# Mapping Vegetation Dynamics in Wyoming: A Multi-Temporal Analysis using Landsat NDVI and Clustering

Nandini Kuppala[1], Chukka Navneet Krishna[1], Sajith Variyar V.V[1], and Ramesh Sivanpillai[2]

[1]Amrita School of Artificial Intelligence, Coimbatore, Amrita Vishwa Vidyapeetham, India.
(knandini7816@gmail.com, ORCID: 0009-0008-8746-1227; navneetkrishna.918@gmail.com, ORCID: 0009-0003-7067-9645; vv_sajithvariyar@cb.amrita.edu, ORCID: 0000-0003-3944-8155)
[2]Wyoming GIS Center, School of Computing, University of Wyoming, Laramie, WY, 82071, USA.
(sivan@uwyo.edu, ORCID: 0000-0003-3547-9464)

**Keywords:** Landsat, non-irrigated pasture, Normalized Difference Vegetation Index (NDVI), Gaussian Mixture Models, Affinity Propagation.

**Abstract**

This research focused on mapping vegetation growth patterns in non-irrigated fields using five early season (2019-2023) Normalized Difference Vegetation Index (NDVI) derived from Landsat imagery. As part of this study, we compared the outputs generated by two unsupervised machine learning algorithms with a conventional image clustering technique. NDVI data were screened for outliers using the interquartile range method. Gaussian Mixture Models (GMM), Affinity Propagation (AP) and a traditional rule-based classification were used for clustering the pixels in the five NDVI images. GMM assigns data points probabilistically, assuming data are generated from a mixture of Gaussian distributions, while AP identifies clusters by finding representative exemplars without needing a predefined number of clusters. To evaluate the performance of these clustering algorithms, we assigned the clusters into six classes based on vegetation growth patterns observed over the five-year period. Class 1 represents five years of good vegetation growth, class 2 represents four years, etc. We used Intersection over Union (IoU) score to measure how well the classes represented in the final products compared to each other. When compared to rule-based classification product, AP generated product had an aggregate IoU score of 0.63, while GMM generated product had 0.59. GMM and AP detected finer NDVI variations that the rule-based method missed. AP's exemplar-based approach provided a better understanding of vegetation pattern compared to GMM and rule-based method. This study highlights the importance of using advanced clustering techniques over traditional approaches for vegetation analysis, contributing to improved environmental monitoring and management decisions.

## 1. Introduction

Monitoring vegetation growth is crucial for understanding environmental changes and promoting sustainable land management. In recent years, many studies have explored the use of advanced machine learning techniques, such as Random Forest (RF), Classification and Regression Trees (CART), and Support Vector Machines (SVM), for classifying vegetation. These approaches have proven effective, but most research has focused on areas with accessible land-use data, such as urban or well-studied regions, leaving significant gaps in our understanding of underrepresented areas (Ouchra et al., 2023; Li et al., 2021). Addressing these gaps is essential to advancing our global understanding of land use and vegetation dynamics.

This study aims to fill one such gap by focusing on non-irrigated pastures—regions where vegetation growth is solely influenced by natural environmental factors, without the influence of irrigation. By examining growth patterns in these areas, we aim to gain insights into how vegetation responds to natural conditions and better understand these ecosystems.

The primary goal of this study is to map vegetation growth in non-irrigated pastures using early-season NDVI images from Landsat satellites, collected between 2019 and 2023. Satellite image classification plays a key role in categorizing land cover types, which helps extract valuable spatial patterns and insights, enabling further applications such as vegetation analysis (A. K. et al., 2023). NDVI, a widely used vegetation health indicator, is calculated as the difference between the near-infrared (NIR) and red spectral bands, normalized by their sum (Radhakrishnan et al., 2022).

This study compares advanced machine learning clustering techniques (GMM and AP) with traditional rule-based classification methods to assess their effectiveness in vegetation mapping. Traditional methods often involve manual feature extraction by domain experts, a time-consuming and complex process (Reddy et al., 2023). By evaluating these approaches, the study aims to identify the most effective method for detecting vegetation patterns and support improvements in land management for non-irrigated regions.

### 1.1 Clustering

Clustering is an unsupervised technique that groups data based on similarities without prior knowledge. Unsupervised classification is commonly used in vegetation mapping due to its simplicity and wide availability in image processing tools (Langley et al., 2001). In this study, we applied GMM and AP to analyze stacked NDVI images from 2019-2023. These methods, which do not rely on prior knowledge of the phenomena, automatically extract patterns from raw data, provided classification accuracy is high (Hsu et al., 2012).

### 1.2 Overview of Gaussian Mixture Model

The GMM is a probabilistic clustering method that assumes data points come from multiple Gaussian distributions, each

with its own parameters. It effectively models complex cluster shapes and provides probabilistic assignments, allowing data points to belong to multiple clusters with certain probabilities. This makes GMM useful for handling overlapping clusters.

