NOVA IMS Information Management School

MDSAA

Master's Degree Program in

Data Science and Advanced Analytics

Comparison of Supervised Image Classification Algorithms

Classifying Diverse Land Cover in California

Miriam Natasha Hadidi Pereira

Dissertation

presented as partial requirement for obtaining the Master's Degree Program in Data Science and Advanced Analytics

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

CLASSIFICATION OF DIVERSE LAND COVER IN CALIFORNIA

by

Miriam Natasha Hadidi Pereira

Dissertation presented as partial requirement for obtaining the Master's degree in Advanced Analytics, with a Specialization in Business Analytics

Supervisor: Pedro Cabral

July 2023

STATEMENT OF INTEGRITY

I hereby declare having conducted this academic work with integrity. I confirm that I have not used plagiarism or any form of undue use of information or falsification of results along the process leading to its elaboration. I further declare that I have fully acknowledge the Rules of Conduct and Code of Honor from the NOVA Information Management School.

Miriam Natasha Hadidi Pereira

Lisboa, Portugal, 13th of July, 2023

ACKNOWLEDGEMENTS

I would like to take this opportunity to thank my advisor for gently guiding me through the process and providing helpful feedback while still allowing freedom to explore and learn a great deal about the field. I would also like to thank my husband for his support and thoughtful input during this process.

This study was supported by the research project MaSOT – Mapping Ecosystem Services from Earth Observations, funded by the Portuguese Science Foundation – FCT [EXPL/CTA-AMB/0165/2021].

ABSTRACT

The research field of machine learning and supervised image classification is quickly developing. There are many studies regarding the different use cases of image classification. However, a comprehensive study on the primary algorithms in ArcGIS Pro has not been assessed for numerous classes. This study attempts to bridge that gap by evaluating the effectiveness of the three primary classification algorithms available in ArcGIS Pro, and to determine an optimal algorithm for the given study area. This scope covers 12 classes of land cover in San Joaquin County, California. Maximum Likelihood, Random Forest, and Support Vector Machine were tested based on their general usability in image classification as well as their proven characteristics through research. The training and ground truth validation data were provided by USGS, in the form of a Landsat 8 image, and crop planning map. The accuracy assessment was performed with a stratified random sampling strategy. Based on the Kappa statistic, this study determines Random Forest (Kappa = 0.68, Accuracy = 0.76) to be the most suitable algorithm for detecting a series of crop types, bodies of water, and urban spaces apart from the rest of the land cover in San Joaquin County, California, USA. In addition to determining a preferred algorithm, it is also apparent that certain parameters when tweaked, produce the optimal classifier for this dataset. In this case, this means most parameters set to default, with an increased spectral detail and a decreased spatial detail. What this indicates for crop planning is that the current algorithms used in California are already quite effective at accurately identifying unique types of land cover. This builds confidence in the field, however parameters could be similarly tweaked to produce an even better classification. This study can be useful for improving crop and water planning.

KEYWORDS

Machine Learning; Remote Image Sensing; Geographic Information Systems; Supervised Image Classification Algorithms

Sustainable Development Goals (SGD):

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ANN	Artificial Neural Networks
DT	Decision Trees
DNN	Deep Neural Networks
DWR	Department of Water Resources
FN	False negatives
FP	False positives
GIS	Geographic Information Systems
KNN	K-Nearest Neighbors
MLC	Maximum Likelihood Classifier
PCC	Probability of Correct Classification
PRE	Proportional Reduction in Error
RF	Random Forest
SVM	Support Vector Machine
TN	True negatives
TP	True positives
USGS	United States Geological Survey
UTM	Universal Transverse Mercator
WGS84	World Geodetic System 1984

1. INTRODUCTION & LITERATURE REVIEW

1.1. BACKGROUND

Climate change has drastically altered the Earth's environment (Hürlimann et al., 2022). The state of California in the United States is particularly susceptible to climate change effects (Cavagnaro et al., 2005, p. 9). Research has shown that California is now facing increased volatility and unpredictability in rainfall (Swain et al., 2018) which greatly affects the various ecosystems in the state. When considering that this state produces 50% of the country's fruits and vegetables and employs 1.1 million people (Cavagnaro et al., 2005, p. 3), environmental changes can quickly become major economic problems if not monitored and properly addressed. For these reasons, California has become an interesting study area prospect. The findings from this study can then be applied to other areas with similar climates.

Remote sensing, a process that determines various parts of the Earth's surface, usually based on satellite imagery (*USGS FAQ*, n.d.), has been useful in determining land cover changes in relation to climate change (Bontemps et al., 2011; Yang et al., 2013). Monitoring of the Earth's surface using remote sensing has been a regular practice for ecological organizations because it is efficient and generally accurate enough off which to base decisions (Bontemps et al., 2011). However, as Yang et al. argues, continued research and testing is needed to evaluate the algorithms and their inputs (Yang et al., 2013). The following project aims to perform this evaluation by assessing three supervised classification algorithms in correctly distinguishing 12 types of land cover.

1.2. LITERATURE REVIEW

Various studies have delved into the effectiveness of supervised classification algorithms in determining land cover types and their relevance to the studies of climate change (Bontemps et al., 2011; Khatami et al., 2016; Potapov et al., 2015). The study locations range all over the world and in various types of ecosystems. Their purposes also range from agricultural planning (Ali et al., 2022a; F. Li et al., 2022) to environmental research (Gessner et al., 2013) to general mapping (Abdollahi et al., 2022). Although there is a plethora of research on the topic, many papers found in the field attempt to improve or predict crop yields, for example. Li et al. (2022) focused on a set of cotton fields in Xinjiang, China as their purpose was to model the cotton yield. Ali et al. (2022) also brings to light many researchers who aim to predict crop yield through remote sensing by measuring soil content and moisture levels. Because many of the studies include only one type of land cover in the image, there is even more reason to study the effectiveness of classification algorithms, as homogenous land cover areas tend to result in the most accurate classification (Gessner et al., 2013; Mishra et al., 2019). This means that since it is a more challenging task to classify heterogenous land areas, there are potentially missed insights into how algorithms behave when applied to more varied land cover.

Other algorithms tend to have a strong tradeoff between accuracy and interpretability, such as Decision Trees (DT) (low accuracy, high interpretability) or Deep Neural Networks (DNN) on the other end of the spectrum (high accuracy, low interpretability) (Sheykhmousa et al., 2020). However, Random Forests (RF) and Support Vector Machines (SVM) tend to balance the two measures as can be seen in Sheykhmousa et al.'s figure A in the annex. For this reason, numerous studies have

evaluated RF and SVM against one another as they have excelled in image classification contexts (Sheykhmousa et al., 2020). A great deal of literature compares SVM, RF, and other algorithms, such as K-Nearest Neighbors (KNN) (Thanh Noi & Kappas, 2017) although many exclude MLC. Some of the most extensive studies have even performed comparisons of seven classification algorithms at once (Osisanwo et al., 2017). Although Osisanwo et al.'s study compared Naïve Bayes, it still did not include MLC. Therefore, this paper aims to cover the research gap of evaluating and comparing three of the most common image classification algorithms when using ArcGIS Pro: RF, SVM, and MLC.

1.3. STUDY SITE

The study site consists primarily of San Joaquin County, of the state of California, in the United States. Figure 1, shown below, displays a map of the study area, provided freely by the United States Geological Survey (USGS).

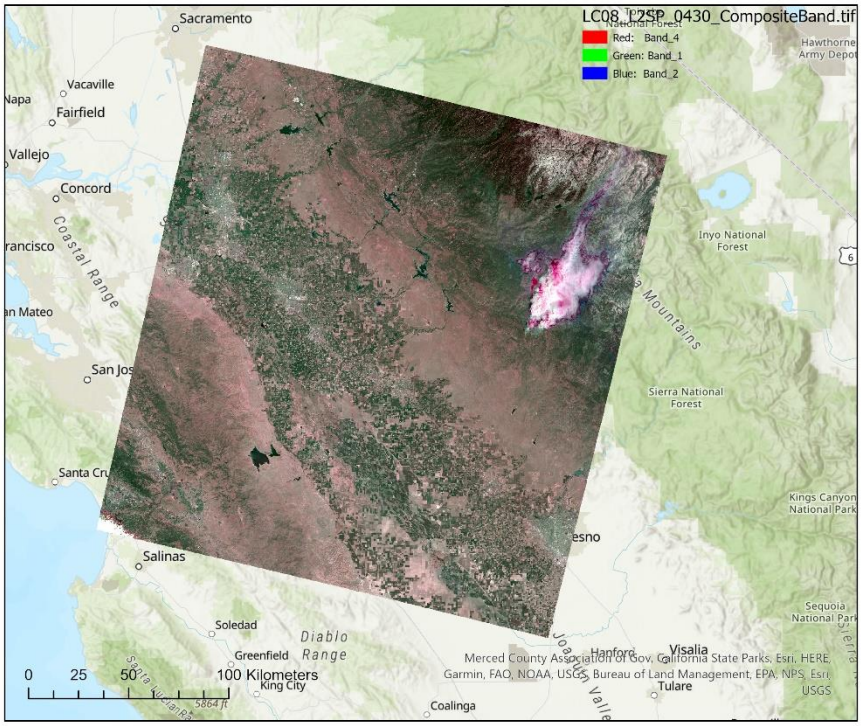

Figure 1: Scope of Satellite Image San Joaquin County, California (USGS, 2022)

There is reliable ground truth data for this county, provided by USGS in 2019, and updated in 2022. Further, this particular county encompasses numerous types of crops as well as bodies of water and urban areas, so more varied land cover will be included in the input, addressing the previously

mentioned research gap. The study attempts to determine the most appropriate algorithm for detection of each crop category. First, the following classes are identified, based on the available crop classes from USGS, which will be used to validate as ground truth data. The 12 classes are defined in Table 1 below.

DESCRIPTION	ABBREVIATION	VALUE
Rice	R	2
Pasture	P	3
Grain and Hay	G	4
Truck Crops	T	5
Field Crops	F	6
Citrus and Subtropical	C	7
Deciduous	D	8
Vineyard	V	9
Water	W	10
Young Perennial	YP	11
Urban	U	12
Unspecified	X	20

Table 1: Land Cover Classes

The results will then be based on the evaluation of the algorithm when looking at accuracy levels regarding all specified classes, including the unspecified class. The unspecified class is a designated class for all other land types which are outside the scope of this study. The crop types were chosen because of the importance of farming in California as well as the availability of accurate and reliable ground truth data. By classifying vegetation in a state with increasingly frequent droughts (Swain et al., 2018), the study will evaluate the performance of each algorithm while contributing to improved monitoring in a volatile state (Swain et al., 2018).

1.4. RESEARCH QUESTION & OBJECTIVES

The research question behind this study is as follows: What is the effectiveness of each tested algorithm in distinguishing classes of land cover when compared with ground truth data?

This then leads to the research objectives, which are two-fold:

1) it aims to evaluate the effectiveness of the three primary classification algorithms available in ArcGIS Pro, and

2) to determine an optimal algorithm for the given study area of San Joaquin County, California, and its respective 12 land cover types, including 9 different crop types.

2. DATA & METHODS

As mentioned in chapter 1, section 1.3 Study Site, there are 12 land cover classes, including water. These land cover classes are based on those available in the 2019 Crop Mapping files (updated most recently prior to the start of this study in August 2022), which will be used as ground truth data in the validation and evaluation phase of this project. The classes were defined by the Department of Water Resources (DWR) for use in crop planning purposes. It is important to note that the initial ground truth files were developed with the Random Forest algorithm. However, equally important is that the DWR had the original classification updated based on local knowledge. Further, this map is regularly updated with the most recent information validated at a ground level (Planning I15_Crop_Mapping_2019, 2022).

2.1. DATA

2.1.1. Criteria

This study will be based on publicly available data from the USGS Earth Explorer. The USGS Earth Explorer allows any user to search through many different types of satellite images using filters. As the Landsat series of images are often used in other similar studies and are one of the best sources for this type of research (Mohammady et al., 2015), this study also uses a Landsat 8 image from USGS. This Landsat 8 image consists of 30m spatial resolution, and the coordinate system is WGS (World Geodetic System) 1984 UTM (Universal Transverse Mercator) Zone 10N. Based on research explained below, there are three filters which are especially important when selecting data for this project.

