Humna Khan[1], Travis Esau[1], Aitazaz Farooque[2], Farhat Abbas[3]

1 Engineering Department, Dalhousie University, Faculty of Agriculture, Truro, Nova Scotia, Canada

2 Faculty of Sustainable Design Engineering, University of Prince Edward Island, Charlottetown, Prince Edward Island, Canada

3 College of Engineering Technology, University of Doha for Science and Technology, Doha, Qatar

Abstract.

The production of wild blueberries (Vaccinium angustifolium.) is contributing 112.2 million dollars yearly to the Canada's revenue which can be further increased through controlling harvest losses. A precise prediction of blueberry harvesting losses is necessary to mitigate such losses. In this study, the performance of two machine learning (ML) models was evaluated to predict the wild blueberry harvest losses on the ground. The data from two commercial fields namely Frank Webb and Tracadie in Atlantic Canada were used for this purpose. Wild blueberry losses (fruit loss on ground, leaf losses, blower losses) and yield were measured manually from randomly selected plots during mechanical harvesting. Wild blueberry plant height, fruit zone, and field slope readings were recorded from each of the plots. Two ML models namely linear regression (LR) and random forest (RF) were used to predict ground losses as a function of plant height, fruit zone, slope, fruit yield, leaf loss, and blower loss. Coefficient of determination (R^2) were used to assess the prediction accuracy of the models. Correlation analysis revealed that fruit yield and other losses (leaf loss, blower loss) had moderate to high correlations judged from the coefficient of correlation (r), i.e., r = 0.32- 0.78. Linear regression model performed best in both fields Frank Webb and Tracadie with R^2= 0.91 and 0.87 as compared to RF model with R^2= 0.53 and 0.78 respectively. The comparison of these algorithms suggested that the LR performed comparatively better for both fields. The results revealed that the LR model could be useful in the prediction of ground losses during the harvesting of wild blueberries in the selected fields.

Keywords.
Machine learning algorithms, harvesting losses, wild blueberries

Summary

Canada produced 161,346 tons of fresh wild blueberries in 2020 making its production more than 50% of the world's wild blueberries (Statistics Canada 2020). Wild blueberries are harvested manually or mechanically and harvesting losses occur while harvesting with hand rakes and can be minimized with improved management practices (Dale et al., 1994). When wild blueberries go through harvest operations in mechanical harvesting, they are tremendously susceptible to mechanical damage because of their soft texture (Fan et al., 2017). Efforts continued for improve mechanical harvesters to reduce harvesting losses. Peterson et al. (1997) redesigned an experimental highbush blueberry harvester and compared it with a commercial rotary-style harvester and reported 6.9 and 8.6% harvesting losses for the experimental and commercial harvester, respectively. Farooque et al. (2014) reported average fruit yield of 8000 kg ha^{-1} in well managed wild blueberry fields located in Central Nova Scotia and observed more than 10% of fruit losses during mechanical harvesting. Traditionally, the wild blueberry farmers depend on their experience and historical data to increase short-term profitability and long-term durability of their operation (Arbuckle and Rosman 2014). New promising technologies such as ML have appeared over the last years that can potentially aid farmers' decision-making (González Sánchez et al., 2014).

Machine learning is a branch of artificial intelligence that focuses on learning from the existing data to help the growers in making informed decisions. The ML studies consist of different challenges when aiming to build a high-performance predictive model. It is crucial to select the right models to solve the problem at hand, and in addition, the models and the underlying platforms need to be capable of handling the volume of data (Klompenburg et al., 2020). Shahhosseini et al. (2019) used four ML models i.e., RF, Least Absolute Shrinkage and Selection Operator (LASSO) regression, ridge regression, and extreme gradient boosting with their ensembles, which were tested to predict maize yield and nitrate losses. Their results showed that the RF model more precisely predicted maize yield and Nitrogen losses. Boroujeni et al., (2019) used support vector regression (SVR) to predict apricot yield and concluded that SVR was able to estimate apricot yield with high accuracy (R^2= 0.81). Abbas et al., (2020) predicted potato yield using four ML models namely LR, elastic net (EN), k-nearest neighbor (k-NN), and SVR concluded that all the algorithms worked very well in explaining the tuber yield having R^2= 0.70, 0.65, 0.64 and 0.72 respectively.

