A Decision-Consistent Framework for Yield Prediction and Genotype Selection using UAV-Based Temporal Phenotyping

Abstract

Accurate and timely identification of superior genotypes is a central objective of modern plant breeding. Recent advances in unmanned aerial vehicle (UAV)-based high-throughput phenotyping and machine learning have improved the estimation of crop yield from remotely sensed data; however, most existing approaches treat the problem as conventional numerical regression. Such formulations often struggle to reflect the operational priorities of breeding programs, where reliable ranking of candidate lines, consistency across advancement thresholds, and robust performance under limited field data are frequently more important than minimizing average prediction error alone. In addition, multi-temporal UAV phenotyping introduces practical challenges related to spatial misalignment, temporal inconsistency, and heterogeneous crop developmental patterns.

This thesis addresses these limitations through a unified decision-aware spatio-temporal learning framework for genotype selection. The framework is organized around four complementary components. First, a spatio-temporal image registration and data harmonization pipeline is developed to ensure geometric consistency of multi-date UAV orthophotos, preserving plot-level identity across the growing season and enabling reliable extraction of phenotypic trajectories for downstream analysis.

Second, to mitigate the scarcity and seasonal limitations of field phenotyping datasets, a yield-conditioned synthetic time-series generation strategy is introduced. By jointly modeling temporal vegetation index dynamics and associated yield outcomes, the proposed augmentation framework expands training diversity while preserving biologically meaningful dependencies between canopy development and productivity. Hybrid training with real and synthetic data consistently improves robustness relative to real-only training.

Third, yield prediction is reformulated as a structured decision problem rather than a purely continuous estimation task. A Boundary-Consistent Ordinal Regression (B-COR) framework is proposed to incorporate agronomic performance tiers directly into the learning objective. By enforcing consistency around decision-critical thresholds, the method reduces boundary leakage and improves categorical reliability while maintaining strong regression performance.

Fourth, a Rank-Aware Multi-Task Learning (RA-MTL) framework is developed to explicitly model relative genotype performance. By jointly optimizing regression, ordinal classification, and pairwise ranking objectives, the proposed model preserves numerical fidelity, boundary consistency, and global ordering among candidate lines, enabling more reliable identification of elite genotypes in top-selection scenarios.

The framework is evaluated across multiple crops, including sunflower, wheat, and rice, under independent seasonal conditions. Results demonstrate strong and consistent performance across regression, ordinal, and ranking objectives, achieving cross-crop R^2 values of 0.62-0.77, ordinal error as low as 0.19, Spearman rank correlations of 0.85-0.88, and Top-20% selection accuracy of 0.77-0.80. Temporal analysis further reveal that reproductive and grain-filling stages provide the most reliable windows for early genotype ranking, offering practical opportunities for accelerated selection prior to harvest maturity.

Collectively, this thesis advances UAV-based phenotyping from isolated yield prediction toward selection-centric intelligence. By integrating data integrity, scarcity-robust augmentation, ordinal-boundary consistency, and ranking-aware optimization within a unified framework, the research provides a scalable and practically deployable methodology for next-generation breeding programs.