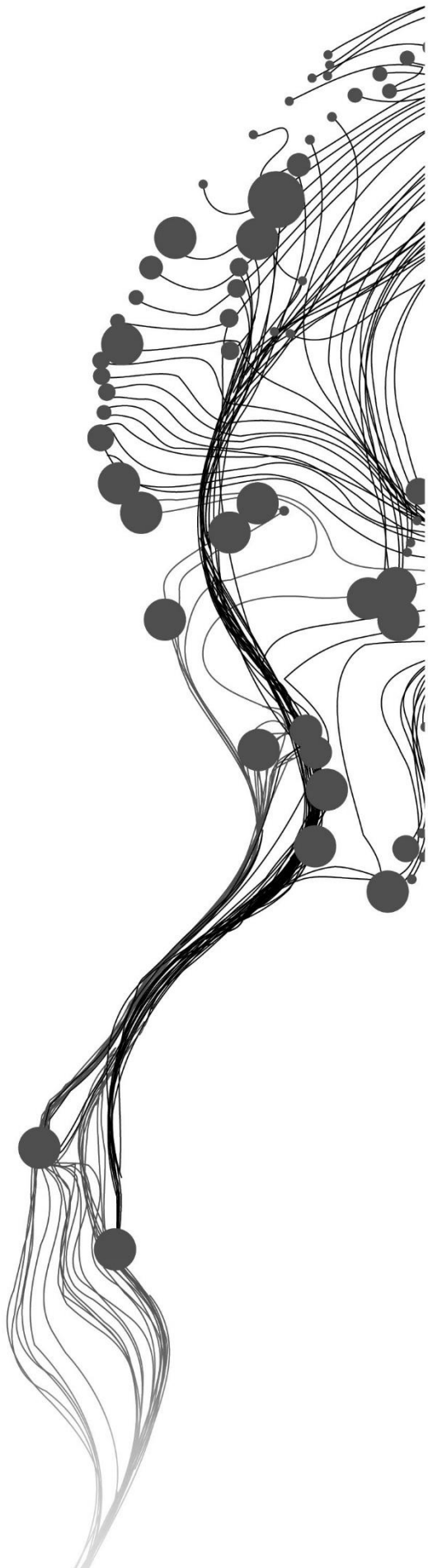
A SUPERVISED, CLOUD-BASED MULTI-CLASSIFIER TO MAP AFRICAN CROPLANDS

ROSA MARIA AGUILAR
February, 2017

SUPERVISORS:
Dr. R.Zurita Milla
Dr. Ir. R.A. de By

ADVISOR
Ms. Dr. E. Izquierdo-Verdiguier



A SUPERVISED, CLOUD-BASED MULTI-CLASSIFIER TO MAP AFRICAN CROPLANDS

ROSA MARIA AGUILAR
Enschede, The Netherlands, February, 2017

Thesis submitted to the Faculty of Geo-Information Science and Earth Observation of the University of Twente in partial fulfilment of the requirements for the degree of Master of Science in Geo-information Science and Earth Observation.
Specialization: Geoinformatics

SUPERVISORS:
Dr. R.Zurita Milla
Dr. Ir. R.A. de By

ADVISOR
Dr. E. Izquierdo-Verdiguier

THESIS ASSESSMENT BOARD:
Prof. Dr. M.J. Kraak (Chair)
Dr. V. Laparra, University of Valencia

ABSTRACT

Reliable crop maps are crucial for informed decision making and for achieving sustainable farming. These maps can also contribute to improving the livelihoods of the farmers and the welfare of their families. A common method to develop crop maps is to classify remotely sensed images. However, conventional classification approaches do not provide the accuracy required for large-scale mapping of smallholder farms in Africa because complex crop-planting patterns, mixed crops, small fields, and undefined field boundaries provide additional challenges to that of mapping crops from space.

The objective of this study is to realize a multi-classifier system that combines several machine learning classification algorithms to produce a more accurate map than the one that can be produced by using each classifier alone. The multi-classifier was implemented in a cloud platform, and this made it possible to handle large sets of spatial and spectral features in a reasonable time. A time series of very high spatial resolution images (Worldview-2) in Mali was used to illustrate the work. A proper configuration of the training strategy, feature space, and complexity of the multi-classifier was determined through experimentation.

After designing the supervised, cloud-based multi-classifier system, it was tested with three types of images: multi-spectral, pan-sharpened and aggregated. Pan-sharpened and aggregated images were produced by fusing the multi-spectral image bands with the corresponding panchromatic band. From these images, ratio bands, vegetation indices and textural features were extracted and included in the multi-classification.

The decision of base classifiers was combined using majority voting and weighted majority voting. The multi-classifier achieved best accuracy when weighted majority voting was applied. Four accuracy metrics of base classifiers were used as weights. These different weights yielded comparable results.

The multi-classification applied to pan-sharpened features set achieved the best accuracy. The kappa value (0.534) was slightly higher than the accuracy obtained by the multi-spectral set (0.531) but retained the spatial details of a panchromatic image. The most informative feature set was the combination of bands, vegetation indices and ratio bands. The result was obtained combining the output of the base classifiers through a weighted majority voting with kappa as weight. No evidence was found that suggest a strong correlation between these accuracy results and diversity in the multi-classifier.

This research demonstrated the potential of multi-classifiers to address complex problems where single classification approaches are not successful. Indeed, the multi-classifier improved up to 21.91% over the accuracy reached by any base classifier. Therefore, more accurate crops in smallholder farms can be produced applying the multi-classifier presented.

The conceptual design and implementation of the multi-classifier is a promising solution. Further research can extend it to include elements not considered due to time constraints. For example, feature selection and texture extraction of vegetation indices and diversity discrimination.

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to my first supervisor, Dr. Raul Zurita-Milla for the opportunity to develop this research project. His precise guidelines enabled me to successfully complete this research. I also learned from him that perseverance is important but taking some time to reflect is equally important.

I also want to offer my gratitude to my second supervisor, Dr. Rolf de By. His strict but positive criticism reinforced the full picture of this research and its main objective.

My deepest gratitude goes to my advisor, Dr. Emma Izquierdo-Verdiguier who not only provided me timely technical assistance but also emotional support in critical stages of my work.

I want to express my gratitude to the STARS project staff that provided me the images and ground data used in this study.

I worship God for the strength, health, and grace he has given me. God extended his mercy in my life every day during the past eighteen months.

My deepest gratitude and love go to my children that suffered my absences without complains. Thanks to my mother and family who continually motivated me and prayed for me. This achievement is dedicated to all of you.

I offer my gratitude to Mary, Lateef, Jody, Pastor Nies and Marsha from my local church "Evangelie Gemeente De Deur Enschede". They encouraged and supported me as a second family. My closest friends: Heloisa and Mila deserve my deepest gratitude as well. They also provided an invaluable motivation during critical phases of this thesis.

My appreciation is well deserved by Norma and Lydia who offered me assistance during my first and most difficult months in Enschede. My fellows of the GFM domain also deserve my gratitude, especially Kaaviya, Nicholus, Armand, Chrisper, Danilla and Raga.

Last but not least, I want to express my gratitude to my best friend, Claudio, who gave me joy and calm when I was sad and sick.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

This section presents the motivation, research objectives and questions that guided this study. The innovation of the study is presented as well, and finally the thesis structure is described.

1.1. Motivation and problem statement

Smallholder farmer in sub-Saharan Africa and South East Asia grow more than 80% of the area dedicated to agriculture activities (Stratoulias et al., 2015). However, smallholder farmers in sub-Saharan Africa, unlike farmers from high income countries, are not getting the most out of technologies such as remote sensing to improve their agricultural practices (STARS-project, 2016). To assist in this matter, the STARS project has collected a multi-spectral satellite image time series of very high spatial resolution over agricultural land in these regions. This dataset can be used to explore the possibilities of remote sensing to contribute to a more sustainable smallholder farming industry. This sustainability should, in turn, benefit their livelihoods and their family welfare.

Remote sensing is a well-established tool for mapping and monitoring crops leading to informed decisions in agricultural management and policy-making. Farmers can benefit from the remote sensing based advice to attain higher crop yields or quality products. Remotely sensed images have been applied to detect soil productive aptitude, stress in plants by disease or insects or adjusting fertilizer according to the specific soil conditions (Khorram, Van der Wiele, Koch, Nelson, & Potts, 2016). Meanwhile, policy makers can use crop growth monitoring and yield forecast systems to prevent or reduce the risk of disproportion between food supply and demand avoiding its negative consequences.

Cloud-free imagery, complicated crop-planting patterns and confusion between natural vegetation and crop land are the main sources of error in crop identification using remote sensed images (Chen et al., 2008). Smallholder farmers have practices such as mixed cropping and often fuzzy field boundaries that provide additional challenges. For instance, a free available dataset such as Landsat lacks spatial resolution to identify agricultural fields (Yan & Roy, 2014); it needs to be less than 2 meters because of the size (below 2 ha) of most African farms. In addition, smallholder agricultural fields are irregularly shaped and the seasonal variation in surface reflectance caused by the crop growth is less obvious because of rainfed farming and mixed cropping.

Supervised image classification is a common approach to crop identification using remotely sensed images (Foody & Mathur, 2004). However, conventional methods for supervised classification like Euclidian distance from the class mean, k-nearest neighbour algorithm, minimum distance classifier or Gaussian maximum likelihood, do not provide the accuracy required for large scale crop maps. In fact, most of these classifiers can only reach around 65% accuracy due to several reasons, among other things, differences in soil brightness, noise or higher variability within classes (Khobragade & Raghuwanshi, 2015).

Several approaches have been followed to improve the accuracy rate in image classification. For example, using multi-temporal images that cover a complete growth season instead of one single image (Debats, Luo, Estes, Fuchs, & Caylor, 2016; HU et al., 2017), combining images from different sensors (Castillejo-González et al., 2009; Zurita Milla et al., 2008) or improving the performance of classification algorithms (Du et al., 2012)

Supervised learning models like Support Vector Machine (SVM; Cortes & Vapnik, 1995) or Random Forest (Breiman, 2001) have proved to be efficient and robust in crop discrimination. For example, Camps-Valls et al.(2004) reached a recognition rate higher than 90% applying SVM and Artificial Neural Network (ANN) on 5 meters spatial resolution hyperspectral data in Spain. Similarly, a Random Forest implementation reached recognition rates above 80% in some areas of China and France but poor rates in Burkina Faso, mainly due to the high heterogeneity in the landscape produced by trees within the fields and the crop mix. It seems that mapping crop fields in smallholder farms requires specialized methods (Debats, Luo, Estes, Fuchs, & Caylor, 2016).

Recently, a combination of multiple machine learning classifiers was applied to image classification for agricultural purposes outperforming single classifiers (Corrales, Figueroa, Ledezma, & Corrales, 2015; Lijun & Chuang, 2011; Dai & Liu, 2011; Li, Yang, & Wang, 2016). A *multi-classifier* combines smartly the output of different base algorithms to improve its capability to model complex (non-linear) decision boundaries.

At the expense of increased complexity, a multi-classifier (also known as *ensemble classifier*) can manage complex feature space and large amount of data. It also reduces the probability of choosing a bad classifier (over fitted or under trained) and increases the possibility of reaching a global minimum. However, to gain these advantages, the individual (base) classifiers must provide higher accuracy than chance (Polikar, 2006) and be diverse. Classifiers are diverse when their decision boundaries are adequately different and consequently, they make errors in different instances. Indeed, diversity is a keystone of multi-classifier design since it is directly related to higher accuracy in majority voting ensembles (Brown & Kuncheva, 2010).

Cloud computing environments provide resources such as storage, computing power, applications and services that make them a suitable option to implement an ensemble of computationally expensive algorithms that process a large amount of data like remote sensing imagery and its derivative spectral and textural features. Google Earth Engine (GEE; Gorelick, 2012) is a novel cloud computing platform available to scientists and researchers who focus on the exploitation of Earth observation data through its extensive catalogue of satellite images as well as analysis tools for developing complex applications.

Since crop mapping in smallholder farms requires specialized methods, it is reasonable to explore GEE capabilities to implement a recent machine learning approach such as multi-classifier for this complex problem

1.2. Innovation aimed at

Most multi-classifier have been used for land cover classification but rarely for crop type recognition in a smallholder context. In addition, few studies have been reported for crop monitoring at large scale (Inglada et al., 2015). The multi-classifier proposed exploits the potential of ensemble learning methods applied to a dataset of very high resolution images from which contextual features, like texture, can be derived.

There is no one single best algorithm or method suitable for all the optimization problems addressed in computational environments (Wolpert & Macready, 1997). Therefore, it makes sense to contribute to the specific problem of crop type recognition in African smallholder farms using a multi-classifier system in a cloud computing environment.

Also, GEE is a relatively new cloud computing platform so to the best of my knowledge there are no references about studies exploiting its capabilities on this or similar research topic.

1.3. Research objectives and questions

The main objective of this research project is to realize a multi-sensor, multi-classifier and cloud-based system for mapping smallholder African croplands. This main objective is accomplished through the following specific objectives and their respective questions:

1. To explore cloud service capabilities to implement a multi-classifier system.
   1-1. Which suitable classifiers for ensemble learning are available in GEE?
   1-2. What are the most promising GEE classifiers to build a multi-classifier system?

2. To evaluate different training strategies to create diversity in the classifiers.
   2-1. What is the impact of sample selection (e.g. k-fold or bagging) on the multi-classifier diversity?
   2-2. Are diversity and accuracy of the multi-classifier correlated?

3. To compare different approaches to combine the output of the classifiers.
   3-1. What is the added value of combining multiple classifiers in comparison to using a single classifier?
   3-2. Which combination rule (majority voting or weighted majority voting) provides the highest accuracy?

4. To evaluate the contribution of multiple Earth observation sensors in the multi-classifier performance.
   4-1. What is the impact of image fusion (e.g. aggregation, pan-sharpening) on the accuracy of the multi-classifier?
   42. What is the added value of combining multi-sensor in comparison to using a single sensor?

1.4. Thesis structure

The thesis is structured in five chapters. Chapter one introduces the topic and the scope of the study. Chapter two summarizes the literature review. Chapter three presents the study area and data used. Chapter four explains the methods used for data pre-processing, design and implementation of the multi-classifier system. Chapter five reports the results obtained during the experimental phase and presents the discussion of them. Finally, in chapter six, conclusions drawn from the research are presented.

2. LITERATURE REVIEW

This section present use cases of indices and textural features exploited in agricultural applications. Image classification methods are also reported and a review of common approaches to generate a multi-classifier is presented.

2.1. Indices and textural features

The use of remotely sensed multi-spectral images for crop classification has been widely reported (Chen et al., 2008). Red, Green and Near InfraRed bands are commonly used in crop mapping, providing indexes related to vegetation health or changes (Jackson & Huete, 1991). Some specific examples of very high spatial resolution images for crop monitoring are given in Tarantino et al. (2012), Bouroubi, Tremblay, Vigneault, & Benoit (2014) and Chellasamy, Ferre, & Greve, (2015).

Multi-spectral bands in the green, red and near infra-red range of the electromagnetic spectrum are convenient for crop identification. However, indices offer a better approach since they are less affected than single bands in isolation by atmospheric conditions and soil background because indices enhance the vegetation signal (Jackson & Huete, 1991). Several indexes have been formulated to assess vegetation health and identification. They have been related to surface biophysical parameters (de la Fuente et al., 2013) such as chlorophyll, moisture content or leaf area index (LAI), that are broadly applied in land cover mapping and crop-growth monitoring.

The development of vegetation indices is based on the difference in the spectral response of the vegetation when illuminated in two (sometimes more) portions of the electromagnetic spectrum. Healthy vegetation transmits or reflects most of the light in the Near Infra-Red (NIR) wavebands with little absorption, whereas in the visible bands (Red, Green, Blue) absorption is strong. The wavebands used to compute a vegetation index are selected such that one decreases and the other increases with growing vegetation cover.

Indices and image bands provide spectral information whereas textural features provide information about the spatial distribution of that spectral information in an image band. Textural features contains useful information about the surface itself and its spatial context for identification purposes. (R. Haralick, Shanmugan, & Dinstein, 1973).

One of the most common methods to compute textural features uses the Gray Level Co-occurrence Matrix (GLCM) (Conners, Trivedi, & Harlow, 1984; R. M. Haralick, 1979). It is a two-dimensional histogram with the probabilities of occurrence of two different gray levels in the spatial neighbourhood (predefined directions and distances) of the pixel under observation expressed in a matrix.

GLCM employs normalized gray levels that reduce the matrix dimension. For example, for an image band of 256 gray levels, a normalized matrix is produced scaling these levels to only eight possible values. This scaling reduces the problem of dimensionality and consequently the computation time.

Several studies have addressed the usefulness of the spatial context for crop discrimination. (Peña-Barragán, Ngugi, Plant, & Six, 2011). Besides, textural features are likely to improve classification accuracy heterogenous area where local variance is high (Li, Yang, & Wang, 2016; Shaban & Dikshit, 2001).

Dynamic in land use or land cover can be modelled by means of multi-temporal remotely sensed images because the changes observed in spectral response have been correlated to biophysical parameters (A. Huete et al., 2002). In this sense, multi-temporal vegetation indices have been exploited in crop identification, crop monitoring or vegetation change because they inform of the seasonal variation in surface reflectance caused by a specific crop phenology (Chellasamy et al., 2015; Misra, Kumar, Patel, & Zurita-Milla, 2014; Tatsumi, Yamashiki, Canales Torres, & Taipe, 2015). Multi-temporal textural features have also shown to be informative to discriminate crops (Peña-Barragán et al., 2011).

2.2. Image fusion

The increasing amount of available remote sensing imagery coming in different spatial, temporal and spectral resolution requires effective methods to combine these images to fully exploit its capabilities. Image fusion is a method to merge several images (two or more) into a single image that maximizes the relevant information from the input images (Haghighat, Aghagolzadeh, & Seyedarabi, 2011). It can be applied for visual enhancement, change or object detection or classification.

Gómez-Chova, Tuia, Moser, & Camps-Valls (2015) explore four approaches of image fusion for classification: sub-pixel level, pixel level, feature level and decision level.

A *subpixel level image fusion* uses higher resolution data to "unmix" lower resolution pixels that usually have higher spectral or temporal resolution using appropriate transformations. It requires a soft classification approach which convert mixed pixel values into numerical subpixel proportions of a few classes instead of an abstract label (Delalieux et al., 2014).

Pixel level image fusion generates a new image where each pixel value is computed from a set of pixel values from the input images.

Feature level image fusion merge objects identified in various input images (Hao et al., 2015). It requires to perform first a feature extraction and poses additional challenges when using an image series due to the inherent shape and size changes of the objects.

Decision level image fusion combine information at a higher level of abstraction. It merges the result of previously applied processing algorithms, possibly different, to the input images. These results are combined through decision rules to yield a final fused decision (Pohl, 1998).

Pan-sharpening is a well-known image fusion method (Zhang, 2010). It uses the panchromatic (PAN) band to increase the spatial resolution (more spatial detail) of multi-spectral (MS) bands. This is suitable when the difference between the PAN and MS spatial resolution is not large. After applying a pan-sharpening fusion, a single resolution classifier can be applied.

Pan-sharpened images has shown to improve crop identification. For example, Gilbertson, Kemp, & van Niekerk (2017), in a study conducted in South Africa, demonstrated that pan-sharpened image series led to a more accurate crop classification result than standard spectral bands of Landsat 8 images.

Hyperspherical Color Space (HCS) pan-sharpening (Padwick, Scientist, Deskevich, Pacifici, & Smallwood, 2010), is an algorithm especially designed for sharpening WorldView-2 images that shown to be more efficient than other similar methods.

A simple strategy to combine images with different spatial resolutions at pixel level is aggregating the finer resolution image to the coarser resolution and then apply the selected method for classification. The usefulness of this approach is to provide redundant information or quantify the contribution of the panchromatic band to the accuracy of image classification (Wang, Sousa, Gong, & Biging, 2004).

