Water Potential and Relative Water Content Prediction of Wild Blueberries during Drought Treatment Using Hyperspectral Sensor and Machine Learning

Trang Tran*[1], Kallol Barai*[2], Yong-Jiang Zhang[2,4], Vikas Dhiman[3], Umesh R. Hodeghatta[1]

[1]CPS-Analytics-Applied Machine Intelligence, Northeastern University, United States
[2]School of Biology and Ecology, University of Maine, United States
[3]Electrical and Computer Engineering, University of Maine, United States
[4]Climate Change Insitute, University of Maine, United States

A paper from the Proceedings of the
16th International Conference on Precision Agriculture
21-24 July 2024
Manhattan, Kansas, United States

Abstract.

Water availability critically influences crop growth, health, and yield. Though accurate, traditional methods for assessing plant water status and water stress, such as leaf water potential (LWP) and relative water content (RWC), are destructive and unsuitable for large-scale monitoring. This study investigates the use of hyperspectral sensing combined with machine learning (ML) to predict water stress in wild blueberries non-destructively during drought treatment. A drought experiment was conducted on wild blueberries in the summer of 2022, using a randomized block design with six genotypes under biochar and drought treatments. Physiological measurements of LWP and RWC were collected alongside hyperspectral data using the SVC HR-1024i sensor, covering the spectral range of 350-2500 nm. Various data mining, feature selection, and feature engineering techniques were implemented to address the imbalanced target variable and high dimensionality issues. We explored the optimal wavelength bands of spectral indices such as simple differences (Rλ1-Rλ2), simple ratios (Rλ1/Rλ2), normalized differences (|Rλ1-Rλ2|/(Rλ1+Rλ2)), and MDATT ((Rλ3−Rλ1)/(Rλ3−Rλ2)) for both LWP and RWC. They also emerged as top predictors for predicting water stress, significantly contributing to the highest-performing models. Our models, particularly Kernel Ridge Regression (KRR), XGBoost, and Gradient Boosting, demonstrated high predictive accuracy for LWP and RWC, with R-squared values ranging from 82% to 95% and normalized root mean square error (NRMSE) values between 7% and 12%. RWC regression predictions consistently outperformed LWP, with the KRR model for RWC achieving an R-squared of 95% and an NRMSE of 7.37%. For LWP, the Gradient Boosting model with selected non-linear features yielded an R-squared of 90% and an NRMSE of 9.47%. In the upper ranges of both target variables, models performed exceptionally well, with R-squared values surpassing 95% and NRMSE values below 3%. However, in the lower range of LWP, all models showed poor performance, with R-squared values below 50%. In contrast, the

lower region of RWC showed more promise, with the XGBoost model achieving a 68% R-squared value. Additionally, Random Forest classification models for binary targets demonstrated accuracy scores of 75% for RWC and 83% for LWP, indicating potential for refining water stress classification using appropriate thresholds. Future research could be done to explore further specific regression models tailored to distinct regions of water availability and to develop an automated neural vegetation index that could work across species and predict different physiological parameters.

Keywords.
Hyperspectral Reflectance, Leaf Water Potential (LWP), Relative Water Content (RWC), Spectral Indices, Machine Learning (ML), Wild Blueberries, Optimal Bands, Drought Treatment.

Introduction

Water availability significantly influences crop growth, development, health, and yield (Taiz et al., 2015). Water deficits can negatively impact crop physiology, resulting in reduced growth and overall production (Aladenola & Madramootoo, 2014; Rossini et al., 2013). It is crucial to accurately measure or estimate plant water status to make informed decisions about managing crop water stress and improving crop production. Although there are many traditional methods for determining soil and plant water status, most of these methods are destructive, labor-intensive, time-consuming, and not suitable for continuously monitoring large commercial fields with varying soil properties (Ihuoma & Madramootoo, 2017). For example, although leaf water potential (LWP) and relative water content (RWC) measurements can accurately indicate crop water status , they are destructive and labor-intensive..

While traditional destructive methods through field sampling provide accurate estimations of water stress indicators such as LWP and RWC, they are not always practical for estimating water stress over a large heterogeneous field. Non-destructive measurement of leaf spectral reflectance offers an instantaneous and practical method for assessing the water stress of plants. This method involves measuring the light reflected by leaves, which varies according to the water availability. By analyzing the spectral reflectance of the leaves, we can determine the water stress of the plants. This non-destructive approach is particularly useful for quick, accurate, large-scale vegetation health evaluations, making it a feasible solution for water status monitoring for precision irrigation in crop fields.

Hyperspectral remote and contact sensing has emerged as a promising avenue for monitoring various plant parameters, including water stress, due to its non-destructive nature and ability to cover large spatial extents (Mulla, 2013). When plants are water-stressed for a prolonged period, their chlorophyll production might be reduced, which will result in decreased absorption in VIS (400–700 nm) and, thus, increased reflectance in the VIS region (Jensen, 2009; Jones & Vaughan, 2010). Due to the decrease in chlorophyll production, a blue shift (towards a shorter wavelength) of the red-edge position may be observed. These properties can be used to monitor the effects of water stress on vegetation. The water content in leaves influences light scattering. Increased scattering due to high water content generally enhances the diffusion and transmittance of leaves, which leads to reduced reflectance in the NIR (700 -1200 nm) and SWIR (1200-2500 nm) in well-watered leaves compared to reduced water leaves (Jensen, 2009; Jones & Vaughan, 2010). So, water-stressed leaves may show overall high reflectance compared to well-water leaves. However, if the stress is prolonged and leads to a reduction in leaf area and biomass, the reflection in the NIR region may decrease, and NIR is sensitive to biomass and canopy density (Jensen, 2009; Jones & Vaughan, 2010). Overall, hyperspectral sensing is promising as a non-destructive water status estimation method. However, the challenge lies in developing water stress prediction models that can accommodate the inherent variability in structural and physiological properties across different crops.

The wild blueberry crop, with its diverse genotypes grown in semi-natural systems, poses a significant challenge for precision water management. To tackle this, we have delved into the application of machine learning (ML) techniques for non-destructive hyperspectral sensing-based water stress detection in wild blueberries. Our objective was to assess the performance of ML models in accurately predicting water stress during drought treatment, taking into account the unique characteristics and variability inherent in wild blueberry genotypes. This research is a significant step towards advancing water stress monitoring applications in agriculture, particularly in the context of wild blueberries, by bolstering the robustness and adaptability of water stress prediction models through the use of hyperspectral sensing and machine learning techniques.

