Predicting Yields of *Pinus taeda* (L.) with Open-Source Time-Series Sentinel 1 and 2 Data with Machine Learning and Neural Networks

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

VIRGINIA TECH.

Presenting:
Gunjan Barua
gunjanb@vt.edu

FOREST PRODUCTIVITY COOPERATIVE

North Carolina State University • Virginia Tech • Universidad de Concepción • Universidade Federal de Santa Catarina

Scan to find this poster on the FPC website:

https://www.forestproductivity.org/event-presentations

Abstract

Continuous monitoring of plantation inventories through traditional field-based methods or active remote sensing (e.g., LiDAR, UAV) can be resource-intensive. As a viable alternative, this study utilizes open-source Sentinel-1 and Sentinel-2 satellite data for continuous growth monitoring to predict yield volumes (m^3) in *Pinus taeda* (L.) plantations. We compared the performance of five machine learning (ML) models (RF, SVR, GBM, XGBoost, LightGBM) and two neural network (NN) architectures (LSTM, GRU). Both approaches were trained on a robust set of inputs, including multitemporal Sentinel-2 multispectral bands (13), numerous vegetation indices (21), and Sentinel-1 SAR bands (VH and VV) with derived indices (12). Monthly time series data from Google Earth Engine were aggregated into seasonal means and calibrated using ground truth data from 258 field plots measured in 2018 across Texas, Georgia, and Florida.

Background

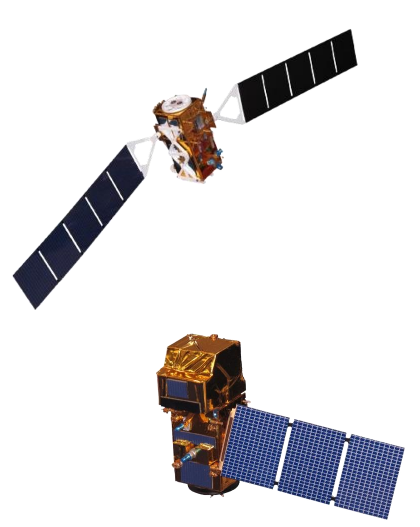

Figure 1: Sentinel 1 and 2 satellites

- **Importance:** Developing cost-effective and scalable methods for accurately predicting yield in Loblolly pine is essential for sustainable forestry and economic planning, addressing the limitations of resource-intensive monitoring.
- **Methodology:** Utilizes open-source, large-scale satellite data (Sentinel-1/2) with advanced ML/ NN models.
- **Goal:** The goal of this study is to enhance yield prediction accuracy by comparing the performance of multiple ML and NN models and identifying the key remote sensing variables that serve as the most effective predictors.

Key Findings

- **Model Performance:** Gated Recurrent Unit (GRU) achieved the highest accuracy in predicting plot-level Loblolly Pine yield using time-series spectral data, followed by RF and LSTM.
- **Impact of Planting Density:** Prediction errors generally decreased as planting density decreased. This effect was statistically significant for all the models.
- **Key Predictors (GRU Model):** Terrain corrected VH 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

