# TARGET-ORIENTED VALIDATION STRATEGIES REVEALS SLIGHT INCREASE IN PREDICTIVE PERFORMANCE OF LAND USE AND LAND COVER CLASSIFICATION USING LIDAR DATA OVER NAIP DATA


mukti.subedi@ttu.edu
suvedimukti@gmail.com

https://scholar.google.com/citations?user=8hXIpL0AAAAJ&hl=en

GitHub

https://github.com/suvedimukti

Mukti Subedi
PhD Candidate
Geospatial Technologies lab
Natural Resources Management
Texas Tech University, Lubbock

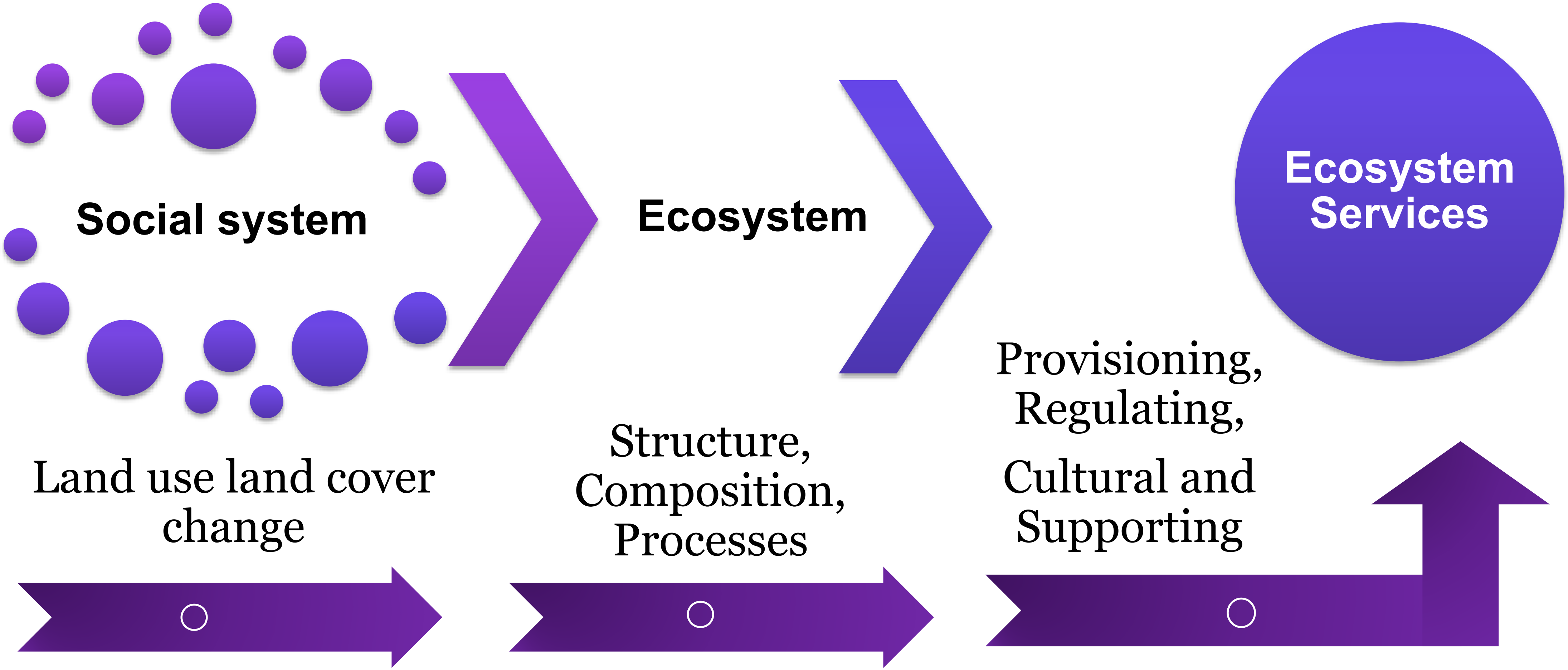
TEXAS TECH UNIVERSITY

# Land Use /Land Cover (LULC)



Land cover is an essential climate variable (Hollmann et al., 2013)

USGS Plans to acquire 3-DEP data across the US by the end of 2023



TEXAS TECH UNIVERSITY

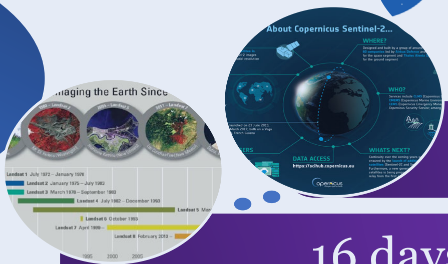
# Accurate LULC Mapping is Important



RS Data

I) Velocity

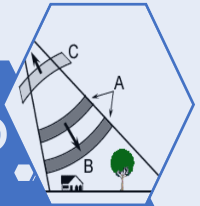
Daily (MODIS)



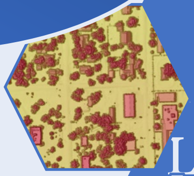
5-days (Sentinel)

16 days (Landsat)

II) Variety



RADAR



LiDAR

UAV

III) Volume

Very High Spatial, Temporal, Spectral

Low Spatial, Temporal, Spectral

IV) Veracity

Quality

Reliability

Uncertainty

TEXAS TECH UNIVERSITY

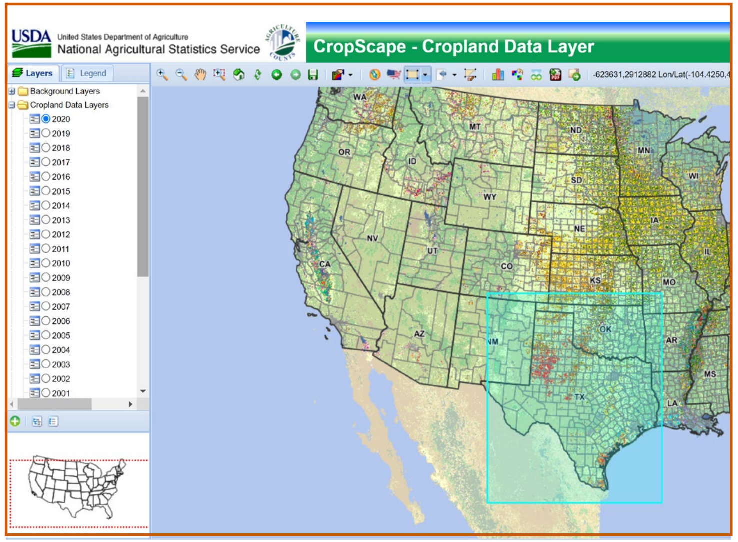


# LULC Products in Texas


Multi-temporal LULC change database from 2001 - 2019 at every 2–3-year intervals.

Machine learning based land cover classifications.

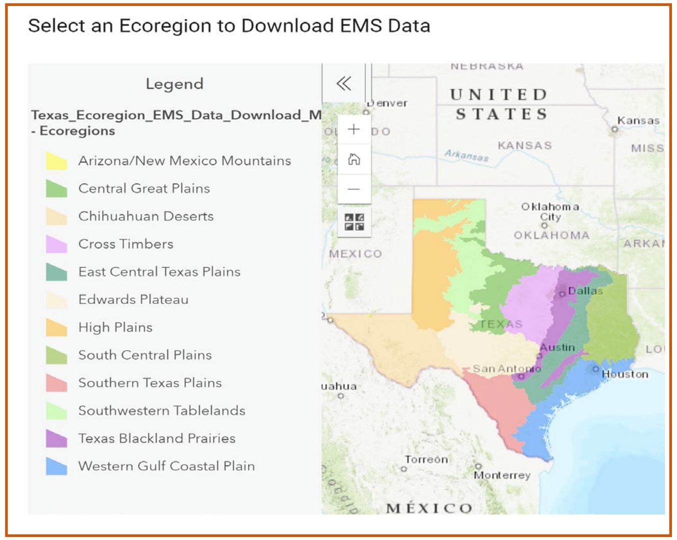
**Availability: FREE**




Cropland Data Layer (CDL). Annual resolution.

Based on moderate resolution satellite imagery and extensive agricultural ground truth.

**Availability: FREE**




Ecological Mapping System[TPWD].

Based on NAIP objects (10 m) and Expert rules.

