In-season Nitrogen Prediction and Evaluation using Airborne Imagery and AI Techniques in Commercial Potato Production

Javed, B.[1,2], Cambouris, A.N.[1], Longchamps, L.[3], Basran, P.S.[4], Duchemin, M.[1], Ziadi, N.[1], Arnold, S.[5], Fenech, A.[5], Karam, A.[2]

[1]Quebec Research and Development Centre, Agriculture and Agri-Food Canada, 2560 Hochelaga Blvd., Quebec City, Quebec, Canada G1V 2J3

[2]Dept. of Soils and Agri-Food Engineering, Paul-Comtois Bldg., Université Laval, Quebec City, Quebec, Canada G1V 0A6

[3]School of Integrative Plant Science, Soil and Crop Sciences Section, Cornell University, 616 Thurston Ave., Ithaca, NY, USA 14853

[4]College of Veterinary Medicine, Cornell University, 930 Campus Road, Box 31, Ithaca, NY, USA 14853

[5]School of Climate Change and Adaptation, 550 University Ave., Charlottetown, Prince Edward Island, Canada C1A 4P3

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

Abstract

*In modern agriculture, timely and precise nitrogen (N) monitoring is essential to optimize resource management and improve trade benefits. Potato (*Solanum tuberosum *L.) is a staple food in many regions of the world, and improving its production is essential to ensure food security and promote related industries. Traditional methods of assessing nitrogen are labor intensive and time consuming, and require subjective observations. To address these limitations, combining the use of multispectral data derived from drone imagery with the use of artificial intelligence models could be a better approach to the non-invasive and high-resolution monitoring of potato fields throughout the growing season. To achieve this objective, a study was conducted at four different commercial potato fields on Prince Edward Island, Canada. Images were taken from an unmanned aerial vehicle (UAV) of plants at the early flowering stage. Multispectral bands were extracted and vegetative indices (VIs) calculated, and were then used to predict petiole nitrate (*NO_3^-*-N) concentrations at the early flowering stage. Various machine learning (ML) algorithms were used in these predictions, including random forest (RF), bagged trees (BT), gradient boosting machines (GBM), support vector machines (SVM) and Bayesian artificial neural network (BANN) algorithms. A prediction model was trained with 75% of the data set and evaluated using the remaining 25% of the data set. The results revealed that the BT model trained on UAV images (relative root mean square error [RRMSE] = 12.7% and relative mean absolute error [RMAE] = 9.6%) performed well in predicting petiole nitrate concentrations. These results suggest that multispectral data has potential as input data for ML algorithms to predict in-season N. This method shows promise in the digital agriculture and smart farming sectors, with the primary objectives of mitigating excessive N applications and optimizing potato production to its full potential.*

The authors are solely responsible for the content of this paper, which is not a refereed publication. Citation of this work should state that it is from the Proceedings of the 16th International Conference on Precision Agriculture. EXAMPLE: Last Name, A. B. & Coauthor, C. D. (2024). Title of paper. In Proceedings of the 16th International Conference on Precision Agriculture (unpaginated, online). Monticello, IL: International Society of Precision Agriculture.

Keywords.
Precision agriculture, remote sensing, artificial intelligence, digital agronomy, feature engineering

Introduction

Nitrogen (N) management in potato (*Solanum tuberosum* L.) plays an important role in promoting physiological growth and tuber yield, but optimizing N rates is quite challenging due to the plant's shallow root system (Muleta and Aga 2019). Another issue aside from the poor root structure is that the optimal N application varies greatly depending on the current growth conditions, which are themselves extremely variable across time and space, eventually leading to over- or under-fertilization with N (Berger et al. 2020). Excessive N in the soil may result in higher vegetative growth, poor skin set, delayed tuber maturation, and leaching of N into groundwater, while N deficiency during critical growth stages leads to premature leaf senescence, low starch content and lower tuber yield (Bohman et al. 2021; Zhou et al. 2022).

