Clarisse Rita Afonso Feio

Licenciada em Ciências da Engenharia
Eletrotécnica e de Computadores

Land Cover mapping based on Hierarchical Decision Trees

Dissertação para obtenção do Grau de Mestre em

Engenharia Eletrotécnica e de Computadores

Orientador: Professor Doutor José Manuel Matos Ribeiro da Fonseca, Professor Associado com Agregação, Universidade NOVA de Lisboa

Júri

Presidente: Professor Doutor Fernando José Almeida Vieira do Coito, FCT-UNL
Arguente: Professor Doutor André Teixeira Bento Damas Mora, FCT-UNL
Vogal: Professor Doutor José Manuel Matos Ribeiro da Fonseca, FCT-UNL

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Julho, 2021

Land Cover mapping based on Hierarchical Decision Trees

Este documento foi gerado utilizando o processador (pdf)LaTeX, com base no template "novathesis" [1] desenvolvido no Dep. Informática da FCT-NOVA [2].
[1] https://github.com/joaomlourenco/novathesis [2] http://www.di.fct.unl.pt

Acknowledgements

Firstly I would like to sincerely thank my advisor, Professor José Fonseca, for being always available and supportive when any problem arose providing me with the guidance needed to improve my work and succeed. I would also like to thank Leonardo Martins and Carlos Garcia for helping me a lot throughout the development of this thesis.

I would also like to thank Faculdade de Ciências e Tecnologia da Universidade Nova de Lisboa (FCT - UNL) for providing me five years of resources and knowledge that were crucial for both the development of the present dissertation and for my personal and academic growth.

I find it essential to also mention professors that stood out and positively impacted me throughout this degree, due to their teaching skills paired with a great relationship with their students, which is crucial for the successful learning of any subject along with cultivating interest and desire to learn more about the subject besides what will be evaluated in an exam. Professor José Fonseca, Professor André Mora, Professor Mário Ventim Neves, Professor Luís Bernardo, Professor Paulo Pinto, I have a lot of respect and look up to you.

To my colleagues, with whom I shared this journey and many stressful days when approaching deadlines, always followed by a lot more moments of success, joy and pride in our accomplished work. Francisco, Lucas, Diogo and David working with all of you was a great pleasure and only motivated me more to better myself, but even greater was the time spent just goofing around, chatting, going for the usual walks around the campus and going for ice cream after exams.

To my family, without whom I would not be where I am today. They always provided me with the means and support so needed to finish this degree, being this thesis the proof of that. A special thanks to my mother, who had the patience and dedication to proofread my entire dissertation.

And I could not forget my friends André, Tiago and Rúben, that although were not as directly involved in the development of this thesis, gave me a lot of good laughs and moments to relax which allowed me to momentarily forget any problems I could be facing, allowing me to tackle them later with a fresh mind.

To all of you, my sincere gratitude.

Abstract

The ability to monitor land cover changes can be very useful for resource management, urban planning, forest fire identification, among plenty of other applications. The topic of remote sensing has been studied for a long time, with many different solutions that typically use satellites or aircraft to obtain multi-spectral imagery and further analyse it.

The entity responsible for monitoring land use and land cover in Portugal is Direção-Geral do Território (DGT) which periodically produces a document called Land Use and Land Cover Map (Carta de Uso e Ocupação do Solo (COS), in Portuguese). This document uses imagery with high spatial resolution of 0,25 m and has a minimum mapping unit of 1 ha, however, it is only produced every few years because it is manually curated by experts. This hinders the ability to closely monitor relevant land changes that occur more frequently or rapidly.

In this dissertation, several classifiers were developed in a hierarchical manner to address some of COS drawbacks. The classifiers used were based on decision trees which were trained using satellite imagery collected from Sentinel-2 satellite constellation. Although having a lower spatial resolution than COS, they can automatically classify land cover in some minutes every time a new set of Sentinel-2 imagery is collected, in this case each 5 days. Cloud coverage might make some of these images unusable but nonetheless, the temporal resolution is still far greater than COS.