**Availability: FREE**

TEXAS TECH UNIVERSITY



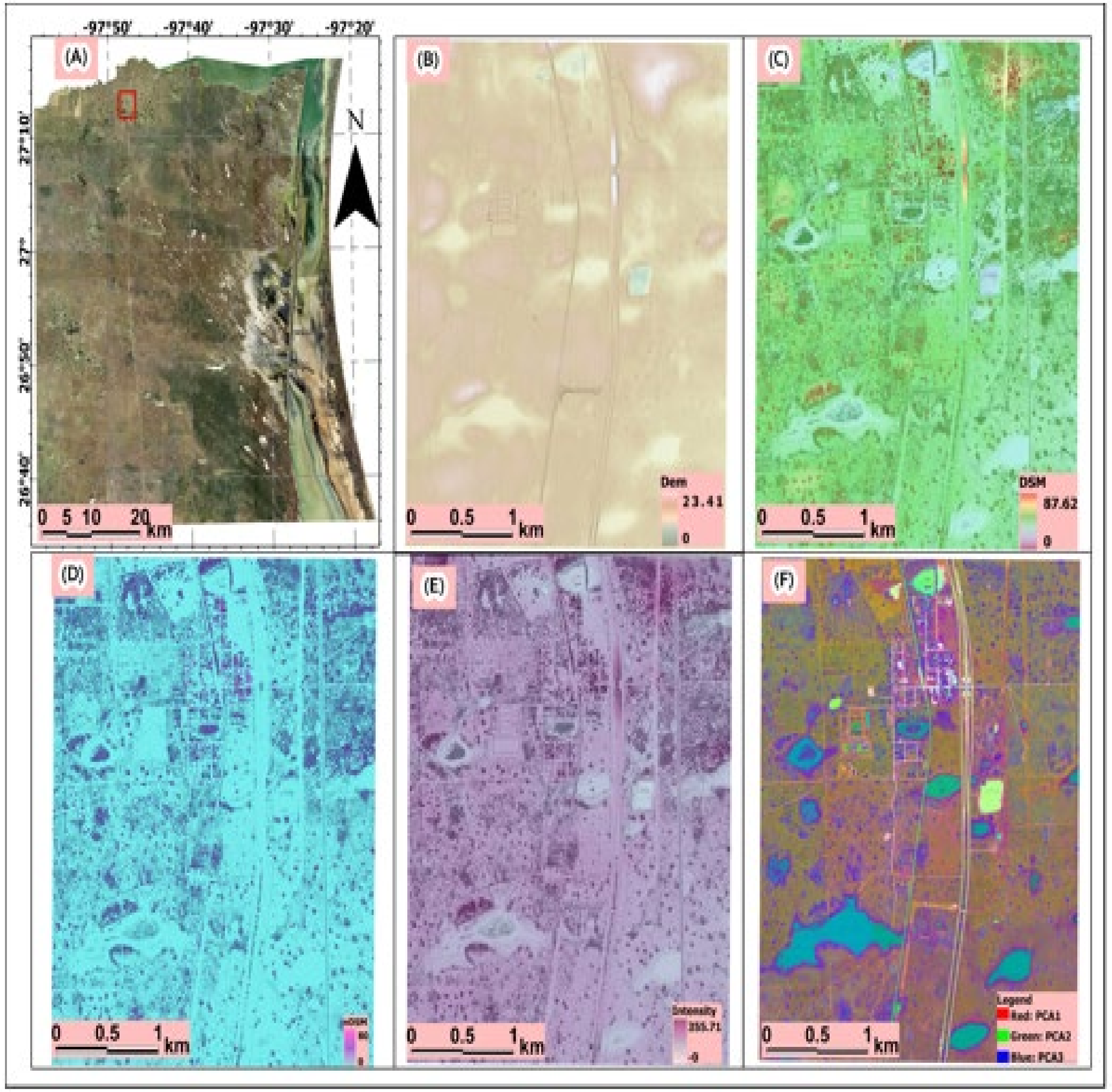
# Objectives

- ❑ To evaluate the classification performance of SVM and RF on NAIP only data and LiDAR products using random sampling strategies.
- ❑ To test if accounting for spatial dependency reduces the over-optimistic performance.
- ❑ To identify suitable classification method(s) for improved LULC classification, emphasizing the characterization of small-scale grassland patches that are deemed suitable for applications in ecology.

TEXAS TECH UNIVERSITY

# Study Area and Data

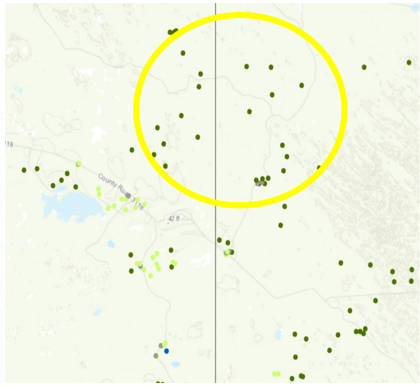
- A. Background
- B. LULC Products
- C. Objective
- D. **Study Area**
- E. Data Processing Workflow
- F. Results: Accuracy
- G. Results: Confusion Matrix
- H. Results: Map Comparison
- I. Results: Validation strategies
- J. Results: Variable Importance
- K. Key Points
- L. Acknowledgements
- M. References




DOQQ : 131 (3.75′× *3.75′)

LiDAR footprints : 2199

Total Samples:- 3719 Classes:- 8



TEXAS TECH UNIVERSITY



# Data Processing Workflow

Pre-Processing → Segmentation → Classification

| Source | Data type | Statistic type | Features | (Count) |
|---|---|---|---|---|
| NAIP | Spectral | Difference | MAXDI | (1) |
| | | Mean | MNRED, MGREN, MBLUE, MNRIR | (4) |
| | | Standard deviation | SNRED, SGREN, SBLUE, SNRIR, | (4) |
| | Indices | Mean | MNDSI, MNDWI, MNDVI, MSAVI | (4) |
| | | Standard deviation | SNDSI, SNDWI, SNDVI, SSAVI | (4) |
| | PCA axis | Mean | MPCA1, MPCA2, MPCA3 | (3) |
| | | Standard deviation | SPCA1, SPCA2, SPCA3 | (3) |
| | GLCM Textures | Mean | MTHOM, MTDIS, MTASM, MTENT, MTSTD | (5) |
| | | Standard deviation | STHOM, STDIS, STASM, STENT, STSTD | (5) |
| LiDAR | Elevation | Mean | MEDEM, MEDSM, MENDSM, | (3) |
| | | Standard deviation | SEDEM, SEDSM, SENDSM | (3) |
| | Intensity | Mean | MINST | (1) |
| | | Standard deviation | SINST | (1) |
| Objects | Geometry | Shape | GCOMP, GDENS, GRECT, GROND, GSHAP, GASYM | (6) |

Icon credits: ArcGIS Pro, eCognition, R Core Team , and Rstudio

TEXAS TECH UNIVERSITY

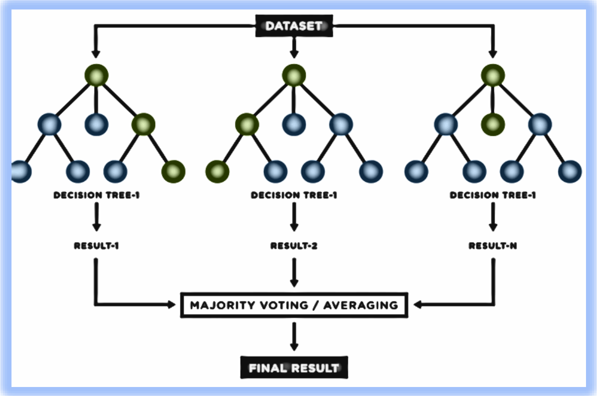


# Data Processing Workflow: contd.

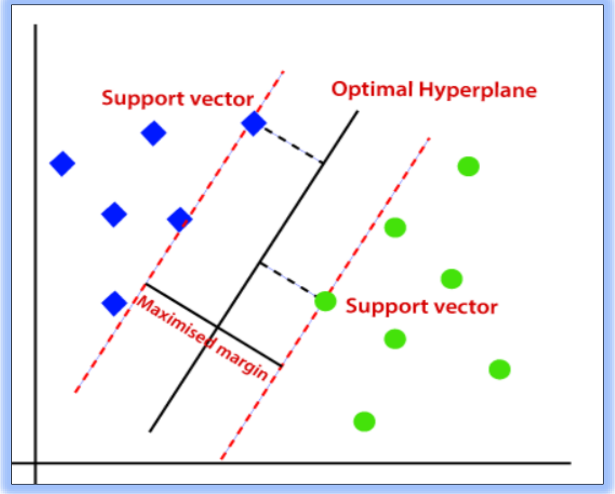
Algorithms → Spatial Dependence Analysis → Validation Strategies

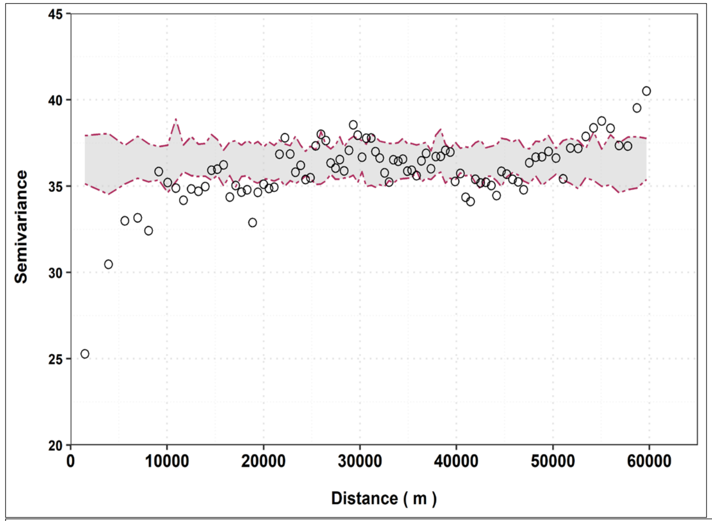
Prediction accuracy does not guarantee spatial accuracy (Mayer et. al, 2018)

