Object-based Classification of High Spatial Resolution Remote Sensing Images in Ethiopia Using Machine Learning Approaches

By Chuying Lu

A thesis submitted in partial fulfillment of the requirements for the degree of Master of Science (School for Environment and Sustainability) at the University of Michigan April 2019

Faculty advisors:

Professor Daniel G. Brown (Chair)

Assistant Professor Meha Jain

Abstract

Remote sensing image classification is the important process of extracting land use and land cover (LULC) information and has been widely used in a range of fields. With the availability of high spatial resolution images, object-based image analysis together with machine learning classification algorithms has received increasing attention and use.

The main goal of this research is to conduct supervised object-based classification experiments based on Random Forest (RF) and Support Vector Machine (SVM) on high spatial resolution images in Benishangul (BG), Gambella (GM), Oromia (OR), Ethiopia. Performance of the classifiers were compared through analyzing the classification results. Multi-variate linear regression models were built to explore the relationships between factors and classification performance. Two questions were addressed: Are SVM or RF appropriate to be applied to mapping LULC in Ethiopia? and What factors influence classification results? Another objective was to explore the possibility to improve classification performance in terms of accuracy of features extracted. Temporal features were included and the effectiveness of which was examined. When trained the data without temporal features, the mean overall accuracy is 0.72 for SVM, 0.74 for RF. The effectiveness of the two classification approaches differed by site. They were significantly difference in OR and GM, where SVM overperformed RF. Because the dataset was unbalanced, SVM had an advantage. The results of the linear regression analysis suggested that the area of class and sample counts had notable impacts on classification performance. Inclusion of temporal features improved results when using SVM, but had little influence on RF.

Acknowledgement

Special thanks to Dr. Daniel G. Brown with all support and guidance on my thesis. Thanks for providing me with the resource, space and work opportunity in ESA lab. I greatly appreciate Dr. Meha Jain who provide help and advice in my research process.

Also, I want to say great thanks to Jonathan Sullivan, Nathan Chesterman, Stephanie Miller, Pablo Nunez, Daniel Wu, Alyssa Roest. Thank you for collaboration in preparing reference data for my experiments. The precious advice you provide help me a lot.

My great appreciation is sent to my friend Yilun Zhao, Zijun Yang, Weiqi Zhou, Huayun Zhou, Qihong Dai for encouragement and help in preparing capstone.

Great thanks to Jennifer Taylor for always patient in helping me solve problems. I thank SEAS and University of Michigan, for providing me with rich of resources and perfect learning environment. Two years' experience here makes me figure it out what kind of person I want to be in the future.

Finally, I want to specially to thank my family, my parents and grandmother, who believe in me, and give me endless love and support.

Table of contents

List of Tables

List of Figures

Introduction

Background and Literature Review

The definition of Land use and Land cover (LULC) covers two separate concepts. Land cover indicates the physical land type such as forest or open water, whereas land use documents how people use the land. (NOAA 2009).

LULC has been regarded as the part of critical information when addressing the impacts and driving forces of LULC change.

Regarding ecological impacts, LULC changes directly affect the status and integrity of global ecosystems and their capacity to supply ecosystem services (Tolessa et al. 2017). For example, Rapid urban sprawl has caused loss of habitat and influence accessibility of food. Under environmental concern, LULC change can also have negative impacts on environment. For instance, increasing agricultural land contributes to the excessive runoff released into water, which cause serious water pollution such as eutrophication. (Ahearn et al. 2005; Keeney and DeLuca 1993; Johnson et al. 1997). Considering the case of Ethiopia, the destruction and fragmentation of shrubland and natural grassland led to the decline of wild plants and, also increased soil erosion, the volume of surface runoff, and sediment transport in the landscape and, consequently, affected the levels and water quality of the lakes found in the rift floor (WoldeYohannes et al. 2018).

Along with impacts, it is necessary to consider driver forces of LULC change, especially for developing countries like Ethiopia. Generally, LULC change is triggered by a complex mixture of political, social, economic and biophysical factors. (Geist et al. 2006). When we investigated Ethiopia, rapid population growth and land policy reform are two critical factors (Gessesse and Bewket 2014). In one aspect, food demand increased in the past few decades as the population experienced exponential growth. In response to the pressure, more and more lands are exploited for grazing and farming. (Urgesa et al. 2016; Nyssen et al. 2004; Gessesse and Bewket 2014). Secondly, land tenure arrangement affects the utilization of land resources and land management investment decisions (Gessesse and Bewket 2014). Before the land reform took place in 1975, Ethiopia had a complex and unsafe tenure system. The local peasants did not own land rights; arbitrary evictions were common. The extreme inequality of the tenure system resulted in the land underutilized and barren. (Deininger 2008). The land reform changed the ownership and tenure rights of land. In detailed, Residents were allocated land use rights and short-term leasing, or sharecropping was allowed. However, land cannot be sold, exchanged, or mortgaged. (Hailu 2016).

Extracting reliable LULC information is essential for scientists from different fields. Remote sensing images have now been widely accepted as the most useful source to extract LULC information. Two major approaches included: manual approaches and computer-assisted approaches. The outputs of manual approaches usually rely on analysts' scientific knowledge, general knowledge of the phenomena as well as their experiences. Some limitations are associated: It is time-consuming when analysts were required to deal with the large quantity of data; the outputs were sometimes affected by analysts' subjective consciousness. (Photointerpretation and Remote Sensing Methodology). The other approach is the computer-assisted approaches which are realized by computer algorithms and able to process remote sensing images automatically. The computer-based approaches solved problems, which existed in manual approaches, are expected to provide reliable outputs.

Commented [db1]: Missing the end of the sentence

In terms of processing targets, computer approaches can be categorized into pixel-based approaches and object-based approaches. Pixel-based approaches focus on each pixel within the extent, while object-based approaches concentrate on an object, an aggregating of pixels which share the similar properties.

The relative merits of pixel-based analysis and object-based analysis have been debated a lot. However, Object-based image analysis (OBIA) is now believed to have advantages compared to pixel-based analysis for the following reasons. Basically, the increased variability implicit within high spatial resolution imagery confuses traditional pixel-based classifiers resulting in lower accuracies (Hay and Castilla 2006). Also, if carefully derived, image objects are closely related to real-world objects. Once these objects are derived, topological relationships with other objects, statistical summaries of spectral and textural values, and shape characteristics can all be employed in the classification procedures (Platt and Rapoza 2008).

Among all computer approaches, traditional classification algorithms such as Maximum Likelihood, K-means, ISODATA, have been used a lot in remote sensing classification issues. Whereas previous techniques based on simple data models, which are insufficient to be applied to complicated cases. Moreover, when dealing with recent data sets like high spatial resolution image, previous techniques can be limited when considering speed, accuracy (Camps-Valls and Bruzzone 2009).

Under this concern, more advanced algorithms were needed to solve remote sensing classification problem. In my research, Support Vector Machine (SVM) and Random Forest (RF) were chosen. In the next section, the reasons why I chose these algorithms would be discussed.

Support Vector Machine

Gualtieri and Cromp (1999) presented the first SVM application on remote sensing images in 1998, conducting a classification experiment on hyperspectral images from AVIRIS imaging spectrometer.

After that, SVM received more and more attention due to its ability to reach good classification results even with limited training samples, a common limitation for remote sensing application (Mountrakis et al. 2011). Not like statistical techniques which rely on the prior assumption of the probability of distribution, SVM can minimize classification error on unseen data, that is why SVM has an advantage when training data size is small. For instance, (Foody and Mathur 2004) showed that only a quarter of the original training samples acquired from SPOT HRV satellite imagery was sufficient to produce an equally high accuracy for a two-crop classifier when used SVM.

Furthermore, comparing with traditional classification algorithms, SVM presents advantages when regarding the classification accuracies. For example, a study focusing on evaluating the performance of SVM, normal Bayes (NB), classification and regression tree (CART) and K nearest neighbor (KNN) when conduct object-based classification. The minimum overall accuracy of SVM is about 7% higher than DT and KNN. (Qian et al. 2014)

Another study is about using SVM, GMM(Gaussian Mixture Model), and ML (Maximum Likelihood) to classify TM images. The result showed that the overall accuracy of SVM is approximate 10% higher than GMM and ML. (Hermes et al. 1999)

Recently, SVM has been applied in high spatial resolution image classifications. A study used SVM and OBIA to map mangroves forest on WorldView-2 and QuikBird images. From their results, overall accuracy is higher than 94%(Heumann 2011). Another example is to extract roads from IKONOS images. In their research, a useful framework consisted of object-oriented spectral-structural information for road extraction based on SVMs are implemented, which allow them to get accuracy as high as 90% (Huang and Zhang 2009).

Random Forest

RF is another classification algorithm which received much attention. Several reasons can explain the popularity of RF. Firstly, RF can perform well with even a small number of samples (Waske et al. 2012). Secondly, the computing time of RF is fast (Belgiu and Drăguţ 2016; Du et al. 2015). Moreover, RF can provide the importance rank of variables, which is useful regarding the difference between classes, especially when dealing with remotely sensed data with high dimensionalities. (Belgiu and Drăguţ 2016).

When compared with traditional algorithms, RF also shows comparative advantages. (Cracknell and Reading 2014) applied RF and other four different algorithms including Naive Bayes, k-Nearest Neighbors, Support Vector Machines, and Artificial Neural Networks, on Landsat_7 and Landsat_8 images for geological mapping. Their results indicated that RF got overall accuracies over 0.9 while the other four classifiers had overall accuracies around 0.8.

A number of studies concentrated on using RF on high spatial resolution satellite images analysis. For example, to identify vegetation species and learn high-density biomass on WorldView-2 (Immitzer et al. 2012; Mutanga et al. 2012; Ramoelo et al. 2015); to map forest structure for wildlife habitat analysis using QuickBird and LiDAR (Hyde et al. 2006). Moreover, (Stumpf and Kerle 2011) used RF and object-based image analysis to extract landslide areas that caused by the earthquake in four different cities, on QuickBird and IKONOS images. Proposed workflow resulted in accuracies between 73% and 87% for the affected areas.

Comparing SVM and RF

Support Vector Machine (SVM) and Random Forests (RF) have been compared in classification issues, especially in object-based remote sensing classification, in terms of the accuracy of the classification results, the training time required (Gislason et al. 2006), stabilities of classifiers to changes in the training samples (Chan and Paelinckx, 2008) and study areas (Vetrivel et al. 2015; Belgiu and Drăguţ 2016).

Relative performance of the two approaches depend on research areas, classification target, scale, pixel or object approach, and sensors used. Table 1~~Table 1~~ shows the examples and results of comparison between RF and SVM.

Table 1 Comparison of SVM and RF in literature

Literature	Pixel-based, /Object based	Sensors
SVM outperformed RF		
Hyperspectral Remote Sensing Classifications: A Perspective Survey (Chutia et al. 2016)	Object-based	Earth Observation
RF outperformed SVM		
Comparison of support vector machine, random forest and neural network classifiers for tree species classification on airborne hyperspectral APEX images (Raczko and Zagajewski 2017)	Pixel-based	APEX sensor
Support vector machines to map rare and endangered native plants in Pacific islands forests (Pouteau et al. 2012)	Pixel-based	Worldview_2
No obvious difference		
Land-use/cover classification in a heterogeneous coastal landscape using RapidEye imagery: evaluating the performance of random forest and support vector machines classifiers (Adam et al. 2014)	Pixel-based	Rapid eye
Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China (Feng et al. 2015)	Pixel-based	Unmanned Aerial Vehicle (UAV)

The application in Ethiopia LULC research

Some scientists were interested in LULC classification in Ethiopia. For example, Kindu and Schneider analyzed land use/land cover (LULC) changes in the landscape of Munessa-Shashemene area of the Ethiopian highlands throughout 39 years (1973–2012) using images from Landsat MSS, TM, ETM+, and RapidEye sensors. (Kindu et al. 2013). (Eggen et al. 2016) conducted SVM classification on Landsat images to map LULC on northwestern Highlands, Ethiopia, which got overall accuracy with 0.55.

Among these researches, most of them are based on images with relatively low spatial resolution. The usage of low spatial resolution images makes it hard to identify classes with complex spatial characters. Moreover, only limited classification approaches are applied when solving LULC in Ethiopia. Under these concerns, the possibility of using advanced classification approaches with high spatial resolutions images is worthy to be discussed.

Research Questions and Objectives

There are two primary questions addressed here.

a. Is SVM or RF appropriate to map LULC in Ethiopia? Which classifier performs better when applied with object-based classification? What factors influence the performance of classifiers?

Many articles have recorded the application of RF and SVM on remote sensing classification. Some of them discussed the comparison between RF and SVM classifiers. However, most of them were associated with the pixel-based classification. The application associated with object-based classification is still uncovered. Also, most research conducted on a small region, usually with clear boundaries between different LULC classes. Whether the classifier works well on Ethiopia which has particular LULC pattern is still uncovered by current research.

In most of the cases, researchers paid more attention on classification results, usually ignored factors which might influence classification results. The answer to this is helpful to see if the specific classifier is appropriate to be used in the real-world case

b. Is there any possibility to improve the object-based performance of classification in Ethiopia?

~~c.~~ features extraction, parameters optimization is two processes in object-based classification. Whereas, not a lot of articles discussed these parts comprehensively in relative research. Moreover, possible improvements associated with two processed are seldomly addressed yet.

Formatted: Indent: Left: 0.31", No bullets or numbering

To answer questions, the following are major objectives in this research

a. Train object-based SVM and RF classifiers on data from Ethiopia. Compare the classification results of classifiers. Discuss the eligibility of them on Ethiopia LULC research. Build multivariate models to analyze the relationship between multiple factors classification results.

b. Analyze the usefulness of temporal information in improving classification performance by comparing the results from different classifiers which trained with different features system.

Method

Study Area

My research focused on Land Use Land Change (LULC) in Ethiopia of Sub-Saharan Africa. Ethiopia borders on South Sudan to the west, Djibouti and Eritrea to the north, Somalia to the east and Kenya to the south. The total territory area is approximate 1126829 km^2 with a total population of around 100 million. (The World Factbook).

Geographically, Ethiopia is a mountainous country with the platinum terrain. It has nine major rivers and twelve large lakes. Tropical climate with wide variation makes Ethiopia an ecologically diverse country. The landscape change from the desert along the east to the tropical forest in the west.

Ethiopia consists of nine ethnically-based administrative regional states. The dominant land use and land cover type is agriculture which accounts for 36% of the total areas. The following is Forest which accounts for 12.2%, and other land use covers 51.1 % of total areas.

My experiment sites are in Benishangul (BG), Gambella (GM), Oromia (OR) (**Error! Reference source not found.**). The total areas that experiment cover are around 865 km^2. There are 11 LULC classes within my research sites. The total areas of each class and the percentage are listed in

Formatted: Left

Table 2~~*Table 2*~~.

Formatted: No underline, Do not check spelling or grammar

Benishangul Gumuz (BG), covering an estimated area of 49,289.46 km2, with population 784345. Two experiment sites located on BG, which cover a total area of 203 km^2. The major LULC class in BG is Small-holder agriculture.

Gambela (GM) has a total population of approximately three hundred thousand, and the estimated area of 29,782.82 km^2(Csa 2007). The experiments cover approximate 509 km^2. Within the experiment sites, Woodland/Savanna and Bare soil cover the high proportion of areas with 26% and 23%.

Oromia is the region covers the area with approximate 284,538 km^2, with the population with thirty million. There are two experiments sites in OR with total areas 52 km^2. Small-holder agriculture accounts for half of the total area.