1) As cloud cover can greatly affect the accuracy in satellite images (Cracknell & Reading, 2014; Lary et al., 2016), only images with less than 10% cloud cover are considered (Kopeć et al., 2020). The algorithms will be trained on the selected satellite images as the USGS provides complete and trustworthy ground truth data along with respective coordinates.
2) Satellite images taken during the summer months are the best candidates for study, as they avoid the problem of snow or ice appearing and altering the ground cover features. Further, and more importantly, the ground truth data being used is most recently updated in August of 2022. To match the validation most accurately, the input must also be taken from the same time period, or as close as possible to most closely resemble the land cover during this time.
3) Lastly, the satellite image should be taken only from the daytime selection to provide the greatest amount of usable data. The data extracted from nighttime images generates more noise and results in less accurate predictions (Ni et al., 2020).

2.2. PREPROCESSING

The true color resolution of the image is made by compiling seven bands of a satellite image, to create a composite image in ArcGIS Pro which is then classified. Please refer to Chapter 1, Figure 1 to

see a composite band of Red (Band 4), Green (Band 1) and Blue (Band 2), which provides a more contrasted and realistic understanding of the land cover than other band combinations.

2.3. METHODS

Recording and pulling from the same coordinates each time, the same area and features are maintained for each test run of the algorithms. Further, by keeping a consistent training baseline, the resulting scores are comparable to understand how effective and accurate each algorithm is. They will be evaluated with a confusion matrix, which then generates the most agreed upon statistic for this measure - Cohen's Kappa statistic (Gómez & Montero, 2011).

This study follows the workflow for supervised image classification, as shown in the flowchart in Figure 2. The *Configure* stage is changed each time the method is performed for a different classification method – three times total for this study. The next step is to train the classifier on the given dataset. When training the algorithm, the initial strategy was to use points, which could vary in pixelation. However, through trial and error, it was found that a polygon specific to the crop area in the map was more effective at training the algorithm, and more accurate as it was better fitted.

Another important point here is for each class to have a proportionate number of training points given its significance in the dataset. To ensure proportionately equal distribution, the "stratified random" option in ArcGIS Pro is selected. Next, the algorithm is processed through the program to classify the entire dataset, which in this case is the composite band image. Using the training data and assigned classes, every pixel in the image is assigned a value. The assignment of this value is explained more thoroughly in each of the classifiers' subsections in chapter 2.4. Finally, before the process is complete, errors can be corrected by reclassifying them to the known class. This step is especially beneficial in real-world use of classification algorithms. However, this study will only make use of this step for major errors in wide swaths of land as the purpose is to study the algorithms with as little effects of interference as necessary. This process is performed three times and is then followed by the accuracy assessment and evaluation which will be explained in chapter 2.5.

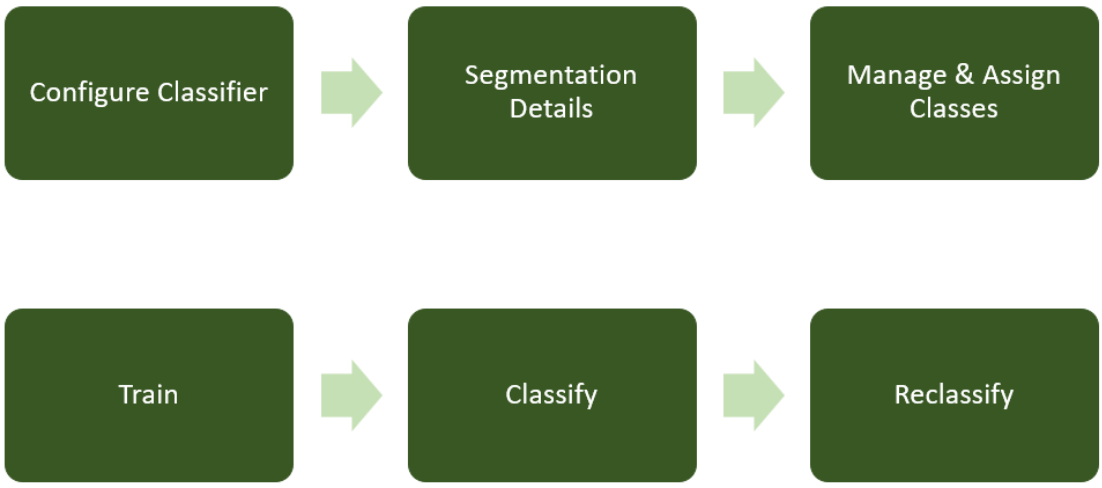

Figure 2 : Classification Workflow

2.4. CLASSIFIERS

This study compares the classification ability of the three primary supervised image classification algorithms in ArcGIS Pro. These are Maximum Likelihood Classifier (MLC), Random Forest (RF), and Support Vector Machines (SVM). These were chosen based on their reliability in ArcGIS software and their capabilities shown in numerous related studies (Cracknell & Reading, 2014; Lawrence & Moran, 2015; Maulik & Chakraborty, 2017; Rao et al., 2022; Thai et al., 2012).

2.4.1. Maximum Likelihood Classifier (MLC)

According to Sun et al., the MLC algorithm is agreed in the field to be the most accurate and most stable classifier. It consistently performs with high levels of precision and accuracy (Rebinth et al., 2021; Sun et al., 2013) when classifying satellite images. It follows the Bayesian theory which has a focus on statistical likelihood. The algorithm used for MLC can be found in the study, *Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method* (Sun et al., 2013). Each pixel is calculated using this formula. However, since it is primarily used as a pixel-based method (Sun et al., 2013), it therefore may not perform as strongly as other algorithms when used in object-based classification. In addition, another study showed that the MLC algorithm struggled to differentiate between agricultural land and rangelands (Mohammady et al., 2015).

2.4.2. Random Forest (RF)

Although MLC is a popular classifier, a study attempting to categorize many classes such as this one may be better fitted to use Random Forest (RF) (Adugna et al., 2022; Cracknell & Reading, 2014). Basing off a very simple classifier of Decision Trees (DT) (Maxwell et al., 2018), RF uses many decision trees to build a stronger classifier. Many trees are grown at once and then a majority vote determines the optimal partition of data to determine the correct class (Cracknell & Reading, 2014; Maxwell et al., 2018). Man et al. provides a diagram helpful in understanding the RF process, displayed in Figure 3:

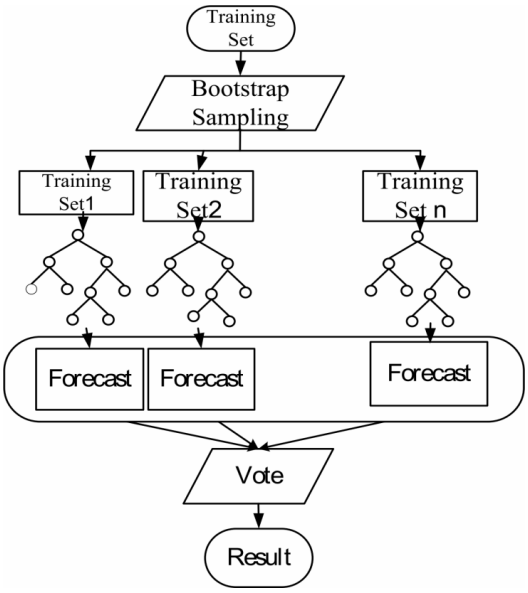

Figure 3: Random Forest Algorithm (Man et al., 2018)

Research has found that RF is a strong performer due to a few reasons: they can handle high dimensionality, they do not require hyper tuning of parameters, and they are able to rank features' importance (Xia et al., 2018). Although much tuning is not required, there are two basic parameters that must be determined for RF to work properly – the number of trees and the number of features at which to split (Thanh Noi & Kappas, 2017). According to some, a greater number of trees, around 100, produced better results (Basheer et al., 2022). However, because solid results have been obtained using the default parameters (Thanh Noi & Kappas, 2017; Zhang & Roy, 2017), this study will also begin by using the same parameters. Further, there is danger in too few trees, such as low accuracy and low precision, or too many trees, which can result in overfitting.

2.4.3. Support Vector Machine (SVM)

The Support Vector Machine (SVM) classifier is generally accepted as one of the most commonly used and high performing algorithms in the field of remote sensing (Cervantes et al., 2020a; Lawrence & Moran, 2015; Thai et al., 2012). SVM is a popular model for image classification because it can handle heavier datasets, or in this case, larger images well (Cervantes et al., 2020b). It is able to detect complex patterns so well because it is able to generalize the image and is less susceptible to noise as compared to many other popular algorithms (Cervantes et al., 2020b). One parameter to be considered is the maximum number of training samples per class. When training the algorithm, the maximum number of training samples per class should equal zero, which ensures that all training samples will be used.

The diagram in Figure 4 below illustrates a simple representation of how the SVM works:

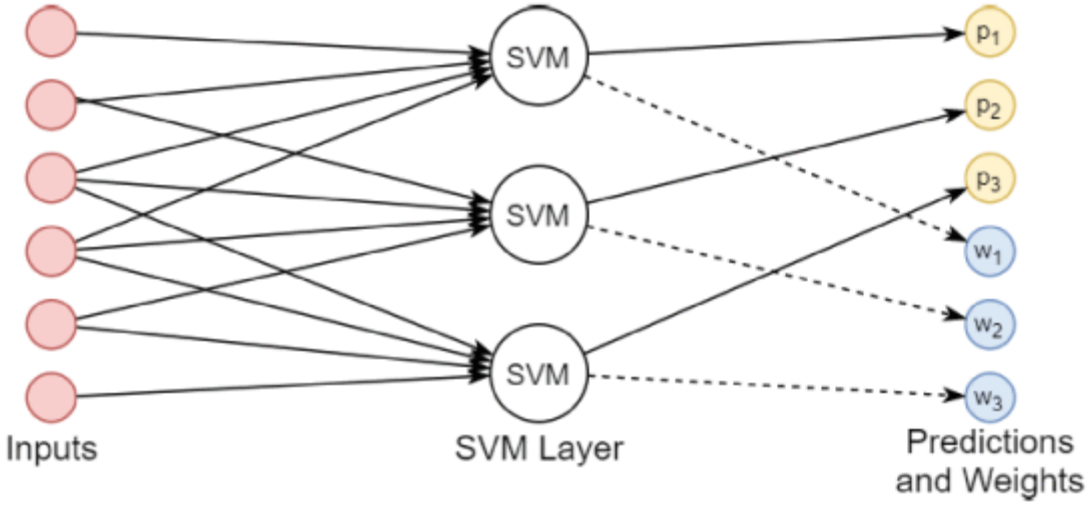

Figure 4: Simple weighted layer of SVM (Goddard & Shamir, 2022)

In addition, the mathematical formula to obtain the support vector can be found in *Non-Parametric Image Classification Based on Convolutional Ensembles of Support Vector Machines for Small Training Sets* (Goddard & Shamir, 2022).

2.5. Evaluation & Accuracy Assessment

When evaluating each algorithm, it is important to consider the relative size of each class to one another. Knowing that classes vary greatly in training sample size and density, the evaluation must also have equally distributed points and randomly across classes to avoid any bias.

Using ArcGIS Pro, once the classification workflow is complete, a table is generated to display the classified and the actual ground truth value in numeric terms for each validation data point. The numeric value is dictated by the value associated with the given class from the training portion of the process. The actual ground truth numeric value is validated using the ground truth map, provided by the California DWR. While validating these values, it is imperative that the classified value is temporarily hidden from the researcher to avoid any possible bias.

In image classification, the commonly and reliably used method of determining the best classifier is by computing the Kappa statistic (Y. Li et al., 2019). The accuracy assessment is performed by summing the predicted and actual values to compute the confusion matrix: totals of true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN). With the confusion matrix, the Kappa statistic can then be computed using the following formula provided by Y. Li et al., 2019.

$$\text{Kappa} = (\text{PCC} - \text{PRE}) / (1 - \text{PRE})$$

The probability of correct classification (PCC) represents accuracy, and is broken down into TP and TN as shown here:

$$\text{PCC} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FP} + \text{TN} + \text{FN})$$

Further, the proportional reduction in error (PRE) is made up of the following formula, once again dictated by TP and TN:

$$PRE = (TP + FP) \times (TP + FN) + (FN + TN) \times (FP + TN) / (TP + TN + FP + FN)^2$$

Using the evaluation measure of Kappa, which takes into account accuracy, the comparison can then be made between the three different algorithms – RF, SVM, and MLC as stated earlier. The highest resulting Kappa score is then proven to be the most appropriate classifier for this context (Anita et al., 2021).

When validating the data for the accuracy assessment, it is important to capture a sufficient amount of information before calculating the evaluation metrics. Therefore, at least 30 validation points per class were used in this project, which is in line with similar studies. This is especially important considering the high number and variability of classes in this study. Earlier accuracy assessment rounds in this project were not successful, partially due to low numbers of validation points.

