ID 2400187

Pre-Harvest Estimation of Alfalfa Quality Traits Using Multi-Type Features and Machine Learning

Tong Yu (tong.yu@wisc.edu)[1], Yijia Xu[1], Jing Zhou[2] , Zhou Zhang[1,*]
[1] Department of Biological Systems Engineering, University of Wisconsin-Madison, [2] Department of Crop and Soil Science, Oregon State University

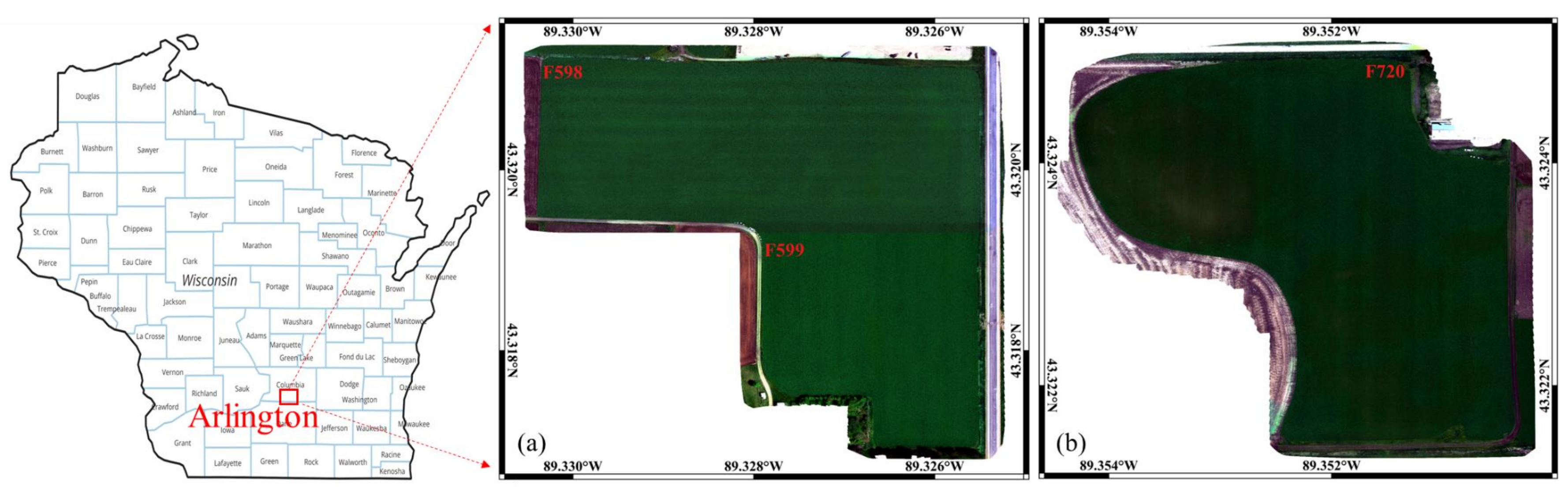
Objectives

- Evaluating the potential of multispectral UAV imagery, environmental data, measured heights, and sampling dates in estimating 16 quality traits of alfalfa.
- Investigating the importance and contribution of multi-type features to different quality traits.
- Assessing the transferability of the method across different fields.

Background

- Alfalfa is a valuable nutritious crop with a comparatively high yield, which is considered the fourth most valuable field crop in the United States and is a commercially grown source of forage and feed in the world [1,2].
- Current research utilizing UAV-based remote sensing technology to monitor field-grown alfalfa primarily focuses on several key traits such as dry matter yield (DMY), crude protein (CP), neutral detergent fiber (NDF) and acid detergent fiber (ADF). Precise estimation of other quality traits such as non-fiber carbohydrates (NFC), fat, and minerals are relatively rare, despite their significant roles in plant growth and livestock development.
- The selection of input features can greatly influence the accuracy of the quality estimation results. Vegetation indices, spectral reflectance, environmental data and various types of features have demonstrated the potential in assessing forage quality, the specific importance and contributions of each type of feature, and even each feature, to different quality traits still require further research and discussion.