Table 2 The total areas and percentage of LULC classes on experiment site

LULC Class	Area(km^2)	Percentage (%)
Small-holder Agriculture	152.43	17.62
Small-holder Agriculture/Settlement	6.50	0.75
Intensive Agriculture	62.79	7.25
Forest	144.08	16.65
Woodland/Savanna	223.91	25.89
Shrubland/Grassland	47.88	5.53
Bare Soil	157.22	18.17
Rural Settlement	4.93	0.56
Development	0.25	0.03
Water	13.62	0.01
Wetland	51.59	5.96
Total	865.18	100

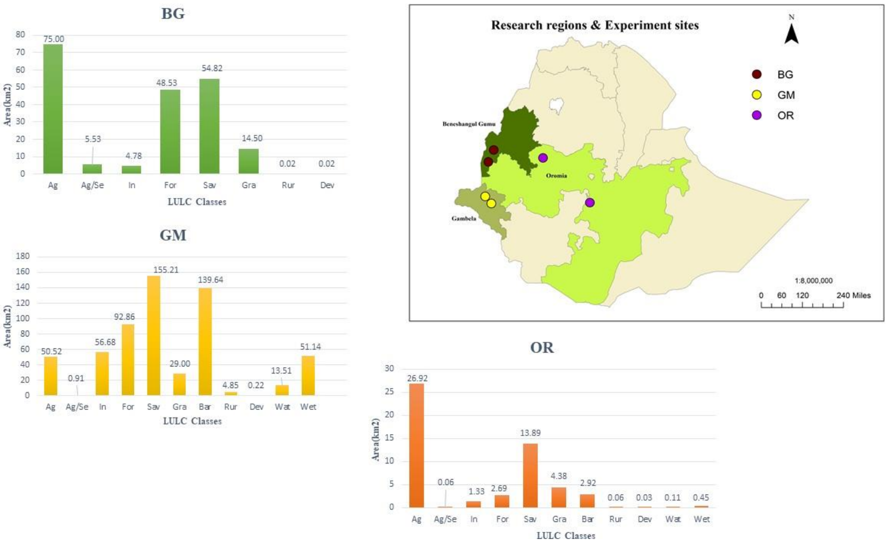


Figure 1 The research regions and experiment sites

Data Sources and Preprocessing

1) High Spatial Resolution Images

There are 28 scenes of high spatial resolution images included in my experiment, which were acquired from QuickBird, WorldView_2, WorldView_1, Geo-eye, IKONOS sensors. These satellite images are provided by the NASA granted project "Large-Scale Land Transactions as Drivers of Land-Cover Change in Sub-Saharan Africa."

There are two main steps in image preprocessing. The first step is orthorectifying, which purpose is to remove the impact of elevation known as relief displacement. This process was accomplished with ERDAS IMAGINE software associated with ground control points extracted from Aster 30 meters global digital elevation model products. The second step is atmospheric correction. Through Atmospheric correction, it firstly converted the DN to top-of-atmosphere radiance and then converted to top of atmosphere reflectance. This process was written as a function in R, and the required parameters were obtained from metadata and Absolute Radiometric Calibration Sheet provided by DigitalGlobe (Kuester et al. 2017).

2) NDVI Time Series Data

Normalized Difference Vegetation Index (NDVI) data were obtained to extract seasonal vegetation change. NDVI data were downloaded from eMODIS collection in Earth Explore website. eMODIS are images composite based on the Moderate Resolution Imaging Spectroradiometer (MODIS) data acquired by the National Aeronautics and Space Administration's (NASA) Earth Observing System (EOS). It provides 10-day interval and 250 meters spatial resolution global products.

The image downloaded have already been orthorectified, and Atmospheric corrected. I sorted and stacked the original NDVI images in ArcMap and then calculate the monthly average NDVI values of each month between 2011 and 2016.

3) Reference Data

The LULC reference data were products of manually labeling and merging segments by researchers in Environment Spatial Analysis Lab in School for Environment and Sustainability, University of Michigan. Considering the seasonal change of Land Use and Land Cover, Google Earth Engine are used as an extra resource to assist researchers to interpret correctly.

Commented [db2]: We have moved toward manual labeling of the OBIA-identified objects. This is more accurate and consistent. If your goal is to test the accuracy of classifications (the labels) it might be better to compare the labels of segments made by humans with those made by machine, so use the segmentation based maps as your reference data.

Image Segmentation

Image segmentation is an image recognition technique which aggregates the pixels with similar characters. As the output of the segmentation process, the entire image is divided into the set of, not overlap, segments.

In this research, I used a segmentation technique named Full Lambda Schedule (FLS) in ERDAS software. FLS segmentation is one of region growth algorithms, which is an efficient way to find the boundary between neighborhood segments and divide the image into homogeneous regions.

The FLS segmentation process was controlled by seven parameters, Shape, Color, Texture, Size, Min, Max, Scale, which are needed to be to be defined by the user. Multiple combinations of parameters result in different segmentation results. The value of parameter represents the relative weight. By giving values, merge cost function in FLS algorithms was determined. The higher values of parameter mean segments would be homogeneous in this parameter. By contrast, lower value means that the segments would be less homogeneous in this parameter.

After the user defining the initial weight, they are standardized, and the sum of these four parameters equals 1. The definition of each parameter was listed in *Table 3*~~*Table 3*~~

To ensure segments is proper for later analysis. I conducted segmentation experiments with multiple sets of parameters and visually check the quality of the segments. Through visually check, the final parameters I applied to all images are listed in *Table 3*~~*Table 3*~~

Table 3 The parameters in FLS segmentation

Parameter	**Definition**	**Optimal parameter value**
Segment Ratio	It determines the average size of the segment. The value means how many pixels that a segment content	1800
Minimum	Specify the minimum size of the segment	10000
Maximum	Specify the minimum size of the segment	200000
Spectral	This is measured as the mean of the values of the pixels in the segments	0.9
Texture	This is measured as the standard deviation of the values of the pixels	0.4
Size	This is measured as the number of pixels in the segment	0.2
Shape	This is a proprietary measurement of the boundary complexity of the segment	0.3

Classification

Classification Features System

Classification features are measurable properties or characteristics between classes. (Bishop 2016). Choosing informative, discriminating and independent classification features is a crucial step in classification. The desirable features should be independent of each other and make classes separable.

Generally, Spectral features, Spatial features the and Shape features were common choices for researchers (Huang and Zhang 2013). In my research, four types of features were considered: 1) Spectral features are statistical summaries of spectral reflectance within an object extent. 2) Location features, which can indirectly reflect the spatial relationship of objects. 3) Shape features, which include area and perimeters of objects were expected to be important in distinguish LULC classes. For example, agriculture has a regular shape and clear boundaries while land covered by forest or woodland have an irregular shape. 4) temporal features, which have not been discussed much yet. In my research, I considered the importance of annulling vegetation difference on different LULC classes, which was indicated by a statistical summary of monthly average NDVI values within the object extent. The Details of each feature was listed in Table 4~~Table 4~~

Table 4 Classification features system

Category	Feature	Statistical summary
Spectral	Blue band spectral reflectance	Mean, Standard deviation, Range, Sum, Min, Max
	Green band spectral reflectance	
	Red band spectral reflectance	
	NIR band spectral reflectance	
Shape	Area(km^2)	
	Perimeters(km)	
Spatial	Centroid Longitude	
	Centroid Latitude	
Temporal	Monthly mean NDVI values	Mean, Standard deviation, Range, Sum, Min, Max

Data Preprocessing and Features Selection

1) Data Preprocessing

Commented [DB3]: It's not clear what you are preprocessing. The pixels in the image? The segments?

In my experiment, the analysis was conducted on objects. Each segment with a unique label of LULC class and sets of feature values was regarded as a single object.

Before the classification stage, data preprocessing was undertaken on objects' features to ensure the data quality. Firstly, features with null values were filled with zero, and the object has outlier feature values that are over two standard deviation distances from mean were removed.

Data normalization is the process to adjust features values under different scales to common scales. The function of data normalization is to align data into a normal distribution and reduce the influence of outliers.

In general, machine learning algorithms benefit from the standardization of the data set. There are several normalization methods in statistics. In this research, I chose the feature scaling method described as the equation below, for each variable

$$X_{standard} = \frac{Xi - X_{min}}{X_{max-}X_{min}} \quad (1)$$

The feature values are scaled between range 0 to 1 now.

2) Feature Selection

Feature selection is a technique, which chooses or converts original features to a subset of indicators that can best reflect the variance in data. Principal components analysis (PCA) is one of common dimension reduction techniques which can be used as a feature selection method. By processing PCA on features, original features can be transferred to multiple non-liner-correlated components. The components are ranked by the variance it explained in data. The number of components is decided by the accumulated variances. Generally, the numbers of components are chosen when the accumulated variances are over 95%.

Support Vector Machine

Support vector machine is a supervised classification algorithm. It can be used to learn the labeled training data and predicts classes of testing data.

The objective of SVM is to find p-1dimension optimal hyperplanes based on training data sets which are thought to maximize the margin-the distance between the hyperplane and its closest point. Generally, SVM performs well in dealing with two classes classification, if the target is multiple classes problem, Integration strategies are needed to extend this method to classifying multiple classes (Huang et al. 2002).

To present how SVM work, a linear classifier is firstly discussed in this section to demonstrate how SVM works.

In a simple linear separable case, n objects which lie on the plane. Each data point has one feature with value xi ,and yi is the label of this point with value either -1 or 1.

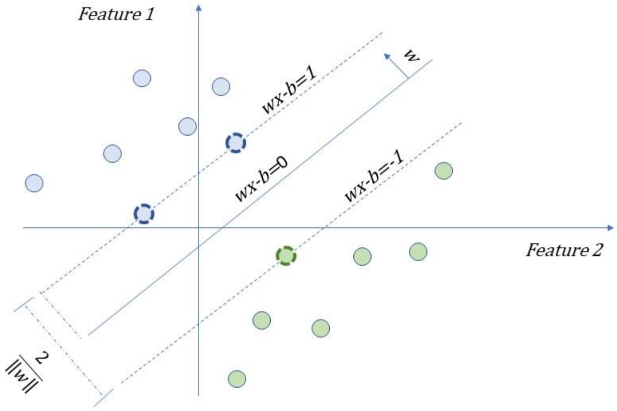


Figure 2 An example of a linear separable classifier

Commented [DB4]: Is this your original figure?

Commented [CL5R4]:

Commented [CL6]: My original figure

Commented [CL7R6]:

There are hyperplane H0 and two plane H1, H2. The vector k is perpendicular to H1 and point to H2 with unit vector w. m is the margin and length of k. The k can be written as:

$$k = \frac{w}{\|w\|} \times m \quad \textit{(2)}$$

And the m can be represented with:

$$m = \frac{2}{\|w\|} \quad \textit{(3)}$$

The purpose of SVM is to find the maximum value of m; it is obvious to see the greater value of |w|, the smaller value of m. The question can be regarded as an optimization question, which is written as:

$$Minimize: F(w) = \frac{1}{2} \times (w'w) \quad \textit{(4)}$$

$$Subject\ to: y_i * (w * x_i + b) - 1 = 0, i = 0,1,\dots,n \quad \textit{(5)}$$

The Lagrangian function is applied to solve the optimization question, which is written as:

$$L(w,b,\alpha)_{primal} = \frac{1}{2}(w'w) - \sum_{i=1}^{n} \alpha_i \{y_i[w'x_i + b] - 1\} \quad \textit{(6)}$$

In this equation, α_i is a positive Lagrangian multiplier. When solve the ∇L(x,y,λ)=0, the minimum value of w can be find when it meets the constraint $y_i * (w * x_i + b) \geq 1$. then minimize Lprimal with respect to w and b to get Wolfe dual Lagrangian (Fletcher 2013) was written as:

$$L(\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \partial_i \partial_j y_i y_j (x_i' x_j) \quad \textit{(7)}$$

The $\partial_i \geq 0$. The training data points lies on planes H1, H2 have ∂_i greater than 0 and are called support vectors. The rest data points have equal to 0, fall on either side of H1 or H2. Then, the solution of w and b are:

$$w = \sum_{i=1}^{nsv} \alpha_i y_i x_i \quad \textit{(8)}$$

In the equation, nsv is the number of support vectors, which is written as:

$$b = -\frac{1}{2} w \times (x_r + x_s) \ \textit{(9)}$$

In this equation, x_r is the data points with y equal to 1. x_s is the data points with y equal to -1 Accordingly, the decision rule to separate two classes which can be derived as:

$$f(x) = sign\left(\sum_{support\ vector} y_i\, \alpha_i^0 (x_i' x) - b^0\right) \quad \textit{(10)}$$

The equation above is called the hard margin formulation; it fits the simple linear separable case. However, no training errors are allowed in the linear separable classifier, and remote sensing classification is a more complicated case. Therefore, linear separable classifier might not eligible. Under this consideration, Kernel Based Non-linear SVM was expected more suitable to be applied**.** In my research, Polynomial and Radius basis function (RBF) are two candidate kernel functions.

Besides, I chose the one-vs-rest strategy to solve the multi-classes problem. In One-vs.-rest strategy a single classifier was trained for per class. Samples from the class were noted as positive while the rest of samples were noted as negative. This strategy requires the base classifiers to produce a real-valued confidence score for its decision, rather than just a class label; discrete class labels alone can lead to ambiguities, where multiple classes are predicted for a single sample (Bishop 2006).

Random Forest

Commented [DB8]:

This is a more appropriate level of detail about them method than all the detail about SVM above.

Random Forest (RF) is an ensemble classification algorithm which is a formed of multiple decision trees, each of tree is trained independently (Du et al. 2015). When training the classifier, a single decision tree was built based on randomly split features and a subset of training samples. In making a prediction, the samples are labeled by each tree in RF, the final class of one sample is the majority vote of the decisions from all trees.

Choosing a subset of data, constructing a single decision tree, obtaining feature importance, are the three most crucial part in RF. The method of each part would be demonstrated in the following paragraphs.

1) Boot-strap Strategies

In a single decision tree, subset data which used as training data is firstly chosen by bootstrap aggregating, which is also known as bagging. This process is utilized to reduce variance, avoid overfitting, thus leads to "improvements for unstable procedures" (Breiman 2001). Supposed we have the overall training dataset H with size n, Subset H' with size n' is generated by sampling from H uniformly and with replacement. As the result, $\frac{2}{3}$ of data in $H'$is expected to contain unique values from H, and rest of the H' are duplicated.

2) Features Split Criterion

When constructing a single decision tree, subset features are random select from all features. Best split of subsets features are assigned to the division of each node. This can decrease the strength of every single tree, but it reduces the correlation between the trees, which reduces the generalization error (Breiman 2001).

Gini impurity and Entropy are two common methods to evaluate the split of features.

a. The Gini impurity index:

$$Gini = 1 - \sum_j p_j^2 \quad (11)$$

b. Entropy:

$$Entropy = -\sum_j p_j log_2 p_j \quad (12)$$

In the above two equations, p_j is the probability that samples from class j being correctly classified. The good split features sets would make two indexes close to 0.

3) Relative Features Importance

Relative feature importance is provided as one part of the result from RF classification. To better understand how feature importance work, it's important to understand out-of-bag (OOB) error. In bootstrap-aggregating, except the chosen subset data, the rest data are used to evaluate the prediction error of RF, in terms of OOB. After OOB being calculated, one of the features will be left out while the rest part stays unchanged, OOB will be calculated again to check whether accuracy decrease. After looping through all features, the rank of feature importance is derived.

In my research, the RF classifier was trained on the major components produced in PCA analysis, so features importance would not be discussed in experiment results part.

Parameters Determination

SVM and RF are both parametric classifiers. A couple of parameters need to be initiated by users. I used a grid search to select the best parameters sets among all candidate sets.

Grid search simulates all possible parameters combination. On each run, the data were split into user-determined folds. Cross-validation was processed, and the main testing over accuracies was ranked. The optimal parameters with the highest mean testing overall accuracy were then determined. For SVM, the multiple combinations of C and gamma were tested. The gamma parameter defines how far the influence of a single training example reaches; the C parameter trades off misclassification of training examples (RBF SVM parameters — scikit-learn 0.1). For RF, several estimators and node spit criterion were combined and tested. The Candidate parameters for SVM and RF were listed in *Table 5*~~*Table 5*~~

Table 5 Candidate parameters for SVM and RF

RF	
Number of estimators	5,10,20,100,1000
Split Criterion	"Entropy" , "Gini"
SVM	
C	1,10,100,1000,10000

gamma	0.001,0.005,0.01,0.05 ,0.1

Classification Assessment and Comparison

1) Classification Assessment

Commented [DB9]: Did you draw point samples and count points to construct the error matrix, or did you overlay the full classifications and sum areas? If the latter, then you don’t need to calculate kappa because your calculation of accuracy is not based on a sample, but on the entire population. In any case, you need to be more clear on how you conducted this analysis.

In the classification process, 50% of data were randomly assigned as training data, the rest of data were assigned as testing data. The classification assessments indexes were calculated after the prediction being made. In order to reduce the effect of randomness, the classifier was trained ten times based on different training data set, and the results that used in assessment are mean values of indexes.

There are two categories of classification assessment indexes. The one is used to evaluate the overall performances of classifiers, which include Overall Accuracies and Kappa Coefficient. The other one is to evaluate the performance of classifiers on single LULC class including Producer Accuracy (PA), User Accuracy (UA), F1 score.

