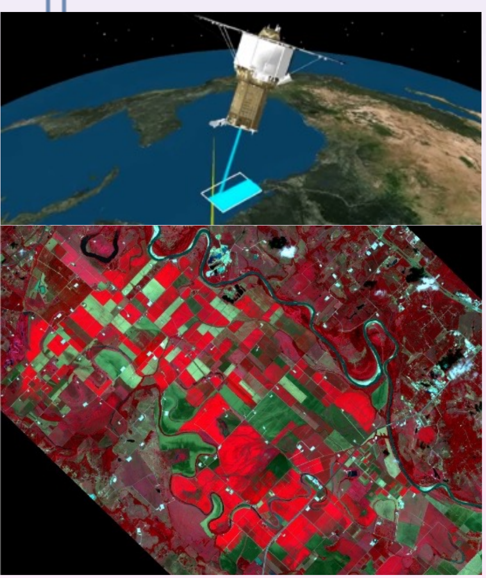
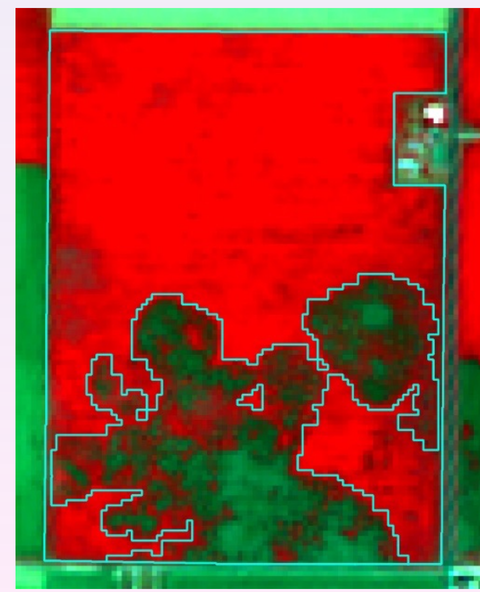
The Electromagnetic Spectrum

# Applying Machine Learning Techniques for Image Classification in Agricultural Applications




Chenghai Yang
Research Agricultural Engineer

USDA-ARS
Aerial Application Technology
Research Unit
College Station, TX

**2024 NAAA Ag Aviation Expo, Fort Worth, TX, Nov. 18-21, 2024**

Aerial Application Technology

# Artificial Intelligence (AI)

- **Artificial Intelligence:** A branch of computer science focused on creating systems that can perform tasks requiring **human intelligence**. These tasks include:
  - Learning from data
  - Recognizing patterns
  - Problem-solving
  - Decision-making
  - Understanding language and visual inputs
- **AI Developments**
  - Has been around since mid-1950s
  - Has had its ups and downs with fastest development in recent years
- **Driving Factors:**
  - Increased computing power: High-performance GPUs and cloud computing
  - Growth of big data: Enormous datasets powering AI training
  - Development of advanced algorithms: Deep learning, reinforcement learning, and neural networks
- **Applications Across Fields:**
  - AI applications are vast and span numerous industries, including agriculture

Aerial Application Technology

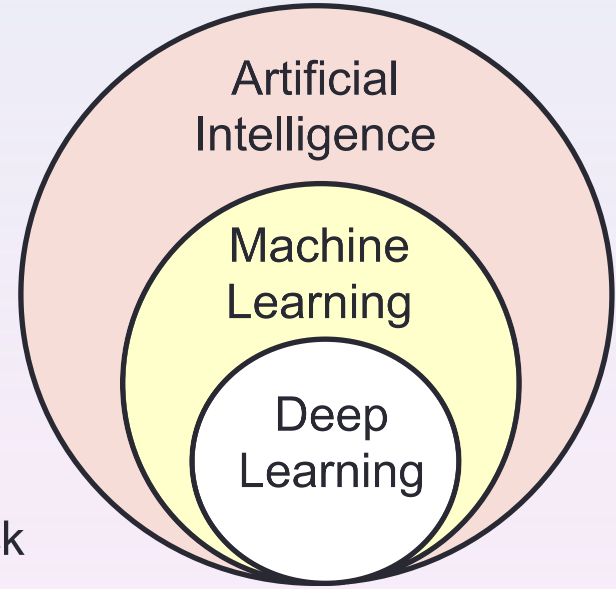
# Key Technologies Behind AI

- **Machine Learning:** A subset of AI that develops algorithms enabling systems to learn from data, identify patterns, and make predictions, widely used for image classification and object detection.
- **Deep Learning:** A specialized branch of machine learning that uses deep neural networks to analyze large and complex datasets, particularly effective in tasks like in image and speech recognition, and self-driving cars.
- **Natural Language Processing (NLP):** Enables machines to understand and respond to human language, forming the basis for chatbots, virtual assistants, and language translation tools.
- **Computer Vision:** Allows machines to interpret visual data, powering applications like facial recognition, object detection, autonomous vehicles, and medical imaging.

Aerial Application Technology

# Types of Machine Learning

- **Supervised Learning:** The algorithm is trained on labeled data (input-output pairs).
  - Examples: Classification, regression
  - Use cases: Spam detection, sentiment analysis, image recognition.
- **Unsupervised Learning:** The algorithm works with unlabeled data, identifying patterns and structures.
  - Examples: Clustering, dimensionality reduction
  - Use cases: Customer segmentation, anomaly detection.
- **Reinforcement Learning:** The model learns by interacting with an environment and receiving feedback (rewards or penalties).
  - Examples: Robotics, game playing (e.g., AlphaGo).
  - Use cases: Self-driving cars, autonomous drones.



# Machine Learning in Agriculture

- **ML applications:**
  - Crop identification
  - Pest detection (disease, weeds, insects)
  - Yield prediction
  - Precision agriculture (site-specific product applications)
  - Soil health monitoring
  - Automated harvesting
- **This presentation illustrates with examples:**
  - Crop identification using supervised learning
  - Disease damage detection using unsupervised learning

# Crop Identification with Supervised Learning

## ❑ Crop Identification

- The process of classifying different types of crops within a given area using various technologies, such as remote sensing, machine learning, and computer vision.
- Supervised learning models are trained using labeled datasets with known crop types.