Image fusion and feature extraction such as indices and textural features address the need of using multiple features to improve image classification. However, the election of an appropriate classifier is also essential in the classification accuracy (Lu & Weng, 2007).

2.3. Multi-classifier systems

A classifier is an algorithm that assigns objects to classes or categories. Classifiers can be *unsupervised*, *supervised* or *semi-supervised*. *Unsupervised* classifiers categorize data without any provided label. *Supervised* classifiers, by contrast, identify/predict to which category a new instance belongs based on a training data set, of which the element memberships are known. *Semi-supervised* classifiers use labelled and unlabelled data for training (Gupta, Sharma, & Jindal, 2016).

A classifier can be *parametric* or *non-parametric*. A classifier is *parametric* if the algorithm estimate statistical parameters such as mean and variance of the data probability density function from the training set to predict to which class an unseen data belongs. A classifier is *non-parametric* if do no use statistic parameters and no assumptions are made over the data probability density function (Richards, 2013).

Several non-parametric algorithms have been applied to the problem of crop classification using remotely sensed image. Each classifier has its own merits and different results can be obtained depending of the classifier chosen (Lu & Weng, 2007).

The *Support Vector Machine* (SVM) was introduced by Cortes & Vapnik, (1995). This type of algorithm aims to find the optimal linear hyperplane that separates the given data in two classes by maximizing the margins between the so-called support vectors (the closest training samples to the optimal hyperplane). When the data present complex patterns (non-linearities) that cannot be separated by a plane, a feature mapping is applied to map the original data to a multidimensional feature space. In a next step, the calculation of the optimal hyperplane takes place, producing non-linear algorithm versions using only the dot products among the mapped samples by applying the kernel trick (Izquierdo-Verdiguier et al., 2015).

Linear, radial and polynomial are common SVM kernel functions. A cost variable c is introduced to assign a penalty for misclassifying instances, with aim to minimize the error. However, there is no agreement on optimal values for this variable, and this is also true for optimality in kernel parameters values (Huang, Davis, & Townshend, 2002). Polynomial kernel has three parameters: gamma, coefficient and degree that are data dependent.

SVM has been applied widely for crop classification outperforming other methods (Beyer, Jarmer, Siegmann, & Fischer, 2015; Mountrakis, Im, & Ogole, 2011; Nitze, Schulthess, & Asche, 2012). SVM has demonstrated robustness to outliers and presents an excellent option when the number of input features is high (Camps-Valls et al., 2004).

Classification and Regression Trees (CART; Breiman, Friedman, Stone, & Richard A., 1985) splits the data in two groups iteratively. Each node of the tree evaluates only one feature and a simple threshold is used in making the decision. As supervised classification, CART uses ground truth data to build the tree that is used afterwards to classify unseen data. CART has been used for crop mapping with multi-source and

multi-temporal data scoring more than 87% accuracy measure values (Shukla, Garg, Srivastava, & Garg, 2016) and above 75% (Patil et al., 2016) for medium resolution images.

Random Forest (RF; Breiman, 2001) is an ensemble of several decision trees. Random Forest split the feature space to create tress by a random selection of instances in the training set or by random feature selection. At each node, only the selected features are searched for the best subsequent split. The ensemble decision is made usually by majority voting.

Random Forest reached around 85% accuracy in crop type classification using a multi-spectral time series of RapidEye images (Nitze et al., 2012) and higher than 80% for a time series of Landsat7 images in homogeneous regions (Tatsumi et al., 2015).

Maximum Entropy (MaxEnt) models was applied widely in classification problems in natural language processing (Mann, Mcdonald, & Silberman, 2009). It computes an approximated probability distribution consistent with the constraints (facts) observed in the data (values of properties) but as uniform as possible (Berger, Pietra, & Pietra, 1996) providing maximum entropy while avoid making assumptions over the unknown.

In a geographic context, it was proposed to predict geographic species distribution and potential habitat (Phillips, Anderson, & Schapire, 2006), however it has been applied to vegetation classification from remote sensing images achieving 90% of accuracy (Evangelista, Stohlgren, Morisette, & Kumar, 2009).

Although individual classifiers showed good performance in specific scenarios. There is no one classifier that perform well in all cases (Woźniak & Graña, 2014). The combination of several classifier or *multi-classifier* was introduced as a new approach for the improvement of classification accuracy reached by individual classifiers.

A *multi-classifier* is a meta-algorithm that fuses under certain rules the decisions of "*base classifiers*", being basic classifications that are combined to produce an ensemble output.

A conventional supervised classification method consists of training a single classifier using (ground) truth data to classify subsequently unseen data. In contrast, in a multi-classifier, several classifiers are trained and applied to the unclassified data to produce a final decision that can be seen as a combination of classifier outputs under certain rules.

An multi-classifier is the result of one of five approaches (Du et al., 2012; Kuncheva, 2004; Polikar, 2006):

1. Training sample manipulation: the training set is sampling to produce multiple versions of a classifier that are combined later. Bagging (Breiman, 1996) and Boosting (Freund & Schapire, 1997) are the most popular methods. Training sample manipulation also includes feature selection as takes place in Random Forest (Breiman, 2001).
2. Parameters diversification: different classifiers models are produced by providing different initialization parameters. For example, by providing different values for the regularization parameter *cost* of SVM, different classifiers will be generated because the decision boundaries are affected by the *cost* value.
3. Series concatenation: several classifier algorithms applied in a chain such that the output of one algorithm is used as input of the following. The overall result is obtained from the final classifier.

4. Parallel concatenation: several classifiers are applied independently and their outputs are combined following a decision rule.
5. Hybrid concatenation: a mixture of parallel and series combination of different classifiers. For example, the outputs of parallel classifiers can be used as input for another classifier to produce a result.

A multi-classifier can be *static* or *dynamic* depending on its behaviour during run-time. In a *static* multi-classifier, also known as a fusion ensemble, each base classifier has knowledge about the complete feature space and decision rules are stablished over all the samples. In a *dynamic* multi-classifier, also known as selection ensemble, by contrast each base classifier has knowledge of only a portion of the feature space and is responsible for the instances in there. This means that the decision rule is adjusted according to the input data (Mousavi & Eftekhari, 2015). Dynamic ensembles have not been shown to be more efficient than static ensembles (Ko, Sabourin, & Britto, 2008).

A multi-classifier produces better results when it combines different classifiers with uncorrelated errors (non-coincident errors). Diversity is a measure of this dissimilarity. It has been related to higher accuracy in ensemble classifier (Brown & Kuncheva, 2010).

One of the most common techniques to obtain diversity is the input data manipulation by means of training strategies. The main idea is that not all classifiers see all the samples to attain generalization capabilities. K-fold and bagging (Leo Breiman, 1996) are recognized training methods in classification algorithms.

K-fold, also known as jack-knife, consists on partitioning the training set in k equal size blocks and train k classifiers with k-1 blocks (i.e. leaving out one block each time). This mechanism reduce overfitting. Bagging creates multiple versions of a classifier by making bootstrap replicates of the training set and using these subsets to train the classifier. Each subset contains between 75% and 100% of the original training set and overlaps substantially, with many instances appearing multiple times.

The decision of a multi-classifier is the fusion of the outputs of base classifiers under certain combination rule. A combination rule defines the parameters, or *weights* for each classifier output. It can be trainable or non-trainable. A trainable rule implies that a separate algorithm is trained to determine the weights, whereas a non-trainable rule does not require this after the base classifiers in the ensemble have been trained individually. A non-trainable combination rule is effective when the base classifiers are trained carefully avoiding overfitting (Duin, 2002) and fixed combination rules are used. Majority voting and weighted majority voting are common combination rules for multi-classifiers.

Multi-classifiers have received substantial attention in recent years. They have been applied in a wide range of areas such as face recognition (Lumini, Nanni, & Brahnam, 2016), discrimination of tumour cells (Gopinath & Shanthi, 2014), land use (Lijun & Chuang, 2011), disease detection (Corrales, Figueroa, Ledezma, & Corrales, 2015) and crop monitoring (Li et al., 2016), among others.

Lijun & Chuang (2011) proposed a method that combines maximum probability and weighted voting based on a threshold for land cover classification. First, an unsupervised classification is performed to estimate prior probabilities. When the prior probability is higher than 85%, a class label is assigned directly; otherwise, a majority voting of the six base classifiers (ML, SVM, ANN, SAM, DTC, ISODATA) outputs is performed to determine the corresponding class label. This method used a single CBERS image and

was implemented in a parallel computing platform. The multi-classifier reached an improvement of 2.22% in comparison to the base classifiers applied.

Different base classifiers (SVM, ANN, Bayesian Networks, Decision trees and K- nearest neighbors) were empirically combined by Corrales et al., (2015) to detect coffee rust in Colombian crops. In this study, an ANN performs a first classification determining which classifier will be applied in a second step. The ANN partition the feature space in two portions that will be processed by a regression trees classifier or a support vector machine. The empirical multi-classifier proposed outperform single base classifiers and conventional multi-classifier approach such as bagging.

Li, Yang, & Wang, (2016) applied adaptive boosting (Freund & Schapire, 1997) to a varying number, from one to twelve, SVM and ANN classifiers. The purpose of the study was to detect corn planting area in northern China. Spectral features of a Gao-Fen-1 satellite image are joined to GLCM and Gabor filter textural features. The multi-classifier yielded a higher accuracy than conventional methods. The utility of textural features for classification. The computational cost is higher and represents a shortcoming of this method.

3. STUDY AREA

Details of the study area and ground truth data used in this research project are given in this section.

3.1. Study Area

The study area is a square area of size 10×10 km located in Mali, a West African country where the STARS project acquired several very high spatial resolution images during the cropping seasons of 2014–2016. This study is based on images from 2014 only. Figure 3-1 shows the relative location of Mali in Africa, an example image of the complete study area on October 18, 2014 and a zoom of this image.

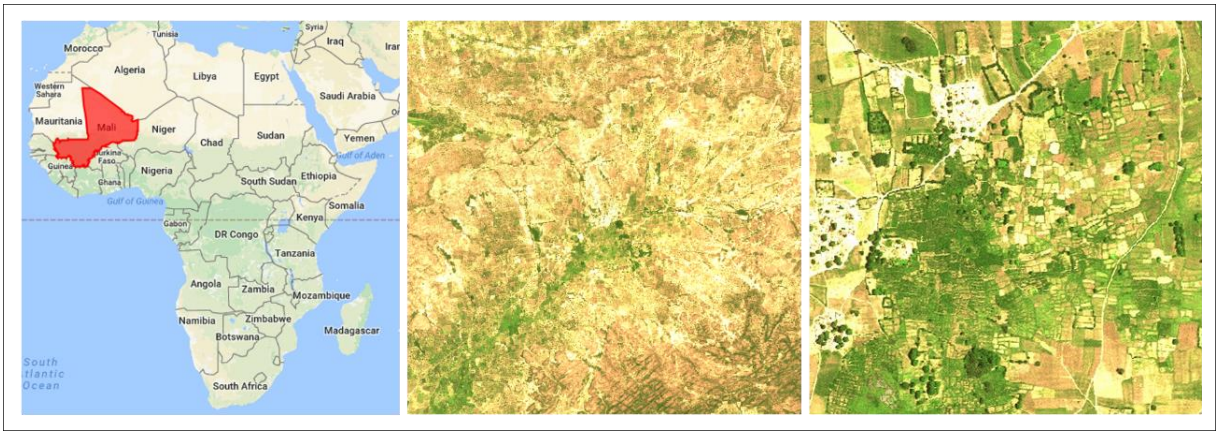


Figure 3-1. Study Area

a) Relative location of Mali in Africa

b) Image of the study area on October 18, 2014 using natural color composite.

c) a zoom of the image shown in b)

The acquisition dates of the images range from May to November. This covers both the beginning and the of the crop growing season, approximately 180 days. Dates of images are: 1) May 22, 2) May 30, 3) Jun 26, 4) Jul 29, 5) October 18, 6) November 1, and 7) November 14. A multi-spectral image of 2m pixel size and a panchromatic image of 0.5m. pixel size were incorporated for each date. The images were captured by the Worldview-2 Satellite sensor. Table 3-1 lists the spectral and spatial resolution of these images.

All the image series were co-registered in the context of the STARS project. For this study, a total of fourteen images were included. In addition, vegetation indices, ratio bands and textural features were extracted from the images and included in the classification process.

Five main crop are cultivated in the study area: Maize, Millet, Peanut, Sorghum and Cotton. The ground truth data required to train the base classifiers was collected by a team of ICRISAT. This data was used to both train base classifiers and to assess the accuracy of the multi-classifier. The ground data collected over the study area is presented in Figure 3-2. Polygons are displayed in red color and an image. Image of the study area on October 18, 2014 in natural color composite is used as background.

Table 3-1. Spectral and spatial resolution of Worldview-2 images

Band	Wavelength (μm)	Spatial Resolution (m)
PAN	0,45 to 0,8	0,5
Band Blue (VIS)	0,45 to 0,51	2
Band Green (VIS)	0,51 to 0,58	2
Band Red (VIS)	0,63 to 0,69	2
Band NIR 1 (NIR)	0,77 to 0,895	2
Band Coastal (VIS)	0,4 to 0,45	2
Band Yellow (VIS)	0,585 to 0,625	2
Band Red Edge (VIS)	0,705 to 0,745	2
Band NIR 2 (NIR)	0,86 to 1,04	2

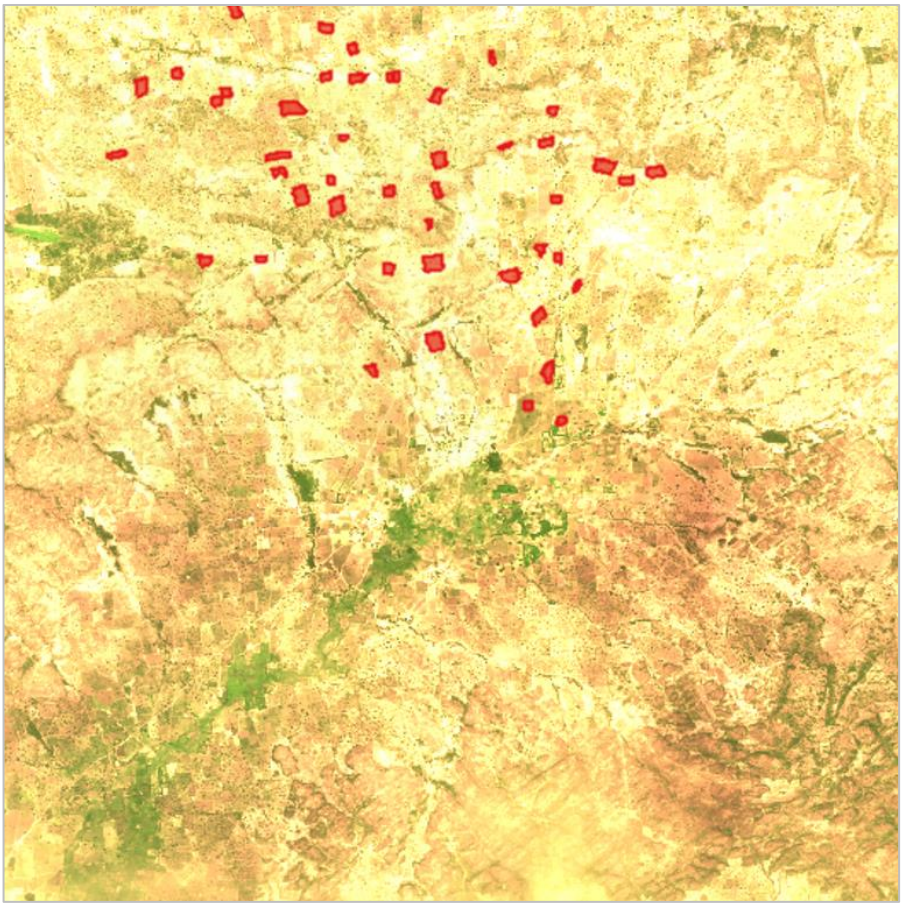


Figure 3-2. Ground data collected over the study area. As background an image of the study area on October 18, 2014 using natural color composite.

4. METHODS

The accuracy of crop mapping using supervised classification methods depends on training data and the classifier performance. Given the complexity and relevance of making crop maps, different approaches to increase their accuracy have been proposed in literature. Some authors addressed the problem of crop mapping from space by adding more informative features such indices or textural features. Others focussed on capturing crops temporal profile by using multi-temporal features and, finally, other authors have used or designed new classification methods, or have studied the combination of classification methods.

In this chapter, the methods used to design, implement and apply the multi-classifier system are described. Three multi-temporal datasets containing bands, indices and textural information were set up. These datasets were obtained from: a) the original multi-spectral images, b) pan-sharpened images, c) aggregated images, and d) vegetation indices and textural features extracted from these images.

4.1. Textural features and indices

Small farms that maintain mixed cropping configure a heterogeneous crop land where textural features can improve the classification accuracy. For this reason, this study exploits 18 GLCM textural features. These features were calculated as the average of their values in four directions angles (0, 45, 90, 135) and using a window of 3 by 3 pixels that means an inter-pixel distance of one pixel. Table 4-1 list those textural features with their corresponding formulas, where:

$p(i,j)$ is the *(i,j)*th entry in a normalized gray tone matrix,

$p_x(i) = \sum_{j=1}^{N_g} P(i,j)$ is the i^{th} entry in the marginal-probability matrix computed by summing the rows of $p(i,j)$,

$p_y(j) = \sum_{i=1}^{N_g} P(i,j)$, is the i^{th} entry in the marginal-probability matrix computed by summing the columns of $p(i,j)$,

N_g, is the number of distinct grey level in the quantized image.

$p_{x+y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)_{i+j=k}$, and $p_{x-y}(k) = \sum_{i=1}^{N_g} \sum_{j=1}^{N_g} p(i,j)_{|i-j|=k}$

Table 4-1. Grey level co-occurrence matrix textural features.

Name/Formula	Name/Formula
Angular Second Moment $f_1 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g}\{p(i,j)\}^2$	*Contrast* $f_2 = \sum_{n=0}^{N_{g-1}} n^2 \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \right\}_{\lvert i-j\rvert=n}$
Correlation $f_3 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{(i,j)p(i,j) - \mu_x\mu_y}{\sigma_x\sigma_y}$	*Variance* $f_4 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} (i-\mu)^2\, p(i,j)$
Inverse Difference Moment $f_5 = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \frac{1}{1+(i-j)^2} p(i,j)$	*Sum Average* $f_6 = \sum_{i=2}^{2N_g} i p_{x+y}(i)$
Sum Variance $f_7 = \sum_{i=2}^{2N_g} (i-f_8)^2 P_{x+y}(i)$	*Sum Entropy* $f_8 = -\sum_{i=2}^{2N_g} p_{x+y}(i) log\{p_{x+y}(i)\}$
Entropy $f_9 = -\sum_{i=1}^{N_g}\sum_{i=2}^{N_g} p(i,j)\log(p(i,j))$	Difference Variance f_{10} = variance of p_{x-y}
Difference Entropy $f_{11} = \sum_{i=0}^{N_{g-1}} p_{x-y}(i) log\{p_{x-y}(i)\}$	*Information Measures of Correlation 1* $f_{12} = \frac{HXY - HXY1}{\max\{HX, HY\}}$ Where, $HXY = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log(p(i,j))$ $HXY1 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j)\log\{p_x(i)p_y(j)\}$ HX and HY are entropies of p_x and p_y
Information Measures of Correlation 2 $f_{13} = (1 - e^{[-2.0(HXY2-HXY)]})^{1/2}$ Where, $HXY2 = -\sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p_x(i)p_y(j)\log\{p_x(i)p_y(j)\}$	*Maximal Correlation Coefficient* f_{14} = (second largest eigen value of Q)$^{1/2}$ Where, $Q_{(i,j)} = \sum_{k=0}^{N_{g-1}} \frac{p(i,k)p(j,k)}{p_x(i)p_y(k)}$
Dissimilarity $f_{15} = \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} \lvert i-j\rvert^2 p(i,j)$	

This study included three features textures proposed by Conners, Trivedi, & Harlow (1984). Table 4-2 shows the name and formula of these features, where:

$s(i,j,\delta,T)$ is the $(i,j)^{th}$ entry in a normalized grey tone matrix, equivalent to $p(i,j)$

T represent the region and shape used to estimate the second order probabilities, and

$\delta = (\Delta x, \Delta y)$ is the displacement vector.