However, automatic classification is not as accurate as manual classification. The produced classifiers did not consider as many classes as COS and had problems distinguishing some types of land cover, due to either poor sample size or spectral signature similarity. Considering Matthews Correlation Coefficient (MCC), water class had the best performance with an average of 91,28%, followed by forest and agriculture class with an average of 47,88% and 42,34%, respectively, and lastly urban areas and bare land class had the worse results averaging 28,03% and 20,53% respectively. Nevertheless, the results obtained were still considered to be good, but with considerable room for improvement.

Keywords: remote sensing, land cover classification, Sentinel-2, decision trees.

Resumo

Acompanhar as mudanças de ocupação de solo tem bastante utilidade para uma correta gestão de recursos, deteção de fogos florestais, e inúmeras aplicações. O tema de deteção remota é estudado há vários anos e tipicamente são usadas imagens multiespectrais obtidas através de satélites e aeronaves que são depois analisadas em detalhe.

A entidade responsável por esta monitorização em Portugal é a Direção-Geral do Território (DGT) que produz a Carta de Uso e Ocupação do Solo (COS), onde identifica o uso e ocupação de solo de Portugal continental. Este documento tem uma resolução espacial muito boa mas a sua resolução temporal é muito baixa, pois só é produzido em alguns anos visto ser feito de forma manual. Isto é prejudicial ao acompanhamento em detalhe das mudanças na ocupação de solo visto muita informação não ser registada.

Nesta dissertação desenvolveram-se vários classificadores, distribuídos de forma hierárquica, para mitigar este problema. Foram usadas árvores de decisão treinadas com imagens recolhidas pela constelação Sentinel-2. Apesar destas imagens terem uma resolução espacial mais fraca, os classificadores conseguem classificar o solo de maneira automática apenas em alguns minutos cada vez que um novo conjunto de imagens é recolhido, neste caso a cada 5 dias. Nem todas as imagens podem ser usadas, devido às condições atmosféricas, mas continua a ter uma resolução temporal superior à COS.

No entanto, esta classificação automática não é tão exata quanto a manual. Também não foram consideradas tantas classes quanto as presentes na COS e os classificadores tiveram dificuldade em diferenciar algumas delas, seja pela amostra ser muito pequena ou pelos valores espetrais serem demasiado semelhantes. Considerando o Matthews Correlation Coefficient (MCC), a classe "water" obteve os melhores resultados com uma média de 91,28%, seguida pelas classes "forest" e "agriculture" com uma média de 47,88% e 42,34%, respetivamente, e por último as classes "urban areas" e "bare land" com uma média de 28,03% e 20,53% respetivamente. Mesmo assim considera-se que os resultados obtidos são satisfatórios, mas com muitas oportunidades de melhoria.

Palavras-chave: deteção remota, classificação de ocupação de solo, Sentinel-2, árvores de decisão.

Contents

List of Figures

List of Tables

Glossary

land use	How a portion of land is used overall, without regard of its current state
land cover	How a portion of land is occupied at its current state, independent of its use
remote sensing	The ability to gather spectral data from a site remotely, usually using satellites and aircrafts
spectral signature	The typical emitted radiation from the various wavelengths from a certain type of land
supervised learning	The training data for a classifier consists of pairs of input values and their desired output
unsupervised learning	The training data for a classifier consists of only the input values without the desired output

Acronyms

BA	Balanced Accuracy
BI	Bare soil Index
CI_{green}	Green Chlorophyll Index
COS	Carta de Uso e Ocupação do Solo
DBSI	Dry Bare-Soil Index
DGT	Direção-Geral do Território
FN	False Negative
FP	False Positive
GNDVI	Green Normalised Difference Vegetation Index
GVI	Greenness Vegetation Index
IBI	Index-based Built-up Index
MCC	Matthews Correlation Coefficient
MNDWI	Modified Normalised Difference Water Index
NDBI	Normalised Difference Built-up Index
NDMI	Normalised Difference Moisture Index
NDVI	Normalised Difference Vegetation Index
NDWI	Normalised Difference Water Index
NIR	Near Infrared
NPV	Negative Predictive Value
OSAVI	Optimised Soil Adjusted Vegetation Index