Plants' N requirements can be assessed during the critical growth stages by in-season sampling, which can be helpful in optimizing N applications (Grell et al. 2021). Traditional N assessment methods require manual tissue sampling and a laborious chemical analysis approach, which often requires significant time and resources (Liang et al. 2019). However, in recent years, the integration of modern machine learning with remote sensing technologies and its application in unmanned aerial vehicle (UAV) image processing have revolutionized the agriculture sector through the optimization of N management (Ennaji et al. 2023).

Numerous studies have suggested the usefulness of UAV imagery, and of machine learning (ML) to process it, in non-invasive in-season N monitoring (Alkhaled et al. 2023; Hunt et al. 2018). For instance, Zhou et al. (2022) successfully predicted petiole nitrate (NO_3^--N) concentrations in potato plots using hyperspectral images, with a reported root mean square error (RMSE) of 0.24%. Similarly, Peng et al. (2021) extracted data from UAV multispectral images to predict plant N uptake using a random forest algorithm, reporting RMSE values ranging between 7.1% and 22.0%. The goal of this study is to use ML algorithms trained and validated using high-resolution UAV data exclusively, in order to predict in-season N at commercial potato growing sites on Prince Edward Island (PEI), Canada.

Material and Methods

Site description

The experiments were conducted in 2020 and 2021 on potato producing commercial sites located in pedoclimatic conditions of PEI, Canada (Fig 1). The four different experimental sites consisted of two sites at Oyster Cove Farms, OC1 (8.6 ha) and OC2 (15 ha), and two sites at Black Pond Farms, BP1 (8.8 ha) and BP2 (8.1 ha) (Fig 1). Soil series for four sites was Charlottetown, having sandy loam texture with a pH ranging from 5.3 to 6.7 (average 6.0). All the sites were on a three year crop rotation strategy and identified as Orthic Humo-Ferric Podzol soil classification (Soil Classification Working Group, 1998). All the sites were on a three-year crop rotation. The experiments were laid out in a generalized randomized design, with three management zones (MZ) and site-specific N treatments.

Fig 1. Experimental sites, Oyster Cove (OC1 & OC2) and Black Pond (BP1 & BP2) (Source : Google Earth®).

Table 1. Experimental layout based on generalized randomized design, and treatments implemented at the Oyster Cove (OC1 and OC2) and Black Pond (BP1 and BP2) sites.

Site	MZ	Block	Treatment	Sampling points
OC1	3	4	Uniform VRA	35
OC2	3	4	Uniform VRA	36
BP1	3	3	Three specific N rates	27
BP2	3	3	Three specific N rates	27

Note: MZ = management zone, Uniform = uniform nitrogen rate application: 168 kg ha^{-1} for OC1 and OC2. VRA = variable nitrogen rate application: 22, 45, 67 kg ha^{-1} for OC1 and 34, 56, 78 kg ha^{-1} for OC2. Three specific nitrogen rates: 22, 50, 78 kg ha^{-1} for BP1 and 45, 73, 101 kg ha^{-1} for BP2.

Ground truth sampling

Sampling points representative of the overall condition of the experimental sites were selected and georeferenced (Table 1). These points were chosen to collect ground truth data on petiole nitrate concentrations through petiole sampling (target stage), which was performed 62 days after planting (DAP) to monitor in-season plant NO_3^--N concentrations. For petiole sampling, the fourth youngest, fully expanded leaf from the top of the plant was sampled from 20 plants at all georeferenced sampling points. The tissue samples were oven-dried at 60°C, ground, and sieved to 2 mm, extracted and analyzed using a continuous flow injection analyzer. N concentrations were calculated from petiole nitrate concentrations as a percentage (NO_3^--N) (Bohman et al. 2019). The data sets from all four sites were subsequently merged (35 + 36 + 27 + 27) (Table 1) into larger data sets, with 125 points provided for ground truthing.