Table 4-2. Textural features textures proposed by Conners included in the multi-classification.

Description	Formula
Inertia	$I(\delta,T) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}(i-j)^2 s(i,j,\delta,T)$
Cluster shade	$A(\delta,T) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left(i+j-\mu_i-\mu_j\right)^3 s(i,j,\delta,T)$
Cluster prominence	$B(\delta,T) = \sum_{i=0}^{L-1}\sum_{j=0}^{L-1}\left(i+j-\mu_i-\mu_j\right)^4 s(i,j,\delta,T)$

Textural features were calculated for each band of each image. As an example, Figure 4-2 shows a subset of one of the multi-spectral images in the study area acquired on October 18, 2014. Figure 4-3 presents *contrast* and *variance* applied to the band five (red channel) of the subset displayed in Figure 4-2. These textures images represent local variance and roughness respectively.

Figure 4-1. Subset of image of the study area on October 18, 2014 using natural color composite.

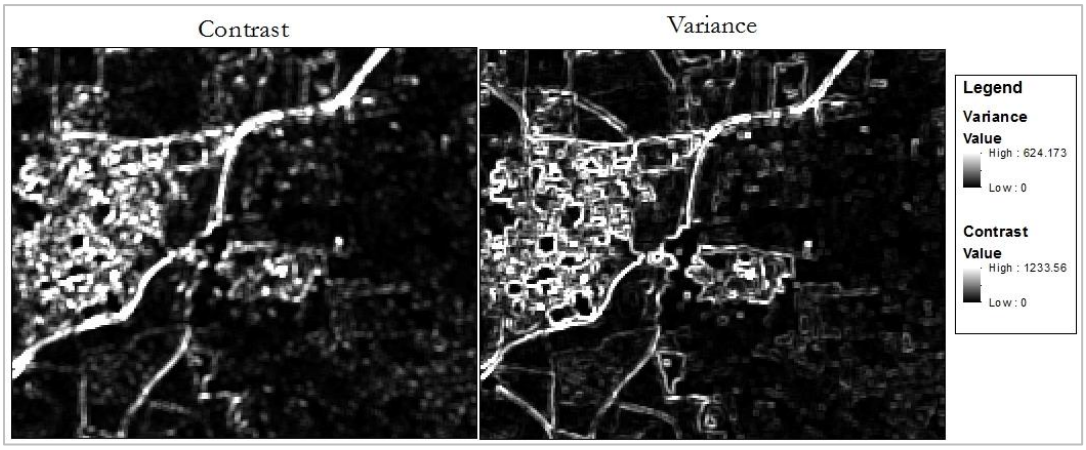


Figure 4-2. *Contrast* and *variance* texture applied to band five (red channel) of image subset drawn in Figure 4-1

Specific vegetation indexes with potential for agricultural studies were selected.
Table 4-3 lists all the selected indices, their formulas and references.

Table 4-3. Vegetation indexes, formulas and author(s). Bands abbreviations are: R = Red, G = Green, B = Blue and NIR = Near InfraRed.

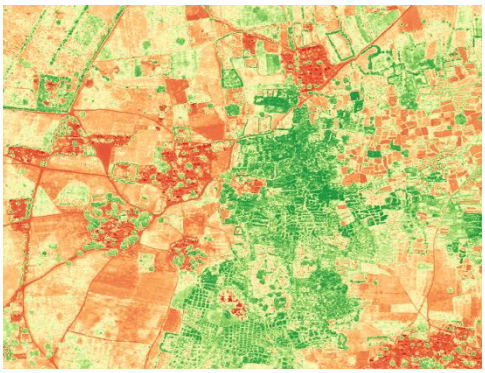
Name	Formula	Author
DVI - Difference Vegetation Index	$DVI = NIR - R$	(Tucker, 1979)
RVI - Ratio Vegetation Index	$RVI = NIR/R$	(Jordan, 1969)
NDVI – Normalized Difference Vegetation Index	$NDVI = (NIR - R)/(NIR + R)$	(Rouse, Haas, & Deering, 1973)
EVI – Enhanced Vegetation Index	$EVI = 2.5 * (NIR - R)/(NIR + 6 * R - 7.5 * B + 1)$	(A. Huete et al., 2002)
SAVI - Soil Adjusted Vegetation Index	$SAVI = (1 + L) * (NIR - R)/(NIR + R + L)$, con L=0.5	(A. R. Huete, 1988)
MSAVI – Modified Soil Adjusted Vegetation Index	$MSAVI = 0.5 * (2 * NIR + 1 - \sqrt{((2 * NIR + 1)^2 - 8 * (NIR - R))})$	(Qi, Chehbouni, Huete, Kerr, & Sorooshian, 1994)
TCARI – Transformed Chlorophyll Absorption in Reflectance Index	$TCARI = 3 * ((RE - R) - 0.2 * (RE - G) * (RE/R))$	(Haboudane, Miller, Tremblay, Zarco-Tejada, & Dextraze, 2002)
GLI – Green Leaf Index	$GLI = (2 * G - R - B)/(2 * G + R + B)$	(Louhaichi, Borman, & Johnson, 2001)
VARI – Visible Atmospherically Resistance Index	$VARI = (G - R)/(G + R - B)$	(Gitelson, Kaufman, Stark, & Rundquist, 2002)

Figure 4-1 displays a subset of one of the Worldview-2 multi-spectral images and the EVIs for three of the acquisition dates. Greenest tones indicate higher values of vegetation whereas reds are lower presence of vegetation It can be observed tones variation that shows differences in vegetation across the crop growth season

a) Subset of image dated October 18 of 2014 in natural color composite

b) EVI of the image dated May 22 of 2014.

c) EVI of the image dated July 29 of 2014.



d)EVI of the image dated November 01 of 2014.

Figure 4-3. Subset of one of the images of the study area and EVI for three dates.

4.2. Image fusion

A pan sharpening and an aggregation strategy were implemented to combine multispectral and panchromatic bands. Due to the small difference between the spatial resolutions of the multispectral bands (2,08 m) and the panchromatic band (0,52 m) both methods are suitable for fusion at pixel level (Gómez-Chova et al., 2015). Pan sharpening is executed by HCS pan-sharpening. Aggregation was performed by averaging pixel values to produce a panchromatic band with the same pixel size than multi-spectral bands. A new image that contains both panchromatic and multi-spectral bands was generated. The effect or relevance of each strategy is analysed through experimentation.

After applied the fusion methods, three datasets were available for experimentation. Table 4-4 presents the description and total number of features available for each dataset. Table 4-5 lists the abbreviations used here in after for those features.

Table 4-4. Description of feature type for multi-spectral, pan-sharpened and aggregated images

Feature type	Number of features			Description
	Original Image	HCS Image	Aggregated Image	
Bands	7	6	8	Bands of the image.
Image bands texture	126	108	144	18 textural features for each band using the mean value. Kernel size: 1.
DVI	21	15	28	All possible combinations for bands, except band 1 reserved for atmospheric correction.
RVI	21	15	28	All possible combinations for bands, except band 1 reserved for atmospheric correction.
NDVI	21	15	28	All possible combinations for bands, except band 1 reserved for atmospheric correction.
EVI – SAVI – MSAVI – TCARI – GLI - VARI	6	6	6	Vegetation indices.
Total	202	165	242	

Table 4-5. Description of abbreviations of multi-temporal feature sets

Feature Set	Number of features			Description
	Original Image	HCS Image	Aggregated Image	
B	49	42	56	Image bands
BVI	91	84	98	Image bands and vegetation indices: EVI, SAVI, MSAVI, TCARI, GLI and VARI.
BTB	931	798	1064	Image bands and texture of bands.
BTBVI	973	840	1106	Bands, texture of bands and vegetation indices.
BVIRA	532	399	638	Image bands, texture and ratio bands.

4.3. Multiclassifier design

Figure 4-4 provides a general overview of the multi-classifier system designed to exploit the multi-temporal set of spatial and spectral features described in Table 4-4.

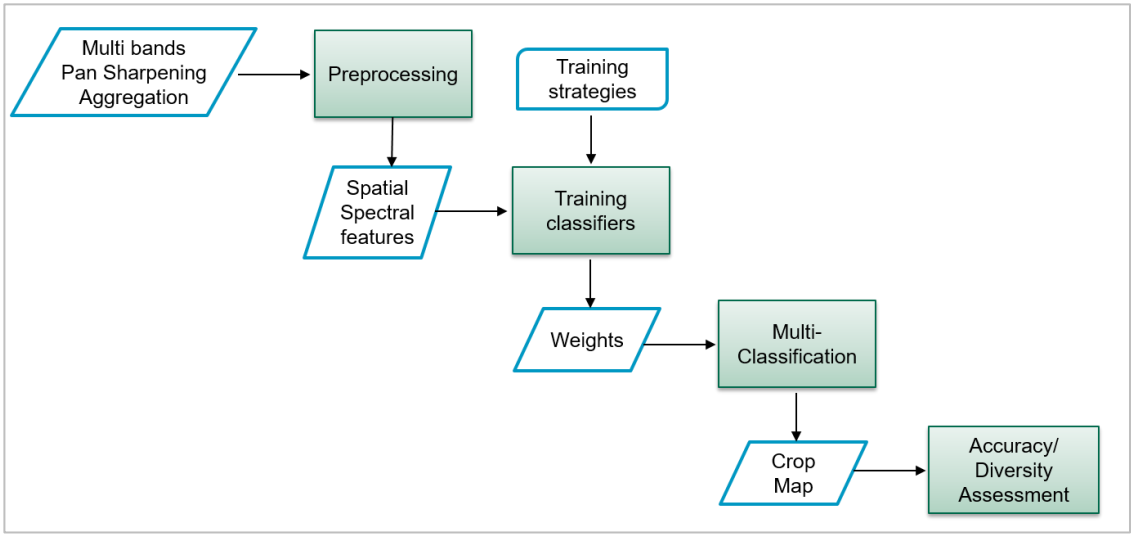


Figure 4-4. Overview of the multi-classifier system

Several choices needed to be made to design the multi-classifier system. Table 4-6 presents the choices that were made following a) best practices found during the literature review phase and b) the existing constraints in terms of the chosen cloud computing platform and of the available time for performing this research. The following paragraphs provide more details about each of the choices.

Table 4-6. Configuration variables for a Multi-classifier

Variable / Factor	Selection
Output	Label
Run-time behavior	Static
Execution	Parallel
Combination Rules	Majority voting Weighted majority voting
Weights	Kappa F-score F-score per class Producer accuracy per class
Training the combiner	No training
Training methods	k-fold Bagging
Feature Set	B BVI BTB BTBVI BVIRA
Base Classifiers	CART Maximum Entropy Models RF SVM linear SVM polynomial

Three types of output can be produced by a classifier: a) abstract form: a single unique class, b) rank level: a rank list of classes based on decreasing likelihood, and c) measurement level: probability values of the classes (Ranawana & Palade, 2006). The output of the multi-classifier is a "hard" classification that predicts a class *label* given a particular instance *x* that belongs to the feature space where the multi-classifier was trained.

A static approach was followed. This implied that each base classifier had knowledge of the whole feature space. Combination rules were fixed and did no change for a particular instance of the input data.

Outputs of base classifiers were combined in a parallel way. This means that each base classifier was trained and applied separately and their output decisions were fused through combination rules. All the classifiers outputs were class labels that are comparable between each other.

For an abstract form output (label), different combination rules can be specified. For an instance x, the decision (output) of the i^{th} classifier out of L, is given as a binary vector of the form (0, 0,…,1,…,0,0), and dimension c: $d_{i,j}$, where $d_{i,j} = 1$ if the classifier labels that instance x belonging to a class c_j and 0 otherwise. The number of classes is denoted by c. A list of most common approaches for combination rules when the classifier output is a label:

a) *Majority voting*. There are three versions of this method: i) unanimity or consensus in which all classifiers agree on the decision; ii) simple majority, in which the decision is made with simple

majority, at least 50% of the voting plus one vote; iii) plurality, also called majority, in which the decision favours the most voted label.

In unanimity voting, all the classifiers agree on the decision, so the decision of the multi-classifier $D_{x,j}$ to choose a class c_j for a given instance x can be described as follows:

$$D_{x,j} = d_{i,j} \text{ if and only if } \forall\, d_{i,j} = 1\,,\ i \in \{1..L\} \text{ and j corresponds with } C_j.$$

Plurality voting for a class C_j can be described as:

$D_{x,j} = max_{\mathrm{j=1..c}} \sum_{i=1}^{L} d_{i,j}$, in which ties are solved arbitrarily.

The same formula applies to simple majority but adds a restriction that ensures that voting reaches more than 50 % of the votes.

b) *Weighted majority voting*: This technique is based on assignation of higher importance, or weight, to the decision of those classifiers with higher expertise. The ensemble's decision $D_{x,j}$ to choose a class c_j for a given instance x can be defined as:

$D_{x,j} = max_{\mathrm{j=1..c}} \sum_{i=1}^{L} w_i d_{i,j}$

Usually weights are specified using the overall accuracy of base classifiers and can be normalized so they sum up to one (1). Another approach is to use the class accuracy of each classifier as a weight for its decision over a certain class (Du et al., 2012).

Simple majority voting (plurality) and weighted majority voting were used as combination rules. The accuracy of the classifiers over the training set were used as weights. In this sense: kappa (Cohen, 1960), f_score (Powers, 2011), producer accuracy per class (Olofsson et al., 2014) and f-score per class are computed and tested. Then, the multi-classification process consists in determining a final class label by following the most voted class or the most weighted voted class from the output decision of base classifiers.

The rules applied were fixed for all the input instances of the multi-classifier. This type of rules is not trainable because they do no *learn* from the decisions of base classifier, it makes a fixed combination of them. This non-trainable strategy was adopted because the use of fixed combination rules is effective when the base classifiers are trained carefully avoiding overfitting (Duin, 2002).

Two training strategies: k-fold and bagging were applied to explore their effect on the diversity and accuracy of the multi-classifier. K-fold was applied using k as the number of instances of base classifier. The k-fold were generated randomly but ensuring a balanced number of pixels per class. Bagging was applied taking randomly 75% of the input training data.

The purpose of testing different feature sets was to explore the discriminative power of these features during the multi-classification. The feature space available for training and testing base classifiers was divided in groups: digital levels (bands), spatial features (textures), ratio bands and vegetation indices. Five groups were configured combining these features as described in Table 4-5. The multi-classifier was applied to each of these feature sets.

Five classifiers were selected among the available algorithms in GEE. This selection was made in two steps. First, all the "trainable classifiers" (i.e. the GEE classifiers that produced an output for our data) were apply to the dataset. This first test used only bands from the original multi-spectral image because it

was the smallest set. Only classifiers that yielded a kappa value higher than 50% over the training data, were preselected since this is a conservative measure of the accuracy required to build a robust multi-classifier system (Ranawana & Palade, 2006). In other words, combining classifiers that yield less than 50% of accuracy will not lead to an improved classification. Table 4-7 presents a list of the classifiers tested in this phase with a short description and reference.

In a second step, the classifiers were run with a largest set of features. Namely, bands, texture bands and vegetation indices. Five instances of each classifier were tested and ranked according to the accuracy obtained. This test also indicated whether five instances of each classifiers finish successfully, in a reasonable time and without raising memory or computation timeout error. The classifiers that finished successfully this second test were chosen to build the multi-classifier. These classifiers are listed in Table 4-8.

Table 4-7. Available classifiers in GEE suitable to apply to the feature datasets

Name	**Description**
CART	CART classifier as described by (L Breiman, Friedman, Stone, & Olshen, 1984)
GmoMaxExt	A google optimization of the maximum entropy model. (Mann et al., 2009) (Mann et al., 2009)
Ikpamir	An optimization of SVM using intersection kernel (Maji, Berg, & Maliks, 2008)
Minimum Distance	Minimum distance classifier. (Wacker & Landgrebe, 1972)
Naïve Bayes	Naive Bayes classifier (Rish, 2001)
Continuos Naïve Bayes	Naïve Bayes classifier for continuous variables. (Bouckaert, 2004)
Pegasos Gaussian	Pegasos classifier with Gaussian Kernel. Pegasos means "Primal Estimated sub-GrAdient SOlver for SVM" (Shalev-Shwartz, Singer, Srebro, & Cotter, 2011)
Pegasos Linear	Pegasos classifier with Linear kernel (Shalev-Shwartz et al., 2011)
Pegasos Polinomial	Pegasos classifier with polinomial kernel (Shalev-Shwartz et al., 2011)
SVM	SVM classifiers. (Cortes & Vapnik, 1995) Kernel available functions are: linear, polynomial, radial and sigmoid.
Perceptron	Generate an ANN (perceptron) classifier described in (Daumé, 2006)
Random Forest	Random Forest as described by (Leo Breiman, 2001)
Winnow	Winnow classifier, an optimization of the perceptron algorithm. (Koppel, 2002)

Table 4-8. Accuracy (kappa) reached by classifier during selection tests.

Classifier	Accuracy (kappa)	
	First test	Second Test
Pegasos Gaussian	**0.9301**	Not finish
CART	**0.8460**	0.9444
SVM Radial	**0.8327**	Not finish
Random Forest	**0.6922**	0.6487
SVM Polinomial	**0.5546**	1.0000
SVM Linear	**0.5449**	0.9400
GmoMaxExt	**0.5241**	0.7219

The use of SVM methods requires optimizing the so-called kernel parameters. The optimal set of kernel parameters for the linear SVM were obtained through in two steps. First, 30% of the training set was used for training and 20% for testing set. Next, an array of c values ranges, in [10, 100, 300, 500] an initial search is performed and the best c that produced the highest kappa value is selected. Then, in a second step a narrower set of c values is executed to fine tune the final value of c. The best c is determined by the highest kappa achieved.

The parameters for the polynomial kernel were determined through a grid search that included all the possible combinations of the following range of values: a) cost c = [10, 100, 300, 500], b) gamma = [0.1, 1, 10], c) degree = [2, 3, 4], and d) coef0 = [1, 10, 100, 1000]. These values are a subset of those used by Nitze et al. (2012). Training and testing set were used in the same proportion than SVM linear optimization.

To improve SVM classifiers performance, input features values (columns) were linearly scaled to [0,1] (Chih-Wei Hsu, Chih-Chung Chang, 2008). Features were scaled using the following formula:

$$feature_{value} = (feature_{value} - min)/(max - min)$$

where min and max are the minimum and maximum value for that attribute respectively.

4.4. Diversity measures

Diversity is associated with accuracy of a multi-classifier. Several metrics have been proposed for diversity. Diversity measures can be pair-wise or global. In a multi-classifier system of L classifiers, L(L-1)/2 pair-wise measures can be calculated and averaged to produce a global measure for the ensemble.

Given two classifiers C_i and C_j, let a be the fraction of instances classified correctly by both classifiers, d the fraction of instances misclassified by both classifiers, b the fraction of instances correctly classified by C_i but not by C_j and d the fraction of instances correctly classified by C_j but not by C_i. Logically, $a + b + c + d = 1$. Table 4-9 illustrates the definition of a, b, c and d.

Table 4-9. Fraction of instances classified by Classifiers (Ci, Cj)

	C_j is correct	C_j is wrong
C_i is correct	a	b
C_i is wrong	c	d

Accepted diversity measures of easy computation in a multi-classifier (Brown & Kuncheva, 2010; Du et al., 2012; Polikar, 2006) are given below:

a) Correlation: indicates the degree of correlation between the classifiers outputs. It is defined as:

$$\rho_{i,j} = \frac{ad-bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}}, 0 \leq \rho_{i,j} \leq 1$$

Higher diversity is reached when $\rho = 0$.

b) Q-statistic: ranges between -1 and 1. It assumes 1 when C_i and C_j classify correctly the same instances and -1 when C_i and C_j misclassify the same instances. Q is defined as:

$$Q_{ij} = \frac{(ad-bc)}{(ad+bc)}, -1 \leq Q \leq 1$$

Higher diversity is reached when $Q = 0$.