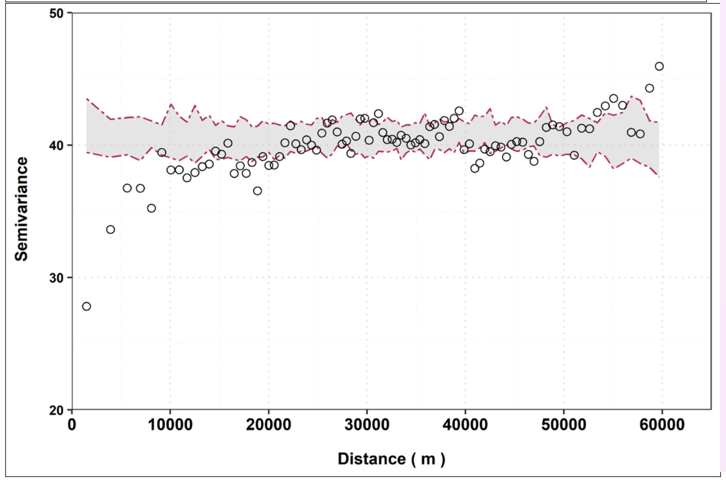
## 1.) Random Forest

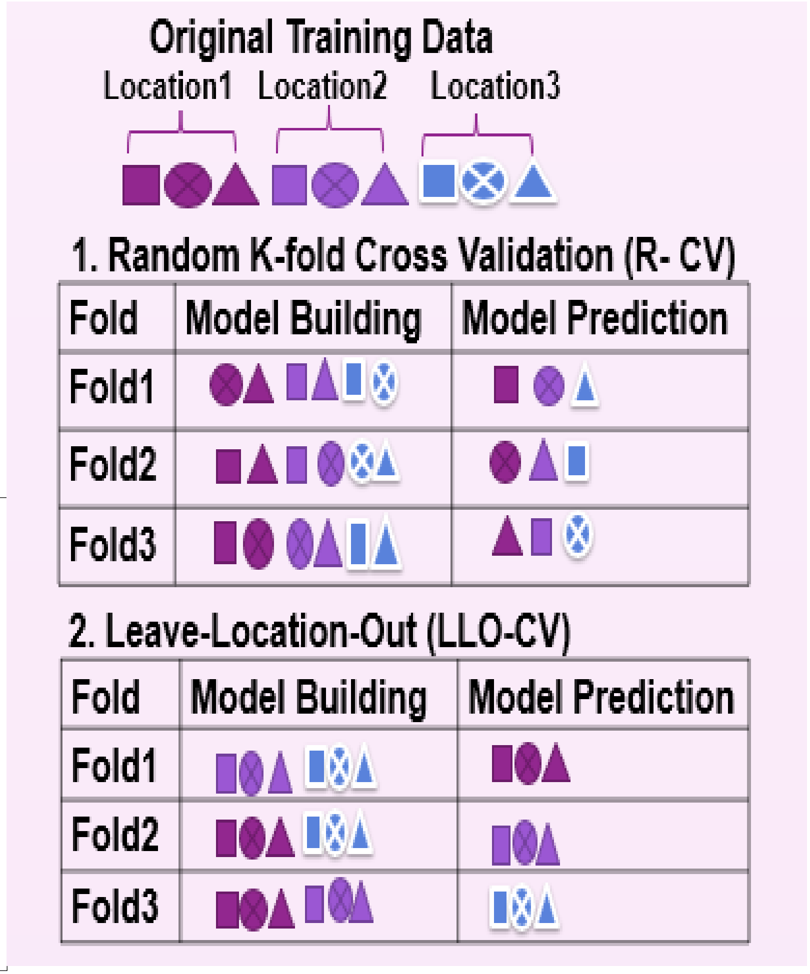



## 2.) Support Vector Machines



TEXAS TECH UNIVERSITY

# Data Processing Workflow: contd.



**STAGE I**

- R-CV
- Random Forest vs SVM

**STAGE II**

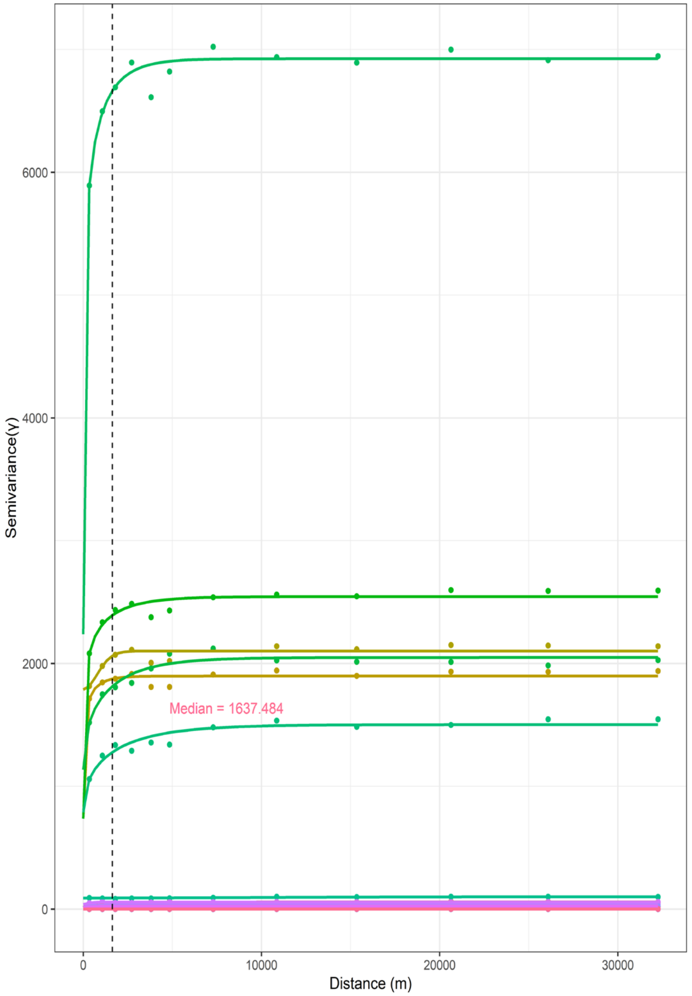
- LLO[Target Oriented Validation Approach]
- Random Forest vs SVM
- Feature Selection Algorithms:
  - Forward Feature Selection: **FFS**
  - Recursive Feature Elimination: **RFE**

**STAGE III**

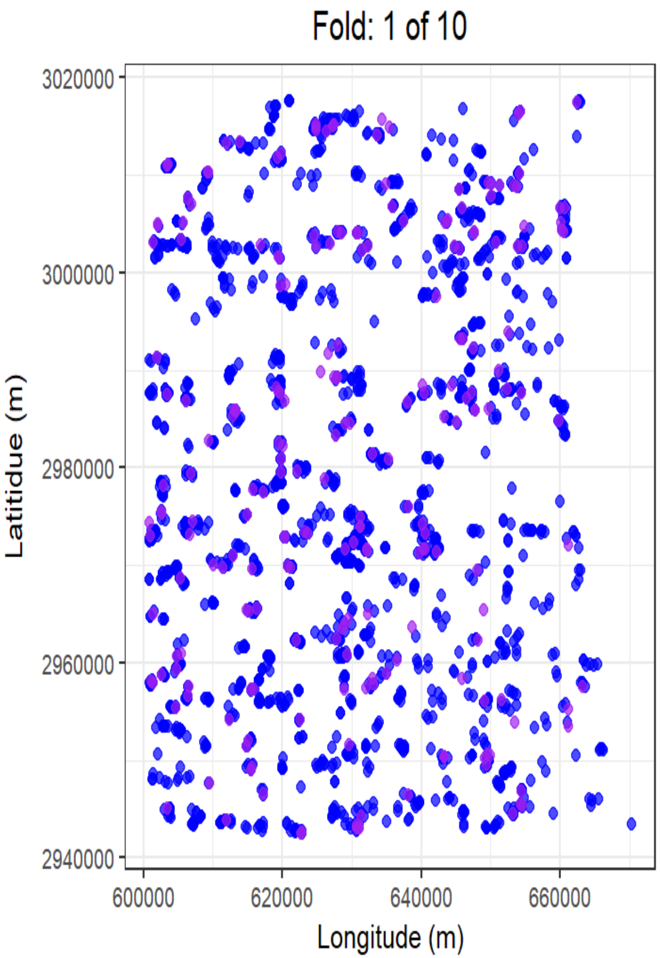
- Compare Algorithms (Confusion Matrices)

