Contribution ID: **277**   Type: **Talk**

Hybrid Stacking and Embedded Regression with Multi-Phase Feature Selection for Explainable Crop Yield Prediction in Botswana

Wednesday, 3 December 2025 15:35 (15 minutes)

Hybrid Stacking and Embedded Regression with Multi-Phase Feature Selection for Explainable Crop Yield Prediction in Botswana

Abstract

In Sub-Saharan Africa's climate instability, inaccurate data, and lack of precision agricultural tools make it extremely difficult to predict crop yields with any degree of accuracy. These restrictions are especially critical in Botswana, where most agricultural activities are rain-fed and highly vulnerable to environmental changes. To provide accurate, comprehensible, and context-specific yield predictions for four staple crops: Maize, Millet, Pulses, and Sorghum. This study uses a hybrid machine learning approach. The approach integrates multiple regression algorithms: Random Forest, XGBoost, Support Vector Regression, and Multi-Layer Perceptron within a stacked ensemble architecture tailored to Botswana's agricultural data context. To optimize predictive power and interpretability, a multi-phase feature selection strategy was applied, combining entropy filtering, mutual information, recursive feature elimination (RFE), and engineered temporal features through lag variables.

This process refined input variables for both the staging models and region-specific selection, ensuring robust model generalization. Model performance was evaluated using historical yield, meteorological, and soil datasets, with R^2, RMSE, and MAE employed as metrics. The Stacking Hybrid Regression Model performed exceptionally well in yield prediction for pulses and sorghum, achieving the best performance with R2 = 0.94, RMSE = 0.60 t/ha, and MAE = 0.32 t/ha. The most significant predictors were rainfall, temperature fluctuation, and lagged yield values, according to a unified interpretability framework that was produced by combining SHapley Additive exPlanations (SHAP) with entropy analysis. Surprisingly, entropy research showed that Sorghum had a greater predictor complexity and shown the ability to adjust to unpredictable weather. Time-horizon stability of the model was confirmed by forward simulations for 2025–2028.

These results confirm that interpretable hybrid ensembles can satisfy precision agriculture's accuracy and transparency requirements when reinforced by multi-phase feature selection. The suggested approach supports climate risk management tactics for Botswana's farmers by providing useful information for early-season production projection and input distribution. Additionally, other sub-Saharan regions with comparable environmental and data-related constraints may find the methodology applicable.

Keywords: predictive crop yield, precision agriculture, Botswana, XAI, multi-phase feature selection, hybrid ensemble models, and SHAP.

Presenting Author

Kalu Ubi Kalu

Email

kaluubi3@gmail.com

Student or Postdoc?

Masters

Institute

University of Botswana

Registered for the conference?

Yes

CHPC User

No

CHPC Research Programme

Primary author: Mr UBI KALU, Kalu (University of Botswana)

Co-authors: Dr ANDERSON, George (Co-Author); Dr MASIZANA, Audrey (University of Botswana)

Presenters: Mr UBI KALU, Kalu (University of Botswana); Dr ANDERSON, George (Co-Author); Dr MA-SIZANA, Audrey (University of Botswana)

Session Classification: DIRISA

Track Classification: DIRISA