DETECTION OF PLOT-LEVEL VEGETATION DISEASE IN HYPERSPECTRAL IMAGERY USING GRAPH MATRIX EIGENDECOMPOSITION

Jason Salmanowitz, Emmett Ientilucci, and David Messinger

Rochester Institute of Technology
Carlson Center for Imaging Science
54 Lomb Memorial Dr., Rochester, NY 14623, USA

ABSTRACT

Detection of vegetation disease in hyperspectral imagery is a problem of continued interest, particularly in difficult imaging conditions such as very high resolution (sub-leaf) and cases with multiple scattering in the leaf structure. To address this problem, this work introduces a new metric, the Eigenvalue Discounted Manhattan Distance (EDMD), which compares the difference between two eigenvalue distributions as calculated from the ISOMAP embedding of two hyperspectral images. The metric was calculated for vegetation plots and compared to spectral angle mapping (SAM) for the detection of plant disease Cercospora Leaf Spot (CLS). EDMD was able to outperform SAM. At an observed CLS disease severity of 26% and a 5% False Alarm Rate (FAR), SAM has a maximum True Positive Rate (TPR) of 50% while EDMD has a minimum observed TPR of 80% and a likely true TPR above 97.5%.

***Index Terms*—** graph theory, manifold, ISOMAP, hyperspectral, vegetation, spectral angle mapping

1. INTRODUCTION

Detection of vegetation disease in hyperspectral imagery is a problem of continued interest, particularly in difficult imaging conditions such as very high resolution (sub-leaf) and cases with multiple scattering in the leaf structure[1]. Due to the non-linear nature of multiple scattering and vegetation disease effects, most modern processing methods rely on machine learning to capture these effects and approximate the underlying manifold of the data[1]. Instead of using machine learning to approximate the underlying manifold, we propose a metric which directly compares the eigenvalues of graph matrices representing the underlying non-linear manifold of hyperspectral vegetation data as a means to identify vegetation disease. In particular, we use the non-linear dimensionality reduction algorithm ISOMAP[2] to embed images into manifolds and compare the eigenvalue distributions the manifolds.

The ISOMAP algorithm[2] aims to find the underlying manifold of a high-dimensional dataset which preserves the geodesic distance (*i.e.*, the shortest distance) between input data points. The algorithm performs eigendecomposition on the geodesic distance matrix, which is a matrix containing the pairwise distances between input graph nodes. The eigenvalues of the geodesic distance matrix, normalized to a sum of 1, correspond to the fraction of variance captured by each latent dimension of the lower-dimensional manifold estimated by ISOMAP, analogous to the eigenvalues for principal component analysis. If two graphs (and therefore manifolds) have a similar shape, then the eigenvalue distributions representing the graphs will be similar. While ISOMAP is often used for dimensionality reduction[3], the resulting eigenvalues have not been used as feature vectors to compare hyperspectral images. We compare this new metric to the popular spectral angle mapping (SAM) algorithm. SAM is popular for plant disease detection, either as a standalone metric[4] or as integrated into a pre-processing framework[1].

2. METHODOLOGY

2.1. Dataset

Saif et al. (2025)[5] collected hyperspectral reflectance imagery of beet root vegetables infected with Cercospora Leaf Spot (CLS), a fungal plant disease that manifests spots on plant leaves. The dataset contains a total of 40 beet root plots, with imagery collected in the 400nm-1000nm wavelength range across 272 bands and over 5 days. The CLS disease severity, scored between 0% and 100%, was evaluated for each plot and observation date by visually estimating the CLS-affected area across leaf samples[6].

To compare healthy vegetation plots with diseased vegetation plots, 18 plots with an observed disease severity of 0% from the first day of data collection were selected to serve as a baseline for healthy vegetation. Representative mean reflectance spectra from the healthy vegetation plots are shown in Figure 1. The mean normalized difference vegetation index (NDVI) for the the healthy vegetation plots ranged from 0.44 to 0.68. The mean reflectance spectra of diseased vegetation were selected from the fifth day of data collection, with observed diseased severities ranging between 10% and 88%

across all 40 plots. The mean NDVI for the diseased vegetation plots ranged from 0.48 to 0.80. NDVI ranges for the healthy vegetation and the diseased vegetation overlap, with many diseased vegetation plots having a higher mean NDVI than the healthy vegetation as shown in Figures 1 and 2.