PPV	Positive Predictive Value
SAVI	Soil Adjusted Vegetation Index
SCI	Soil Composition Index
SWIR	Shortwave Infrared
TCARI	Transformed Chlorophyll Absorption in Reflectance Index
TIR	Thermal Infrared
TN	True Negative
TNR	True Negative Rate
TP	True Positive
TPR	True Positive Rate
UI	Urban Index

Chapter 1

Introduction

1.1 Motivation

The surveying and mapping of land use and land cover are very important for various sectors, such as forest and water bodies monitoring, precision agriculture and urban planning. Traditional methods include *in situ* measurements accompanied by manual classification of those same measurements. This process has many flaws, as it is extremely time-consuming and expensive, making it difficult to effectively follow the temporal evolution of land change (Calvao and Palmeirim (2004)). Besides these, a manual process is always subjective to personal interpretation, producing different results for the same area, causing discrepancies (Zha et al. (2003)).

To tackle these problems, satellite remote sensing and automatic classification have been abundantly used for these kinds of applications, providing the so needed higher temporal availability at a lower cost, while usually sacrificing accuracy and precision of class identification.

Satellites have the ability to continuously monitor Earth's surface (Huang et al. (2018)) providing consistent, accurate and reliable digital images to be used in computer-based classification techniques. These readings are multi-spectral, meaning they will produce an image at each band the sensor is able to capture, which is extremely useful as some elements of the land can have a signature high/low reflectance in certain bands that will help to identify and differentiate them from their surroundings. They can also have different spatial resolutions, that will dictate the level of detail they can capture, usually ranging from 1000 m to 1 m, depending on the satellite (Huang et al. (2018)), with technology always evolving to allow even finer spatial resolution (e.g. WorldView-4 with a spatial resolution of 0,3 m for its panchromatic band).

Satellites also have the advantage of being able to reach otherwise inaccessible areas

if done through ground measurements (Etteieb et al. (2013)). All these at a very low, sometimes free, cost.

However, they also present some drawbacks. Multi-spectral sensors are not able to penetrate clouds obscuring some areas of the lands, especially on overcast days (Huang et al. (2018); Xue and Su (2017)). The spatial or temporal resolution might not always be enough depending on the study being performed and when one of them is high, the other tends to be low, making it difficult to find a good balance between them (Huang et al. (2018)).

Nevertheless, its applications are countless and of great importance.

Surveying water bodies can help to identify changes in water level over the year, allowing better management of this resource for human activity, such as agriculture, consumption, industrial purposes and energy production, among others (Huang et al. (2018)).

Forests are another critical resource that should be monitored as they are important for both human activity and the environment. They have a lot of economic, aesthetic and recreational values for humans (Etteieb et al. (2013)) and are crucial for biodiversity, global atmospheric cycles, carbon biogeochemical cycling (Brown et al. (2013)), etc.

Besides natural resources, there is also a need for remote sensing of man-made structures, being them crop fields or urban buildings. When remote sensing is applied to agriculture improvement it is usually given the name of precision agriculture. It can improve effectiveness by monitoring crops' health and identifying any lack of micronutrients or diseases, so adequate fertilisers can be used at early growth stages, or by predicting the crop's yield at an advanced development stage (Haboudane et al. (2002)). It is also used for crop discrimination, crop prediction (Bannari et al. (1995)) and crop frequency changes (Brown et al. (2013)) so agricultural maps can be created to better manage the available land (Karakizi et al. (2016)).

Remote sensing in urban environments can be used for map updating, urban development analysis, military reconnaissance and disaster management (Tian et al. (2018)). Besides these, both built-up and bare land have to be closely monitored as they have been replacing natural vegetation cover over the years. This is damaging the environment and results in less precipitation, higher temperatures and an increase in water pollutants (Assyakur et al. (2012); Xu (2008)).

1.2 Statement of the Problem

In Portugal the mapping and management of soil is done by Direção-Geral do Território (DGT) and is called Land Use and Land Cover Map (Carta de Uso e Ocupação do Solo (COS), in Portuguese). COS maps and classifies all the territory in continental Portugal, excluding the archipelagos of Azores and Madeira. The images used are collected by aerial means at a height of 3200 m to 4300 m with a spatial resolution of 0,25 m (DGTerritório (2019)). However, they only capture four bands: Red, Green, Blue and Near Infrared (DGTerritório (2019)).