The confusion matrix is used to explore how many samples from different classes been classified. It stored the classification results in an n*n matrix. After normalization, values on diagonal represent the percentage of samples that were classified correctly, while other values mean the percentage of samples been wrongly classified.

The details of each assessment index are introduced below.

Overall Accuracy is a common method to evaluate overall performance despite the influence of a single class. The labels of lands which is predicted by classifiers would be compared with labels of reference data. A number of samples being correctly classified are counted then. The Overall Accuracy can be calculated by dividing the total number of correctly classified samples by the total number of testing samples. Overall Accuracy can be written as

$$OA = \frac{\#\, Total\ number\ correctly\ classified\ smaples}{\#\, Total\ number\ of\ testing\ samples} \quad (13)$$

However, a significant problem with OA for is that some cases may have been allocated to the correct class purely by chance (Congalton 1991; Pontius 2000; Rosenfield and Fitzpatrick-Lins 1986; Türk 1979). Out of this concern, KAPPA coefficient is efficient in solving the effect caused by occasionally. It can be calculated as

$$\widehat{K} = \frac{N\sum_{i=1}^{r} x_{ii} - \sum_{i=1}^{r}(x_{i+} * x_{+i})}{N^2 - \sum_{i=1}^{r}(x_{i+} * x_{+i})} \quad (14)$$

In this equation, N is the total number of samples in the testing dataset. r is the total number of rows; x_{ii} represents the value at diagonal, x_{i+} represent the values on row i except xii. Similarly, x_{+i} represents the values on column i except x_{ii}.

Producer accuracy is used to measure the omission error that how many samples been wrongly classified in other classes. It can be calculated as:

$$PA = \frac{\#\, Samples\ from\ class\ been\ correctly\ classified}{\#\, total\ smaples\ within\ class} \quad (15)$$

User accuracy is used to measure the commission error that data with other classes mistakenly classified in this class. It can be calculated as:

$$UA = \frac{\#\, Samples\ from\ class\ been\ correctly\ classified}{\#\, Total\ samples\ been\ classified\ in\ that\ class} \quad (16)$$

F-1 score is the index which can balance the UA and PA. It provides a user clear way to evaluate the performance of the classifier on individual class. It can be calculated as:

$$F1 = 2 \times \frac{UA \times PA}{UA + PA} \quad (17)$$

2) Classification Accuracy Comparison

McNemar’s test is applied to check whether the classifiers perform differently. McNemar’s test is a nonparametric test based on standardized normal test statistic calculated from error matrices of the two classifiers (Ricotta 2004; de Leeuw et al. 2006). It can be written as:

$$Z = \frac{f_{12} - f_{21}}{\sqrt{f_{12} + f_{21}}} \quad (18)$$

In (33), f_{12} denotes the number of samples been correctly classifier in classifier 1 but wrongly classified in classifier 2. Conversely, f_{21} denotes the number of samples been correctly classified in classifier 2 but misclassified in classifier 1. After that, chi-squared distribution is referenced to check whether two classifiers are significantly different under one degree of freedom. (Abdel-Rahman et al. 2014)

X2, in this case, can be represented as:

$$X^2 = \frac{(f_{12} - f_{21})^2}{f_{12} - f_{21}} \quad (19)$$

Effects on Classification Performance

Another question proposed is the effects on classification performance. The total areas of class, sample counts of class, area standard deviation within the class, are considered as three possible individual factors that can influence classification performance. The syntheses of influence from areas and counts are also considered. The null hypothesis for this question is that the LULC class accuracy has no relationship with any factors mentioned here.

Multi-variate linear regression models are built for different classifiers to test the relationships. Which can be written as:

$$LULC\ class\ accuracy \sim Area + Count + Std_{site\_area} + \ Area * Count \quad (20)$$

Results

Principal Component Analysis

Commented [DB10]: Need some description of what you're trying to do here and what you found.

The purpose of Principal Component Analysis (PCA) is to convert original features to nonlinear correlated components. I conducted PCA on data which contain temporal features, which results are shown in Table 6~~Table 6~~, and features without temporal features separately. The results are shown in Table 7~~Table 7~~.

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Table 6 PCA results of features contain temporal features

Principal Component	BG	GM	OR
1	0.3590	0.4712	0.6988
2	0.1940	0.1439	0.0763
3	0.0825	0.1026	0.0464
4	0.0683	0.0872	0.0322
5	0.0572	0.0416	0.0280
6	0.0491	0.0238	0.0244
7	0.0297	0.0218	0.0243
8	0.0244	0.0178	0.0204
9	0.0234	0.0137	0.0147
10	0.0209	0.0125	0.0109

The values in the table represent the variance that each component explain

Table 7 PCA results of features without temporal features

Principal Component	BG	GM	OR
1	0.5184	0.5693	0.6569

2	0.2079	0.1553	0.1148
3	0.0943	0.1383	0.0691
4	0.0628	0.0413	0.0543
5	0.0356	0.0270	0.0427
6	0.0235	0.0173	0.0202
7	0.0152	0.0138	0.0149
8	0.0095	0.0084	0.0084
9	0.0085	0.0063	0.0063
10	0.0053	0.0054	0.003

The values in the table represent the variance that each component explain

In my research, I chose a number of features which make accumulative variance higher than 95%.

In Table 6~~Table 6~~, The accumulative variances were over 0.95 when chose ten principal components. The results are same for three sites. In Table 7~~Table 7~~, the accumulative variance was over 0.95 when chose five principal components.

Formatted: Font: 12 pt

Formatted: Font: 12 pt

Therefore, ten principal components were selected when trained classifiers based on features which contain temporal features, while five features principal components were selected and used in training classifier based on data without temporal features.

Parameters Determination

Support Vector Machine

In parameters determination, 5-fold cross-validation was processed on all candidate objects. Table 8~~Table 8~~ presents the best parameter set when applying SVM of both cases. The RBF kernel performs better than the other two options in all sites. The selection of C values depends on the specific case. BG; GM (contain temporal features); GM (without temporal features) had their best classifier when C was 1000. BG (without temporal features), OR (contain temporal features), OR (without temporal features) had their best classifiers, when C is 10000. In most of the situations, the construction of best classifiers with gamma 0.1, OR (without temporal features) had its best classifier with gamma 0.05.

Formatted: Font: 12 pt

Table 8 Best parameter sets to build SVM classifier by Grid-search analysis

Site	Contain temporal features or not	C	Gamma parameters	Kernel	Training time(sec)	Best mean test Overall Accuracy (OA)
BG	Features_T	1000	0.1	RBF	1691	0.7
	Features_NT	10000	0.1	RBF	412	0.66
GM	Features_T	1000	0.1	RBF	154	0.75

	Features_NT	1000	0.1	RBF	107	0.73
OR	Features_T	10000	0.05	RBF	158	0.83
	Features_NT	10000	0.1	RBF	117	0.71

Features_T means data contain temporal features, Features_NT means data without temporal features.

Different composite of parameters results in different accuracies. Figure~~Figure~~3, Figure 4~~Figure 4~~,Figure 5~~Figure 5~~, presents the pattern of testing accuracies versus parameters of BG, OR, GM respectively. The patterns are similar regardless of sites and temporal features. When using Polynomial kernel, accuracy change abruptly from very low accuracy to higher when C and gamma increase. When using Linear, the accuracies did not change gradually, it increased as C increase but not influenced by gamma values. When using the RBF kernel, the accuracy changed gradually. It changed from lower values to higher values as the C and gamma increase. Commonly, the highest accuracy occurred with either the highest C value or the highest gamma value.

Formatted: Font: 12 pt, Not Italic

Formatted: Font: 12 pt, Not Italic

Formatted: Font: 12 pt, Not Italic, Do not check spelling or grammar

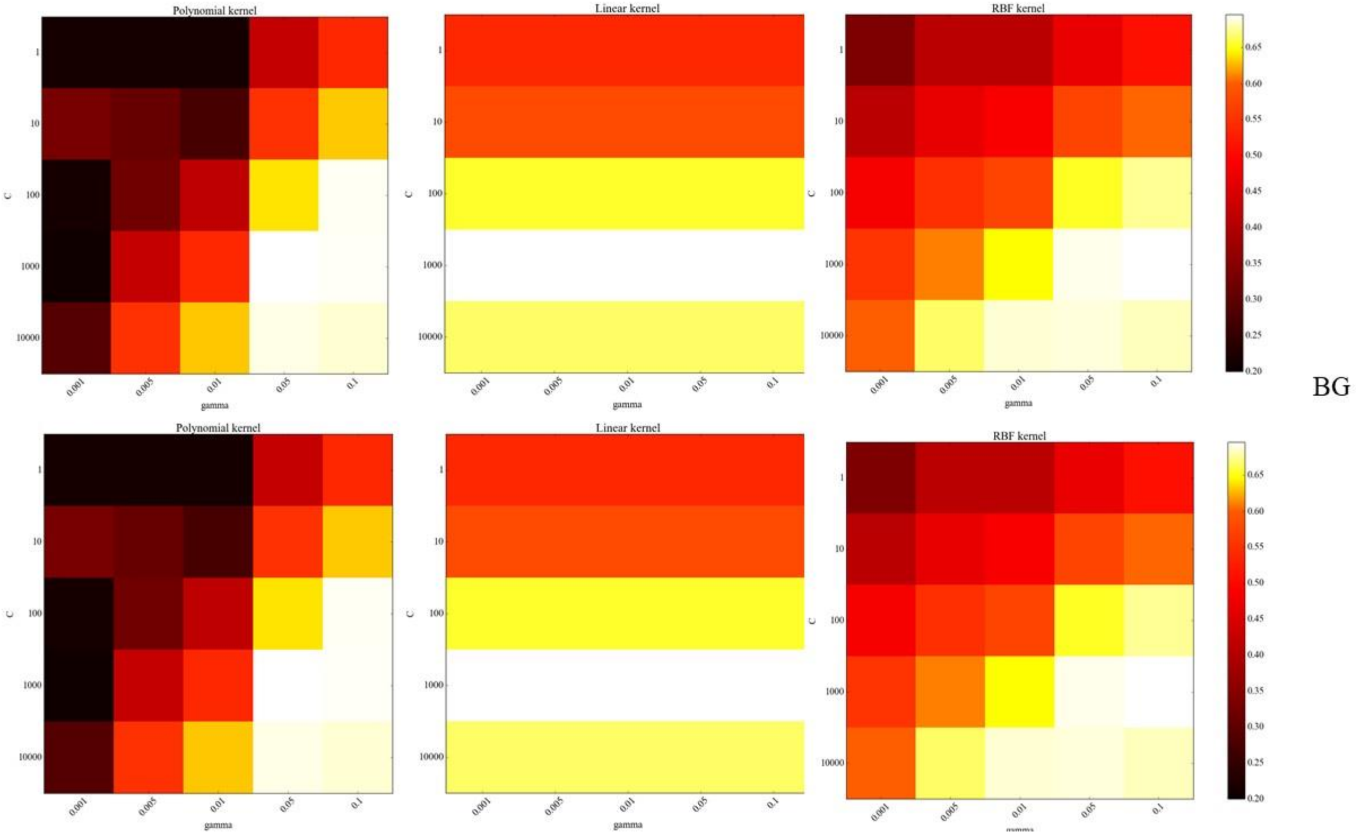


Figure 3 Patterns of mean test accuracy versus parameters in BG site. Data contain temporal features (Upper), Data without temporal features(bottom); Polynomial(left), Linear(middle)r, RBF (right).

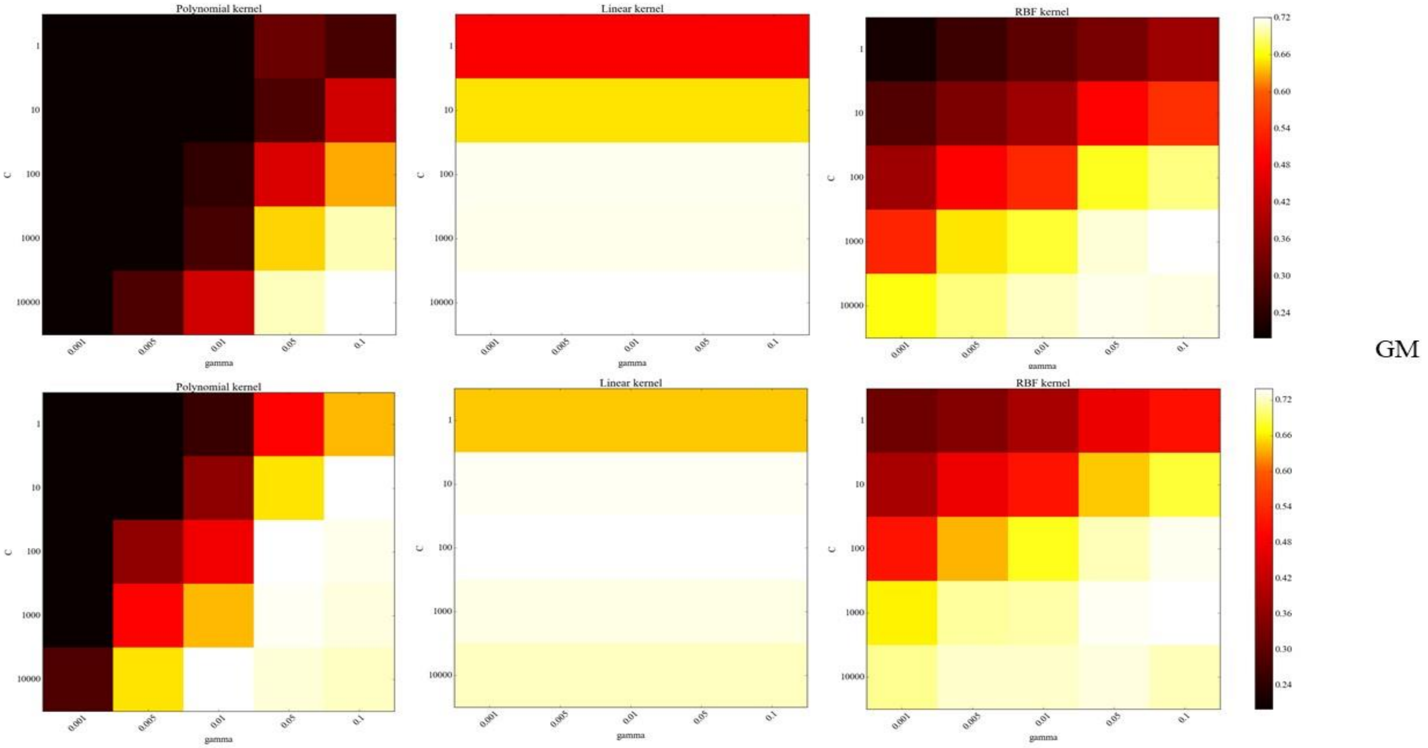


Figure 4 Patterns of mean test accuracy versus parameters in GM site. Data contain temporal features (Upper), Data without temporal features(bottom); Polynomial(left), Linear(middle)r, RBF (right).