Materials and Method

In the summer of 2022, a drought experiment was conducted on wild blueberries at Rogers Farm Greenhouse in Old Town, Maine, USA (Longitude: -68.69° N, Latitude: 44.93° W). The experiment

utilized a randomized block design with two factors (biochar treatment and drought treatment) and six genotypes implemented in pots within the greenhouse. The experiment spanned from July 3rd to August 2nd, with data collection occurring every three days or longer, based on measurements of leaf water potential (LWP) and relative water content (RWC), and the rate of plant drying. The control crop was irrigated twice a day, and the drought treatment crop underwent natural dry-down by withholding irrigation. The control crops were irrigated in the morning and early evening using irrigation lines.

Plant Physiological Measurements

Physiological data was collected from different genotypes and soil blocks with various crop irrigation treatments, with the target variables being LWP and RWC. At least 12 leaf samples were taken on each sampling date. All samples were collected between 11:00 and 14:00, within a 30-minute window, and were then placed in sealed bags, kept in a dark cooler, and transported to the University of Maine Plant Physiology Laboratory within 10 to 15 minutes of collection. Midday LWP was measured using a leaf pressure chamber (Model 1505D; PMS Instrument Company, Corvallis, OR USA). RWC measurements were calculated using the formula below.

$$\mathrm{RWC}(\%) = \frac{\mathrm{W-DW}}{\mathrm{TW-DW}}$$

where W = sample fresh weight, TW = sample turgid weight, DW = sample dry weight.

Fresh leaf samples were initially weighed to determine the leaf sample weight (W). Subsequently, the samples were hydrated to full turgidity for 4 hours under normal room light and temperature. After this hydration period, the samples were taken out of the water, gently dried with filter paper to eliminate surface moisture, and promptly reweighed to establish the fully turgid weight (TW). Following this, the samples underwent oven-drying at 80°C for 72 hours and were then weighed to ascertain the dry weight (DW).

Hyperspectral Reflectance Measurements

Hyperspectral data for each leaf sample were collected in coordination with physiological measurements using a handheld hyperspectral sensor (SVC HR-1024i). Spectral measurements were obtained prior to any physiological assessments. The SVC HR-1024i operates over a spectral range of 350-2500 nm.

Data Preprocessing

For our research, we employed the 'specdal' Python package to parse individual SIG-formatted files containing hyperspectral data, subsequently organizing them into a structured data frame. This hyperspectral data was then integrated with corresponding Leaf Water Potential (LWP) and Relative Water Content (RWC) datasets using shared columns such as “Date”, “Drought Status”, “Genotype”, “Block”, and “Biochar Treatment”.

Regarding the LWP target variable, our dataset includes 121 observations with the 'no biochar' treatment, of which 53 samples have a 'drought' status. For the RWC target variable, the dataset comprises 176 samples across both treatment groups, with 38 samples displaying a 'drought' status, all of which belong to the 'no biochar' treatment group. Each dataset contains 994 columns, representing reflectance rates across 994 wavelength bands ranging from 339 nm to 2516 nm.

This research aims to investigate the relationship between spectral characteristics and the LWP and RWC of wild blueberries under drought conditions. To achieve this, we selectively refine the training dataset to include only leaves experiencing drought stress without any biochar treatment.