## ❑ Steps Involved:

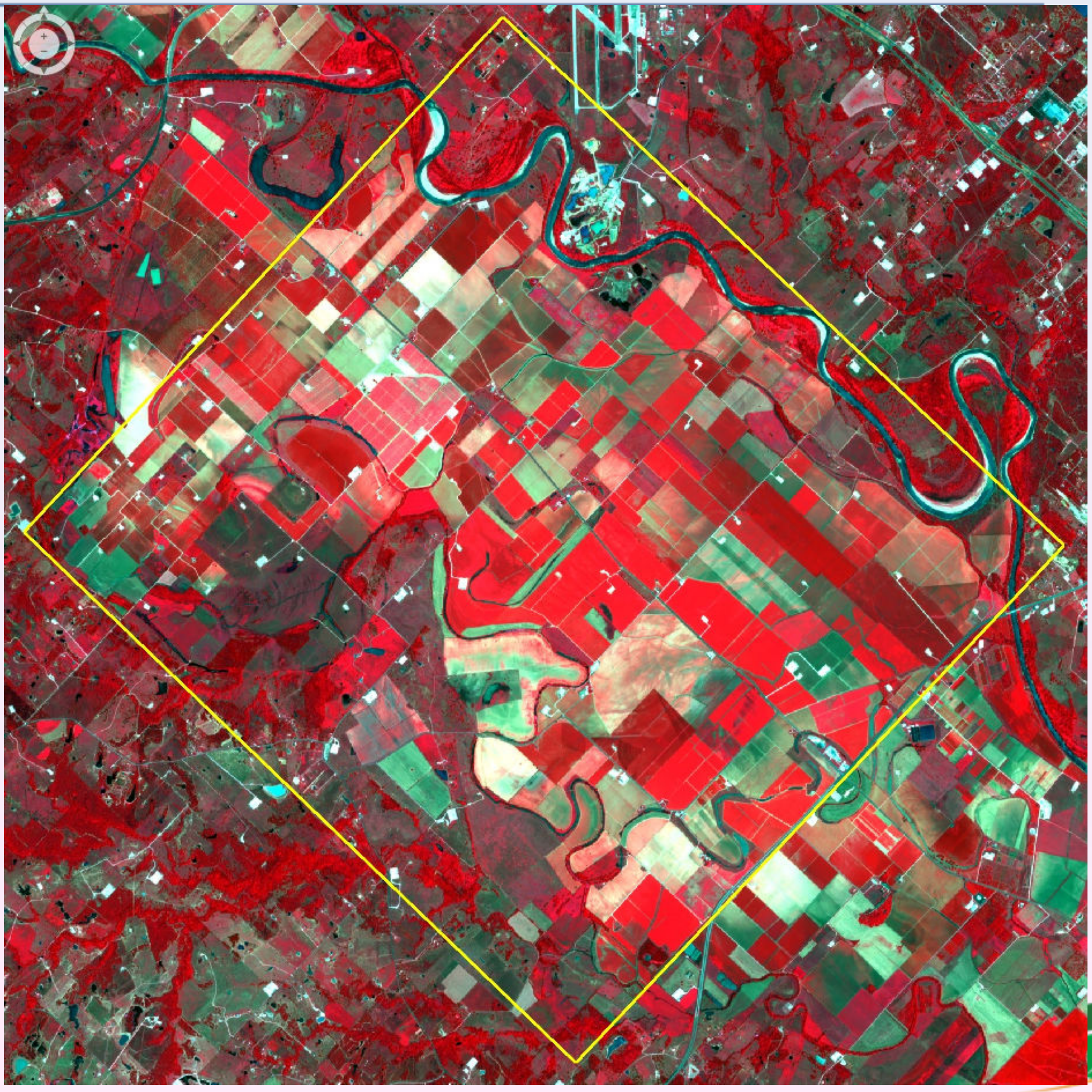
- **Training data collection:** Gather a labeled dataset of images, each assigned a category, such as crop types.
- **Data preprocessing:** Preprocess training data (resizing, normalization, augmentation).
- **Model training:** Train a classification model on labeled image data.
- **Model validation:** Assess model accuracy using test image data.
- **Deployment:** Use the model for real-time crop identification.

# Supervised Learning Algorithms

- **Random Forest (RF):** An ensemble method that uses multiple decision trees to classify crops.
- **Support Vector Machines (SVM):** Effective for classification in high-dimensional spaces.
- **K-Nearest Neighbors (KNN):** A simple, instance-based learning algorithm where the target class is assigned based on the majority class of the K closest data points.
- **Convolutional Neural Networks (CNNs):** Deep learning models that automatically learn spatial features from images.
- **Naive Bayes:** A probabilistic classifier based on Bayes' theorem, assuming that features are conditionally independent given the class.
- **XGBoost:** An optimized version of gradient boosting that is faster and more efficient, particularly in handling large datasets.