Multispectral image acquisition and preprocessing

Remote sensing (RS) hyperspectral imagery data were recorded at the selected target stage (petiole sampling, 62 DAP) using a UAV (Matrice 210 RTK V2 UAV; DJI, Shenzhen, China) equipped with a MicaSense Altum multi-spectral camera (MicaSense Inc., Seattle, WA, USA). The UAV collected imagery in five spectral bands, including blue (475 nm and 32 nm bandwidths), green (560 nm and 27 nm bandwidths), red (668 nm and 14 nm bandwidths), rededge (717 nm and 12 nm bandwidths) and near-infrared (NIR) (842 nm and 57 nm bandwidths). The hyperspectral images were processed using Pix4Dmapper software (Pix4D S.A., Prilly, Switzerland, version 4.2.27), and the final mosaic images were resampled to a 5-cm spatial resolution. At the georeferenced sampling points, mean reflectance values for each spectral band were extracted using the simple feature (sf) package in R, for later use in the calculation of vegetative indices (VIs) (ver. 4.3.1; R Core Team, 2023). In addition, ten different VIs—consisting of the optimized soil adjusted index (OSAVI), green and red ratio vegetation index (GRVI), normalized difference red-edge index (NDRE), modified soil-adjusted vegetation index (MSAVI), canopy chlorophyll content index (CCCI), chlorophyll vegetation index (CVI), chlorophyll index (CIgreen), chlorophyll index (CIrededge), green normalized vegetation index (GNDVI) and transformed chlorophyll absorption reflectance index (TCARI)—were chosen and calculated based on an extensive review of the specific literature on potato production (Goffart et al. 2023; Peng et al. 2021; Sun et al. 2022; Yang et al. 2021).

Feature optimization and model development

Since the VIs were calculated from data extracted from the five multispectral bands, multicollinearity among predictors may occur that can lead to unnecessary noise in ML algorithms, causing an increase in standard errors, unstable coefficients and difficulty in selecting important features (Kuhn and Johnson 2013). Highly correlated predictors were eliminated using the findCorrelation function in R, followed by the Boruta feature selection algorithm in the R Boruta package, in order to identify the most relevant features associated with the target stage (petiole nitrate concentration). Following a careful feature selection, in order to obtain robust and unbiased ML models, the data set was randomly partitioned into training data (75%) and model validation data (25%). For model validation, a five-fold cross validation technique was employed using the caret package in R studio (ver. 4.3.1; R Core Team, 2023). Five different ML models, random forest (RF), bagged trees (BT), gradient boosting machines (GBM), support vector machines (SVM) and Bayesian artificial neural network (BANN), were employed for the prediction process. Each ML model underwent an intensive iterative process in order to achieve hyperparameter optimization, maximizing the model’s ability to detect complex pattens in the data set.

Model performance evaluation

The final predictive model was evaluated using a validation test set. To screen the goodness of fit for all models, several metrics were computed, including the root mean square error (RMSE), relative root mean square error (RRMSE [%]), mean absolute error (MAE) and relative mean absolute error (RMAE [%]). The selection of the best performing ML model was based on the minimal error percentage closest to zero (Chen et al. 2011).

$$\mathrm{RMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n}} \quad (1)$$

$$\mathrm{RRMSE} = \sqrt{\frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{n\hat{y}_i}} \quad (2)$$

$$\mathrm{MAE} = \frac{1}{n}\sum_{i=1}^{n} \left(|y_i - \hat{y}_i|\right) \quad (3)$$

$$\mathrm{RMAE} = \frac{1}{n}\sum_{i=1}^{n} \left(\left|\frac{y_i - \hat{y}_i}{\hat{y}_i}\right|\right) \quad (4)$$

Result and discussion

Relationship among predictors and target stage

Understanding the complex associations between multispectral bands, VIs and biochemical traits such as petiole NO_3^--N concentrations is essential for understanding overall ML model performance and potato crop health (Ahmad and Sharma 2023; Peng et al. 2021). A moderate but nevertheless significant correlation was found between predictors (multispectral bands and VIs) and petiole NO_3^--N concentrations (Fig 2). Among the predictors, NDRE showed the highest correlation ($r = -0.42$, $p = 0.001$) with petiole NO_3^--N concentrations. Conversely, TCARI showed the lowest correlation ($r = -0.19$, $p = < 0.001$) (Fig 2). These findings are in agreement with those obtained by Zhou et al. (2022), who reported a significant relationship between rededge or rededge derived VIs and petiole NO_3^--N concentrations. This significant relationship could be due to the sensitivity of rededge region wavelengths to leaf chlorophyll content, which is eventually related to the plant's nitrogen status (Morier et al. 2015; Wu et al. 2007).