# Data Processing Workflow: contd.

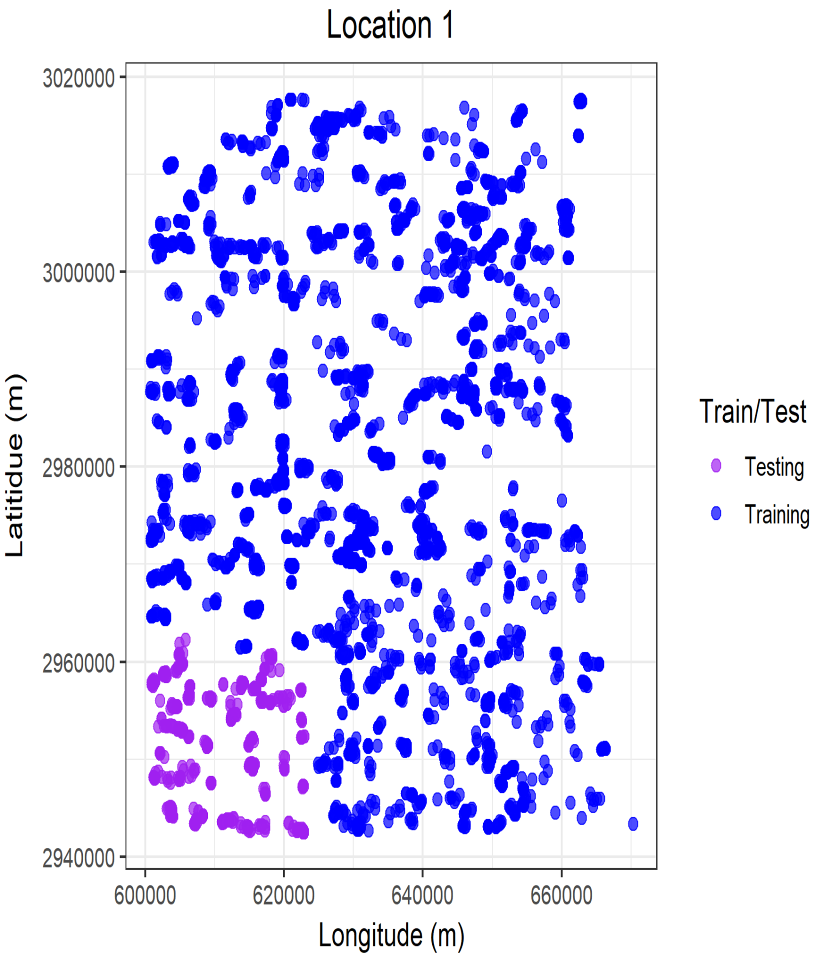
## Semivariance: LiDAR




## Random Cross Validation: (R-CV)



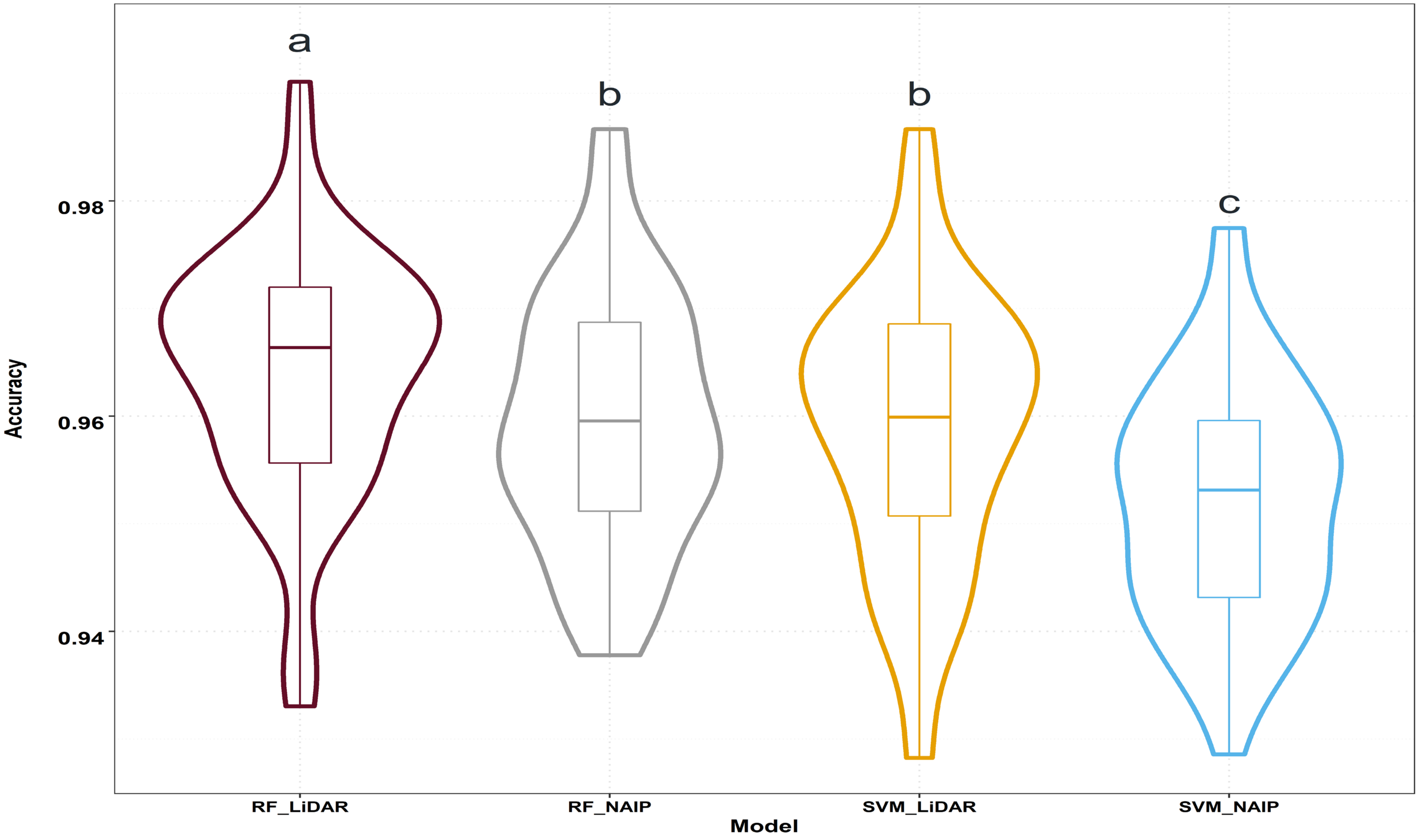

## Leave Location Out: (LLO)



TEXAS TECH UNIVERSITY

# Results: STAGE I [Accuracy (R-CV)]





# Results: STAGE I [Class-Based Accuracy]

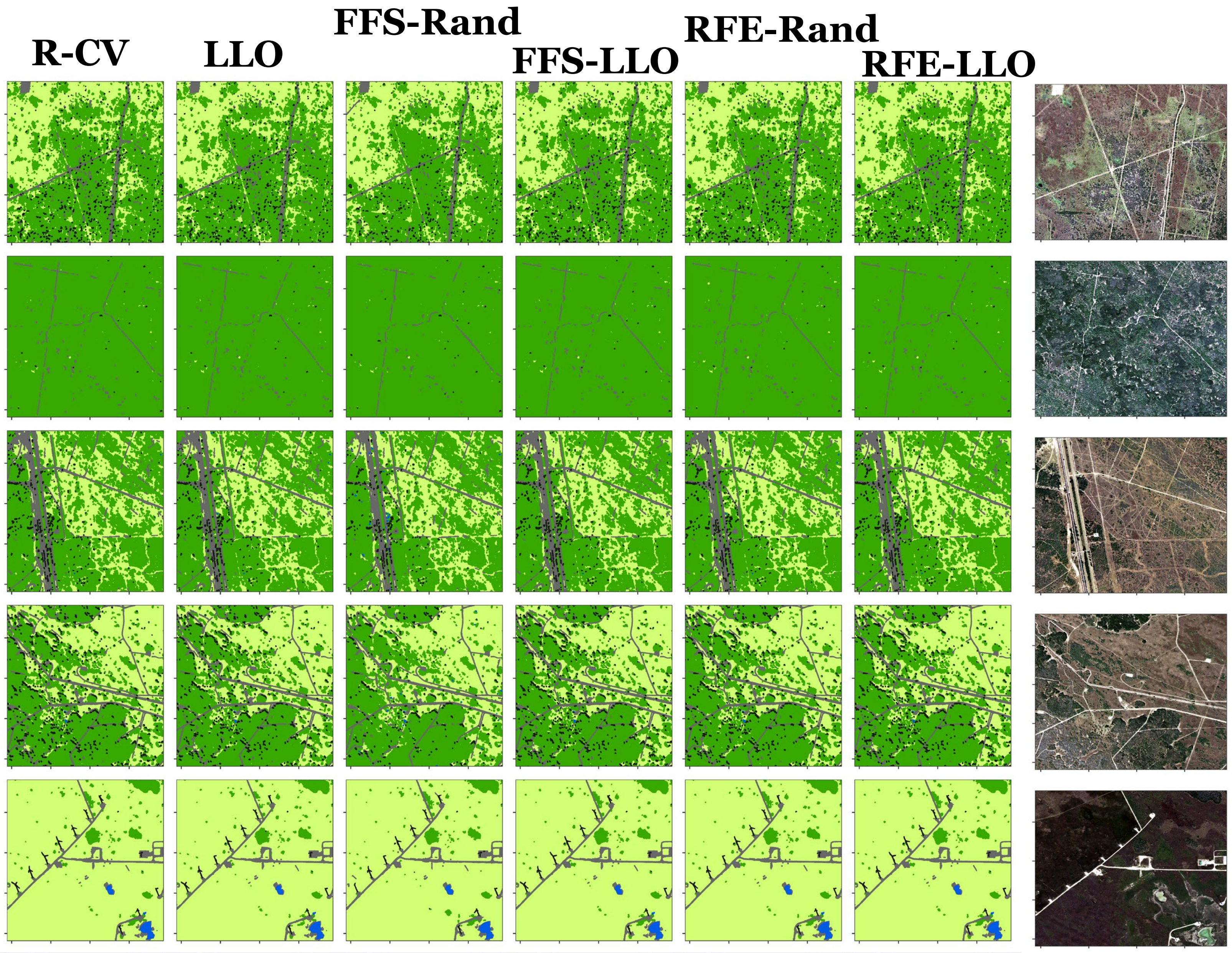


| | | Producer's Accuracy (%) | | | User's Accuracy (%) | | | F1-Score (%) | | |
|---|---|---|---|---|---|---|---|---|---|---|
| **Data** | **Class** | **RF** | **SVM** | **Change (D)** | **RF** | **SVM** | **(D)** | **RF** | **SVM** | **(D)** |
| NAIP | Cropland | 84.38 | 62.50 | -21.88 | 96.43 | 86.96 | -9.47 | 90.00 | 72.73 | -17.27 |
| | Fallow land | 78.26 | 86.96 | 8.70 | 94.74 | 95.24 | 0.50 | 85.71 | 90.91 | 5.20 |
| | Grassland | 98.23 | 98.72 | 0.49 | 95.63 | 95.35 | -0.28 | 96.91 | 97.00 | 0.09 |
| | Shrubland | 95.06 | 95.06 | 0.00 | 96.86 | 97.78 | 0.92 | 95.95 | 96.40 | 0.45 |
| | Built-up1 | 94.27 | 96.95 | 2.68 | 95.37 | 93.73 | -1.64 | 94.82 | 95.31 | 0.49 |
| | Built-up2 | 86.49 | 79.73 | -6.76 | 81.01 | 78.67 | -2.34 | 83.66 | 79.19 | -4.47 |
| | Water | 96.43 | 94.64 | -1.79 | 100.0 | 98.15 | -1.85 | 98.18 | 96.36 | -1.82 |
| | Shadow | 95.00 | 80.00 | -15.00 | 97.44 | 100.00 | 2.56 | 96.20 | 88.89 | -7.31 |
| **Overall accuracy** | | **95.44** | **94.90** | **-0.54** | | | | | | |
| **Kappa statistic** | | **93.63** | **93.09** | **-0.54** | | | | | | |
| NAIP+ LiDAR | Cropland | 81.25 | 71.88 | -9.37 | 96.30 | 95.83 | -0.47 | 88.14 | 82.14 | -6.00 |
| | Fallow land | 78.26 | 82.61 | 4.35 | 94.74 | 95.00 | 0.26 | 85.71 | 88.37 | 2.66 |
| | Grassland | 99.36 | 99.04 | -0.32 | 96.72 | 96.56 | -0.16 | 98.02 | 97.78 | -0.24 |
| | Shrubland | 97.22 | 97.22 | 0.00 | 99.06 | 99.06 | 0.00 | 98.13 | 98.13 | 0.00 |
| | Built-up1 | 98.09 | 96.18 | -1.91 | 96.62 | 95.09 | -1.53 | 97.35 | 95.64 | -1.71 |
| | Built-up2 | 87.84 | 83.78 | -4.06 | 92.86 | 81.58 | -11.28 | 90.28 | 82.67 | -7.61 |
| | Water | 96.43 | 95.54 | -0.89 | 98.18 | 96.40 | -1.78 | 97.30 | 95.96 | -1.34 |
| | Shadow | 97.50 | 90.00 | -7.50 | 97.50 | 97.30 | -0.20 | 97.50 | 93.51 | -3.99 |
| **Overall accuracy** | | **97.11** | **96.04** | **-1.07** | | | | | | |
| **Kappa statistic** | | **96.08** | **94.64** | **-1.44** | | | | | | |