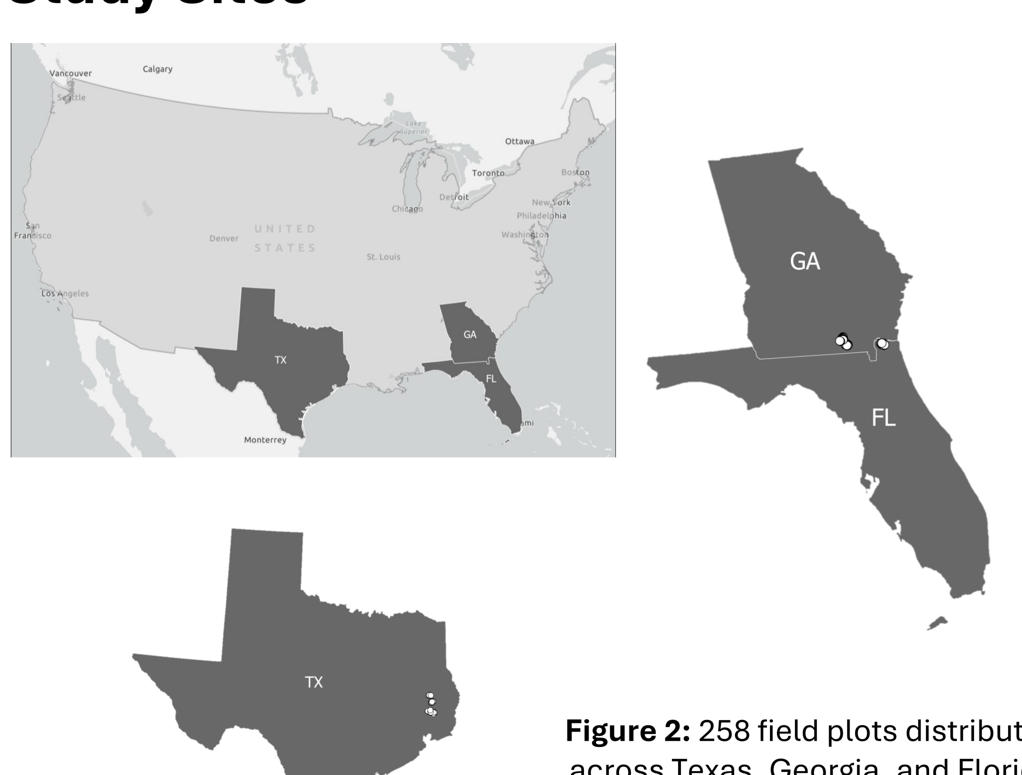

Figure 2: 258 field plots distributed across Texas, Georgia, and Florida

Methodology

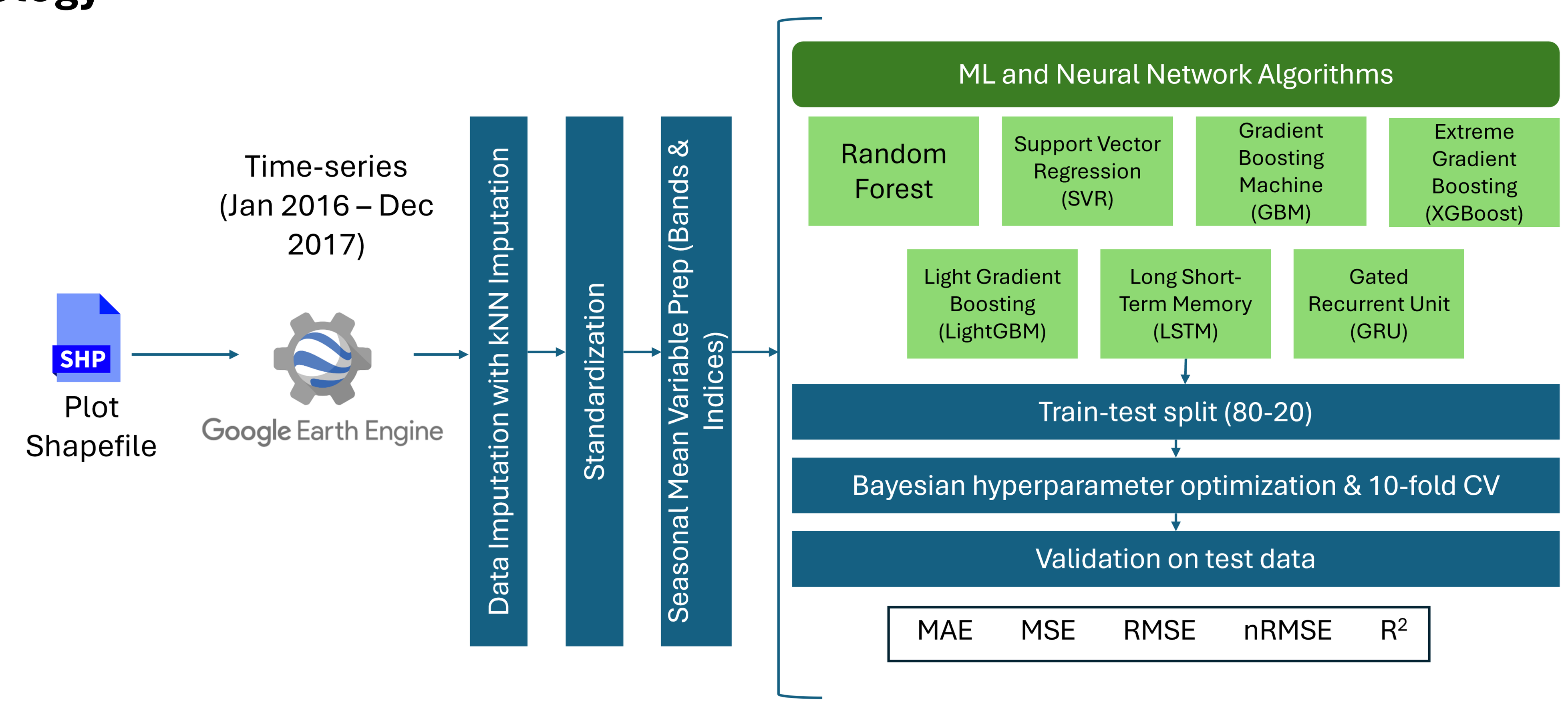

Figure 3: Methodological flow of the modeling process

Remote Sensing Bands and Indices

Sentinel 2
Level-1C (Top of Atmosphere)
13 bands: (B1 to B12)
21 Vegetation Indices