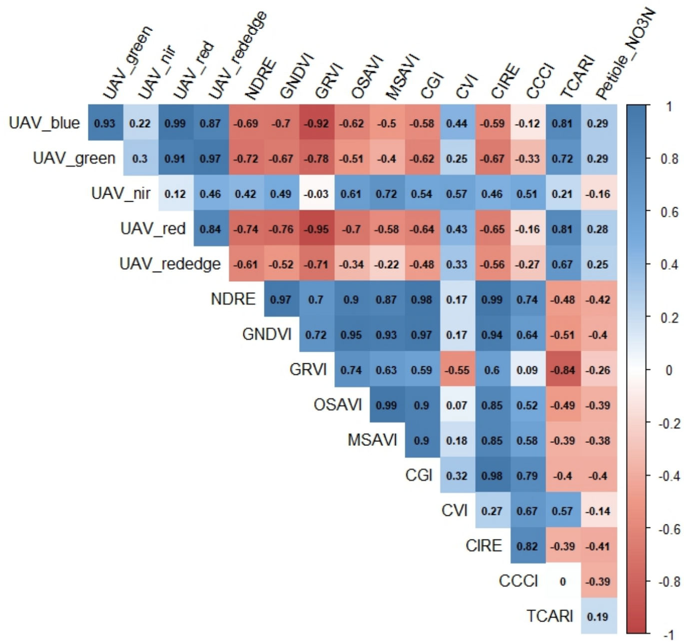

Fig 2. Correlation matrix between multispectral bands, vegetation indices and petiole NO3$^-$-N concentrations at 62 DAP.

Feature selection for model building

The hybrid feature selection approach (correlation threshold for predictors and Boruta algorithm) proved useful in identifying the most relevant features influencing the predictive power of ML models among the 15 different bands and VIs. Using this approach, the CCCI was found to be most important feature in predicting petiole nitrate concentrations (Fig 3). These findings corroborate those by Liu et al. (2021), who found that rededge region wavelengths and VIs derived from rededge wavelengths are associated with petiole nitrate concentrations.

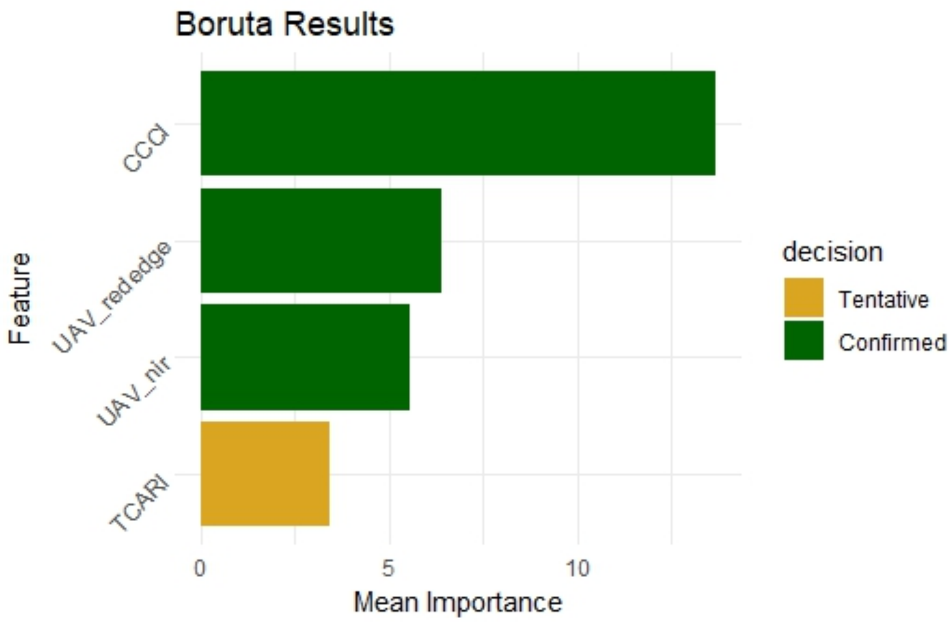

Fig 3. Ranking of feature importance based on the Boruta feature selection algorithm.