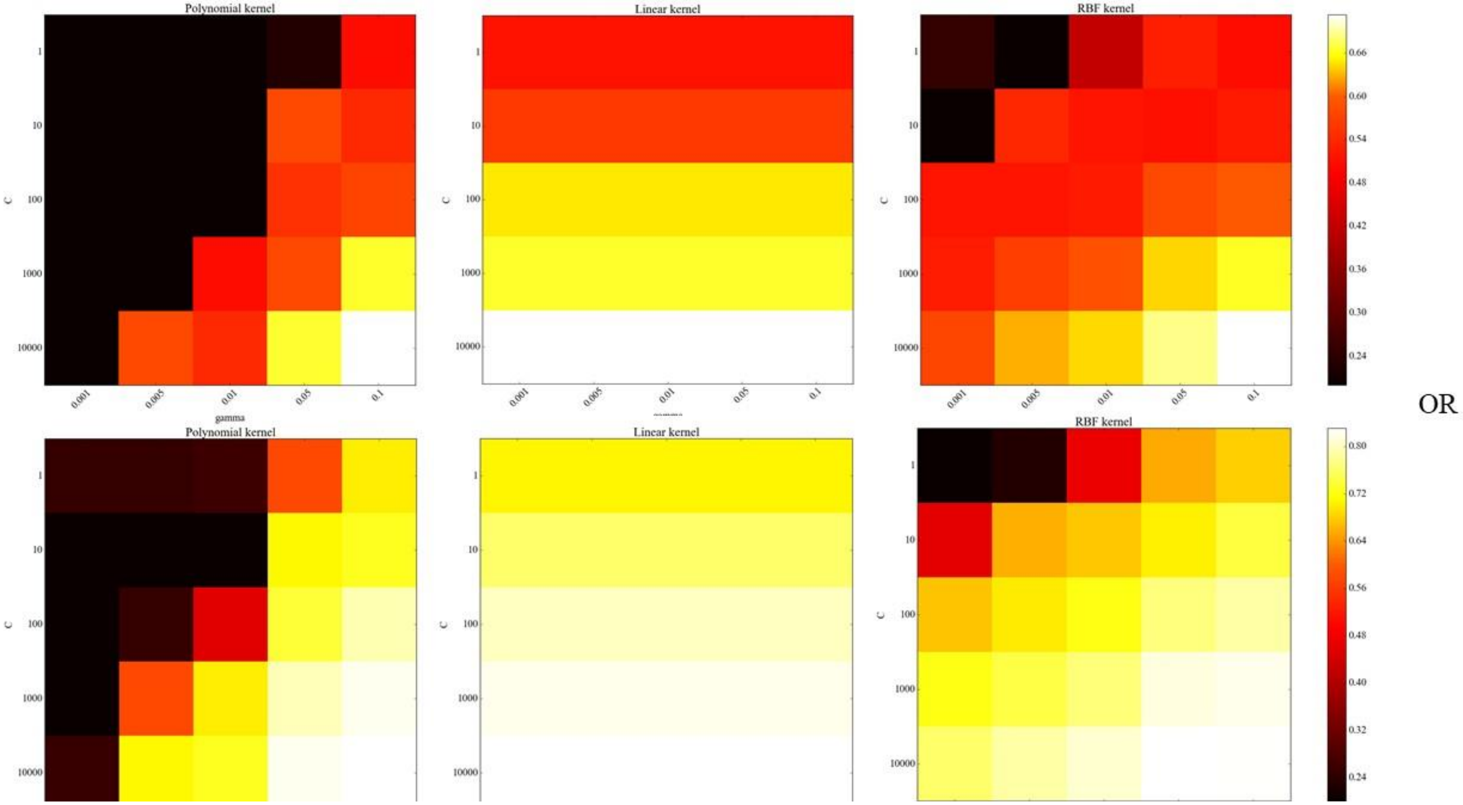


Figure 5 Patterns of mean test accuracy versus parameters in OR site. Data contain temporal features (Upper), Data without temporal features(bottom); Polynomial(left), Linear(middle)r, RBF (right)

Random Forest

Table 9~~*Table 9*~~ presented the best parameter sets for RF. In most cases, best classifiers were built with entropy criterion and 1000 of trees.

Table 9 Best parameter sets to build SVM classifier by Grid-search analysis

Site	Contain temporal features or not	Split criterion	Number of trees	Training time(sec)	Best mean test Overall Accuracy (OA)
BG	Features_T	entropy	1000	410	0.70
	Features_NT	entropy	1000	149	0.68
GM	Features_T	gini	1000	127	0.75
	Features_NT	entropy	1000	101	0.75
OR	Features_T	entropy	100	99	0.86
	Features_NT	entropy	1000	89	0.82

Features_T means data contain temporal features, Features_NT means data without temporal features.

The pattern of parameters versus testing accuracies is plotted in Figure 6~~Figure 6~~. The testing accuracies increased as the number of trees increased. Primarily, with the number of trees increased from 10 to 100, testing accuracies increased steely. With the number of trees increased from 100 to 1000, accuracies did not increase a lot. About node split criterion, it is hard to tell which one performs better. Since, half of them showed that using" Gini" criterion was better, while the rest of them did oppositely. However, only slight differences appeared when using two different criterions.

Formatted: Font: 12 pt, Not Italic

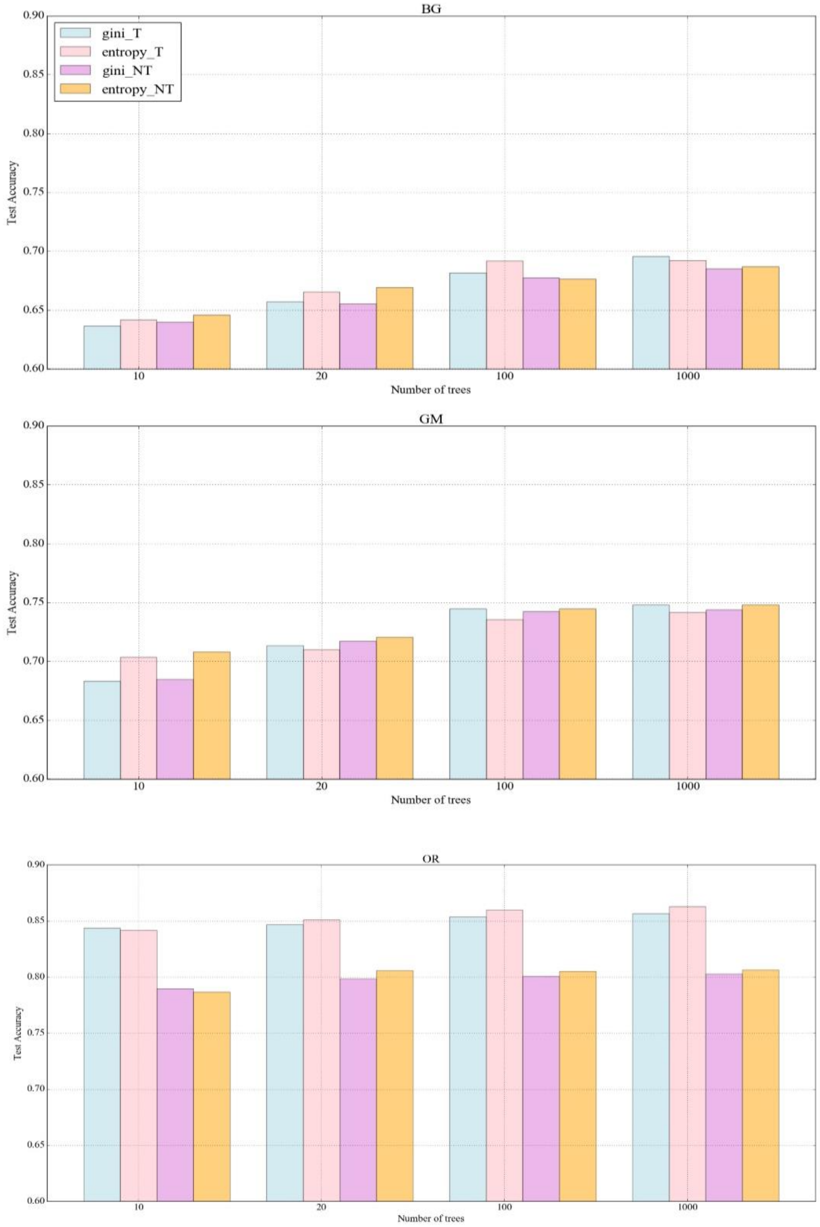


Figure 6 Patterns of testing accuracies versus parameters of RF in BG, GM, OR sites

Classification Results Analysis

Overall Classification Accuracy Assessment

Overall Accuracies and KAPPA coefficients of SVM and RF were listed in *Table 10*~~*Table 10*~~.

Table 10 Overall Accuracy and KAPPA coefficient of different classifiers.

Site	Number of training data	Number of testing data	SVM				RF			
			Features_NT		Features_T		Features_NT		Features_T	
			OA	KAPPA	OA	KAPPA	OA	KAPPA	OA	KAPPA
BG	2718	2718	0.67	0.56	0.71	0.60	0.68	0.54	0.68	0.54
GM	1675	1675	0.74	0.69	0.75	0.70	0.75	0.68	0.74	0.68
OR	1662	1662	0.76	0.51	0.84	0.62	0.80	0.46	0.81	0.47
Total/Average	6055	6055	0.72	0.59	0.76	0.64	0.74	0.56	0.74	0.56

Feature _T represents the features which include temporal features, Feature_NT represents features without temporal features; OA represents overall accuracy, KAPPA represents kappa coefficients.

All classifiers had mean overall accuracies (0.68-0.84). SVM (contain temporal features) reached the highest mean overall accuracy 0.76; the following were RF (contain temporal features) with an accuracy of 0.74, RF (without temporal features) with a mean accuracy of (0.74), SVM (without temporal features) with an accuracy of 0.72.

Regarding the mean KAPPA coefficient, all classifiers got moderate values (0.4-0.6), which means some correctly labeled samples were still classified by chance. SVM (contain temporal features) had the highest KAPPA values (0.64), While SVM (without temporal features) had a little bit lower Kappa (0.59). Two RF classifiers had the lowest KAPPA (0.56) when compared overall accuracy and kappa. The overall accuracies were significantly higher than kappa.

Both classifiers achieved the highest mean overall accuracies in the OR site; the GM site had relatively lower mean overall accuracies, the BG site had the lowest mean overall accuracies for SVM and RF.

Site Classification Accuracy Assessment

In the following section, classification results were analyzed site by site. F1 scores are assisted in evaluating the performance of different classifiers on LULC classes. Also, Normalized confusion matrixes are used to analyze misclassification between classes.

Benishangul-Gumuz (BG)

The classification results of SVM and RF in the BG site are shown in Table 11~~Table 11~~

Table 11 User Accuracies (UA), Producer Accuracies (PA), F1-scores of different classifiers in BG site

LULC Class	**Total number of samples**	**SVM**						**RF**					
		Features_NT			Features_T			Features_NT			Features_T		
		PA	UA	F1	PA	UA	F1	PA	UA	F1	PA	UA	F1
Smallholder Agriculture	1269	0.73	0.74	0.73	0.71	0.80	0.75	0.65	0.80	0.72	0.65	0.80	0.71
Smallholder Agriculture/Settlement	193	0.42	0.44	0.43	0.60	0.63	0.61	0.72	0.23	0.35	0.75	0.23	0.35
Intensive Agriculture	66	0.21	0.24	0.22	0.46	0.53	0.49	0.5	0.03	0.06	0.75	0.05	0.09
Forest	2160	0.85	0.77	0.81	0.82	0.82	0.82	0.77	0.86	0.81	0.76	0.86	0.81
Wood land/Savanna	1143	0.57	0.60	0.59	0.62	0.60	0.61	0.57	0.57	0.57	0.57	0.57	0.57
Shrubland/Grassland	322	0.32	0.44	0.37	0.45	0.35	0.39	0.40	0.16	0.23	0.43	0.19	0.26
Bare Soil	261	0.36	0.37	0.36	0.52	0.37	0.43	0.48	0.18	0.26	0.48	0.18	0.26
Rural Settlement	9	0.27	0.33	0.30	0.43	0.33	0.38	0.00	0.00	0.00	0.00	0.00	0.00
Development	13	0.59	0.31	0.38	0.89	0.62	0.73	0.00	0.00	0.00	1.00	0.08	0.14
Water	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Wetland	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Average		**0.68**	**0.67**	**0.67**	**0.70**	**0.71**	**0.70**	**0.66**	**0.68**	**0.65**	**0.66**	**0.68**	**0.65**

Feature _T represents the data which include temporal features, Feature_NT represents data without temporal features; PA represents the producer accuracy, UA represents the user accuracy

SVM

For both cases, Forest, Smaller-holder Agriculture, Savanna got three highest F1. By contrast, Intensive Agriculture, Rural settlement had lowest F1-scores.

Most of the classes experienced an increase of F1 scores, except for Shrubland/Grassland when temporal features were involved. Small-holder Agriculture/Settlement, Intensive Agriculture, Development had their F1 scores increased significantly.

Figure 7~~Figure 7~~ shows how misclassification occurred in BG site, using SVM.

Formatted: Font: 12 pt

When temporal features were not included, the obvious pattern are: 1) Small-holder Agriculture/Settlement had 28% of samples being misclassified as Small-holder Agriculture; 2) Rural Settlement had 33% of samples being misclassified as Small-holder Agriculture, 33% of samples being misclassified as Small-holder Agriculture/Settlement; 3) Development had 31% of samples being misclassified as Bare Soil.

When temporal features are included, the obvious pattern are: 1) Shrubland/ Grassland had 27% of samples being misclassified as Woodland/Savanna; 2) Rural Settlement had 33% of samples being misclassified as Small-holder Agriculture, 33% of samples being misclassified as Small-holder Agriculture/Settlement;

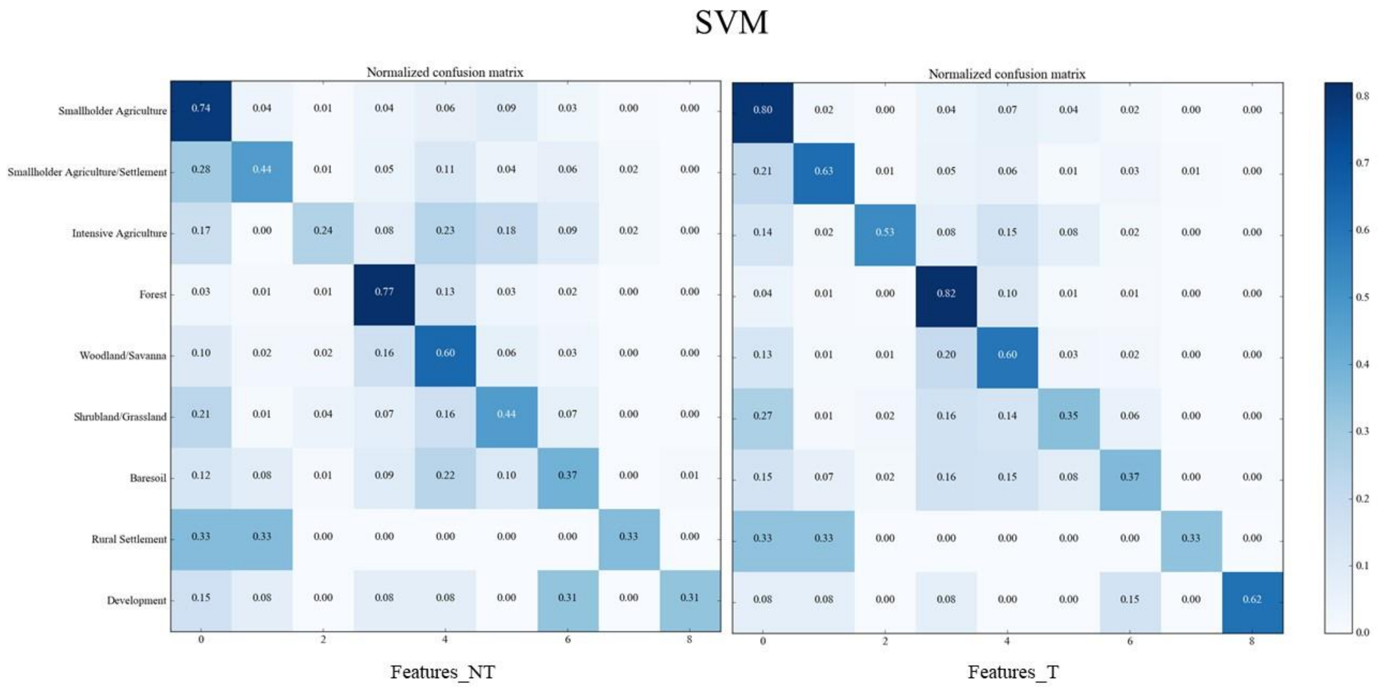


Figure 7 Normalized confusion matrix of SVM classifiers in BG site. Feature _T represents the features which include temporal features(left), Feature_NT represents features which did not include temporal features(right).

RF

As shown in Table 11~~Table 11~~, extreme variances of F1scores occurred between classes, when appliing RF (without temporal features), Forest, Small-holder Agriculture, and Woodland-Savanna got the highest scores with 0.77,0.72,0.65 respectively. Rural Settlement and Development got the lowest scores of F1-scores, which is close to 0.

Formatted: Font: 12 pt

When applied RF (contain temporal features), F1-Score of Intensive Agriculture, Shrubland/Grassland, increased slightly and F1-Score of rest classes did not change at all.

Figure 8~~Figure 8~~ shows how misclassification occurred in BG site, using RF. When temporal features were not included, the obvious pattern are: 1) All classes have samples being misclassified as Small-holder Agriculture, of which, 46% of Small-holder Agriculture, 44% of samples Intensive Agriculture, 56% of Rural Settlement, 0.69 of Development; 2) All classes had samples being misclassified as Woodland/Savanna, of which, 30% of Intensive Agriculture, 28% of Bare Soil.