TEXAS TECH
UNIVERSITY

# Final Classified Map: [STAGE II-LiDAR]

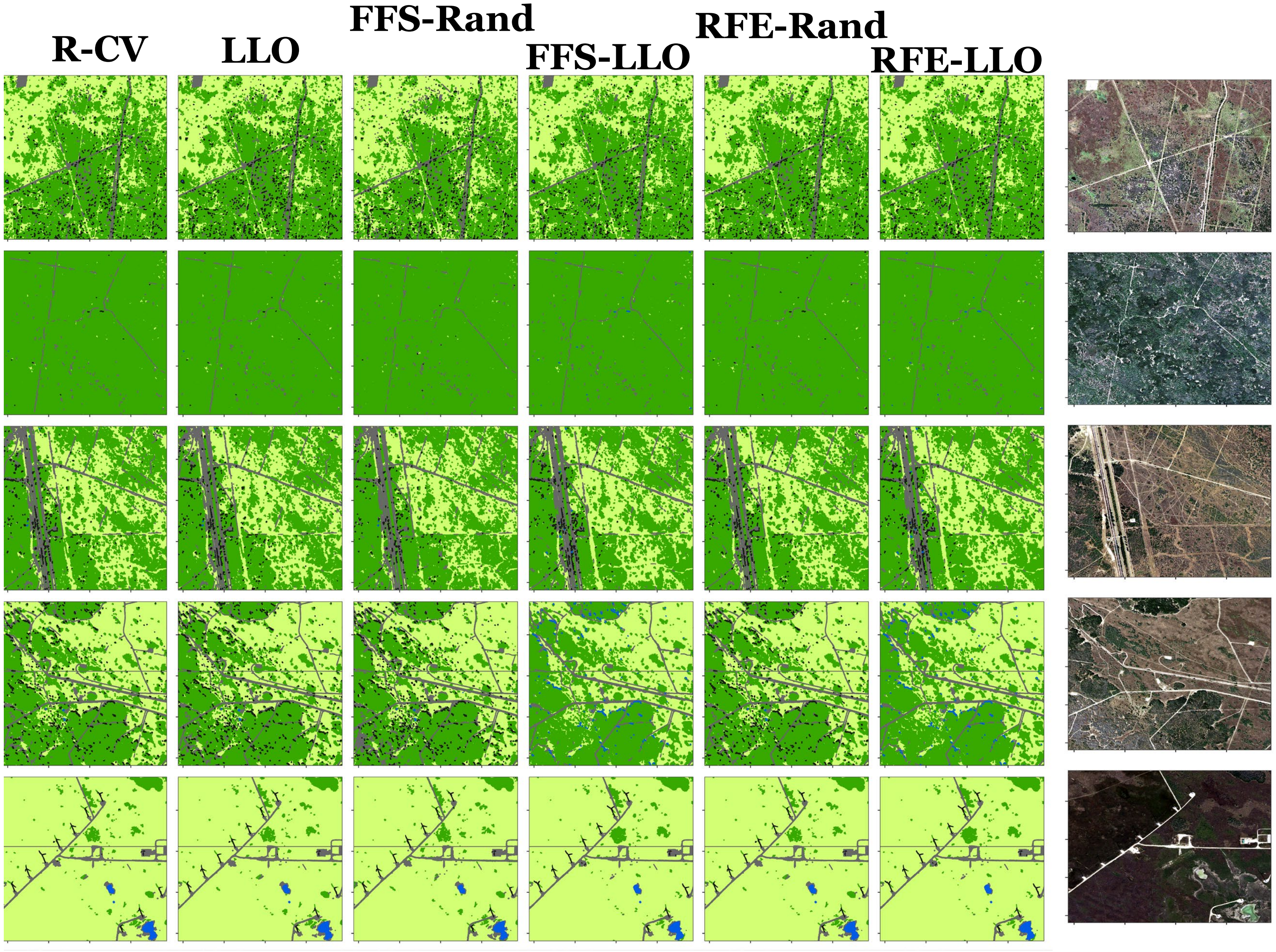




TEXAS TECH UNIVERSITY

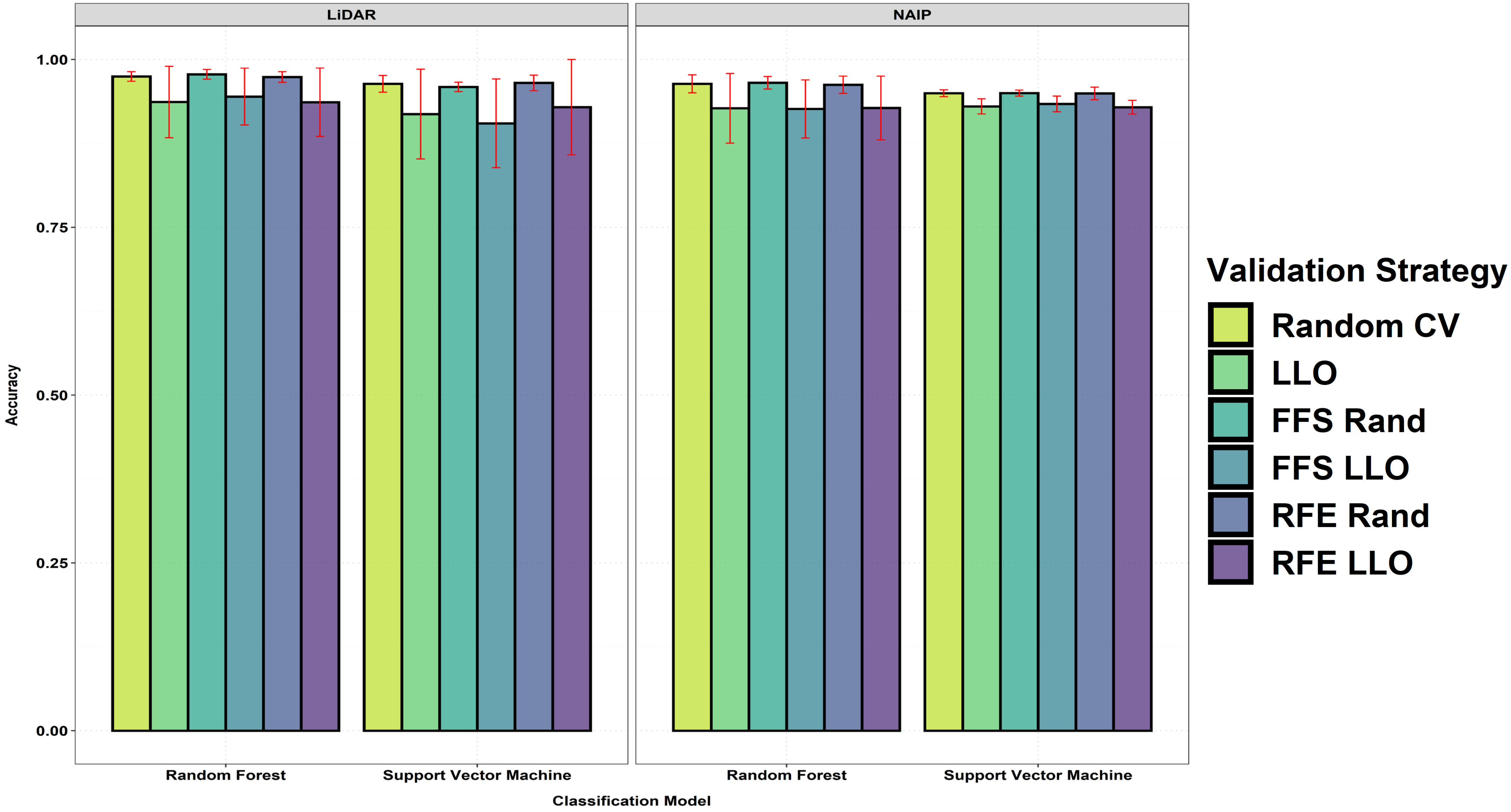


# Final Classified Map: STAGE II-NAIP



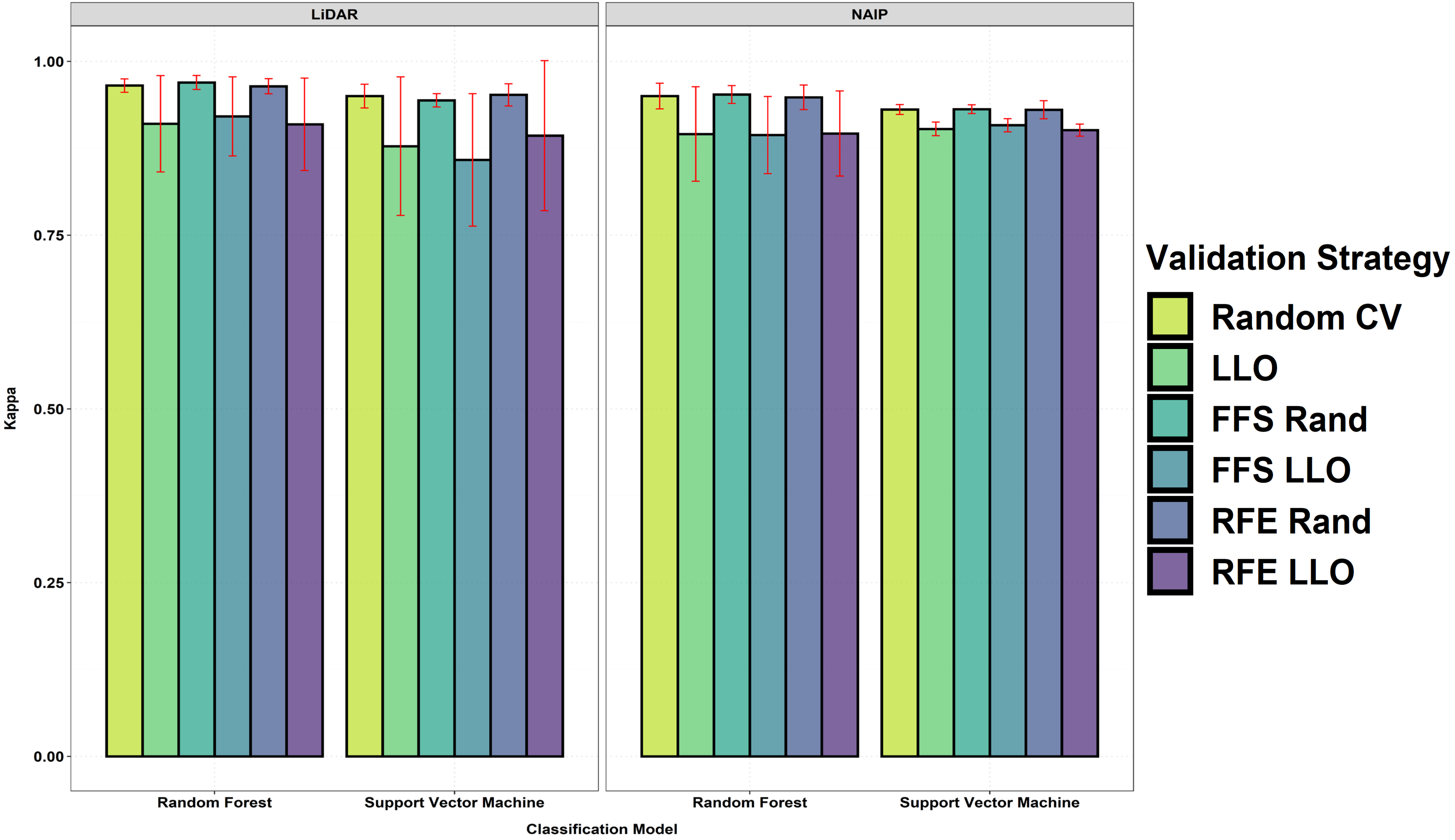
# Accuracy by Validation strategies





# Accuracy by Validation strategies

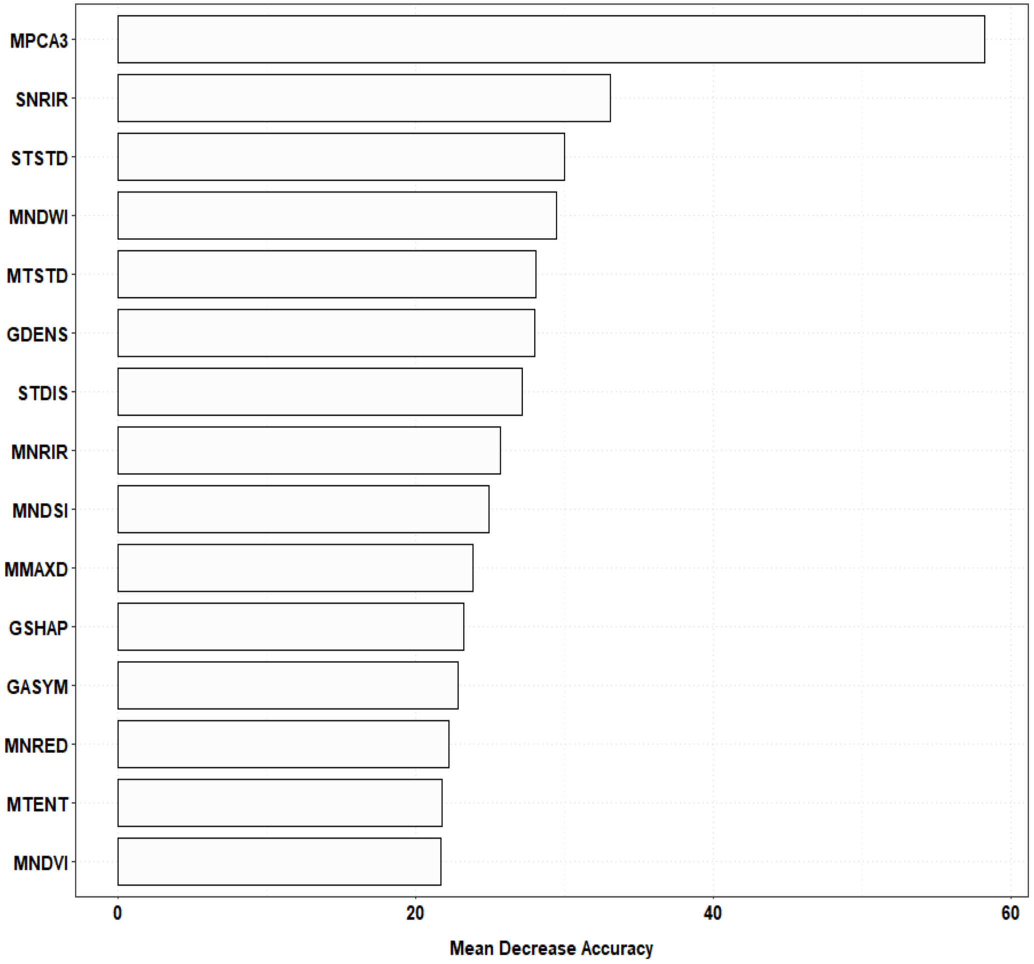
A. Background
B. LULC Products
C. Objective
D. Study Area
E. Data Processing Workflow
F. Results: Accuracy
G. Results: Confusion Matrix
H. Results: Map Comparison
I. Results: Validation strategies
J. Results: Variable Importance
K. Key Points
L. Acknowledgements
M. References