Machine learning model performance and validation

Using the selected input variables, five different ML predictive models were built to predict petiole nitrate concentrations in order to compare their performance: RF, BT, GBM, SVM and BANN. All models were trained using k-fold cross validation (k = 10) and validated using the test data set. For validation purposes, scatter plots for observed and predicted petiole nitrate concentration values were fitted using a bootstrapping technique to reveal complex patterns in the data set (Fig 4).

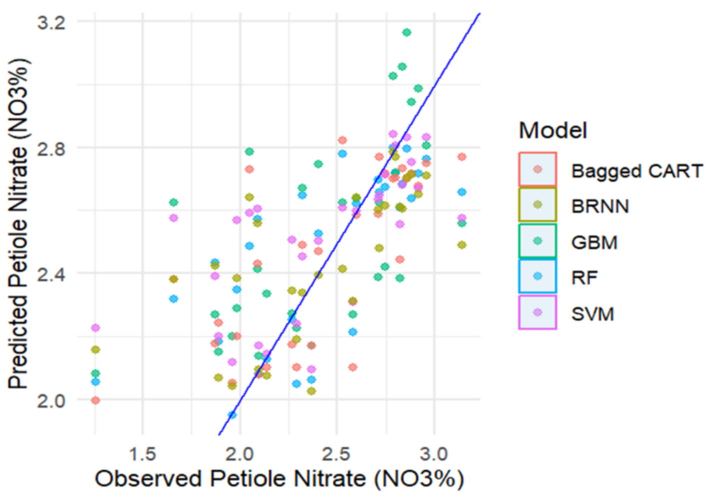

Fig 4. Validation of the machine learning models by comparing observed and predicted petiole nitrate concentrations (NO3-N). The five different colors of the points represent the different machine learning models.

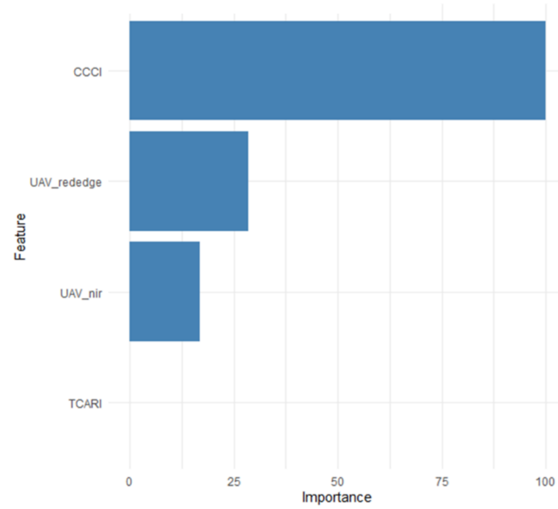

Fig 5. Importance of different bands and vegetation indices according to machine learning algorithm in explaining petiole nitrate concentration.

Overall, all ML models exhibited RMSE values ranging between 0.31 and 0.36 (NO_3^--N) and RRMSE values between 12.7% and 15.4%, while the MAE for all ML models ranged between 0.23 and 0.28 (NO_3^--N), with a RMAE below 12%. In particular, the BT model had the best performance, with an RMSE of 0.30 (NO_3^--N) (12.7% RRMSE) and a MAE of 0.23 (NO_3^--N) (9.6% RMAE) (Table 2). The most relevant feature in the BT model was CCCI, followed by rededge, NIR and TCARI (Fig 5). The study results are in agreement with those obtained by Zhou et al. (2022), who predicted petiole NO_3^--N concentrations using three different remote-sensing platforms (DJI phantom 4, Parrot Sequoia+ and MicaSense RedEdge MX), and reported RMSEs ranging between 0.13 and 4.6 (NO_3^--N). Moreover, Zhang et al. (2022) reported that rededge spectra and rededge derived VIs such as the canopy chlorophyll content index (CCCI) are more sensitive to chlorophyll content and eventually to leaf nitrogen levels than broad wavebands consisting of the blue or red bands or a mixture of visible and NIR light.