Research Methodology

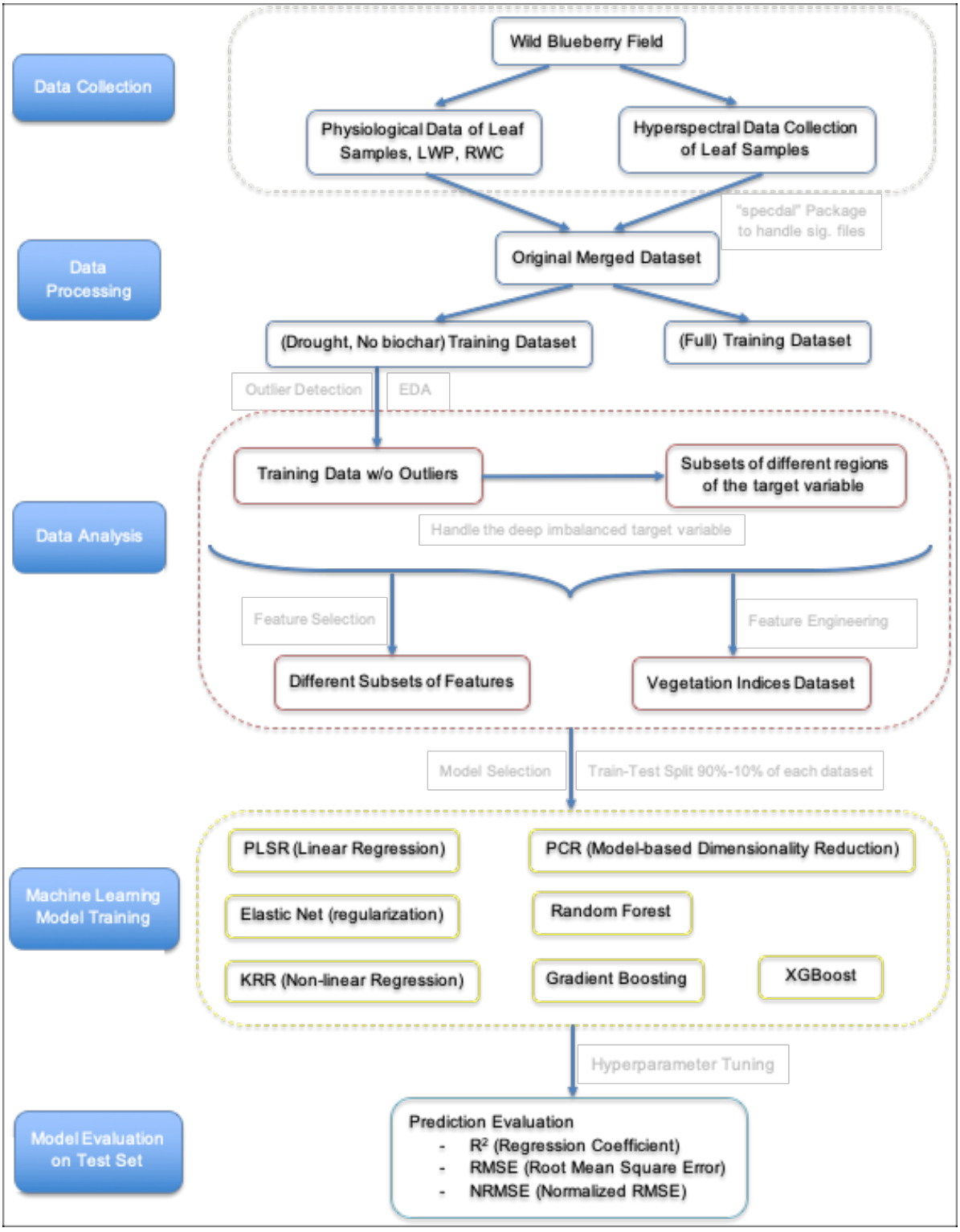

Figure 1: Scheme of methodology used in this study

In this study, we employed a systematic approach to extract refined training datasets. After many experiments on various approaches, we have come up with the appropriate method which brings the best results in this data field. Initially, we conducted outlier detection and exploratory data analysis to clean the raw data and identify potential issues that could hinder subsequent model training. Following this, we performed feature selection and feature engineering to create various

subsets of features and develop new training datasets based on spectral vegetation indices. Further details are presented in Figure 1.

Exploratory Data Analysis

Figure 2 illustrates the changes in average Leaf Water Potential (LWP) and Relative Water Content (RWC) over ten days of drought treatment. The data reveal an overall downward trend throughout the period, with a slight increase around the two-thirds mark (7/21/22) before experiencing a significant decline in the final days.

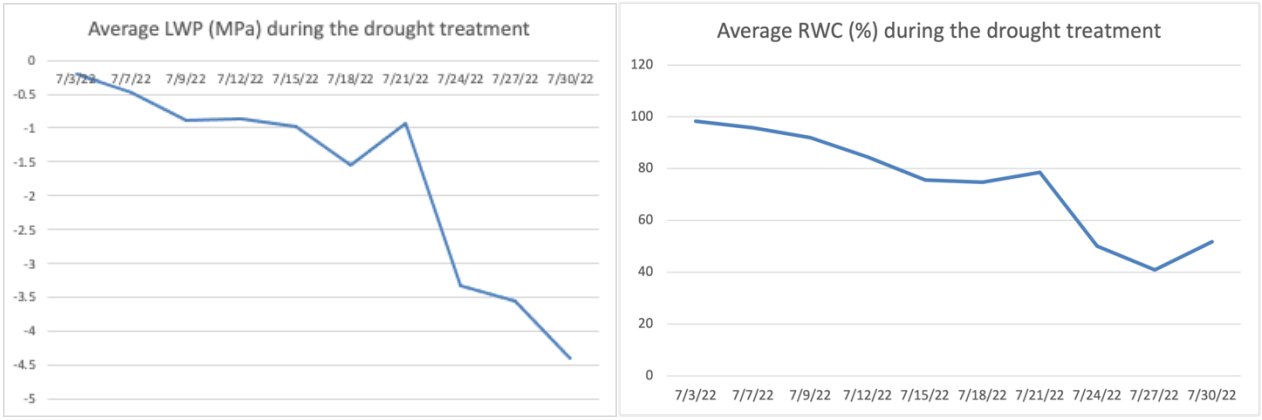

Figure 2: Changes in (a) Leaf Water Potential (LWP) and (b) Relative Water Content (RWC) of wild blueberry plants during the drought treatment

Figure 3 presents the mean reflectance rates across all wavelengths within the 339 nm to 2516 nm range over a ten-day period. The plot reveals a significant divergence in the trend on the final day (7/30/22), as highlighted.

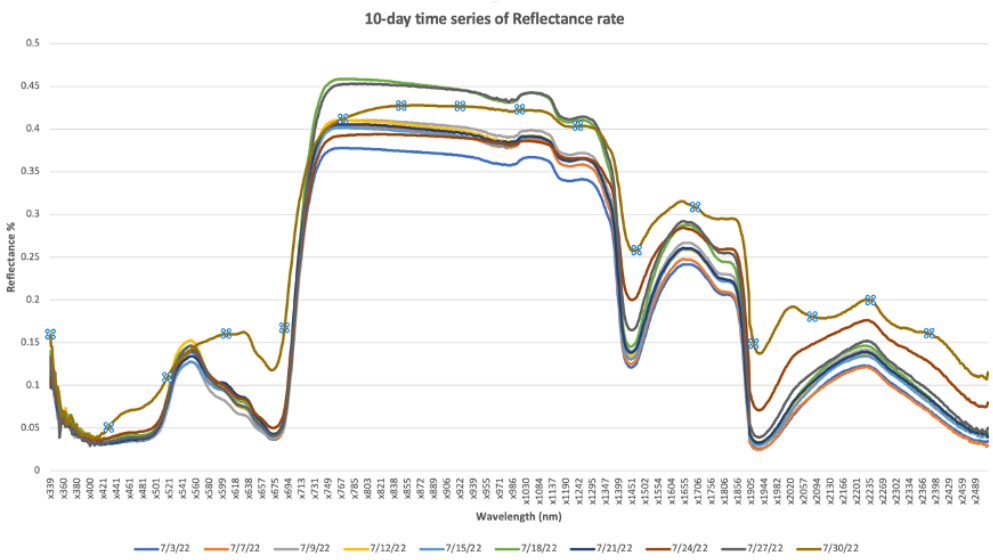

Figure 3: Time series of reflectance rate (%) of wild blueberry leaves during the drought treatment

We aim to examine the noise within the training data, which may affect the model training and prediction processes. Figure 4 shows the mean reflectance across all wavelength bands during the drought period. Our analysis indicates that there is no substantial noise within this dataset. We applied the Savitzky-Golay filter, Gaussian kernel smoother, and Wiener filter to assess their impact on noise reduction (see Appendix). However, due to the inherently low noise in the dataset, these filters did not significantly transform the original data.