Aerial Application Technology

# Sentinel-2 Imagery for a Cropping Area near College Station, TX

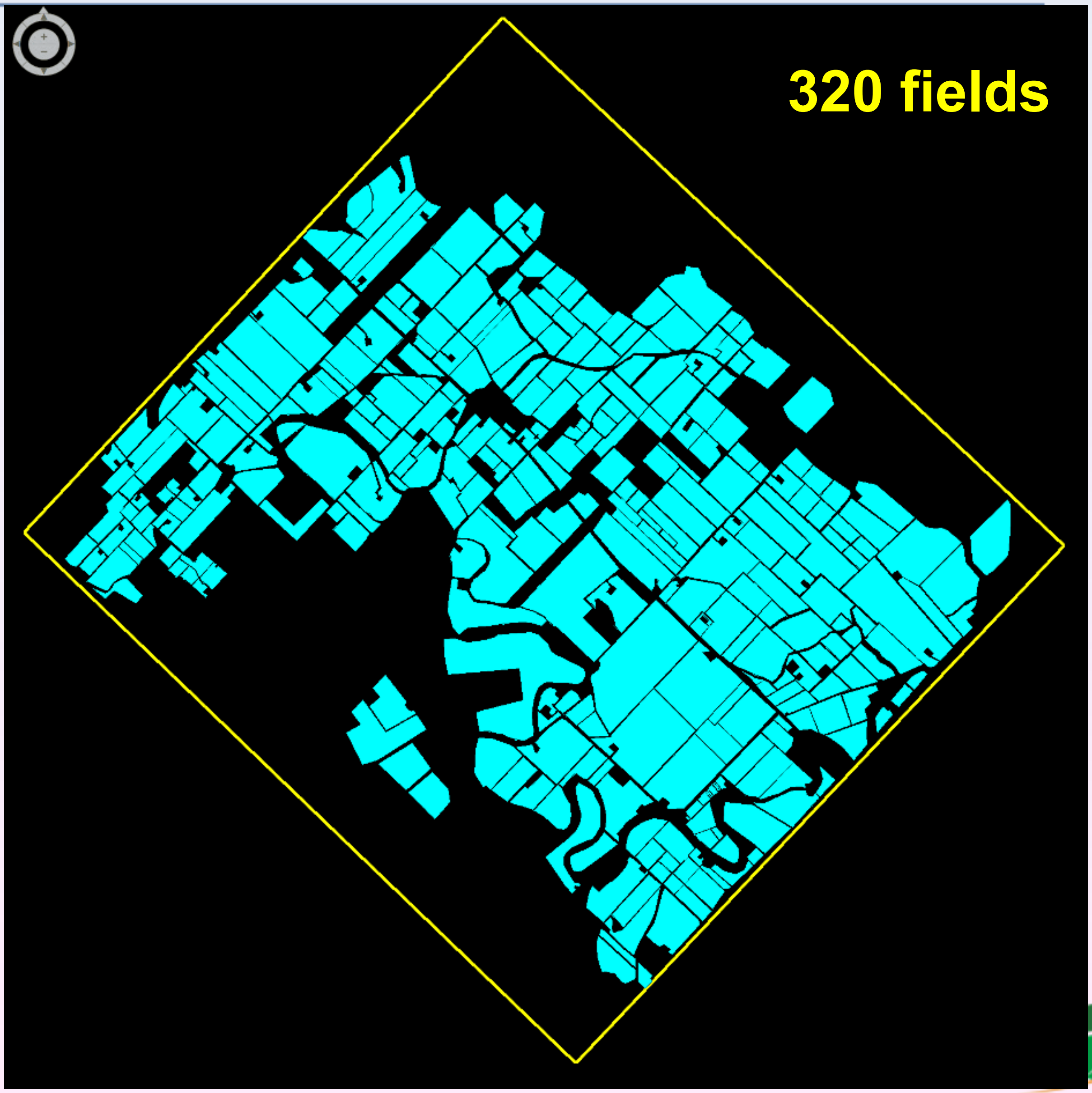


**6 mi x 6.6 mi**
**25,300 ac**



Aerial Application Technology

# Crop Fields and Boundaries



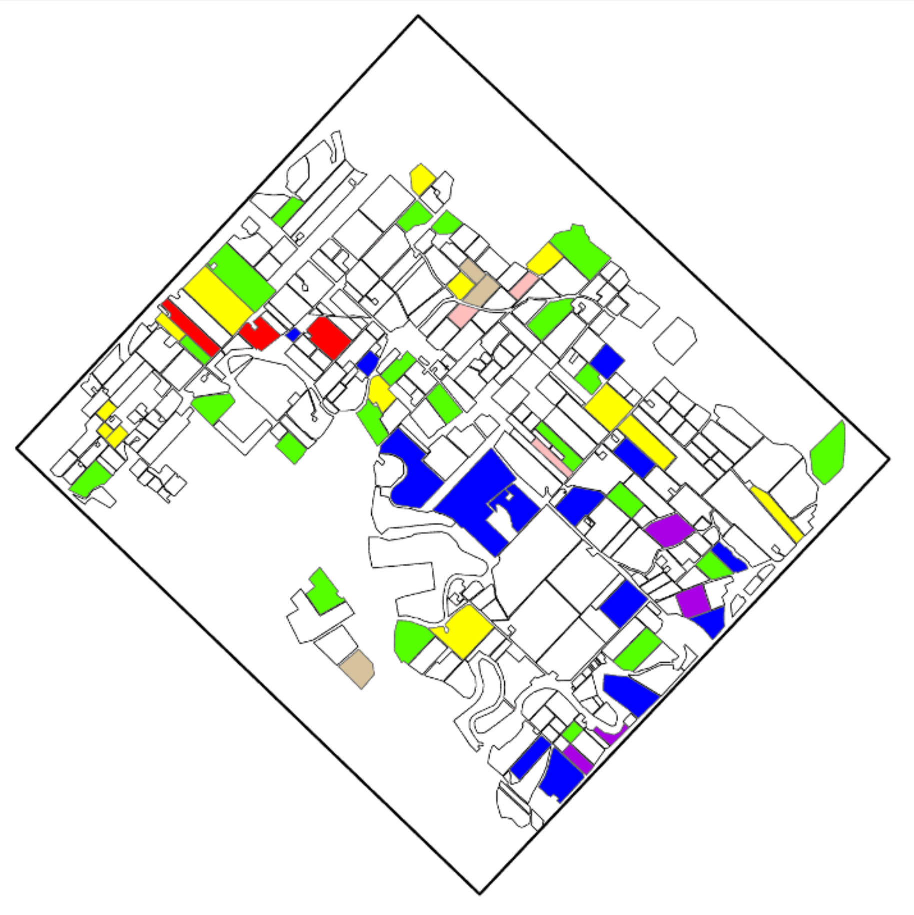

USDA
Farm
Service