The literature review has shown that various ML models have been used for the prediction of crop yield and losses. However, limited work has been done using ML models for the prediction of wild blueberry fruit losses during harvesting. Therefore, the objective of this research was to predict wild blueberry ground losses during harvesting using ML models. The data from two commercial wild blueberry fields of Atlantic Canada were used for this study including Frank Webb and Tracadie having field areas of 2.57 and 1.6 ha, respectively. The selected fields were harvested using a commercial blueberry harvester from early August to early September each year to simulate early and late season harvesting. Wild blueberry losses (fruit loss on ground, leaf losses, blower losses) and yield were measured manually from randomly selected plots during mechanical harvesting. The blower loss was collected by attaching a bucket under the blower fan of the harvester. Berries on the ground were manually picked from each plot for calculating the ground loss. For the leaf loss, the leaves and debris were separated from the collected good berries and measured to determine the actual weight of fruit yield and losses. Wild blueberry plant height, fruit zone, and field slope readings were recorded from each of the plots. Two ML models namely LR and RF were used to predict ground losses as a function of plant height, fruit zone, slope, fruit yield, leaf loss, and blower loss. Coefficient of determination were used to assess the prediction accuracy of the models. Correlation matrices were developed to identify the relationships between ground losses and other input variables. The results of the Pearson correlation revealed that fruit yield and other losses (leaf loss, blower loss) had moderate to high

correlations with the ground loss with r ranging from 0.32- 0.78. Linear regression model performed best in both fields Frank Webb and Tracadie with R^2= 0.91 and 0.87 as compared to RF model with R^2= 0.53 and 0.78 respectively.

The comparison of these algorithms suggested that the LR performed comparatively better for both fields. The LR performed better because it uses data to learn by minimizing loss such as MAE and R^2 (Ray, 2019). The finding of this study also emphasizes the better performance of LR compared to RF due to its better optimization techniques for a high number of variables (Drucker et al., 1997). Linear regression performance was best in both fields as compared to RF. Based on the results of this study, LR model is suggested to predict the ground losses in the selected blueberry fields. These results will further help in optimizing the harvesting techniques.

References

Abbas, F., Afzaal, H., Farooque, A. A., Tang, S. (2020). Crop yield prediction through proximal sensing and machine learning algorithms. *Agronomy,* 10, 1046.

Arbuckle, J., Gand, Rosman, H. (2014). Iowa farmers' nitrogen management practices and perspectives Extension Community and Economic Development Publications (Ames, IA: Iowa State University Extension and Outreach) Brown, G. K., N. L. Schulte, E. J. Timm, R. M. Deaudry, J. F. Hancock, D. L. Peterson, and F. Takeda. 1996. Estimates of mechanization effects on fresh blueberry quality. *Appl. Eng in Agric.* 12, 21-26.

Boroujeni, E., Karimi, E., Shirani, H., Esmaeilizadeh, M., Mosleh, Z. (2019). Yield prediction of apricot using a hybrid particle swarm optimizationimperialist competitive algorithm- support vector regression (PSO-ICA-SVR) Method. *Scientia Horticulturae,* 257, 108756.

Dale, A., Hanson, E. J., Yarborough, D. E., McNicol, R. J., Stang, E. J., Brennan, R., ... Hergert, G. B. (1994). Mechanical harvesting of berry crops. In J. Janick (Ed.), *Horticultural Reviews* (Vol. 16, pp. 255-382). Oxford, U. K.: John Wiley & Sons. http://dx.doi.org/10.1002/9780470650561.ch8.

Drucker, H., Surges, C. J. C., Kaufman, L., Smola, A., Vapnik, V. (1997). Support vector regression machines. In Advances in Neural Information Processing Systems; The MIT Press: Cambridge, MA, USA. 9, 155–161.

Farooque, A. A., Zaman, Q. U., Groulx, D., Schumann, A. W., Yarborough, D. E., Nguyen-Quang, T. (2014). Effect of ground speed and header revolutions on the picking efficiency of a commercial wild blueberry harvester. *Appl. Eng. in Agric. 30,* 535-546.

Fan, S., Li, C., Huang, W., Chen, L. (2017). Detection of blueberry internal bruising over time using NIR hyperspectral reflectance imaging with optimum wavelengths. *Postharvest Biol. Technol.* 134, 55–66.

González Sánchez, A., Frausto, Solís, J., Ojeda Bustamante, W. (2014). Predictive ability of machine learning methods for massive crop yield prediction. 12, 313–28.

Klompenburg, T., Kassahun, A., Catal, C. (2020). Crop yield prediction using machine learning: A systematic literature review. *Comput Electron Agric.* 0168-1699.

Peterson, D. L., Wolford, S. D., Timm, E. J., Takeda, F. (1997). Fresh market quality blueberry harvester. *Transactions of the ASAE.* 40, 535-540.

Ray, S. (2019). A Quick Review of Machine Learning Algorithms. 2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (Com-IT-Con), India, 14th -16th Feb 2019.

Shahhosseini, M., Martinez-Feria, R. A., Hu, G., Archontoulis, S. V. (2019). Maize yield and nitrate loss prediction with machine learning algorithms. *Environ. Res. Lett.* 14, 124026.

Statistics Canada (2020). Table 32-10—364-01. Area, production and farm gate value of marketed fruits. DOI: https://doi.org/10.25318/3210036401-eng.