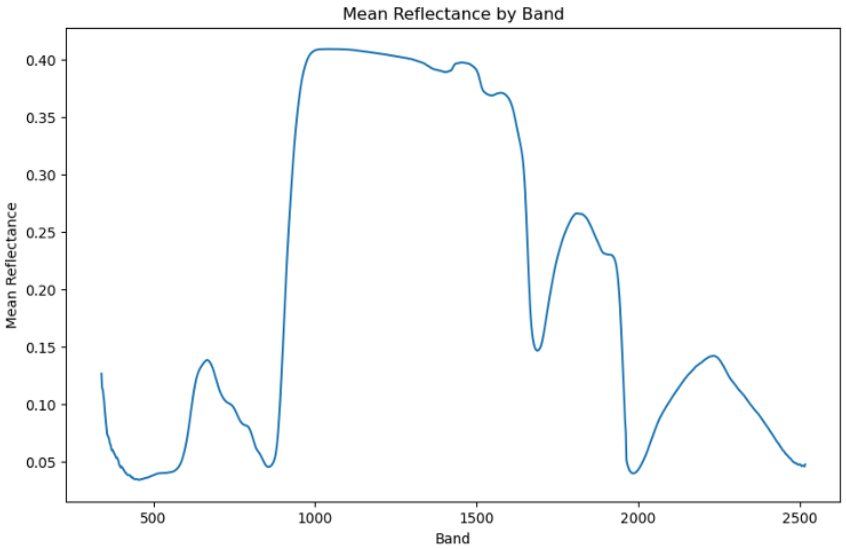

Figure 4: Mean reflectance rate (%) by Wavelength Band (nm)

Handling imbalanced continuous target variables

Managing a deeply imbalanced continuous target variable in machine learning requires robust strategies to ensure effective model learning. Imbalanced data can lead to biased models that fail to generalize well. This study addressed this issue by combining regression and classification techniques.

Initially, we trained regression models directly on the continuous target variables. However, due to the severe imbalance, we applied a binning method to discretize the continuous target variable into distinct bins based on specific thresholds. This discretization transformed the regression problem into a classification task, simplifying handling imbalanced data (Fawcett & Provost, 1997).

By converting the continuous target into bins, we trained classification models to predict the bin each observation belongs to. After classification, we applied secondary regression models to predict the exact value within each bin. This two-step approach, classification followed by regression, improved predictive accuracy and robustness (Hastie et al., 2009).

We selected thresholds that divided the data into more equitable bins to ensure balanced sample distribution across bins. For LWP, the threshold was -0.7 MPa; for RWC, it was 80%, effectively separating the data into high and low regions for each target and with a relative balance of the number of samples for training (Figure 5).

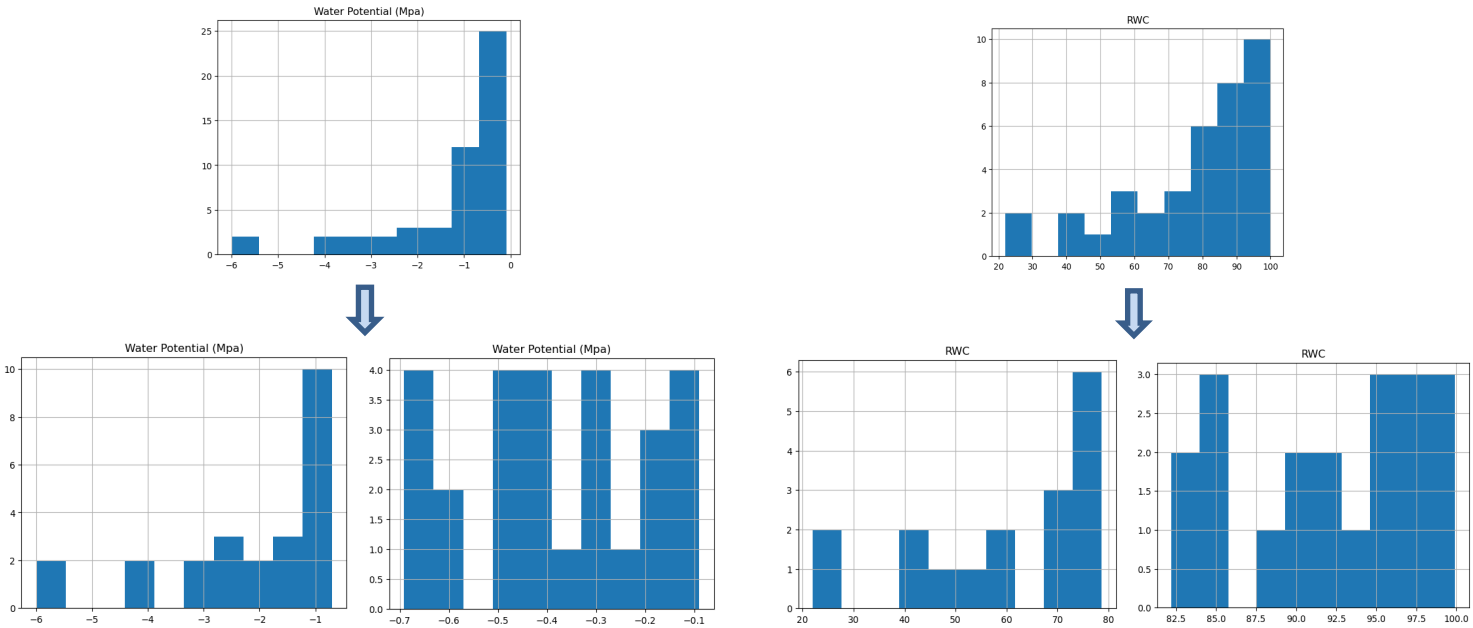

Figure 5: Binning method for the imbalanced continuous targets (a) LWP with threshold -0.7 MPa (b) RWC with threshold 80%

Feature Selection

Hyperspectral sensors capture data across numerous narrow spectral bands with minimal separation, leading to high collinearity. This complicates statistical and machine-learning applications, making model estimates unreliable and unstable (Hastie et al., 2009). Additionally, the large number of predictor variables (wavelength bands) can cause overfitting, capturing noise instead of the true signal and reducing model generalizability (Guyon & Elisseeff, 2003). Effective feature selection strategies are crucial to reduce dimensionality while preserving informative features for model training (Fodor, 2002).

