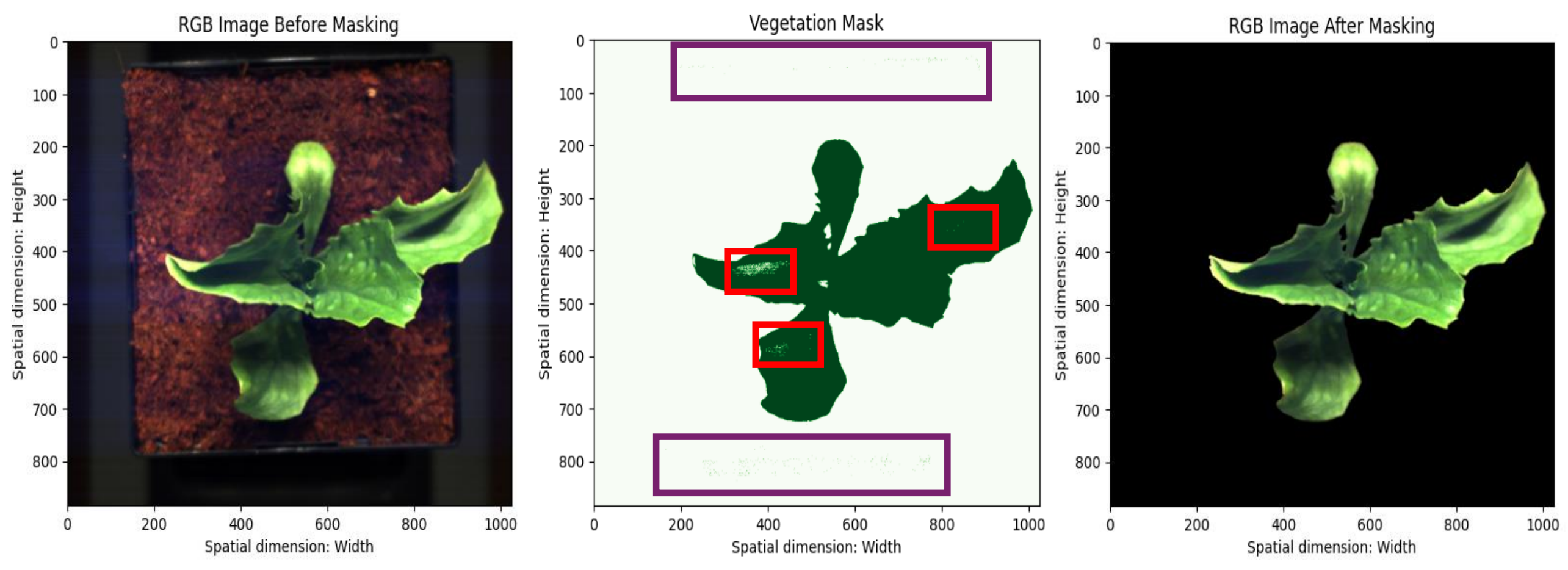


Figure 2. Segmentation process.