Non-pair diversity measures consider the behaviour of all classifiers at once. Entropy is one of the most popular metric (Brown & Kuncheva, 2010; Du et al., 2012; Polikar, 2006). Under the assumption that a higher diversity is reached when the half of classifiers differ from the others, for a given set of L classifiers, entropy is defined as:

$$E = \frac{1}{N}\frac{2}{L-1}\sum_{i=1}^{N} \min\{\zeta_i, (L-\zeta_i)\}, \ 0 \leq E \leq 1$$

where N is the sample cardinality and ζ_i is the number of classifiers that misclassified an instance given Xi. Higher diversity is reached for $E = 1$.

The interrater agreement is another global metric for diversity. It measures the level of agreement between classifiers and is defined as follows (Ranawana & Palade, 2006):

$K = 1 - \frac{1}{2\bar{\rho}\,(1-\bar{\rho})} Dis_{avg}$, where $\bar{\rho}$ is the average accuracy of classifiers and Dis_{avg} is the average disagreement calculated from the pair-wise disagreement measure described previously.

The relation between diversity and global accuracy of the multi-classifier was analysed. Pair-wise diversity measures of the multi-classifier: Q-statistic and correlation (ρ) were computed and compared with the accuracy of the multi-classifier. Entropy and Interrater agreement were also tested although they are not pair-wise measures.

4.5. Multiclassifier implementation

An algorithm using the API interface of GEE was coded in python to execute the base classification step. Results from this step were exported to the cloud. Afterwards, another algorithm, running in a local python environment, fused the decisions of base classifiers using intermediate results and accuracy measures computed in the previous step. Also, diversity is calculated in the second step.

Before applying the multi-classifier, multi-temporal features sets were generated and pre-processed in several steps that are illustrated in Figure 4-5 and enumerated as follows:

1. Aggregation and pan-sharpening process were executed
2. An available feature extraction algorithm from STARS project was applied to each image in each dataset to generate training and testing sets per image per dataset

3. Files generated in step two were concatenated to produce multi-temporal series of training and testing sets.
4. A filter by rows was applied to ensure that all sets have same number of pixels per class and this number must be greater than 300 pixels for training (Löw, Conrad, & Michel, 2015) and more than 100 pixels for testing (Khobragade & Raghuwanshi, 2015). 400 pixels were selected per class to train base classifiers. 225 pixels per class were selected to test the accuracy of the multi-classification result.
5. A filter by columns was applied to generated the feature sets described in Table 4-4.
6. Training and testing sets were upload to the cloud to made available for use in GEE.

Base classification process is presented in Figure 4-6. This process is the same for any feature set or image type, then a generic case is illustrated. Given a fixed number of instances for each base classifier [1,3,5] base classifiers, the algorithm of base classification is done as follows:

1. Training data is sampled using k-fold or bagging.
2. Train first instance of: CART, GmoMaxEnt, RF, SVM linear and SVM polynomial. Accuracy measures to use as weights are generated in this step.
3. The test set is presented to the trained base classifiers to generate the suite accuracy metrics for base classifiers.
4. The class label assigned by each base classifier is stored as attribute. Four binary vectors are calculated using a) the label assigned by the base classifier and b) the weights: kappa, f-score, producer accuracy per class and f_score per class. These weights are obtained in step 2.
5. If number of instances executed is lower than fixed number of instances for each classifier, go back to step one.
6. The accuracy metrics of each base classifier (training and testing sets) are exported as csv files.
7. Features classified (tables are exported

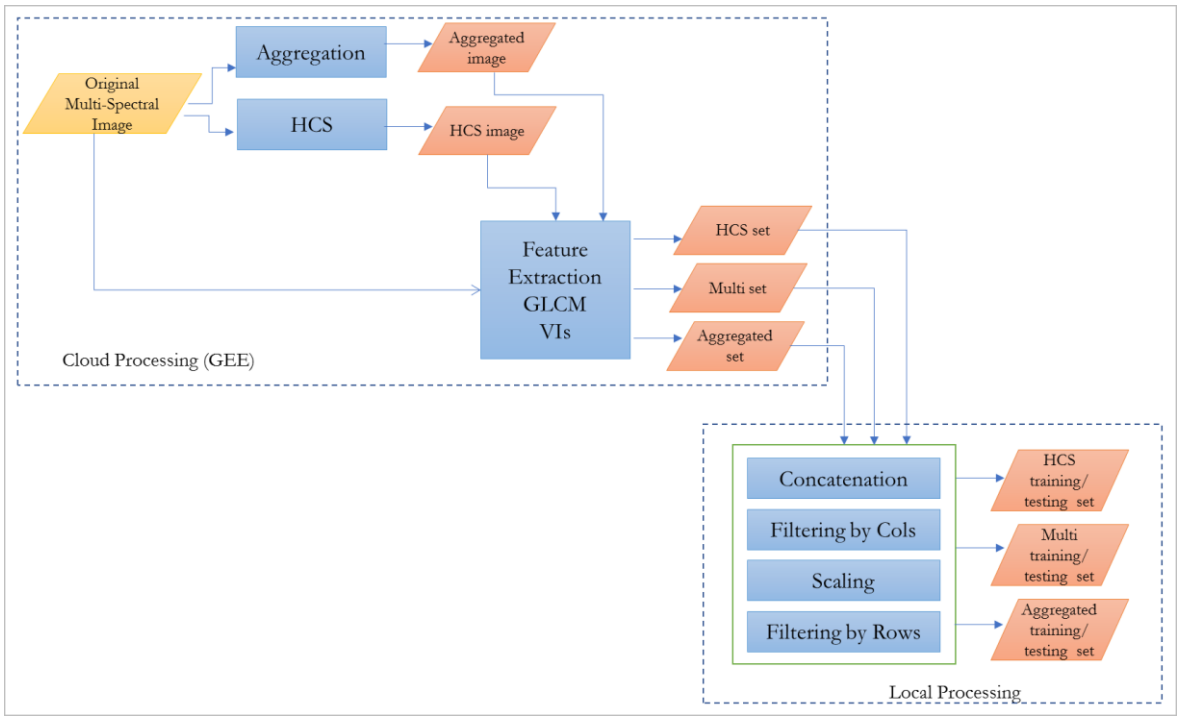


Figure 4-5. Pre-processing of Multi-temporal image series to configure three features datasets.

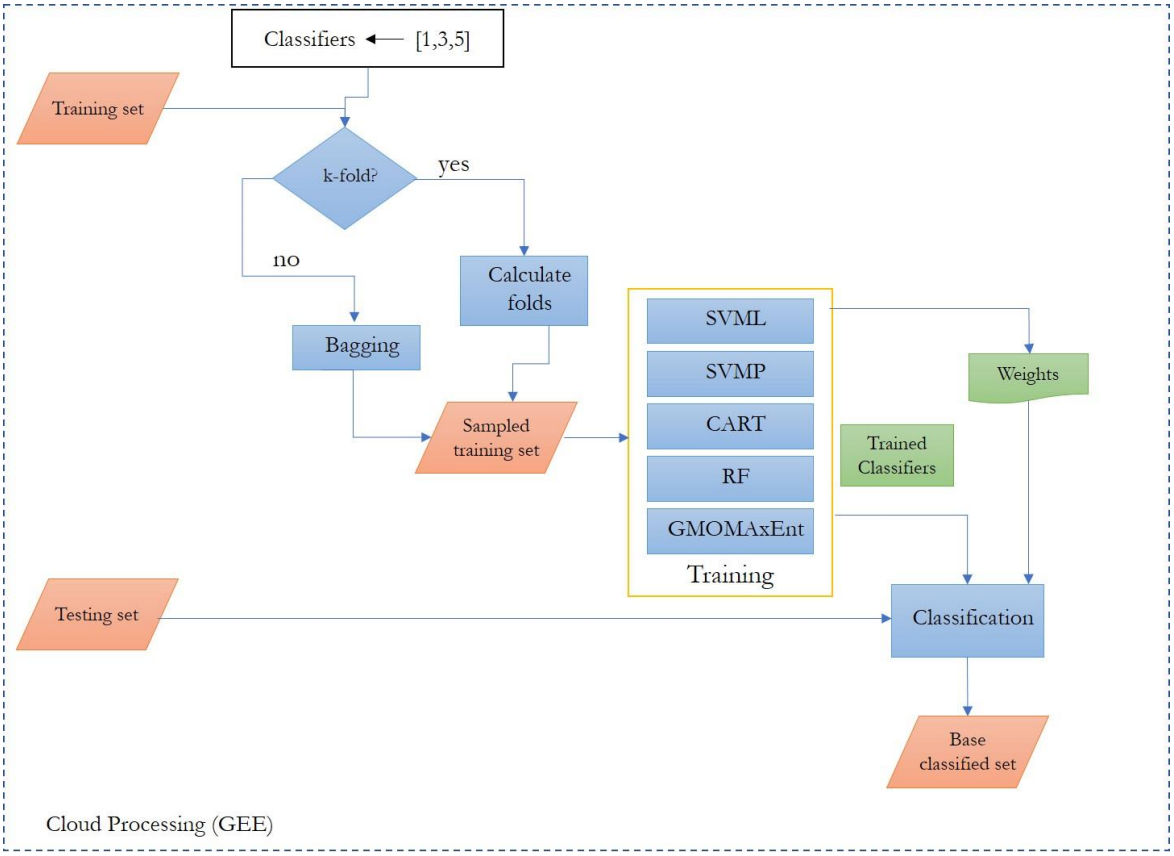


Figure 4-6. Base classification overview

The output files of the base classifiers were processed locally to apply the combination rules (voting and weighted majority voting) that produced the multi-classification. This processing, summarized in Figure 4-6, can be described as follows:

1. For each pixel, the most voted (mode) class is determined and stored as an attribute.
2. For each pixel and for each base classifier, binary vectors produced during base classification with kappa, f-score, producer accuracy per class and f_score per class as weights respectively, are summed into four binary vectors.
3. The class label is assigned to the class that obtained the maximum weight.
4. Multi-classifier accuracy measures are computed and exported as csv file.
5. Pair-wise diversity metrics are computed according to the formulas listed in section 2.4.5. Entropy metric is calculated as well.
6. Multi-classified features and diversity measures are exported as csv files

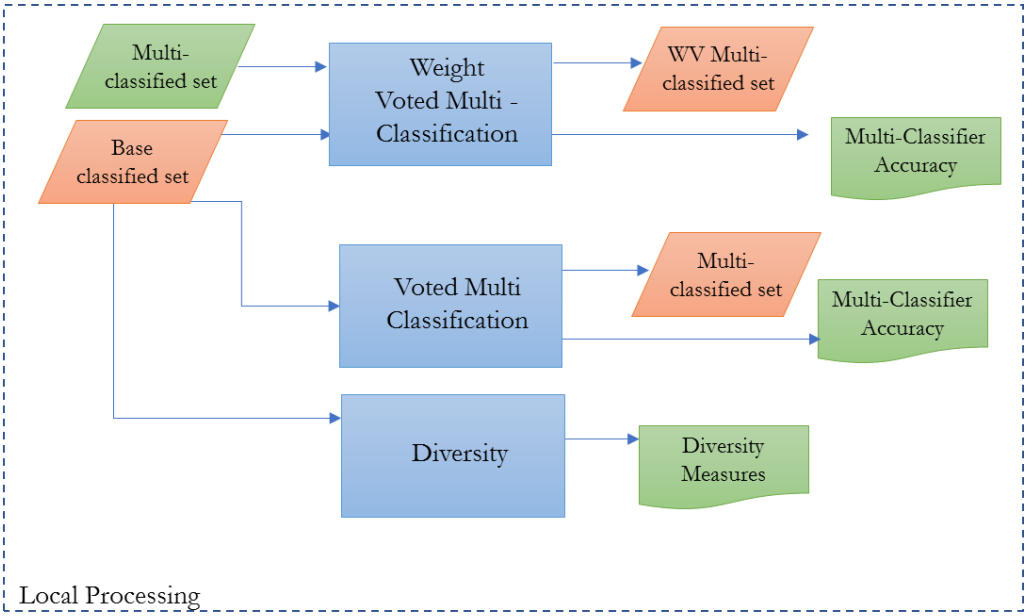


Figure 4-7. Multi-classification process

The best combination of image type, feature set, training strategy, voting rule and number of base classifiers determined by means of empirical experimentation can be used to produce the crop map.

4.6. Design of experiments

The purpose of the experimental phase is to determine the best combination of training strategy, features set, number of base classifiers and combination rule that yield to the best multi-classifier accuracy (kappa). Experiment were designed to investigate the influence of these four factors when the multi-classifier was applied to the multispectral, pan-sharpened and aggregated dataset. Table 4-10 shows the experimental design factors (variables) and their levels (possible values) used in the experimentation phase.

A full factorial design was selected to test all the possible levels of each factor. This design allows to analyse the multi-classifier performance under every possible combination and discriminate the contribution of each factor in the achieved result.

The selected configuration generated 150 combinations for each image series (5x2x3x5=150) that produced 450 experiments. However, the algorithm computed simultaneously all the combination rules listed in Table 4-10, then 30 experiments were planned for each image series reaching a total of 90 experiments. The experiment results are presented and discussed in chapter five.

Table 4-10. Factors and levels for experiments

Factor / Variable	Levels (Fixed)
Feature Set	1. Bands (B) 2. Bands + Vis (BVI) 3. Bands + VIs + Ratio Bands (BVIRA) 4. Bands + Textures (BTB) 5. Bands + Textures + VIs (BTBVI)
Training Strategy	1. Bagging 2. K-fold
Number of base classifiers	5 classifiers = 1 instance of each base classifier 15 classifiers = 3 instances of each base classifier 25 classifiers = 5 instances of each base classifier
Rule combination	1. Majority Voting (Plurality) 2. Weighted Majority Voting (kappa as weights) 3. Weighted Majority Voting (f-score average as weights) 4. Weighted Majority Voting (accuracy per class as weights) 5. Weighted Majority Voting (f-score per class as weights)

5. EXPERIMENTAL RESULTS AND DISCUSSION

In this chapter the experimental results of applying the multi-classifier to three multi-temporal datasets are explained and discussed in detail. First, general results are reported for multispectral, pan-sharpened and aggregated image series. Second, the influence of training strategy, feature set and number of classifiers in the multi-classification accuracy is discussed. Third, results yielded using different combination rules area compared. Four, the accuracy of base classifiers is analysed and compared to the accuracy of the multi-classifier. Five, the influence of training strategies over diversity is explored as well as the relation between diversity and accuracy. Finally, a crop map example is provided to illustrate the application of the multi-classifier.

5.1. Multi-classifier accuracy for multi-spectral dataset

The first dataset tested was the multi-spectral image series. Table A-1presents the accuracy (kappa) obtained by the multi-classifier applied to this dataset for each combination of factors. Highest accuracy per group are bold. No results are listed when twenty-five classifiers were applied to the feature set consisting of bands, vegetation and ratio bands because there was no possible way to perform this test due to time constraints.

Best accuracy was achieved by the combination of fifteen classifiers with a k-fold training strategy applied to a feature set comprised of image bands, vegetation indices and ratio bands using weighted majority voting. In contrast, poorest accuracy was achieved by using twenty-five classifiers and a bagging training applied to the same feature set. Besides, it is noticeable that there are no relevant differences between the mean kappa obtained by the voting (majority) and the weighted voting methods. In addition, minor differences are observed between the kappa values obtained applying f-score, f-score per class, kappa and producer accuracy per class as weights, this is explained by the column standard deviation (SD) of mean kappa that is close to zero. For all the features sets tested k-fold led to a better accuracy than bagging, except when only bands were used.

5.2. Multi-classifier accuracy for pan-sharpened dataset

The second dataset tested was the fusion image series obtained by pan-sharpening. Table A-2 lists the kappa values obtained when the multi-classifier was applied to this dataset. Highest accuracy per group are highlighted. Best accuracy is achieved when fifteen classifiers were combined applied to the feature set comprised of bands, vegetation indices and ratio bands using a k-fold training strategy. No substantial differences are observed between the voting and the weighted voting method. Also, no differences are appreciated between distinct factors used as weights.

Influence of training strategies, k-fold and bagging, for the pan-sharpened image series was also analysed. Similarly, to the result obtained for multi-spectral images, k-fold achieved better accuracy than bagging for all feature sets excluding the feature set composed only by bands.

5.3. Multi-classifier accuracy for aggregated dataset

The third dataset analysed was the aggregated image series. Table A-3 provides the results obtained by the multi-classifier when was applied to this dataset. The feature set comprised by bands, vegetation indices and ratio bands lead the best result using k-fold and a weighted voting rule. Besides, k-fold achieved better accuracy than bagging for all feature sets, including the feature set composed only by bands.

5.4. Relation between training strategy and multi-classifier accuracy

The effect of training strategies in the multi-classification accuracy was reported during the experiments. K-fold provided higher accuracy (kappa values) than bagging. Figure 5-1 illustrates the mean values of kappa obtained by the multi-classifier when bagging and k-fold training methods were applied to multi-spectral, pan-sharpened and aggregated images. K-fold provided better accuracy than bagging with a discrete difference.