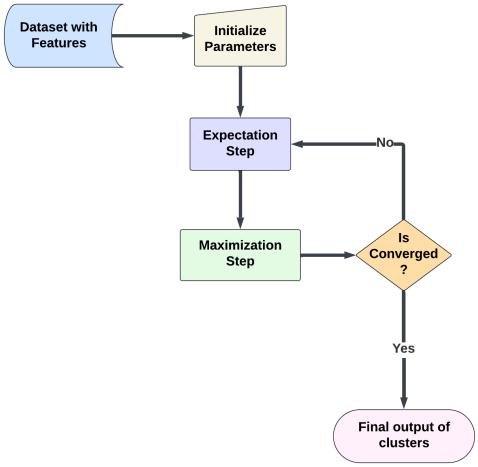


Figure 1. Workflow of GMM

The GMM process begins with the initialization of the parameters. First, the desired number of clusters, denoted as $K$, is determined, typically through prior knowledge or model selection criteria such as the Bayesian Information Criterion (BIC). Next, initial values for the cluster parameters are randomly chosen.

These parameters include the mean ($\mu_k$), representing the center of each cluster, the covariance ($\Sigma_k$), representing the shape or spread of each cluster, and the mixing coefficient ($\pi_k$), indicating the relative contribution of each cluster to the total model. The core of the GMM fitting process is the Expectation-Maximization (EM) algorithm, an iterative optimization technique that alternates between the Expectation (E-Step) and Maximization (M-Step) until convergence.

**E-Step**

In the E-Step, the algorithm calculates the "responsibility" $\gamma_{ik}$, which is the probability that a given data point $x_i$ belongs to each cluster $k$, based on the current estimates of the parameters. This responsibility is computed using the formula:

$$\gamma_{ik} = \frac{\pi_k \mathcal{N}(x_i \mid \mu_k, \Sigma_k)}{\sum_{j=1}^{K} \pi_j \mathcal{N}(x_i \mid \mu_j, \Sigma_j)} \tag{1}$$

where $\mathcal{N}(x_i \mid \mu_k, \Sigma_k)$ represents the Gaussian probability density function for cluster $k$, evaluated at data point $x_i$, and $\pi_k$ denotes the mixing coefficient for cluster $k$.

**M-Step**

In the M-Step, the parameters are updated to maximize the likelihood of the data given the responsibilities calculated in the E-Step. The parameters updated include the mixing coefficients, means, and covariances for each cluster as follows.

The mixing coefficient $\pi_k$ is updated to reflect the proportion of data points assigned to each cluster:

$$\pi_k = \frac{1}{N} \sum_{i=1}^{N} \gamma_{ik} \tag{2}$$

where $N$ is the total number of data points.

The mean $\mu_k$ is updated to represent the weighted average of the data points assigned to cluster $k$:

$$\mu_k = \frac{\sum_{i=1}^{N} \gamma_{ik} x_i}{\sum_{i=1}^{N} \gamma_{ik}} \tag{3}$$

The covariance matrix $\Sigma_k$ is adjusted to capture the variance of data points in cluster $k$:

$$\Sigma_k = \frac{\sum_{i=1}^{N} \gamma_{ik} (x_i - \mu_k)(x_i - \mu_k)^T}{\sum_{i=1}^{N} \gamma_{ik}} \tag{4}$$

The E-Step and M-Step are repeated iteratively until the change in parameter values falls below a specified threshold, signifying that the model has converged and reached an optimal configuration. At this point, the Gaussian Mixture Model is fully trained and can be applied to predict the cluster membership of new data points based on their likelihood under each Gaussian component.

### 1.3 Overview of Affinity Propagation

AP is a clustering algorithm that groups data points by identifying representative exemplars, or cluster centers, from the data itself rather than requiring a predetermined number of clusters. Unlike other clustering methods that rely on iterative updates of cluster centroids, AP operates through a unique message-passing mechanism where data points exchange messages regarding their suitability as potential exemplars. This method is advantageous in scenarios where the optimal number of clusters is unknown, as it automatically determines both the number of clusters and their structure.

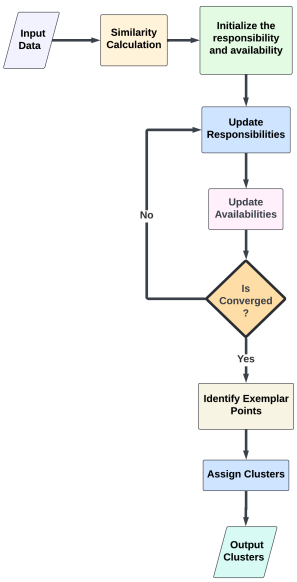


Figure 2. Workflow of Affinity Propagation

**Similarity Calculation**

The similarity between two data points $x(i)$ and $x(k)$ quantifies their closeness, which can be expressed as the negative squared Euclidean distance between them. This similarity score, $S(i,k)$, forms the basis for determining if one point can serve as an exemplar for the other:

$$S(i,k) = -\|x(i) - x(k)\|^2 \tag{5}$$

where $\|x(i)-x(k)\|^2$ is the squared Euclidean distance between points $x(i)$ and $x(k)$. Here, a lower distance (and thus higher similarity) implies a greater likelihood for one point to serve as the exemplar for the other.