Sentinel 1
C-band SAR data
2 bands: VV and VH
13 SAR Indices

Results

Metrics	GRU (1st)	RF (2nd)	LSTM (3rd)	XG Boost	SVR	Light GBM	GBM
MAE (m^3ha^{-1})	34.29	33.91	35.04	36.92	37.93	37.55	37.30
MSE (m^3ha^{-1})	290.24	291.37	292.75	325.53	340.72	334.95	354.54
RMSE (m^3ha^{-1})	60.41	60.53	60.66	63.92	65.43	64.80	66.69
nRMSE (%)	19.95	19.99	20.04	21.13	21.62	21.43	22.05
R^2	0.49	0.488	0.486	0.43	0.40	0.41	0.38

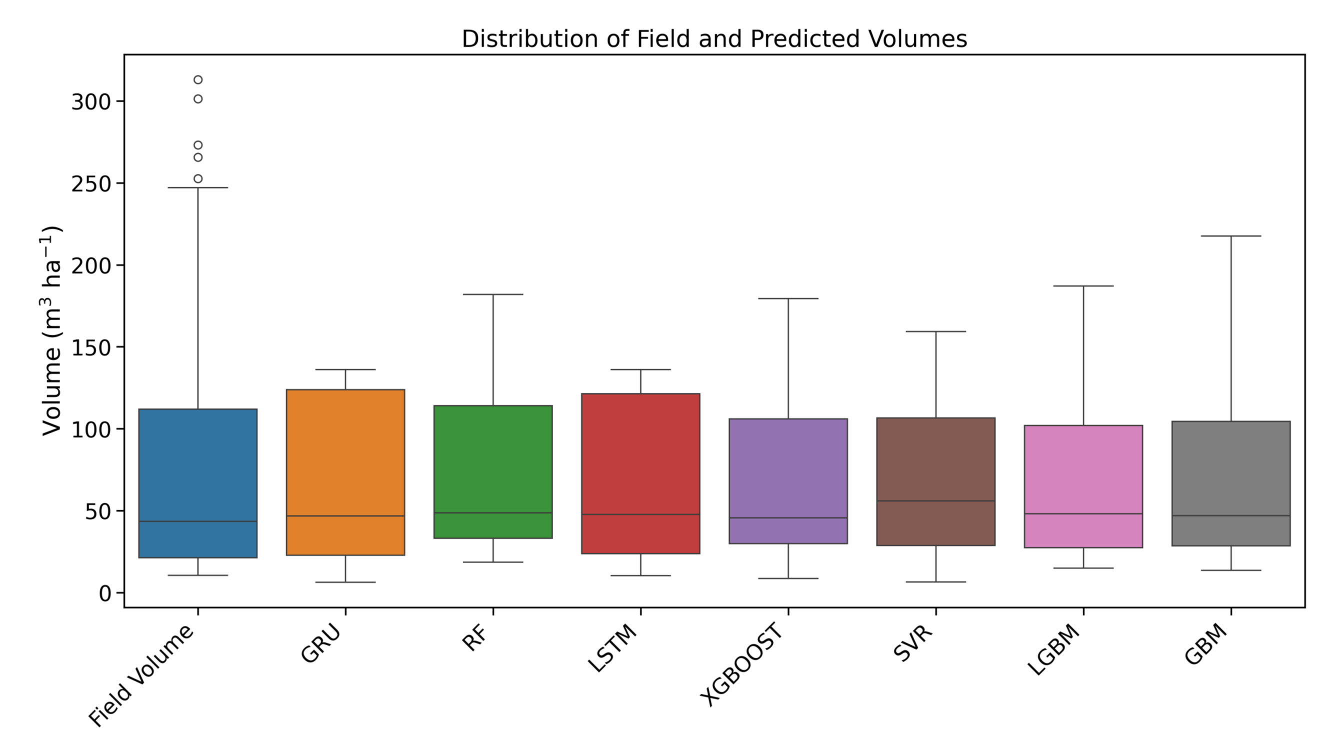

Figure 4: Field and predicted plot volume per hectare distribution

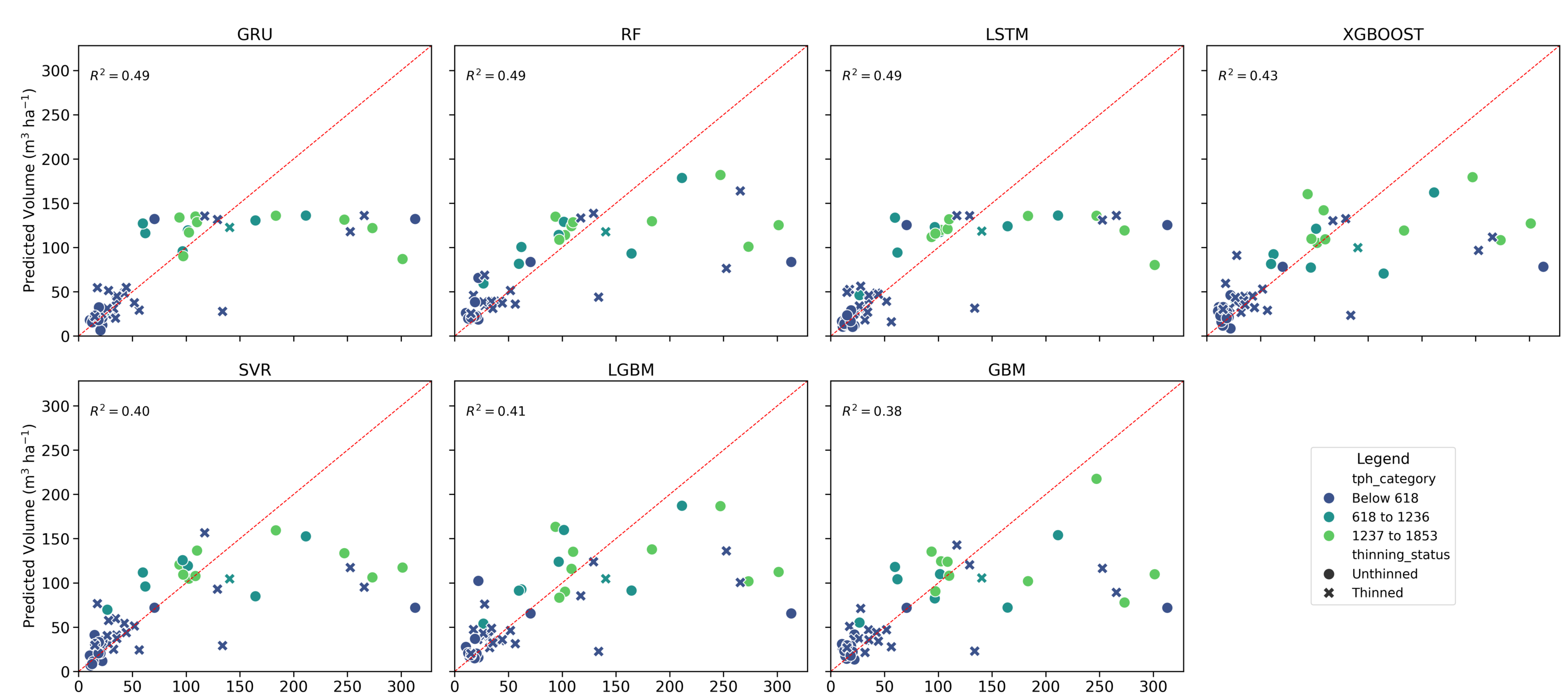

Figure 5: Scatter plot for field vs predicted volumes per hectare for all models showing planting densities and thinning conditions. Thinned and lower TPH plots seem to be close to the 1:1 line

Results (Contd.)