After collecting the necessary orthophotographs they are evaluated and classified manually. There are currently a total of 83 different classes, arranged in a hierarchical structure, with 9 mega-classes as their base and 4 levels of detail (DGTerritório (2019)). Due to its manual classification, its temporal resolution is very low, needing several months for a single COS (COS2018 took 9 months to complete), only producing a new one every few years. Currently there is a COS for 1995, 2007, 2010, 2015 and 2018 (DGTerritório (2019)). Besides this, although the orthophotos have a very high spatial resolution of 0,25 m the proceeding classification will only identify land parcels with a minimum mapping unit of 1 ha where the class constitutes a minimum of 75% of said land (DGTerritório (2019)).

Another problem is the fact that these orthophotos are only collected during a portion of the year, not providing enough information if one desires to study the land change over time. For COS2018 images were collected in June, July, August, September and October of 2018 (DGTerritório (2019)), leaving winter and spring months out, which can hide crucial information for certain studies.

1.3 Objectives

Although COS is a very useful information, it has many problems as described above. The aim of the current work is to tackle some of these problems, starting with the classification method.

Manual classification is too time-consuming for time-sensitive matters such as forest monitoring, prevention and fighting of forest fires, precision agriculture, and so on. For this reason, the main goal of this project is the development and training of automatic classifiers through supervised learning to classify land cover more briefly and frequently so said time-sensitive matters can be detected and monitored more efficiently. There are several supervised learning classification procedures, such as artificial neural networks, nearest neighbour and decision trees, just to name a few. Decision Trees were chosen due to their simplicity and the abundance of examples found in the literature of remote sensing (Wang et al. (2020)).

The training data was obtained using the satellite constellation Sentinel-2, and COS was considered as ground truth. When compared to COS, Sentinel-2 has a lower, but still high spatial resolution of 60 m, 20 m and 10 m, a higher temporal resolution of 10 days per satellite, or 5 days if both are considered, and 9 additional bands to the ones used in COS, totalling 13 bands. This presents many advantages over the methods used for COS. However, due to its lower spatial resolution it is not possible to gather as much detail of the land cover and so the nine mega-classes from COS were grouped into five different classes: water, forest, agriculture, urban areas and bare land.

The final objective of the present work is to develop five different decision trees, one for each of the classes described above, and a final decision tree based on these. These classifiers will take advantage of the higher temporal resolution and the higher number of bands of Sentinel-2 satellite constellation. The 13 bands available are useful for the

computation of indices (water indices, vegetation indices, etc) to help to recognise certain land features, and the higher temporal resolution will allow for these classifiers to be used all year round.

1.4 Dissertation Structure

The present dissertation is structured as follows:

Introduction The present chapter gives the reader an introductory review of the work developed and its motivation.

State of the Art Chapter 2 studies and presents many examples found in the literature of automatic classification in remote sensing. It explores various articles in the five different classes and aims to understand what kind of indices and classifiers are used and the results obtained.

Methodology This chapter presents the methodology used, in particular, how the training dataset was built, the calculated indices and the decision tree parameters used.

Results On Chapter 4 the accuracy assessment metrics are presented and final results are shown and analysed in detail for each of the six classifiers.

Conclusions and Future work Chapter 5 concludes this document reflecting over the methodology used and results obtained, discussing future work that could be done to further improve them as well as presenting some drawbacks felt during the execution of this work.

Chapter 2

State of the Art

This chapter will describe some of the studies found in the literature that could aid the development of the current dissertation. Each will be succinctly described, presenting their methodology, namely, the satellites used, indices calculated, type of classifiers developed and lastly the results obtained. Since the present dissertation focuses on the detection of five different classes (water, forest, agriculture, urban areas and bare land) this chapter will be divided as such, with each section presenting articles from the corresponding class.