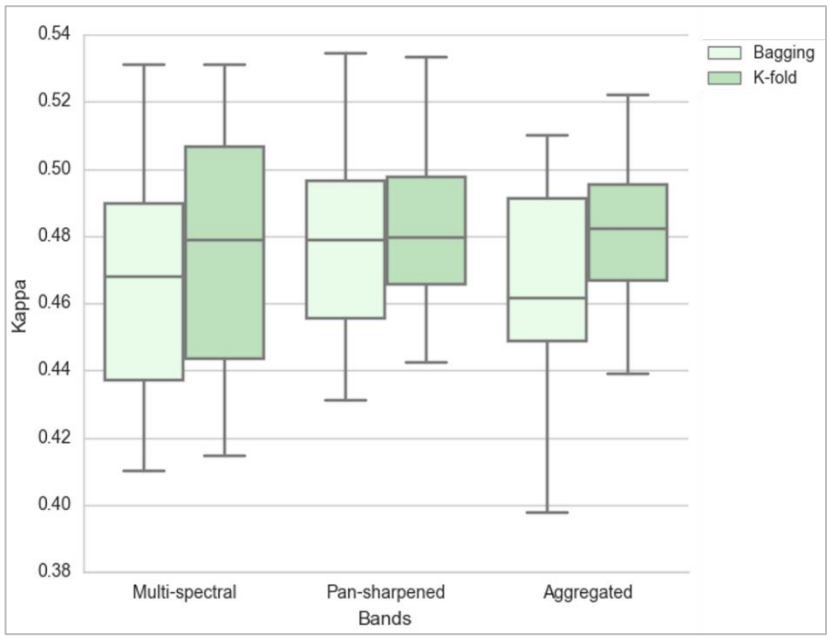


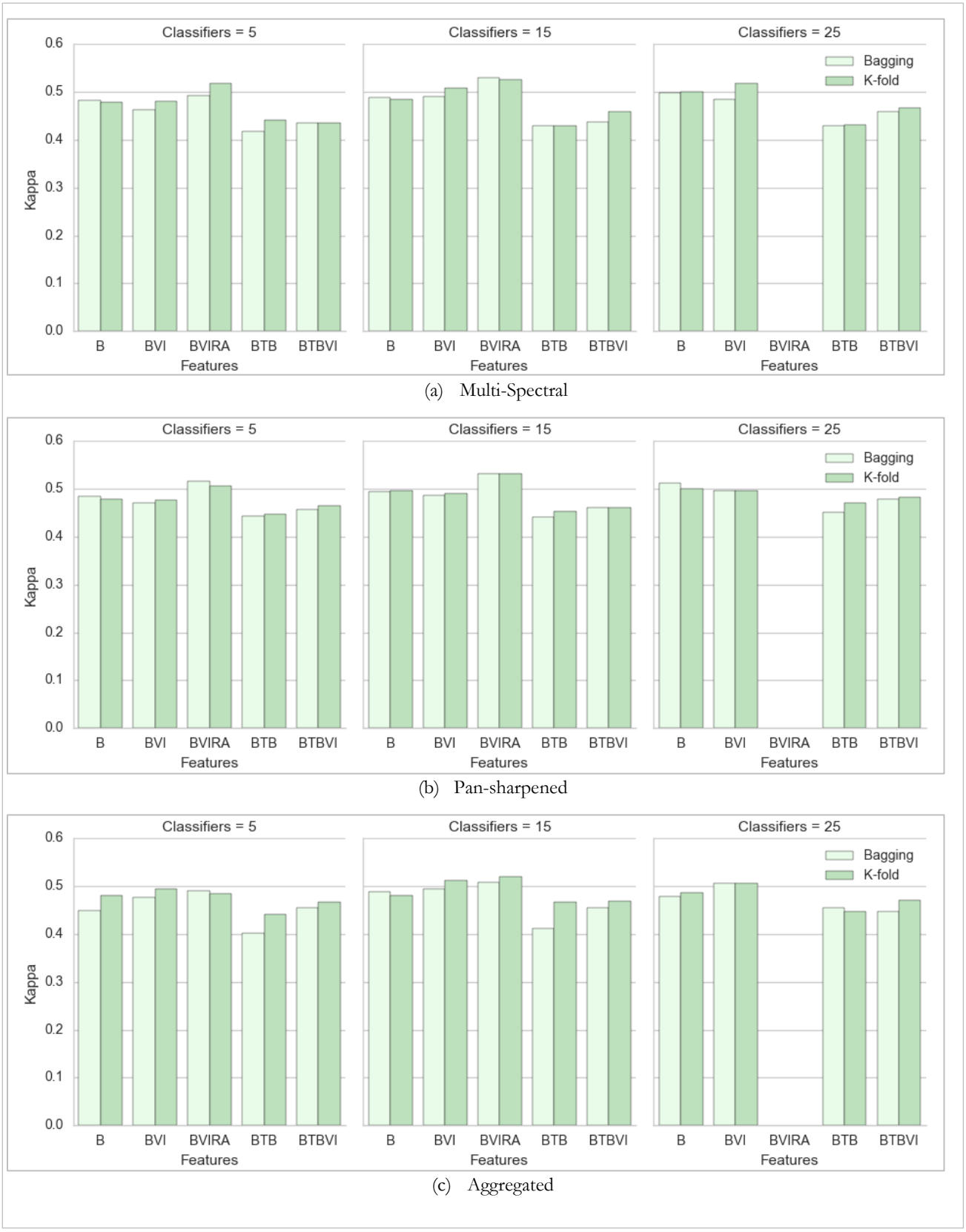
Figure 5-1. Kappa values for each training strategy applied to multi-spectral, pan-sharpened and aggregated images.

In general, k-fold provided better accuracy than bagging with a discrete difference. The highest value for k was five as the number of instances of base classifiers. Further studies can explore whether a higher number of folds (i.e. ten) increase the accuracy (Kohavi,1995) reached by k-fold and the difference with results yielded by bagging.

5.5. Relation between feature set and multi-classifier accuracy

With respect to the discriminative power of features tested, highest classification rates were achieved by the features set comprised by bands, vegetation indices and ratio bands. The three datasets reached comparable results that are represent in Figure 5-2. Lower accuracy was obtained by the feature set composed by bands and texture of bands.

Textures of bands do no increase classification accuracy but increase the dimensionality of the problem providing more complexity. Cossu, (1988) found that some of the GLCM textural features are highly correlated between each other. The window size used in this study was 3x3 that means one interpixel distance. Rao et al.,(2002) optimized the GLCM textures feature extraction and found that the most discriminative textural features were generated using different inter-pixel distance. Further studies can exploit this approach with an efficient feature selection to discriminate the most informative textural features. Vegetation indices and ratio bands showed to increase the discrimination ability of the multi-classifier. Further studies are needed to explore the discriminative power of textures of vegetation indices in high resolution images because they demonstrated to be useful in crop identification using medium resolution images (Shukla et al., 2016).



(a) Multi-Spectral

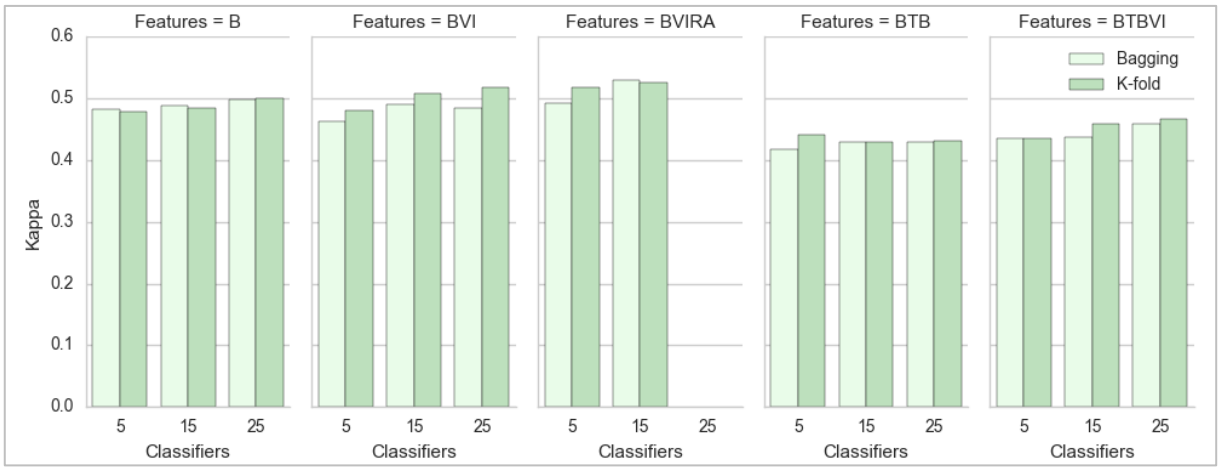
(b) Pan-sharpened

(c) Aggregated

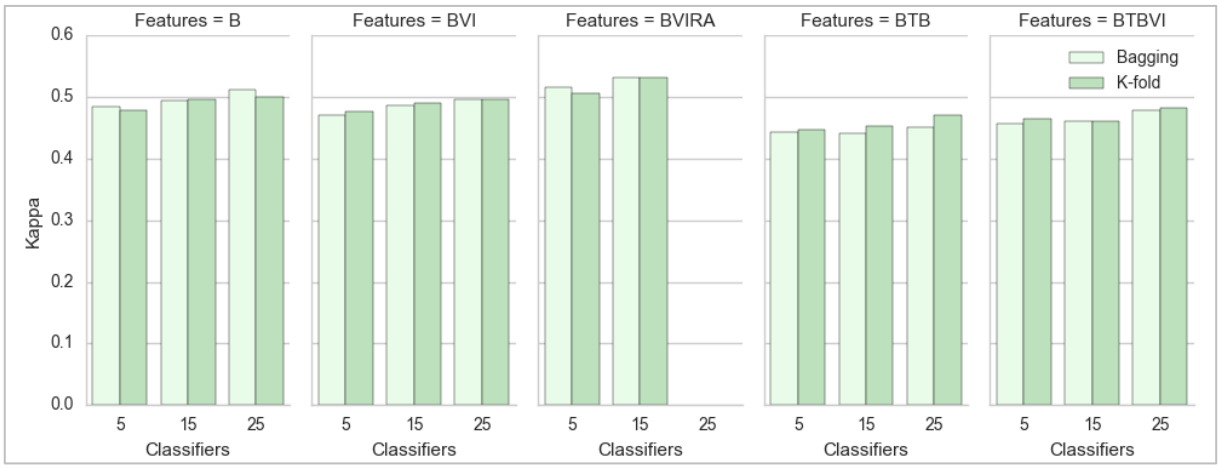
Figure 5-2. Kappa values and feature type using 5, 15 and 25 classifiers for a) multi-spectral, b) pan-sharpened, and c) aggregated dataset.

5.6. Relation between number of classifiers and multi-classifier accuracy

The number of base classifiers included in the multi-classifier is another factor analysed and related to its performance in terms of accuracy (kappa values). Figure 5-3 displays the mean kappa values grouped per feature set, obtained when a different number of classifiers were applied to multi-spectral, pan-sharpened and aggregated dataset.



(a) Multi-Spectral



(b) Pan-sharpened

(c) Aggregated

Figure 5-3. Kappa values and number of classifiers for each type of feature set for a) multi-spectral images, b) pan-sharpened, and c) aggregated dataset.

For the multi-spectral dataset, the multi-classifier accuracy improved as the number of classifiers increased. The feature set composed by bands and texture of bands showed a different pattern, reaching lower accuracy as the number of classifiers increased from five to fifteen when k-fold training was used. Besides, no relevant changes in kappa values for this group was observed when the number of classifiers shift from fifteen to twenty-five with both training methods.

Pan-sharpened dataset showed an accuracy improvement as the number of classifiers increased for all feature set. In contrast, the accuracy of the multi-classifier applied to the aggregated dataset presented an improvement when the number of classifier increased from five to fifteen but no changes or poorer kappa values when the number of classifiers shifted from fifteen to twenty-five excluding the bands features that achieved better accuracy as the number of classifiers increased from five to fifteen and from fifteen to twenty-five.

5.7. Relation between combination rules and multi-classifier accuracy

Weighted voting method led to a better accuracy than majority voting for each dataset tested, as can be observed in Figure 5-4. However, no relevant dissimilarities were observed between different weights. Thus, the multi-classifier was comparable for all the weights selected.

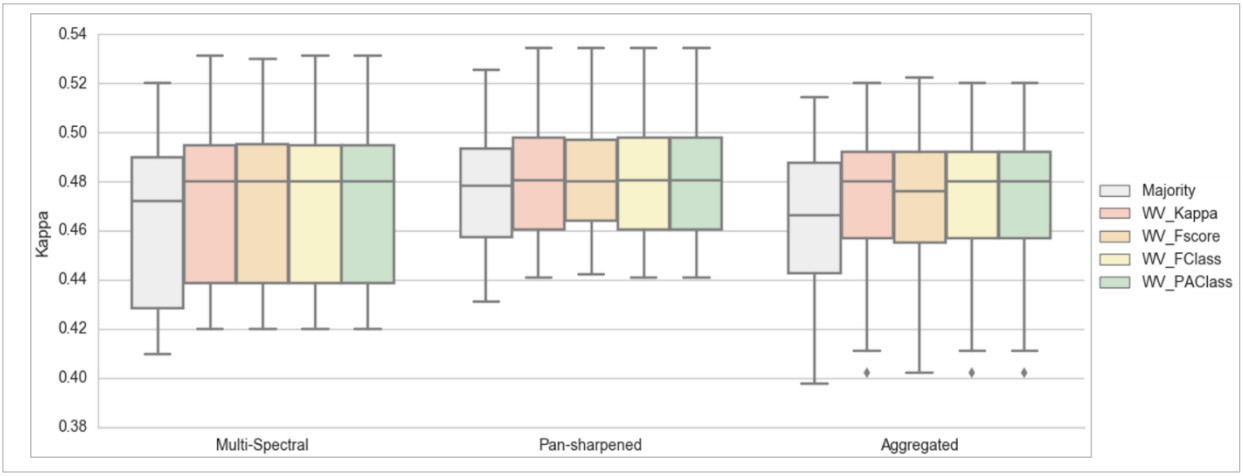


Figure 5-4. Comparison between accuracy (kappa values) and voting method.

The best performance achieved by the multi-spectral and pan-sharpened dataset may look similar, the pixel size of pan-sharpened images is 0.5m which represent and improvement in term of spatial details over an equivalent accuracy but at a coarser pixel size.

Weighted majority voting outperformed simple majority or plurality. However, there was not observed relevant differences between different weights used as measures as represented in Figure 5-7. No major differences were noticed between the accuracy led by the four weighted voting schemes used.

5.8. Accuracy of base classifiers

Base classifiers accuracy is presented in Table 5-1 and illustrated in Figure 5-5. It can be observed that linear SVM outperform other classifiers, reaching a mean around 0.44 and 0.41 when is applied to pan-sharpened and aggregated dataset respectively. Linear SVM achieved results comparable to polynomial SVM when was applied to the multi-spectral dataset. Besides, linear SVM did not show high sensitive to sampling training data, showing low variability in its performance. This results agree with findings of

Khobragade, Raghuwanshi, & Malik (2016) who revealed that not all kernel functions perform well over an specific data.

Table 5-1. Base classifiers Accuracy. Classifier abbreviations are, GMOMaxEnt = Google Margin Optimization of Maximum Entropy Models, SVML= Linear SVM, SVMP = Polynomial SVM.

Dataset	Classifier	Kappa			
		Mean	STD	Minimum	Maximum
Multi-Spectral	CART	0.3408	0.0307	**0.2756**	0.4156
	GMOMaxEnt	0.3890	0.0515	0.3100	**0.4767**
	RF	0.3848	0.0318	0.3267	0.4422
	SVML	0.4032	0.0325	0.3422	0.4733
	SVMP	**0.4090**	0.0390	0.3278	0.4656
Pan-sharpened	CART	0.2906	0.0218	**0.2344**	0.3467
	GMOMaxEnt	0.4271	0.0298	0.3733	0.4944
	RF	0.3461	0.0407	0.2811	0.4711
	SVML	**0.4381**	0.0344	0.3756	**0.4989**
	SVMP	0.4137	0.0225	0.3733	0.4644
Aggregated	CART	0.3456	0.0332	**0.2644**	0.4256
	GMOMaxEnt	0.3997	0.0387	0.3378	0.4778
	RF	0.3778	0.0350	0.2767	0.4733
	SVML	**0.4110**	0.0446	0.3322	**0.4867**
	SVMP	0.3973	0.0278	0.3500	0.4500

The lowest accuracy was obtained by CART that reached a minimum kappa value of 0.2344 when was applied to the pan-sharpened dataset. Lower variance of its kappa values suggests that is not sensitive to sampling training data. Random Forest followed second in low accuracy for all datasets.

Although kappa values of base classifiers did not show drastic differences. It was observed that polynomial SVM outperform when it was applied to the multi-spectral dataset and behaved close to linear SVM when it applied to pan-sharpened SVM.

GMOMaxEnt obtained kappa values close to the kappa values achieved by linear SVM when was applied to pan-sharpened and aggregated datasets. Similarly, GMOMaxEnt accuracy was comparable to linear SVM when it was applied to multi-spectral dataset. GMOMaxEnt showed to be sensitive in change to parameters having the highest variance. This results revealed GMOMaxEnt as a promising classifier and a good candidate for a multi-classifier for complex problem as the addressed in this study.

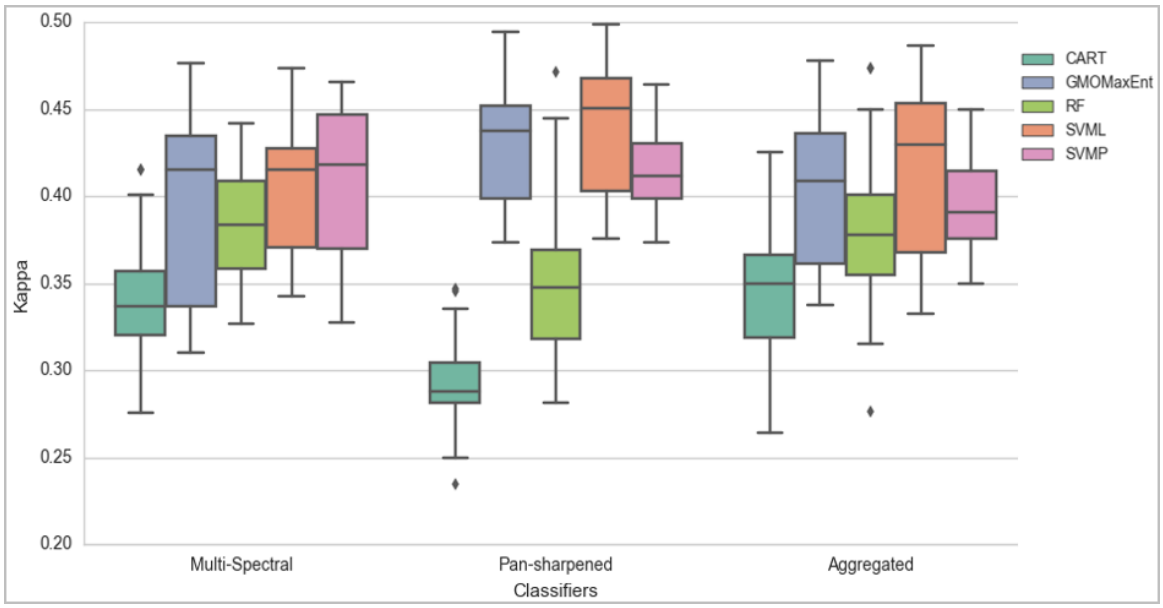


Figure 5-5. Comparison of base classifiers accuracy. Classifier abbreviations are: GMOMaxEnt = Google Margin Optimization of Maximum Entropy Models, SVML= Linear SVM, SVMP = Polynomial SVM.

5.9. Comparison between base and multi-classifier accuracy

A multi-classifier is expected to perform better than singe base classifiers, tests results probed than indeed for the three datasets, the multi-classifier achieved better accuracy than any of the base classifiers applied. Figure 5-7 shows a comparison between the kappa values yielded by the multi-classifier and the kappa values yielded by the base classifiers.

A specific case of the multi-classifier applied to the multi-spectral image is display in Figure 5-6. In this case, the feature set was comprised only by image bands and the best result was 0,52 with a weighted voting method based on kappa values over the training set. Main diagonal values are highlighted. It can be observed that multi-classification presented an improvement in the classification accuracy for all the classes compare to all the classifiers accuracy. The highest classification rate was achieved by class five (Cotton) whereas the lowest was achieved by class two (Millet).

Only for class two, the multi-classifier performed as well as one of the base classifier (RF). For the rest of cases, the multi-classifier presented an improved from 7% until 51% for this specific case.