Formatted: Font: 12 pt

When temporal features are included, the pattern were similar to what mentioned above.

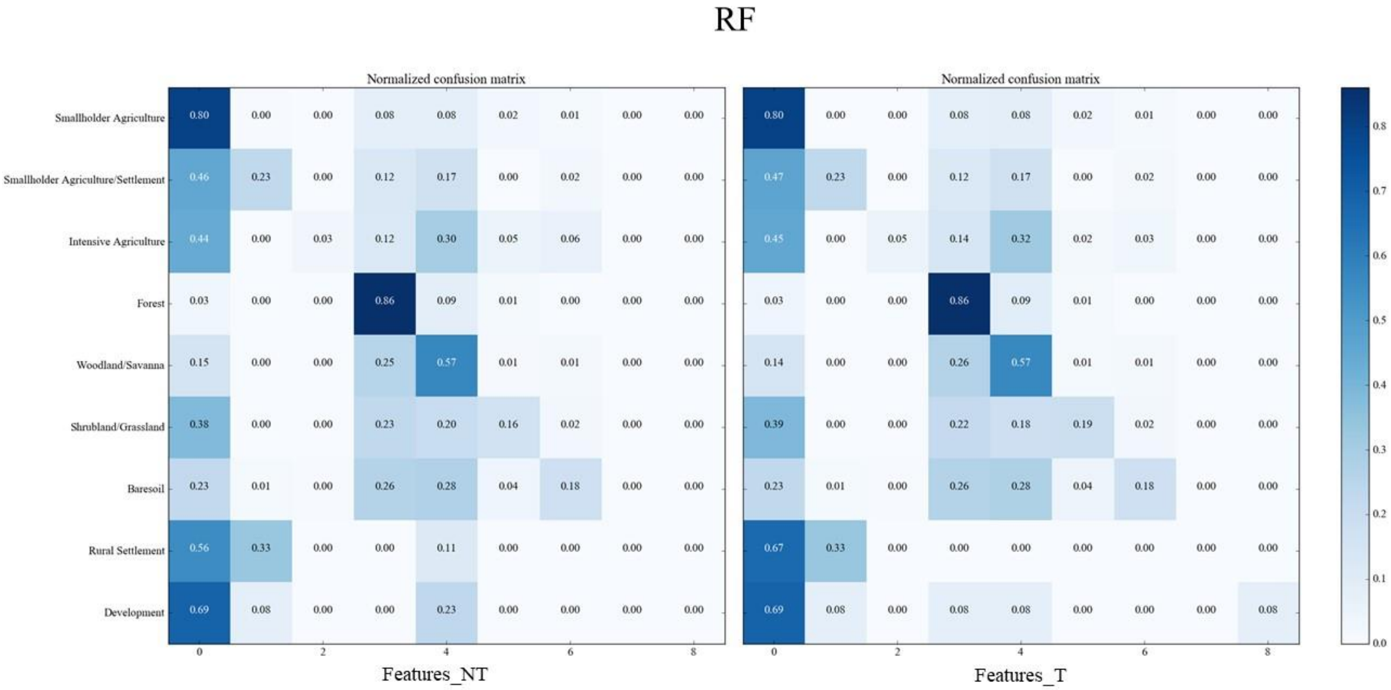


Figure 8 Normalized confusion matrix of RF classifiers in BG site. Feature _T represents the features which include temporal features(left), Feature_NT represents features which didn't include temporal features(right).

Gambela (GM)

The classification results of SVM and RF in the GM site are shown in Table 12~~Table 12~~

Formatted: Font: 12 pt, Do not check spelling or grammar

Table 12 User Accuracies (UA), Producer Accuracies (PA), F1-scores of different classifiers in GM site

	Total number of samples	**SVM**						**RF**					
		Features_NT			Features_T			Features_NT			Features_T		
		PA	UA	F1	PA	UA	F1	PA	UA	F1	PA	UA	F1
Smallholder Agriculture	372	0.57	0.69	0.63	0.66	0.76	0.71	0.65	0.69	0.67	0.64	0.70	0.67
Smallholder Agriculture/Settlement	12	0.33	0.75	0.46	0.33	0.50	0.40	0.50	0.17	0.25	0.40	0.17	0.24
Intensive Agriculture	243	0.76	0.79	0.77	0.87	0.83	0.85	0.87	0.75	0.80	0.88	0.73	0.80
Forest	630	0.86	0.89	0.88	0.83	0.85	0.84	0.88	0.82	0.85	0.87	0.83	0.85
Wood land/Savanna	841	0.73	0.70	0.72	0.69	0.73	0.71	0.68	0.76	0.71	0.67	0.75	0.71
Shrubland/Grassland	162	0.59	0.51	0.54	0.63	0.52	0.57	0.71	0.36	0.48	0.71	0.37	0.49
Bare Soil	676	0.78	0.78	0.78	0.81	0.78	0.79	0.77	0.84	0.80	0.76	0.84	0.80
Rural Settlement	54	0.70	0.65	0.67	0.75	0.76	0.75	0.76	0.54	0.63	0.78	0.54	0.64
Development	11	0.29	0.18	0.22	0.50	0.09	0.15	0	0	0	0	0	0
Water	67	0.74	0.52	0.61	0.76	0.46	0.57	0.93	0.39	0.55	0.93	0.39	0.55
Wetland	282	0.79	0.70	0.74	0.76	0.67	0.71	0.69	0.78	0.73	0.70	0.76	0.73
Average		**0.75**	**0.74**	**0.74**	**0.75**	**0.75**	**0.75**	**0.75**	**0.75**	**0.74**	**0.75**	**0.74**	**0.74**

Feature _T represents the data which include temporal features, Feature_NT represents data without temporal features; PA represents the producer accuracy, UA represents user accuracy.

SVM

As shown, When SVM (data without temporal features) was applied, Forest got the highest F1-score (0.88). The following are Bare Soil, Intensive Agriculture, Wetland. They have lower scores than forest, but all above 0.7. Rural Settlement and Water had F1-score above 0.6, Shrubland/Grassland had F1-score above 0.5. Development had the lowest F1-score (0.22).

Small-holder Agriculture, Intensive Agriculture, Rural settlement experienced a significant increase in F1-score (more than 0.07) when temporal features are included, Shrubland/Grassland, Bare soil experienced a slight increase in F1-score (0.01-0.04). Inversely, the decrease of F1-score happened in Small-holder Agriculture/ Settlement, Forest, Woodland/Savanna, Development, Water, Wetland. Of which, Woodland/Savanna, Wetland experienced an only a slight decrease, with 0.01 and 0.03 respectively; the other classes experienced more than 0.04 decrease. Above all. It is hard to conclude whether temporal features are useful when training data from GM.

Figure 9~~Figure 9~~ shows how misclassification occurred in GM site, using SVM. When temporal features were not included, the obvious patterns are: 1) Development had 0.22 of samples being misclassified as Bare Soil; 2) Water had 0.28 of samples being misclassified as Bare Soil.

Formatted: Font: 12 pt

When temporal features were included, the obvious patterns are: 1) Small-holder Agriculture/Settlement had 33% samples being misclassified as Small-holder Agriculture; 2) Development had 64% samples being misclassified as Rural Settlement.

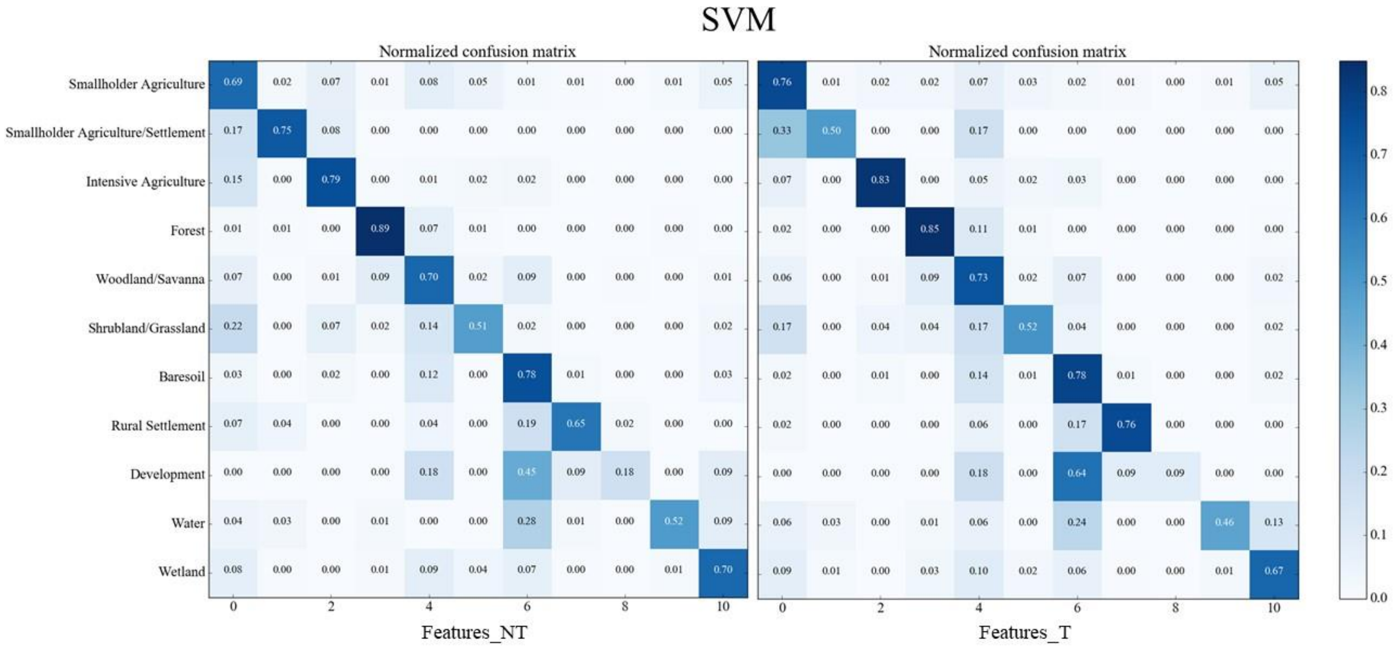


Figure 9 Normalized confusion matrix of SVM classifiers in GM site. Feature _T represents the data which include temporal features(left), Feature_NT represents features without temporal features(right).

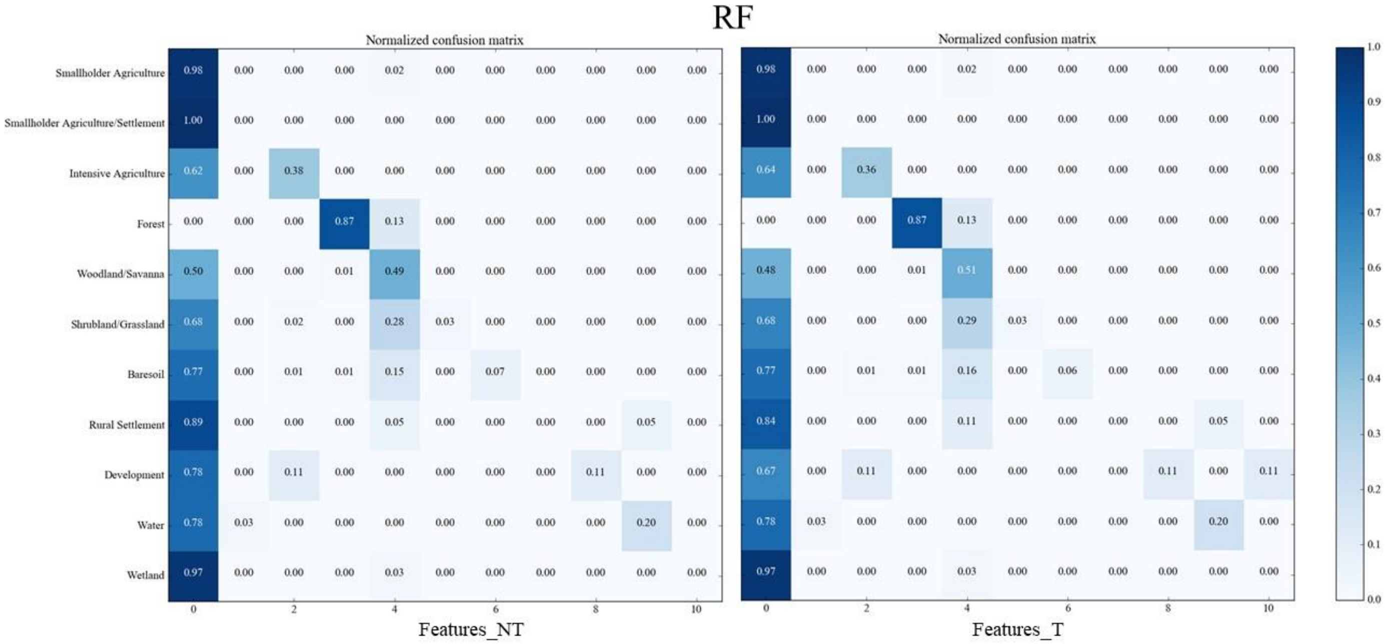
RF

Table 12~~Table 12~~ shows the results when using RF (data without temporal features) to make a prediction, Forest, Intensive Agriculture, Bare Soil had the three highest F1-scores, with 0.85,0.80,0.80 respectively. The following were Small-holder Agriculture, Woodland, Rural settlement, Wetland, which have F1-scores above 0.6. Water and Shrubland had 0.55, 0.48 separately. Small-holder Agriculture/Settlement had low F1-score with 0.25. There was one unexpected result happened on Development, with F1-score equal to 0.

When temporal features were involved, the results show that only Small-holder Agriculture/ Settlement, Shrubland/Grassland, Rural Settlement experienced a slight increase. Other classes stayed unchanged in F1-score.

As shown in Figure 10~~Figure 10~~, the most significant pattern is that except Forest and Small-holder Agriculture itself, samples from other classes were misclassified in Small-holder Agriculture.

Formatted: Font: 12 pt



Smallholder Agriculture	0.98	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Smallholder Agriculture/Settlement	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Intensive Agriculture	0.62	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Forest	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.00	0.00	0.00
Woodland/Savanna	0.50	0.00	0.00	0.01	0.49	0.00	0.00	0.00	0.00	0.00	0.00
Shrubland/Grassland	0.68	0.00	0.02	0.00	0.28	0.03	0.00	0.00	0.00	0.00	0.00
Baresoil	0.77	0.00	0.01	0.01	0.15	0.00	0.07	0.00	0.00	0.00	0.00
Rural Settlement	0.89	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.05	0.00
Development	0.78	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
Water	0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00
Wetland	0.97	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00

0.98	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.64	0.00	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.00	0.00	0.00
0.48	0.00	0.00	0.01	0.51	0.00	0.00	0.00	0.00	0.00	0.00
0.68	0.00	0.00	0.00	0.29	0.03	0.00	0.00	0.00	0.00	0.00
0.77	0.00	0.01	0.01	0.16	0.00	0.06	0.00	0.00	0.00	0.00
0.84	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.05	0.00
0.67	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.11
0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00
0.97	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00

Figure 10 Normalized confusion matrix of RF classifiers in GM site. Feature _T represents the features which include temporal features(left), Feature_NT represents features which did not include temporal features(right)

Oromia (OR)

The classification results of SVM and RF in the OR site are shown in Table 13~~Table 13~~.

Formatted: Font: 12 pt

Table 13 User Accuracies (UA), Producer Accuracies (PA), F1-scores of different classifiers in GM site

	Total number of samples	**SVM**						**RF**					
		Features_NT			Features_T			Features_NT			Features_T		
		PA	UA	F1	PA	UA	F1	PA	UA	F1	PA	UA	F1
Smallholder Agriculture	2295	0.89	0.82	0.85	0.91	0.93	0.92	0.81	0.98	0.89	0.81	0.98	0.89
Smallholder Agriculture/Settlement	21	0.14	0.14	0.14	0.21	0.14	0.17	0.00	0.00	0.00	0.00	0.00	0.00
Intensive Agriculture	45	0.62	0.69	0.65	0.81	0.84	0.83	0.71	0.38	0.49	0.73	0.36	0.48
Forest	62	0.93	0.92	0.93	0.86	0.89	0.87	0.90	0.87	0.89	0.90	0.87	0.89
Wood land/Savanna	520	0.59	0.71	0.64	0.72	0.70	0.71	0.75	0.49	0.59	0.75	0.51	0.60
Shrubland/Grassland	65	0.22	0.15	0.18	0.33	0.28	0.30	1.00	0.03	0.06	1.00	0.03	0.06
Bare Soil	115	0.29	0.50	0.37	0.58	0.50	0.53	0.89	0.07	0.13	1.00	0.06	0.11
Rural Settlement	19	0.10	0.16	0.12	0.47	0.47	0.47	0.00	0.00	0.00	0.00	0.00	0.00
Development	9	0.38	0.33	0.35	0.80	0.44	0.57	0.50	0.11	0.18	0.50	0.11	0.18
Water	40	0.62	0.50	0.56	0.63	0.42	0.51	0.80	0.20	0.32	0.80	0.20	0.32
Wetland	30	0.25	0.3	0.27	0.36	0.39	0.38	0.00	0.00	0.00	0.00	0.00	0.00
Average		**0.78**	**0.76**	**0.77**	**0.83**	**0.84**	**0.84**	**0.79**	**0.80**	**0.76**	**0.79**	**0.81**	**0.76**

Feature _T represents the features which include temporal features, Feature_NT represents features which did not include temporal features; PA represents the producer accuracy, UA represents the user accuracy.