**Responsibility Update**

The responsibility $r(i,k)$ reflects the suitability of point $k$ to be the exemplar for point $i$. This value is calculated by considering the similarity $s(i,k)$ between points $i$ and $k$ and subtracting the influence of alternative exemplars for point $i$. Mathematically, the responsibility is updated as:

$$r(i,k) = s(i,k) - \max_{k' \neq k}\{a(i,k') + s(i,k')\} \tag{6}$$

Here, $s(i,k)$ is the similarity between points $i$ and $k$, while the term $\max_{k' \neq k}\{a(i,k') + s(i,k')\}$ represents the highest combined availability and similarity score for alternative exemplars $k'$. By subtracting this maximum value, the algorithm ensures that $r(i,k)$ is high only if $k$ is a strong candidate compared to other potential exemplars for $i$.

**Availability Update**

The availability $a(i,k)$ indicates how appropriate point $k$ is as an exemplar, given the responsibilities of other points. Availability is influenced by both the self-responsibility of $k$ (indicating whether $k$ prefers itself as an exemplar) and the aggregate responsibility from other points toward $k$. The availability is calculated as:

$$a(i,k) = \min\left(0, r(k,k) + \sum_{i' \neq i} \max(0, r(i',k))\right) \tag{7}$$

In this formula, $r(k,k)$ denotes the self-responsibility of point $k$, showing how strongly $k$ considers itself an exemplar. The term $\sum_{i' \neq i} \max(0, r(i',k))$ adds the positive responsibilities of all other data points $i'$ toward $k$. By taking the minimum of this sum with zero, the algorithm ensures that the availability remains constrained and non-positive, emphasizing cases where $k$ is truly a suitable exemplar for $i$.

**Convergence and Cluster Assignment**

The iterative updates of responsibilities and availabilities continue until the algorithm stabilizes, or converges, meaning that the changes in these values across iterations become minimal. Upon convergence, exemplar points (i.e., cluster centers) are identified based on the final values of the responsibility and availability matrices. Cluster assignments are then determined by linking each data point to its corresponding exemplar, resulting in the final cluster structure without requiring the number of clusters as a predefined parameter.

## 2. Experiments and Methods

### 2.1 Study Area and Data Acquisition:

Non-irrigated pastures located in Albany County (Wyoming) were selected for this study. Winter precipitation in the form of snow was the primary source of moisture for these fields. Cloud-free satellite imagery from Landsat 8 and Landsat 9, covering two fields over a five-year period (2019–2023) during the early growing season in July, was acquired for analysis. The spatial resolution of the images is 30 m by 30 m. Quality assurance and quality control measures were implemented during the acquisition process to ensure the reliability and accuracy of the data.

The Landsat imagery underwent visual inspection to identify and mitigate potential artifacts, such as cloud cover, atmospheric distortion, and sensor anomalies. Cloud-free scenes were prioritized to minimize data contamination and ensure the integrity of the dataset. From this Landsat dataset, NDVI data covering the same five-year period was derived for both fields and was used for further research. NDVI serves as a pivotal indicator of vegetation health and density, offering valuable insights into temporal changes in vegetation dynamics.

### 2.2 Preprocessing and outlier removal

The NDVI images spanning the years 2019-2023 of two fields underwent preprocessing and outlier removal to ensure data quality and consistency for subsequent analysis. NDVI values were scaled down by dividing each pixel value by 10,000, standardizing the data for comparability across different images and datasets. Pixels with invalid or unreliable NDVI values, such as those affected by non vegetation, sensor noise or atmospheric interference, were masked. NDVI values below 0 were replaced with a null value (None) to denote missing or invalid data.

To identify outliers in the NDVI data, a robust statistical method based on the interquartile range (IQR) was applied. Outliers were defined as pixel values falling outside the range (Figure 4):

$$\text{Lower bound} = Q1 - 1.5 \times \text{IQR} \tag{8}$$

$$\text{Upper bound} = Q3 + 1.5 \times \text{IQR} \tag{9}$$

where $Q1$ and $Q3$ represent the first (25th percentile) and third (75th percentile) quartiles, respectively, and IQR is the interquartile range.