Classifier: CART
kappa: 0.336667
CFM :

	Class1	Class2	Class3	Class4	Class5
Class1	**112**	37	30	19	27
Class2	36	**87**	45	31	26
Class3	16	39	**112**	30	28
Class4	6	34	28	**108**	49
Class5	19	27	31	39	**109**

Classifier: GMO
Kappa: 0.441111
CFM:

	Class1	Class2	Class3	Class4	Class5
Class1	**118**	31	21	14	41
Class2	45	**68**	63	39	10
Class3	12	17	**149**	32	15
Class4	10	24	18	**133**	40
Class5	9	16	20	26	**154**

Classifier: RF
Kappa: 0.398889
CFM:

	Class1	Class2	Class3	Class4	Class5
Class1	**128**	26	37	15	19
Class2	46	**93**	45	22	19
Class3	15	34	**133**	21	22
Class4	14	30	32	**116**	33
Class5	12	25	28	46	**114**

Classifier: SVM Linear
Kappa: 0.434444
CFM:

	Class1	Class2	Class3	Class4	Class5
Class1	**115**	37	28	21	24
Class2	44	**75**	64	29	13
Class3	20	41	**134**	23	7
Class4	8	27	20	**139**	31
Class5	6	21	22	23	**153**

Classifier: SVM Polynomial
Kappa: 0.447778
CFM:

	Class1	Class2	Class3	Class4	Class5
Class1	**127**	17	33	25	23
Class2	57	**78**	56	17	17
Class3	22	25	**135**	20	23
Class4	10	24	26	**129**	36
Class5	8	14	21	23	**159**

Classifier: Multi-classifier
Kappa: 0.52
CFM:

	Class1	Class2	Class3	Class4	Class5
Class1	**139**	24	21	21	20
Class2	42	**93**	53	22	15
Class3	18	28	**147**	19	13
Class4	5	28	18	**149**	25
Class5	5	17	21	17	**165**

Figure 5-6. Comparison of Confusion Matrix of each classifier and confusion matrix of the multi-classifier

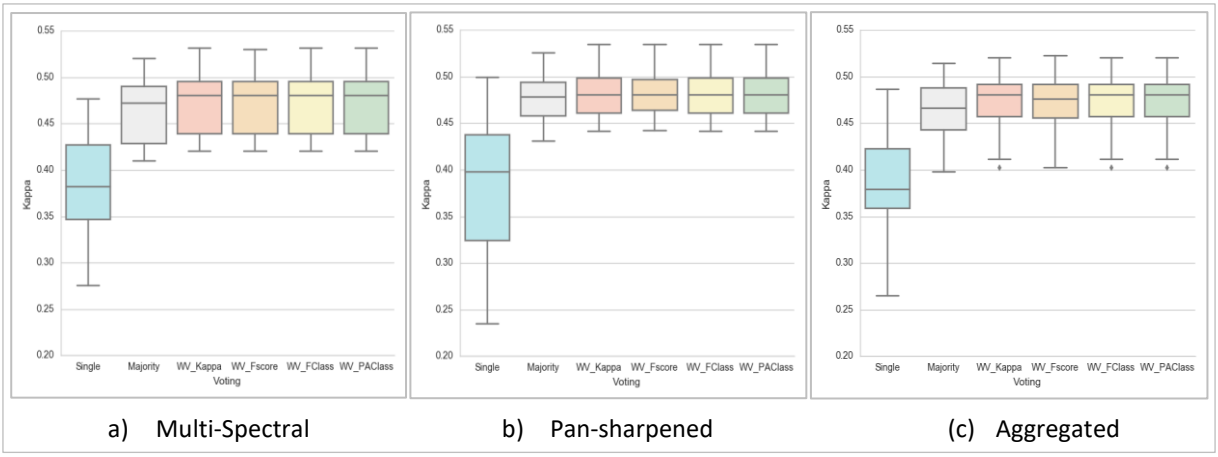


Figure 5-7. Comparison between base classifiers and the Multi-classifier accuracy with different voting methods.

The best kappa value was 0.534 that was obtained when the multi-classifier was applied to a pan-sharpened dataset (see Table A-2. Kappa values for the multi-classifier built with the pan-sharpened images.). This represents an improvement up to 21.91% in comparison to the average kappa values (0.438) reached by base classifiers applied to this dataset (see Table 5-1). Multi-classifier accuracy improvement in this research is higher than the accuracy improvement obtained by Löw, Conrad, & Michel (2015) and Lijun et al. (2011). The former presented a multi-classifier that reached an overall accuracy of 56% increasing in 6% the performance of its base classifiers. The latter reported a multi-classifier that improved the accuracy of its base classifiers in 2.22%.

The multi-classifier presented is considerably complex in comparison to the multi-classifier reported by Li et al.,(2016) who applied a multi-classifier to features set formed by images bands, GLCM and Gabor features textures of images up to 1024x1024 pixels.

Figure 5-7 reveals that kappa values of base classifier are lower than 50% but nevertheless the multi-classifier outperformed in all cases. Kappa value is a conservative the highest accuracy obtained by a single classifier is lower than the best multi-classifier accuracy. This is true for the three datasets tested.

5.10. Relation between training strategies and diversity

The impact of training strategies over diversity is illustrated in Figure 5-8. Mean values of Q-statistic, correlation and entropy are reported when k-fold and bagging were applied to multispectral, pan-sharpened and aggregated image series. Lower values of Q-statistic and correlation imply higher diversity whereas higher values of entropy mean higher diversity.

Results shows that Q-statistic and correlation, for pan-sharpened and aggregated images, had better (lower) values when k-fold was applied. In contrast, for multi-spectral images, Q-statistic achieved better values when bagging was applied. Values of correlation for both strategies were comparable for this dataset.

K-fold led to better (higher) values of entropy when was applied to multispectral and pan-sharpened images. For aggregated images, entropy was slightly lower when k-fold was applied.

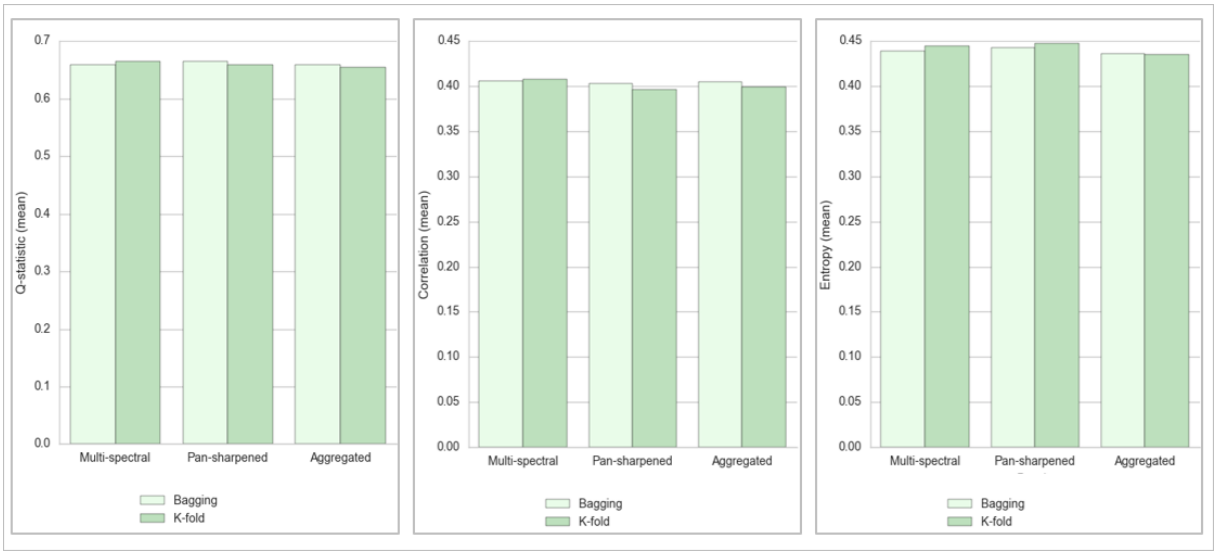


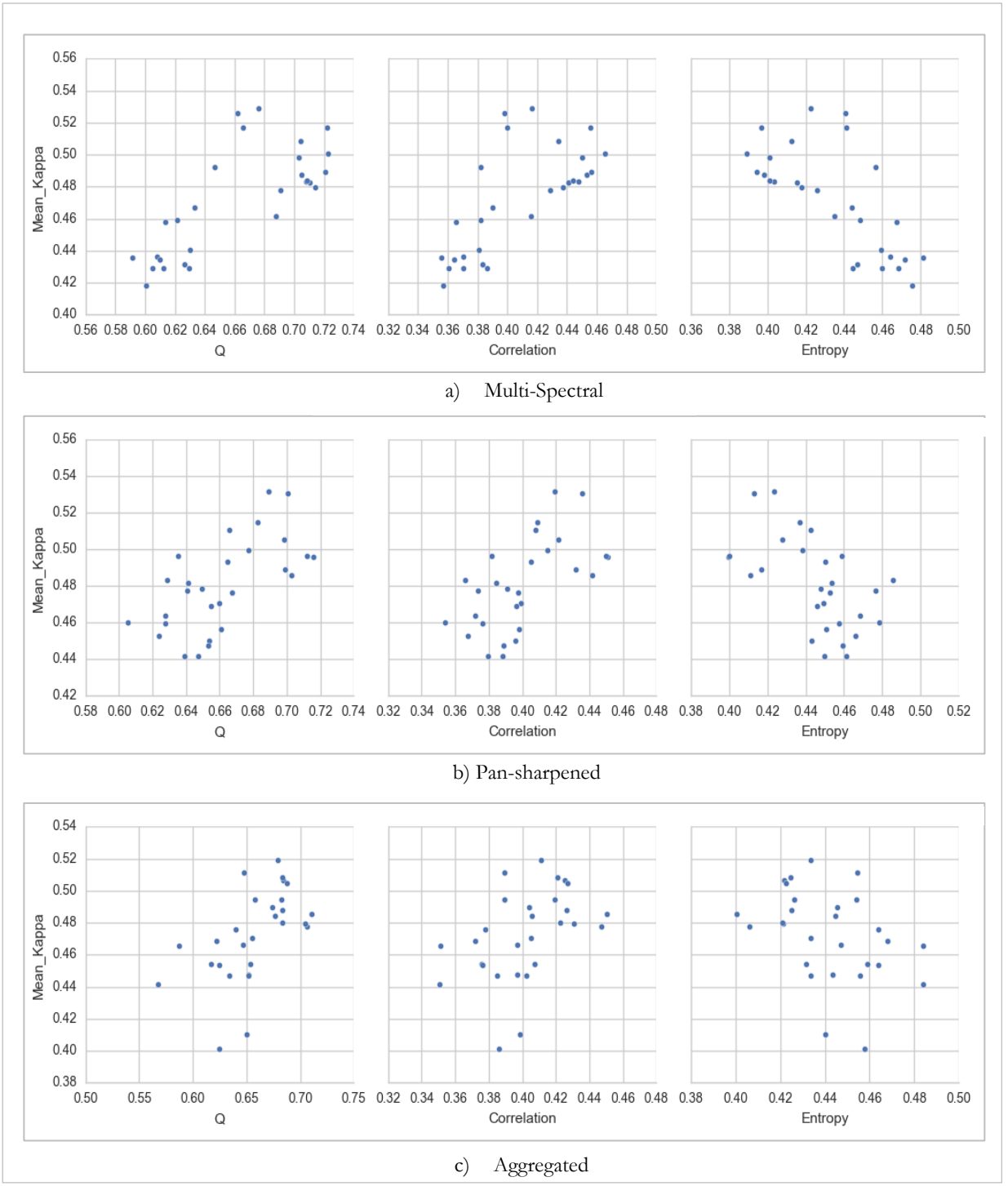
Figure 5-8. Diversity metrics: Q-statistic, correlation and entropy for multispectral, pan-sharpened and aggregated images using two different training strategies

5.11. Relation between diversity and accuracy

Regarding the relation between diversity and multi-classifier accuracy, results obtained are counter-intuitive since there is not a clear relation or pattern between these two elements as it is display in Figure 5-9. Best values of q-statistic, correlation or entropy are not coincident with best values of kappa.

No clear correlation was observed between diversity and feature type set. In the multi-spectral dataset, whereas the features set comprised by bands reached the best q-statistic value in pan-sharpened series. Aggregated series reached the best value of q-statistic using bands and textures bands.

Experiments results did not provide insights about correlation between accuracy and diversity in the multi-classifier. Although, it may be counter-intuitive, previous works drawn similar conclusions (Bi, 2012; Kuncheva & Whitaker, 2003; Tang et al., 2006). Diversity can be discriminate between "good" and "bad" diversity. A higher value of the former decrease the majority voting error while a higher value of the latter increase the majority voting error (Brown & Kuncheva, 2010). For this reason, a deeper analysis of "bad" and "good" diversity can be performed to discriminate the obtained diversity and analyse its effects on the multi-classifier accuracy.



a) Multi-Spectral

b) Pan-sharpened

c) Aggregated

Figure 5-9. Comparison of diversity measures and accuracy per each dataset

5.12. Crop map

Figure 5-10 presents a subset of a classified image, result of applying the multi-classifier to the pan-sharpened dataset. Selected features are bands, vegetation indices and ratio bands. Five classifiers were applied and their decisions were combined by weighting majority voting with kappa accuracy as weight. A pan-sharpened image in a natural color composition is used as a background.

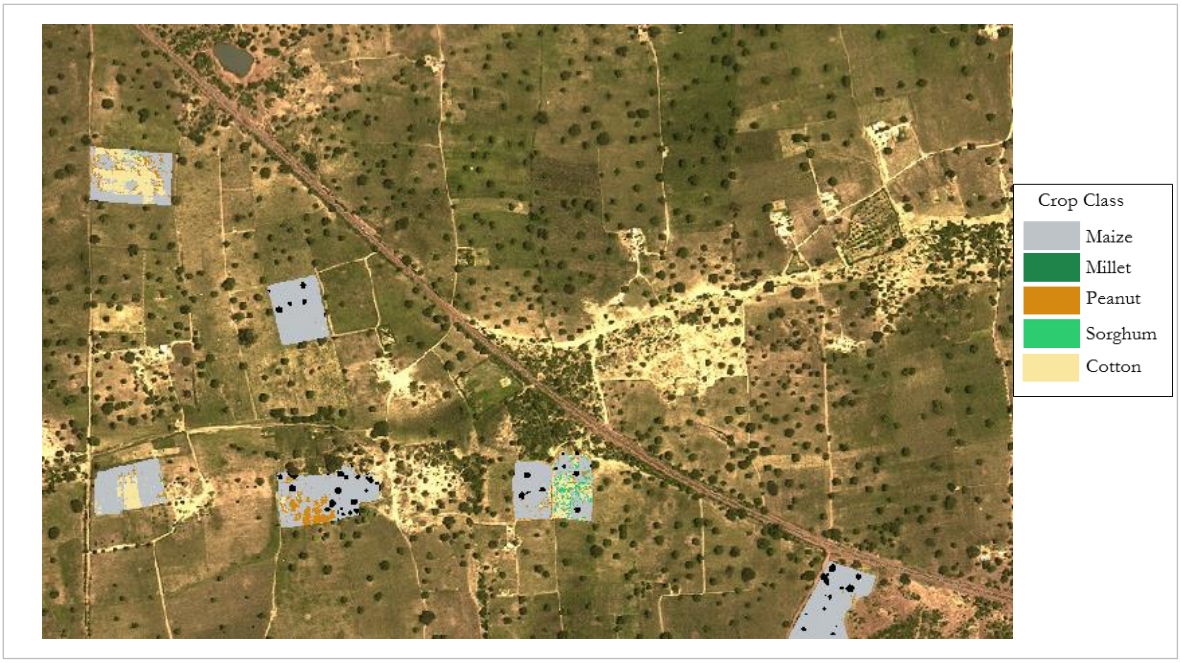


Figure 5-10. Subset of a multi-classified pan-sharpened dataset. A pan-sharpened image on October 18, 2014, in natural color composite, as background

6. CONCLUSIONS AND RECOMMENDATIONS

6.1. Conclusions

There is no doubt about the relevance and complexity of the problem addressed in this MSc thesis because smallholder farmers represent more than 80% of the area dedicated to agriculture activities in Sub-Saharan Africa. Reliable crop maps are crucial for an efficient agriculture because they provide relevant information about crop patterns that can be used to address properly current and future required resources, leading to better crop yields and consequently to an enhanced food security.

This research has demonstrated the potential of multi-classifiers to address complex problems. Indeed, the multi-classifier reached higher accuracy (0.534) than any of single classifiers. This represents an improvement up to 21.91% in comparison to the average kappa values (0.438) of base classifiers. Because of this, the conceptual design of the multi-classifier system, together with the associated cloud-based code, constitutes an important result of this research.

GEE is a powerful and dynamic platform that evolves constantly incorporating new functionalities. Although, this platform is optimized for image processing and table and vector operations are not straightforward, GEE has tremendous potential. to map crop types. This is because it can deal with time series of very high spatial resolution satellite images as well as with large amounts of data and with high dimensional problems.

During this research project, a considerable amount of data was processed requiring an important effort in terms of human resources, time and computation power. Working with such kind of data required organization, analytical and programming skills. Feature selection is still needed even though a robust cloud based platform is used. For example, a feature set of bands, vegetation indices, ratio and texture bands would sum more than nine hundred bands. A feature selection strategy discriminates the most informative features and help to minimize the number of experiments that is a time-consuming task.

The multi-classification was tested with three multi-temporal datasets that included image bands, ratio bands, vegetation indices and textural features. These features were extracted from the multi-spectral, pan-sharpened and aggregated images series. The best kappa value (0.534) was obtained with the pan-sharpened images when the output of base classifiers was combined using weighted majority voting.

Diversity in the multi-classifier was analysed through pair-wise and global metrics. Experimental results did not suggest a correlation between diversity and accuracy for this specific case study.

Fusion of images panchromatic and multi-spectral trough the aggregation of the panchromatic band to a coarser resolution increased the complexity of the problem but did not improve the multi-classifier performance. In general, the aggregated dataset achieved the poorest accuracy compared to the multi-spectral and pan-sharpened image series. This implies that adding the panchromatic band averaged to a coarser resolution did not improve the discrimination ability of a multi-spectral image. In contrast, pan-sharpened image series retained the accuracy of the multi-spectral dataset improving spatial detail.

Crop mapping in small farms with high resolution images is not a common research topic. Methods and techniques applied successfully to medium and lower spatial resolution images do not necessarily reach

success in such a complex environment. Given accuracies of base classifiers and the multi-classifier, crop classification in smallholder farms with mixing crops remains challenging.

6.2. Recommendations for future work

This research aim to provide a novel approach to a real and complex problem. Indeed, experiments carried out demonstrated the potential of the solution presented. Further work can extend this study and include other aspects not considered because time constraints. Some recommendations are given for future work:

- ✓ Apply an efficient feature selection strategy to filter the most discriminative features. This is especially relevant for textural features.
- ✓ Include textural features of vegetation indices in the feature sets.
- ✓ Limit the number of weights options using a global measure such as kappa and a class measure such as producer accuracy per class.
- ✓ Discriminate 'good' and 'bad' diversity in the multi-classifier.
- ✓ Increase the number of folds to ten. This means than more training features should be derived from the available ground data.
- ✓ Conduct more experiments to generalize conclusions.

