IECAG 2025 Conference

The 5th International Electronic Conference on Agronomy

15–18 December 2025 | Online

MDPI

AI-Driven Paddy (*Oryza sativa*) Yield Forecasting Using Open Satellite Data, Weather APIS and Historical Data for Sri Lankan Agro zones

D. Senevirathne 1, W.M.A.C. Premasiri 2

Department of Manufacturing and Industrial Engineering, Faculty of Engineering, University of Peradeniya, Peradeniya, Sri Lanka 1
Department of Crop Science, Faculty of Agriculture, University of Ruhuna, Matara, Sri Lanka 2

INTRODUCTION & AIM

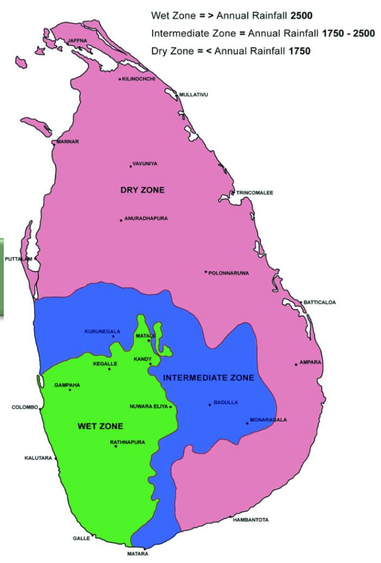
- ***Oryza sativa* (Paddy)** is a widely consumed staple food worldwide and Sri Lanka has been producing paddy for centuries
- A growing population and increasing food demand make timely and accurate paddy yield prediction essential for national food security and resource management
- Conventional yield estimation methods are often manual, time-consuming and limited in accuracy, highlighting the need for modern approaches
- This research introduces a stacked ensemble learning approach using XGBoost, integrating multiple base and meta models, achieving an accuracy (R^2 = 0.9986) for district-level seasonal paddy yields.

Aim –

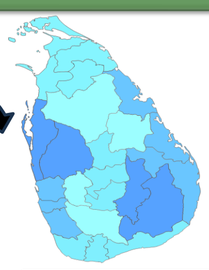
- This study uniquely combines multiple openly available datasets: CHIRPS 2.0 rainfall, NASA POWER climate variables and RRDI soil maps to create a comprehensive dataset for district-level paddy yield forecasting in Sri Lankan agro zones

***Oryza sativa* (Paddy)**

Paddy cultivation in Sri Lanka



METHOD



Study area

Data collection

- ✓ Rainfall, climate, soil maps, historical yield and crop calendar
- ✓ (2004–2024, 25 districts, 3 agro-climatic zones)

Dataset Construction

- ✓ Integrated climate, soil, and historical yield data with simulated crop harvesting patterns
- ✓ Created 12 heterogeneous datasets

Machine Learning Modeling

- ✓ (XGBoost Base + Meta Models)
- ✓ Trained 12 base models on datasets
- ✓ Out-of-fold predictions used to train intermediate meta models
- ✓ Final XGBoost stacked ensemble combined outputs for prediction

Yield Prediction & Evaluation

- ✓ Predicted district-level seasonal paddy yields
- ✓ Performance: RMSE = 3,535 MT, R^2 = 0.9986, NRMSE = 0.76%

RESULTS & DISCUSSION

Model Performance

Final XGBoost stacked model achieved:

- RMSE = 3,535 MT
- R^2 = 0.9986
- NRMSE = 0.76%

Interpretation:
The model predicts district-level seasonal paddy yield with very high accuracy, showing the power of combining climate, soil and historical yield data with crop calendar simulation

Paddy Yield Trends

Time period- (2004-2024)

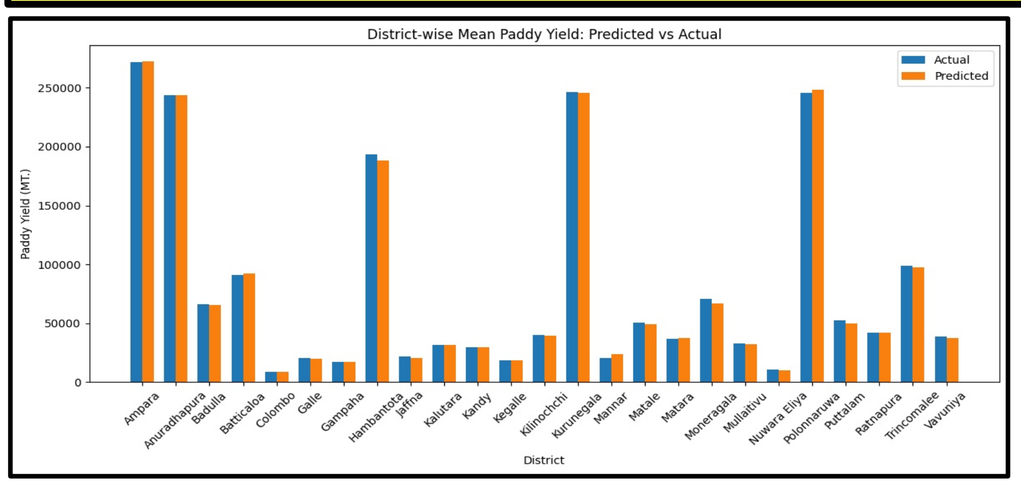
Lowest Yield:
Mannar district, Yala 2006 (~185 MT)
Highest Yield:
Anuradhapura district, Maha 2019–2020 (~530,356 MT)
Observation:
Seasonal and district-wise variations reflect agro-climatic influences on paddy production

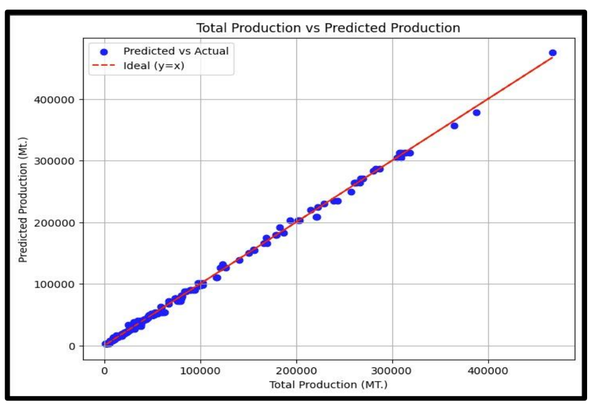
Impact of Factors

- Soil and climatic factors
- Model captured rainfall, temperature, pH, and salinity effects on yield
- Highlights potential for climate-informed agricultural planning

Novelty & Implications

- ➢ **Integrated heterogeneous data + stacked ensemble learning approach is a novel approach for Sri Lankan district-level paddy forecasting**
- ➢ **Can support data-driven decision-making for sustainable paddy production and resource allocation**





CONCLUSION

- ❖ The XGBoost-based stacked ensemble model can accurately predict district-level seasonal paddy yields in Sri Lanka
- ❖ Integrating climate data, soil properties, and historical yields provides a robust and data-driven framework for forecasting
- ❖ The model's high accuracy (R^2 = 0.9986, NRMSE = 0.76%) demonstrates its potential to support sustainable paddy production planning
- ❖ This approach can be applied globally to paddy and other staple crops, supporting data-driven decisions for food security

FUTURE WORK / REFERENCES

- Expand the framework to other staple crops or regions worldwide by integrating real-time satellite data for near-real-time yield predictions
- Develop a user-friendly decision support platform for farmers and policymakers

Sri Lankan Rice Statistics: Department of Census and Statistics, Sri Lanka (2024)
NASA POWER: Stackhouse et al., 2017, *IEEE Geoscience and Remote Sensing*
XGB & Ensemble Learning: Chen & Guestrin, 2016, *Proceedings of KDD*
CHIRPS 2.0: Funk et al., 2015, *Remote Sensing of Environment*