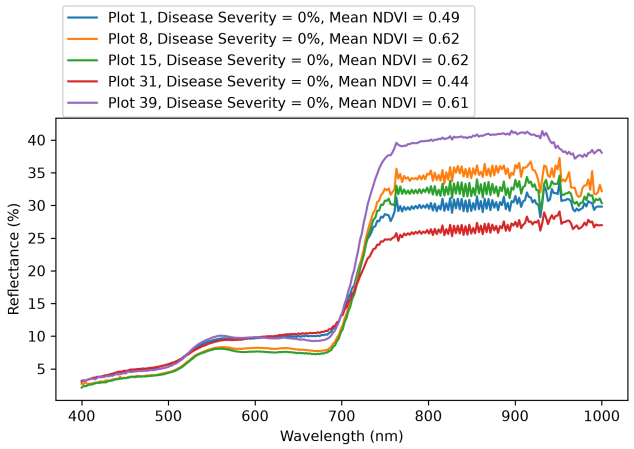


Fig. 1. Example mean reflectance spectra from healthy beet root plots collected by Saif et al.(2025)[5]. Healthy plots correspond to plots with an observed Cercospora Leaf Spot (CLS) disease severity of 0%. The NDVI of the healthy vegetation plots ranged from 0.44 to 0.68.

2.2. Detection Metrics

For this work, graphs were generated from the pixels in hyperspectral images of each individual beet root plot discussed in Section 2.1 using $k = 10$ nearest neighbors and a weighted connectivity, using cosine similarity as the distance metric to facilitate comparison with SAM. To enable the direct comparison of the eigenvalue distributions between graphs (*i.e.*, graphs 1 and 2), we introduce the Eigenvalue Discounted Manhattan Distance (EDMD). The EDMD is defined as

$$\text{EDMD}(\vec{\nu}_1, \vec{\nu}_2) = \sum_{i=1}^{n} \gamma^n |v_{1i} - v_{2i}|, \tag{1}$$

where $\vec{\nu}_1$ is the eigenvalue distribution of graph 1 (consdiered as a vector of eigenvalues), $\vec{\nu}_2$ is the eigenvalue distribution of graph 2, and $|\cdot|$ represents the euclidean norm. γ is a hyperparameter which can be tuned to assign higher weights in the sum to either lower index ($\gamma < 1$) or higher index ($\gamma > 1$) eigenvalues. For this work, $\gamma = 1.1$ was empirically chosen because the differences between eigenvalue distributions for diseased vegetation occur at higher index eigenvalues as demonstrated in Figure 3.

To serve as a baseline for comparison, the EDMD was compared to the Spectral Angle Mapping (SAM) algorithm,

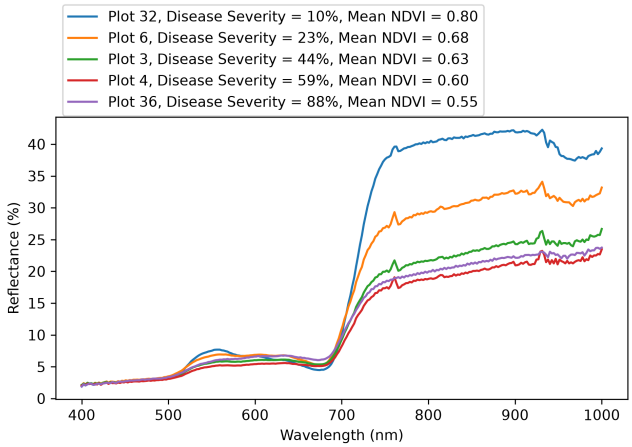


Fig. 2. Example mean reflectance spectra from diseased beet root plots collected by Saif et al.(2025)[5]. The diseased vegetation plots have an observed Cercospora Leaf Spot (CLS) disease severity between 10% and 88%. The NDVI of the diseased vegetation plots ranged from 0.48 to 0.80.

which was used to calculate the spectral angle distance between two mean reflectance spectra. The SAM between two reflectance spectra $\vec{\rho_1}$ and $\vec{\rho_2}$ is defined as

$$\text{SAM}(\vec{\rho_1}, \vec{\rho_2}) = \arccos\left(\frac{\langle \vec{\rho_1}, \vec{\rho_2} \rangle}{|\vec{\rho_1}||\vec{\rho_2}|}\right), \tag{2}$$

where $\langle \cdot, \cdot \rangle$ represents the dot product.