For each year's NDVI image, the 25th percentile ($Q1$) and 75th percentile ($Q3$) values were calculated. These values were then averaged across all five years to establish consistent threshold values. Specifically:

$$\text{IQR} = Q3_{\text{avg}} - Q1_{\text{avg}} \tag{10}$$

where $Q1_{\text{avg}}$ represents the average of the 25th percentile values and $Q3_{\text{avg}}$ represents the average of the 75th percentile values. Subsequently, the average lower and upper outlier thresholds were calculated using the following formulas:

$$\text{Average lower threshold} = Q1_{\text{avg}} - 1.5 \times \text{IQR} \tag{11}$$

$$\text{Average upper threshold} = Q3_{\text{avg}} + 1.5 \times \text{IQR} \tag{12}$$

This method removes extreme values, focusing on the middle 50% of the data to ensure an accurate representation of central tendency. The IQR approach is particularly robust against skewed distributions (Figure 3), offering more reliable boundaries for identifying outliers compared to methods based solely on the mean and standard deviation.

This preprocessing step was crucial for cleaning and normalizing the NDVI data, significantly enhancing the reliability and validity of the clustering analysis that followed.

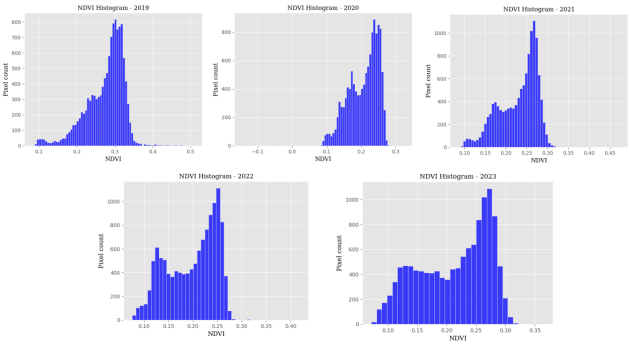


Figure 3. Histogram of NDVI values (2019-2023), showing noticeable skewness.

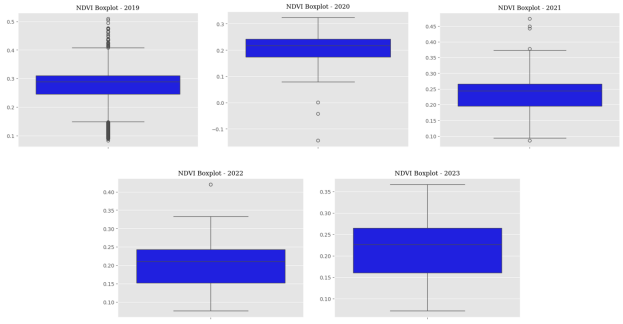


Figure 4. Box plot showing NDVI distribution (2019-2023), highlighting outliers beyond the whiskers (IQR) and they are removed to maintain dataset integrity for analysis.

### 2.3 Calculation of Mean NDVI Over Five Years

After preprocessing, the mean NDVI value was calculated for each NDVI image in the dataset. This involved computing the average NDVI value across all pixels within each image, providing an overall representation of vegetation health for each temporal period. Using the calculated mean NDVI values for each image, five binary maps were generated to visualize vegetation health. The five NDVI images were stacked to create a composite image encompassing the entire study period (Figure 5). Subsequently, the mean NDVI value for each pixel across the five years was computed, providing insights into vegetation dynamics over the entire time span. Finally, a binary map was generated based on the stacked average NDVI value (Figure 6).

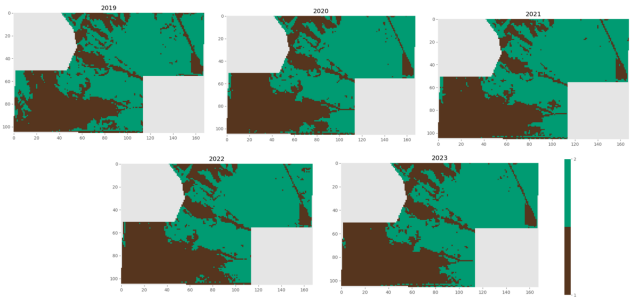


Figure 5. NDVI pixel values across the study period (2019-2023). Green pixels highlight areas with more vegetation and brown pixels indicate less vegetation or barren land.

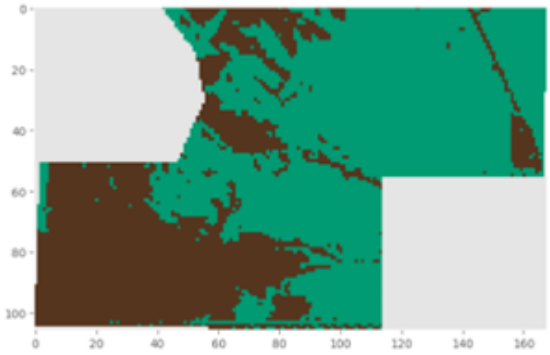

Figure 6. Stacked NDVI values (2019-2023) show green for areas with consistently high vegetation and brown for areas with lower vegetation health over time.