The accuracy assessment is performed with a stratified random sampling strategy in an attempt to cover all classes equally and in an unbiased manner. As explained by Basheer et al., "stratified sampling is to divide the dataset or strata according to the characteristics of its attribute" (Basheer et al., 2022). By doing so, each group can be trained and validated in a more efficient way as the samples are already grouped by similar features (Basheer et al., 2022).

The initial accuracy assessment table generated in ArcGIS Pro contains a default "GrndTruth" value of -1 for every point. Each point is then evaluated manually by comparing to the ground truth dataset and the correct value is input alongside the classified value. Points outside of the scope of the ground truth data were removed from consideration during the accuracy assessment. Further, when evaluating each in-scope point, all classified values are hidden from view in order to avoid any confirmation bias. Upon completing validation, the confusion matrix can then be computed by the ArcGIS Pro software. The resulting confusion matrices, as explained in the previous chapter, are displayed and explained in chapter 3.3.

3. RESULTS

3.1. FINAL SPECIFIED PARAMETERS

This section presents the final parameters specified for each algorithm. Each round of testing was started with the defaults and altered to understand better how the algorithm might improve. One important note here is that the common parameters among all three classifiers were kept consistent in order to ensure a fair comparison. For example, the segmentation section of the classification process, shown in the first section of each table, contain the same values for all three classifiers. In addition, the conditions for the accuracy assessment, namely the number of assessment points and the sampling strategy, remain the same for all classifiers.

3.1.1. Maximum Likelihood Classifier

Table 2, below, displays the final parameters used to test and compare the MLC algorithm. The starting point was the default value in each section, which was then tweaked to improve the performance, in this case the spectral detail and spatial detail.

Segmentation	**Parameter**	**Final Value**	**Default Value**
	Spectral Detail	18	15.5
	Spatial Detail	5	15
	Minimum Segment Size in Pixels	20	20
Train	**Segment Attributes**	Active chromaticity color	Active chromaticity color
		Mean digital number	Mean digital number
Accuracy Assessment	**Number of random points**	394	N/A
	Sampling Strategy	Stratified Random	N/A

Table 2: Maximum Likelihood Classifier Parameters

3.1.2. Random Forest

The final parameters for RF are shown here in Table 3. The parameters were purposefully maintained as similar to the other algorithms as possible. However, the training segment was tweaked slightly when the default maximum number of samples per class of 1000 resulted in poor accuracy and a value of 100 here was shown in other studies to produce improved results.

Segmentation	**Parameter**	**Final Value**	**Default Value**
	Spectral Detail	18	15.5
	Spatial Detail	5	15
	Minimum Segment Size in Pixels	20	20
Train	**Maximum number of trees**	50	50
	Maximum Tree Depth	30	30
	Maximum number of samples per class	100	1000
	Segment Attributes	Active chromaticity color	Active chromaticity color
		Mean digital number	Mean digital number
Accuracy Assessment	**Number of random points**	394	N/A
	Sampling Strategy	Stratified Random	N/A

Table 3: Random Forest Parameters

3.1.3. Support Vector Machine

The SVM classifier parameters, shown in Table 4, were maintained similarly to MLC.

Segmentation	Parameter	Final Value	Default Value
	Spectral Detail	18	15.5
	Spatial Detail	5	15
	Minimum Segment Size in Pixels	20	20
Train	**Maximum number of samples per class**	500	500
	Segment Attributes	Active chromaticity color	Active chromaticity color
		Mean digital number	Mean digital number
Accuracy Assessment	**Number of random points**	394	N/A
	Sampling Strategy	Stratified Random	N/A

Table 4: Support Vector Machine Parameters

3.2. CLASSIFICATION RESULTS

Below the final resulting classified images are displayed for each tested algorithm. In Figure 5, one can see the classified image according to MLC.

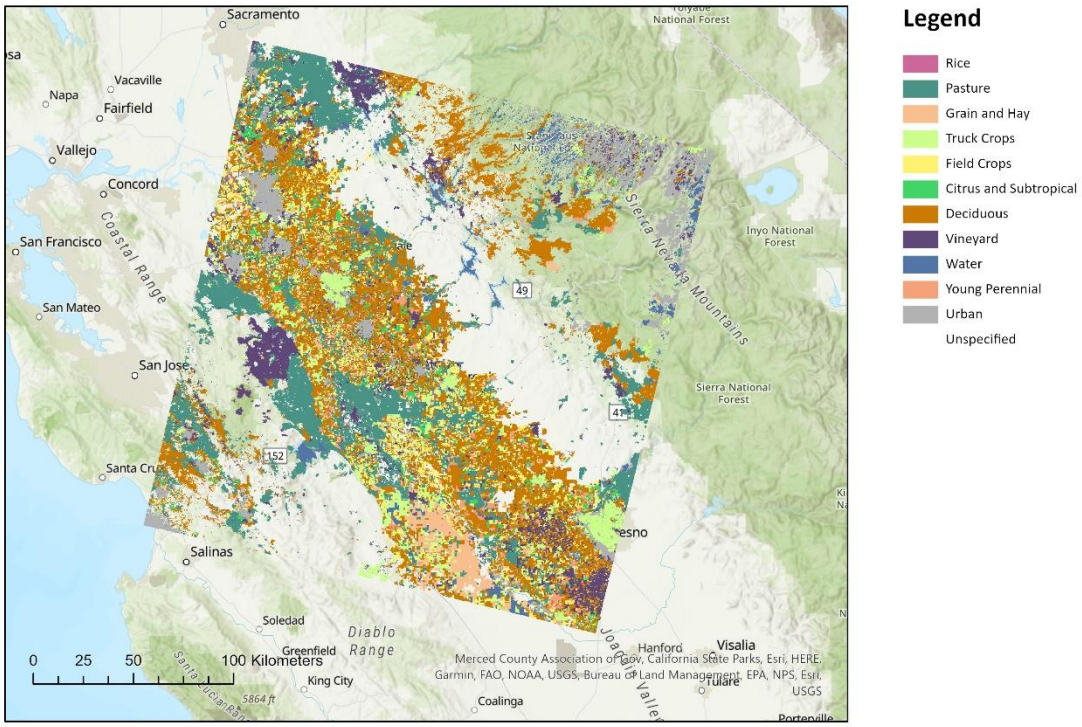

Figure 5: Maximum Likelihood Classifier Final Result

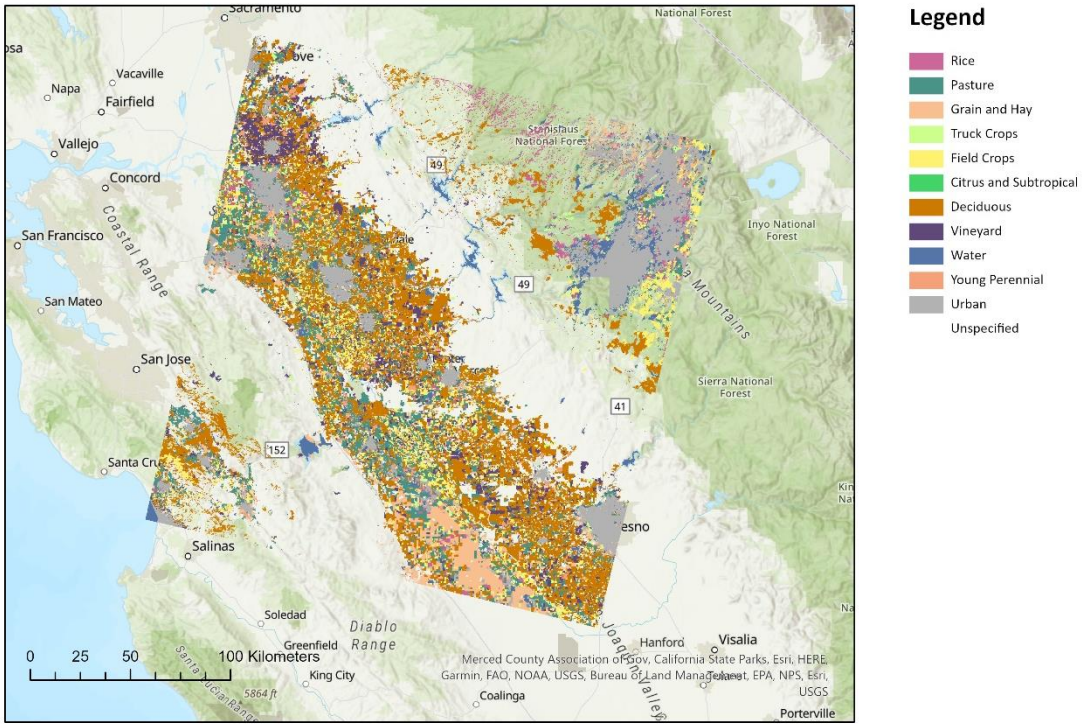

Figure 6: Random Forest Classifier Final Result

Figure 6 illustrates the RF final classification while Figure 7 shows the SVM final classification. As can be seen by comparing the images, one interesting finding is the difference between Random Forest and Support Vector Machine when deciding how to classify vastly different features. The majority of classes are representing various types of crops. However, water, urban, and unspecified classes are also included. When examining the mountainous areas of the county, RF found the closest in composition which is the urban class. This follows the logic that stone resembles concrete more than any crop type. This can be seen categorized by the vast swaths of grey. RF is a more cautious classifier than SVM. When attempting to classify part of the image in which the classifier may not have a clear match, SVM attempts to find the nearest possible class and assign it, thereby classifying the entire image as one class or another.

RF at first glance appears to have the most accurate and precise results. However, one must be careful in the case of over-fitting. If there are too many trees, the algorithm can potentially over-fit to the training dataset. A common sign of this is high accuracy in validation and low accuracy when applying to new test data.

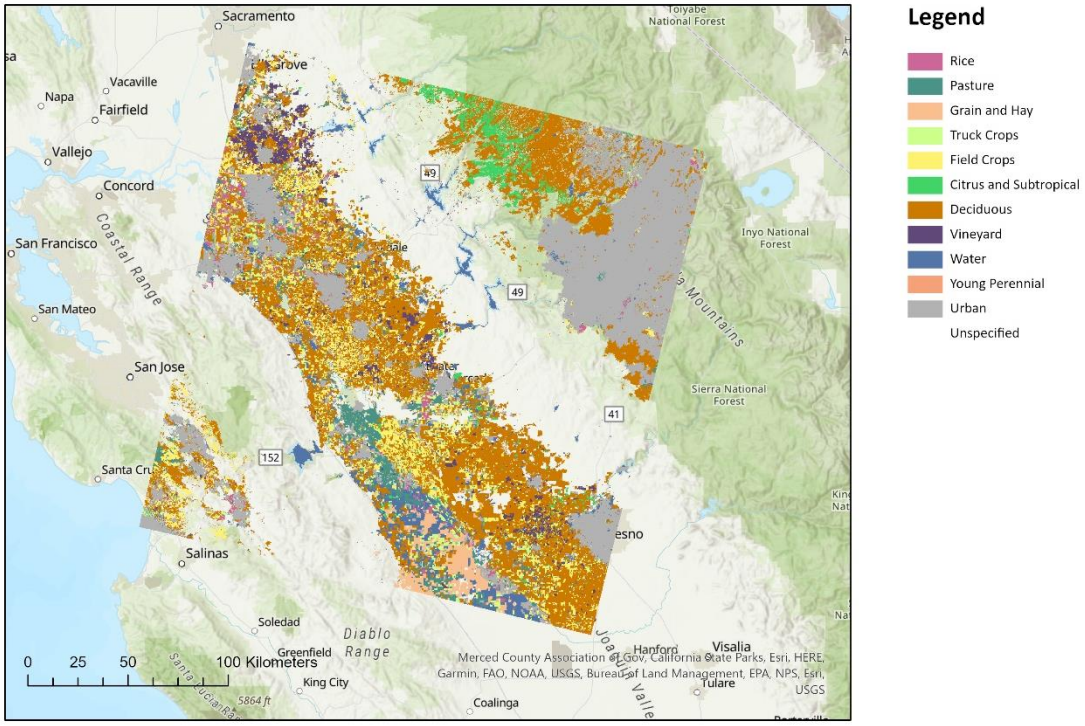

Figure 7: Support Vector Machine Classifier Final Result

3.3. Accuracy Assessment

This section displays the confusion matrices for each of the three tested classifiers as well as the percentages of total area in the image per class. Class values correspond to the crops listed in Table 1, with the accuracy score and kappa statistic calculated by ArcGIS Pro according to the formulas in chapter 2.5.