2.1 Water Detection

Water is a crucial resource for all life and ecological systems. It sustains all sorts of different habitats for a lot of species of fauna and flora and its hydrologic cycle is critical for the global ecosystem and climate system (Du et al. (2016); Huang et al. (2018)). Its use is also indispensable in modern society, being utilised for human consumption, also as a way of transportation, industrial purposes, agriculture, electricity production, recreational activities, and so on (Huang et al. (2018)).

Due to these reasons, it is imperative that water resources are monitored for their better management. Surface water bodies are not static and do not maintain the same shape and volume of water all year round. These changes occur due to both natural and human influence and can have disastrous effects if extreme, like drought or flooding (Huang et al. (2018)).

In order to prepare and prevent these extreme events, it is important for monitoring to be done with high spatial and temporal resolutions. A high spatial resolution will allow for more precise detection of water boundaries, enabling detection of smaller changes that will aid the prediction of future events. High temporal resolution is just as important for

these same reasons. More frequent observations allow the detection of smaller changes and more samples can help to correctly predict the following increases and decreases in water level.

Despite this, both high and low spatial resolution satellites have been used for this purpose.

Some examples of coarser resolution satellites include Advanced Very High Resolution Radiometer that is onboard of National Oceanic and Atmospheric Administration satellites (NOAA/AVHRR), Moderate Resolution Imaging Spectroradiometer (MODIS), Visible Infrared Imaging Radiometer Suite onboard Suomi National Polar-orbiting Partnership (Suomi NPP-VIIRS) and MEdium Resolution Imaging Spectrometer (MERIS) (Du et al. (2016); Huang et al. (2018)). All these satellites have a spatial resolution above 200 m, not providing as much detail as other satellites but having a broad coverage and higher revisit frequency between 4 and 1 days (Huang et al. (2018)).

Other more commonly used satellites are the Landsat series, which include the sensors Thematic Mapper (TM), Enhanced Thematic Mapper Plus (ETM+) and Operational Land Imager (OLI), Système Probatoire d'Observation de la Terre (SPOT) and Sentinel-2, which is a constellation of 2 satellites (Du et al. (2016); Garcia et al. (2020); Huang et al. (2018); Yang et al. (2017)). These have a higher spatial resolution between 5 m and 100 m which makes them a more popular choice among this kind of studies.

Lastly, some satellites with even greater spatial resolution are also used, however, they have very low coverage and temporal resolution and usually do not have free access as opposed to the majority of the previously stated satellites. These high resolution satellites include IKONOS, RapidEye, Worldview, ZY-3, Quickbird, and GF-1/2 and their multispectral bands have a spatial resolution below 5 m, being able to capture a lot of detail and map even small bodies of water like pools and small rivers (Du et al. (2016); Huang et al. (2018)).

After imagery have been collected from satellites, plenty of techniques can be used to identify water bodies.

Water indices are abundantly used as they are simple to calculate, require low computational power and help to differentiate water from non-water terrain (Du et al. (2016); Huang et al. (2018)). Many different indices have been proposed, some gaining more popularity than others. Some examples include Normalised Difference Water Index (NDWI), that was proposed by McFeeters (1996), which uses the Green and Near Infrared (NIR) reflectance bands; Modified Normalised Difference Water Index (MNDWI), developed by Xu (2006), replaced the NIR band by Shortwave Infrared (SWIR) band in NDWI formula, resulting in a more reliable index (Huang et al. (2018); Yang et al. (2017)); Automated Water Extraction Index (AWEI) which was created by Feyisa et al. (2014) and includes two variations $AWEI_{sh}$ and $AWEI_{nsh}$ for images with or without shadows, respectively.

Besides these, some other indices can also be used, such as Normalised Difference Vegetation Index (NDVI), a vegetation index, to further help differentiate between water

and surrounding land (Garcia et al. (2020); Huang et al. (2018)).

With these indices calculated a simple method to extract water bodies is binary segmentation. A threshold is applied so that any land whose resulting index has a value below that threshold is considered as non-water, and those above it are considered water bodies (Du et al. (2016); Huang et al. (2018); Yang et al. (2017)). This threshold is usually set to 0. However, it may not always be suitable depending on the satellites used, their sensors and the time of the year images were taken. A simple solution is applying adaptive threshold algorithms such as OTSU (Du et al. (2016)), although these also present some drawbacks such as uneven luminosity or when the water is not present in the image.