Materials and Methods

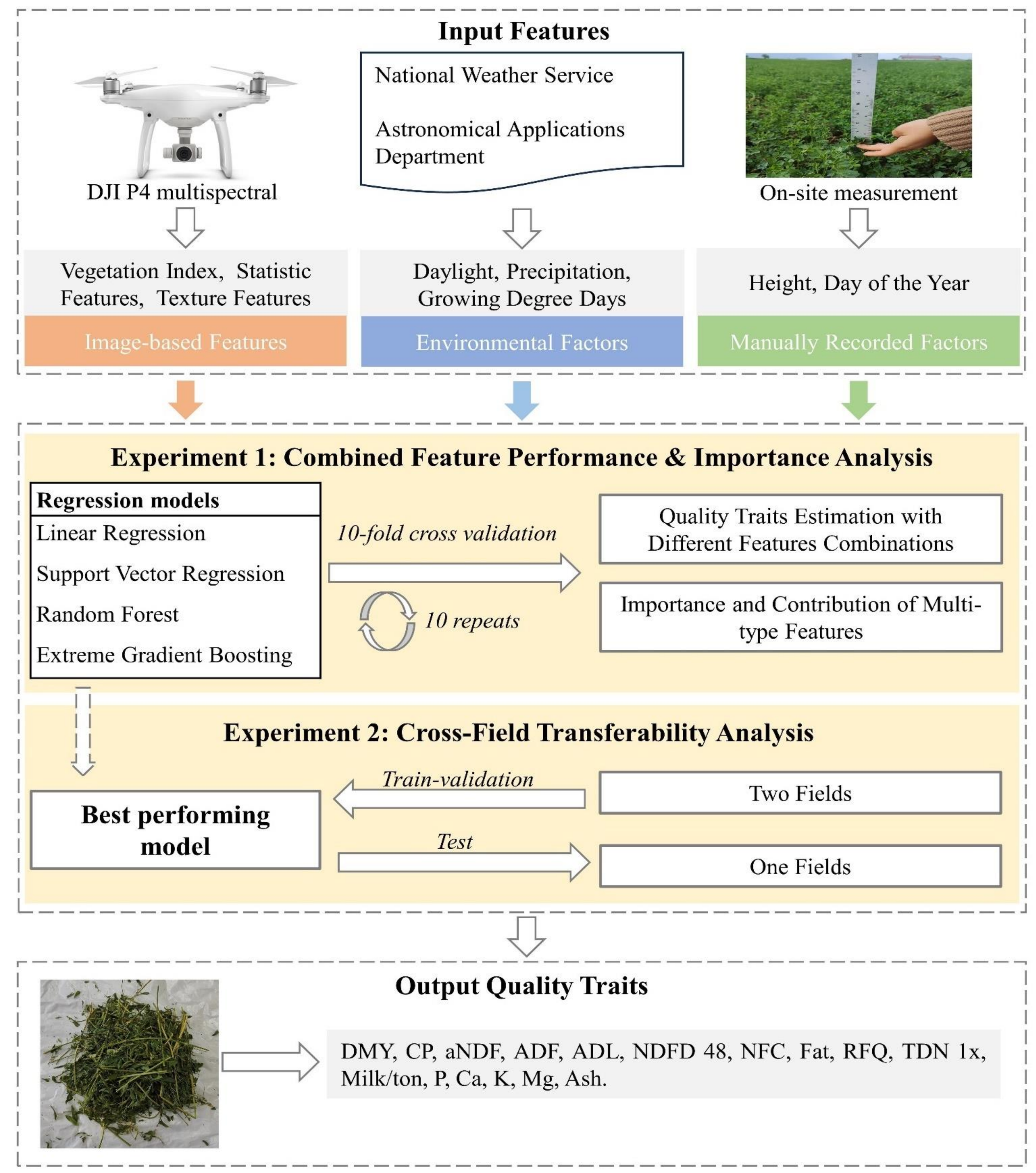


Study area in Arlington Agricultural Research Station

- **Sample Collection**: Sampling was conducted three to four times per month for each field from May to August in 2022 and 2023. Five to seven sites were selected for each sampling within each field. At each site, four subsamples were marked and collected, evenly spaced around a central point with a radius of 7.5 meters.
- **UAV Imagery**: Multispectral data was acquired by a DJI phantom 4 multispectral drone. The data was then processed using Pix4D software. Images covering each site were cropped using manually drawn shapefiles.
- **Environmental Factors**: Temperature, precipitation and daylight duration data were collected from [3] and [4]. Growing degree days (GDD) were calculated from max and min temperature.

This work was supported by the United States Department of Agriculture (USDA) National Institute of Food and Agriculture, Alfalfa and Forage Program under Grant 1027160.

