Multi-Sensor Remote Sensing: An AI-Driven Framework for Predicting Sugarcane Feedstock [1]

Marcelo Barbosa[1,2], Dulis Duron[1], Fagner Rontani[1], Giulia Bortolon[1], Bruno Moreira[2,3], Luan Oliveira[4], Kyle Holland[5], Tri Setiyono[1], Luciano Shiratsuchi[1], Rouverson Silva[2]

[1] AgCenter, School of Plant, Environmental and Soil Sciences, Louisiana State University, Baton Rouge, LA, USA.
[2] Department of Engineering and Mathematical Sciences, São Paulo State University, Jaboticabal, SP, Brazil.
[3] Department of Crop and Soil Sciences, University of Georgia, Athens, GA, USA.
[4] Department of Horticulture, University of Georgia, Tifton, GA, USA.
[5] Holland Scientific Inc, Lincoln, NE, USA.

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

Abstract

Predicting saccharine and bioenergy feedstocks in sugarcane enables growers and industries to determine the precise time and location for harvesting a better-quality product in the field. On one hand, Brix, Purity, and total recoverable sugars (TRS) can provide meaningful and reliable indicators of high-quality raw materials for first-generation (1 G) bioethanol. Conversely, Cellulose, Hemicellulose, and Lignin are the primary constituents of straw, directly contributing to second-generation (2 G) bioethanol. However, analyzing these materials in the laboratory is a time-consuming and non-scalable task. Therefore, we propose an approach based on a multi-sensor framework, which includes multispectral unmanned aerial vehicle (UAV) imagery, thermal, photosynthetic active radiation (PAR), and chlorophyll fluorescence (ChlF) data, along with machine learning (ML) algorithms namely random forest (RF), multiple linear regression (MLR), decision tree (DT), and support vector machine (SVM), to develop a non-invasive and predictive framework for mapping sugarcane feedstocks. We collected samples of stalks and leaves/straw during the maturity stage while simultaneously collecting remote sensing data. The ML models played a crucial role in predicting 1 G (R^2 = 0.88–0.93) and 2 G (R^2 = 0.56–0.82) feedstocks. Our study marks a significant advancement in the industrial-scale prediction of sugarcane feedstocks, providing stakeholders with invaluable prescriptive harvesting strategies for both primary products and by-products.

Keywords

Sugar content, Lignocellulosic content, Multispectral imagery, Machine learning, Active sensor, 1G and 2G bioethanol.

The full paper was published in the journal *Industrial Crops and Products*: https://doi.org/10.1016/j.indcrop.2024.118627

Introduction

Sugarcane (*Saccharum* spp.) is cultivated commercially across more than 100 countries, yielding an annual production exceeding 1.91 billion tons (FAOSTAT 2020). As the leading biomass crop globally, it stands as the principal source of sugar, contributing to over 80% of the total global sugar production (FAOSTAT 2020). Furthermore, sugarcane serves as a critical supplier for bioenergy and bioethanol, owing to its substantial cellulose content (Palliprath et al. 2023; Sica et al. 2023). Sugar-energy mills typically harvest and process sugarcane stalks to extract juice, which is used to produce sugar and first-generation (1 G) bioethanol (Freitas et al. 2021). As a result, residual bagasse generated during milling is usually incinerated and serves as a primary raw material for second-generation (2 G) bioethanol (Chandel et al. 2019; Mustafa et al. 2023; Padella et al. 2019). This strategic integration ensures a sustainable bioethanol production pathway while aligning with broader objectives encapsulated in the Sustainable Development Goals (SDGs). These efforts specifically contribute to SDG#7 (Affordable and Clean Energy), SDG#9 (Industry, Innovation, and Infrastructure), SDG#12 (Responsible Consumption and Production), and SDG#13 (Climate Action).

In 1 G bioethanol production, the efficiency of the harvested product depends on the mechanical harvest schedule. Harvesting typically begins when crops reach maturity, which is characterized by the highest sugar content. This sugar content can be measured through Brix, Purity, and total recoverable sugar (TRS). However, field teams usually estimate these feedstocks. While this approach has proven to be effective, it is also costly, labor-intensive, and time-consuming. Therefore, integrating remote sensing technology serves as a feasible alternative to the traditional sampling methodology.