LIST OF REFERENCES

Berger, A. L., Pietra, V. J. D., & Pietra, S. a. D. (1996). A maximum entropy approach to natural language processing. *Computational Linguistics*, *22*(1), 39–71. https://doi.org/10.3115/1075812.1075844

Beyer, F., Jarmer, T., Siegmann, B., & Fischer, P. (2015). Improved crop classification using multitemporal RapidEye data. In *2015 8th International Workshop on the Analysis of Multitemporal Remote Sensing Images, Multi-Temp 2015* (pp. 1–4). IEEE. https://doi.org/10.1109/Multi-Temp.2015.7245780

Bi, Y. (2012). The impact of diversity on the accuracy of evidential classifier ensembles. *International Journal of Approximate Reasoning*, *53*(4), 584–607. https://doi.org/10.1016/j.ijar.2011.12.011

Bouckaert, R. R. (2004). Naive Bayes Classifiers that perform well with continuous variables. *AI 2004: Advances in Artificial Intelligence*, 1089–1094. https://doi.org/10.1007/978-3-540-30549-1_106

Bouroubi, Y., Tremblay, N., Vigneault, P., & Benoit, M. (2014). Linear spectral unmixing for crop and soil information extraction from a single worldview-2 image. In *2014 IEEE Geoscience and Remote Sensing Symposium* (pp. 5103–5106). IEEE. https://doi.org/10.1109/IGARSS.2014.6947645

Breiman, L. (1996). Bagging predictors. *Machine Learning*, *24*(421), 123–140. https://doi.org/10.1007/BF00058655

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

Breiman, L., Friedman, J. H., Stone, C. J., & Richard A., O. (1985). Classification and regression trees: The Wadsworth Statistics/Probability Series, Wadsworth, Belmont, 1984, x + 358 pages. *European Journal of Operational Research*. North-Holland. https://doi.org/10.1016/0377-2217(85)90321-2

Brown, G., & Kuncheva, L. I. (2010). ``Good" and ``Bad" diversity in majority vote ensembles. In N. El Gayar, J. Kittler, & F. Roli (Eds.), *Multiple Classifier Systems: 9th International Workshop, MCS 2010, Cairo, Egypt, April 7-9, 2010. Proceedings* (pp. 124–133). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-12127-2_13

Camps-Valls, G., Gómez-Chova, L., Calpe-Maravilla, J., Martín-Guerrero, J. D., Soria-Olivas, E., Alonso-Chordá, L., & Moreno, J. (2004). Robust support vector method for hyperspectral data classification and knowledge discovery. *IEEE Transactions on Geoscience and Remote Sensing*, *42*(7), 1530–1542. https://doi.org/10.1109/TGRS.2004.827262

Castillejo-González, I. L., López-Granados, F., García-Ferrer, A., Peña-Barragán, J. M., Jurado-Expósito, M., de la Orden, M. S., & González-Audicana, M. (2009). Object- and pixel-based analysis for mapping crops and their agro-environmental associated measures using QuickBird imagery. *Computers and Electronics in Agriculture*, *68*(2), 207–215. https://doi.org/10.1016/j.compag.2009.06.004

Chellasamy, M., Ferre, T. P. A., & Greve, M. H. (2015). An ensemble-based training Data Refinement for Automatic Crop Discrimination Using WorldView-2 Imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *8*(10), 4882–4894. https://doi.org/10.1109/JSTARS.2015.2459754

Chen, Z., Li, S., Ren, J., Gong, P., Zhang, M., Wang, L., … Jiang, D. (2008). Monitoring and Management of Agriculture with Remote Sensing. In *Advances in Land Remote Sensing* (pp. 397–421). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6450-0_15

Chih-Wei Hsu, Chih-Chung Chang, and C.-J. L. (2008). A Practical Guide to Support Vector Classification. *BJU International*, *101*(1), 1396–400. https://doi.org/10.1177/02632760022050997

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, *20*(1), 37–46. https://doi.org/10.1177/001316446002000104

Conners, R. W., Trivedi, M. M., & Harlow, C. A. (1984). Segmentation of a high-resolution urban scene using texture operators. *Computer Vision, Graphics, and Image Processing*, *25*(3), 273–310. https://doi.org/10.1016/0734-189X(84)90197-X

Corrales, D. C., Figueroa, A., Ledezma, A., & Corrales, J. C. (2015). An Empirical Multi-classifier for Coffee Rust Detection in Colombian Crops. In O. Gervasi, B. Murgante, S. Misra, L. M. Gavrilova, C. A. M. A. Rocha, C. Torre, … O. B. Apduhan (Eds.), *Computational Science and Its Applications -- ICCSA 2015: 15th International Conference, Banff, AB, Canada, June 22-25, 2015, Proceedings, Part I* (pp. 60–74). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-319-21404-7_5

Cortes, C., & Vapnik, V. (1995). Support Vector Networks. *Machine Learning*, *20*(3), 273–297. https://doi.org/10.1007/BF00994018

Daumé, H. C. (2006). *Practical Structured Learning Techniques for Natural Language Processing*. UNIVERSITY OF SOUTHERN CALIFORNIA. Retrieved from http://www.umiacs.umd.edu/~hal/docs/daume06thesis.pdf

de la Fuente, D., Suarez, J., Yague, J., & Pedrazzani, D. (2013). Potentiality of World-View 2 data for precision agriculture. In *2013 IEEE International Geoscience and Remote Sensing Symposium - IGARSS* (pp. 2825–2828). IEEE. https://doi.org/10.1109/IGARSS.2013.6723412

Debats, S. R., Luo, D., Estes, L. D., Fuchs, T. J., & Caylor, K. K. (2016). A generalized computer vision approach to mapping crop fields in heterogeneous agricultural landscapes. *Remote Sensing of Environment*, *179*, 210–221. https://doi.org/10.1016/j.rse.2016.03.010

Delalieux, S., Zarco-Tejada, P. J., Tits, L., Bello, M. A. J., Intrigliolo, D. S., & Somers, B. (2014). Unmixing-based fusion of hyperspatial and hyperspectral airborne imagery for early detection of vegetation stress. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, *7*(6), 2571–2582. https://doi.org/10.1109/JSTARS.2014.2330352

Du, P., Xia, J., Zhang, W., Tan, K., Liu, Y., & Liu, S. (2012). Multiple classifier system for remote sensing image classification: a review. *Sensors (Basel)*, *12*(4), 4764–4792. https://doi.org/10.3390/s120404764

Evangelista, P. H., Stohlgren, T. J., Morisette, J. T., & Kumar, S. (2009). Mapping Invasive Tamarisk (Tamarix): A Comparison of Single-Scene and Time-Series Analyses of Remotely Sensed Data. *Remote Sensing*, *1*(3), 519–533. https://doi.org/10.3390/rs1030519

Foody, G. M., & Mathur, A. (2004). Toward intelligent training of supervised image classifications: directing training data acquisition for SVM classification. *Remote Sensing of Environment*, *93*(1), 107–117. https://doi.org/10.1016/j.rse.2004.06.017

Freund, Y., & Schapire, R. (1997). A desicion-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, *55*(1), 119–139. https://doi.org/doi:10.1006/jcss.1997.1504

Gershoff, M., & Schulenburg, S. (2007). Collective behavior based hierarchical XCS. *Proceedings of the 2007 GECCO Conference Companion on Genetic and Evolutionary Computation - GECCO '07*, 2695. https://doi.org/10.1145/1274000.1274064

Gilbertson, J. K., Kemp, J., & van Niekerk, A. (2017). Effect of pan-sharpening multi-temporal Landsat 8 imagery for crop type differentiation using different classification techniques. *Computers and Electronics in Agriculture*, *134*, 151–159. https://doi.org/10.1016/j.compag.2016.12.006

Gitelson, A. A., Kaufman, Y. J., Stark, R., & Rundquist, D. (2002). Novel algorithms for remote estimation of vegetation fraction. *Remote Sensing of Environment*, *80*(1), 76–87. https://doi.org/10.1016/S0034-4257(01)00289-9

Gómez-Chova, L., Tuia, D., Moser, G., & Camps-Valls, G. (2015). Multimodal Classification of Remote Sensing Images: A Review and Future Directions. *Proceedings of the IEEE*, *103*(9), 1560–1584. https://doi.org/10.1109/JPROC.2015.2449668

Gopinath, B., & Shanthi, N. (2014). Development of an Automated Medical Diagnosis System for Classifying Thyroid Tumor Cells using Multiple Classifier Fusion. *Technology in Cancer Research & Treatment*. https://doi.org/10.7785/tcrt.2012.500430

Gupta, P., Sharma, A., & Jindal, R. (2016, September). Scalable machine-learning algorithms for big data analytics: a comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*. Wiley Periodicals, Inc. https://doi.org/10.1002/widm.1194

Haboudane, D., Miller, J. R., Tremblay, N., Zarco-Tejada, P. J., & Dextraze, L. (2002). Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture. *Remote Sensing of Environment*, *81*(2), 416–426. https://doi.org/10.1016/S0034-4257(02)00018-4

Haghighat, M. B. A., Aghagolzadeh, A., & Seyedarabi, H. (2011). A non-reference image fusion metric based on mutual information of image features. *Computers & Electrical Engineering*, *37*(5), 744–756. https://doi.org/10.1016/j.compeleceng.2011.07.012

Hao, P., Wang, L., Niu, Z., Tilman, D., Balzer, C., Hill, J., … He, Y. (2015). Comparison of Hybrid Classifiers for Crop Classification Using Normalized Difference Vegetation Index Time Series: A Case Study for Major Crops in North Xinjiang, China. *PLOS ONE*, *10*(9), e0137748. https://doi.org/10.1371/journal.pone.0137748

Haralick, R. M. (1979). Statistical and structural approaches to texture. *Proceedings of the IEEE*, *67*(5), 786–

804. https://doi.org/10.1109/PROC.1979.11328
Haralick, R., Shanmugan, K., & Dinstein, I. (1973, November). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*. https://doi.org/10.1109/TSMC.1973.4309314
HU, Q., WU, W., SONG, Q., LU, M., CHEN, D., YU, Q., & TANG, H. (2017). How do temporal and spectral features matter in crop classification in Heilongjiang Province, China? *Journal of Integrative Agriculture*, *16*(2), 324–336. https://doi.org/10.1016/S2095-3119(15)61321-1
Huang, C., Davis, L. S., & Townshend, J. R. G. (2002). An assessment of support vector machines for land cover classii cation. *Int. J. Remote Sensing*, *23*(4), 725–749. https://doi.org/10.1080/01431160110040323
Huete, A., Didan, K., Miura, T., Rodriguez, E. P., Gao, X., & Ferreira, L. G. (2002). Overview of the radiometric and biophysical performance of the MODIS vegetation indices. *Remote Sensing of Environment*, *83*(1), 195–213. https://doi.org/10.1016/S0034-4257(02)00096-2
Huete, A. R. (1988). A Soil-Adjusted Vegetation Index (SAVI). *REMOTE SENSING OF ENVIRONMENT*, *25*, 295–309.
Inglada, J., Arias, M., Tardy, B., Hagolle, O., Valero, S., Morin, D., … Koetz, B. (2015). Assessment of an operational system for crop type map production using high temporal and spatial resolution satellite optical imagery. *Remote Sensing*, *7*(9), 12356–12379. https://doi.org/10.3390/rs70912356
Izquierdo-Verdiguier, E., Gómez-Chova, L., Camps-Valls, G., Izquierdo-Verdiguier, E., Gómez-Chova, L., & Camps-Valls, G. (2015). Kernels for Remote Sensing Image Classification. In *Wiley Encyclopedia of Electrical and Electronics Engineering* (pp. 1–23). Hoboken, NJ, USA: John Wiley & Sons, Inc. https://doi.org/10.1002/047134608X.W8252
Jackson, R. D., & Huete, A. R. (1991). Interpreting vegetation indices. *Preventive Veterinary Medicine*, *11*, 185–200.
Jordan, C. F. (1969). Derivation of Leaf-Area Index from Quality of Light on the Forest Floor. *Ecology*, *50*(4), 663–666. https://doi.org/10.2307/1936256
Khobragade, A. N., & Raghuwanshi, M. M. (2015). Contextual Soft Classification Approaches for Crops Identification Using Multi-sensory Remote Sensing Data: Machine Learning Perspective for Satellite Images (pp. 333–346). https://doi.org/10.1007/978-3-319-18476-0_33
Khobragade, A. N., Raghuwanshi, M. M., & Malik, L. (2016). Evaluating Kernel Effect on Performance of SVM Classification using Satellite Images. *International Journal of Scientific & Engineering Research*, *7*(3). Retrieved from http://www.ijser.org
Ko, A. H. R., Sabourin, R., & Britto, A. S. (2008). From dynamic classifier selection to dynamic ensemble selection. *Pattern Recognition*, *41*(5), 1735–1748. https://doi.org/10.1016/j.patcog.2007.10.015
Kohavi, R. (1995). A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection. *International Joint Conference on Artificial Intelligence*, *14*(12), 1137–1143. https://doi.org/10.1067/mod.2000.109031
Koppel, M. (2002). Automatically Categorizing Written Texts by Author Gender. *Literary and Linguistic Computing*, *17*(4), 401–412. https://doi.org/10.1093/llc/17.4.401
Kuncheva, L. I. (2004). *Base Classifiers*. *Combining Pattern Classifiers: Methods and Algorithms*. https://doi.org/10.1109/TNN.2007.897478
Kuncheva, L. I., & Whitaker, C. J. (2003). Measures of diversity in classifier ensembles and their relationship with the ensemble accuracy. *Machine Learning*, *51*(2), 181–207. https://doi.org/10.1023/A:1022859003006
Li, D., Yang, F., & Wang, X. (2016). Study on Ensemble Crop Information Extraction of Remote Sensing Images Based on SVM and BPNN. *Journal of the Indian Society of Remote Sensing*, 1–9. https://doi.org/10.1007/s12524-016-0597-y
Lijun, D., & Chuang, L. (2011). Research on remote sensing image of land cover classification based on multiple classifier combination. *Wuhan University Journal of Natural Sciences*, *16*(4), 363–368. https://doi.org/10.1007/s11859-011-0764-5
Louhaichi, M., Borman, M. M., & Johnson, D. E. (2001). Spatially Located Platform and Aerial Photography for Documentation of Grazing Impacts on Wheat. *Geocarto International*, *16*(1), 65–70. https://doi.org/10.1080/10106040108542184
Löw, F., Conrad, C., & Michel, U. (2015). Decision fusion and non-parametric classifiers for land use mapping using multi-temporal RapidEye data. *ISPRS Journal of Photogrammetry and Remote Sensing*, *108*, 191–204. https://doi.org/10.1016/j.isprsjprs.2015.07.001
Lu, D., & Weng, Q. (2007). A survey of image classification methods and techniques for improving

classification performance. *International Journal of Remote Sensing*, *28*(5), 823–870. https://doi.org/10.1080/01431160600746456
Lumini, A., Nanni, L., & Brahnam, S. (2016). Ensemble of texture descriptors and classifiers for face recognition. *Applied Computing and Informatics*. https://doi.org/10.1016/j.aci.2016.04.001
Maji, S., Berg, A. C., & Maliks, J. (2008). Classification using intersection kernel support vector machines is efficient. In *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR* (pp. 1–8). IEEE. https://doi.org/10.1109/CVPR.2008.4587630
Mann, G., Mcdonald, R., & Silberman, N. (2009). Efficient Large-Scale Distributed Training of Conditional Maximum Entropy Models. In *Advances in Neural Information Processing Systems* (pp. 1–9).
Misra, G., Kumar, A., Patel, N. R., & Zurita-Milla, R. (2014). Mapping a Specific Crop-A Temporal Approach for Sugarcane Ratoon. *Journal of the Indian Society of Remote Sensing*, *42*(2), 325–334. https://doi.org/10.1007/s12524-012-0252-1
Mountrakis, G., Im, J., & Ogole, C. (2011). ISPRS Journal of Photogrammetry and Remote Sensing Support vector machines in remote sensing : A review. *ISPRS Journal of Photogrammetry and Remote Sensing*, *66*(3), 247–259. https://doi.org/10.1016/j.isprsjprs.2010.11.001
Mousavi, R., & Eftekhari, M. (2015). A new ensemble learning methodology based on hybridization of classifier ensemble selection approaches. *Applied Soft Computing*, *37*, 652–666. https://doi.org/10.1016/j.asoc.2015.09.009
Nitze, I., Schulthess, U., & Asche, H. (2012). Comparison of Machine Learning Algorithms Random Forest, Artificial Neural Network and Support Vector Machine to Maximum Likelihood for Supervised Crop Type Classification. In P. S. Radek Silhavy, Roman Senkerik, Zuzana Kominkova Oplatkova, Zdenka Prokopova (Ed.), *Fourth international conference on Geographic Object-Based Image Analysis (GEOBIA) Conference* (pp. 35–40). Rio de Janeiro.
Olofsson, P., Foody, G. M., Herold, M., Stehman, S. V., Woodcock, C. E., & Wulder, M. A. (2014). Good practices for estimating area and assessing accuracy of land change. *Remote Sensing of Environment*, *148*, 42–57. https://doi.org/10.1016/j.rse.2014.02.015
Padwick, C., Scientist, P., Deskevich, M., Pacifici, F., & Smallwood, S. (2010). WorldView-2 pan-sharpening. *Asprs 2010*, *48*(1), 26–30. Retrieved from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:WorldView-2+Pan-Sharpening#0
Patil, V. C., Al-Gaadi, K. A., Madugundu, R., Tola, E., Zeyada, A. M., Marey, S., & Biradar, C. M. (2016). CART and IDC – based classification of irrigated agricultural fields using multi-source satellite data. *Geocarto International*, *0*(0), 1–19. https://doi.org/10.1080/10106049.2016.1232312
Peña-Barragán, J. M., Ngugi, M. K., Plant, R. E., & Six, J. (2011). Object-based crop identification using multiple vegetation indices, textural features and crop phenology. *Remote Sensing of Environment*, *115*(6), 1301–1316. https://doi.org/10.1016/j.rse.2011.01.009
Phillips, S. J., Anderson, R. P., & Schapire, R. E. (2006). Maximum entropy modeling of species geographic distributions. *Ecological Modelling*, *190*(3), 231–259. https://doi.org/10.1016/j.ecolmodel.2005.03.026
Pohl, C. (1998). Multisensor image fusion in remote sensing : Concepts , methods and applications, (August 2016).
Polikar, R. (2006). Ensemble based systems in decision making. *Circuits and Systems Magazine, IEEE*, *6*(3), 21–45. https://doi.org/10.1109/MCAS.2006.1688199
Powers, D. (2011). Evaluation: From Precision, Recall and F-Measure To Roc, Informedness, Markedness & Correlation. *Journal of Machine Learning Technologies*, *2*(1), 37–63. https://doi.org/10.1.1.214.9232
Qi, J., Chehbouni, A., Huete, A. R., Kerr, Y. H., & Sorooshian, S. (1994). A modified soil adjusted vegetation index. *Remote Sensing of Environment*, *48*(2), 119–126. https://doi.org/10.1016/0034-4257(94)90134-1
Ranawana, R., & Palade, V. (2006). Multi-Classifier Systems: Review and a roadmap for developers. *International Journal of Hybrid Intelligent Systems*, *3*(1), 35–61. https://doi.org/10.3233/HIS-2006-3104
Rao, P. V. N., Sai, M. V. R. S., Sreenivas, K., Rao, M. V. K., Rao, B. R. M., Dwivedi, R. S., & Venkataratnam, L. (2002). Textural analysis of IRS-1D panchromatic data for land cover classification. *International Journal of Remote Sensing*, *23*(17), 3327–3345. https://doi.org/10.1080/01431160110104665
Richards, J. (2013). *Remote Sensing Digital Image Analysis An Introduction*. *Methods* (fifth). Berlin: Springer-Verlag Berlin Heidelberg. https://doi.org/10.1007/978-3-642-30062-2

Rish, I. (2001). An empirical study of the naive Bayes classifier. *In Proceedings of IJCAI- 01 Workshop on Empirical Methods in AI, International Joint Conference on Artifi- Cial Intelligence*, *335*, 41–46. Retrieved from http://www.cc.gatech.edu/~isbell/%0Aclasses/reading/papers/Rish.pdf

Rouse, W., Haas, R. H., & Deering, D. W. (1973). MONITORING VEGETATION SYSTEMS IN THE GREAT PLAINS WITH ERTS. *Proceedings of the Earth Resources Technology Satellite Symposium NASA*, *1*, 309–317.