Overall Workflow



Results

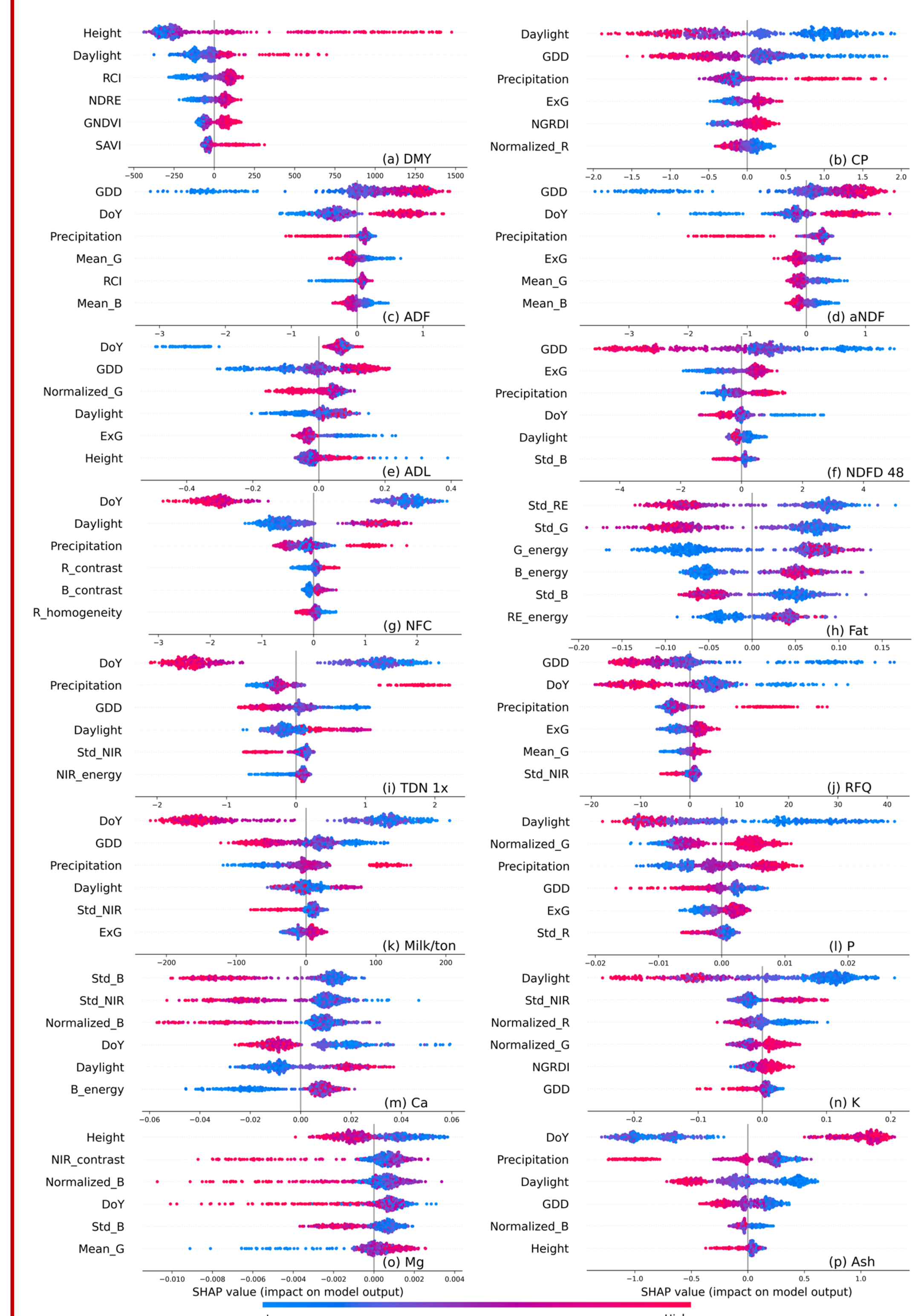
- Combinations of EF and MF, as well as the combination of all three types of features, exhibited better performance.
- For DMY and CP, all three types of features contributed positively, and combining them all achieved the highest R^2 of 0.921 (RMSE = 300 kg/ha) and 0.695 (RMSE = 1.885 % DM).
- For fiber content, the role of IF significantly diminished, even negatively impacting the estimation.

The R^2 for predicting 16 quality traits using multi-type features

Quality	IF+EF	IF+MF	EF+MF	EF+IF+MF
DMY	0.908	0.917	0.915	**0.921**
CP	0.690	0.627	0.660	**0.695**
ADF	0.413	0.397	**0.463**	0.427
aNDF	0.418	0.408	**0.455**	0.429
ADL	0.412	0.375	**0.461**	0.421
NDFD 48	0.620	0.564	0.624	**0.625**
NFC	0.786	0.757	0.784	**0.795**
Fat	0.864	0.850	**0.870**	0.865
TDN 1x	0.657	0.626	**0.677**	0.652
RFQ	0.565	0.537	**0.575**	0.573
Milk/ton	0.643	0.613	**0.669**	0.649
P	0.768	0.672	0.760	**0.770**
Ca	0.606	0.601	**0.622**	0.612
K	0.588	0.525	0.568	**0.596**
Mg	0.302	0.292	0.301	**0.307**
Ash	**0.757**	0.716	0.750	0.757