Agency

# Training Fields for Seven Classes

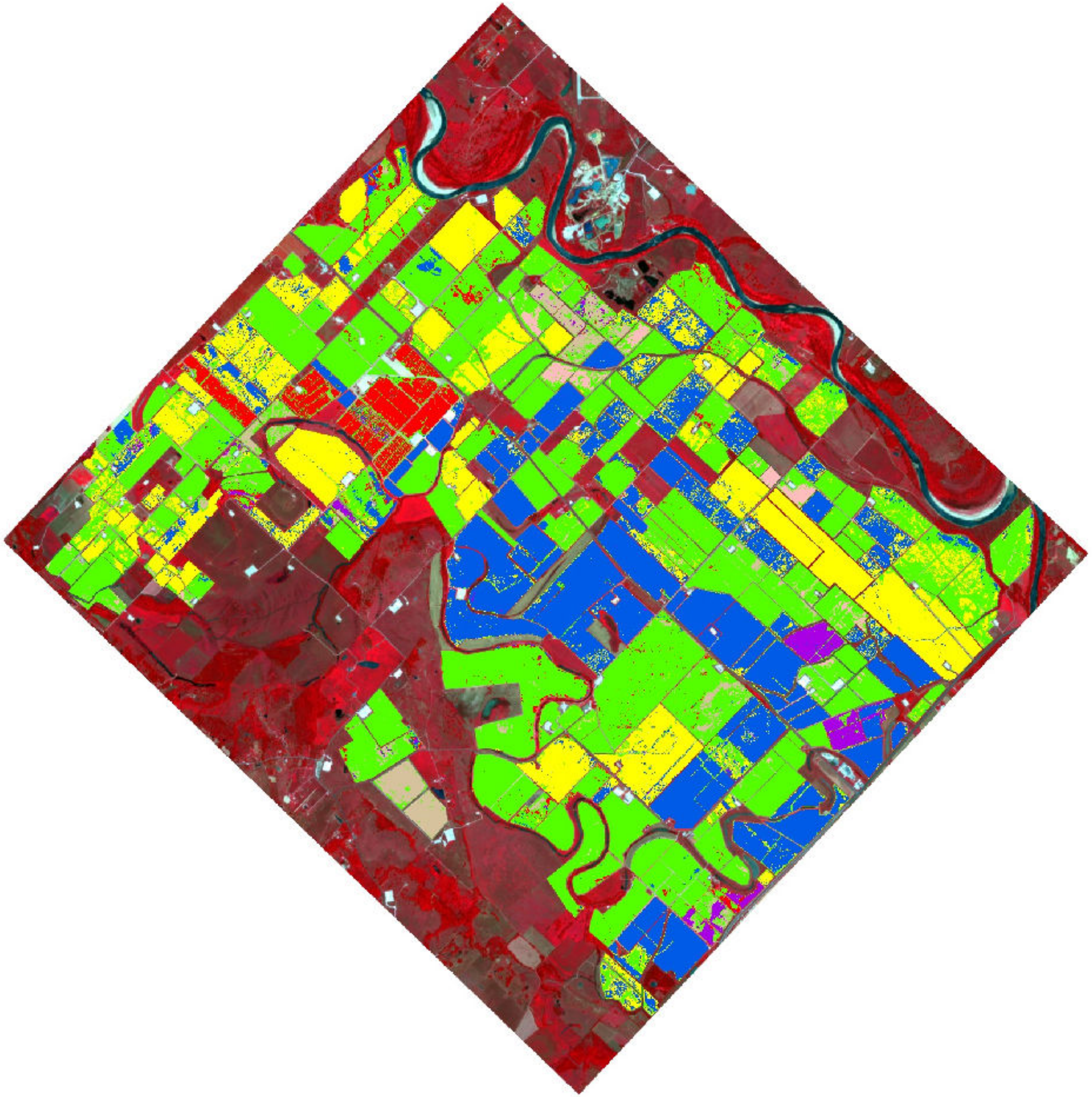
- Six crops & fallow

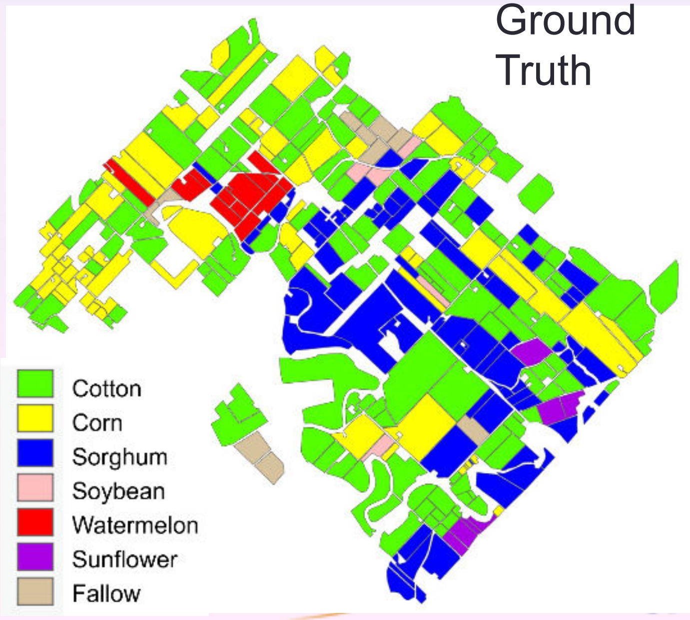
| | | |
|---|---|---|
| | Cotton | 22 |
| | Corn | 13 |
| | Sorghum | 14 |
| | Soybean | 3 |
| | Watermelon | 3 |
| | Sunflower | 4 |
| | Fallow | 3 |