The adoption of mechanical harvesting stands as a pivotal element in sustaining an intensive sugarcane production system. Nevertheless, this method yields a significant agri-residue in the form of straw, amounting to 10–20 t ha^{-1} $year^{-1}$ (Bilatto et al. 2020; Freitas et al. 2021; Michelin et al. 2023). Straw is mainly composed of dry leaves (60%) and green tops (40%) after the sugarcane stalks are harvested (Aguiar et al. 2021). Additionally, straw is basically composed of Cellulose, Hemicellulose and Lignin in stoichiometric proportions of 40, 28 and 21% respectively (Pereira et al. 2015). In particular, Cellulose is the polymer of interest in the production of 2 G bioethanol by enzymatic hydrolysis. Therefore, quantifying it, spatially and temporally, is crucial to the success of collecting raw materials from the production field and industrial processing. Consequently, there is a convincing need for new studies that specifically address the real-time estimation of those 2 G feedstocks concentrations in sugarcane straw during the crop growing period using remote sensing.

Therefore, we conducted a comprehensive field study to evaluate whether an integrated sensing and ML algorithms can be useful for predicting 1 G (Brix, Purity, and TRS) and 2 G (Cellulose, Hemicellulose, and Lignin) sugarcane feedstocks during the maturity stage, presenting a significant opportunity to strengthen and refine the industry process.

Material and Methods

Site Study

The study was conducted in a commercial sugarcane field in the city of Napoleonville, Louisiana, USA. The sugarcane cultivar L 01–299 was planted in September 2021 and harvested for the first time in October 2022; hence, our research was conducted on the first ratoon, from August-October 2023. We started the data collection at the maturity stage by sampling the stalks, tops, and using sensors. For each data collection we defined 22 sample plots regularly distributed throughout the field. Each plot had equal dimension of 10 × 10 m. For both 1 G analysis (Brix, Purity, and TRS) and 2 G analysis (Cellulose, Hemicellulose, and Lignin), the data collection was performed five times during the maturity stage with an interval of ∼15 days. Data collection for both 1 G and 2 G involved sampling 10 millable stalks and green top leaves (∼1 kg of fresh material), respectively, on five different dates at each sample plot.

Laboratory Analysis

The analysis of 1 G (Brix, Purity, and TRS) and 2 G (Cellulose, Hemicellulose, and Lignin) feedstocks was conducted as part of routine procedures in dedicated sugar and forage laboratories, respectively. These analyses employed wet chemistry methods. For the 1 G analysis, samples comprising 10 stalks each were individually analyzed, adhering to the methodology prescribed by Legendre (1992). While for the 2 G analysis, green top leaves and straw samples were processed through a sequential procedure involving neutral detergent fiber (NDF), acid detergent fiber (ADF), and acid detergent lignin (ADL), as outlined by the methodology established by Goering (1970).

Multi-Sensor Framework and Data Collection

The multi-sensor framework was composed of aerial and proximal sensors. For the aerial sensor we used a multi-rotor UAV (DJI Matrice 300 RTK, Shenzhen, China) as remote sensing platform equipped with a multispectral camera (MicaSense RedEdge-MX, MicaSense Inc., Seattle, WA, USA). For the proximal sensors we used on-the-go sensors. For chlorophyll fluorescence (ChlF) we used an active sensor (Crop Circle ACS-527, Holland Scientific, Lincoln, NE, USA). Finally, we used a multi-parameter sensor (Crop Circle DAS43X, Holland Scientific, Lincoln, NE, USA) to quantify air temperature, canopy temperature, and incoming and reflected photosynthetic active radiation (PAR). For our analysis, we subtracted air temperature from canopy temperature to generate delta temperature (DTemp); and divided reflected PAR by incoming PAR to generate fractional PAR (fPAR). To better response plant properties, we used a chlorophyll concentration meter (MC-100, Apogee Instruments Inc., Logan, UT, USA) for soil plant analysis development (SPAD) analysis. Three plants were selected in each sample plot for SPAD readings. The readings were performed in the top visible dewlap (TVD) leaves.

Data Analysis

The dataset was randomly split into train and test subsets in the proportion of 70% (n = 77) and 30% (n = 33), respectively. In this study, we used 4 ML algorithms to perform the predictive models namely ML, RF, DT, and SVM. The assessment of models precise and accuracy were conducted based on key metrics, namely the coefficient of determination (R^2), mean absolute error (MAE), and root mean square error (RMSE), applied to the test dataset.

Results

The performance results of all models across different feedstocks are systematically presented in **Table 1**, encompassing precision (R^2) and accuracy metrics (MAE and RMSE). All the models were executed using all the available features, and notably, each exhibited commendable performance across diverse feedstocks. However, apparent enhancements in predictive capabilities, justifying emphasis, were observed in the case of MLR, RF, and DT. Despite its simplicity in statistical complexity, MLR demonstrated a slightly superior performance in predicting Brix, exhibiting an approximately 6% enhancement in accuracy. Conversely, the DT model produced superior results for Purity, excelling in both precision and accuracy metrics. Noteworthy is the superior prediction of 1 G feedstock TRS, with RF better, particularly in terms of MAE. Similarly, Cellulose feedstock prediction experienced improvement, although moderated, with the implementation of the RF algorithm. In contrast, the DT model emerged as the optimal choice for Hemicellulose and Lignin prediction, achieving superior results, particularly in the context of 2 G feedstock. Conversely, despite slight differences, the SVM exhibited no comparable proficiency in representing any feedstock, lagging behind the others predictive models. This underscores the importance of considering the dataset trend to determine the most suitable algorithm.