3.3.1. Maximum Likelihood Classifier:

In the MLC confusion matrix, shown below in Table 5, it can be deduced that in addition to having a lower overall accuracy and kappa statistic, the scores among individual classes are highly imbalanced. This gives the impression that the algorithm was perhaps overfitted for determining C_20 and C_8, unspecified and deciduous classes respectively. An interesting point here is that these two classes took up two of the largest shares of ground truth land cover in the study. Table 6 displays the percentages of all classes in relation to the total for further consideration.

	OID	ClassValue	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_20	Total	U_Accuracy	Kappa
1	0	C_2	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0
2	1	C_3	0	17	0	0	0	0	0	1	2	0	2	21	43	0.395349	0
3	2	C_4	0	0	9	1	0	0	1	0	0	3	0	1	15	0.6	0
4	3	C_5	0	0	2	10	4	0	4	2	0	1	4	2	29	0.344828	0
5	4	C_6	0	3	0	1	9	0	2	0	0	0	1	0	16	0.5625	0
6	5	C_7	0	0	1	0	0	0	1	1	0	0	0	0	3	0	0
7	6	C_8	0	0	0	2	5	0	55	6	0	1	1	7	77	0.714286	0
8	7	C_9	0	1	0	0	1	0	1	4	1	0	1	12	21	0.190476	0
9	8	C_10	0	0	0	0	0	0	0	0	3	0	0	2	5	0.6	0
10	9	C_11	0	0	0	1	1	0	2	1	1	0	1	2	9	0	0
11	10	C_12	0	0	0	0	0	0	0	0	1	0	19	8	28	0.678571	0
12	11	C_20	0	0	0	0	0	0	0	0	1	0	0	146	147	0.993197	0
13	12	Total	0	21	12	15	21	0	66	15	9	5	29	201	394	0	0
14	13	P_Accuracy	0	0.809524	0.75	0.666667	0.428571	0	0.833333	0.266667	0.333333	0	0.655172	0.726368	0	0.690355	0
15	14	Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.591036

Table 5: MLC Confusion Matrix

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

MLC	
Class	**Percent of Total**
C_2	0%
C_3	11%
C_4	4%
C_5	7%
C_6	4%
C_7	1%
C_8	20%
C_9	5%
C_10	1%
C_11	2%
C_12	7%
C_20	37%

Table 6: MLC Percent of Total Image Area by Class

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

3.3.2. Random Forest:

The RF confusion matrix, in Table 8, on the other hand, shows more consistency among classes with more of the class values resulting in an accuracy score above 0.5. This translates well to the overall accuracy and kappa score, which rank highest among the three classifiers. According to the original kappa statistic benchmarks as shown in Table 7, 0.61-0.80 is considered a “substantial” agreement (Landis & Koch, 1977). Therefore, the RF classifier, with a kappa score of 0.68 falls comfortably within a range dictating substantial agreement. Percentages of each class can be assessed in Table 9.

Kappa Statistic	Strength of Agreement
< 0.00	Poor
0.00-2.00	Slight
0.21-0.40	Fair
0.41-0.60	Moderate
0.61-0.80	Substantial
0.81-1.00	Almost Perfect

Table 7: Kappa Statistic Benchmarks (Landis & Koch, 1977)

	OID	ClassValue	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_20	Total	U_Accuracy	Kappa
1	0	C_2	2	0	0	0	0	0	2	0	0	0	0	7	11	0.181818	0
2	1	C_3	0	10	1	1	0	0	6	2	0	2	0	2	24	0.416667	0
3	2	C_4	0	0	11	0	3	0	1	0	0	0	0	2	17	0.647059	0
4	3	C_5	0	1	1	6	1	0	0	3	0	0	0	1	13	0.461538	0
5	4	C_6	0	0	2	3	5	0	3	0	0	0	1	4	18	0.277778	0
6	5	C_7	0	0	0	0	0	11	0	0	0	0	0	0	11	1	0
7	6	C_8	0	0	0	0	2	0	51	2	0	1	3	1	60	0.85	0
8	7	C_9	0	1	1	0	0	0	0	11	0	0	0	3	16	0.6875	0
9	8	C_10	0	0	0	0	1	0	0	0	8	0	1	3	13	0.615385	0
10	9	C_11	0	0	0	0	0	0	4	1	1	4	0	1	11	0.363636	0
11	10	C_12	0	0	0	2	0	0	0	0	0	0	15	13	30	0.5	0
12	11	C_20	0	0	1	0	0	0	1	0	0	0	1	165	168	0.982143	0
13	12	Total	2	12	17	12	12	11	68	19	9	7	21	202	392	0	0
14	13	P_Accuracy	1	0.833333	0.647059	0.5	0.416667	1	0.75	0.578947	0.888889	0.571429	0.714286	0.816832	0	0.762755	0
15	14	Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.678598

Table 8: RF Confusion Matrix

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

RF	
Class	**Percent of Total**
C_2	3%
C_3	6%
C_4	4%
C_5	3%
C_6	5%
C_7	3%
C_8	15%
C_9	4%
C_10	3%
C_11	3%
C_12	8%
C_20	43%

Table 9: RF Percent of Total Image Area by Class

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

3.3.3. Support Vector Machine:

Lastly, the SVM confusion matrix, shown in Table 10, provides competition to RF. The Kappa score is only 0.00246 lower than that of RF and so it still provides a substantial strength of agreement in data. In addition, the overall accuracy of all classes amounts to 0.75, and only 0.008948 less than that of RF. However, when looking at individual classes, it becomes more apparent that SVM is not as successful as RF. C_2 (rice) and C_11 (young perennial) which show an accuracy of 0, a value that would be unacceptable for the purposes of crop, water, and general resource planning. These results suggest that RF is more adaptable to a diverse dataset, while SVM may be better suited to specific sets of features. Corresponding percentages of each class to make up the total area are listed out in Table 11.

	OID	ClassValue	C_2	C_3	C_4	C_5	C_6	C_7	C_8	C_9	C_10	C_11	C_12	C_20	Total	U_Accuracy	Kappa
1	0	C_2	0	0	2	0	2	0	5	0	0	0	0	1	10	0	0
2	1	C_3	0	6	0	0	0	0	4	0	0	0	0	0	10	0.6	0
3	2	C_4	0	0	8	2	0	0	0	0	0	0	0	0	10	0.8	0
4	3	C_5	0	0	0	3	3	0	2	1	0	1	0	0	10	0.3	0
5	4	C_6	0	1	0	2	9	0	6	1	0	0	0	2	21	0.428571	0
6	5	C_7	0	0	0	0	1	6	0	1	1	0	0	1	10	0.6	0
7	6	C_8	0	2	0	3	4	0	53	7	1	3	1	2	76	0.697368	0
8	7	C_9	0	0	0	0	1	0	1	7	0	0	1	0	10	0.7	0
9	8	C_10	0	0	0	1	1	0	1	0	2	1	0	5	11	0.181818	0
10	9	C_11	0	0	0	3	1	0	2	0	0	0	0	4	10	0	0
11	10	C_12	0	1	1	0	0	0	0	0	0	0	51	4	57	0.894737	0
12	11	C_20	0	1	1	0	0	0	0	0	1	0	4	152	159	0.955975	0
13	12	Total	0	11	12	14	22	6	74	17	5	5	57	171	394	0	0
14	13	P_Accuracy	0	0.545455	0.666667	0.214286	0.409091	1	0.716216	0.411765	0.4	0	0.894737	0.888889	0	0.753807	0
15	14	Kappa	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.676138

Table 10: SVM Confusion Matrix

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

SVM	
Class	**Percent of Total**
C_2	3%
C_3	3%
C_4	3%
C_5	3%
C_6	5%
C_7	3%
C_8	19%
C_9	3%
C_10	3%
C_11	3%
C_12	14%
C_20	40%

Table 11: SVM Percent of Total Image Area by Class

(C_2 = Rice, C_3 = Pasture, C_4 = Grain and Hay, C_5 = Truck Crops, C_6 = Field Crops, C_7 = Citrus and Subtropical, C_8 = Deciduous, C_9 = Vineyard, C_10 = Water, C_11 = Young Perennial, C_12 = Urban, C_20 = Unspecified)

4. DISCUSSION

4.1. CHALLENGES & MITIGATION STRATEGIES

4.1.1. Data Collection & Quality

One challenge which arose early on during the study was the data collection in terms of the search parameters, weather conditions, time of year, and so on. It was important to obtain the best quality image possible, which ruled out winter months, because of potential snow; cloud cover greater than 10% (Kopeć et al., 2020) because of the obstruction of view, and it had to be during the same time of year as the ground truth data – as land cover composition can vary greatly from month to month depending on climate.

4.1.2. Class Imbalance & Unlabeled Data

One of the major challenges of image classification is the vast amount of unlabeled data (Qi & Luo, 2022). This pairs hand in hand with the challenge of class imbalance which existed naturally in this study's dataset. As the goal was to identify various crop types, and the satellite image contained only a limited amount of crop land cover, the resulting issue was that large swaths of the image contained unspecified land cover within the scope of the study. The proposed mitigation strategy for this problem was initially to follow an under sampling of the unspecified regions and oversampling of the labeled data. However, results from these initial attempts proved to misclassify the unspecified areas as one of the 11 land cover types (see Appendix figures A-H for a visual representation). When comparing with Figures 10-12 in the results section, one can see a clear difference when the training samples then changed to a stratified random sampling, a method which would ensure the same proportion of training samples to actual data.

4.1.3. Researcher Impartiality in Assessment

A secondary challenge occurred throughout the accuracy assessment. The assessment was performed by manually checking the ground truth data. However, to avoid bias in the manual check, the column containing the classified value needed to be hidden from view to maintain complete objectivity.

4.1.4. Overfitting

The last major challenge of this project was to balance between under and overfitting: overfitting being when the algorithm is highly trained and fitted to a specific dataset but does not perform as well when applied to new and unprocessed data (Hesamian et al., 2019). Between the primary and secondary trial runs, the training size had to be increased, and as the training dataset increased in proportion, the risk of overfitting increased as well.

4.2. FINDINGS & DISCUSSION

From the beginning of the classification workflow, it was found that object-based classification tends to outperform pixel-based (Fu et al., 2017). This was supported by a diverse set of datasets. Further, it is generally agreed upon, and found in this study as well, that greater training samples provide a better fit to the algorithm and thus more effective learning (Basheer et al., 2022). By training the algorithm with polygon shapes which are specific to their classes, results improved from the initial round of testing which used only datapoints, which were circular in shape and not fitted to the data. The polygon shapes appear to have helped the algorithm distinguish boundaries in addition to setting a lower spatial detail. The lower spatial detail value helped to segment the image into distinct classes before the classification step and is more suited to distinguishing impervious features in the satellite image. Finally, the increased spectral detail helped create some distinction between land cover patches which are similar in feature characteristics and close together. Adjusting the spatial and spectral detail parameters was especially helpful in this study as the classes included many related crop types with potentially overlapping characteristics.

These tweaks in training and parameters result in visible changes in the final classified results. One can view a dramatic improvement from early testing (round 1, seen in Appendix Figure A) to the final testing (seen in Results section Figure 10) regarding MLC, in addition to the other classifiers. Initial testing, performed with round datapoints in the training set and all parameters set to the default values, amounted to Kappa values of around 0.2. Final testing was performed with polygon training samples, a greater number of training samples, and tuned parameters, the Kappa value reached 0.68.

As supported by many other researchers, it was also found that a greater number of training and validation samples greatly improved the classification output. Initial testing was performed with ten samples per class in a schema of 12 classes, to understand the parameters and their effects. However, the initial results could not be used to determine any conclusions due to low scores. Subsequent tests were performed with larger and more specific training samples as well as an increased number of validation points; at least 30 validation samples per class ensured a sufficient amount of data to validate against, especially considering the high number of classes in the confusion matrix. According to Ramezan et al, it was found that “RF, the algorithm with the highest overall accuracy, was notable for its negligible decrease in overall accuracy, 1.0%, when training sample size decreased from 10,000 to 315 samples” and that SVM was very sensitive to a decrease in sample size (Ramezan et al., 2021).

5. CONCLUSION

5.1. SUMMARY

This study seeks to address two research objectives:

1) it aims to evaluate the effectiveness of the three primary classification algorithms available in ArcGIS Pro, and

2) to determine an optimal algorithm for the given study area of San Joaquin County, California, and its respective 12 land cover types, including 9 different crop types.