- A total of 62 training fields

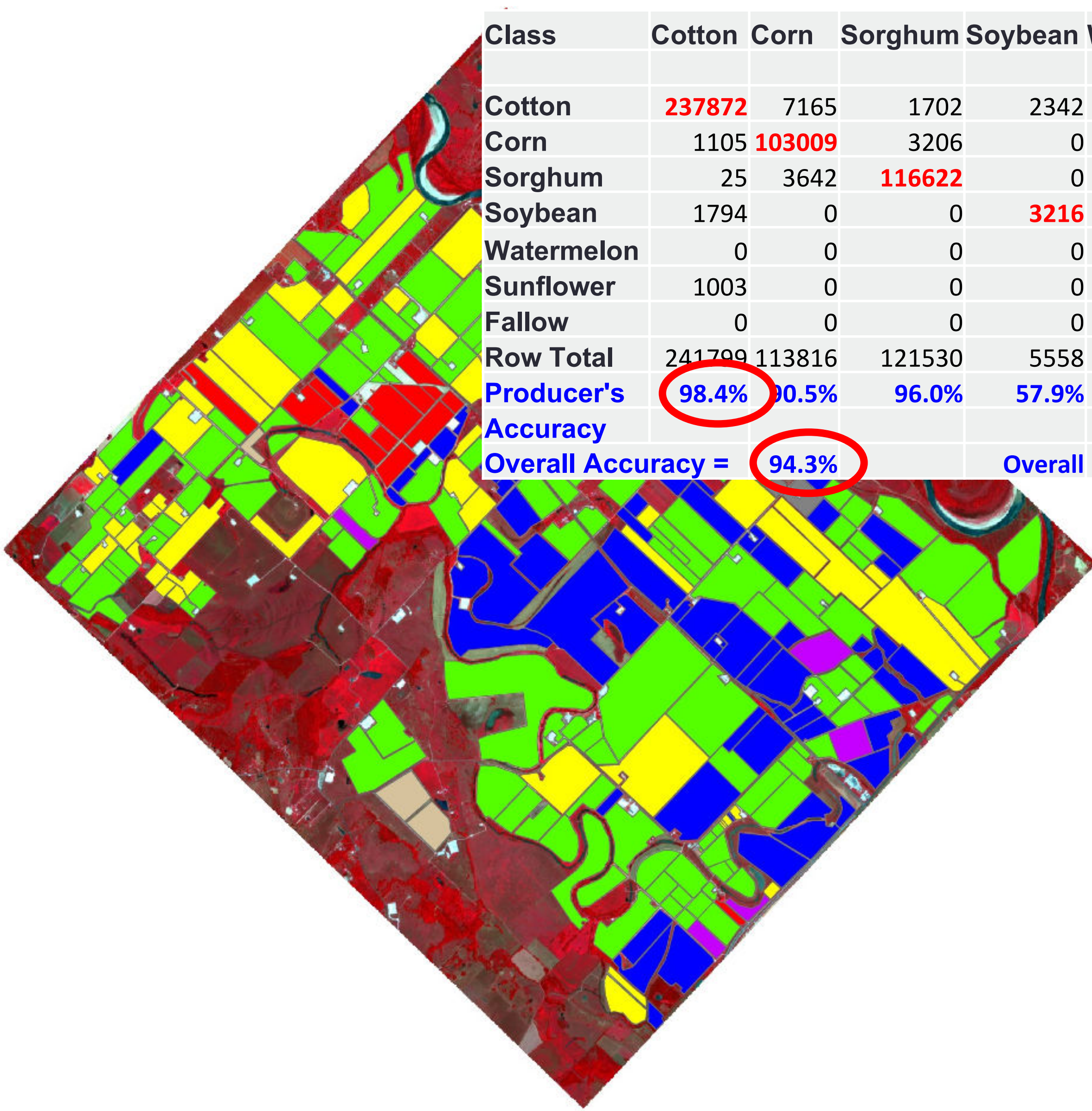


# Random Forest-based Classification Map Clipped with Field Boundaries

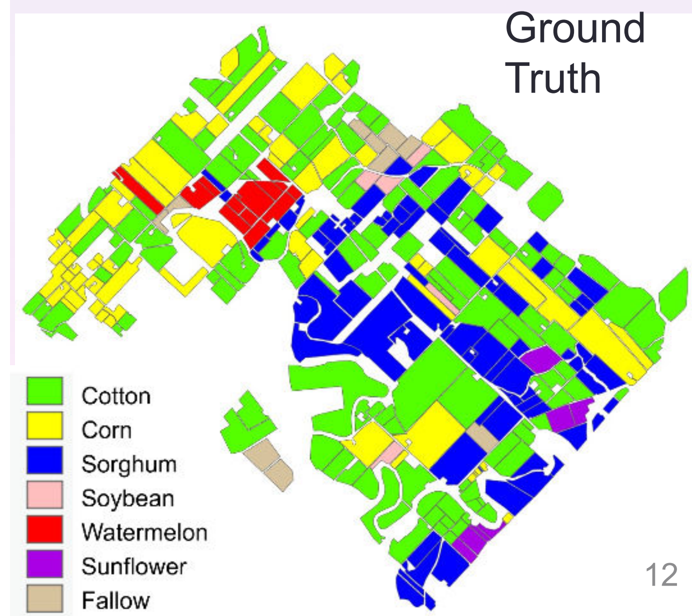




# Random Forest-based Classification Map with One Dominant Class Assigned to Each Field



| Class | Cotton | Corn | Sorghum | Soybean | Watermelon | Sunflower | Fallow | Column Total | User's Accuracy |
|---|---|---|---|---|---|---|---|---|---|
| Cotton | 237872 | 7165 | 1702 | 2342 | 686 | 1556 | 3942 | 255265 | 93.2% |
| Corn | 1105 | 103009 | 3206 | 0 | 0 | 0 | 0 | 107320 | 96.0% |
| Sorghum | 25 | 3642 | 116622 | 0 | 0 | 1468 | 0 | 121757 | 95.8% |
| Soybean | 1794 | 0 | 0 | 3216 | 0 | 0 | 0 | 5010 | 64.2% |
| Watermelon | 0 | 0 | 0 | 0 | 16842 | 396 | 0 | 17238 | 97.7% |
| Sunflower | 1003 | 0 | 0 | 0 | 0 | 6480 | 0 | 7483 | 86.6% |
| Fallow | 0 | 0 | 0 | 0 | 0 | 0 | 10197 | 10197 | 100.0% |
| Row Total | 241799 | 113816 | 121530 | 5558 | 17528 | 9900 | 14139 | 524270 | |
| Producer's Accuracy | 98.4% | 90.5% | 96.0% | 57.9% | 96.1% | 65.5% | 72.1% | | |
| Overall Accuracy = | | 94.3% | | Overall Kappa = | | 0.915 | | | |