SVM

The range of F1-score was from 0.12 of Rural Settlement to 0.93 of Forest, when I applied SVM (data contain temporal features), the results in The classification results of SVM and RF in the OR site are shown in Table 13.

Formatted: Font: 12 pt

~~The classification results of SVM and RF in the OR site are shown in Table 13.~~

indicates that Forest and Small-holder Agriculture had high F1-score with 0.93 and 0.85 separately. Intensive Agriculture, Woodland/Savanna, Water, had above 0.5 scores. The rest of the classes had F1-score lower than 0.4.

The effects of temporal features are evident in the OR site when applying SVM. The F1-scores of Rural Settlement, Development increased 0.35, 0.22 respectively. Intensive agriculture, Bare soil, Shrubland/Grassland increased with 0.18,0.16, and 0.12. F1-scores of Small-holder agriculture, Small-holder agriculture/Settlement was also had a slight increase.

Figure 11~~Figure 11~~ shows how misclassification occurred in OR site, using SVM. When temporal features were not included, the obvious patterns are: 1) Development had 0.22 of samples being misclassified as Bare Soil; 2) Water had 0.28 of samples being misclassified as Bare Soil.

Formatted: Font: 12 pt

When temporal features were included, the obvious patterns were: 1) Except Forest and Small-holder Agriculture, other classes all had samples being misclassified as Small-holder Agriculture, of which, 48% of Small-holder Agriculture/Settlement, 45% of Grassland/Shrubland, 44% of Development; 2) Small-holder Agriculture had 33% samples being misclassified as Woodland/Savanna.

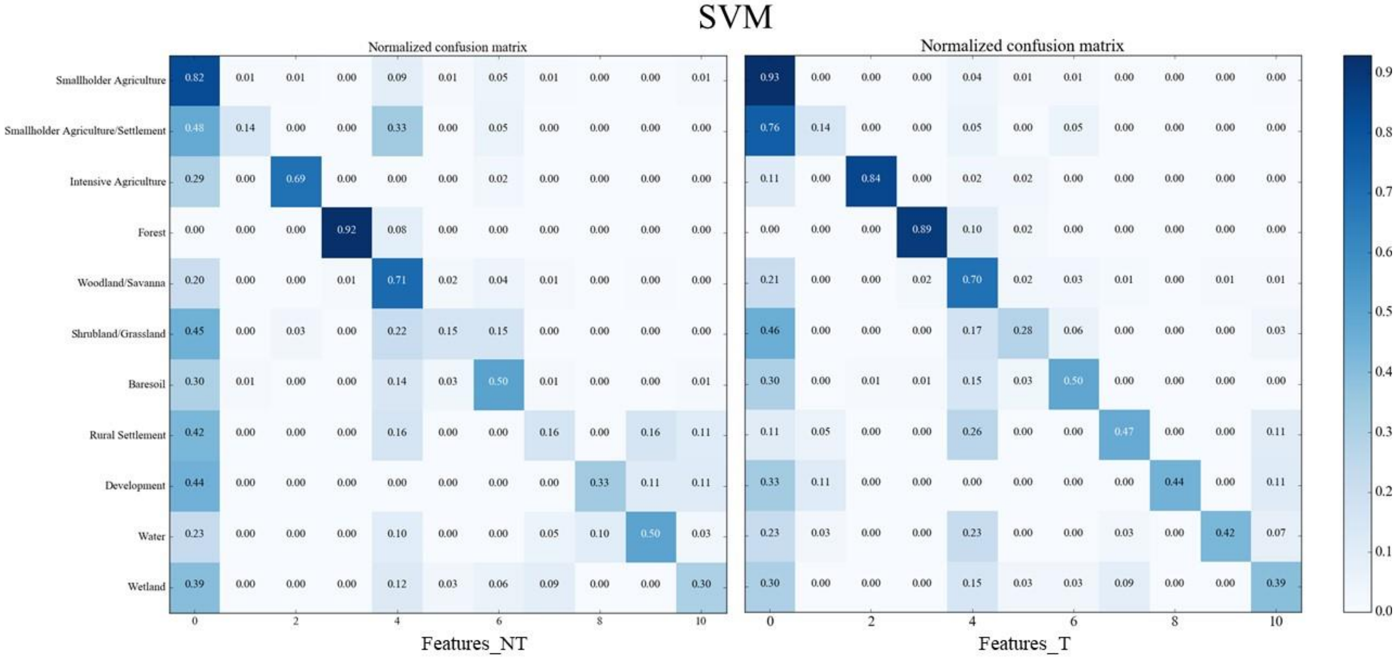


Figure 11 Normalized confusion matrix of SVM classifiers in OR site; left: Features_NT (data without temporal features); right: Features_T (data contain temporal features)

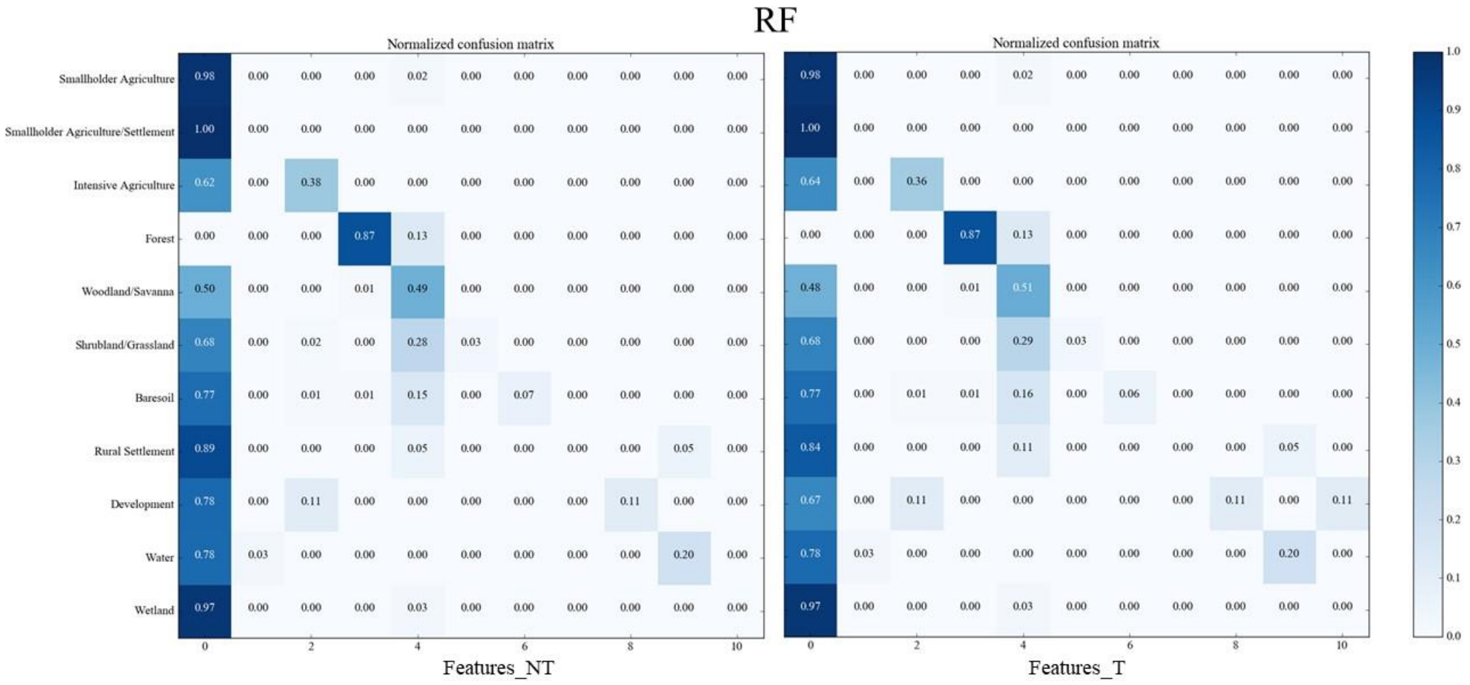
RF

When RF (data without temporal features) was applied on data, results in Table 13~~Table 13~~ indicate that severe disparity of F1 scores appeared between classes. Small-holder Agriculture and Forest had much higher F1 score than other classes. By contrast, Bare Soil, and Development had only 0.13,0.18 of F1-score. Moreover, samples from Small-holder Agriculture/Settlement, Rural Settlement, Wetland, were misclassified into other classes. When the effect of temporal features was analyzed, we can see that there is no increase of F1 score for any class.

Formatted: Font: 12 pt

Figure 12~~Figure 12~~ shows that Small-holder Agriculture had dominant impacts on most of the classes, which made it difficult to separate them apart, under this situation. Only one exception is Forest.

Formatted: Font: 12 pt



	0		2		4		6		8		10
Smallholder Agriculture	0.98	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
Smallholder Agriculture/Settlement	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Intensive Agriculture	0.62	0.00	0.38	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Forest	0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.00	0.00	0.00
Woodland/Savanna	0.50	0.00	0.00	0.01	0.49	0.00	0.00	0.00	0.00	0.00	0.00
Shrubland/Grassland	0.68	0.00	0.02	0.00	0.28	0.03	0.00	0.00	0.00	0.00	0.00
Baresoil	0.77	0.00	0.01	0.01	0.15	0.00	0.07	0.00	0.00	0.00	0.00
Rural Settlement	0.89	0.00	0.00	0.00	0.05	0.00	0.00	0.00	0.00	0.05	0.00
Development	0.78	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.00
Water	0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00
Wetland	0.97	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00

0		2		4		6		8		10
0.98	0.00	0.00	0.00	0.02	0.00	0.00	0.00	0.00	0.00	0.00
1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.64	0.00	0.36	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
0.00	0.00	0.00	0.87	0.13	0.00	0.00	0.00	0.00	0.00	0.00
0.48	0.00	0.00	0.01	0.51	0.00	0.00	0.00	0.00	0.00	0.00
0.68	0.00	0.00	0.00	0.29	0.03	0.00	0.00	0.00	0.00	0.00
0.77	0.00	0.01	0.01	0.16	0.00	0.06	0.00	0.00	0.00	0.00
0.84	0.00	0.00	0.00	0.11	0.00	0.00	0.00	0.00	0.05	0.00
0.67	0.00	0.11	0.00	0.00	0.00	0.00	0.00	0.11	0.00	0.11
0.78	0.03	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.20	0.00
0.97	0.00	0.00	0.00	0.03	0.00	0.00	0.00	0.00	0.00	0.00

Figure 12 Normalized confusion matrix of SVM classifiers in OR site; left: Data did not include temporal features (Feature_NT); right: Data include temporal features (Features_NT)

Classification Performance Comparison

To test whether the classifiers perform significantly different, I conducted the McNemar's tests on paired classifiers.

The results can refer to the chi-squared table with one-degree freedom. So, if the χ2 statistic is over 3.84, the difference is at 0.05 level significant. The test results are shown in Table 14~~Table 14~~. The highlight values represent significant differences between classifiers at 0.05 level.

Formatted: Font: 12 pt

Table 14 McNemar's test between classifiers of BG, GM, OR sites

Site	Classifiers	χ^2
BG	SVM(Features_T) vs SVM(Features_NT)	0.7025
	RF(Features_T) vs RF(Features_NT)	1.7857
	SVM(Features_T) vs RF(Features_T)	**4.1896**
	SVM(Features_NT) vs RF(Features_NT)	0.8120
GM	SVM(Features_T) vs SVM(Features_NT)	0.1632
	RF(Features_T) vs RF(Features_NT)	0.4285
	SVM(Features_T) vs RF(Features_T)	0.1633
	SVM(Features_NT) vs RF(Features_NT)	1.0464
OR	SVM(Features_T) vs SVM(Features_NT)	**72.1153**
	RF(Features_T) vs RF(Features_NT)	1.6
	SVM(Features_T) vs RF(Features_T)	**38.0802**
	SVM(Features_NT) vs RF(Features_NT)	**11.4131**

Feature _T represents the features which include temporal features, Feature_NT represents features which didn't include temporal features.

Regarding the difference between RF and SVM, the difference was significant at 0.05 level when applied BG (data contain temporal features), OR (data contain temporal features), OR (data without temporal features).

Regarding the effects of temporal features on classifiers, the difference is only significant when applied SVM on the OR site.

Multiple Effects on Classification Accuracy

To investigate the relationship between multiple factors and classification accuracies, I built four multivariate linear regression models. The results were shown in Table 15~~Table 15~~

Table 15 The relationships between classification accuracies and variables

Classifier	Variable	Coefficient	t value	p-value
SVM	Area(km^2)	0.55	2.476	**0.0207**
(Features_T)	Counts	0.0002	2.327	**0.0287**
	Site_Area_std	-0.208	-1.055	0.3018
	Area*Count	0.0006	-1.849	**0.0768**
	(Intercept)	0.54	8.488	**1.09e-08**
	R^2: 0.373	**F-statistic:3.57**	**p-value:0.0202**	
SVM	Area	0.62	2.791	**0.0101**
(Features_NT)	Counts	0.0003	2.714	**0.0121**
	Site_Area_std	-0.021	-0.107	0.9158
	Area*Count	-0.0006	-2.108	**0.0457**
	(Intercept)	0.39	6.031	**3.15e-06**
	R^2: 0.4741	**F-statistic:5.409**	**p-value:0.0029**	
RF	Area	0.6440	2.472	**0.0209**
(Features_T)	Counts	0.0003	2.823	**0.0094**
	Site_Area_std	0.1987	0.837	0.4106
	Area*Count	-0.0007	-1.955	**0.0623**
	(Intercept)	0.2553	3.314	**0.0029**
	R^2: 0.5031	**F-statistic:6.074**	**p-value:0.001595**	
RF	Area	0.0064	3.890	**0.0006**
(Features_NT)	Counts	0.0003	2.488	**0.0202**
	Site_Area_std	0.0014	2.717	**0.0120**
	Area*Count	7.243e-06	0.605	0.5510
	(Intercept)	0.29	-1.918	0.0670
	R^2: 0.4881	**F-statistic:5.72**	**p-value:0.002222**	

Feature _T represents the features which include temporal features, Feature_NT represents features which didn’t include temporal features.

To analyze the effects on different cases. The results of each model were discussed separately: 1) SVM (data contain temporal features), The overall p-values was less than 0.05, which means accumulated effects are significant. The R2 was 0.373 indicated that 37.3% of the variance is explained by four variables. Based on the single p-values of each variable, it can be indicated that area, count and area*count have a significant influence on classification accuracies. Of which, area have much more influence than counts and area*counts. 2) SVM (data without temporal features), The accumulate influence from four variables was significant as p-value was lower than 0.05. Except for site_area_std, other three variables had a significant influence on classification accuracy. The area had a higher positive effect on accuracy when compared with other variables. All variables accounted for 47% of the variance in accuracies. 3) RF (data contain temporal features), accumulated effects from four variables was significant as p-value was 0.0015, and 48% difference between accuracies are explained. area, Moreover, counts area and count had significant effects on classification accuracies, while site_area_std was not significant. 4) RF (data without temporal features), the collective effects of four variables was significant and explained 48% of the variance between accuracies. Exclude area*count, other variables had significant but only a slight influence on accuracies.

Alone with the effects on different classifiers, the effect caused by the individual variable is also worthy to consider. 1) Area, the p-values of the area in four models all indicated that area had a significant influence on classification accuracies. It had much more influence than the other three variables. 2) Counts, the p-values of counts in four models all indicated that count is significantly related to classification accuracies. However, the effect on accuracies are small 3) Site_Area_std, this variable is expected to reflect the variance of the area within the class. It only had a significant influence when applying RF (data without temporal features). 4)Area*Count, the effects are significant, except the case which applying RF (data without temporal features). However, the difference was small.