2.3. Statistics

For metrics EDMD and SAM, a baseline "healthy" distribution was determined by calculating the each metric across every pair of healthy plots. For 18 plots, this corresponded to 153 unique samples. Each sample set was converted into a histogram which was then fit to a folded normal distribution, *i.e.*, a distribution where the negative x-axis portion of a normal distribution is "folded" onto the positive portion of the x-axis. A folded normal distribution was selected because the mean for EDMD and SAM across all healthy plot pairs was expected to be near-zero and both EDMD and SAM are bounded to be greater than or equal to 0. A decision boundary was assigned 2 standard deviations away from the mean, corresponding to a 5% False Alarm Rate (FAR).

The EDMD and SAM distributions for each diseased plot were calculated using the following procedure: Each diseased plot had a unique disease severity. The EDMD and SAM were calculated between the diseased plot and each individual healthy plot, resulting in 18 samples per disease severity. As with the healthy distribution, the sample set for each metric was converted into a histogram which was fit to a folded normal distribution. The mean and standard deviation from the

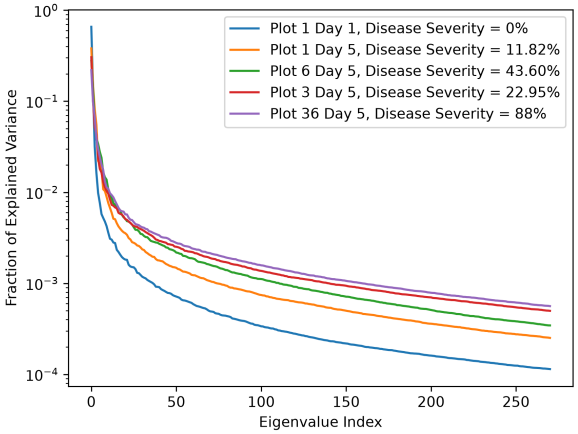


Fig. 3. Example eigenvalue distributions for graphs representing hyperspectral images of beet root plots across various disease severities. Note that the differences between eigenvalue distributions for diseased vegetation become more visible as eigenvalue index increases.

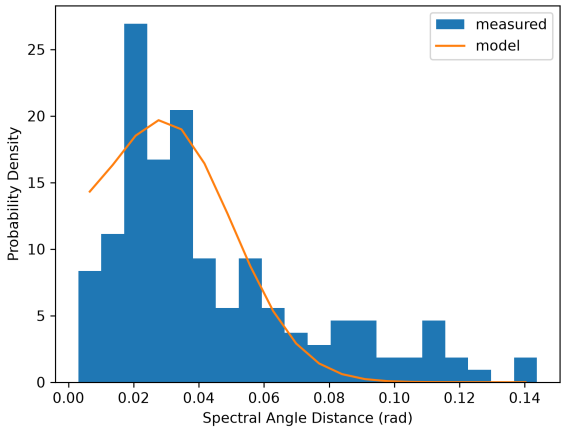


Fig. 4. Measured and modeled probability density of the spectral angle distance between each pair of mean reflectance spectra of healthy beet root plots.

fit were used to generate 95% confidence intervals centered on the mean for each disease severity. Cases where the confidence interval for a given disease severity lies entirely above the decision boundary for healthy vegetation correspond to a True Positive Rate (TPR) of at least 97.5%.

Plots were generated showing EDMD and SAM vs disease severity, containing confidence intervals for each disease severity and a line along the y-axis marking the decision boundary. Linear regression lines and corresponding R^2 values were calculated for each plot. A validation of the folded normal distribution fits is presented in Section 3.1 and a demonstration of using these distributions to detect plant disease is presented in Section 3.2.