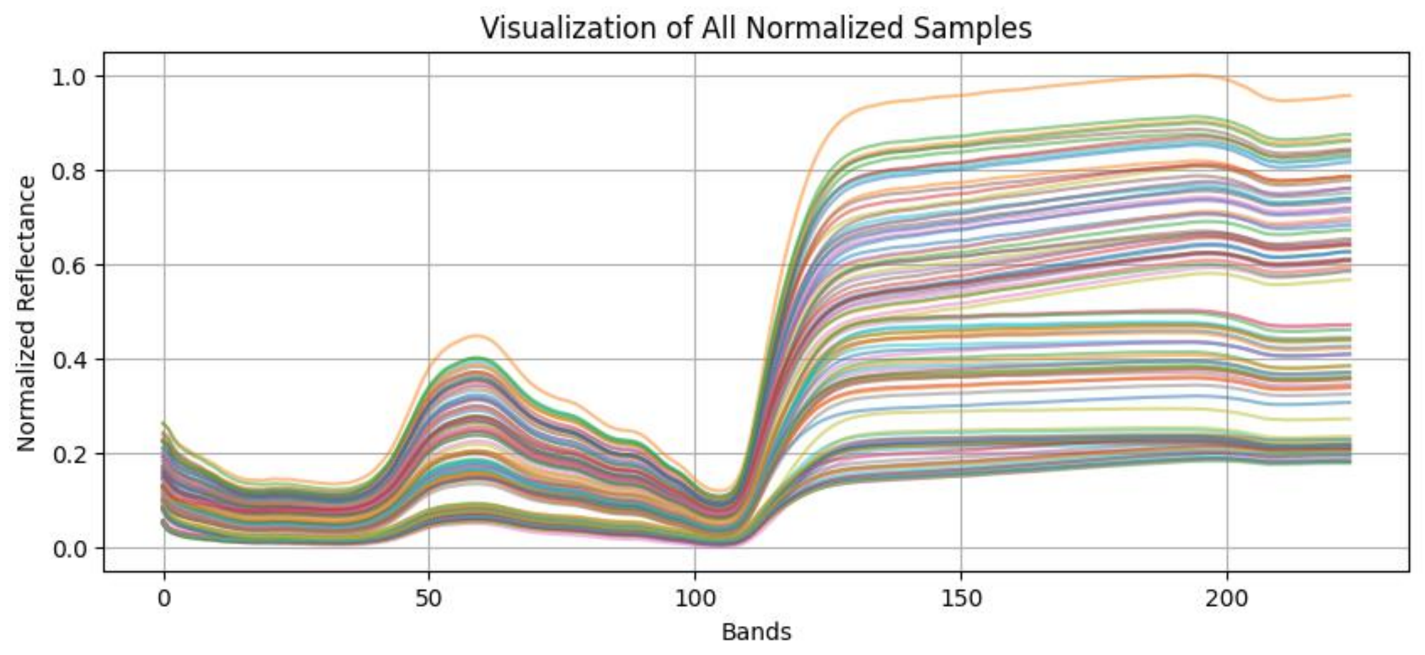