### 2.4 Gaussian Mixture Models

A key aspect of the GMM algorithm is the requirement to specify the number of clusters beforehand. In our study, the number of components was determined using the Bayesian Information Criterion (BIC). BIC balances model fit and complexity, aiding in the selection of an optimal number of clusters. After extensive experimentation and evaluation, the GMM was configured with 68 components. This number was chosen because it provided a good balance between capturing the complexity of the data and avoiding overfitting. Additionally, to ensure reproducibility, the model was initialized with a random state of 42, a commonly used value that provides a consistent starting point for random number generation, ensuring that the results can be replicated across different runs.

The fitting process involves the Expectation-Maximization (EM) algorithm, which iteratively updates the parameters of the Gaussian distributions to maximize the likelihood of the observed data. After fitting the model, each pixel in the NDVI image

was assigned a cluster label based on the highest probability of cluster membership. This process segments the NDVI image into 68 distinct regions, each representing a unique cluster with specific vegetation characteristics. This allowed for a detailed and nuanced clustering of the NDVI data, capturing the variability in vegetation patterns.

### 2.5 Affinity Propagation

A key aspect of the AP algorithm is the preference parameter, which influences the number of exemplars chosen by the algorithm. A higher preference value results in more clusters, while a lower value results in fewer clusters. Additionally, the damping factor helps stabilize the algorithm by controlling the update rate of responsibilities and availabilities during the clustering process. In our study, a damping factor of 0.5 was used.

The implementation process involved normalizing the stacked NDVI data using StandardScaler to facilitate the convergence of the AP algorithm, applying AP with a preference parameter of -10, and then assigning cluster labels to each pixel in the NDVI image based on the identified exemplars. This approach enabled the identification of clusters without prior knowledge of their number, leveraging the algorithm's ability to adapt to the data's intrinsic structure.

### 2.6 Comparison of GMM and AP

In our analysis, we used both Gaussian Mixture Model (GMM) and Affinity Propagation (AP) to assign cluster labels to pixels in the stacked NDVI image, facilitating a detailed examination of spatial vegetation patterns. Each algorithm offers distinct advantages based on data characteristics.

GMM works best when the number of clusters is known beforehand and when data follows a Gaussian distribution. It models clusters as overlapping Gaussian components, making it ideal for NDVI imagery with regular patterns and slight overlaps, allowing for precise classification.

On the other hand, AP determines the number of clusters automatically by analyzing the data's structure. It identifies cluster centers, or exemplars, through an iterative process, making it well-suited for NDVI datasets with irregular or complex patterns.

### 2.7 Cluster Analysis

In this section, we delve deeper into the analysis of the clusters formed by GMM and AP. The cluster labels assigned to each pixel allowed us to segment the NDVI data spatially and temporally, providing insights into vegetation dynamics over the study period. The statistical metrics calculated for each cluster, including Mean (Figure 7, Figure 8) of NDVI values, were crucial for characterizing the health and variability of vegetation within each cluster.

The clusters identified by GMM and AP captured subtle variations in NDVI values that were not evident in the binary classification, highlighting the importance of using advanced clustering techniques for vegetation analysis.

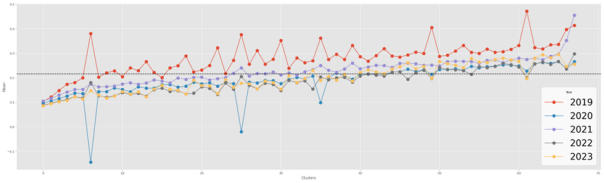


Figure 7. The plot uses five lines to show yearly mean values (2019-2023) for individual clusters calculated by AP, with a dashed black line marking the overall average of 0.2151.

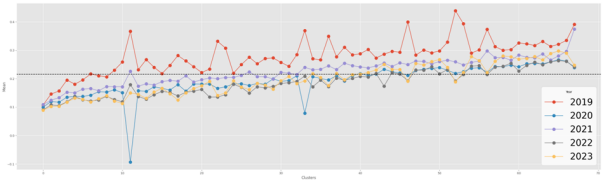


Figure 8. The plot uses five lines to show yearly mean values (2019-2023) for individual clusters calculated by GMM, with a dashed black line marking the overall average of 0.2154.

Using the AP clustering method, the mean NDVI values for each cluster were tracked annually. The data reveals intriguing trends, particularly the variations in cluster behavior between 2019 and subsequent years.

As shown in Figure 7 , the mean NDVI values are represented as distinct lines for each year, with a dashed black line indicating the cutoff value of 0.2151. The red line, representing 2019, displays significant fluctuations, with pronounced peaks and valleys that mostly remain above the cutoff. This suggests that 2019 experienced higher variability in vegetation health or environmental conditions. From 2020 onward, however, the trends smooth out, indicating a more stable pattern of vegetation dynamics, likely influenced by consistent climatic or land management factors.