3. RESULTS

3.1. Statistics Validation

The measured and modeled SAM distributions for healthy plots are shown in Figure 4. For the healthy distributions, we observed the measured distribution had a sharper peak and a heavier tail as compared to the modeled distribution, which is characteristic of a folded Student's t-distribution. Considering that a Student's t-distribution converges to a normal distribution with increasing sample size and the modeled mean appeared well-centered, the folded normal distribution is deemed to be an appropriate model for the distribution SAM distribution of healthy vegetation.

An example measured and modeled SAM distribution measured between a diseased plot and all healthy plots is shown in Figure 5. Similar to the healthy distributions, we observed a sharper peak and heavier tail in the measured distribution as compared to the modeled distribution, once again characteristic of a folded Student's t-distribution. In contrast to the healthy distributions, the measured distribution showed zero data points below the mean, suggesting the distribution was right-skewed, which was not incorporated into the model. Due to the skewness of the measured distribution, the model may have underestimated the lower bound of the confidence interval. Because the lower bound is being compared to the decision boundary to find the TPR, this error may have been significant and therefore the mean was prioritized for analyzing the TPR in Section 3.2.

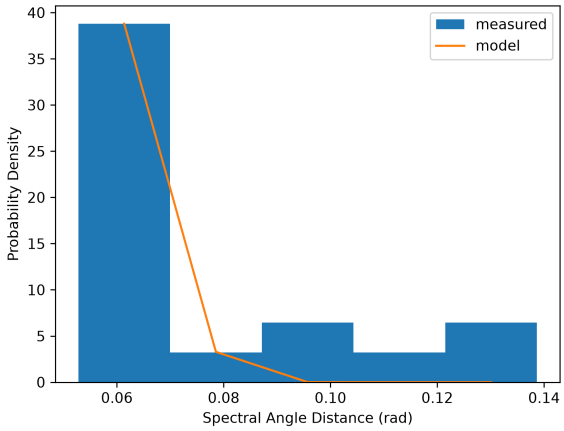


Fig. 5. Example measured and modeled probability density of the spectral angle distance between the mean reflectance spectrum a plot with an assessed disease severity of 23% and the mean reflectance spectrum of healthy beet root plots.

The measured and modeled EDMD distributions for healthy plots are shown in Figure 6 and an example measured and modeled EDMD distribution measured between a diseased plot and all healthy plots is shown in Figure 7. As with the SAM distributions, the measured healthy distribution for EDMD showed characteristics of a folded Student's t-distribution, which demonstrates that a folded normal distribution is an appropriate model since the t-distribution converges to a normal distribution as sample increases. In contrast to the example SAM distribution for a diseased vegetation plot, the example EDMD distribution had few values above the mean, which indicates that the distribution is left-skewed. Therefore, the model may have overestimated the upper bound of the confidence interval, which was not used to calculate the TPR.

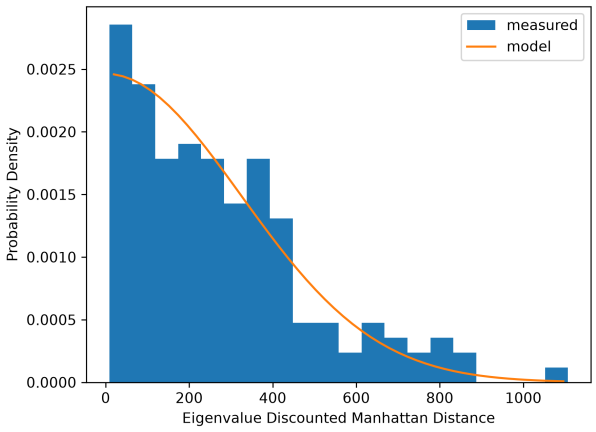


Fig. 6. Measured and modeled probability density of the Eigenvalue Discounted Manhattan Distance between each pair of eigenvalue distributions representing healthy beet root plots.