# Disease Detection with Unsupervised Learning

- **Disease Detection:**
  - The task of detecting plant disease or damage without labeled data, by identifying anomalies or unusual patterns.
  - Unsupervised learning can cluster images into groups based on visual patterns and anomalies, allowing for disease detection.
- **Steps Involved:**
  - **Data collection:** Collect images of healthy and diseased crops.
  - **Data preprocessing:** Prepare image data for training.
  - **Model training:** Select unsupervised techniques to detect natural patterns and anomalies.
  - **Cluster labeling:** Interpret or label the clusters based on expert knowledge and ground truth data. For example, you might identify one cluster as "healthy," another as "mild disease," and another as "severe disease."
  - **Model validation:** Assess model accuracy using test image data.
  - **Deployment:** Detect disease and damage in new images.

# Unsupervised Learning Algorithms

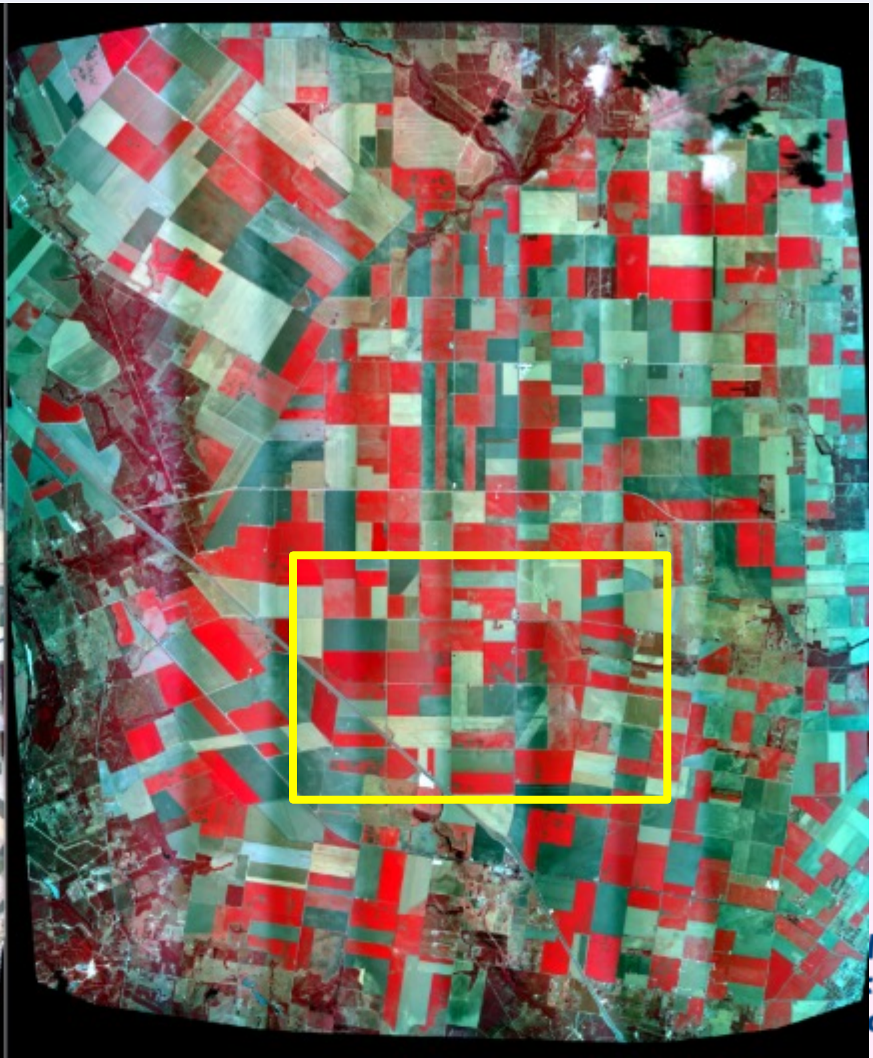
- **K-Means Clustering:** Groups similar data points (e.g., pixels or image patches) into clusters, which can help identify regions of interest or anomalies related to crop diseases.
- **Hierarchical Clustering:** Builds a hierarchy of clusters in a tree-like structure (dendrogram) and is useful for identifying clusters of diseased crops based on their similarity in spectral or image features.
- **Principal Component Analysis (PCA):** A dimensionality reduction technique that can be used to identify patterns and variations in hyperspectral or multispectral image data, which can reveal disease symptoms that may not be immediately obvious in raw data.
- **Autoencoders (Deep Learning):** Neural networks trained to encode data into a smaller representation and then reconstruct it. Anomalies such as disease symptoms can be detected by comparing reconstruction errors between healthy and diseased crops.
- **Isolation Forest:** An anomaly detection algorithm that isolates anomalies (such as diseased crops) rather than profiling normal data, making it useful for detecting outliers in agricultural data (e.g., images or sensor data).