Besides segmentation, both supervised learning and unsupervised learning methods are also popular. Some of them are Decision Trees (Garcia et al. (2020); Huang et al. (2018)), Minimum distance (Garcia et al. (2020)), Nearest neighbour (Huang et al. (2018)), Regression (Huang et al. (2018)), Random forests (Yang et al. (2017)), Deep learning (Yang et al. (2017)), etc. All having their benefits and drawbacks.

2.2 Forest Detection

Similarly to water, forests have a very important role for both ecosystems and human life. They cover about 30% of terrestrial land all across the globe and provide a suitable habitat for almost two-thirds of Earth's species (Huete (2012)). Forests also have an immense impact in global atmospheric cycles (Etteieb et al. (2013)) and their decline leads to an increase in temperature. As previously stated, forests have a great economic value, as wood products play an important role in today's society, just as aesthetic and recreational values for us humans.

However, deforestation has become an increasing problem, with close to 7 million ha being destroyed every year (Huete (2012)) mainly due to economic and political reasons. This is unsustainable and many problems arise due to it, such as climate change, loss of habitat and endangerment of plenty of different species (Huete (2012)).

Therefore it is indispensable to monitor and manage forests globally in a timely and consistent manner. Satellite remote sensing has been increasingly popular in this field, proving its ability to accurately identify vegetation (Etteieb et al. (2013); Huete (2012)). This is due to their high spectral, spatial and temporal resolution.

To further improve vegetation detection, vegetation indices are used. These are widely adopted when multi-spectral imagery is available and will usually use a band with high absorbing characteristics in relation to the desired feature (in this case vegetation) and a low absorbing band (Huete (2012)), but there are also several indices that will use more than two bands. These indices measure the vigour of vegetation and will make it stand out from surrounding land, making its identification easier.

Numerous studies that use satellite remote sensing and vegetation indices have been conducted.

Huete (1997) used Landsat TM imagery to simulate MODIS, a satellite that was launched the year after (1998). This was done to better understand the changes, especially calculating vegetation indices, that would occur when shifting from the broadband AVHRR to the narrower band MODIS satellite. This shift would improve vegetation detection and monitoring as MODIS had a greater spatial and spectral resolution than AVHRR. AVHRR had visible and NIR bands overlapping in the red region of the spectrum (Huete (1997)). This was not ideal for vegetation detection because that is the chlorophyll absorption region and various indices use both the NIR and red portion of the spectrum. AVHRR also had a lower spatial resolution of 1 km to 4 km compared to the 250 m to 500 m resolutions of MODIS (Huete (1997)).

Vegetation indices studied include the most widely used NDVI; Soil Adjusted Vegetation Index (SAVI), created by Huete (1988); atmospherically resistant vegetation index (ARVI) and soil and atmospherically resistant vegetation index (SARVI), developed by Kaufman and Tanre. (1992); modified normalised difference vegetation index (MNDVI), by Liu and Huete (1995) and soil and atmosphere resistant vegetation index (SARVI2) produced by Huete et al. (1996). It was observed that NDVI would saturate in certain scenarios while SARVI2 did not show any saturation. It was shown that SARVI2, SARVI and SAVI would extend the range of sensitivity on various scenarios where NDVI would saturate (Huete (1997)).

Arroyo-Mora et al. (2005) conducted a study on secondary forest detection using Landsat 7 ETM+ and IKONOS imagery. Images from IKONOS with 4 m resolution and 4 spectral bands were collected during the dry season and from Landsat ETM+, with 28,5 m spatial resolution and 7 spectral bands, during the transition period from rainy to dry season. The vegetation indices examined were simple ratio (SR); NDVI; infrared index (IRI) and mid-infrared index (MIRI), both by McMorrow (2001). And some of the techniques used were quadratic classifier (qdc), k-nearest neighbour classifier (knnc) and feed-forward neural network classifier along with the threshold approach for vegetation indices.