Discussion

Is RF or SVM appropriate to applied on LULC mapping in Ethiopia?

I conducted experiments on Benishangul (BG), Gambella (GM), Oromia (OR), Ethiopia, applying RF and SVM on high spatial resolution satellite images. For each algorithm, two classifiers were built depend on whether data included temporal features or not.

Regarding overall accuracy, the average overall accuracy is 0.72 for SVM (data contain temporal features), and 0.76 for SVM (without temporal features). Both scores are lower than findings from research, which concentrated on applying object-based SVM and high spatial resolution images on smaller scale classification, with overall accuracies around 0.9. (Heumann 2011; Li et al. 2010; Li et al. 2011). However, the overall accuracy of SVM is higher than findings from research, which focused on using SVM to solve large-scale LULC classification in Ethiopia, with an overall accuracy of 0.55 (Eggen et al. 2016). As for RF, the average was 0.74 which are lower than findings from other research. For example, (Rodriguez-Galiano et al. 2012) applied object-based Random Forest approach to mapping LULC classes, Mediterranean. The research areas occupy 12,635 km2, and overall accuracy was 0.92. (Watts and Lawrence 2008)
applied the objected-based RF approaches to map dryland cropping practices within north-central Montana. They got over accuracies over 0.9.

The McNemar's test decided whether the two algorithms performed equally or not by considering the overall disagreement from all samples regardless of the classes. According to the results, the answer is case dependent. Based on my findings, the only significant difference between RF and SVM occurred when fitted to data with temporal features in OR site, which SVM out-competed RF.

Along with the overall performances of classifiers, it is crucial to know how misclassification happened. Some common patterns occurred regardless sites and classifiers: 1) Forest achieved high accuracy due to it outstanding characteristics in spectral features; 3) Among three classes of agriculture, Small-holder agriculture got the highest accuracy, Small-holder Agriculture/Settle and Intensive agriculture were misclassified as Small-holder Agriculture 4) Woodland/Savanna are likely to be mixed with Forest, and Shrubland/Grassland are likely to be mixed with Woodland/Savanna. 5) Bare Soil was easy to be misclassified as savanna. The possible reasons which caused misclassification can be summarized by the following points. Firstly, the misclassification occurred in classes which share similar feature values. For example, three different classes of agriculture lands are hard to be separated due to their similar spectral characters. The second reason is that lands with complicate structures frustrate the classifier to make correct decisions. For example, woody plants grow on agriculture lands can always confuse the classifiers. The third reason is related to blur boundaries between classes. In detailed, one LULC class transit into another LULC class gradually, adjacent areas increase the difficulty of the classifier in identifying. For example, there is usually no clear boundaries between Forest and Savanna.

When analyzing the performances of classifiers, I noticed that samples' distribution influenced the classification accuracy. In my research, the area and count of classes are relatively balanced in GM. By contrast, BG and OR both have dominant LULC classes which make the samples unbalanced. From my experiment results, when fitting the classifier on balanced data, the variance of classification accuracies between classes are small for both SVM and RF. Whereas, the extreme situation happened when used RF to fitted unbalanced data: The dominant classes achieved remarkably high accuracy, while other classes had very low accuracies. This finding supports the idea that SVM has the advantage in dealing with unbalanced data.

On the other hand, when training RF classifiers, training samples were randomly split to the node in a decision tree. Random split eliminates the effect of minority class in making a prediction. A similar finding has already been reached in other research. For example, Porter and YvesMeyer indicated that SVM over-competed RF when used in mapping rare and endangered native plants in Pacific islands forests(Pouteau et al. 2012). Out of this consideration, SVM might be a better choice other than RF, when encountering unbalanced data.

To further exploring what factors can influence classification performance. I build multivariate linear models for each classifier to investigate the relationship between accuracy of each class and four independent variables including total area, counts of segments, the standard deviation of segment areas; combination effects which include area and counts. According to the regression results, I found that the area has significant effects on accuracies, the larger areas that class cover, the higher accuracy that class can reach. Moreover, counts of samples were also proved to have a significant influence on accuracy, but the influence is slight. The effects from the combination of area and counts are also significant but ignorable.

The standard deviation of the segment area did not have a significant influence on accuracy. My findings are consistent with the conclusions reached by Waldner and Jacques. They claimed that the class proportion of the calibration samples, had a stronger impact on classification accuracy than the total number of calibration samples when using machine learning algorithms (Waldner et al. 2017).

There are about 50% of the total variance of accuracies are explained by the four factors mentioned above. Therefore, other factors are possible to influence the classification performance. For example, different scenarios of each site might account for differences. Firstly, three sites have unique geographical conditions which result in a difference of distribution and the forms of LULC classes. For example, Baro river flowing across the GM region, which creates a lot of seasonal wetlands also benefit the distribution of Small-holder Agriculture along the river bank. Furthermore, different forms occurred in the class.

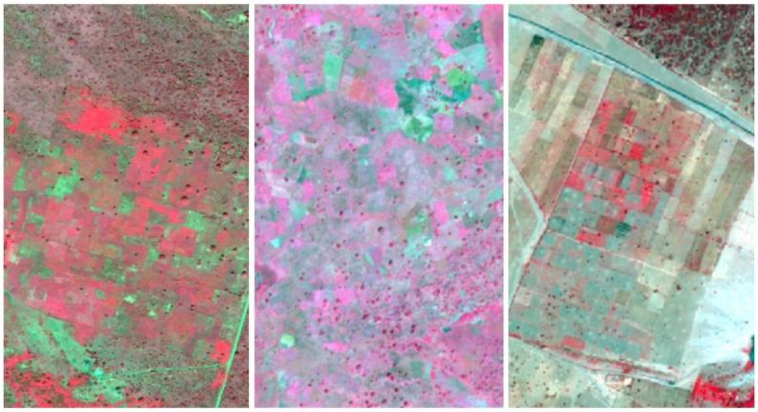


Figure 13 The Small-holder Agriculture in BG, GM, OR

The example (**Error! Reference source not found.**) presents the patterns of Small-holder Agriculture in different sites. In BG and GM, the Small-holder Agriculture had irregularly shaped parcels and had blurred boundaries with surroundings, while in OR, the shape of parcels is regular, well sorted and the had clear boundary with others. Besides the geographical conditions, the satellites images covered three sites were from multiple remote sensors and acquired in different time. Even the digital numbers of all images had been converted to spectral reflectance, the variance from sensors and time are still possible to influence the results.

Thoroughly, whether the SVM or RF is appropriate to map the LULC classes in Ethiopia depends on the specialty of the research site and purposes. As unbalance problem always associated with large-scale LULC classification. If the purpose is to map a continuous and large area of land. SVM had an advantage. Otherwise, if the purpose is to identify the major class within the research area, RF is an excellent choice due to its' time efficiency. It is worthy to notice the differences existed in different locations. When faced with large-scale LULC mapping, these differences are unignorable. The recommended solution is to investigate the variance in a large region and then convert the large-scale questions to multiple smaller scale questions and find a specific strategy for each division.

Is there any chance to improve classification performance, in aspect to the LULC in Ethiopia?

There are two directions were considered in my research to improve the classification performance on high spatial resolution satellite images.

The first direction is to conduct parameters determination experiments before training classifiers. This attempt was motivated by the following reasons: Firstly, SVM and RF are constructed based on multiple parameters, the different combination of parameters can directly influence the performance of algorithms. Under this consideration, Parameters determination is an important process which helps researchers to get best parameter set when fitting different data set. However, this process was seldom discussed in most research, when using advanced machine learning approaches in remote sensing applications. Secondly, parameters determination is restricted by software which the researchers usually choose. This software provides the ready-to-use tools, which make remote sensing analysis easy to undertake. However, it also blocks the chance of researchers, who have interests to know more details of how classifiers work. Moreover, obtaining the authority of software also frustrate researchers to get started.

Fortunately, there are several open-source tools available for researchers to conduct machine learning analysis now, such as Scikit-Learn, Tensor-Flow. These tools are easy to get access to and have already been applied in a wide range of data analysis efficiently. Moreover, they provide tools to conduct parameters determination, which make analysis flexible. In my research, I used grid-search tool provided by Scikit-Learn API to conduct parameters determination.

Based on the results, I found patterns occurred when change parameters. For SVM, the accuracies were significantly influenced by gamma and C which represent the influence of a single sample and the tradeoff of misclassification in training samples respectively. The accuracies increased as C and gamma increased. For RF, the accuracies were influenced by some decision trees and split criterions. The former one had great impacts on accuracies, but the later one only had little influence on accuracy. However, researchers should be cautious when select parameters. Some problems can happen if the parameters are out of specific ranges. When researchers training SVM, if gamma and C are too high, the classifier would be overfitting. If too many trees are involved in RF, the processing speed could be tardy.

Overall, parameters determination is an important process in using machine learning approached to solve remote sensing problems. Open-source tools provide researchers with reliable tools to conduct experiments, also provide researchers with insights into how to get the optimal classification results. However, this process should be undertaken carefully and depend on the researcher's knowledge of algorithms.

The second direction is to add temporal features in original features. There is two motivation for raising this question. Firstly, the information extracted from a single image is limited. It failed to provide how LULC classes change during a year. For instance, when the agriculture lands in fallow seasons, no crops growing on the land. Agricultural lands have no difference with bare soil when considering spectral characters. Under this situation, continuous temporal information is expected to be useful.

In my experiment, two classifiers were established for each algorithm. One of them trained with data contain temporal features, while another trained with data without temporal features. By comparing the classification results, I was able to conclude the effectiveness of temporal features. According to my findings, overall accuracy increased when applying SVM but stayed unchanged when applying RF, which indicating that temporal features work when used in SVM but influence a little when applied to RF. The possible explanation is the split process in establishing decision trees, break down the continuous time series features, so combined effect of temporal features was weakened.

Results from McNamara’s test showed that the effect of temporal features is site dependent. Temporal features significantly improved the classification performances, when trained SVM in BG and OR. In both sites, the dominant classes had great impacts on classifying other classes. The possible reason is that temporal differences assist classifiers to better separate minority classes apart from dominant classes. Whereas the effect of temporal features are not significant in GM when used SVM; the possible reason is that GM had relative balance dataset, the SVM trained with features not include temporal features was already useful to classify classes, thus temporal features might not help.

The effects of adding temporal features on different LULC class were not identical. The effect on Intensive agriculture was apparent which agreed with results from three sites. Notably, some misclassified samples of Intensive Agriculture were extracted from Smallholder Agriculture. Similarly, the effects on Rural-Settlement were also proved useful in most of the cases. Other classes did not show a uniform pattern, but their classification accuracies also experienced increases in either of one site.

Incorporating temporal features in remote sensing classification is still a new topic. From current research, it has been proved an efficient way to map canopy (Karlson et al. 2015). The experiment focused on including temporal features in LULC mapping in Ethiopia has not been covered yet.

Research limitations and Future work

1) The ambiguous definition of classes might influence the classification results. The quantitative definition failed to represent the variances between LULC classes. For example, it only used ranges of vegetation coverage rate to distinguish Forest, Woodland/Savanna, and Shrubland/ Grassland. However, this overlooked the complicated vertical on land. In my research area, there is land which had a vertical structure with grassland on the bottom of layers and forest on the top of layers. Classes with detailed and precise definition are needed. In one aspect, it can guide researchers who participate in manual digitization to make a correct judgment, in

another aspect, it makes it easier for "machine" to understand the boundaries between classes.

2) In my experiments, all classes are treated at the same level — for example, the subdivision of Agriculture, including Small-holder Agriculture, Small-holder Agriculture, Intensive Agriculture, which were treated with the same with other classes like Water, Bare soil. The differences which were expected to distinguish Small-holder Agriculture and Small-holder Agriculture/Settlement could be diminished; when compared with the differences between Agriculture and Water, this can increase the difficulties to form optimal hyperplane in SVM. The possible solution to this is to build multiple hierarchies LULC system and execute classification on classes at the same level.
3) Another problem is associated with the grid-search process. In the grid-search process, even it can provide the optimal parameter sets which reached the highest testing accuracies. The number of parameters sets still need to be pre-defined by users who require users to have a comprehensive understanding of the mechanism behind algorithms — also, the selection based on overall accuracies. Through my analysis, even with high accuracies, it did not mean the performance of the classifier is "good".

Literature Cited

Abdel-Rahman, Elfatih M., Onisimo Mutanga, Elhadi Adam, and Riyad Ismail. 2014. “Detecting Sirex Noctilio Grey-Attacked and Lightning-Struck Pine Trees Using Airborne Hyperspectral Data, Random Forest and Support Vector Machines Classifiers.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 88 (February): 48–59.

Adam, Elhadi, Onisimo Mutanga, John Odindi, and Elfatih M. Abdel-Rahman. 2014. “Land-Use/cover Classification in a Heterogeneous Coastal Landscape Using RapidEye Imagery: Evaluating the Performance of Random Forest and Support Vector Machines Classifiers.” *International Journal of Remote Sensing* 35 (10): 3440–58.

Ahearn, Dylan S., Richard W. Sheibley, Randy A. Dahlgren, Michael Anderson, Joshua Johnson, and Kenneth W. Tate. 2005. “Land Use and Land Cover Influence on Water Quality in the Last Free-Flowing River Draining the Western Sierra Nevada, California.” *Journal of Hydrology* 313 (3): 234–47.

Belgiu, Mariana, and Lucian Drăguţ. 2016. “Random Forest in Remote Sensing: A Review of Applications and Future Directions.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 114 (April): 24–31.

Bishop, Christopher M. 2006. “Graphical Models.” *Pattern Recognition and Machine Learning* 4: 359–422.

———. 2016. *Pattern Recognition and Machine Learning*. Springer New York.

Breiman, Leo. 2001. “Random Forests.” *Machine Learning* 45 (1): 5–32.

Camps-Valls, Gustau, and Lorenzo Bruzzone. 2009. *Kernel Methods for Remote Sensing Data Analysis*. John Wiley & Sons.

Chan, Jonathan Cheung-Wai, and Desiré Paelinckx. 2008. “Evaluation of Random Forest and Adaboost Tree-Based Ensemble Classification and Spectral Band Selection for Ecotope Mapping Using Airborne Hyperspectral Imagery.” *Remote Sensing of Environment* 112 (6): 2999–3011.

Chutia, D., D. K. Bhattacharyya, K. K. Sarma, R. Kalita, and S. Sudhakar. 2016. “Hyperspectral Remote Sensing Classifications: A Perspective Survey.” *Transactions in GIS* 20 (4): 463–90.

Congalton, Russell G. 1991. “A Review of Assessing the Accuracy of Classifications of Remotely Sensed Data.” *Remote Sensing of Environment* 37 (1): 35–46.

Cracknell, Matthew J., and Anya M. Reading. 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 & Geosciences* 63 (February): 22–33.

Csa, Csa. 2007. “The 2007 Population and Housing Census of Ethiopia.” CSA Addis Ababa.

Deininger, Klaus. 2008. “Implementing Low-Cost Rural Land Certification : The Case of Ethiopia,” February. http://hdl.handle.net/10986/9528.

Du, Peijun, Alim Samat, Björn Waske, Sicong Liu, and Zhenhong Li. 2015. “Random Forest and Rotation Forest for Fully Polarized SAR Image Classification Using Polarimetric and Spatial Features.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 105 (July): 38–53.

Du, Shihong, Fangli Zhang, and Xiuyuan Zhang. 2015. “Semantic Classification of Urban Buildings Combining VHR Image and GIS Data: An Improved Random Forest Approach.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 105 (July): 107–19.

Eggen, Michael, Mutlu Ozdogan, Benjamin Zaitchik, and Belay Simane. 2016. “Land Cover Classification in Complex and Fragmented Agricultural Landscapes of the Ethiopian Highlands.” *Remote Sensing* 8 (12): 1020.

Feng, Quanlong, Jiantao Liu, and Jianhua Gong. 2015. “Urban Flood Mapping Based on Unmanned Aerial Vehicle Remote Sensing and Random Forest Classifier—A Case of Yuyao, China.” *WATER* 7 (4): 1437–55.

Fletcher, R. 2013. *Practical Methods of Optimization*. John Wiley & Sons.

Foody, Giles M., and Ajay Mathur. 2004. “Toward Intelligent Training of Supervised Image Classifications: Directing Training Data Acquisition for SVM Classification.” *Remote Sensing of Environment* 93 (1): 107–17.