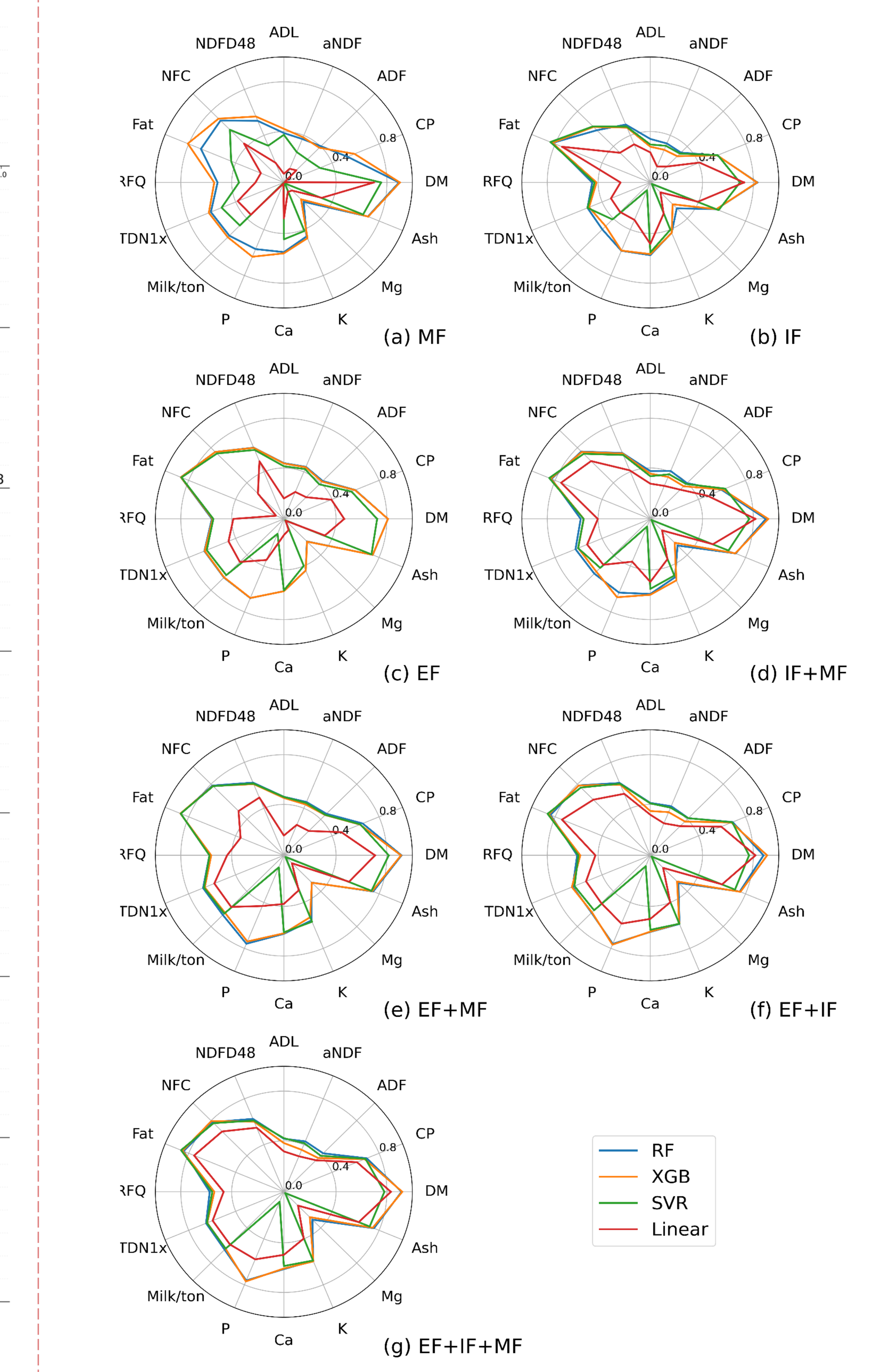
The highest R^2 for each quality trait is in bold and the second highest is underlined.

- The specific contributions of each feature are represented by SHAP values.
- EF contributed more significantly, with GDD, precipitation, and daylight being among the foremost contributors to most quality traits.
- DoY stood out in the estimation of the three quality indices, TDN 1x, RFQ, and Milk/ton.
- The height feature only shows a significant contribution in DMY and Mg.



SHAP value for the top 6 contributing features across 16 quality traits

- RF exhibited exceptional performance when incorporating all three feature types
- XGB followed closely behind RF, demonstrating a similar trend of improved accuracy with the inclusion of more feature types.
- SVR and the LR exhibited lower R^2 values across the traits, with the LR showing the least capacity to capture the complexity of the data.



The R^2 of four ML models with different input features

- Sixteen alfalfa quality traits can be effectively estimated with multispectral UAV imagery, environmental data, and manual records.
- The coefficient of determination of DMY reached 0.921.
- Environmental factors significantly enhance the accuracy of quality traits estimation.
- Ensemble learning models exhibit superior performance with multi-type input features.

More figures, tables and references can be found in: