OGIS Symposium 2025

Evaluating Machine Learning and Neural Networks for Yield Prediction in *Pinus taeda* (L.) Plantations Using Time-series Sentinel-1 and 2 Data

Gunjan Barua[1], David Carter[2], Valerie Thomas[1]

[1]Department of Forest Resources and Environmental Conservation, College of Natural Resources and Environment, Virginia Tech

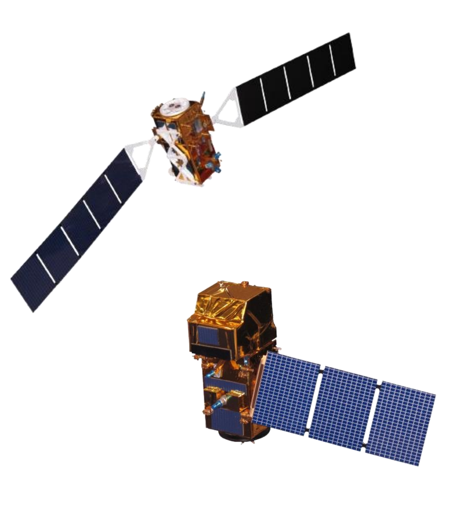
[2]Department of Forestry, College of Agriculture & Natural Resources, Michigan State University

College of Natural Resources and Environment
Forest Resources and Environmental Conservation
Virginia Tech

Abstract

This study compares the performance of advanced machine learning and neural network models for predicting yield volumes (m^3) of Pinus taeda (L.) plantations using satellite remote sensing data. Both approaches employed multitemporal Sentinel-2 multispectral bands (13 bands), a comprehensive suite of vegetation indices (21 indices), and Sentinel-1 SAR bands (VH and VV) with their derived indices (12 indices). Field measurements (258 plots) from 2018 in Texas, Georgia, and Florida provided ground truth data, while monthly time series from Google Earth Engine were aggregated into seasonal means. Five machine learning models (RF, SVR, GBM, XGBoost, LightGBM) and two neural network models (LSTM, GRU) were developed using standardized inputs. Bayesian hyperparameter optimization with 10-fold cross-validation was employed, and an 80-20 train-test split was used to rigorously assess performance using MAE, MSE, RMSE, nRMSE, and r-squared metrics.

Background



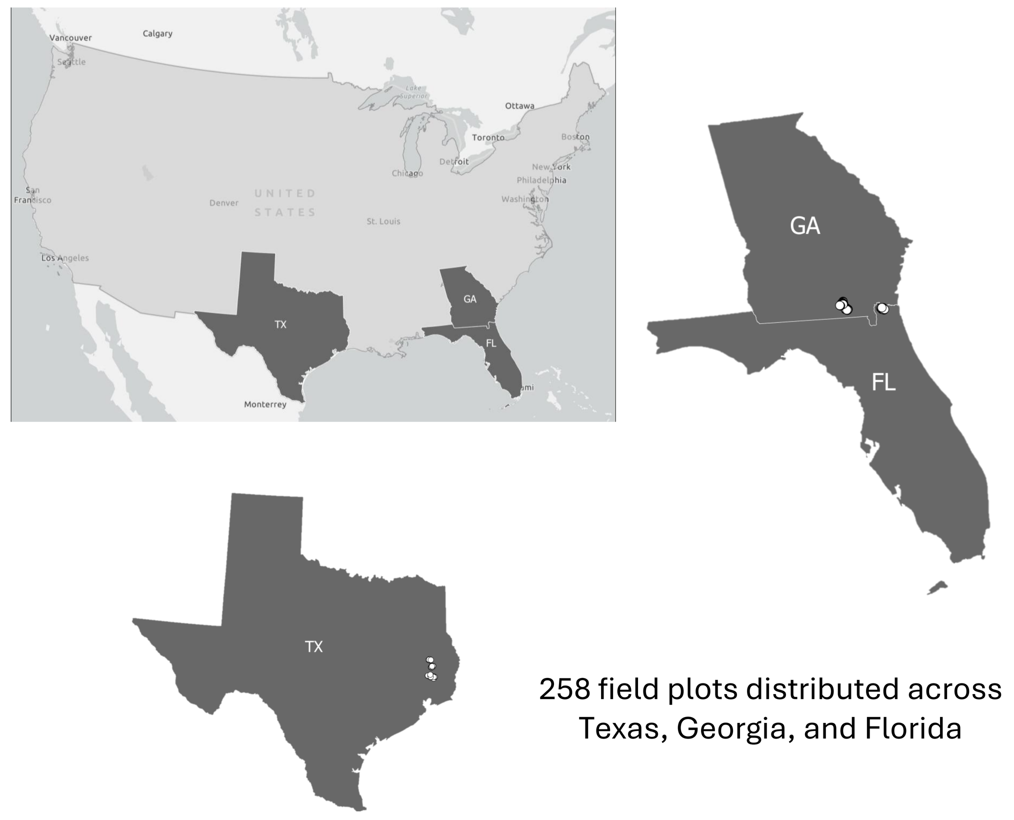
Sentinel 1 and 2 satellites

- **Importance:** Accurate yield prediction for key species (e.g., Loblolly pine) is crucial for sustainable forestry and economics.
- **Methodology:** Utilizes cost-effective, large-scale satellite remote sensing (Sentinel-1/2 data) with advanced machine learning/neural network models.
- **Goal:** Compare model performance to identify the best approach for improving yield prediction accuracy using remote sensing.

Key Findings

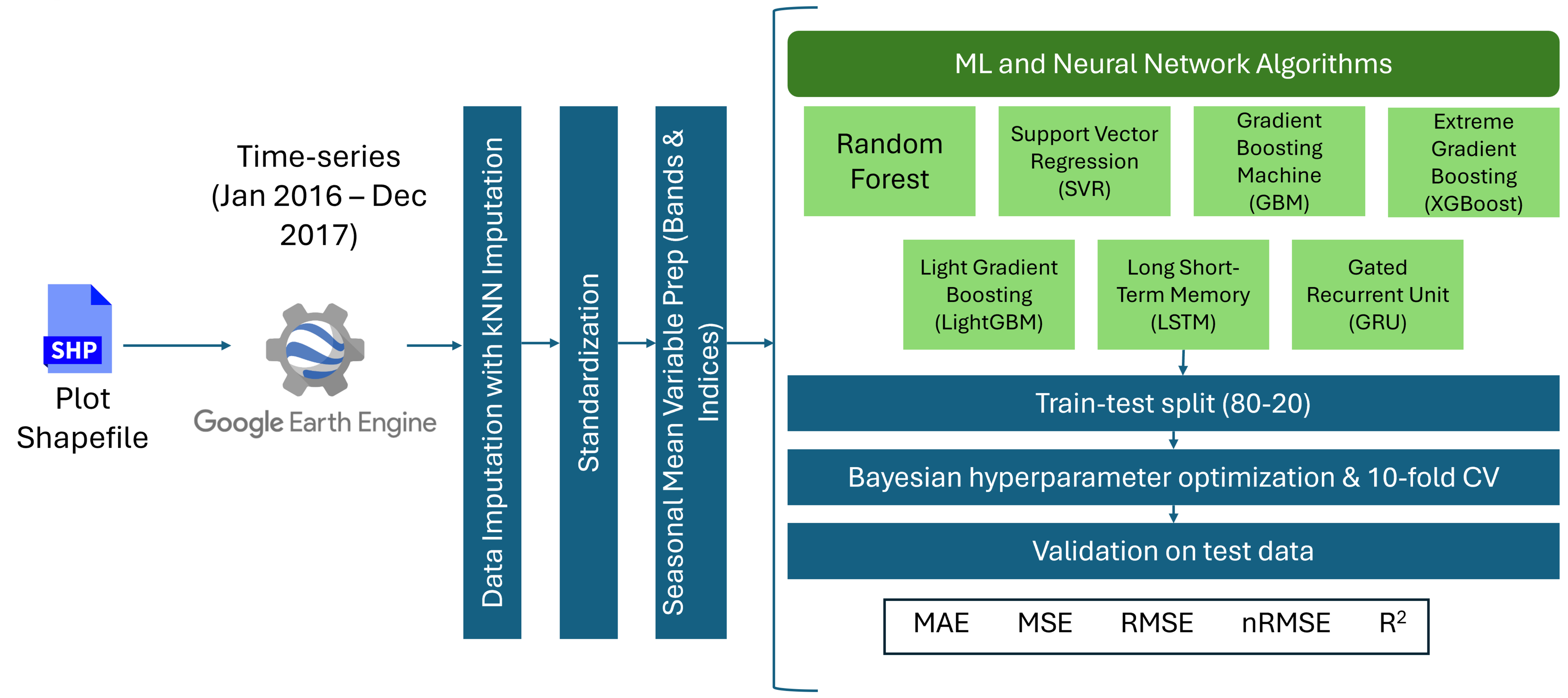
- **Model Performance:** Random Forest (RF) achieved the highest accuracy in predicting plot-level Loblolly Pine yield using time-series spectral data, followed by XGBoost and LSTM.
- **Impact of Planting Density:** Prediction errors generally decreased as planting density decreased. This effect was statistically significant for the superior RF model ($F_{4,64}$ = 16.64, p < 0.001).
- **Key Predictors (RF Model):** Shortwave Infrared 1 (SWIR 1 - Band 11) and Water Vapor (Band 9) were identified as the most influential features, though further analysis is recommended.
- **Areas for Further Research:** The specific roles of key bands and the impact of thinning status on yield prediction accuracy warrant future investigation.

Study Sites



258 field plots distributed across Texas, Georgia, and Florida

Methodology



Remote Sensing Bands and Indices

Sentinel 2

Level-1C (Top of Atmosphere)
13 bands: (B1 to B12)

NDVI (Normalized Difference Vegetation Index)	(NIR -Red)/(NIR+Red)
EVI (Enhanced Vegetation Index)	2.5 ∗(NIR -Red)/(NIR+6 ×Red -7.5 ×Blue+1)
SAVI (Soil-adjusted Vegetation Index)	(NIR-Red)/(NIR+Red+0.5)×1+0.5
MSAVI2 (Modified Soil-adjusted Vegetation Index 2)	$0.5 \times (2 \times NIR + 1) - \sqrt{(2 \times NIR + 1)^2 - 8 \times (NIR - Red}$
OSAVI (Optimized Soil-Adjusted Vegetation Index)	(NIR-Red)/(NIR+Red+0.16)
NDWI_McFeeters (Normalized Difference Water Index (McFeeters version))	(Green-NIR)/(Green+NIR)
MSI (Moisture Stress Index)	SWIR1/NIR
NDRE (Normalized Difference Red Edge Index)	(NIR -RedEdge1)/(NIR+RedEdge1)
SIPI3 (Structure Insensitive Pigment Index)	(NIR -Blue)/(NIR-Red)
SR (Simple Ratio Index)	NIR/Red
DVI (Difference Vegetation Index)	NIR-Red
REIP (Red Edge Inflection Point)	700+40×(((Red+RedEdge3)/2-RedEdge1)/(RedEdge2-RedEdge1))

LCI (Leaf Chlorophyll Index)	(NIR-RedEdge1)/(NIR+Red)
NDII (Normalized Difference Infrared Index)	(NIR -SWIR1)/(NIR+SWIR1)
NDLI (Normalized Difference Lignin Index)	$\frac{\log(SWIR1) - \log(SWIR2)}{\log(SWIR1) + \log(SWIR2)}$
NMDI (Normalized Multi-band Drought Index)	(NIR -(SWIR1-SWIR2))/(NIR+(SWIR1-SWIR2))
GNDVI (Green Normalized Difference Vegetation Index)	(NIR -Green)/(NIR+Green)
CVI (Chlorophyll vegetation index)	$9 \times \frac{RedEdge1}{Green^2}$
GLI (Green Leaf Index)	(2Green-Red-Blue)/(2Green+Red+Blue)
TC_Brightness (Tasselled Cap – Brightness)	0.3037×Blue+0.2793×Green+0.4743×Red+0.5585×NIR+0.5082×Cirrus+0.1863×SWIR2
TC_GVI (Tasselled Cap Green Vegetation Index)	-0.283×Green-0.660×Red+0.577×RedEdge2+0.388×Water Vapor

Sentinel 1

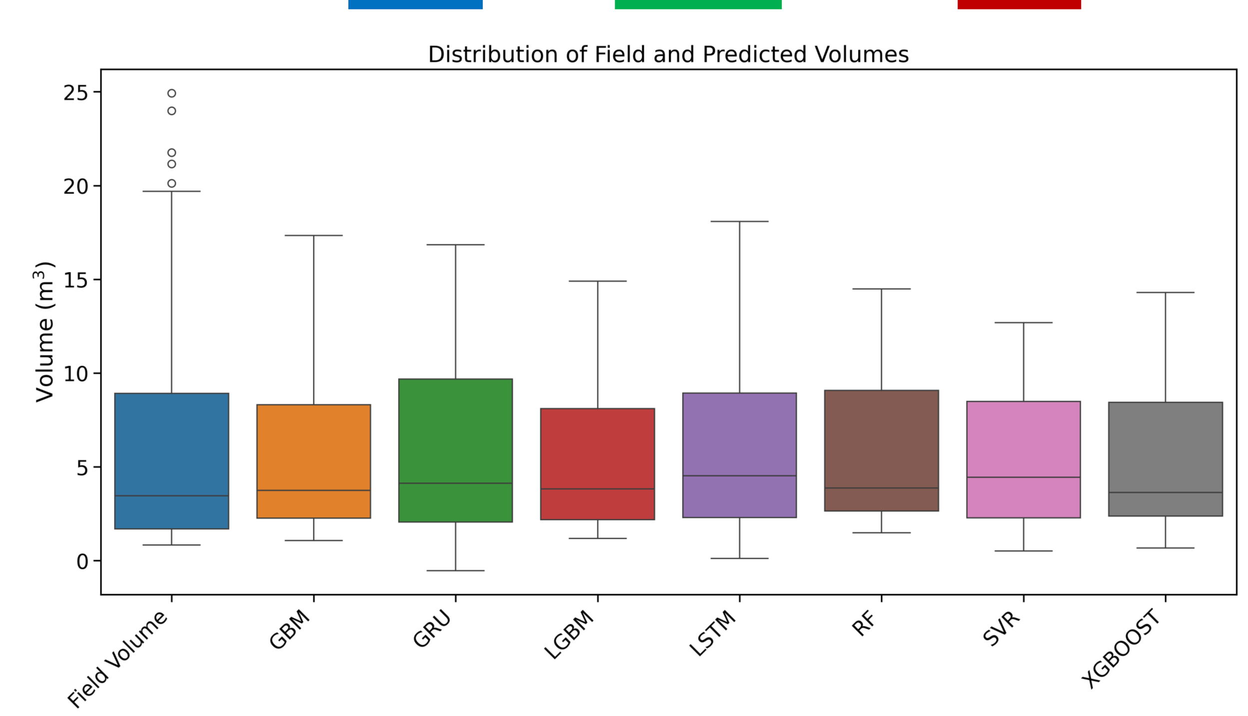
C-band SAR data
2 bands: VV and VH

VV_VH_Ratio	VV/VH
VH_VV_Ratio	VH/VV
VV_VH_Difference	VV - VH
VH_VV_Difference	VH -VV
Normalized_VVVH	(VV - VH)/(VV + VH)
Sum_VVVH	VV + VH
Product_VVVH	VV * VH
DPDD (Dual-polarization difference descriptor)	(VV + VH) / √2
Gamma_nought_VH	VH/cos(angle * (π)/180)
Gamma_nought_VV	VV/cos(angle * (π)/180)
RVI (Radar Vegetation Index)	(4 * VH)/ (VV + VH)
VDDPI (Volume density dual-polarization index)	(VV * VH) / VV

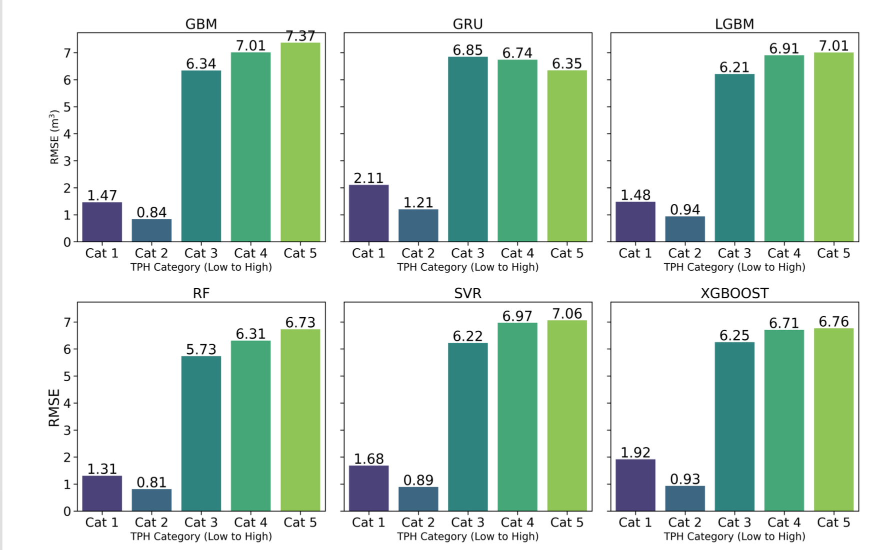
Results

		1st		2nd		3rd	
Metrics	SVR	RF	GBM	XGBoost	LightGBM	LSTM	GRU
MAE (m^3)	3.02	2.70	2.97	2.94	2.99	3.18	3.30
MSE (m^3)	27.13	23.20	28.23	25.92	26.67	26.35	26.79
RMSE (m^3)	5.21	4.82	5.31	5.09	5.16	5.13	5.17
nRMSE (%)	21.62	19.99	22.05	21.13	21.43	21.31	21.48
R^2	0.40	0.49	0.38	0.43	0.41	0.42	0.41

Distribution of Field and Predicted Volumes

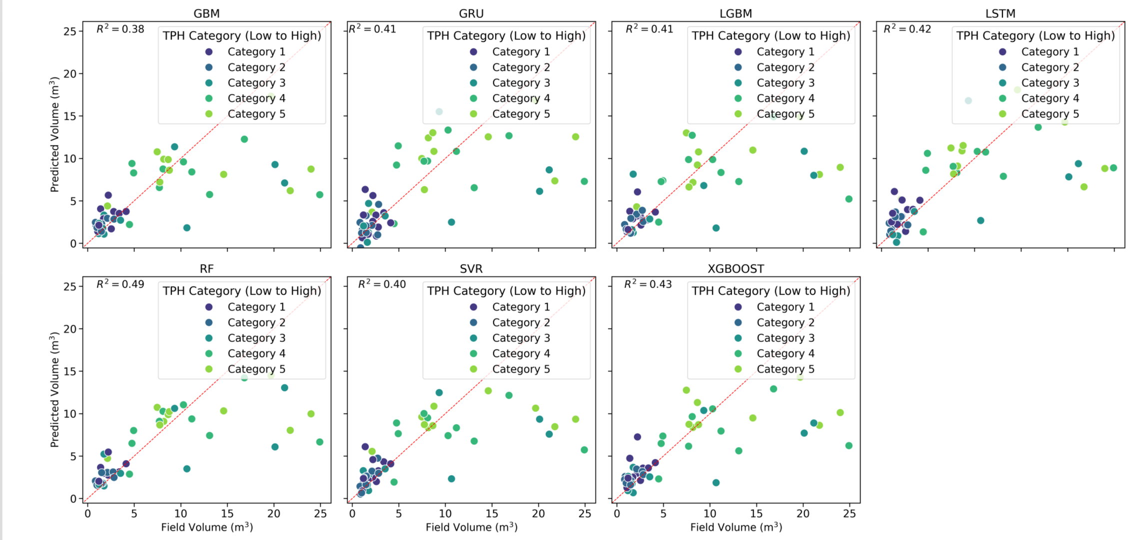


Results (Contd.)

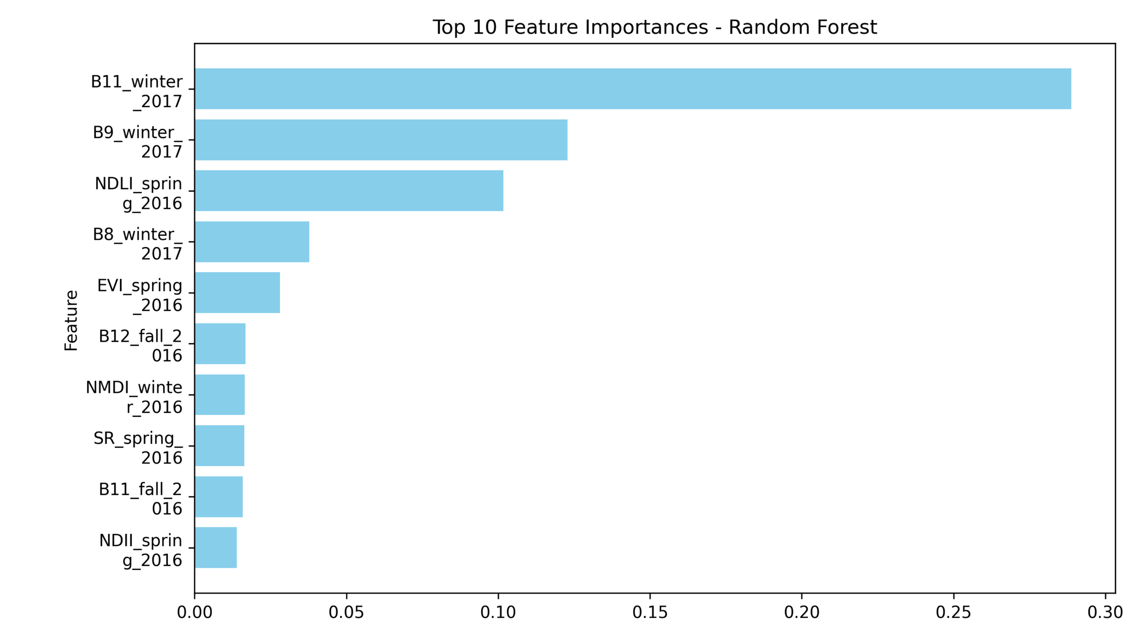


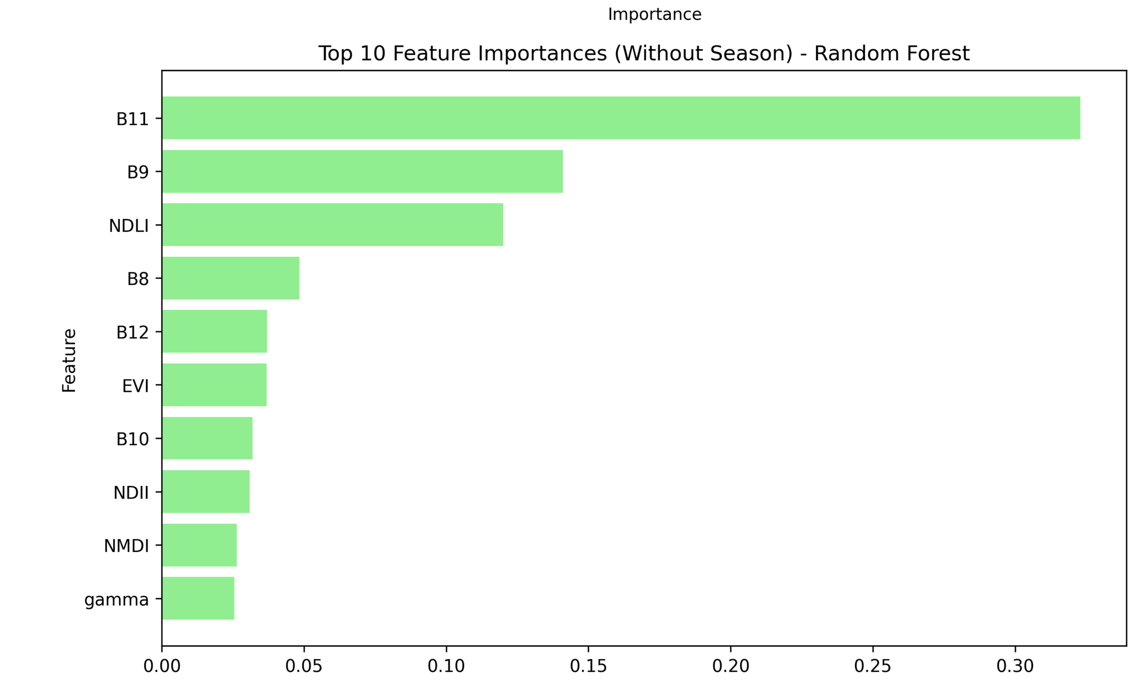
Category 1: 12.56 - 100.46
Category 2: 100.46 - 138.16
Category 3: 150.71 - 489.78
Category 4: 502.36 - 1193.13
Category 5: 1218.28 - 1682.99

RMSE of 5 different planting densities across all models



Scatterplot of all models with variation in 5 different planting densities





The **Random Forest** model best predicts plot-level yield for *Pinus taeda*, followed by **XGBoost** and **LSTM.** The yield results for different planting densities in the Random Forest model are significantly different *($F_{4,64}$ = 16.64, p < 0.001)*

Acknowledgment

I am grateful to Forest Research Cooperative (FPC) for providing the field-level data for this research.

References