To answer these questions, three image classification algorithms were researched, tested, tweaked, and compared to find optimal parameters and a best fit to the data. These three algorithms (Maximum Likelihood, Random Forest, and Support Vector Machine) were chosen based on their general usability in image classification as well as their proven characteristics through research. The data was in the form of a Landsat 8 image, provided by USGS, with 7 bands which were then processed and formed into a composite band for use in training. The ground truth validation data was also provided by USGS through the form of a crop planning map, which had categorized a wide swath of San Joaquin County into nine different crop types, as well as water, urban, and an unspecified group.

Using the final confusion matrices as presented in chapter 3, the most effective algorithm tested was RF with an accuracy score of 0.75 and a Kappa statistic of 0.68. Although not in perfect agreement, the Kappa statistic of 0.68 still represents substantial agreement and can be used with some confidence. This result was obtained after multiple trials with varying parameters. The parameters for each algorithm were also tested and one can further deduce from this study that an increased spectral detail and a decreased spatial detail parameter enabled more effective learning. These were the most altered parameters, while the other parameters' default values were sufficient to remain.

5.2. INTERPRETATION & IMPLICATION

Although RF and SVM both had very close Kappa scores of 0.68, their accuracy scores varied slightly with RF in the lead and SVM at 0.75. Therefore, the real implication is that both algorithms will provide similarly accurate and quality results even though RF is the chosen algorithm in this particular study. It is also important to note the other scores in the confusion matrix. If certain features are more important to classify, the ultimate algorithm may differ. For example, the SVM algorithm performed the best in terms of solely identifying urban spaces, with an accuracy score of 0.89 on the class value of 20 (urban). Because of this, SVM would be best suited to differentiating between vegetation and urban land if the goal is only to have a two-class output.

5.3. LIMITATIONS

As to be expected, the main issue in remote image sensing is incorrect classification. A major source of misclassification GIS is heterogeneous patches of land cover (Gessner et al., 2013). This project was a solid case study in classification in heterogeneous land cover as a single county in California consisted of 12 different classes and within those 12, nine of them were crop types. Some crop types contain more similar spectral detail to each other and were thus more difficult to distinguish from each other. This can be seen in earlier trials of the algorithms and even in the final confusion matrices, when two classes, for example classes 5 and 6 – truck crops and field crops, were more frequently misclassified with each other than with the other classes.

Fadhillah et al. (2022) points out some limitations of the machine learning algorithms being used in this field. Many of these complicated algorithms lack information-extraction features and require long processing times (Fadhillah et al., 2022, p. 464). Their proposed solution involves hybrid models and optimization algorithms. Further, their results utilize correlation statistics with physical conditions in the study area to effectively determine landslide potential. Similarly, Thai et al. combined SVM and Artificial Neural Networks (ANN) to create and test a hybrid model and Okwuashi & Ndehedehe integrated Deep Neural Network (DNN) with SVM. Both draw on respective strengths of one learning model to address the gaps in the other. Combined and hybrid models as well as other complex solutions were out of scope in this study and therefore limited in the extent of answering how optimized certain solutions could be.

Despite the pitfalls of supervised image classification, they remain the most accurate and efficient method. Traditional and manual methods of crop assessment are time consuming, labor intensive, and expensive (Ali et al., 2022b). Even the ground truth data for this study was done in part by machine learning methods – specifically Random Forest – before being aided by manual validation. As seen in the Appendix, in the second round Support Vector Machine testing, misclassification was obvious even prior to the accuracy assessment as large swaths of the image were misclassified as water. This, as opposed to human errors through manual work, was much easier to spot and correct through reclassification tools. Further, the use of remote sensing can classify much larger segments and much faster than any other method. Using a combination of supervised classification and manual assessment was a way of mitigating some of the natural limitations of either method alone.

5.4. RECOMMENDATIONS FOR FUTURE WORKS

In future works, it may be more effective to focus on fewer classes and more parameter testing. This study accepted a wide range of features which contributed to its limitations and resulted in potentially lower accuracy and Kappa scores than what it could have otherwise performed if there were fewer classes and greater distinction between them.

REFERENCES

Abdollahi, A., Liu, Y., Pradhan, B., Huete, A., Dikshit, A., & Nguyen Tran, N. (2022). Short-time-series grassland mapping using Sentinel-2 imagery and deep learning-based architecture. *Egyptian Journal of Remote Sensing and Space Science*, *25*(3), 673–685. https://doi.org/10.1016/j.ejrs.2022.06.002

Adugna, T., Xu, W., & Fan, J. (2022). Comparison of Random Forest and Support Vector Machine Classifiers for Regional Land Cover Mapping Using Coarse Resolution FY-3C Images. *Remote Sensing*, *14*(3). https://doi.org/10.3390/rs14030574

Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022a). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). In *Egyptian Journal of Remote Sensing and Space Science* (Vol. 25, Issue 3, pp. 711–716). Elsevier B.V. https://doi.org/10.1016/j.ejrs.2022.04.006

Ali, A. M., Abouelghar, M., Belal, A. A., Saleh, N., Yones, M., Selim, A. I., Amin, M. E. S., Elwesemy, A., Kucher, D. E., Maginan, S., & Savin, I. (2022b). Crop Yield Prediction Using Multi Sensors Remote Sensing (Review Article). In *Egyptian Journal of Remote Sensing and Space Science* (Vol. 25, Issue 3, pp. 711–716). Elsevier B.V. https://doi.org/10.1016/j.ejrs.2022.04.006

Anita, N., Sukojo, B. M., Meisajiwa, S. H., & Romadhon, M. A. (2021). OIL PATTERN IDENTIFICATION ANALYSIS USING SEMANTIC DEEP LEARNING METHOD from PLEIADES-1B SATELIITE IMAGERY with ARCGIS PRO SOFTWARE (Case Study: Village "a"). *IOP Conference Series: Earth and Environmental Science*, *936*(1). https://doi.org/10.1088/1755-1315/936/1/012021

Basheer, S., Wang, X., Farooque, A. A., Nawaz, R. A., Liu, K., Adekanmbi, T., & Liu, S. (2022). Comparison of Land Use Land Cover Classifiers Using Different Satellite Imagery and Machine Learning Techniques. *Remote Sensing*, *14*(19). https://doi.org/10.3390/rs14194978

Bontemps, S., Herold, M., Kooistra, L., van Groenestijn, A., Hartley, A., Arino, O., Moreau, I., & Defourny, P. (2011). Global land cover dataset for climate models Revisiting land cover observations to address the needs of the climate modelling community Global land cover dataset for climate models. *Biogeosciences Discuss*, *8*, 7713–7740. https://doi.org/10.5194/bgd-8-7713-2011

Cavagnaro, T., Jackson, L., Scow, K., & Org, E. (2005). *UC Berkeley Climate Change Title Climate Change: Challenges And Solutions For California Agricultural Landscapes Permalink https://escholarship.org/uc/item/9c0618tq Publication Date*. https://escholarship.org/uc/item/9c0618tq

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020a). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Cervantes, J., Garcia-Lamont, F., Rodríguez-Mazahua, L., & Lopez, A. (2020b). A comprehensive survey on support vector machine classification: Applications, challenges and trends. *Neurocomputing*, *408*, 189–215. https://doi.org/10.1016/j.neucom.2019.10.118

Cracknell, M. J., & Reading, A. M. (2014). Geological mapping using remote sensing data: A comparison of five machine learning algorithms, their response to variations in the spatial distribution of training data and the use of explicit spatial information. *Computers and Geosciences*, *63*, 22–33. https://doi.org/10.1016/j.cageo.2013.10.008

Fadhillah, M. F., Hakim, W. L., Panahi, M., Rezaie, F., Lee, C. W., & Lee, S. (2022). Mapping of landslide potential in Pyeongchang-gun, South Korea, using machine learning meta-based optimization algorithms. *Egyptian Journal of Remote Sensing and Space Science*, *25*(2), 463–472. https://doi.org/10.1016/j.ejrs.2022.03.008

Fu, B., Wang, Y., Campbell, A., Li, Y., Zhang, B., Yin, S., Xing, Z., & Jin, X. (2017). Comparison of object-based and pixel-based Random Forest algorithm for wetland vegetation mapping using high spatial resolution GF-1 and SAR data. *Ecological Indicators*, *73*, 105–117. https://doi.org/10.1016/j.ecolind.2016.09.029

Gessner, U., Machwitz, M., Conrad, C., & Dech, S. (2013). Estimating the fractional cover of growth forms and bare surface in savannas. A multi-resolution approach based on regression tree ensembles. *Remote Sensing of Environment*, *129*, 90–102. https://doi.org/10.1016/j.rse.2012.10.026

Goddard, H., & Shamir, L. (2022). SVMnet: Non-Parametric Image Classification Based on Convolutional Ensembles of Support Vector Machines for Small Training Sets. *IEEE Access*, *10*, 24029–24038. https://doi.org/10.1109/ACCESS.2022.3154405

Gómez, D., & Montero, J. (2011). *Determining the accuracy in image supervised classification problems*.

Hesamian, M. H., Jia, W., He, X., & Kennedy, P. (2019). Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, *32*(4), 582–596. https://doi.org/10.1007/s10278-019-00227-x

Hürlimann, M., Guo, Z., Puig-Polo, C., & Medina, V. (2022). Impacts of future climate and land cover changes on landslide susceptibility: regional scale modelling in the Val d'Aran region (Pyrenees, Spain). *Landslides*, *19*(1), 99–118. https://doi.org/10.1007/s10346-021-01775-6

Jet, A., & O, H. J. (2017). Supervised Machine Learning Algorithms: Classification and Comparison. *International Journal of Computer Trends and Technology*, *48*. https://doi.org/10.14445/22312803/IJCTT-V48P126

Khatami, R., Mountrakis, G., & Stehman, S. v. (2016). A meta-analysis of remote sensing research on supervised pixel-based land-cover image classification processes: General guidelines for practitioners and future research. *Remote Sensing of Environment*, *177*, 89–100. https://doi.org/10.1016/j.rse.2016.02.028

Kopeć, A., Trybała, P., Głąbicki, D., Buczyńska, A., Owczarz, K., Bugajska, N., Kozińska, P., Chojwa, M., & Gattner, A. (2020). Application of remote sensing, gis and machine learning with geographically weighted regression in assessing the impact of hard coal mining on the natural environment. *Sustainability (Switzerland)*, *12*(22), 1–26. https://doi.org/10.3390/su12229338

Landis, J. R., & Koch, G. G. (1977). *The Measurement of Observer Agreement for Categorical Data* (Vol. 33, Issue 1). https://about.jstor.org/terms

Lary, D. J., Alavi, A. H., Gandomi, A. H., & Walker, A. L. (2016). Machine learning in geosciences and remote sensing. *Geoscience Frontiers*, *7*(1), 3–10. https://doi.org/10.1016/j.gsf.2015.07.003

Lawrence, R. L., & Moran, C. J. (2015). The AmericaView classification methods accuracy comparison project: A rigorous approach for model selection. *Remote Sensing of Environment*, *170*, 115–120. https://doi.org/10.1016/j.rse.2015.09.008

Li, F., Bai, J., Zhang, M., & Zhang, R. (2022). Yield estimation of high-density cotton fields using low-altitude UAV imaging and deep learning. *Plant Methods*, *18*(1). https://doi.org/10.1186/s13007-022-00881-3

Li, Y., Peng, C., Chen, Y., Jiao, L., Zhou, L., & Shang, R. (2019). A Deep Learning Method for Change Detection in Synthetic Aperture Radar Images. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(8), 5751–5763. https://doi.org/10.1109/TGRS.2019.2901945

Maulik, U., & Chakraborty, D. (2017). Remote Sensing Image Classification: A survey of support-vector-machine-based advanced techniques. In *IEEE Geoscience and Remote Sensing Magazine* (Vol. 5, Issue 1, pp. 33–52). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/MGRS.2016.2641240

Maxwell, A. E., Warner, T. A., & Fang, F. (2018). Implementation of machine-learning classification in remote sensing: An applied review. In *International Journal of Remote Sensing* (Vol. 39, Issue 9, pp. 2784–2817). Taylor and Francis Ltd. https://doi.org/10.1080/01431161.2018.1433343

Mishra, V. N., Prasad, R., Rai, P. K., Vishwakarma, A. K., & Arora, A. (2019). Performance evaluation of textural features in improving land use/land cover classification accuracy of heterogeneous landscape using multi-sensor remote sensing data. *Earth Science Informatics*, *12*(1), 71–86. https://doi.org/10.1007/s12145-018-0369-z

Mohammady, M., Moradi, H. R., Zeinivand, H., & Temme, A. J. A. M. (2015). A comparison of supervised, unsupervised and synthetic land use classification methods in the north of Iran. *International Journal of Environmental Science and Technology*, *12*(5), 1515–1526. https://doi.org/10.1007/s13762-014-0728-3

Ni, Y., Li, X., Ye, Y., Li, Y., Li, C., & Chu, D. (2020). An Investigation on Deep Learning Approaches to Combining Nighttime and Daytime Satellite Imagery for Poverty Prediction. *IEEE Geoscience and Remote Sensing Letters*, *18*(9), 1545–1549. https://doi.org/10.1109/lgrs.2020.3006019

Planning i15_Crop_Mapping_2019. (n.d.).