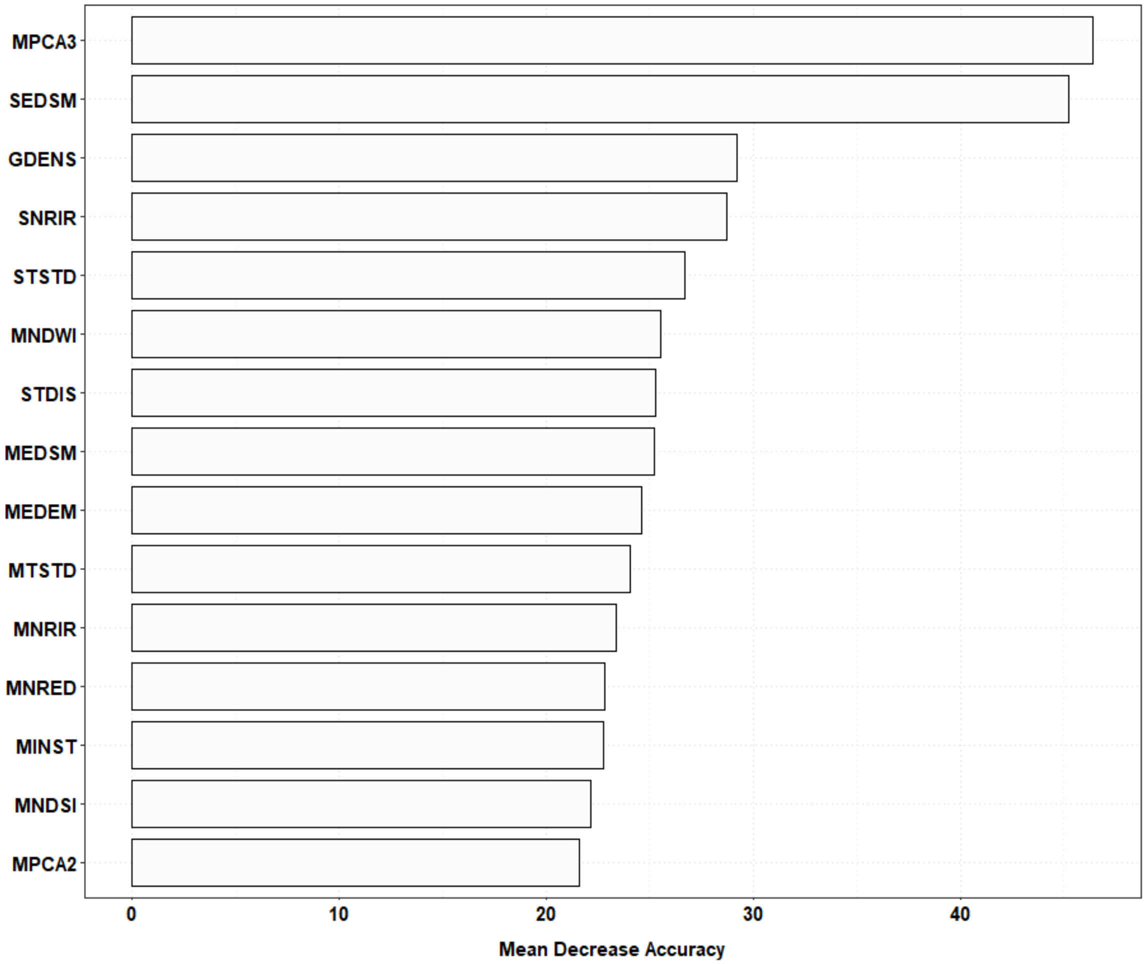
TEXAS TECH UNIVERSITY



# Variable Importance



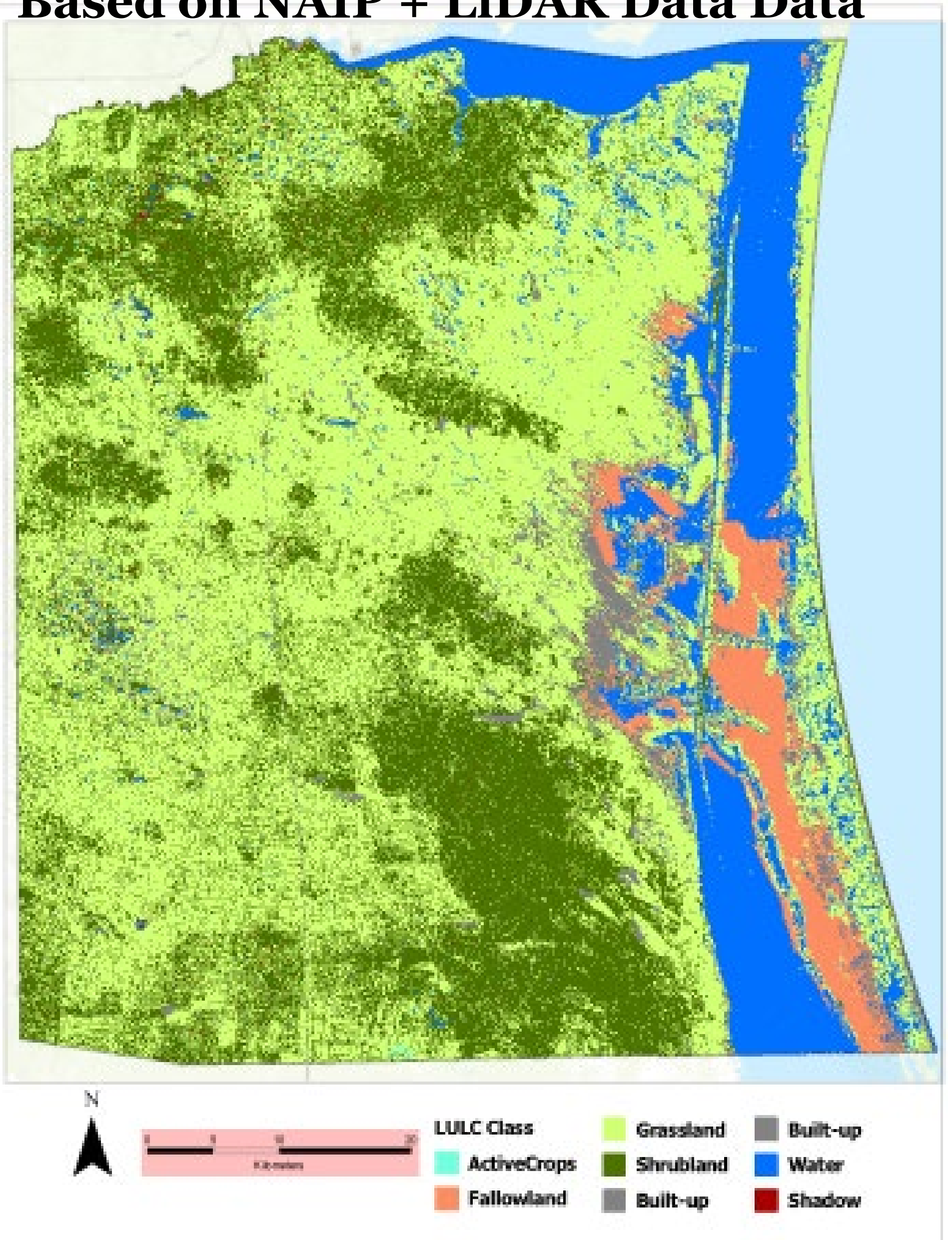

NAIP: LLO



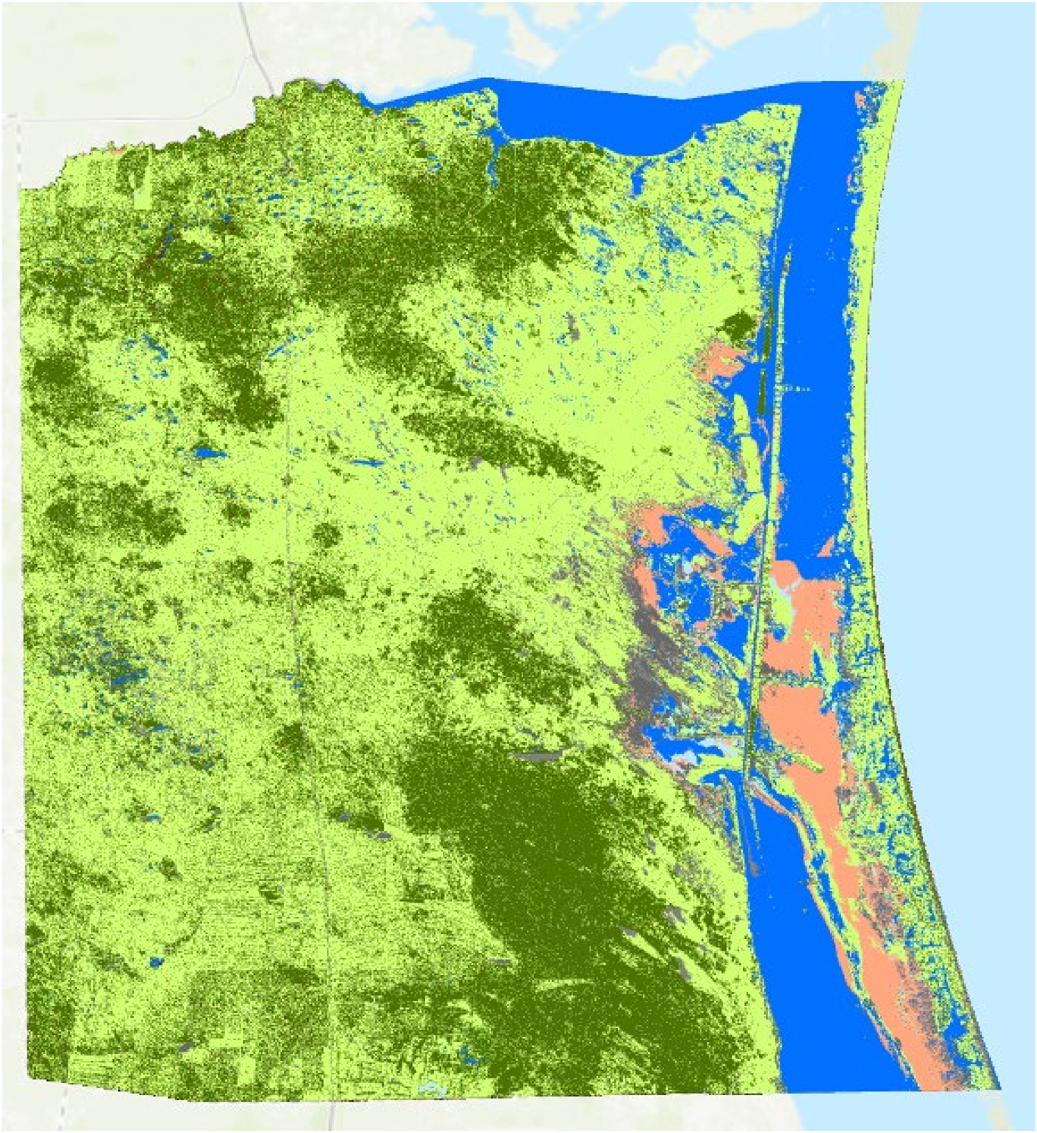

NAIP+LiDAR: LLO

# Final Classified Map

### Based on NAIP + LiDAR Data Data




### Based on NAIP Data



# Key Points

- Random Forest performed the best, while SVM produced better accuracy statistics in certain classes.
- Higher predicted accuracy based on random sampling strategies does not necessarily guarantee higher spatial accuracy. Evidence from target-oriented validation.
- Creating spatially independent data improve spatial accuracy, especially when used with target-oriented validation strategies. Removes- autocorrelation and avoids pseudo-replication.
- Accounting for the computation cost associated with LiDAR Data, we did not find the practical significance of LiDAR data. An addition of LiDAR data on top of original NAIP bands and features generated on these bands marginally improved the classification.
- When spatially independent data are utilized in the spatial validation strategies, RFE, FFS and LLO produce similar accuracy metrics.

TEXAS TECH UNIVERSITY