3.2. Diseased Vegetation Detection

The distribution of SAM for healthy vegetation had a modeled mean of 0.03 and a modeled standard deviation 0.02. The decision boundary corresponding to a 5% FAR was 0.07. Figure 8 shows the healthy mean, the decision boundary (*i.e.*, the healthy 95% confidence interval), and the modeled SAM distributions vs. disease severity along with a linear regression line ($R^2 = 0.38$) based on the SAM distribution means. The line intersects the decision boundary at 26% disease severity.

The distribution of EDMD for healthy vegetation had a modeled mean of 1 and a modeled standard deviation 324. The decision boundary corresponding to a 5% FAR was 629. Figure 9 shows the healthy mean, the decision boundary (*i.e.*, the healthy 95% confidence interval), and the modeled EDMD distributions vs. disease severity along with a linear regression based on the EDMD distribution means. In contrast to the SAM distributions, the size of the confidence interval was roughly constant across disease severity, with a mean of 451 and a standard deviation of 33 (*i.e.*, 7% variation). The linear regression line ($R^2 = 0.20$) was a visibly poor fit for the underlying data.

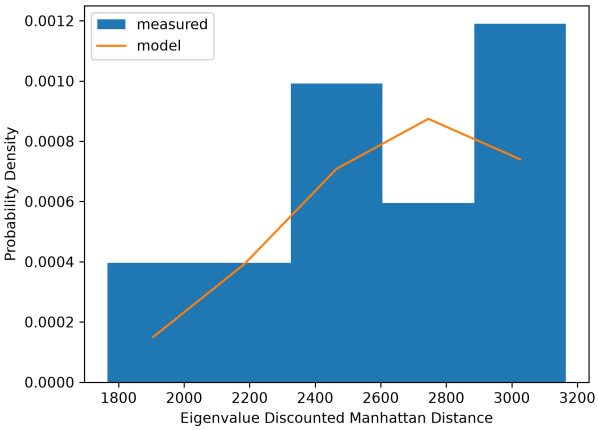


Fig. 7. Example measured and modeled probability density of the Eigenvalue Discounted Manhattan Distance between the eigenvalue distribution representing a plot with an assessed disease severity of 23% and the eigenvalue distributions representing healthy beet root plots.

4. DISCUSSION

As discussed in Section 3.1, the right-skewness of the SAM distributions for diseased vegetation may have resulted in an underestimated lower bound for the distribution's 95% confidence interval. Using only the means to find the linear regression line and assuming the real distribution is centered on the linear regression line, the intersection with the decision boundary corresponds to a 50% TPR and a 5% FAR at 26% disease severity. The R^2 value of the line, 0.38, indicates that the majority of the variance is not explained by this linear model and that the modeled TPR may be an overestimate. For example, note that certain means shown in Figure 8 appear below the decision boundary up to 65% disease severity.

In contrast, with one exception, every confidence interval representing the diseased vegetation EDMD distribution (as shown in Figure 9) lies above the decision boundary beginning at 25% disease severity, corresponding to a minimum TPR of 97.5%. That is, absent the single outlier at 33% disease severity where the distribution overlaps the decision boundary, at a 26% disease severity, SAM results in a modeled 50% TPR (with the TPR likely being lower) while EDMD results in an empirical 97.55% TPR at minimum. Considering the outlier, that distribution would have an ap-

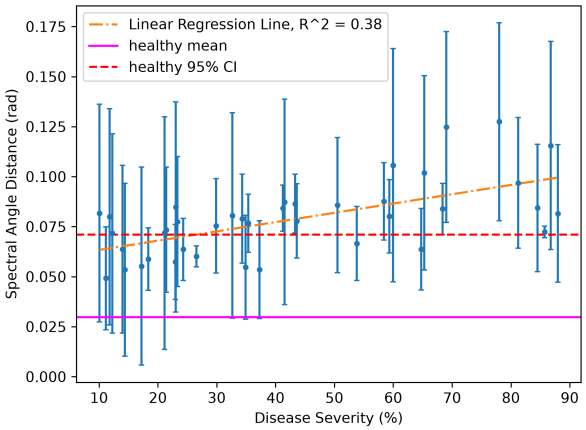


Fig. 8. Healthy vegetation distribution mean, healthy distribution 95% confidence interval (corresponding to the decision boundary for a 5% FAR), and diseased distributions vs. disease severity for spectral angle mapping (SAM).