Potapov, P. v., Turubanova, S. A., Tyukavina, A., Krylov, A. M., McCarty, J. L., Radeloff, V. C., & Hansen, M. C. (2015). Eastern Europe's forest cover dynamics from 1985 to 2012 quantified from the full Landsat archive. *Remote Sensing of Environment*, *159*, 28–43. https://doi.org/10.1016/j.rse.2014.11.027

Qi, G. J., & Luo, J. (2022). Small Data Challenges in Big Data Era: A Survey of Recent Progress on Unsupervised and Semi-Supervised Methods. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *44*(4), 2168–2187. https://doi.org/10.1109/TPAMI.2020.3031898

Ramezan, C. A., Warner, T. A., Maxwell, A. E., & Price, B. S. (2021). Effects of training set size on supervised machine-learning land-cover classification of large-area high-resolution remotely sensed data. *Remote Sensing*, *13*(3), 1–27. https://doi.org/10.3390/rs13030368

Rao, P., Wang, Y., Liu, Y., Wang, X., Hou, Y., Pan, S., Wang, F., & Zhu, D. (2022). A comparison of multiple methods for mapping groundwater levels in the Mu Us Sandy Land, China. *Journal of Hydrology: Regional Studies*, *43*. https://doi.org/10.1016/j.ejrh.2022.101189

Rebinth, A., Kumar, S. M., Kumanan, T., & Varaprasad, G. (2021). Glaucoma Image Classification Using Entropy Feature and Maximum Likelihood Classifier. *Journal of Physics: Conference Series*, *1964*(4). https://doi.org/10.1088/1742-6596/1964/4/042075

Sheykhmousa, M., Mahdianpari, M., Ghanbari, H., Mohammadimanesh, F., Ghamisi, P., & Homayouni, S. (2020). Support Vector Machine Versus Random Forest for Remote Sensing Image Classification: A Meta-Analysis and Systematic Review. In *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing* (Vol. 13, pp. 6308–6325). Institute of Electrical and Electronics Engineers Inc. https://doi.org/10.1109/JSTARS.2020.3026724

Sun, J., Yang, J., Zhang, C., Yun, W., & Qu, J. (2013). Automatic remotely sensed image classification in a grid environment based on the maximum likelihood method. *Mathematical and Computer Modelling*, *58*(3–4), 573–581. https://doi.org/10.1016/j.mcm.2011.10.063

Swain, D. L., Langenbrunner, B., Neelin, J. D., & Hall, A. (2018). Increasing precipitation volatility in twenty-first-century California. *Nature Climate Change*, *8*(5), 427–433. https://doi.org/10.1038/s41558-018-0140-y

Thai, L. H., Hai, T. S., & Thuy, N. T. (2012). Image Classification using Support Vector Machine and Artificial Neural Network. *International Journal of Information Technology and Computer Science*, *4*(5), 32–38. https://doi.org/10.5815/ijitcs.2012.05.05

Thanh Noi, P., & Kappas, M. (2017). Comparison of Random Forest, k-Nearest Neighbor, and Support Vector Machine Classifiers for Land Cover Classification Using Sentinel-2 Imagery. *Sensors (Basel, Switzerland)*, *18*(1). https://doi.org/10.3390/s18010018

USGS FAQ. (n.d.).

Xia, J., Ghamisi, P., Yokoya, N., & Iwasaki, A. (2018). Random forest ensembles and extended multiextinction profiles for hyperspectral image classification. *IEEE Transactions on Geoscience and Remote Sensing*, *56*(1), 202–216. https://doi.org/10.1109/TGRS.2017.2744662

Yang, J., Gong, P., Fu, R., Zhang, M., Chen, J., Liang, S., Xu, B., Shi, J., & Dickinson, R. (2013). The role of satellite remote sensing in climate change studies. *Nature Climate Change*, *3*(10), 875–883. https://doi.org/10.1038/nclimate1908

Zhang, H. K., & Roy, D. P. (2017). Using the 500 m MODIS land cover product to derive a consistent continental scale 30 m Landsat land cover classification. *Remote Sensing of Environment*, *197*, 15–34. https://doi.org/10.1016/j.rse.2017.05.024