Similarly, GMM clustering method provides another perspective on these dynamics. The mean NDVI values for 2019–2023, grouped using GMM, are illustrated in Figure 8 . Here, the cutoff value is slightly higher at 0.2154, derived from the algorithm's calculations. Again, 2019 shows the greatest variability, with its mean NDVI values fluctuating significantly yet largely staying above the cutoff. From 2020 to 2023, the trends become steadier, with minimal variability, reinforcing the observations from the AP analysis.

These findings, consistent across both clustering methods, highlight how vegetation clusters have changed over time. The variability observed in 2019 may reflect external factors such as extreme weather events or land-use changes, while the stability in later years could indicate improved environmental or management practices.

### 2.8 Tools and Libraries

The analysis was performed using Python, with key libraries including OpenCV for image processing, NumPy for numerical operations, scikit-learn for clustering algorithms, and Matplotlib and Seaborn for visualization. These tools provided a comprehensive environment for data analysis and visualization, ensuring reproducibility and accuracy.

This contribution has been peer-reviewed.
https://doi.org/10.5194/isprs-archives-XLVIII-M-5-2024-87-2025 | © Author(s) 2025. CC BY 4.0 License. 

## 3. Results and Discussion

### 3.1 Affinity Propagation

Figure 9. Binary Map generated using AP. Cluster averages using AP were calculated, and their overall average was used as a cutoff to divide clusters into high (green) and low (brown) classes.

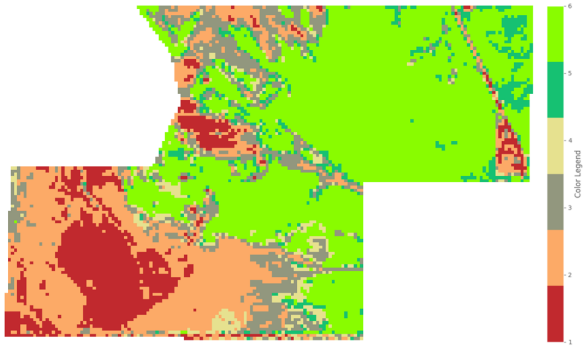

Figure 10. Map generated using AP. Light green shows pixels above the cutoff for all five years, dark green for one year above, and red for all years below, with other colors following the same pattern.

Using a threshold of 0.215, the analysis identified six clusters with varying pixel counts and areas (Figure 10). Class 1 covers 1,306,800 m² (322.92 acres) with 1452 pixels, Class 2 spans 2,324,700 m² (574.45 acres) with 2583 pixels, and Class 3 totals 1,342,800 m² (331.81 acres) with 1492 pixels. Class 4 occupies 444,600 m² (109.86 acres) with 494 pixels, Class 5 covers 594,000 m² (146.78 acres) with 660 pixels, and the largest, Class 6, spans 5,114,700 m² (1263.87 acres) with 5683 pixels.

| Class | Pixel Count | Area (*square meters*) | Area (*acres*) |
|---|---|---|---|
| 1 | 1452 | 1,306,800 | 322.92 |
| 2 | 2583 | 2,324,700 | 574.45 |
| 3 | 1492 | 1,342,800 | 331.81 |
| 4 | 494 | 444,600 | 109.86 |
| 5 | 660 | 594,000 | 146.78 |
| 6 | 5683 | 5,114,700 | 1263.87 |

Table 1. Class-wise pixel count, area in square meters, and area in acres

### 3.2 Gaussian Mixture Model

Figure 11. Binary Map generated using GMM. Cluster averages using GMM were calculated, and their overall average was used as a cutoff to divide clusters into high (green) and low (brown) classes.

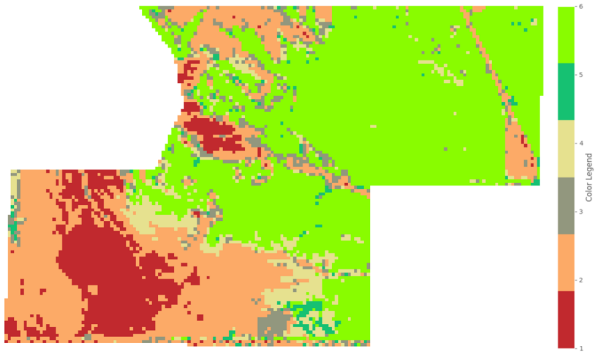

Figure 12. Map generated using GMM. Light green shows pixels above the cutoff for all five years, dark green for one year above, and red for all years below, with other colors following the same pattern.