Etteieb et al. (2013) used hyper-spectral imagery to map Mediterranean forests. They used Hyperion for this study, which has much greater spectral resolution, at a spatial resolution of 30 m, than commonly used satellites with the same spatial resolution, such as Landsat and Sentinel. It has 242 narrow spectral bands, however there is no need in using all of them considering that bands closer to each other are very similar and can add redundancy. For this reason 44 of the 242 bands were left out for this study and the remaining were grouped into 9 different band combinations for further testing (Etteieb et al. (2013)). Several indices were used in this study, being them NDVI, Difference Vegetation Index (DVI) , (NIR - Green) index, SAVI, and SWIR_VI.

This study used both pixel-based classification, where each individual pixel is assigned a class based on its spectral characteristics disregarding surrounding pixels, and object-oriented classification, where the image is segmented into objects and the class is attributed to each object as a whole, taking into account spatial information (Etteieb

et al. (2013)). The algorithm used for pixel-based classification was Spectral Angle Mapper (SAM) which is a supervised learning classification algorithm. For object-oriented classification, segmentation was first applied to the images and then a nearest neighbour classifier and a membership function classifier were used.

The results obtained by testing all band combinations with all classification methods achieved an overall accuracy between 78.74% and 82.95% (Etteieb et al. (2013)). The band combination that fared better against all was 8, which produced the best results with all three classifiers. This band combination resembled Landsat data and the results obtained were similar to previous studies performed with Landsat imagery (Etteieb et al. (2013)). It proved that for this kind of applications spatial resolution is more important than increasing spectral resolution (Etteieb et al. (2013)).

From the three classifiers, nearest neighbour was the best classifier for coniferous forests, with an accuracy ranging from 84,62% to 87,42%, and the membership function classifier produced the best results for broad-leaf forests with a user accuracy of 91,67% (Etteieb et al. (2013)). Nearest neighbour and membership function classifier were overall better than SAM classifier, proving object-oriented approaches to be better in this context than pixel-based approaches, even though pixel-based approaches have been used and refined for longer (Etteieb et al. (2013)).

Finally, Oliveira et al. (2021) used satellite remote sensing to identify eucalyptus trees in continental Portugal. Eucalyptuses are a major concern in Portugal, because although they have an interesting commercial value (e.g. paper industry) due to their fast growing nature, they are a significant water-demanding plant, out-competing other native vegetation for this resource and soil nutrients (Oliveira et al. (2021)). The satellite used in this study was Sentinel-2 and the vegetation indices considered were NDVI; Normalised Difference Moisture Index (NDMI) created by Wilson and Sader (2002); Green Normalised Difference Vegetation Index (GNDVI) by Gitelson et al. (1996); Green Chlorophyll Index (CI_{green}) proposed by Gitelson et al. (2005); Soil Composition Index (SCI) (also called Normalised Difference Built-up Index (NDBI) by Zha et al. (2003)) and Green Vegetation Moisture Index (GVMI). Different classification techniques were studied, such as decision trees, artificial neural networks, fuzzy logic, among others but the final choice was fuzzy logic, specifically Fuzzy Information Fusion (FIF) (Mora et al. (2017); Oliveira et al. (2021); Ribeiro et al. (2014)).

2.3 Agriculture Detection

When it comes to vegetation, remote sensing is not only useful for the detection and monitoring of forests but also for agriculture, namely, precision agriculture. Agriculture is one of the oldest human inventions that helped us thrive as a species and it is still today a crucial part in any society. Agriculture is essential for many different sectors, being direct human consumption, livestock feeding, textile industry, some of them. For this

reason, there is a great economic interest in increasing productivity, maximising yield and optimising profitability (Haboudane et al. (2002)).

Some uses and applications for agriculture detection include greenhouse mapping and detection, as well as identification of the crops growing inside them (Aguilar et al. (2015)), identification of crop types and crop rotation dynamics (Aguilar et al. (2015)), seasonal evolution of single crops and their use over time (Solano-Correa et al. (2017)), crop mapping, crop inventory, crop monitoring, crop forecasting and yield predictions, crop condition and status, crop disease, crop micronutrient deficiency (Haboudane et al. (2002)), monitoring land use and land cover change, evaluating the implications for carbon biogeochemical cycling, deforestation trends and effectiveness of agri-environmental governance systems (Brown et al. (2013)), and many more could be listed.