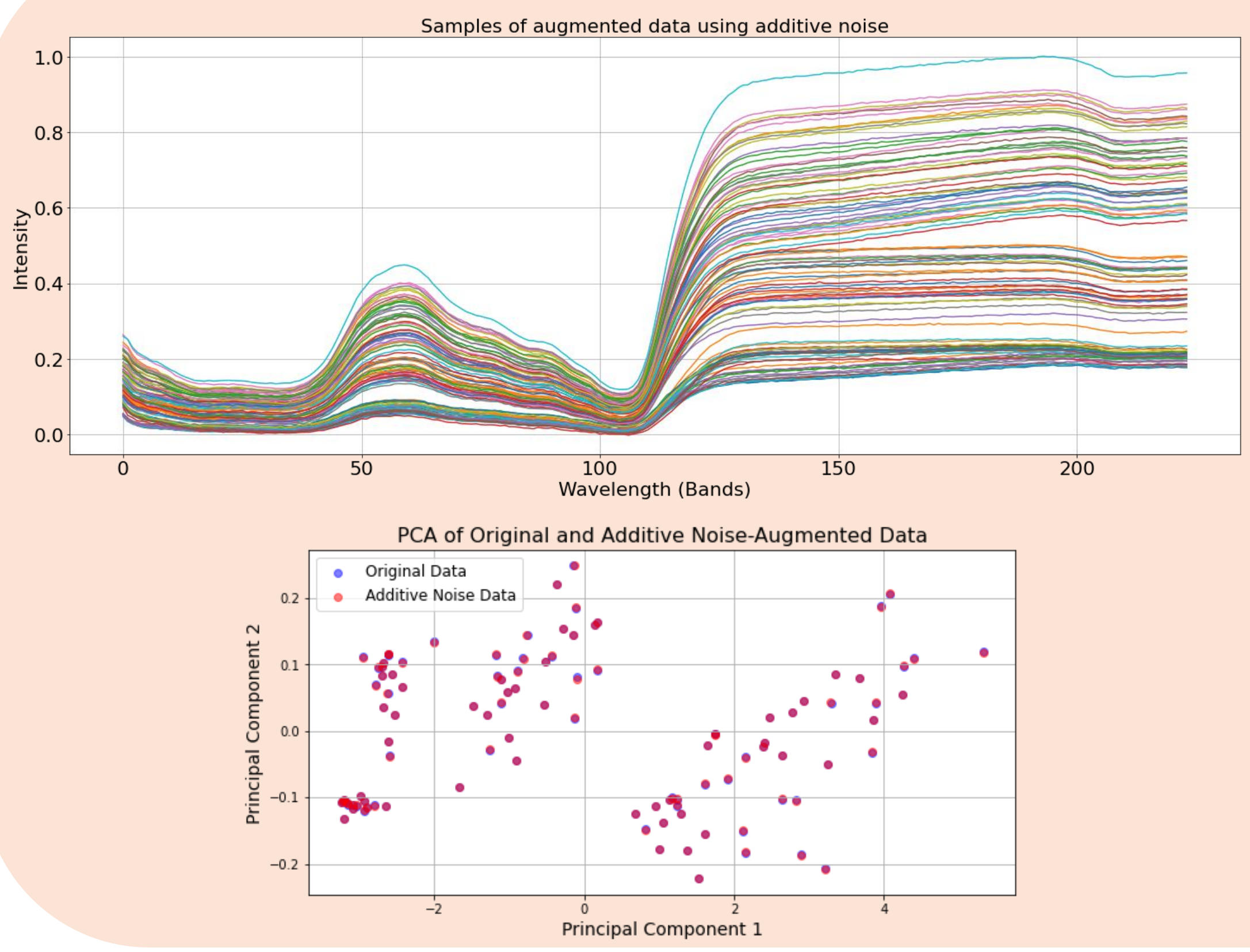