APPENDIX

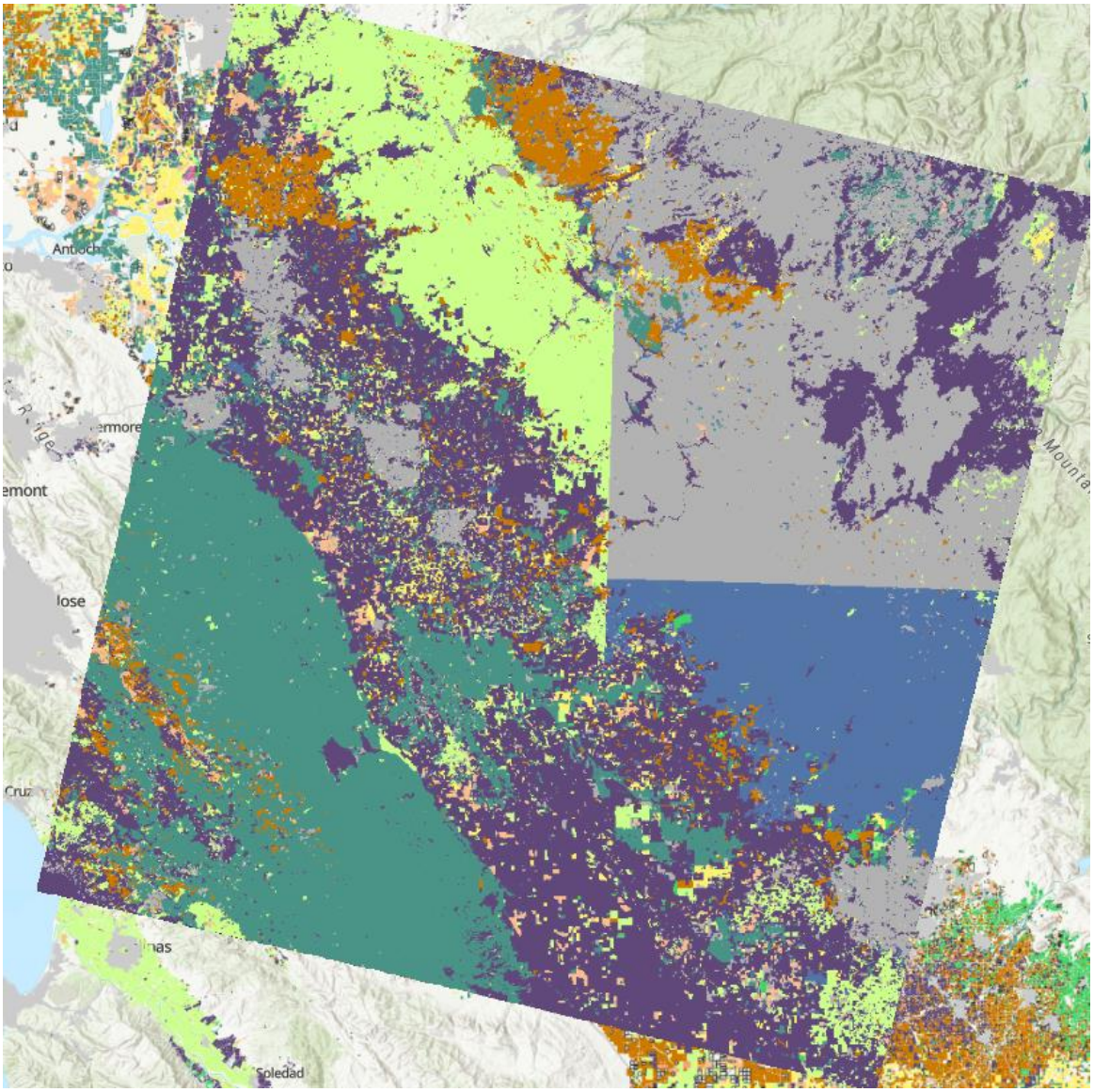

Figure A: Maximum Likelihood trial 1 result

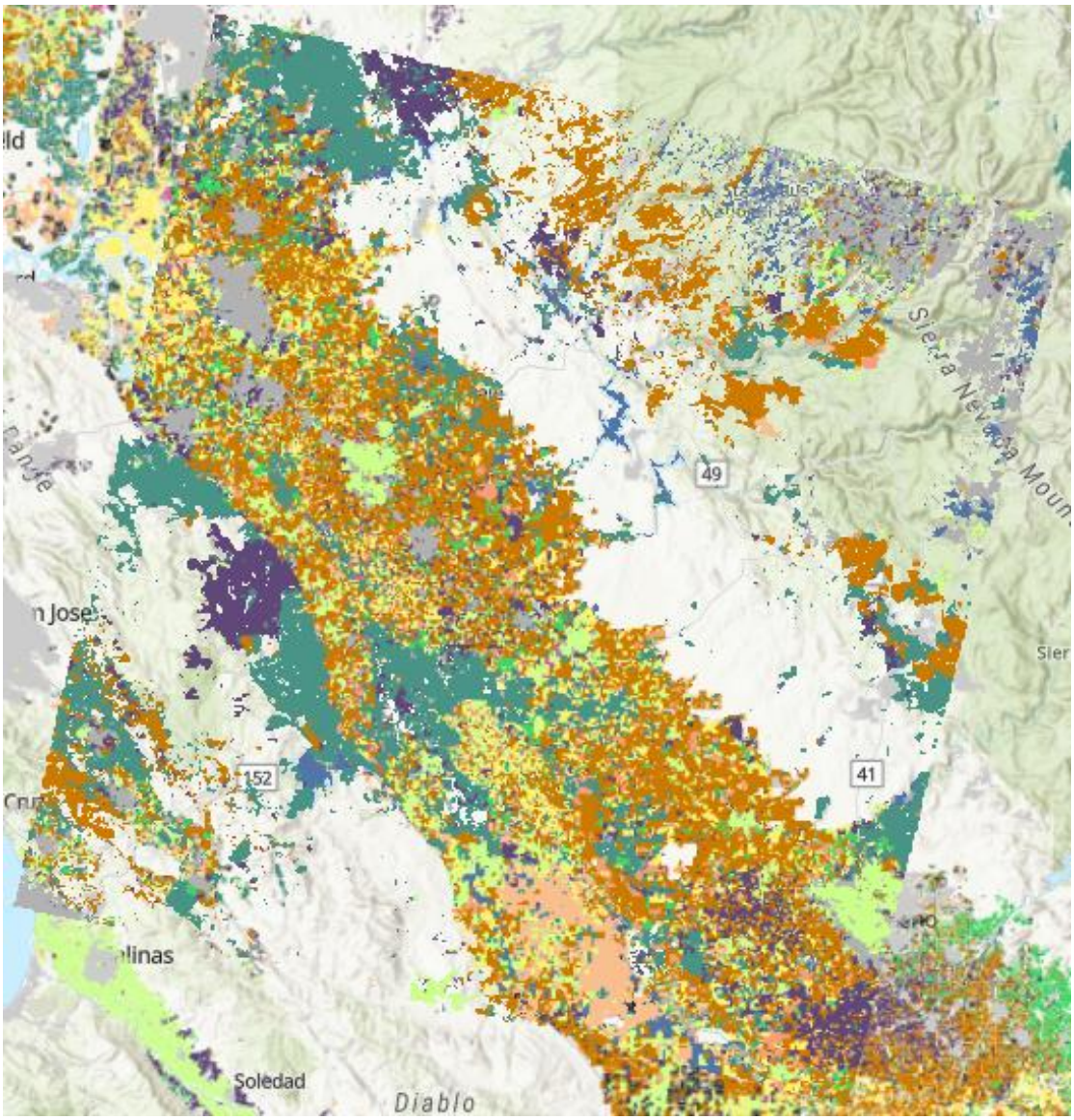

Figure B: Maximum Likelihood trial 2 result

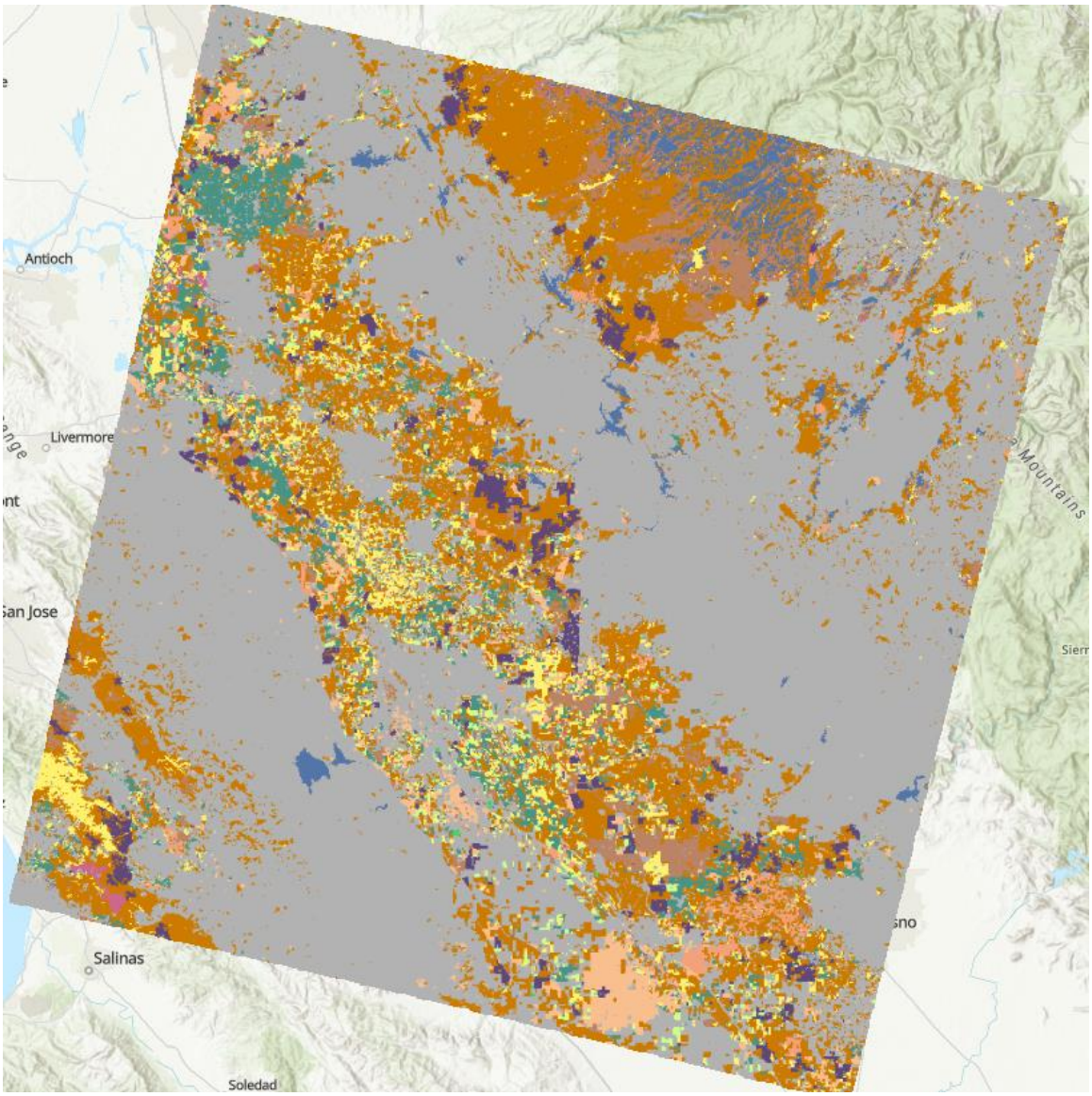

Figure C: Random Forest trial 1 preliminary result (1000 maximum number of samples per class)

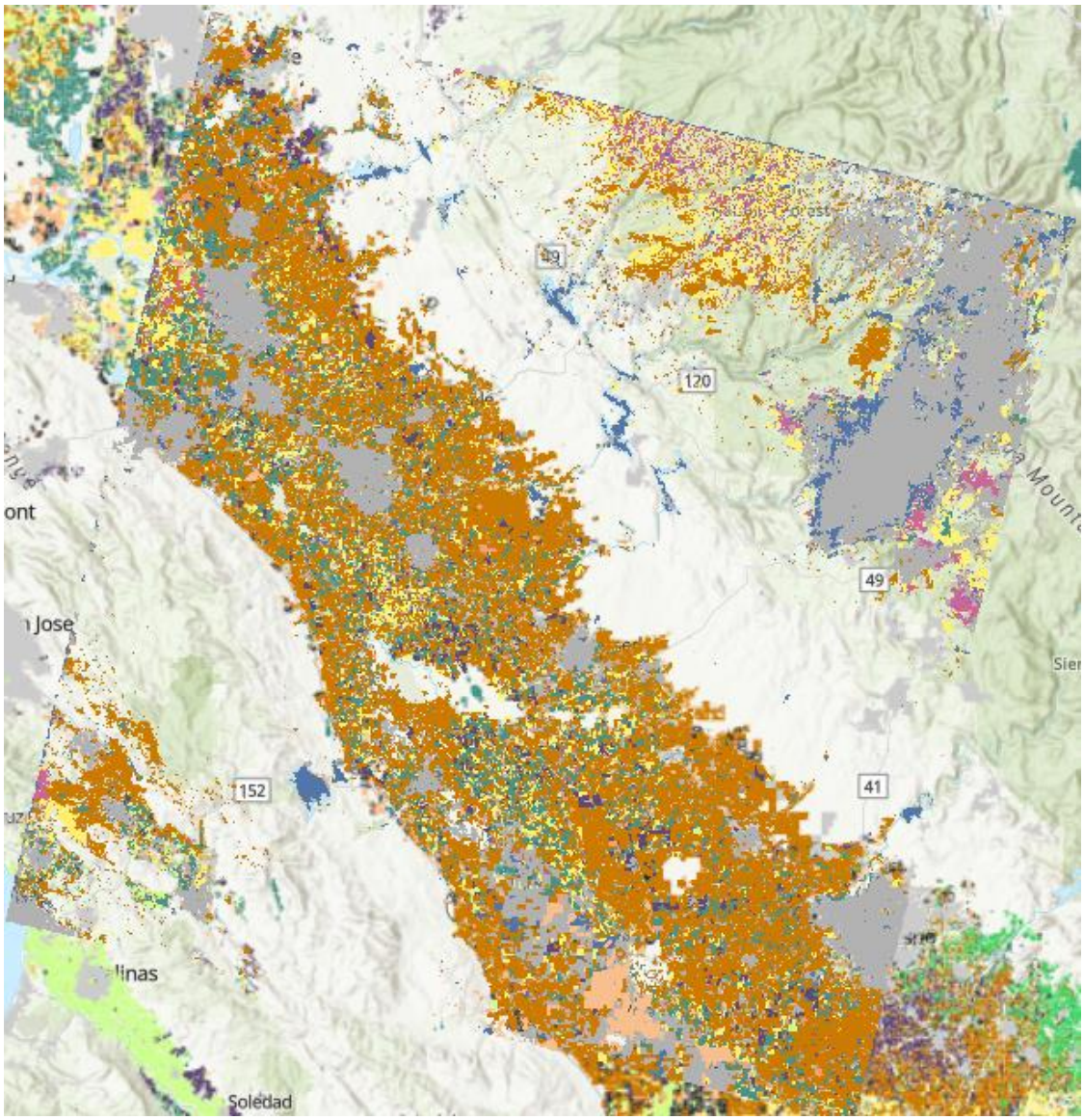

Figure D: Random Forest trial 2 preliminary result (500 maximum number of samples per class)

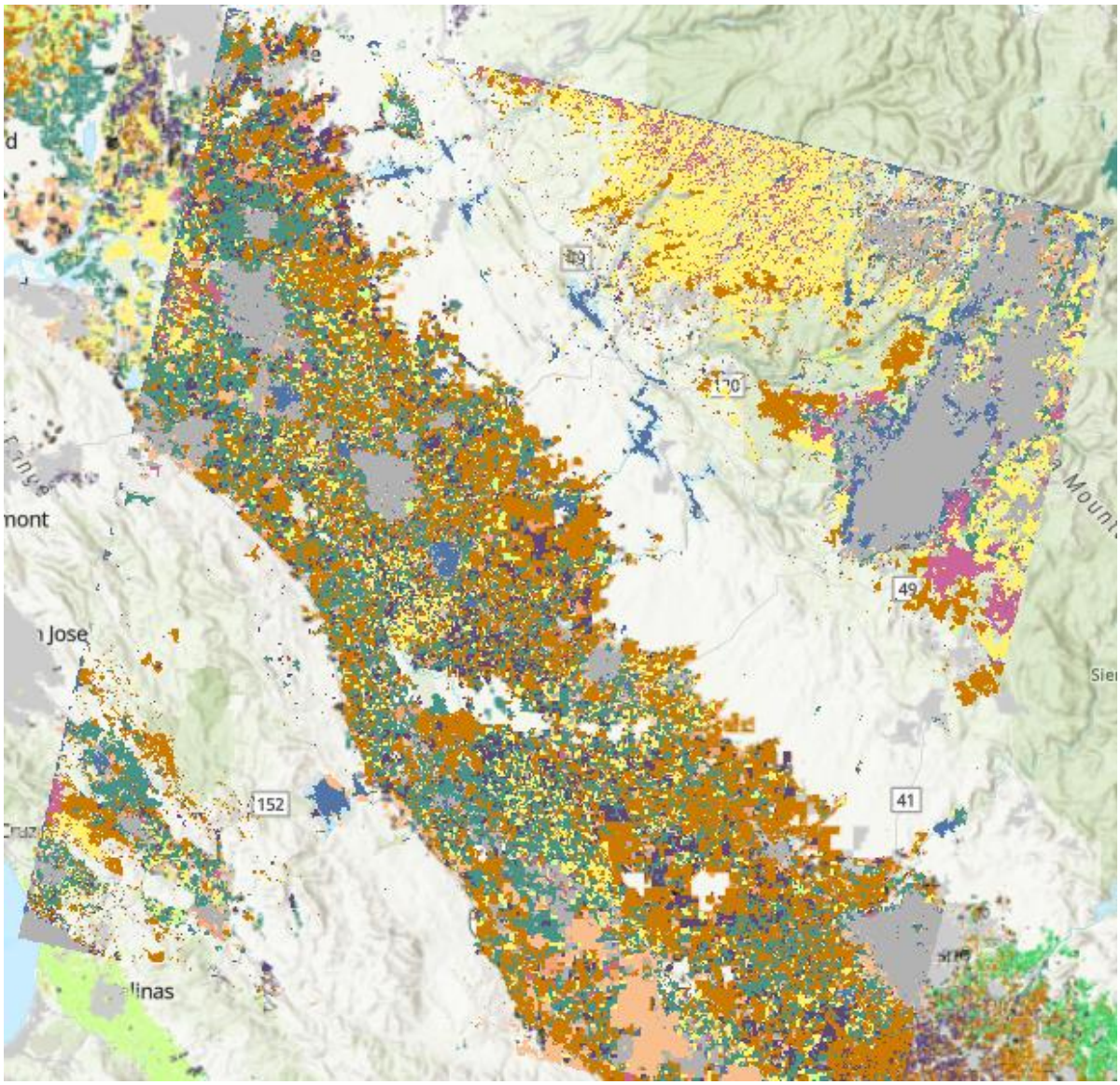

Figure E: Random Forest trial 2 preliminary result (100 maximum number of samples per class)