Aerial Application Technology

# Mosaicked Aerial Images for Cotton Growing Area near Corpus Christi, TX

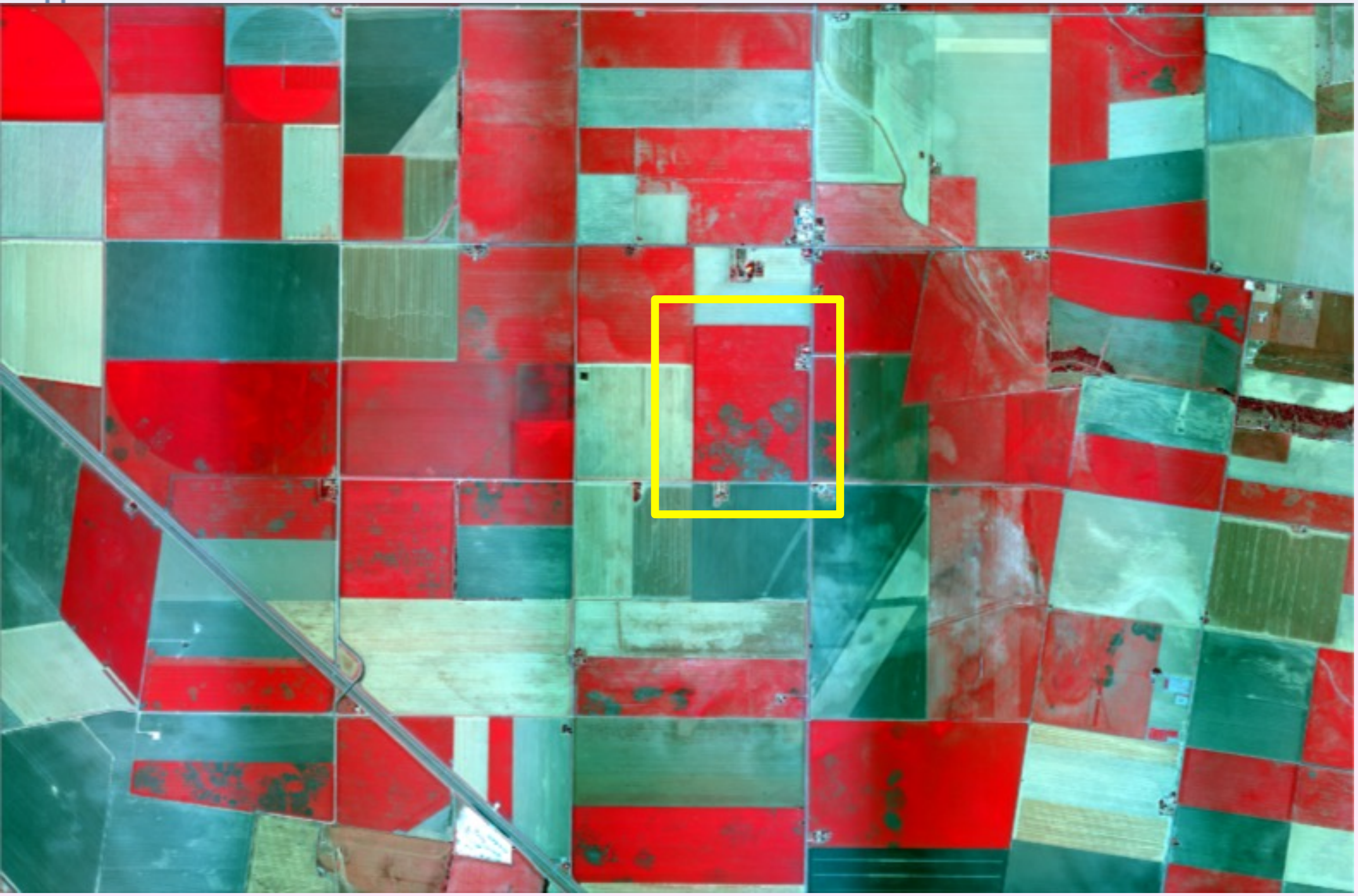
**RGB**

**Color-infrared (CIR)**



# Subset Color-infrared Image

**6 mi x 4 mi (15,000 ac)**

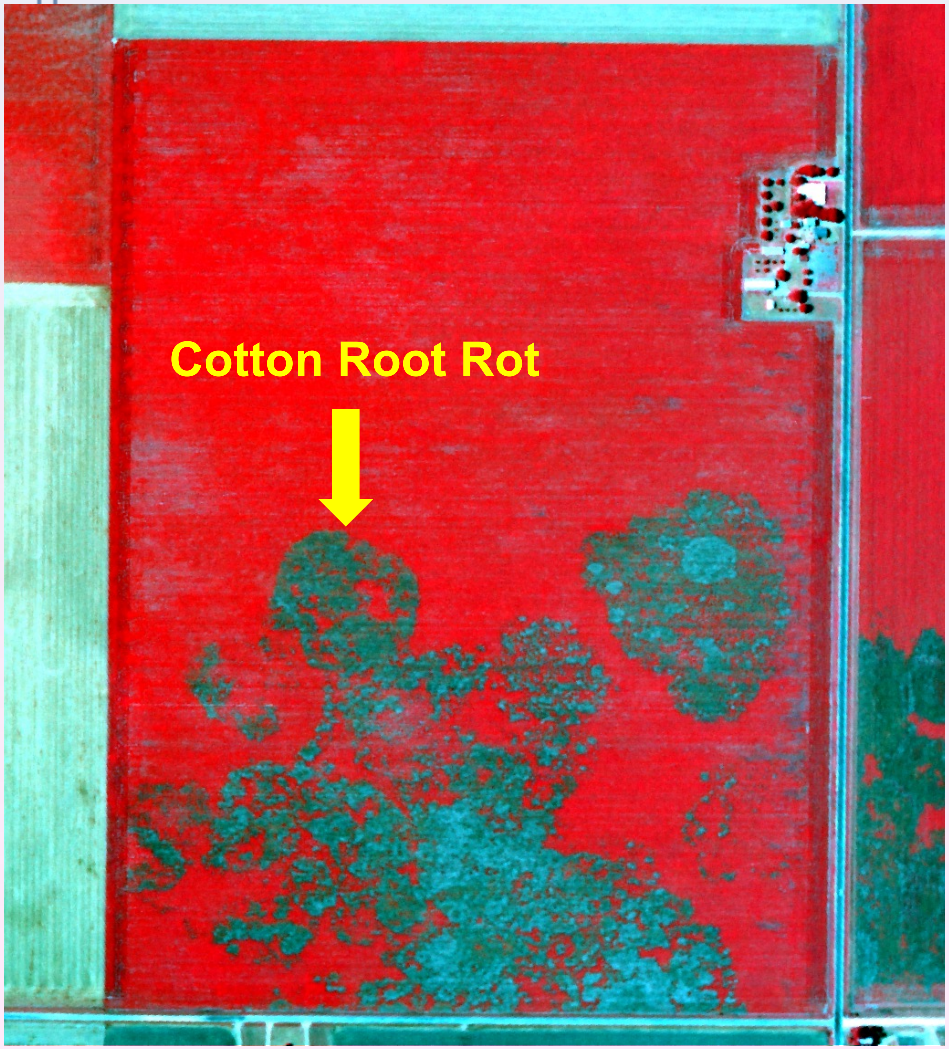


**Cotton root rot**

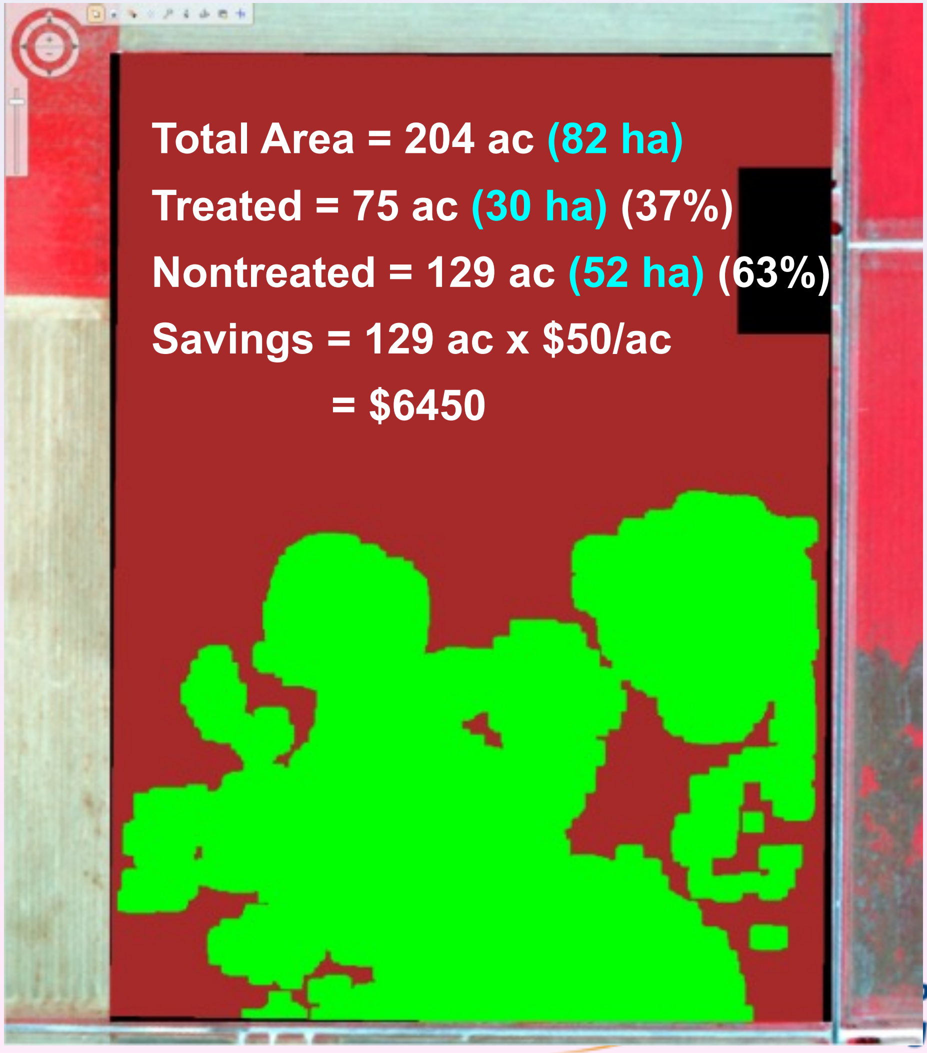
Aerial Application Technology

# CIR Image and Prescription Map

**CIR Image**




**Prescription Map**



# Challenges in Crop Identification and Disease Detection

- Crop Identification (Supervised Learning):
  - Field variability: Crop images can vary depending on growth stages, management practices, and weather conditions.
  - Model generalizability: Overfitting, underfitting, and regional differences in cropping conditions can limit performance across environments.
- Disease Detection (Unsupervised Learning):
  - Noise and variability: Healthy plants can have features that resemble diseases, leading to false positives.
  - Coexisting stresses: Multiple stresses (biotic and abiotic) often co-occur, and their symptoms can overlap, making it challenging to isolate the exact cause through remote sensing data alone.
  - Lack of ground truth: Difficult to assign the clusters to actual disease classes without large and high-quality ground datasets.

# Future Directions and Trends in AI for Agriculture

- **Improved Data Fusion Techniques:** Combining satellite, aerial, and ground-level data enhances the accuracy of crop classification and disease detection models.