To enhance the illustration of the models' performance, **Figure 1** provides a visual comparison of R^2, MAE, and RMSE values derived from the test dataset for each feedstock. Generally, the 1 G models demonstrated a slight tendency towards over and underestimation in comparison to their 2 G feedstocks. Concerning 2 G, the Cellulose model exhibited commendable performance for low and medium values; however, for higher values, the models' trend began to underestimate.

Table 1. Machine learning models performance for the sugarcane feedstocks.

ML Algorithm	Feedstock	R^2	MAE	RMSE
MLR	**Brix**	**0.88**	**0.27**	**0.33**
	Purity	0.90	1.27	1.54
	TRS	0.93	3.42	4.15
	Cellulose	0.37	0.80	0.96
	Hemicellulose	0.73	0.79	0.99
	Lignin	0.78	0.30	0.41
RF	Brix	0.87	0.29	0.35
	Purity	0.90	1.26	1.52
	TRS	**0.93**	**3.23**	**4.17**
	Cellulose	**0.56**	**0.70**	**0.81**
	Hemicellulose	0.81	0.68	0.87
	Lignin	0.77	0.32	0.43
DT	Brix	0.88	0.27	0.35
	Purity	**0.91**	**1.20**	**1.48**
	TRS	0.92	3.92	4.78
	Cellulose	0.48	0.68	0.86
	Hemicellulose	**0.82**	**0.69**	**0.84**
	Lignin	**0.79**	**0.31**	**0.40**
SVM	Brix	0.81	0.35	0.41
	Purity	0.89	1.31	1.59
	TRS	0.93	3.72	4.37
	Cellulose	0.54	0.75	0.84
	Hemicellulose	0.78	0.76	0.91
	Lignin	0.77	0.31	0.41

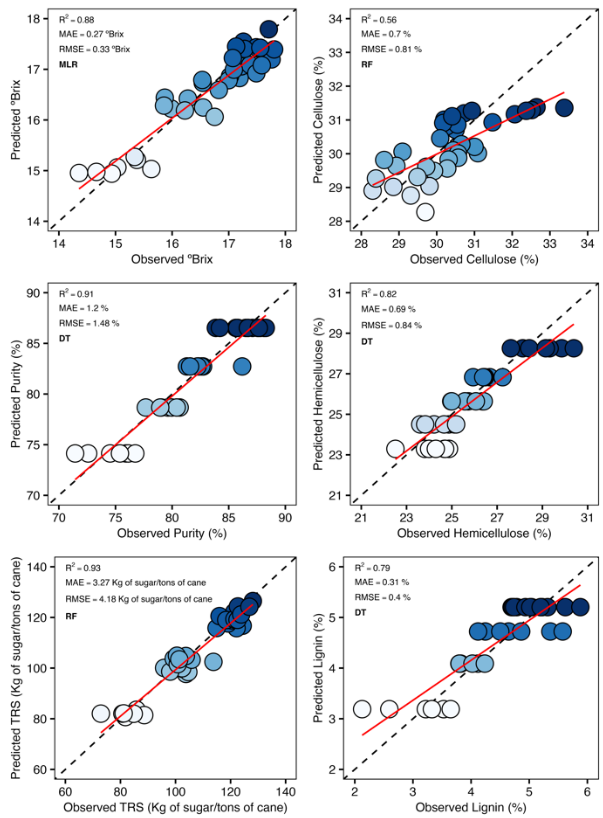

Figure 1. Scatterplots for the bests ML models performance in predicting the 1 G and 2 G feedstocks. Blue points merge from light to dark blue, representing the low and high values, respectively.

Conclusion

This study introduces a valuable methodology wherein ML algorithms are applied to remote sensing data as features to construct predictive models for 1 G and 2 G sugarcane feedstocks. The results obtained are robust, affirming the feasibility of predicting both types of feedstocks. Particularly, TRS exhibited enhanced predictability in relation to 1 G feedstock, while Hemicellulose demonstrated superior performance for 2 G. This information is invaluable for researchers and specialists, providing practical insights into deploying ML algorithms for the accurate prediction of sugarcane feedstocks. The study highlights variations in the improvement rate of prediction performance in ML algorithms, emphasizing the importance of adapting the approach according to the characteristics of the specific ML algorithm used. The outcomes of this work offer timely and remote prediction results for sugarcane feedstock quality. These findings can serve as a foundation for further research and advisory activities, enabling them to optimize farm management strategies for sugarcane. This, in turn, contributes to the development of more agricultural systems by enhancing the efficiency and precision of feedstock predictions.