With a threshold of 0.215, the analysis of the imagery reveals a total of 12,364 pixels classified across six categories (Figure 12). Class 1 occupies 1,375 pixels, totaling 1,237,500 m² (305.79 acres). Class 2 encompasses 3,179 pixels, covering 2,861,100 m² (706.99 acres). Class 3 spans 679 pixels, corresponding to 611,100 m² (151.01 acres). Class 4 covers 811 pixels, with an area of 729,900 m² (180.36 acres). Class 5 consists of 197 pixels, totaling 177,300 m² (43.81 acres). Finally, Class 6 is represented by 6,123 pixels, equating to 5,510,700 m² (1,361.72 acres). All of these values are summarized in Table 2.

| Class | Pixel Count | Area (*square meters*) | Area (*acres*) |
|---|---|---|---|
| 1 | 1375 | 1237500 | 305.79 |
| 2 | 3179 | 2861100 | 706.99 |
| 3 | 679 | 611100 | 151.01 |
| 4 | 811 | 729900 | 180.36 |
| 5 | 197 | 177300 | 43.81 |
| 6 | 6123 | 5510700 | 1361.72 |

Table 2. Class-wise pixel count, area in square meters, and area in acres

### 3.3 Rule Based

Figure 13. Binary Map generated using rule-based method, with the cutoff determined by the overall mean of pixel averages across five years, classifying them into high (green) and low (brown) classes.

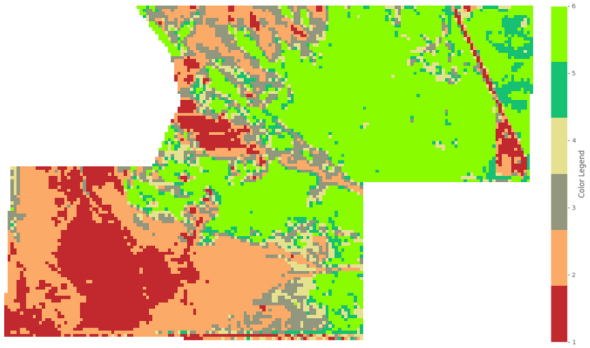

Figure 14. Map generated using rule-based method, with the cutoff based on the average of yearly pixel means. Light green shows pixels above the cutoff for all five years, dark green for one year above, and red for all years below, with other colors following the same pattern.

In the analysis of the imagery with a threshold of 0.224, the classification results indicate significant areas for each class (Figure 14): Class 1 covers 1,906 pixels, totaling 1,715,400 m² (423.88 acres). Class 2 comprises 2,775 pixels, corresponding to 2,497,500 m² (617.15 acres). Class 3 occupies 1,309 pixels, equating to 1,178,100 m² (291.11 acres). Class 4 encompasses 763 pixels, with an area of 686,700 m² (169.69 acres). Class 5 spans 753 pixels, covering 677,700 m² (167.46 acres). Finally, Class 6 is the largest with 4,858 pixels, covering 4,372,200 m² (1,080.39 acres). All of these values are summarized in Table 3.

| **Class** | **Pixel Count** | **Area (*square meters*)** | **Area (*acres*)** |
|---|---|---|---|
| 1 | 1906 | 1715400 | 423.88 |
| 2 | 2775 | 2497500 | 617.15 |
| 3 | 1309 | 1178100 | 291.11 |
| 4 | 763 | 686700 | 169.69 |
| 5 | 753 | 677700 | 167.46 |
| 6 | 4858 | 4372200 | 1080.39 |

Table 3. Class-wise pixel count, area in square meters, and area in acres

### 3.4 Rule-Based Classification vs. Clustering Methods

The rule-based classification approach offered a straightforward and interpretable method for classifying NDVI values into high and low categories. It relied on a predefined threshold (0.22423 mean of stacked image) and did not involve iterative clustering processes. This method was efficient and easy to implement but lacked the ability to capture complex patterns and interactions within the data.

In comparison, both GMM and Affinity Propagation provided more sophisticated clustering results, capturing the underlying structure of the NDVI data. GMM's probabilistic framework and Affinity Propagation's exemplar-based approach offered detailed insights into vegetation health, surpassing the binary classification of the rule-based method.

### 3.5 Model Evaluation Using IOU :

| **Label** | **IOU** |
|---|---|
| 1 | 0.78 |
| 2 | 0.03 |
| 3 | 0.16 |
| 4 | 0.32 |
| 5 | 0.66 |
| 6 | 0.66 |
| Aggregate IOU: **0.59** | |

Table 4. IoU scores were obtained by comparing the Rulebased and GMM classified images.

| **Label** | **IOU** |
|---|---|
| 1 | 0.85 |
| 2 | 0.18 |
| 3 | 0.15 |
| 4 | 0.43 |
| 5 | 0.65 |
| 6 | 0.76 |
| Aggregate IOU: **0.64** | |

Table 5. IoU scores were obtained by comparing the RuleBased and AP classified images.

| **Label** | **IOU** |
|---|---|
| 1 | 0.87 |
| 2 | 0.06 |
| 3 | 0.19 |
| 4 | 0.32 |
| 5 | 0.67 |
| 6 | 0.73 |
| Aggregate IOU: **0.65** | |

Table 6. IoU scores were obtained by comparing the AP and GMM classified images.