To address these challenges, we implemented KBest Feature Selection, a type of filter method for feature selection. This technique ranks features based on their statistical significance in relation to the target variable. We employed both linear methods (such as the ANOVA F-test) and non-linear methods (such as mutual information) to capture a wide spectrum of relevant features. By selecting the top k features, we aimed to retain those with the highest predictive power while eliminating redundant or irrelevant ones (Chandrashekar & Sahin, 2014; Saeys et al., 2007). We generated distinct subsets of features that were subsequently used for model training for each feature selection method.

Feature Engineering

Feature engineering is critical for enhancing the performance of machine learning models, particularly with complex datasets like hyperspectral data. This study focused on identifying the optimal wavelength bands for several vegetation indices frequently used to assess chlorophyll and water content. These indices include Single Difference (SD), Single Ratio (SR), Normalized Difference (ND), and the Modified Datt Index (MDATT).

To find the optimal wavelength bands, we analyzed their relationship with LWP and RWC target variables using Pearson's Correlation Coefficient. This helped us determine which combinations of wavelength bands had the highest correlation with LWP and RWC. Using NumPy, we quantitatively identified these optimal bands, ensuring our models utilized the most informative features. The results in Table 1 highlight the optimal combination of bands that exhibited the strongest correlations.

Table 1: Computed optimal bands for spectral indices

Spectral Index	Formula	Computed Optimal Bands (nm)
LWP Data		
SD	$R_{\lambda1} - R_{\lambda2}$	R_{2188}, R_{2245}
SR	$R_{\lambda1} / R_{\lambda2}$	R_{1756}, R_{1749}
ND	$\lvert R_{\lambda1} - R_{\lambda2} \rvert / (R_{\lambda1} + R_{\lambda2})$	R_{1749}, R_{1756}
MDATT	$(R_{\lambda3} - R_{\lambda1}) / (R_{\lambda3} - R_{\lambda2})$	R_{1428}, R_{1848}, R_{1852}
RWC Data		
SD	$R_{\lambda1} - R_{\lambda2}$	R_{1938}, R_{1941}
SR	$R_{\lambda1} / R_{\lambda2}$	R_{2318}, R_{2334}
ND	$\lvert R_{\lambda1} - R_{\lambda2} \rvert / (R_{\lambda1} + R_{\lambda2})$	R_{354}, R_{1894}
MDATT	$(R_{\lambda3} - R_{\lambda1}) / (R_{\lambda3} - R_{\lambda2})$	R_{837}, R_{842}, R_{892}

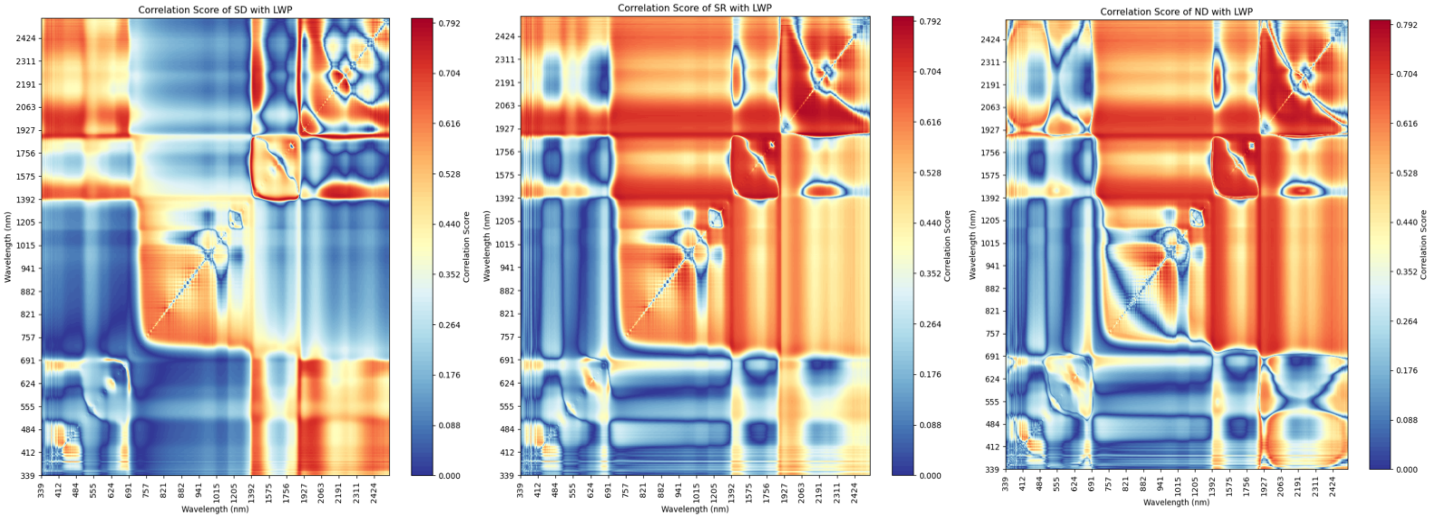

Figure 6: Contour map of correlation score between (a) SD (b) SR (c) ND indices and LWP