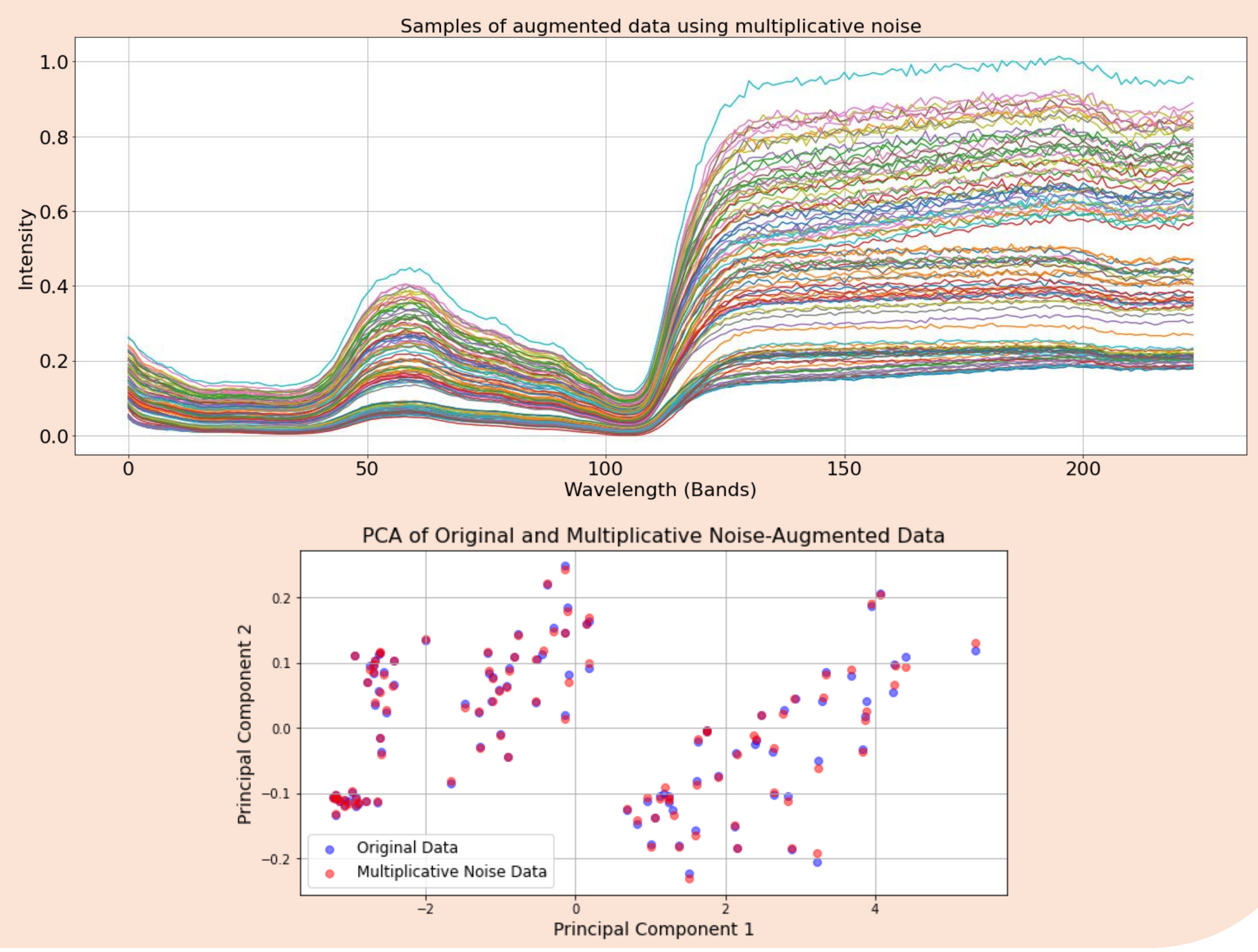
Figure 3. Normalized spectral signatures