Table 2. Model performance insights and statistical analysis indices of RMSE, MAE, RRMSE and RMAE for petiole nitrate concentration (NO_3^--N).

Model	RMSE	MAE	RRMSE	RMAE
RF	0.31	0.23	13.0	9.40
GBM	0.36	0.28	15.2	11.8
BT	0.30	0.23	12.7	9.60
SVM	0.35	0.25	14.7	10.3
BANN	0.33	0.24	15.4	10.9

Note: RMSE and MAE in (NO_3^--N) percentage, whereas RRMSE and RMAE are presented as error percentage.

Conclusion

The goal of this project was to utilize multispectral remote sensing data obtained with a UAV to train different ML models to predict in-season N via petiole nitrate concentrations. The fitted ML models proved to be robust and useful in predicting in-season N values and in identifying the most relevant features for determining in-season petiole nitrate concentrations. Moreover, this work also provided a complete methodology — from image processing, and feature extraction and selection to model development in R — which can be combined with more advanced ML techniques such as ensemble model to improve the accuracy of predictions in the future. The study results have improved and better informed the existing art of in-season N monitoring in potato cultivation. However, future attempts should incorporate increased sample sizes in combination with additional sources of information (agronomic, weather or soil properties) in order to develop highly robust and generalized ML models.

Acknowledgement

We would like to thank Andrée-Dominique Baillargeon, Camille Lambert-Beaudet, Jeff Daniel Steeven Nze Memiaghe, Audrey-Kim Minville, David Andres Ramirez-Gonzalez, Sylvie Côté (Cambouris' Precision Agriculture teams, AAFC Quebec), Ryan Barrett (PEI Potato Board), Jennifer MacLeod, the team of Scott Anderson, the team of Dr. Nyiraneza (AAFC Charlottetown), the Kensington North Watershed Association, East Prince Agri Environment Association, the Souris and Area Branch of the PEI Wildlife Federation, and Luke Meloche and Andy MacDonald (drone pilots from the UPEI Canadian Centre for Climate Change and Adaptation) for their field assistance and technical support.

Funding

This project was funded by the Living Laboratories Initiative of Agriculture and Agri-Food Canada (AAFC, project number J-002269).

Conflicts of interest

The authors declare no conflict of interest.

References

Ahmad, U., and Sharma, L. 2023. A review of Best Management Practices for potato crop using Precision Agricultural Technologies. Smart Agricultural Technology. **4**: 100220.

Alkhaled, A., Townsend, P.A., and Wang, Y. 2023. Remote Sensing for Monitoring Potato Nitrogen Status. American Journal of Potato Research. **100**: 1-14.

Berger, K., Verrelst, J., Féret, J.-B., Wang, Z., Wocher, M., Strathmann, M., Danner, M., Mauser, W., and Hank, T. 2020. Crop nitrogen monitoring: Recent progress and principal developments in the context of imaging spectroscopy missions. Remote Sensing of Environment. **242**: 111758.

Bohman, B.J., Rosen, C.J., and Mulla, D.J. 2019. Evaluation of Variable Rate Nitrogen and Reduced Irrigation Management for Potato Production. *Agronomy Journal*. **111**: 2005-2017.

Bohman, B.J., Rosen, C.J., and Mulla, D.J. 2021. Relating nitrogen use efficiency to nitrogen nutrition index for evaluation of agronomic and environmental outcomes in potato. Field Crops Research. **262**: 108041.

Canadian Agricultural Services Coordinating Committee. Soil Classification Working Group, National Research Council Canada, Canada. Agriculture and Agri-Food Canada. Research Branch, 1998. The Canadian system of soil classification (No. 1646). NRC Research Press.

Chen, J.-L., Liu, H.-B., Wu, W., and Xie, D.-T. 2011. Estimation of monthly solar radiation from measured temperatures using support vector machines – A case study. Renewable Energy. **36**: 413-420.