# Acknowledgments


Texas Comptroller of Public Accounts

Dr. Carlos Portillo-Quintero [NRM, TTU)

Dr. Nancy McIntyre [Biology, TTU]

Dr. Gad Perry [NRM, TTU]

Dr. Robert Cox [NRM, TTU]

Dr. Samantha Kahl [Blackburn College, IL]


# References


- Hollmann, R., Merchant, C.J., Saunders, R., Downy, C., Buchwitz, M., et al., 2013. The ESA climate change initiative: Satellite data records for essential climate variables. Bull. Am. Meteorol. Soc. 94, 1541–1552.
- Meyer, H., Reudenbach, C., Hengl, T., Katurji, M., & Nauss, T. (2018). Improving performance of spatio-temporal machine learning models using forward feature selection and target-oriented validation. Environmental Modelling and Software,**101**:1–9.
- ESRI, 2020. ArcGIS Pro. Redland, California; HEXAGON, 2020. ERDAS IMAGINE 2020
- R Core Team, 2020. R: A language and environment for statistical computing.
- Trimble, 2020. eCognition Developer 9; Sunnyvale, CA, USA.
- Landsat (www.usgs.gov)
- Sentinel (sentinel.esa.int)
- Modis (www.uvm.edu)
- Radar (Canadian Remote Sensing)
- UAV (Wikipedia)
- USGS (3D Elevation Program)

Thank

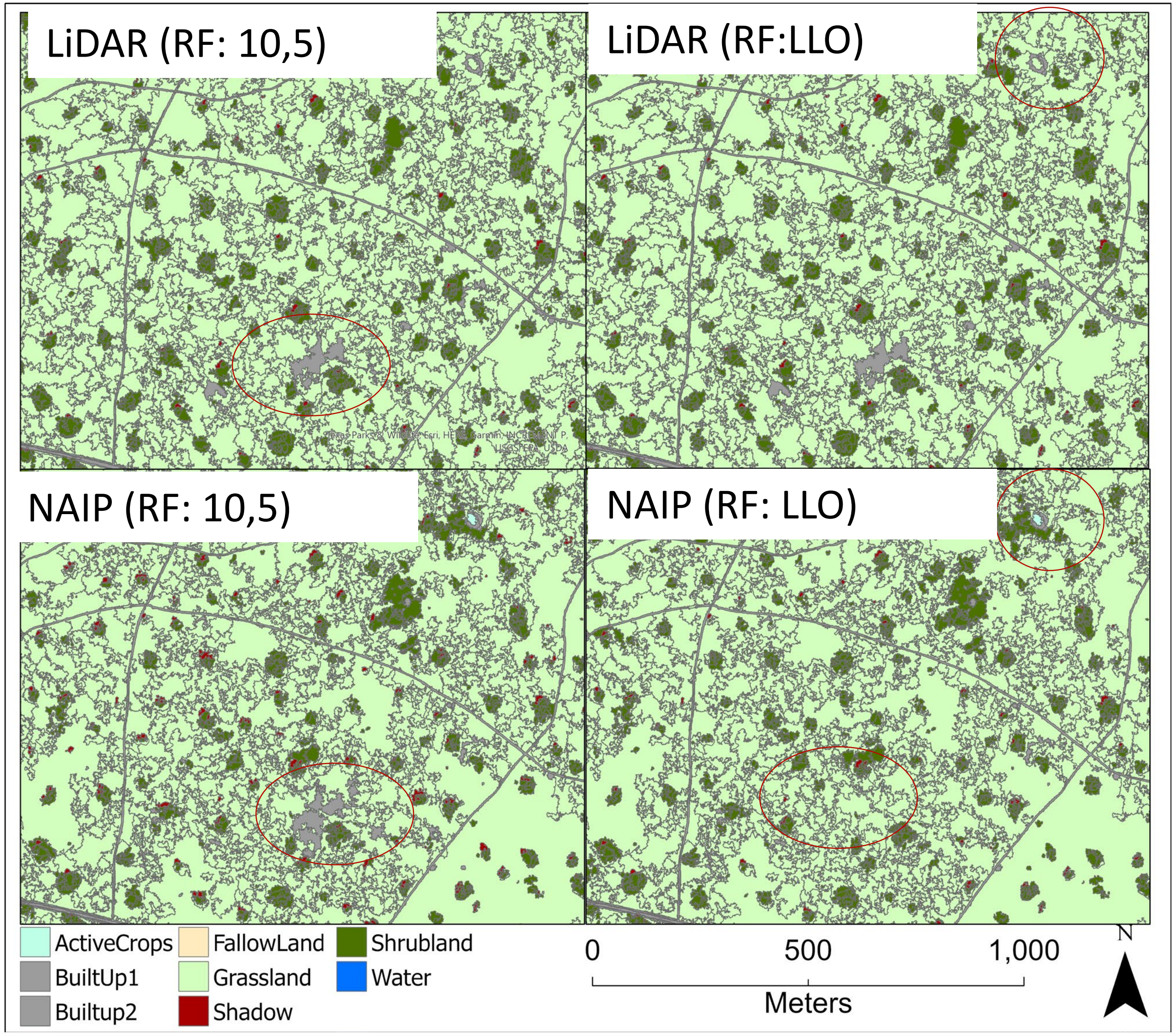
You

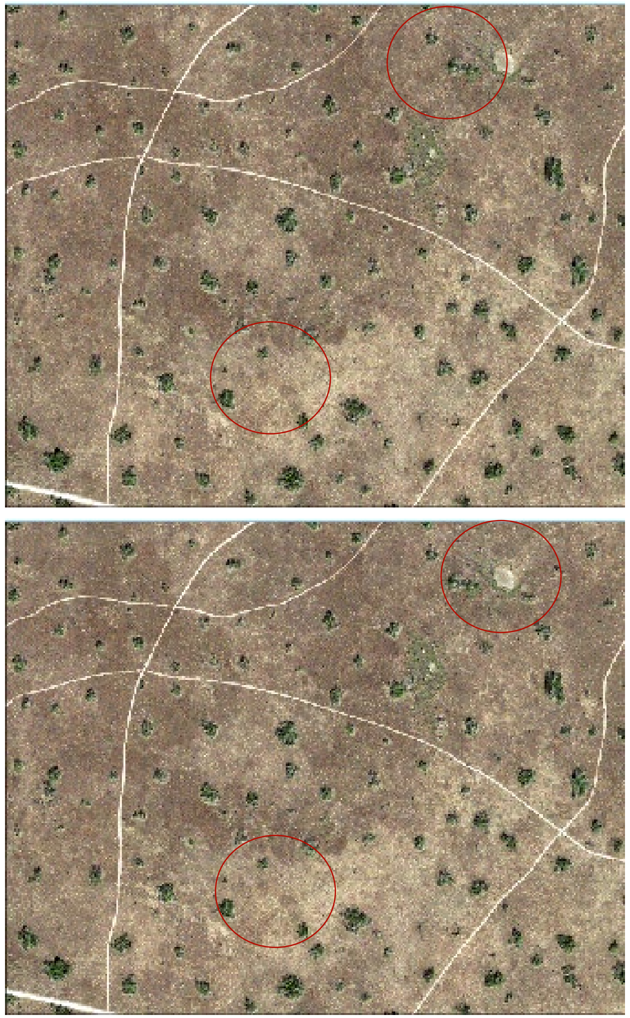
# Questions?

