YIELD ESTIMATION IN HYDROPONIC SPROUTS FROM HIGH-FREQUENCY RGB IMAGING

VYTAUTO DIDŽIOJO UNIVERSITETAS
Informatikos fakultetas

AUTORIAI:

Danylo Abramov
danylo.abramov@vdu.lt

Eimantas Zaranka
eimantas.zaranka@vdu.lt

Rūta Juozaitienė
ruta.juozaitiene@vdu.lt

Andrius Grigas
andrius.grigas@vdu.lt

Tomas Krilavičius
tomas.krilavicius@vdu.lt

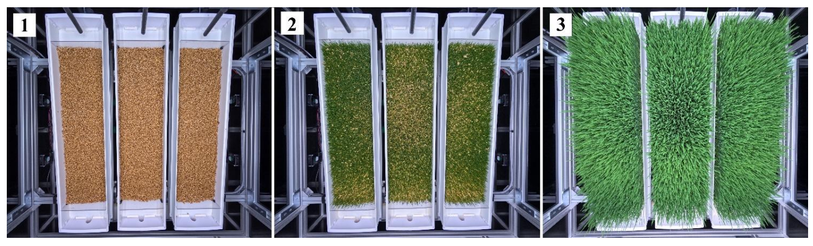
1 MOTIVATION

- Hydroponic farming requires continuous and accurate yield monitoring, but traditional methods are often labour-intensive or rely on expensive sensing equipment.
- Affordable RGB cameras are widely available, yet their use for precise, automated biomass estimation remains limited, creating a need for a low-cost AI-based solution.

2 GOAL

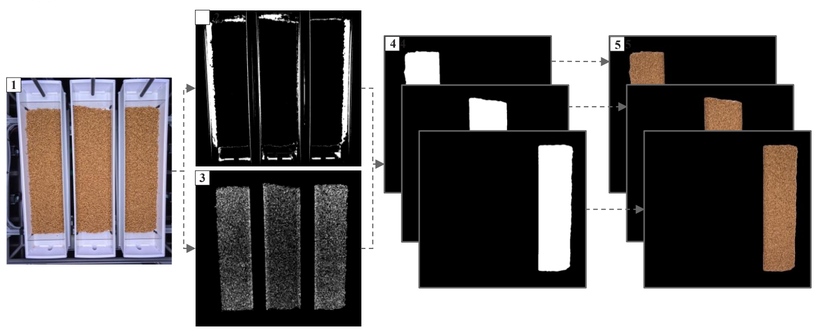
- Develop an automated, non-invasive RGB-based pipeline to estimate hydroponic wheat biomass using synchronized image and weight data.
- Use machine-learning models on extracted visual features to determine whether RGB-only imaging can achieve accurate yield prediction.

3 GROWTH STAGES



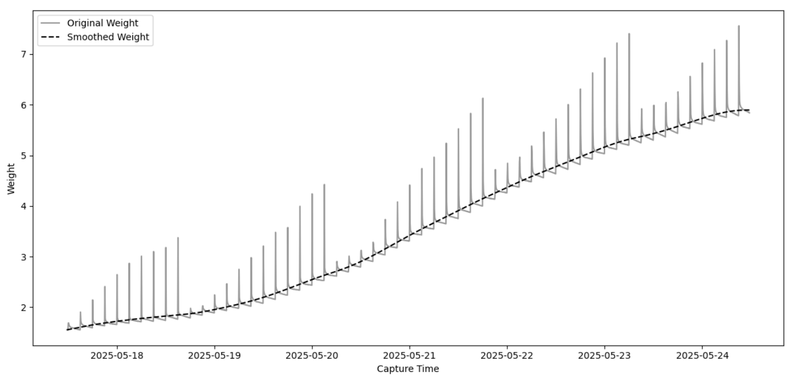
Representative dataset images used for AI model training, illustrating (1) day 1 (early), (2) day 5 (mid-growth), and (3) day 7 (mature) stages of wheat sprouts.