Shaban, M. A., & Dikshit, O. (2001). Improvement of classification in urban areas by the use of textural features: The case study of Lucknow city, Uttar Pradesh. *International Journal of Remote Sensing*, *22*(4), 565–593. https://doi.org/10.1080/01431160050505865

Shalev-Shwartz, S., Singer, Y., Srebro, N., & Cotter, A. (2011). Pegasos: Primal estimated sub-gradient solver for SVM. *Mathematical Programming*, *127*(1), 3–30. https://doi.org/10.1007/s10107-010-0420-4

Shukla, G., Garg, R. D., Srivastava, H. S., & Garg, P. K. (2016). Performance analysis of different predictive models for crop classification across an aridic to ustic area of Indian states. *Geocarto International*, *0*(0), 1–20. https://doi.org/10.1080/10106049.2016.1240721

Stratoulias, D., By, R. A. De, Zurita-milla, R., Bijker, W., Tolpekin, V., & Schulthess, U. (2015). THE POTENTIAL OF VERY HIGH SPATIAL RESOLUTION REMOTE SENSING IN APPLICATIONS IN SMALLHOLDER AGRICULTURE. *Proceedings of ACRS 2015 : The 36th Asian Conference on Remote Sensing : Fostering Resilient Growth in Asia*, 1–7. Retrieved from https://www.itc.nl/resumes/Zurita-Milla#conference

Tang, E. K., Suganthan, P. N., Yao, X., & Fawcett K Tang P N Suganthan, T. E. (2006). An analysis of diversity measures. *Mach Learn*, *65*, 247–271. https://doi.org/10.1007/s10994-006-9449-2

Tarantino, C., Adamo, M., Pasquariello, G., Lovergine, F., Blonda, P., & Tomaselli, V. (2012). 8-Band Image Data Processing of the Worldview-2 Satellite in a Wide Area of Applications. *Earth Observation*, 137–152. https://doi.org/10.5772/27499

Tatsumi, K., Yamashiki, Y., Canales Torres, M. A., & Taipe, C. L. R. (2015). Crop classification of upland fields using Random forest of time-series Landsat 7 ETM+ data. *Computers and Electronics in Agriculture*, *115*, 171–179. https://doi.org/10.1016/j.compag.2015.05.001

Tucker, C. J. (1979). Red and photographic infrared linear combinations for monitoring vegetation. *Remote Sensing of Environment*, *8*(2), 127–150. https://doi.org/10.1016/0034-4257(79)90013-0

Wacker, A. G., & Landgrebe, D. A. (1972). Minimum Distance Classification in Remote Sensing Minimum Distance Classification in Remote Sensing The Laboratory for Applications of Remote Sensing MINIMUM DISTANCE CLASSIFICATION IN REMOTE SENSING*. Retrieved from http://docs.lib.purdue.edu/larstech

Wang, L., Sousa, W. P., Gong, P., & Biging, G. S. (2004). Comparison of IKONOS and QuickBird images for mapping mangrove species on the Caribbean coast of Panama. *Remote Sensing of Environment*, *91*(3), 432–440. https://doi.org/10.1016/j.rse.2004.04.005

Wolpert, D. H., & Macready, W. G. (1997). No free lunch theorems for optimization. *IEEE Transactions on Evolutionary Computation*, *1*(1), 67–82. https://doi.org/10.1109/4235.585893

Woźniak, M., & Graña, M. (2014). A survey of multiple classifier systems as hybrid systems. *Information Fusion*, *16*, 3–17. https://doi.org/10.1016/j.inffus.2013.04.006

Zhang, J. (2010). Multi-source remote sensing data fusion: status and trends. *International Journal of Image and Data Fusion*, *1*(1), 5–24. https://doi.org/10.1080/19479830903561035

Zurita Milla, R., Kaiser, G., Clevers, J. G. P. W., Schneider, W., Schaepman, M. E., Zurita-Milla, R., … Schaepman, M. E. (2008). Monitoring vegetation dynamics using MERIS fused images. In H. Lacoste & L. Ouwehand (Eds.), *Proceedings of the 2nd MERIS / (A)ATSR User Workshop, Frascati, Italy, 22 - 26 August, 2008* (Vol. SP-666, p. 3). 79, , : ESA. Retrieved from http://edepot.wur.nl/398

A. APPENDIX A

Table A-1. Kappa values for the multi-classifier built with the original multi-spectral images.

Training	Classifiers	Features	Voting	Weighted Voting					
			Kappa	Mean Kappa	SD	F-score	F-score per Class	Kappa	Producer Accuracy
K-fold	5	Bands	0.4733	0.4789	0.0000	0.4789	0.4789	0.4789	0.4789
		Bands + Vis	0.4744	0.4811	0.0000	0.4811	0.4811	0.4811	0.4811
		Bands + VIs + Ratio Bands	**0.5044**	**0.5200**	**0.0000**	**0.5200**	**0.5200**	**0.5200**	**0.5200**
		Bands + Textures	0.4300	0.4433	0.0000	0.4433	0.4433	0.4433	0.4433
		Bands + Textures + Vis	0.4156	0.4389	0.0000	0.4389	0.4389	0.4389	0.4389
	15	Bands	0.4800	0.4842	0.0010	0.4833	0.4844	0.4844	0.4844
		Bands + Vis	0.5067	0.5086	0.0010	0.5078	0.5089	0.5089	0.5089
		Bands + VIs + Ratio Bands	**0.5078**	**0.5303**	**0.0029**	**0.5278**	**0.5311**	**0.5311**	**0.5311**
		Bands + Textures	0.4156	0.4322	0.0038	0.4289	0.4333	0.4333	0.4333
		Bands + Textures + VIs	0.4456	0.4611	0.0000	0.4611	0.4611	0.4611	0.4611
	25	Bands	0.4900	0.5031	0.0010	0.5022	0.5033	0.5033	0.5033
		Bands + VIs	**0.5067**	**0.5192**	**0.0029**	**0.5167**	**0.5200**	**0.5200**	**0.5200**
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4144	0.4356	0.0077	0.4289	0.4378	0.4378	0.4378
		Bands + Textures + VIs	0.4589	0.4689	0.0000	0.4689	0.4689	0.4689	0.4689
Bagging	5	Bands	0.4744	0.4844	0.0000	0.4844	0.4844	0.4844	0.4844
		Bands + VIs	0.4589	0.4622	0.0000	0.4622	0.4622	0.4622	0.4622
		Bands + VIs + Ratio Bands	**0.4867**	**0.4933**	**0.0000**	**0.4933**	**0.4933**	**0.4933**	**0.4933**
		Bands + Textures	0.4100	0.4200	0.0000	0.4200	0.4200	0.4200	0.4200
		Bands + Textures + VIs	0.4233	0.4389	0.0000	0.4389	0.4389	0.4389	0.4389
	15	Bands	0.4911	0.4864	0.0010	0.4856	0.4867	0.4867	0.4867
		Bands + VIs	0.4800	0.4914	0.0029	0.4889	0.4922	0.4922	0.4922
		Bands + VIs + Ratio Bands	**0.5200**	**0.5308**	**0.0010**	**0.5300**	**0.5311**	**0.5311**	**0.5311**
		Bands + Textures	0.4122	0.4336	0.0029	0.4311	0.4344	0.4344	0.4344
		Bands + Textures + VIs	0.4311	0.4375	0.0010	0.4367	0.4378	0.4378	0.4378
	25	Bands	**0.4900**	**0.5003**	**0.0010**	**0.5011**	**0.5000**	**0.5000**	**0.5000**
		Bands + VIs	0.4711	0.4872	0.0058	0.4822	0.4889	0.4889	0.4889
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4167	0.4322	0.0038	0.4289	0.4333	0.4333	0.4333
		Bands + Textures + VIs	0.4522	0.4611	0.0038	0.4644	0.4600	0.4600	0.4600

Table A-2. Kappa values for the multi-classifier built with the pan-sharpened images.

Training	Classifiers	Features	Voting	Weighted Voting					
			Kappa	Mean kappa	SD	F-score	F-score per Class	Kappa	Producer Accuracy
K-fold	5	Bands	0.4722	0.4789	0.0000	0.4789	0.4789	0.4789	0.4789
		Bands + VIs	0.4678	0.4789	0.0000	0.4789	0.4789	0.4789	0.4789
		Bands + VIs + Ratio Bands	**0.5011**	**0.5067**	**0.0000**	**0.5067**	**0.5067**	**0.5067**	**0.5067**
		Bands + Textures	0.4422	0.4489	0.0000	0.4489	0.4489	0.4489	0.4489
		Bands + Textures + VIs	0.4578	0.4656	0.0000	0.4656	0.4656	0.4656	0.4656
	15	Bands	0.4944	0.4972	0.0029	0.5022	0.4956	0.4956	0.4956
		Bands + VIs	0.4811	0.4914	0.0005	0.4922	0.4911	0.4911	0.4911
		Bands + VIs + Ratio Bands	**0.5256**	**0.5333**	**0.0000**	**0.5333**	**0.5333**	**0.5333**	**0.5333**
		Bands + Textures	0.4533	0.4525	0.0014	0.4500	0.4533	0.4533	0.4533
		Bands + Textures + VIs	0.4544	0.4614	0.0005	0.4622	0.4611	0.4611	0.4611
	25	Bands	**0.5000**	**0.4994**	**0.0010**	**0.4978**	**0.5000**	**0.5000**	**0.5000**
		Bands + VIs	0.4933	0.4972	0.0010	0.4956	0.4978	0.4978	0.4978
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4644	0.4706	0.0010	0.4689	0.4711	0.4711	0.4711
		Bands + Textures + VIs	0.4800	0.4819	0.0005	0.4811	0.4822	0.4822	0.4822
Bagging	5	Bands	0.4789	0.4844	0.0000	0.4844	0.4844	0.4844	0.4844
		Bands + VIs	0.4600	0.4733	0.0000	0.4733	0.4733	0.4733	0.4733
		Bands + VIs + Ratio Bands	**0.5122**	**0.5156**	**0.0000**	**0.5156**	**0.5156**	**0.5156**	**0.5156**
		Bands + Textures	0.4311	0.4444	0.0000	0.4444	0.4444	0.4444	0.4444
		Bands + Textures + VIs	0.4600	0.4556	0.0000	0.4556	0.4556	0.4556	0.4556
	15	Bands	0.4933	0.4933	0.0019	0.4967	0.4922	0.4922	0.4922
		Bands + VIs	0.4778	0.4878	0.0000	0.4878	0.4878	0.4878	0.4878
		Bands + VIs + Ratio Bands	**0.5167**	**0.5344**	**0.0000**	**0.5344**	**0.5344**	**0.5344**	**0.5344**
		Bands + Textures	0.4422	0.4414	0.0005	0.4422	0.4411	0.4411	0.4411
		Bands + Textures + VIs	0.4567	0.4603	0.0024	0.4644	0.4589	0.4589	0.4589
	25	Bands	**0.4956**	**0.5147**	**0.0034**	**0.5089**	**0.5167**	**0.5167**	**0.5167**
		Bands + VIs	0.4878	0.4981	0.0014	0.4956	0.4989	0.4989	0.4989
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4444	0.4517	0.0010	0.4500	0.4522	0.4522	0.4522
		Bands + Textures + VIs	0.4789	0.4786	0.0005	0.4778	0.4789	0.4789	0.4789

Table A-3. Kappa values for the multi-classifier built with the aggregated images

Training	Classifiers	Features	Voting	Weighted Voting					
			Kappa	Mean kappa	SD	F-score	F-score per Class	kappa	Producer Accuracy
K-fold	5	Bands	0.4678	0.4822	0.0000	0.4822	0.4822	0.4822	0.4822
		Bands + VIs	**0.4900**	**0.4956**	**0.0000**	**0.4956**	**0.4956**	**0.4956**	**0.4956**
		Bands + VIs + Ratio Bands	0.4822	0.4844	0.0000	0.4844	0.4844	0.4844	0.4844
		Bands + Textures	0.4389	0.4422	0.0000	0.4422	0.4422	0.4422	0.4422
		Bands + Textures + VIs	0.4633	0.4667	0.0000	0.4667	0.4667	0.4667	0.4667
	15	Bands	0.4733	0.4819	0.0005	0.4811	0.4822	0.4822	0.4822
		Bands + VIs	0.5000	0.5144	0.0019	0.5111	0.5156	0.5156	0.5156
		Bands + VIs + Ratio Bands	**0.5144**	**0.5206**	**0.0010**	**0.5222**	**0.5200**	**0.5200**	**0.5200**
		Bands + Textures	0.4600	0.4667	0.0000	0.4667	0.4667	0.4667	0.4667
		Bands + Textures + VIs	0.4622	0.4700	0.0000	0.4700	0.4700	0.4700	0.4700
	25	Bands	0.4789	0.4869	0.0014	0.4844	0.4878	0.4878	0.4878
		Bands + VIs	**0.4956**	**0.5069**	**0.0014**	**0.5044**	**0.5078**	**0.5078**	**0.5078**
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4411	0.4481	0.0005	0.4489	0.4478	0.4478	0.4478
		Bands + Textures + VIs	0.4544	0.4744	0.0000	0.4744	0.4744	0.4744	0.4744
Bagging	5	Bands	0.4433	0.4489	0.0000	0.4489	0.4489	0.4489	0.4489
		Bands + VIs	0.4667	0.4778	0.0000	0.4778	0.4778	0.4778	0.4778
		Bands + VIs + Ratio Bands	**0.4833**	**0.4911**	**0.0000**	**0.4911**	**0.4911**	**0.4911**	**0.4911**
		Bands + Textures	0.3978	0.4022	0.0000	0.4022	0.4022	0.4022	0.4022
		Bands + Textures + VIs	0.4367	0.4578	0.0000	0.4578	0.4578	0.4578	0.4578
	15	Bands	0.4867	0.4883	0.0010	0.4867	0.4889	0.4889	0.4889
		Bands + VIs	0.4944	0.4944	0.0000	0.4944	0.4944	0.4944	0.4944
		Bands + VIs + Ratio Bands	**0.5011**	**0.5100**	**0.0000**	**0.5100**	**0.5100**	**0.5100**	**0.5100**
		Bands + Textures	0.4078	0.4111	0.0000	0.4111	0.4111	0.4111	0.4111
		Bands + Textures + VIs	0.4411	0.4572	0.0010	0.4556	0.4578	0.4578	0.4578
	25	Bands	0.4656	0.4803	0.0034	0.4744	0.4822	0.4822	0.4822
		Bands + VIs	**0.4967**	**0.5086**	**0.0024**	**0.5044**	**0.5100**	**0.5100**	**0.5100**
		Bands + VIs + Ratio Bands	NA	NA	NA	NA	NA	NA	NA
		Bands + Textures	0.4522	0.4544	0.0000	0.4544	0.4544	0.4544	0.4544
		Bands + Textures + VIs	0.4411	0.4481	0.0014	0.4456	0.4489	0.4489	0.4489

Table A-4. Diversity measures for multi-spectral image series

Training	Classifiers	Features	Diversity Measure			
			Q	Correlation	Interrater Agreement	Entropy
K-fold	5	Bands	0.6908	0.4285	0.4252	0.4258
		Bands + VIs	0.7142	0.4372	0.4343	0.4178
		Bands + VIs + Ratio Bands	0.6659	0.3997	0.3974	0.4413
		Bands + Textures	0.6301	0.3811	0.3805	0.4591
		Bands + Textures + VIs	**0.6099**	**0.3644**	**0.3637**	**0.4716**
	15	Bands	0.7077	0.4476	0.4439	0.4033
		Bands + VIs	0.7040	0.4341	0.4322	0.4127
		Bands + VIs + Ratio Bands	0.6619	0.3982	0.3954	0.4409
		Bands + Textures	**0.6049**	**0.3608**	**0.3612**	**0.4684**
		Bands + Textures + VIs	0.6138	0.3657	0.3655	0.4673
	25	Bands	0.7225	0.4655	0.4627	0.3891
		Bands + VIs	0.7222	0.4554	0.4530	0.3969
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	**0.6266**	**0.3836**	**0.3829**	**0.4469**
		Bands + Textures + VIs	0.6329	0.3902	0.3895	0.4439
Bagging	5	Bands	0.7105	0.4412	0.4381	0.4156
		Bands + VIs	0.6875	0.4160	0.4112	0.4351
		Bands + VIs + Ratio Bands	0.6469	0.3819	0.3800	0.4564
		Bands + Textures	0.6005	0.3569	0.3571	0.4756
		Bands + Textures + VIs	**0.5917**	**0.3554**	**0.3530**	**0.4813**
	15	Bands	0.7050	0.4533	0.4493	0.3981
		Bands + VIs	0.7209	0.4560	0.4530	0.3947
		Bands + VIs + Ratio Bands	0.6762	0.4166	0.4156	0.4225
		Bands + Textures	0.6121	0.3707	0.3705	0.4598
		Bands + Textures + VIs	**0.6082**	**0.3704**	**0.3698**	**0.4640**
	25	Bands	0.7030	0.4504	0.4478	0.4011
		Bands + VIs	0.7088	0.4443	0.4426	0.4013
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	0.6294	0.3865	0.3855	0.4445
		Bands + Textures + VIs	**0.6214**	**0.3820**	**0.3816**	**0.4484**

Table A-5. Diversity measures for pan-sharpened images

Training	Classifiers	Features	Diversity Measure			
			Q	Correlation	Interrater Agreement	Entropy
K-fold	5	Bands	0.6408	0.3733	**0.3661**	**0.4764**
		Bands + VIs	0.6673	0.3975	0.3900	0.4524
		Bands + VIs + Ratio Bands	0.6983	0.4213	0.4151	0.4280
		Bands + Textures	0.6531	0.3887	0.3859	0.4591
		Bands + Textures + VIs	**0.6276**	**0.3718**	0.3692	0.4684
	15	Bands	0.6352	0.3817	0.3771	0.4590
		Bands + VIs	0.6990	0.4317	0.4262	0.4169
		Bands + VIs + Ratio Bands	0.6890	0.4194	0.4164	0.4234
		Bands + Textures	0.6240	0.3676	0.3654	0.4662
		Bands + Textures + VIs	**0.6054**	**0.3537**	**0.3521**	**0.4785**
	25	Bands	0.6771	0.4150	0.4090	0.4385
		Bands + VIs	0.7120	0.4497	0.4454	0.4000
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	0.6548	0.3967	0.3940	0.4460
		Bands + Textures + VIs	**0.6413**	**0.3845**	**0.3822**	**0.4537**
Bagging	5	Bands	**0.6288**	**0.3660**	**0.3573**	**0.4853**
		Bands + VIs	0.6596	0.3993	0.3934	0.4493
		Bands + VIs + Ratio Bands	0.6828	0.4091	0.4073	0.4369
		Bands + Textures	0.6388	0.3795	0.3756	0.4613
		Bands + Textures + VIs	0.6608	0.3982	0.3950	0.4507
	15	Bands	0.6649	0.4053	0.3984	0.4500
		Bands + VIs	0.7029	0.4419	0.4374	0.4110
		Bands + VIs + Ratio Bands	0.7007	0.4358	0.4311	0.4131
		Bands + Textures	0.6474	0.3881	0.3850	0.4497
		Bands + Textures + VIs	**0.6275**	**0.3761**	**0.3739**	**0.4573**
	25	Bands	0.6657	0.4076	0.4017	0.4427
		Bands + VIs	0.7158	0.4509	0.4458	0.3999
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	0.6536	0.3959	0.3938	0.4429
		Bands + Textures + VIs	**0.6495**	**0.3911**	**0.3893**	**0.4479**

Table A-6. Diversity measures for aggregated images

Training	Classifiers	Features	Diversity Measure			
			Q	Correlation	Interrater Agreement	Entropy
K-fold	5	Bands	0.7049	0.4306	0.4275	0.4213
		Bands + VIs	0.6578	0.3895	0.3862	0.4542
		Bands + VIs + Ratio Bands	0.6763	0.4058	0.4001	0.4449
		Bands + Textures	**0.5672**	**0.3509**	**0.3502**	**0.4840**
		Bands + Textures + VIs	0.6468	0.3970	0.3974	0.4471
	15	Bands	0.6837	0.4227	0.4200	0.4211
		Bands + VIs	0.6477	0.3893	0.3856	0.4546
		Bands + VIs + Ratio Bands	0.6792	0.4112	0.4071	0.4335
		Bands + Textures	**0.5874**	**0.3510**	**0.3499**	**0.4839**
		Bands + Textures + VIs	0.6219	0.3717	0.3710	0.4679
	25	Bands	0.7107	0.4503	0.4472	0.4007
		Bands + VIs	0.6879	0.4269	0.4243	0.4227
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	**0.6341**	**0.3851**	**0.3846**	**0.4559**
		Bands + Textures + VIs	0.6556	0.4053	0.4049	0.4335
Bagging	5	Bands	0.6516	0.3973	0.3923	0.4436
		Bands + VIs	0.6398	0.3780	**0.3733**	**0.4640**
		Bands + VIs + Ratio Bands	0.6741	0.4043	0.3965	0.4453
		Bands + Textures	**0.6243**	0.3862	0.3866	0.4578
		Bands + Textures + VIs	0.6249	**0.3762**	0.3760	0.4640
	15	Bands	0.6833	0.4263	0.4233	0.4253
		Bands + VIs	0.6825	0.4195	0.4175	0.4263
		Bands + VIs + Ratio Bands	0.6829	0.4208	0.4174	0.4246
		Bands + Textures	**0.6504**	**0.3986**	**0.3988**	**0.4401**
		Bands + Textures + VIs	0.6538	0.4072	0.4069	0.4316
	25	Bands	0.7064	0.4471	0.4439	0.4061
		Bands + VIs	0.6839	0.4254	0.4225	0.4217
		Bands + VIs + Ratio Bands	NA	NA	NA	NA
		Bands + Textures	**0.6172**	**0.3755**	**0.3750**	**0.4588**
		Bands + Textures + VIs	0.6520	0.4022	0.4018	0.4335