Results

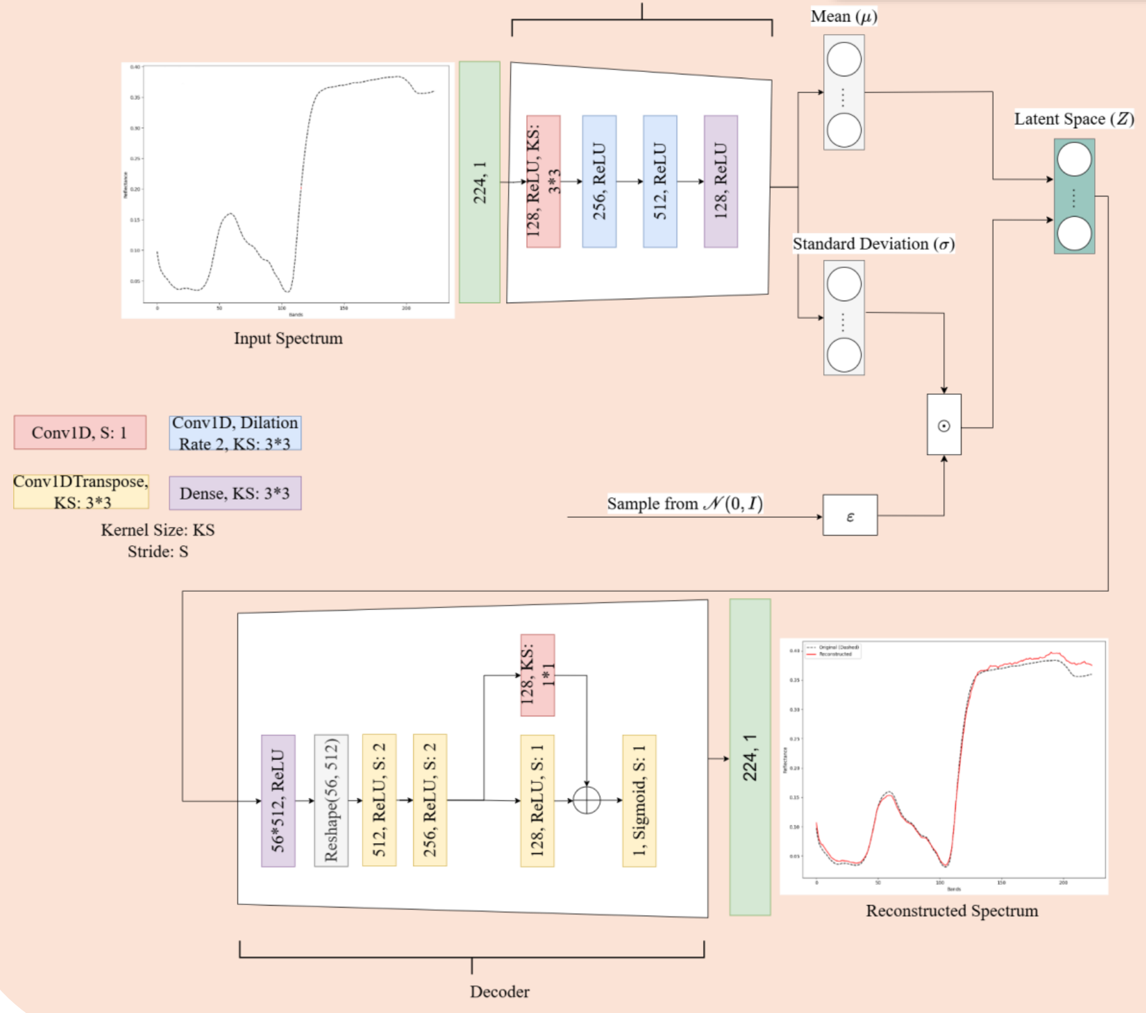
Additive noise: Data + $N(0, \sigma^2)$

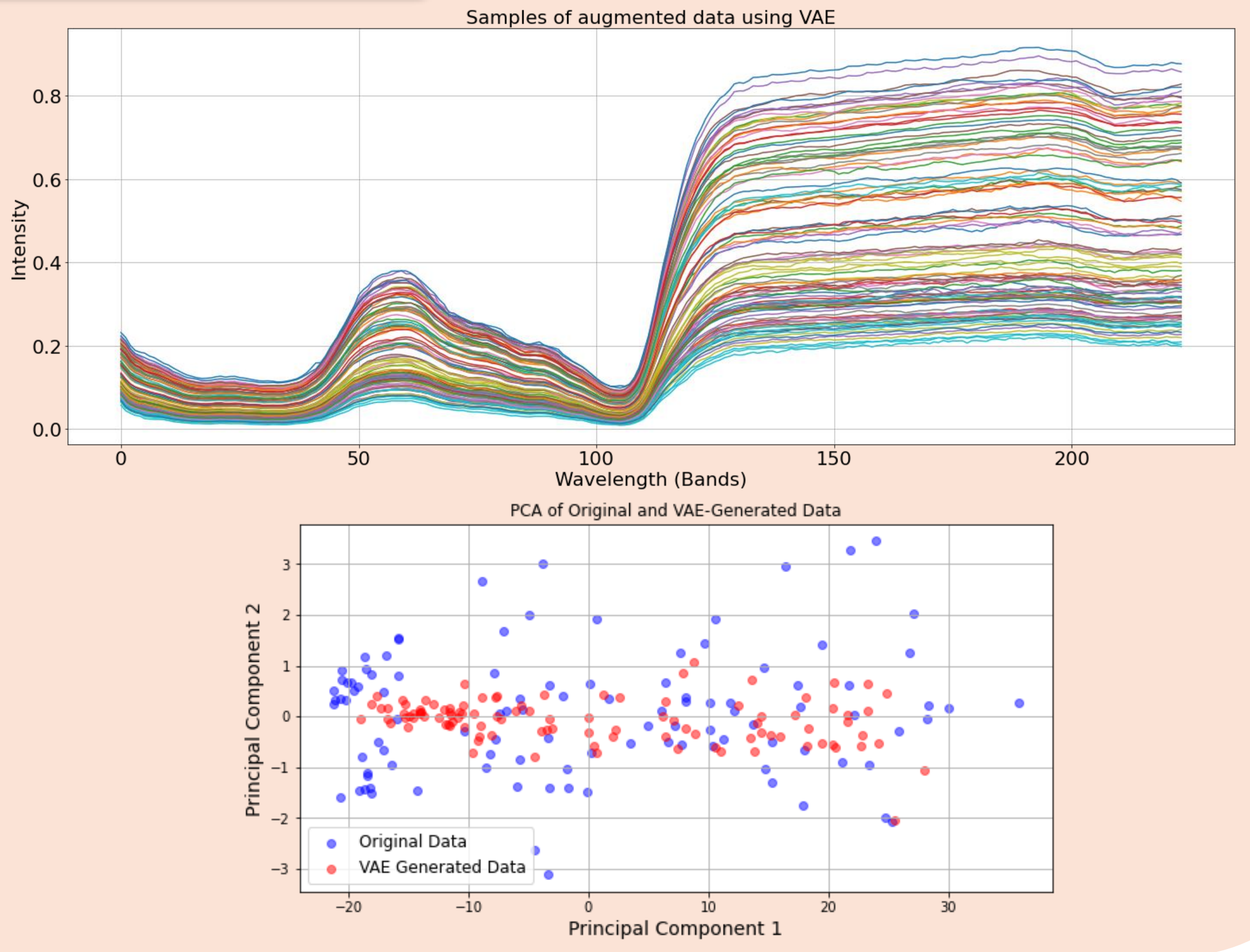


Multiplicative noise: Data × $\exp(N(\mu, \sigma^2))$

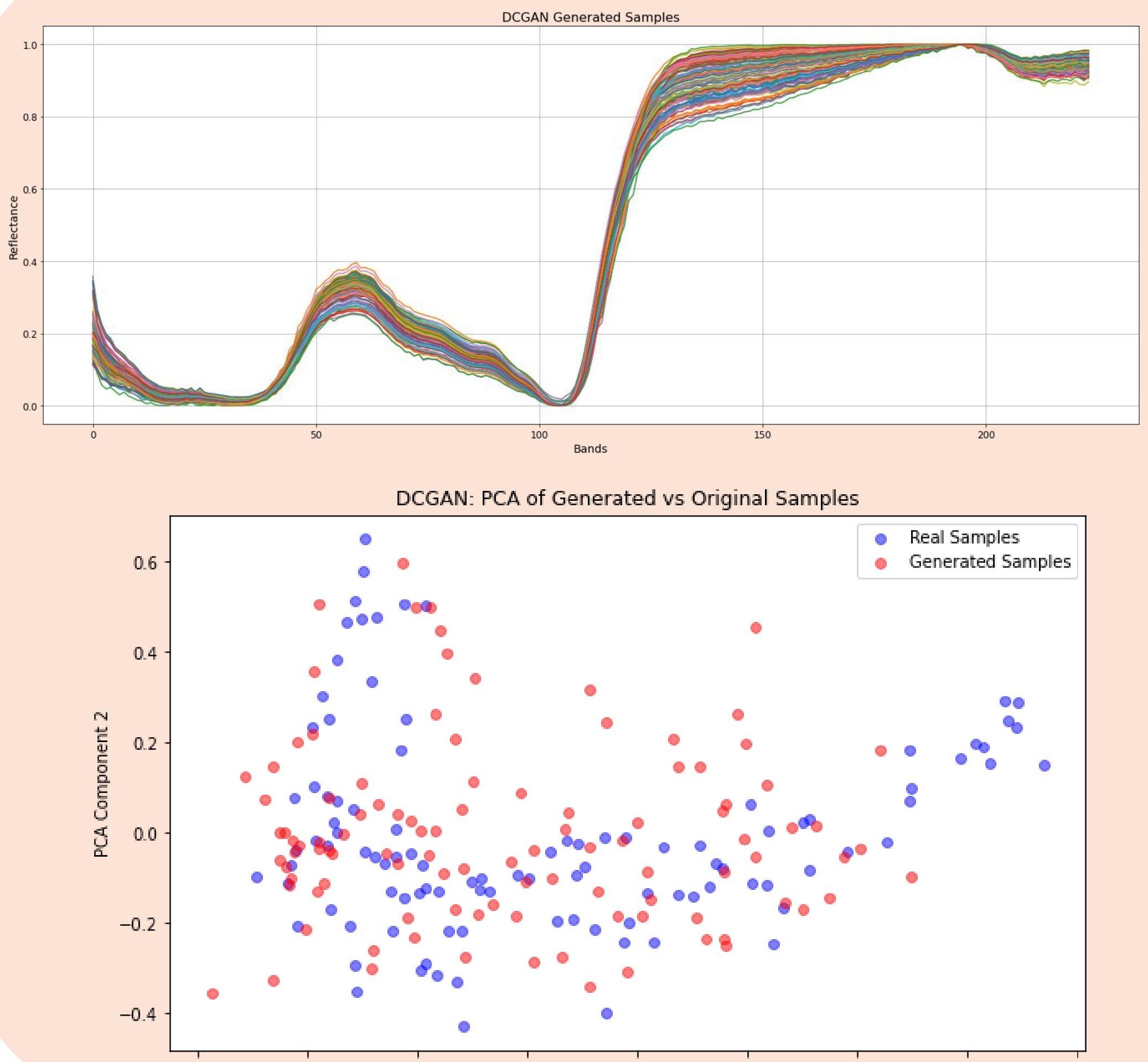


A Customized VAE

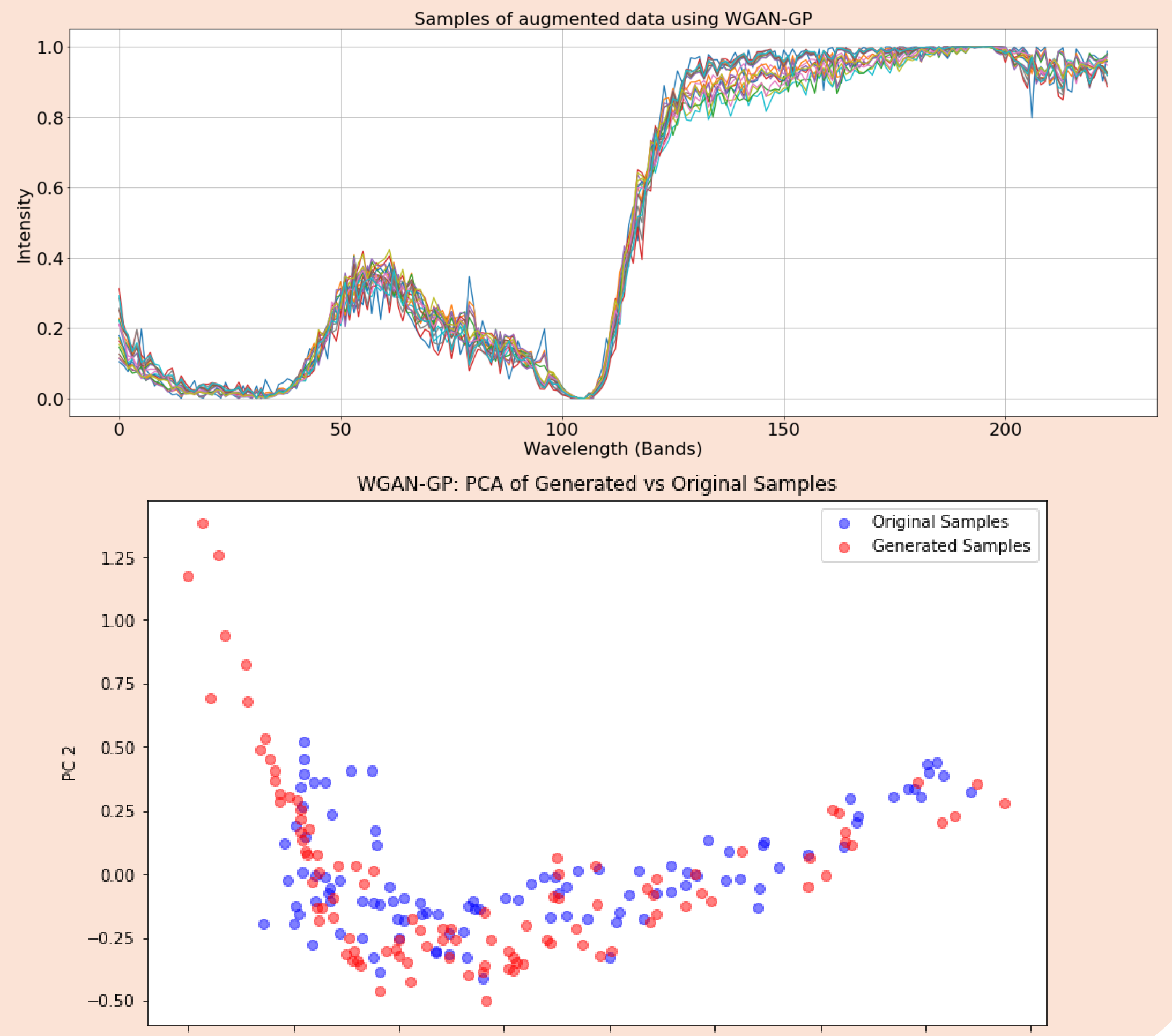




DCGAN



WGAN-GP



Performance Comparison of Data Augmentation Techniques

Metric	Additive Noise	Multiplicative Noise	VAE	WGAN-GP	DCGAN
Cosine Similarity	0.9972	0.9972	0.9985	0.9974	0.9985
Jensen-Shannon Divergence	0.0052	0.0132	0.1908	0.0652	0.0359