Geist, Helmut, William McConnell, Eric F. Lambin, Emilio Moran, Diogenes Alves, and Thomas Rudel. 2006. “Causes and Trajectories of Land-Use/Cover Change.” In *Land-Use and Land-Cover Change: Local Processes and Global Impacts*, edited by Eric F. Lambin and Helmut Geist, 41–70. Berlin, Heidelberg: Springer Berlin Heidelberg.

Gessesse, B., and W. Bewket. 2014. “Drivers and Implications of Land Use and Land Cover Change in the Central Highlands of Ethiopia: Evidence from Remote Sensing and Socio-Demographic Data Integration.” *Ethiopian Journal of the Social Sciences and Humanities* 10 (2): 1–23.

Gislason, Pall Oskar, Jon Atli Benediktsson, and Johannes R. Sveinsson. 2006. "Random Forests for Land Cover Classification." *Pattern Recognition Letters* 27 (4): 294–300.

Hailu, Zerfu. 2016. *Land Governance Assessment Framework Implementation in Ethiopia*. Other Rural Study. World Bank.

Hay, G. J., and G. Castilla. 2006. "Object-Based Image Analysis: Strengths, Weaknesses, Opportunities and Threats (SWOT)." In *Proc. 1st Int. Conf. OBIA*, 4–5.

Hermes, L., D. Frieauff, J. Puzicha, and J. M. Buhmann. 1999. "Support Vector Machines for Land Usage Classification in Landsat TM Imagery." In *IEEE 1999 International Geoscience and Remote Sensing Symposium. IGARSS'99 (Cat. No.99CH36293)*, 1:348–50 vol.1.

Heumann, Benjamin W. 2011. "An Object-Based Classification of Mangroves Using a Hybrid Decision Tree—Support Vector Machine Approach." *Remote Sensing* 3 (11): 2440–60.

Huang, C., L. S. Davis, and J. R. G. Townshend. 2002. "An Assessment of Support Vector Machines for Land Cover Classification." *International Journal of Remote Sensing* 23 (4): 725–49.

Huang, X., and L. Zhang. 2013. "An SVM Ensemble Approach Combining Spectral, Structural, and Semantic Features for the Classification of High-Resolution Remotely Sensed Imagery." *IEEE Transactions on Geoscience and Remote Sensing: A Publication of the IEEE Geoscience and Remote Sensing Society* 51 (1): 257–72.

Hyde, Peter, Ralph Dubayah, Wayne Walker, J. Bryan Blair, Michelle Hofton, and Carolyn Hunsaker. 2006. "Mapping Forest Structure for Wildlife Habitat Analysis Using Multi-Sensor (LiDAR, SAR/InSAR, ETM+, Quickbird) Synergy." *Remote Sensing of Environment* 102 (1): 63–73.

Immitzer, Markus, Clement Atzberger, and Tatjana Koukal. 2012. "Tree Species Classification with Random Forest Using Very High Spatial Resolution 8-Band WorldView-2 Satellite Data." *Remote Sensing* 4 (9): 2661–93.

Johnson, Lucinda, Carl Richards, George Host, and John Arthur. 1997. "Landscape Influences on Water Chemistry in Midwestern Stream Ecosystems." *Freshwater Biology* 37 (1): 193–208.

Karlson, Martin, Madelene Ostwald, Heather Reese, Josias Sanou, Boalidioa Tankoano, and Eskil Mattsson. 2015. "Mapping Tree Canopy Cover and Aboveground Biomass in Sudano-Sahelian Woodlands Using Landsat 8 and Random Forest." *Remote Sensing* 7 (8): 10017–41.

Keeney, D. R., and T. H. DeLuca. 1993. "Des Moines River Nitrate in Relation to Watershed Agricultural Practices: 1945 Versus 1980s." *Journal of Environmental Quality* 22: 267–72.

Kindu, Mengistie, Thomas Schneider, Demel Teketay, and Thomas Knoke. 2013. “Land Use/Land Cover Change Analysis Using Object-Based Classification Approach in Munessa-Shashemene Landscape of the Ethiopian Highlands.” *Remote Sensing* 5 (5): 2411–35.

Kuester, Michele A., Miguel Ochoa, Alberto Dayer, Jared Levin, David Aaron, Dennis L. Helder, Larry Leigh, et al. 2017. “Absolute Radiometric Calibration of the DigitalGlobe Fleet and Updates on the New WorldView-3 Sensor Suite.” In *Report of JACIE Civil Commercial Imagery Evaluation Workshop of DigitalGlobe Inc.; DigitalGlobe Inc.: Westminster, CO, USA*. https://calval.cr.usgs.gov/wordpress/wp-content/uploads/JACIE2016_Kuester_V2.pdf.

Leeuw, J. de, H. Jia, L. Yang, X. Liu, K. Schmidt, and A. K. Skidmore. 2006. “Comparing Accuracy Assessments to Infer Superiority of Image Classification Methods.” *International Journal of Remote Sensing* 27 (1): 223–32.

Li, Haitao, Haiyan Gu, Yanshun Han, and Jinghui Yang. 2010. “Object-Oriented Classification of High-Resolution Remote Sensing Imagery Based on an Improved Colour Structure Code and a Support Vector Machine.” *International Journal of Remote Sensing* 31 (6): 1453–70.

Li, M., X. Zhou, X. Wang, and B. Wu. 2011. “Genetic Algorithm Optimized SVM in Object-Based Classification of Quickbird Imagery.” In *Proceedings 2011 IEEE International Conference on Spatial Data Mining and Geographical Knowledge Services*, 348–52.

Mountrakis, Giorgos, Jungho Im, and Caesar Ogole. 2011. “Support Vector Machines in Remote Sensing: A Review.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 66 (3): 247–59.

Mutanga, Onisimo, Elhadi Adam, and Moses Azong Cho. 2012. “High Density Biomass Estimation for Wetland Vegetation Using WorldView-2 Imagery and Random Forest Regression Algorithm.” *International Journal of Applied Earth Observation and Geoinformation* 18 (August): 399–406.

NOAA. 2009. “What Is the Difference between Land Cover and Land Use?” September 10, 2009. https://oceanservice.noaa.gov/facts/lclu.html.

Nyssen, Jan, Jean Poesen, Jan Moeyersons, Jozef Deckers, Mitiku Haile, and Andreas Lang. 2004. “Human Impact on the Environment in the Ethiopian and Eritrean Highlands—a State of the Art.” *Earth-Science Reviews* 64 (3): 273–320.

Platt, Rutherford V., and Lauren Rapoza. 2008. “An Evaluation of an Object-Oriented Paradigm for Land Use/Land Cover Classification.” *The Professional Geographer: The Journal of the Association of American Geographers* 60 (1): 87–100.

Pontius, Robert G. 2000. “Quantification Error versus Location Error in Comparison of Categorical Maps.” *Photogrammetric Engineering and Remote Sensing* 66 (8): 1011–16.

Pouteau, Robin, Jean-Yves Meyer, Ravahere Taputuarai, and Benoît Stoll. 2012. “Support Vector Machines to Map Rare and Endangered Native Plants in Pacific Islands Forests.” *Ecological Informatics* 9 (May): 37–46.

Qian, Yuguo, Weiqi Zhou, Jingli Yan, Weifeng Li, and Lijian Han. 2014. “Comparing Machine Learning Classifiers for Object-Based Land Cover Classification Using Very High Resolution Imagery.” *Remote Sensing* 7 (1): 153–68.

Raczko, Edwin, and Bogdan Zagajewski. 2017. “Comparison of Support Vector Machine, Random Forest and Neural Network Classifiers for Tree Species Classification on Airborne Hyperspectral APEX Images.” *European Journal of Remote Sensing* 50 (1): 144–54.

Ramoelo, Abel, M. A. Cho, R. Mathieu, S. Madonsela, R. van de Kerchove, Z. Kaszta, and E. Wolff. 2015. “Monitoring Grass Nutrients and Biomass as Indicators of Rangeland Quality and Quantity Using Random Forest Modelling and WorldView-2 Data.” *International Journal of Applied Earth Observation and Geoinformation* 43 (December): 43–54.

“RBF SVM Parameters — Scikit-Learn 0.19.2 Documentation.” n.d. Accessed August 13, 2018. http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html.

Ricotta, C. 2004. “Evaluating the Classification Accuracy of Fuzzy Thematic Maps with a Simple Parametric Measure.” *International Journal of Remote Sensing* 25 (11): 2169–76.

Rodriguez-Galiano, V. F., M. Chica-Olmo, F. Abarca-Hernandez, P. M. Atkinson, and C. Jeganathan. 2012. “Random Forest Classification of Mediterranean Land Cover Using Multi-Seasonal Imagery and Multi-Seasonal Texture.” *Remote Sensing of Environment* 121 (June): 93–107.

Rosenfield, George H., and Katherine Fitzpatrick-Lins. 1986. “A Coefficient of Agreement as a Measure of Thematic Classification Accuracy.” *Photogrammetric Engineering and Remote Sensing* 52 (2): 223–27.

Stumpf, André, and Norman Kerle. 2011. “Object-Oriented Mapping of Landslides Using Random Forests.” *Remote Sensing of Environment* 115 (10): 2564–77.

“The World Factbook — Central Intelligence Agency.” n.d. Accessed July 12, 2018. https://www.cia.gov/library/publications/the-world-factbook/geos/et.html.

Tolessa, Terefe, Feyera Senbeta, and Moges Kidane. 2017. “The Impact of Land Use/land Cover Change on Ecosystem Services in the Central Highlands of Ethiopia.” *Ecosystem Services* 23 (February): 47–54.

Türk, Goksel. 1979. “Gt Index: A Measure of the Success of Prediction.” *Remote Sensing of Environment* 8 (1): 65–75.

Urgesa, Alemayehu A., Assefa Abegaz, Asmamaw L. Bahir, and Diogenes L. Antille. 2016. “Population Growth and Other Factors Affecting Land-Use and Land-Cover Changes in North-Eastern Wollega, Ethiopia.” *Tropical Agriculture* 93 (4): 298–311.

Vetrivel, Anand, Markus Gerke, Norman Kerle, and George Vosselman. 2015. “Identification of Damage in Buildings Based on Gaps in 3D Point Clouds from Very High Resolution Oblique Airborne Images.” *ISPRS Journal of Photogrammetry and Remote Sensing: Official Publication of the International Society for Photogrammetry and Remote Sensing* 105 (July): 61–78.

Waldner, François, Damien C. Jacques, and Fabian Löw. 2017. “The Impact of Training Class Proportions on Binary Cropland Classification.” *Remote Sensing Letters* 8 (12): 1122–31.

Waske, Björn, Sebastian van der Linden, Carsten Oldenburg, Benjamin Jakimow, Andreas Rabe, and Patrick Hostert. 2012. “imageRF – A User-Oriented Implementation for Remote Sensing Image Analysis with Random Forests.” *Environmental Modelling & Software* 35 (July): 192–93.

Watts, J. D., and R. L. Lawrence. 2008. “Merging Random Forest Classification with an Object-Oriented Approach for Analysis of Agricultural Lands.” *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences* 37 (Pat B7): 2006–9.

WoldeYohannes, Ashebir, Marc Cotter, Girma Kelboro, and Wubneshe Dessalegn. 2018. “Land Use and Land Cover Changes and Their Effects on the Landscape of Abaya-Chamo Basin, Southern Ethiopia.” *Land* 7 (1): 2.

Appendix I

Proposed LULC classification for large-scale land transaction sites located in Benishangul Gumuz, Gambella and Oromia.

Code	Land use/cover	Description
11	Small-holder Agriculture	Areas allotted to rain fed crop production, mostly of cereals in subsistence farming. Characterized by small cultivated areas (<10ha) with a mosaic of different crop types, fallow area, and cycle of crop maturity.
12	Small-holder Agriculture/ Settlement	Small-cultivated areas within a mosaic of rural settlements and sparse tree cover. Cultivated areas in this case tend to be "garden" plots adjacent to households. 13
13	Intensive Agriculture	Large areas under mono-cropped patterns. From imagery this can seen as large cultivated fields (>10ha) with the same pattern of vegetation indicating similar sow and harvest cycles.
21	Forest	Areas covered with dense growth of trees that formed nearly closed canopies (70–100%).
23	Woodland/ Savannah	Areas with sparse trees mixed with short bushes, grasses and open areas; less dense than the forest with approximately 10-70% cover.
31	Shrubland/ Grassland	Areas covered with grasses shrubs, bushes and very sparse, small trees (<10%)
32	Bare/ Exposed Soil	Grassy areas as well as bare land that has no other vegetation cover.

41	Rural Settlement	Composed of impermeable surfaces including roads, factories and dense housing.
42	Development	This includes major roads or infrastructure that does not fit into a category of urban settlements or rural settlements. The minimum mapping unit is 30m.
51	Water	Natural and artificial water bodies such as river, lakes or reservoirs with a minimum mapping unit of 30m.
52	Wetland	Areas that are waterlogged and swampy in the wet season, and dry in the dry season. These are very important for grazing during the dry season

Appendix II

The information of high spatial resolution images used in experiments

SENSOR	ACQ_TIME	BANDS	ROWS	COLUMNS	CLOUDCOVER	SUN_ELEV	OFF_NADIR	SPEC_TYPE	COUNTRY
WV02	2011-11-07T08:37:43.982850	4	8192	9216	0	61.1	2	Multispectral	ET
WV02	2011-11-07T08:38:11.997850	4	8192	9216	0	61.4	12.5	Multispectral	ET
WV02	2011-11-07T08:38:10.755250	4	8192	9216	0	61.3	12.7	Multispectral	ET
WV02	2011-11-07T08:37:45.304050	4	8192	9216	0	61.2	2.2	Multispectral	ET
WV02	2011-11-07T08:37:46.625050	4	8192	9216	0	61.3	2.5	Multispectral	ET
QB02	2006-11-08T08:36:57.294203	4	7312	6876	0.164	60.8	9.6	Multispectral	ET
QB02	2006-04-24T08:30:12.269855	4	7162	6876	0.099	72.9	13.5	Multispectral	ET
QB02	2006-11-08T08:36:53.905797	4	7312	6876	0.058	60.7	8.3	Multispectral	ET
QB02	2006-11-13T08:42:12.032754	4	7201	6876	0.068	59.9	19.4	Multispectral	ET
QB02	2006-11-13T08:42:08.701159	4	7201	6876	0.024	59.8	18.7	Multispectral	ET
QB02	2006-04-24T08:30:08.958551	4	7161	6876	0.196	72.9	13.1	Multispectral	ET
QB02	2011-11-30T07:42:50.668696	4	8192	7168	0	49	5.5	Multispectral	ET
QB02	2006-12-09T08:32:42.900580	4	7240	6876	0	54.2	14.3	Multispectral	ET
QB02	2006-12-09T08:32:43.295362	4	7240	6876	0	54.2	14.3	Multispectral	ET
QB02	2011-11-30T07:42:54.007826	4	8192	7168	0	49.1	4.3	Multispectral	ET
QB02	2006-12-09T08:32:46.647826	4	7241	6876	0	54.3	14.3	Multispectral	ET
QB02	2009-11-03T08:16:56.381159	4	7168	7168	0	61.6	13.2	Multispectral	ET
QB02	2006-12-09T08:33:26.840290	4	4127	6876	0.001	55.8	19.6	Multispectral	ET
QB02	2006-11-21T08:32:24.618841	4	5697	6876	0.038	59	12.6	Multispectral	ET
OV	2006-11-21T08:32:24.618841	4	5697	6876	0.038	59	12.6	Multispectral	ET
WV02	2016-04-13T08:36:57.294203	8	7312	6876	0.164	60.8	9.6	Multispectral	ET
WV02	2016-01-08T08:30:12.269855	4	7162	6876	0.099	72.9	13.5	Multispectral	ET

WV02	2016-01-27T08:36:53.905797	4	7312	6876	0.058	60.7	8.3	Multispectral	ET
WV02	2016-04-13T08:42:12.032754	8	7201	6876	0.068	59.9	19.4	Multispectral	ET
WV02	2016-01-08T08:42:08.701159	4	7201	6876	0.024	59.8	18.7	Multispectral	ET
QB02	2008-03-01T08:20:21.482609	4	7168	7168	0.002	65.5	12.7	Multispectral	ET
QB02	2006-10-24T08:21:40.137101	4	6787	6876	0	67.2	19.3	Multispectral	ET
QB02	2006-10-24T08:21:37.019420	4	6787	6876	0	67.1	18.3	Multispectral	ET