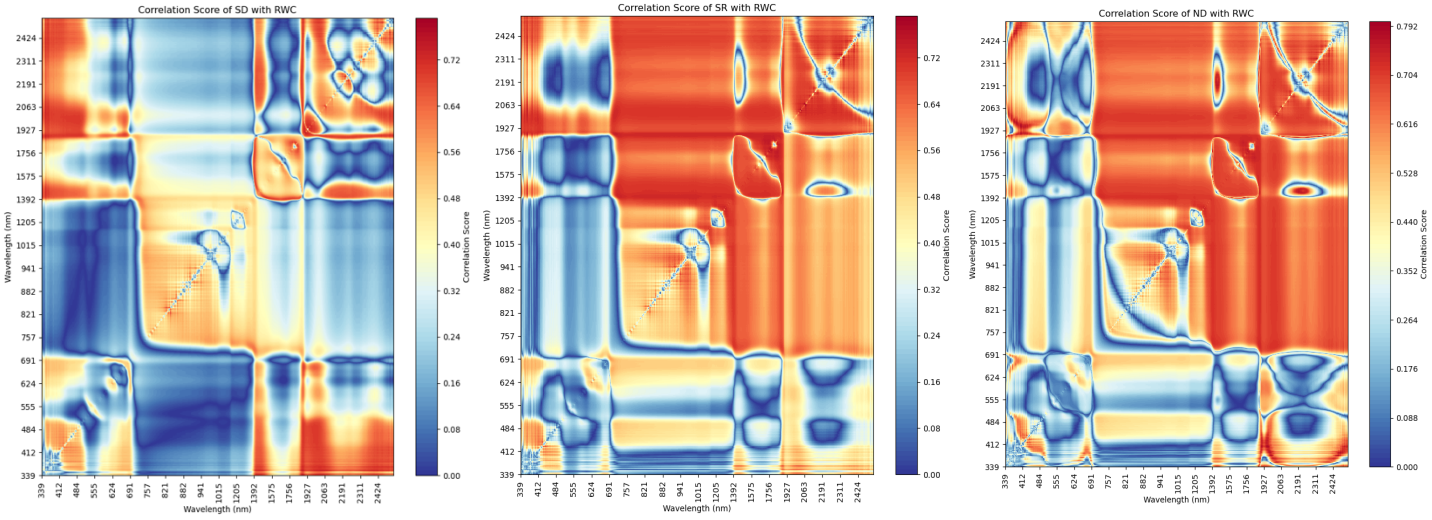

Figure 7: Contour map of correlation score between (a) SD (b) SR (c) ND indices and RWC

Previous studies have made significant strides in developing spectral indices tailored to various aspects of plant health. Building upon these advancements, our research focuses on enhancing these indices by integrating them with optimized vegetation indices (SD, SR, ND, and MDATT). This integration aims to create a unique dataset that complements the training process. Common spectral indices for water content estimation include the Disease Water Stress Index (DWSI), Moisture Stress Index (MSI), Leaf Water Vegetation Index 1 & 2 (LWVI), Normalized Difference Infrared Index (NDII), Normalized Difference Water Index (NDWI), Water Band Index (WBI), and Normalized Multi-band Drought Index (NMDI) (Table 2).

Table 2: Other spectral indices used in this study

Spectral Index	Formula	References
Disease Water Stress Index (DWSI)	$(R_{802} - R_{547})/(R_{1657}+R_{682})$	Galvao et al., (2005)
Moisture Stress Index (MSI)	R_{1599}/R_{819}	Ceccato et al., (2001)
Leaf Water Vegetation Index 1 (LWVI1)	$(R_{1094} - R_{983})/(R_{1094}+R_{983})$	Galvao et al., (2005)
Leaf Water Vegetation Index 2 (LWVI2)	$(R_{1094} - R_{1205})/(R_{1094}+R_{1205})$	Galvao et al., (2005)
Normalized Difference Infrared Index (NDII)	$(R_{819} - R_{1649})/(R_{819}+R_{1649})$	Hardisky et al.,(1983)
Normalized Difference Water Index (NDWI)	$(R_{857} - R_{1241})/(R_{857}+R_{1241})$	Gao (1995)
Water Band Index (WBI)	R_{970}/R_{900}	Penuelas et al., (1993)
Normalized Multi-band Drought Index (NMDI)	$(R_{860} - (R_{1640} - R_{2130}))/(R_{860}+(R_{1640} - R_{2130}))$	Jackson et al., (2004); Lu et al., (2018)

Model Selection and Training

Our objective was to develop a robust series of base learners by leveraging various machine learning models on diverse hyperspectral datasets. We delved into both linear and non-linear regression relationships between hyperspectral bands and target variables, LWP and RWC. We incorporated model-based dimensionality reduction and regularization techniques to enhance model performance and handle high-dimensional data challenges. Furthermore, ensemble models were employed to improve predictive accuracy and robustness.

Due to the limited sample size, we conducted a thorough 90/10 train-test split on each dataset to ensure rigorous evaluation of the models. We utilized the "Optuna" package for efficient parameter tuning across all models, streamlining the optimization process. Subsequently, a comprehensive analysis of model performance was conducted, highlighting the top-performing models across all relevant datasets. Detailed results are presented in Tables 3 and 4, showcasing the optimal models for each dataset.

Drawing from insights from prior studies on hyperspectral sensing data analysis of plants, Kernel Ridge Regression (KRR) and Partial Least Squares Regression (PLSR) are identified as promising candidate methods. These methods have demonstrated efficacy in capturing complex relationships between hyperspectral data and plant physiological parameters (Mohd et al., 2022; Ge et al., 2016; Yeh et al., 2016; Weber et al., 2012; Vigneau et al., 2011; Mo et al., 2015; Rapaport et al., 2015; Nguyen and Lee, 2006).

Results and Discussion

In evaluating prediction metrics of regression models, it became evident that RWC data outperformed LWP data, indicating that spectral bands present a more effective way of predicting RWC than LWP. Among the models assessed, Kernel Ridge Regression (KRR), XGBoost, and Gradient Boosting emerged as optimal choices for both the complete datasets of LWP and RWC, featuring original continuous targets. These models consistently achieved coefficient of determination (R-squared) values ranging from 82% to 95%, highlighting their robust predictive capabilities. Concurrently, the normalized root mean square error (NRMSE) values, falling within

the range of 7% to 12%, underscored the accuracy and reliability of these models.

With the full range of LWP (-7 to 0 MPa), the subset of non-linear selected features yielded the highest R-squared result (89.6%) and lowest NRMSE (9.47%) on the Gradient Boosting model. The top three bands for prediction are 355nm,1894nm and1898nm.

With the full range of RWC (17 to 100), the vegetation indices dataset has the highest prediction performance with the KRR model, yielding an R-squared of 94.9% and NRMSE of 7.37%. The top vegetation indices predictors are LWVI1, ND, and MSI. The original dataset also performed very well, with 92.2% R-squared and 9.17% NRMSE. The top 3 corresponding predictors are bands 373-342-357 nm.