- **Integration of AI and IoT:** Real-time disease detection using smart sensors and machine learning models.
- **Autonomous Crop Monitoring:** Drones and robots equipped with ML models to automatically identify crops and detect diseases.
- **Improved Algorithms:** Hybrid models that combine supervised and unsupervised learning for more robust predictions.
- **Self-Learning Models with Transfer Learning:** Transfer learning allows models trained on one crop or region to be adapted for other crops or areas with minimal additional data, enhancing model versatility.
- **Deep Learning for Complex Disease Identification:** Advanced neural networks, especially convolutional neural networks (CNNs), are being fine-tuned to identify complex disease patterns and subtle symptoms in crops.
- **Precision Agriculture:** Use of AI to optimize irrigation, fertilization, and pest control.

Aerial Application Technology

# Takeaways: Machine Learning is Changing Agriculture

- Machine Learning offers powerful tools for improving agricultural practices, from identifying crops to detecting disease.
- Supervised Learning is ideal for crop classification, while unsupervised learning is effective for disease and anomaly detection.
- The future of agriculture will rely on AI-driven solutions to increase efficiency, reduce costs, and ensure better crop management.
- AI presents both opportunities and challenges, it is crucial to approach its adoption thoughtfully.

# Questions?

Thank You!

chenghai.yang@usda.gov