Intersection over Union calculates the overlap between the predicted areas with the actual areas to assess the degree of their similarity(S. Aswin et al., 2021) [1]. This helps us compare advanced algorithms with traditional methods to see which one is best for analyzing non-irrigated fields. We calculated IoU scores for each label (1-6) across three comparisons: RuleBased vs. GMM (Table 4), RuleBased vs. AP (Table 5), and AP vs. GMM (Table 6), along with the aggregate IoU scores. The results showed that label 1 and 6 had consistently high IoU scores across all three comparisons. For RuleBased vs GMM(Table 4), the highest IoU score is 0.78 for label 1, followed by label 6 with a score of 0.66, which has remained below the cutoff. For RuleBased vs. AP(Table 5), the highest match is 0.85 for label 1, followed by 0.76 for label 6. In the AP vs. GMM comparison(Table 6), label 1 again has the highest match at 0.87, with

label 6 at 0.73. These results suggest that areas with maximum vegetation (above the cutoff) over the past five years were effectively detected by the algorithms, with the second highest also performing well. The highest aggregate IoU score, 0.65, is between AP and GMM

### 3.6 Conclusion

In conclusion, our study provided valuable insights into vegetation growth patterns in non-irrigated fields using Landsat NDVI imagery. While Rule-Based Classification and Gaussian Mixture Models each offered useful perspectives, Affinity Propagation stood out as the most effective method.

The results indicate that advanced clustering techniques can be more beneficial than traditional methods, as they are more effective and easier to implement on new datasets. The cluster maps produced in this study can assist landowners in identifying areas of good growth and less fertile regions. This allows them to focus resources on more productive areas and reduce operational costs in less fertile zones. Ultimately, our study offered valuable insights into vegetation dynamics and supported better land management practices.

## 4. References

A. K., B. R., and P. S. Neha, 2023: Remote Sensing Image Classification Based on Confidence Score of Ensemble Machine Learning Classifiers. *Int. Conf. on Evolutionary Algorithms and Soft Computing Techniques (EASCT)*, Bengaluru, India, 2023, pp. 1-6. doi.org/10.1109/EASCT59475.2023.10392571.

S. Aswin et al., 2021: Effect of Annotation and Loss Function on Epiphyte Identification using Conditional Generative Adversarial Network. *Int. Conf. on Advances in Electrical, Computing, Communication and Sustainable Technologies (ICAECT)*, Bhilai, India, 2021, pp. 1-6. doi.org/10.1109/ICAECT49130.2021.9392478.

D. Hsu, S. M. Kakade, and T. Zhang, 2012: A spectral algorithm for learning Hidden Markov Models. *J. of Computer and System Sciences*, vol. 78, no. 5, pp. 1460-1480. doi.org/10.1016/j.jcss.2011.12.025.

S. K. Langley, H. M. Cheshire, and K. S. Humes, 2001: A comparison of single date and multitemporal satellite image classifications in a semi-arid grassland. *J. of Arid Environments*, vol. 49, no. 2, pp. 401-411. doi.org/10.1006/jare.2000.0771.

Y. Li, C. Liu, J. Zhang, P. Zhang, and Y. Xue, 2021: Monitoring Spatial and Temporal Patterns of Rubber Plantation Dynamics Using Time-Series Landsat Images and Google Earth Engine. *IEEE J. of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 14, pp. 9450-9461. doi.org/10.1109/JSTARS.2021.3110763.

S. K. Manoharan, R. K. Megalingam, and S. P. Sreeji, 2024: Machine Learning Based Classification of Coconut Trees Based on Tree Parameters. 2nd World Conf. on Communication & Computing (WCONF), Raipur, India, 2024, pp. 1-6. doi.org/10.1109/WCONF61366.2024.10692268.

H. Ouchra, A. Belangour, and A. Erraissi, 2023: Machine Learning Algorithms for Satellite Image Classification Using Google Earth Engine and Landsat Satellite Data: Morocco Case Study. *IEEE Access*, vol. 11, pp. 71127-71142. doi.org/10.1109/ ACCESS.2023.3293828.

S. Radhakrishnan and P. Geetha, 2022: Urban Sprawl Assessment Using Remote Sensing and GIS Techniques: A Case Study of Ernakulam District. In A.K. Nagar, D.S. Jat, G. Marín-Raventós, and D.K. Mishra (eds), *Intelligent Sustainable Systems*, Lecture Notes in Networks and Systems, vol. 334. Springer, Singapore. doi.org/10.1007/978-981-16-6369-7_26.

O. C. Reddy, I. D. Kumar, P. Sathvika, V. V. Sajith Variyar, V. Sowmya, and R. Sivanpillai, 2023: Effect of hyperparameters on DeeplabV3+ performance to segment water bodies in RGB images. *Int. Arch. Photogramm. Remote Sens. Spatial Inf. Sci.*, vol. XLVIII-M-3-2023, pp. 203–209. doi.org/10.5194/isprs-archives-XLVIII-M-3-2023-203-2023.