4 TRAY SEGMENTATION



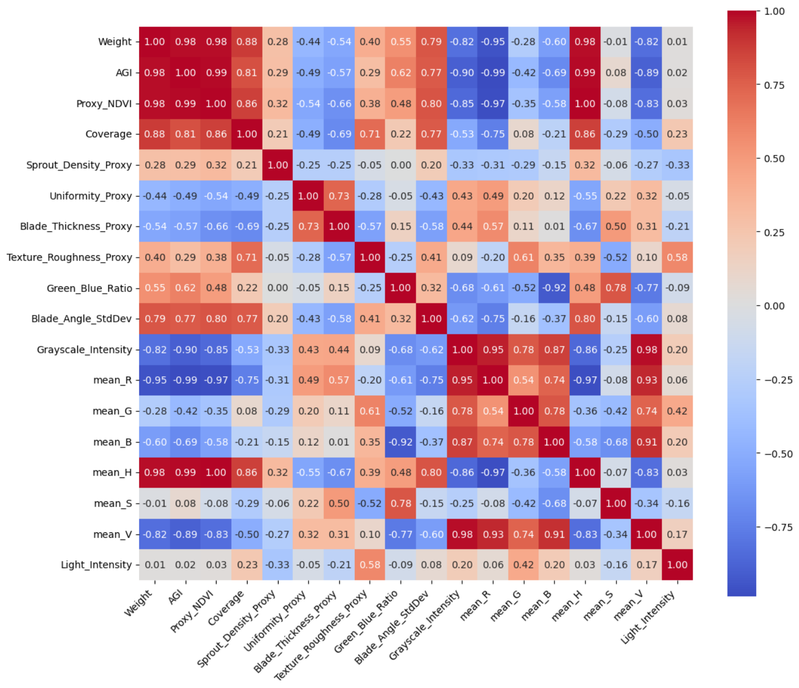
1. Raw tray image
2. White-channel mask highlighting tray boundaries
3. Brown-channel mask enhancing sprout texture
4. SAM generated segmentation masks
5. Final extracted tray regions used for feature computation

5 WEIGHT SMOOTHING



Sprout weight curve smoothed with a Bessel filter to reduce noise and reveal growth trends.

6 FEATURE CORRELATION



Correlation heatmap illustrating dependencies between image-derived features and sprout weight.

7 FEATURE SELECTION

Evaluation of five feature-selection strategies using an identical 5-fold linear-regression baseline.

Subset	MAE	RMSE	sMAPE (%)
Elastic net	0.068	0.087	2.18
Correlation	0.077	0.098	2.53
Mutual information	0.086	0.111	2.67
Random forest	0.109	0.138	3.46
VIF	0.324	0.396	11.07

Elastic Net selected features:

- AGI
- Blade_Thickness_Proxy
- Coverage
- Green_Blue_Ratio
- mean_G
- mean_H
- mean_R
- mean_S
- mean_V
- Proxy_NDVI
- Sprout_Density_Proxy
- Texture_Roughness_Proxy
- Uniformity_Proxy

8 MODEL PERFORMANCE

Predictive performance of eight machine-learning models trained on the Elastic net feature set, with Random Forest achieving the lowest error metrics.

Model	MAE	RMSE	sMAPE
Random Forest	0.006	0.018	0.22 %
LightGBM	0.008	0.015	0.27 %
XGBoost	0.008	0.017	0.27 %
MLP (Neural Net)	0.019	0.030	0.67 %
SVR (RBF)	0.053	0.063	1.74 %
Linear Regression	0.068	0.087	2.18 %
Ridge Regression	0.070	0.090	2.30 %
Lasso Regression	0.078	0.098	2.62 %

9 CONCLUSIONS

- RGB-only imaging with synchronized weight data enables accurate, non-invasive biomass estimation in hydroponic sprouts.
- Elastic Net produced the most reliable feature subset by reducing multicollinearity and preserving informative variables, achieving a baseline model sMAPE of 2.18%.
- Random Forest and boosting models achieved the best predictive performance, with sMAPE values of 0.22% and 0.27 %, respectively. The strongest predictors were greenness, color, and coverage features.

10 FUTURE PLANS

- Expand the dataset across more growth cycles and varying environmental conditions.
- Explore deep learning models to learn biomass-relevant representations directly from full RGB images.
- Develop an automated crop-monitoring and decision-support pipeline for controlled-environment agriculture.

Co-funded by the European Union

SustAIn Liv Work

CARD
CENTRE FOR APPLIED RESEARCH AND DEVELOPMENT