Upon closer examination, particularly within the upper region – class 1 (beyond -0.7 MPa of LWP and beyond 80% RWC) of both target variables, the results exhibited remarkable performance, boasting R-squared values surpassing 95% and maintaining NRMSE values below 3% across various models such as PLSR, KRR, Elastic Net, and XGBoost. However, contrasting patterns emerged within the lower region – class 0 of LWP (under -0.7 MPa), where all models yielded poor performance, consistently registering R-squared values below 50%. Consequently, these outcomes were excluded from the tabulated results due to their limited predictive utility.

Conversely, the lower region of RWC (under 80%) showcased more promising outcomes, with the XGBoost model achieving a commendable 68% R-squared value. This disparity in performance between the lower regions of LWP and RWC indicates the nuanced relationships between spectral indices and plant physiological parameters, highlighting the importance of targeted model development and refinement to address specific areas of interest and variability within plant health assessment.

Table 3: Regression models with high results

Dataset	Regression Model	R-squared	RMSE	NRMSE (%)	Top 3 band predictors (nm) in the desc. order
LWP Full Data					
Indices_df	XGBoost	0.821	0.846	12.44%	SR SD WBI
nonlinear_selected_features	Gradient Boosting	0.896	0.644	9.47%	R_{1894} R_{355} R_{1898}
LWP Class 1 (> -0.7 MPa)					
indices_df_1	PLSR	0.973	0.017	2.42%	ND SD SR
indices_df_1	Elastic Net	0.955	0.022	3.13%	SR ND NDII
indices_df_1	KRR	0.969	0.018	2.60%	ND WBI NMDI
indices_df_1	XGBoost	0.989	0.011	1.55%	SD SR LWVI1
RWC Full Data					
Original_df	KRR	0.922	7.616	9.17%	R_{373} R_{342} R_{357}
Indices_df	KRR	0.949	6.125	7.37%	LWVI1 ND MSI

linear_selected_features	KRR	0.878	9.515	11.45%	R_{1894} R_{1891} R_{1887}
nonlinear_selected_features	KRR	0.887	9.173	11.04%	the subset has only 2 bands: R_{409} R_{2507}
RWC Class 1 (>80 RWC)					
Original_df_1	PLSR	0.959	0.817	1.41%	R_{341} R_{2516} R_{357}
Original_df_1	Elastic Net	0.999	0.092	0.16%	R_{2516} R_{418} R_{2484}
indices_df_1	PLSR	0.950	0.899	1.55%	SD ND SR
indices_df_1	Elastic Net	0.930	1.068	1.84%	SD ND SR
indices_df_1	KRR	0.974	0.648	1.12%	LWVI2 NDWI MSI
indices_df_1	Gradient Boosting	0.817	1.723	2.97%	SR SD ND
indices_df_1	XGBoost	0.999	0.110	0.19%	SR SD ND
RWC Class 0 (<80 RWC)					
indices_df_0	XGBoost	0.680	2.487	12.44%	MSI ND SD

Additionally, the Random Forest classification models for binary target variables demonstrate promising performance, achieving accuracy scores of 75% for RWC and 83.3% for LWP binary data, as detailed in Table 4. These results highlight the potential to refine classification predictions of water stress levels using appropriate thresholds. There is an opportunity to explore further specialized regression models tailored to specific regions of water availability, enhancing the precision of water stress assessments.

Table 4: Classification model results with binary target variables

Dataset	Classification Model	Accuracy	Top 3 band predictors (nm) in the desc. order
LWP Binary Data (0, 1)	RF Classifier	0.833	R_{2457} R_{597} R_{613}
RWC Binary Data (0, 1)	RF Classifier	0.75	R_{2499} R_{358} R_{586}

Conclusion

Our study demonstrated the efficacy of hyperspectral sensing in assessing water stress through spectral reflectance analysis. By leveraging various machine learning regression models trained

on a comprehensive set of LWP and RWC targets, we achieved robust predictions for water availability and a clear comparison of these two water status variables in prediction models. Models such as KRR, XGBoost, Gradient Boosting, PLSR, and Elastic Net consistently exhibited high R-squared values, indicating their strong predictive capabilities for water stress levels.

This exemplary performance across diverse regression models underscores the effectiveness of spectral bands in capturing plant physiological traits, particularly in regions associated with optimal plant health. Our approach addressed challenges like imbalanced data and high-dimensional feature spaces through binning, feature selection, and feature engineering techniques. By transforming continuous target variables into discrete regions, we explored the potential of combining classification and regression for water stress prediction, suggesting further research on applying specific regression models to distinct ranges of water availability indices. Additionally, we enhanced model performance and generalizability by selecting informative spectral features and leveraging both non-optimized vegetation indices from previous studies and optimized spectral indices from this study.

Our findings highlight the potential of integrating hyperspectral sensing with machine learning for precise water stress monitoring in agriculture, especially for challenging crops like wild blueberries. Further refinement and validation of our models can facilitate practical implementation in real-world agricultural settings, fostering sustainable water management practices and enhancing crop security. As potential future work, incorporating the neural vegetation index and up-scaling the models for UAV-based hyperspectral data could significantly enhance our monitoring capabilities. This would allow for more precise and scalable water stress assessments across larger agricultural fields, enabling real-time decision-making and further advancing the practical applications of our research.

Acknowledgment

We would like to thank Wyman's for providing plant materials. We would also like to acknowledge Abby Novak for helping with the data collection. This project was supported by the USDA National Institute of Food and Agriculture, Hatch Project Number ME0-22021, through the Maine Agricultural and Forest Experiment Station. This research was also supported by the Wild Blueberry Commission of Maine, the Maine Department of Agriculture, Conservation and Forestry (SCBGP; 20200918*0996), the NASA Maine Space Grant Consortium grant (SG-24-06), and the UMaine Faculty Summer Research Award.

References

Aladenola, O., & Madramootoo, C. (2014). Response of greenhouse-grown bell pepper (Capsicum annuum L.) to variable irrigation. Canadian Journal of Plant Science, 94(2), 303–310. https://doi.org/10.4141/cjps2013-048

Ceccato, P., Flasse, S., Tarantola, S., Jacquemoud, S., & Grégoire, J.-M. (2001). Detecting vegetation leaf water content using reflectance in the optical domain. Remote Sensing of Environment, 77, 22–33. https://doi.org/10.1016/S0034-4257(01)00191-2

Champagne, C., Staenz, K., Bannari, A., & McNairn, H. (2001). Mapping crop water status: Issues of scale in the detection of crop water stress using hyperspectral indices. In Proceedings of the 8th International Symposium on Physical Measurements and Signatures in Remote Sensing, Aussois, France (pp. 79–84).

Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16-28.

Fawcett, T., & Provost, F. (1997). Adaptive fraud detection. Data Mining and Knowledge Discovery, 1(3), 291-316.

Fodor, I. K. (2002). A survey of dimension reduction techniques. Center for Applied Scientific Computing, Lawrence Livermore National Laboratory.

Galvao, L. S., dos Santos, J. R., Roberts, D. A., de Moura, Y. M., & Soares, J. V. (2005). Discrimination of sugarcane varieties in Southeastern Brazil with EO-1 Hyperion data. Remote Sensing of Environment, 94, 523–534. https://doi.org/10.1016/j.rse.2004.11.010

Gao, B. (1995). Normalized difference water index for remote sensing of vegetation liquid water from space. In Proceedings of SPIE 2480 (pp. 225–236). https://doi.org/10.1117/12.210877

Ge, Y., et al. (2016). Using a hybrid genetic algorithm for parameter estimation in spectral mixture analysis. Remote Sensing Letters, 7(5), 499-508.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. Journal of Machine Learning Research, 3(Mar), 1157-1182.

Hardisky, M., Klemas, V., & Smart, R. (1983). The influences of soil salinity, growth form, and leaf moisture on the spectral reflectance of Spartina alterniflora canopies. Photogrammetric Engineering and Remote Sensing, 49, 77–83.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: data mining, inference, and prediction. Springer Science & Business Media.

Hunt Jr., E., & Rock, B. (1989). Detection of changes in leaf water content using near- and middle-infrared reflectances. Remote Sensing of Environment, 30, 43–54. https://doi.org/10.1016/0034-4257(89)90046-1

Ihuoma, S. O., & Madramootoo, C. A. (2017). Recent advances in crop water stress detection. Computers and Electronics in Agriculture, 141, 267–275. https://doi.org/10.1016/j.compag.2017.07.026

Jackson, T., Chen, D., Cosh, M., Li, F., Anderson, M., Walthall, C., ... & Doraiswamy, P. (2004). Vegetation water content mapping using Landsat data derived normalized difference water index for corn and soybeans. Remote Sensing of Environment, 92, 475–482. https://doi.org/10.1016/j.rse.2003.10.021

Jensen, J. R. (2009). Remote sensing of the environment: An earth resource perspective 2/e. Pearson Education India.

Jones, H. G., & Vaughan, R. A. (2010). Remote sensing of vegetation: Principles, techniques, and applications. Oxford University Press, USA.

Lu, S., Zhou, D., & Zhou, J. (2018). A robust vegetation index for remotely assessing chlorophyll content of dorsiventral leaves across several species in different seasons. Plant Methods, 14, 15. https://doi.org/10.1186/s13007-018-0281-z

Maimaitiyiming, M., Ghulam, A., Bozzolo, A., Wilkins, J. L., & Kwasniewski, M. T. (2017). Early detection of plant physiological responses to different levels of water stress using reflectance spectroscopy. Remote Sensing, 9, 745. https://doi.org/10.3390/rs9070745

Mohd Asaari, M. S., Mishra, P., Mertens, S., Dhondt, S., Inzé, D., Wuyts, N., ... & Aasen, H. (2022). Non-destructive analysis of plant physiological traits using hyperspectral imaging: A case study on drought stress. Computers and Electronics in Agriculture, 195, 106806. https://doi.org/10.1016/j.compag.2022.106806

Mo, S., et al. (2015). Estimating leaf water content of rice using MODIS EVI data. Remote Sensing, 7(8), 9794-9815.

Mulla, D. J. (2013). Twenty five years of remote sensing in precision agriculture: Key advances and remaining knowledge gaps. Biosystems Engineering, 114(4), Article 4.

Nguyen, H. T., & Lee, B. W. (2006). A survey of spectral unmixing algorithms. EURASIP Journal on Advances in Signal Processing, 2006(1), 075291.

Penuelas, J., Filella, I., Biel, C., Serrano, L., & Save, R. (1993). The reflectance at the 950–970 nm region as an indicator of plant water status. International Journal of Remote Sensing, 14, 1887–1905. https://doi.org/10.1080/01431169308954010

Rapaport, L., et al. (2015). Using spectral mixture analysis for soil organic carbon estimation. European Journal of Soil Science, 66(5), 842-853.

Rossini, M., Fava, F., Cogliati, S., Meroni, M., Marchesi, A., Panigada, C., Giardino, C., Busetto, L., Migliavacca, M., Amaducci, S., & Colombo, R. (2013). Assessing canopy PRI from airborne imagery to map water stress in maize. ISPRS Journal of Photogrammetry and Remote Sensing, 86, 168–177. https://doi.org/10.1016/j.isprsjprs.2013.10.002

Saeys, Y., Inza, I., & Larrañaga, P. (2007). A review of feature selection techniques in bioinformatics. Bioinformatics, 23(19), 2507-2517.

Taiz, L., Zeiger, E., Møller, I. M., & Murphy, A. (2015). Plant physiology and development. https://www.cabidigitallibrary.org/doi/full/10.5555/20173165866

Vigneau, N., et al. (2011). Vegetation indices derived from Sentinel-2 for precision agriculture applications. Proceedings of SPIE, 7824, 78240F.

Wang, L., & Qu, J. (2007). NMDI: A normalized multi-band drought index for monitoring soil and vegetation moisture with satellite remote sensing. Geophysical Research Letters, 34, L20405. https://doi.org/10.1029/2007GL031021

Wang, L., & Qu, J. (2008). Forest fire detection using the normalized multi-band drought index (NMDI) with satellite measurements. Agricultural and Forest Meteorology, 148(11), 1767–1776. https://doi.org/10.1016/j.agrformet.2008.06.001

Weber, V., et al. (2012). Estimating biophysical parameters of barley using spectral vegetation indices and partial least squares regression. ISPRS Journal of Photogrammetry and Remote Sensing, 69, 97-111.

Yeh, M. L., et al. (2016). Leaf water content estimation using hyperspectral remote sensing. International Journal of Applied Earth Observation and Geoinformation, 52, 219-229.