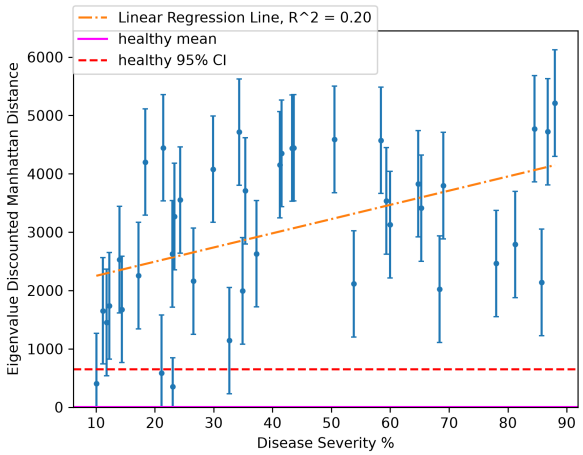


Fig. 9. Healthy vegetation distribution mean, healthy distribution 95% confidence interval (corresponding to the decision boundary for a 5% FAR), and diseased distributions vs. disease severity for eigenvalue discounted manhtaan distance (EDMD).

proximately 80% TPR at the decision boundary, which still outperforms SAM. The roughly constant size of the confidence interval also suggests that the distribution of EDMD can be predicted. While a linear regression was visibly a poor fit to the measured EDMD distribution, a sigmoid function fit and a known standard deviation may enable the predictive modeling of the EDMD distribution.

5. CONCLUSION

The Eigenvalue Discounted Manhattan Distance (EDMD), a new metric which compares the difference between two eigenvalue distributions as calculated from the ISOMAP embedding of two hyperspectral images, outperforms the spectral angle mapping (SAM) algorithm for the detection of plant disease Cercospora Leaf Spot (CLS) in hyperspectral imagery of the vegetation plots analyzed here. At a disease severity of 26%, SAM has a maximum True Positive Rate (TPR) of 50% while EDMD has a minimum observed TPR of 80% and a likely true TPR above 97.5%. Future work is needed to evaluate the utility of non-parametric statistical tests in order to generalize this approach to generic and non-gaussian data. Future work is also required to validate this approach on other plant disease datasets and to explore other use cases, such as agriculture and environmental monitoring.

6. REFERENCES

[1] R Selvaprakash, NS Rabi, K Aundy, R Rajeev, and C Viswanathan, "Early detection of rice blast disease using hyperspectral remote sensing," *Plant Science Today*, vol. 13, pp. 11111, 2026.

[2] JB Tenenbaum, V de Silva, and JC Langford, "A global geometric framework for nonlinear dimensionality reduction," *Science*, vol. 290, no. 5500, pp. 2319–2323, 2000.

[3] X Zhao, S Zhang, W Yan, and X Pan, "Multi-temporal grass hyperspectral classification via full pixel decomposition spectral manifold projection and boosting active learning model," *International Journal of Remote Sensing*, vol. 44, no. 2, pp. 469–491, 2023.

[4] H Li, WS Lee, R Wang, R Ehsani, and C Yang, "Spectral angle mapper (sam) based citrus greening disease detection using airborne hyperspectral imaging," in *Proc. 11th Intl Conf. on Precision Agriculture. Monticello, Ill.: International Society of Precision Agriculture*, 2012.

[5] MS Saif, R Chancia, SP Murphy, S Pethybridge, and J van Aardt, "Exploring UAS imaging modalities for precision agriculture: predicting table beet root yield and estimating disease severity using multispectral, hyperspectral, and LiDAR sensing," in *Algorithms, Technologies, and Applications for Multispectral and Hyperspectral Imaging XXXI*, Miguel Velez-Reyes and David W. Messinger, Eds. International Society for Optics and Photonics, 2025, vol. 13455, p. 1345509, SPIE.

[6] MS Saif, R Chancia, P Sharma, SP Murphy, S Pethybridge, and J van Aardt, "Estimation of cercospora leaf spot disease severity in table beets from uas multispectral images," *Computers and Electronics in Agriculture*, vol. 248, pp. 111807, 2026.