Ennaji, O., Vergütz, L., and El Allali, A. 2023. Machine learning in nutrient management: A review.

Artificial Intelligence in Agriculture. **9**: 1-11.

Goffart, J.P., Ben Abdallah, F., Goffart, D., Curnel, Y., and Planchon, V. 2023. Potato Crop Nitrogen Status Monitoring for Sustainable N Fertilisation Management: Last 15 Years and Future-Expected Developments with Reference Method and Use of Optical Sensors. Potato Research. **66**: 1257-1303.

Grell, M., Barandun, G., Asfour, T., Kasimatis, M., Collins, A.S.P., Wang, J., and Güder, F. 2021. Point-of-use sensors and machine learning enable low-cost determination of soil nitrogen. Nature Food. **2**: 981-989.

Hunt, E.R., Horneck, D.A., Spinelli, C.B., Turner, R.W., Bruce, A.E., Gadler, D.J., Brungardt, J.J., and Hamm, P.B. 2018. Monitoring nitrogen status of potatoes using small unmanned aerial vehicles. Precision Agriculture. **19**: 314-333.

Kuhn, M., and Johnson, K. 2013. An Introduction to Feature Selection. Pages 487-519 in M. Kuhn and K. Johnson, eds. Applied Predictive Modeling. Springer New York, New York, NY. Available: https://doi.org/10.1007/978-1-4614-6849-3_19.

Liang, K., Jiang, Y., Nyiraneza, J., Fuller, K., Murnaghan, D., and Meng, F.-R. 2019. Nitrogen dynamics and leaching potential under conventional and alternative potato rotations in Atlantic Canada. Field Crops Research. **242**: 107603.

Liu, N., Townsend, P.A., Naber, M.R., Bethke, P.C., Hills, W.B., and Wang, Y. 2021. Hyperspectral imagery to monitor crop nutrient status within and across growing seasons. Remote Sensing of Environment. **255**: 112303.

Morier, T., Cambouris, A.N., and Chokmani, K. 2015. In-season nitrogen status assessment and yield estimation using hyperspectral vegetation indices in a potato crop. Agronomy Journal. **107**: 1295-1309.

Muleta, H.D., and Aga, M.C. 2019. Role of nitrogen on potato production: a review. Journal of plant sciences. **7**: 36-42.

Peng, J., Manevski, K., Kørup, K., Larsen, R., and Andersen, M.N. 2021. Random forest regression results in accurate assessment of potato nitrogen status based on multispectral data from different platforms and the critical concentration approach. Field Crops Research. **268**: 108158.

Sun, C., Zhou, J., Ma, Y., Xu, Y., Pan, B., and Zhang, Z. 2022. A review of remote sensing for potato traits characterization in precision agriculture. *Frontiers in Plant Science*. **13**.

Wu, J., Wang, D., Rosen, C.J., and Bauer, M.E. 2007. Comparison of petiole nitrate concentrations, SPAD chlorophyll readings, and QuickBird satellite imagery in detecting nitrogen status of potato canopies. Field Crops Research. **101**: 96-103.

Yang, H., Li, F., Hu, Y., and Yu, K. 2021. Hyperspectral indices optimization algorithms for estimating canopy nitrogen concentration in potato (*Solanum tuberosum* L.). International Journal of Applied Earth Observation and Geoinformation. **102**: 102416.

Zhang, H., Li, J., Liu, Q., Lin, S., Huete, A., Liu, L., Croft, H., Clevers, J.G.P.W., Zeng, Y., Wang, X., Gu, C., Zhang, Z., Zhao, J., Dong, Y., Mumtaz, F., and Yu, W. 2022. A novel red-edge spectral index for retrieving the leaf chlorophyll content. Methods in Ecology and Evolution. **13**: 2771-2787.

Zhou, J., Wang, B., Fan, J., Ma, Y., Wang, Y., and Zhang, Z. 2022. A Systematic Study of Estimating Potato N Concentrations Using UAV-Based Hyper- and Multi-Spectral Imagery. Agronomy. **12**: 2533.