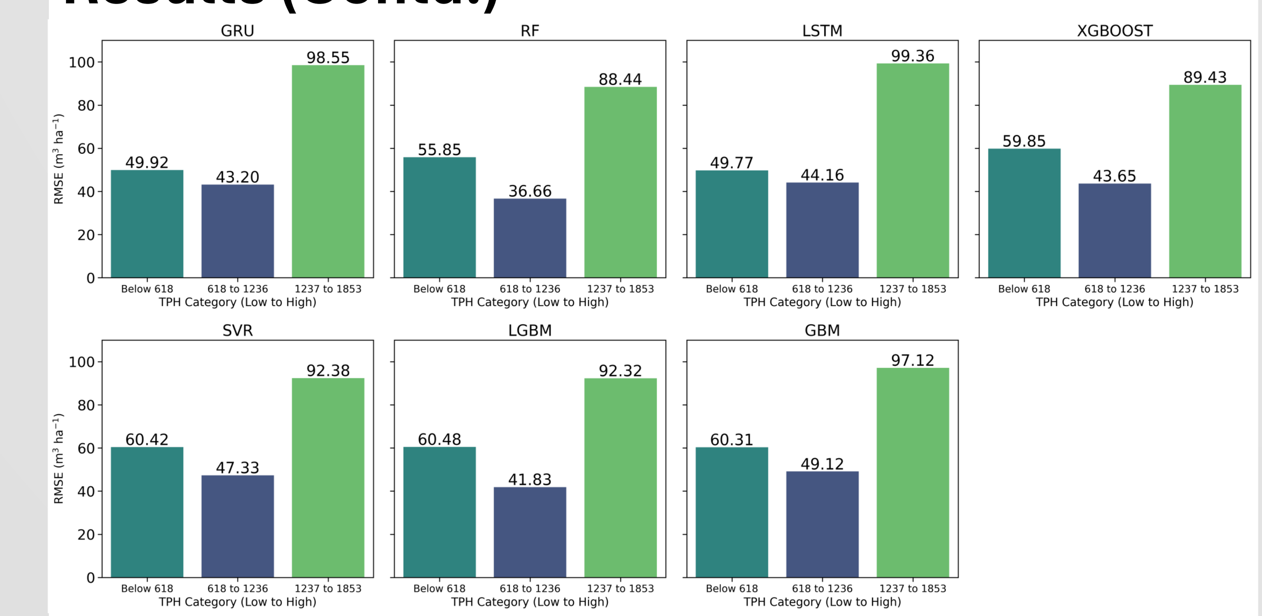

Figure 6: RMSE results across models for 3 different planting densities, showing that TPH category 618 to 1236 has the lowest RMSE

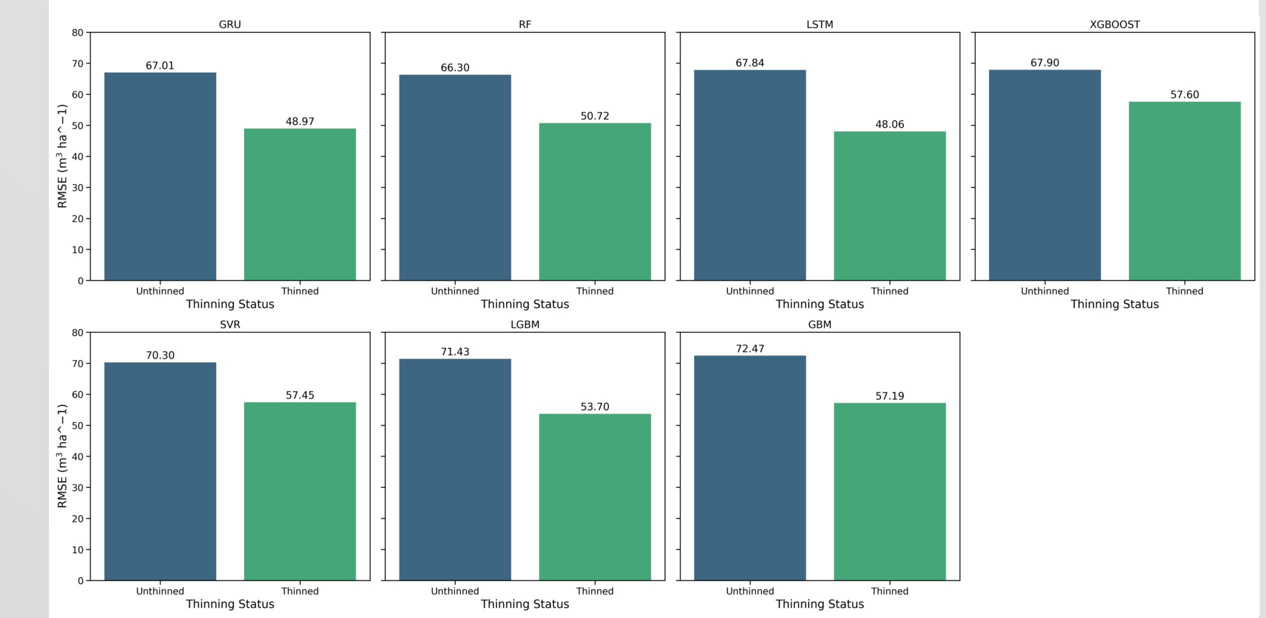

Figure 7: RMSE results across models for different thinning conditions showing thinned plots have lower RMSE