KL Divergence	0.0001	0.0007	0.2194	0.0163	0.0053
Nearest Neighbor Distance	0.0149	0.0495	0.1488	0.3759	0.1539

Discussion and Future Work

The comparison of augmentation techniques highlights that additive noise ensures high accuracy in preserving the original data distribution but offers limited diversity, making it suitable for maintaining core data characteristics. In contrast, VAE introduces greater diversity at the cost of data consistency, whereas DCGAN achieves a balance between the two, offering moderate variability with alignment to the original data distribution. While quantitative metrics provide valuable insights, their practical relevance must be validated through downstream tasks such as classification or regression. Building on our findings with deep generative models for spectral data augmentation, future work will focus on integrating spatial information from hyperspectral cubes to capture the full spectral-spatial dependencies of plant data. To achieve this, we propose leveraging transformer-based architectures in combination with diffusion models to generate high-quality synthetic hyperspectral data that accurately represents both spectral and structural characteristics.

References

[1] https://doi.org/10.1016/j.atech.2023.100316
[2] https://doi.org/10.48550/arXiv.1312.6114
[3] https://doi.org/10.48550/arXiv.1511.06434
[4] https://doi.org/10.48550/arXiv.1704.00028
[5] https://hdl.handle.net/10680/2127
[6] https://doi.org/10.3389/fpls.2018.01182

Learn More: Visit the TerraByte Research Group website for all ongoing research projects and related publications.

Scan to connect with Nooshin Noshiri on LinkedIn.