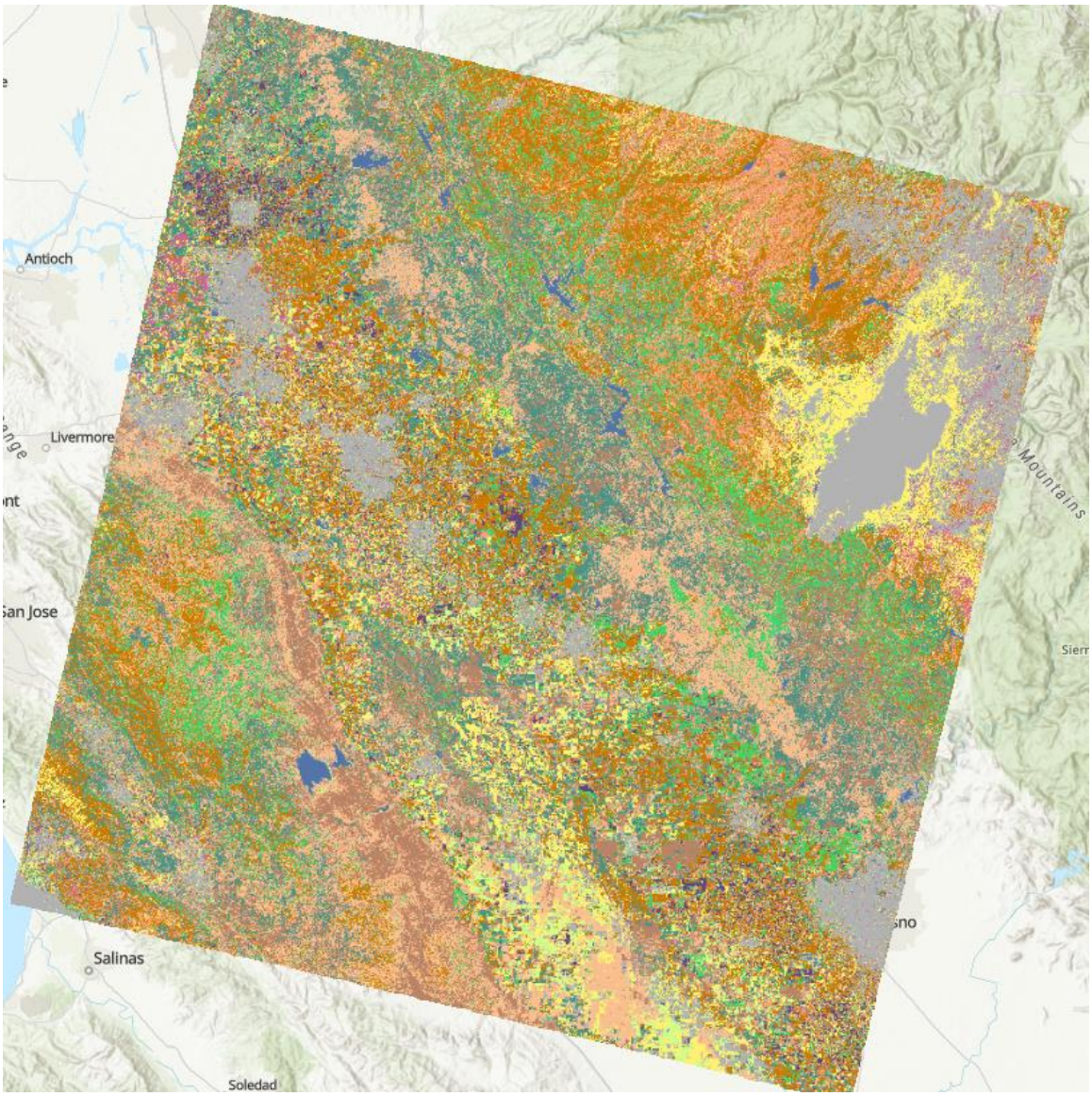

Figure F: Support Vector Machine trial 1 result

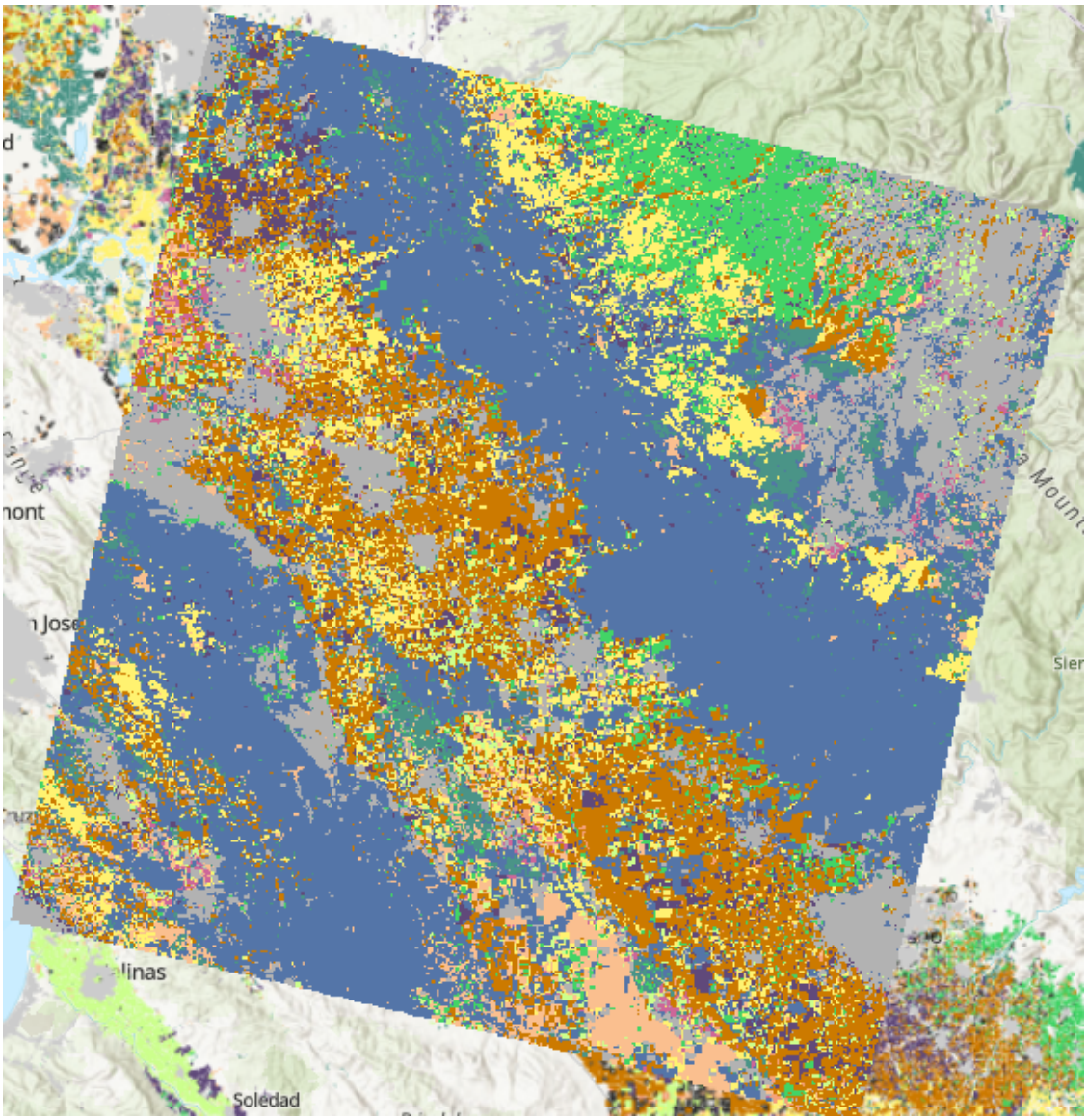

Figure G: Support Vector Machine trial 2 result

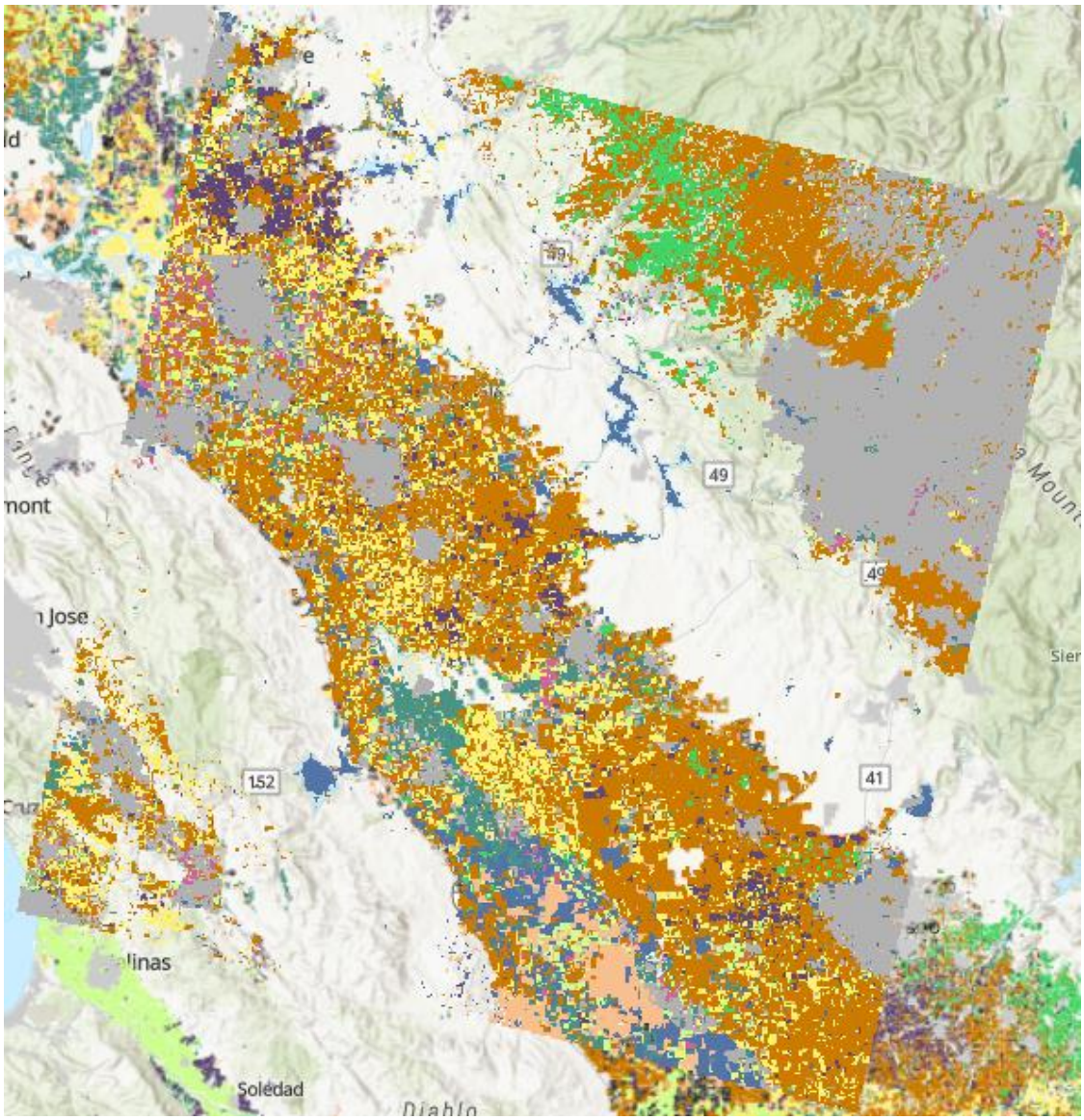

Figure H: Support Vector Machine trial 3 result

ANNEX

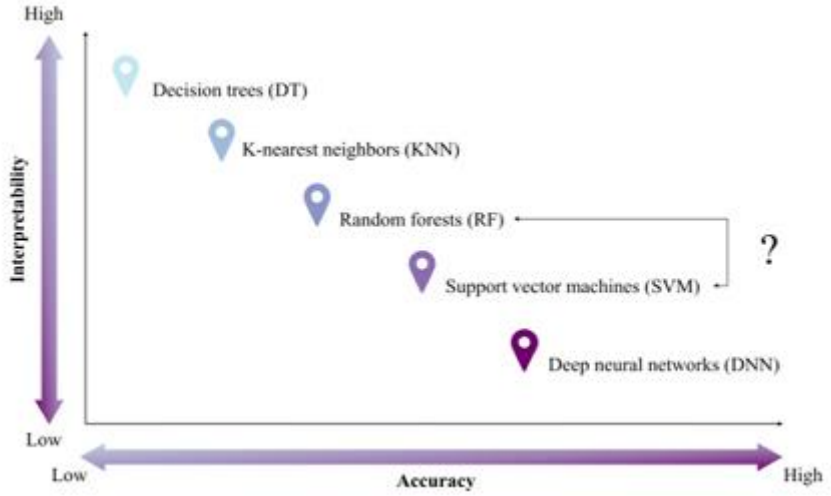

Figure A: Interpretability-accuracy tradeoff in machine learning classification algorithms

(Sheykhmousa et al., 2020)

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação

Universidade Nova de Lisboa