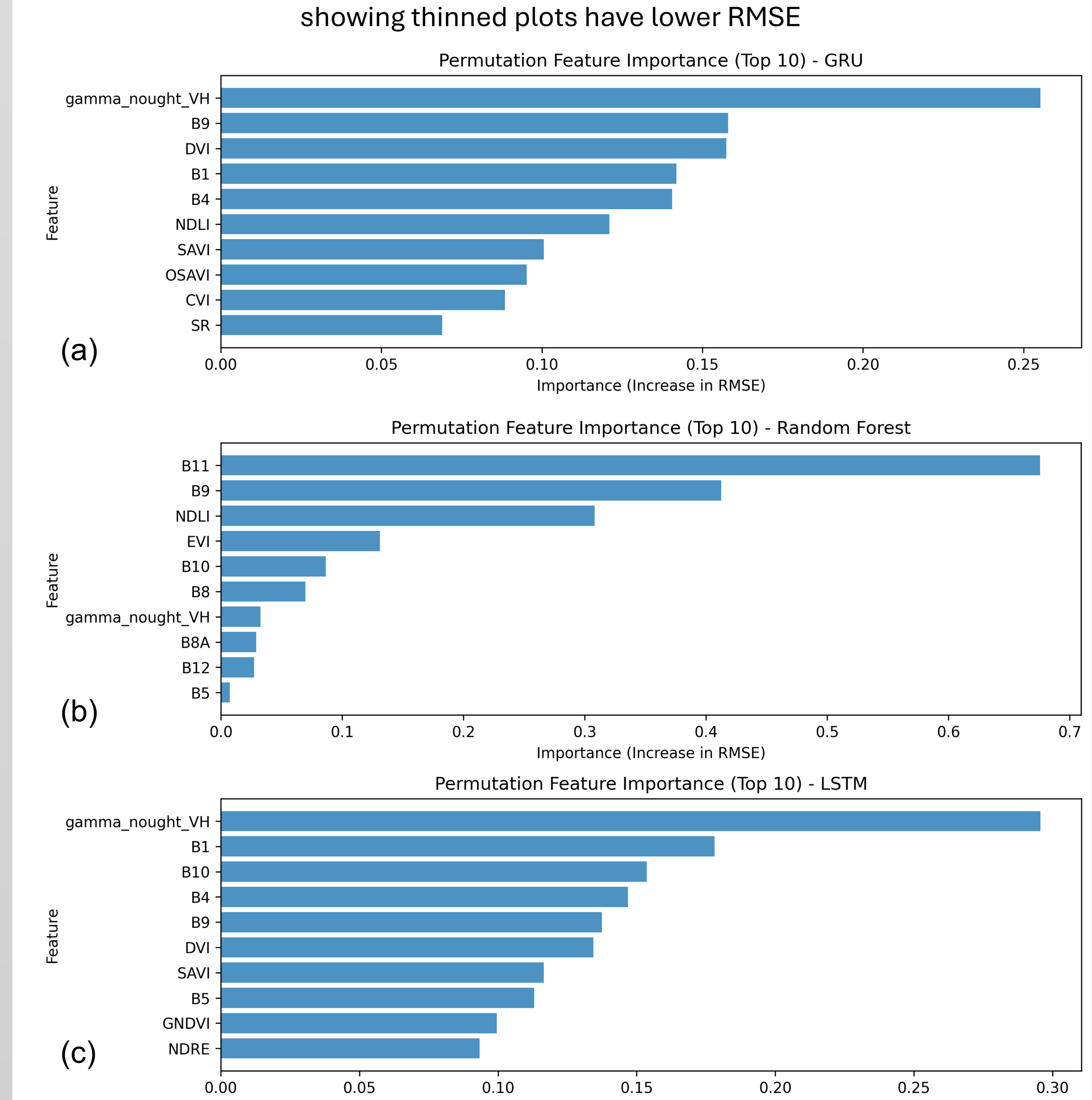

Figure 8: Permutation feature importance for top 3 models, i.e., (a) GRU, (b) RF, and (c) LSTM

Acknowledgment

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

References