Acknowledgements

We would like to acknowledge the São Paulo Research Foundation (Fapesp), for the scholarship (2022/13992–0) to the first author; the laboratories of Precision Agriculture and Digital Agriculture of the School of Plant, Environment and Soil Sciences (SPESS) for the infrastructural support; the Holland Scientific for the Crop Circle sensors; the Sugar Research Station Laboratory of the LSU AgCenter for the technological analysis; the Forage Laboratory of the SPESS for the bromatological analysis; and the Laboratory of Machinery and Agricultural Mechanization (LAMMA) of the Department of Engineering and Mathematical Sciences. This work was also supported by Precision Agriculture for Sugarcane Production Project (John Deere: BG-008042).

References

Aguiar, A., Milessi, T. S., Mulinari, D. R., Lopes, M. S., da Costa, S. M., & Candido, R. G. (2021). Sugarcane straw as a potential second generation feedstock for biorefinery and white biotechnology applications. *Biomass and Bioenergy*, *144*, 105896. https://doi.org/10.1016/j.biombioe.2020.105896

Bilatto, S., Marconcini, J. M., Mattoso, L. H. C., & Farinas, C. S. (2020). Lignocellulose nanocrystals from sugarcane straw. *Industrial Crops and Products*, *157*, 112938. https://doi.org/10.1016/j.indcrop.2020.112938

Chandel, A. K., Albarelli, J. Q., Santos, D. T., Chundawat, S. P., Puri, M., & Meireles, M. A. A. (2019). Comparative analysis of key technologies for cellulosic ethanol production from Brazilian sugarcane bagasse at a commercial scale. *Biofuels, Bioproducts and Biorefining*, *13*(4), 994–1014. https://doi.org/10.1002/bbb.1990

FAOSTAT. (2020). FAOSTAT. *Crops and Livestock Products. Food and Agriculture Organization of the United Nations.* http://www.fao.org/faostat/en/#data/QCL

Freitas, J. V., Bilatto, S., Squinca, P., Pinto, A. S. S., Brondi, M. G., Bondancia, T. J., et al. (2021). Sugarcane biorefineries: potential opportunities towards shifting from wastes to products. *Industrial Crops and Products*, *172*, 114057. https://doi.org/10.1016/j.indcrop.2021.114057

Michelin, M., Ximenes, E., M. Polizeli, M. de L. T., & Ladisch, M. R. (2023). Inhibition of enzyme hydrolysis of cellulose by phenols from hydrothermally pretreated sugarcane straw. *Enzyme and Microbial Technology*, *166*, 110227. https://doi.org/10.1016/j.enzmictec.2023.110227

Mustafa, G., Arshad, M., Bano, I., & Abbas, M. (2023). Biotechnological applications of sugarcane bagasse and sugar beet molasses. *Biomass Conversion and Biorefinery*, *13*(2), 1489–1501. https://doi.org/10.1007/s13399-020-01141-x

Padella, M., O'Connell, A., & Prussi, M. (2019). What is still Limiting the Deployment of Cellulosic Ethanol? Analysis of the Current Status of the Sector. *Applied Sciences*, *9*(21), 4523. https://doi.org/10.3390/app9214523

Palliprath, S., Poolakkalody, N. J., Ramesh, K., Mangalan, S. M., Kabekkodu, S. P., Santiago, R., & Manisseri, C. (2023). Pretreatment of sugarcane postharvest leaves by γ-valerolactone/water/FeCl3 system for enhanced glucan and bioethanol production. *Industrial Crops and Products*, *197*, 116571. https://doi.org/10.1016/j.indcrop.2023.116571

Pereira, S. C., Maehara, L., Machado, C. M. M., & Farinas, C. S. (2015). 2G ethanol from the whole sugarcane lignocellulosic biomass. *Biotechnology for Biofuels*, *8*(1), 44. https://doi.org/10.1186/s13068-015-0224-0

Sica, P., de Castro Mattos, E., Silveira, G. M., Abdalla, J. P., Alves, V. K., Borges, I. S., et al. (2023). Quantitative and qualitative evaluation of novel energy cane accessions for sugar, bioenergy, 1 G, and 2 G ethanol production. *Industrial Crops and Products*, *203*, 117163. https://doi.org/10.1016/j.indcrop.2023.117163