TEXAS TECH UNIVERSITY

# Classification Comparison: Subset



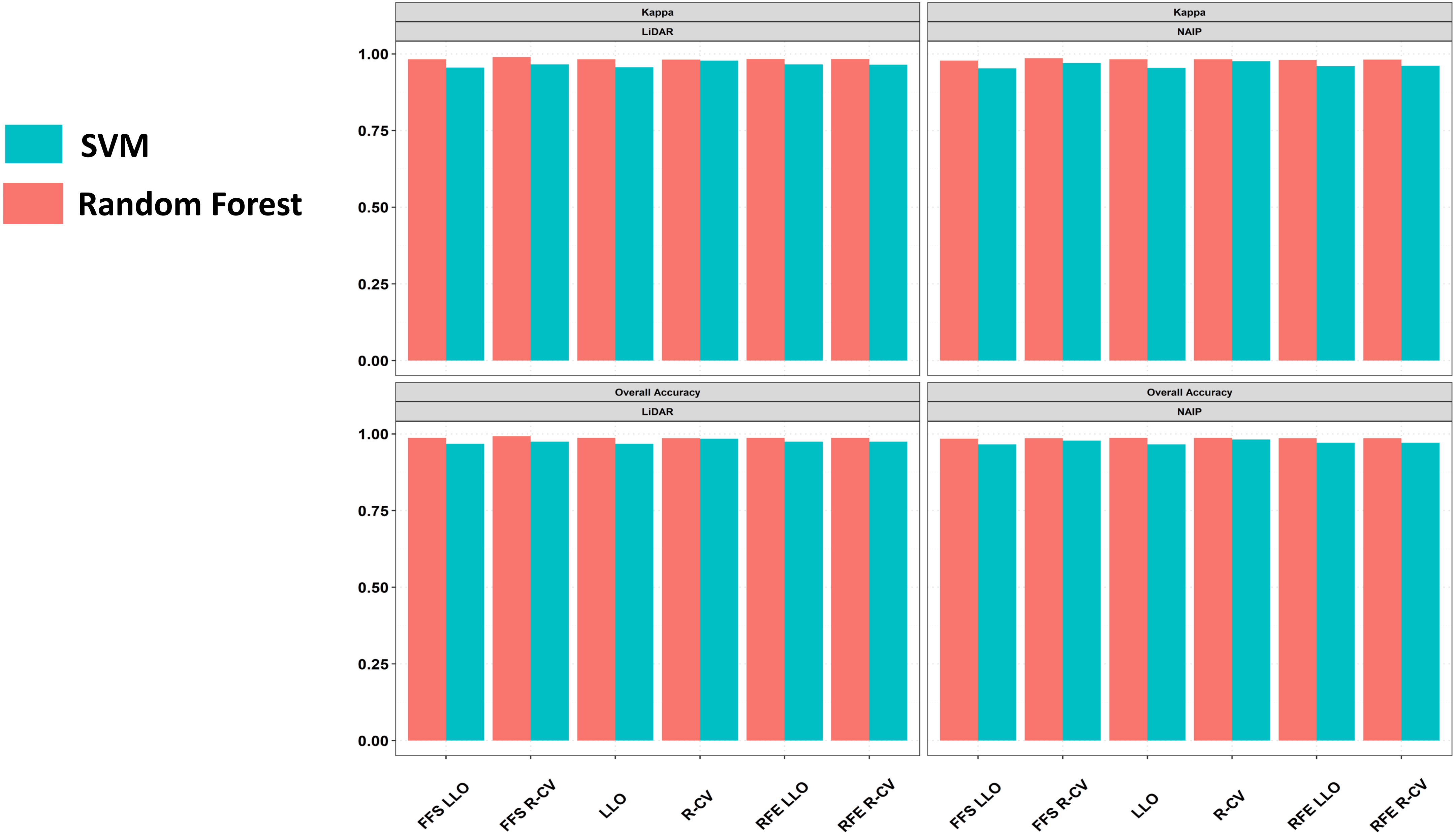


TEXAS TECH UNIVERSITY

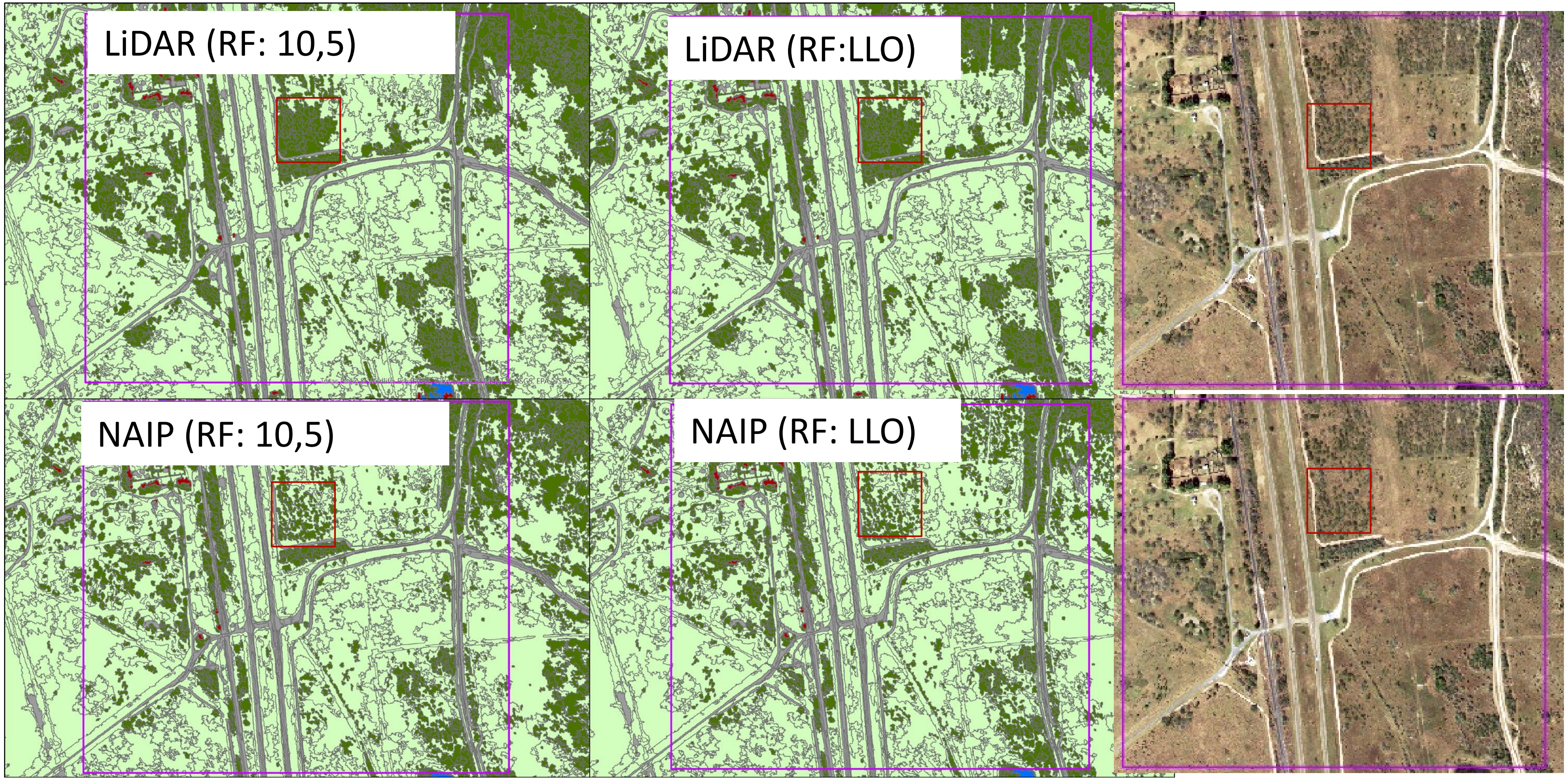
# Classification Accuracy: Validation Data

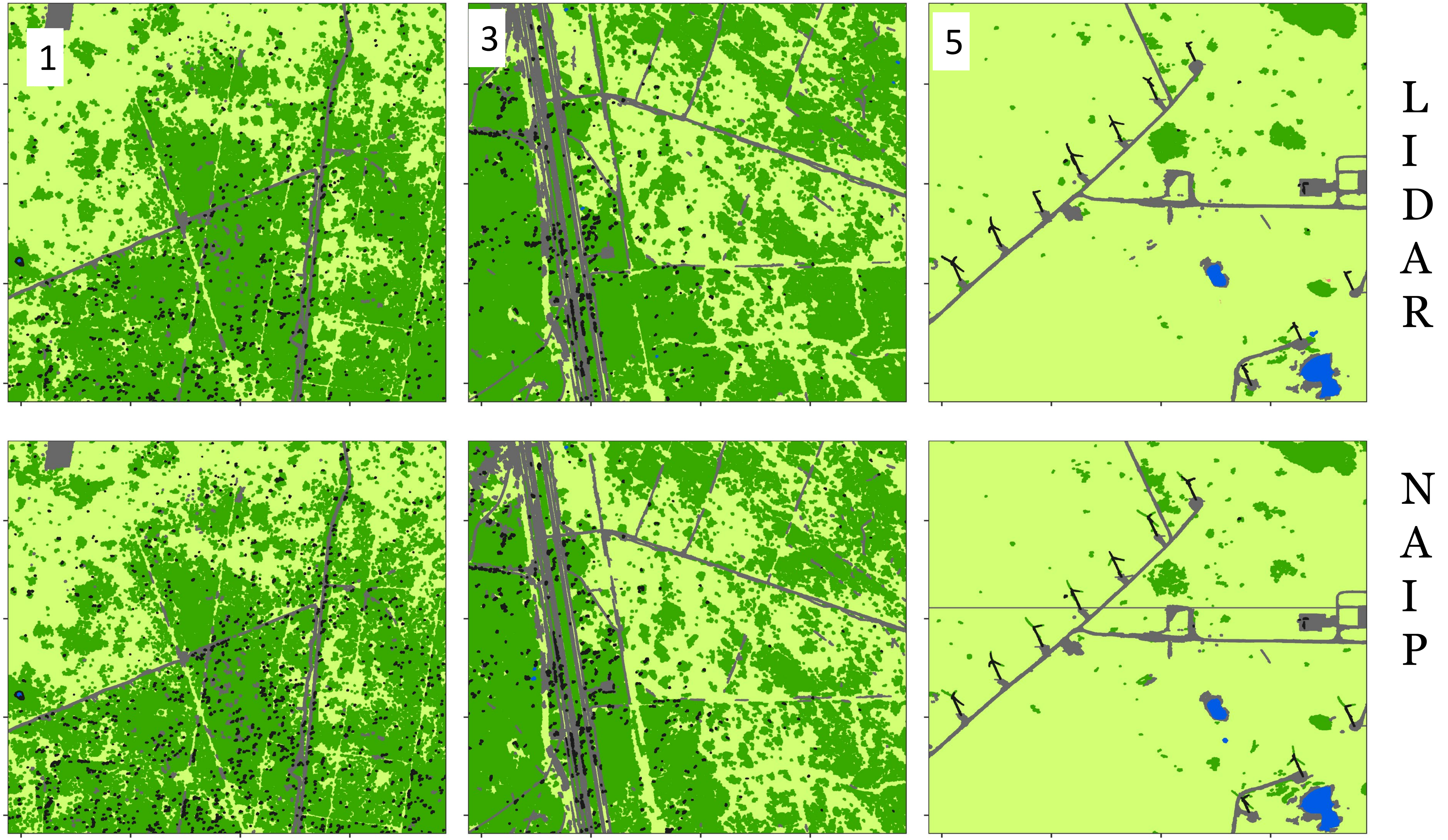




# Classification Comparison: subset






1
3
5
LIDAR
NAIP