The information gathered from these applications can then be used to make better management decisions, for example, when to apply fertilisers and the correct amount to use.

Satellites are often used for this kind of applications, however, precision agriculture usually needs higher detail because simply identifying vegetation is not enough, instead, the correct identification of their species is highly desirable. This makes the use of satellites with finer spatial resolution mandatory. Some other means are also used, such as aerial imagery and local measurements.

Similarly to the previous section, Forest Detection, vegetation indices are widely used to highlight vegetation and minimise disturbance from background reflectance.

Wiegand et al. (1991) studied the relationship between vegetation indices and crop assessment. They used SPOT satellite along with an airborne video system and studied the vegetation indices Greenness Vegetation Index (GVI); Perpendicular Vegetation Index (PVI), created by Richardson and Wiegand (1977), Transformed Soil Adjusted Vegetation Index (TSAVI), proposed by Baret et al. (1989) and NDVI. They found these vegetation indices to be highly correlated to chlorophyll content, leaf area index (LAI), plant height, yield, percentage ground cover by vegetation, among others (Wiegand et al. (1991)) and could be used to calculate other parameters such as green leaf density, photosynthesis rate, amount of photosynthetically active tissue, and others (Wiegand et al. (1991)). They also discovered the measurements of these indices to be best done during the middle of the plant's life, as measures taken too early or too late were found to not correlate well with yield (Wiegand et al. (1991)).

Haboudane et al. (2002) further explored the correlation of vegetation indices and crops' chlorophyll content. The study revolved around two vegetation indices: Transformed Chlorophyll Absorption in Reflectance Index (TCARI), proposed in this study, and Optimised Soil Adjusted Vegetation Index (OSAVI) by Rondeaux et al. (1996) and the ratio between them, although Modified Chlorophyll Absorption in Reflectance Index (MCARI) was also used as a mean of comparison. Satellite imagery was not used but instead aerial by the Compact Airborne Spectrographic Imager (CASI), with a spatial resolution of 1 m for multi-spectral mode and 2 m for hyper-spectral mode, along with

ground measurements, which are, collection of leaf tissue samples to be further analysed in a laboratory, use of spectrometer to measure leaf reflectance and transmittance, LAI and growth measurements (Haboudane et al. (2002)). CASI data went through some processing stages, for example, atmospheric corrections and removal of aircraft motion effects. MCARI was found to perform poorly on low chlorophyll concentration levels, emphasising the benefits of using TCARI which exhibited better sensitivity even at low levels. It was found that the use of TCARI and OSAVI offered great potential for estimating crops' chlorophyll content. Chlorophyll content can help to determine nitrogen concentration as it relates to photosynthesis activity which chlorophyll content is an indicator of, serving as a measure of the crop response to nitrogen application (Haboudane et al. (2002)).

Brown et al. (2013) developed a multiyear classifier for agricultural land use in Brazil. Data was acquired by MODIS satellite, NDVI and Enhanced Vegetation Index (EVI) 250 m datasets were used, as well as *in situ* interviews of farmers and farm managers. The final datasets had 5 years of field data and corresponding MODIS data. Some preprocessing was done to improve future results, which include smoothing, filtering and interpolation.

Then the type of classifier was chosen. Decision Trees are a very common classifier in remote sensing and this study took a similar approach using Boosted Decision Trees. Boosting will generate multiple decision trees rather than a single one, iteratively optimising them. Cross-validation was then used to train the classifier, leaving one year out of the training dataset in the first approach, and leaving 20% of random data out in the second approach. Data filtering was also applied to NDVI and EVI datasets in three different levels: no filtering, "85/15" and "80/20" which used identical filtering procedures but with a different filter parameter value (Brown et al. (2013)). The results of this study showed very little difference between NDVI and EVI datasets with only 0,1% difference in percentage accuracy when using the second approach of cross-validation method. Overall final results increased as filtering became more rigorous and using the year approach to cross-validation and filtering of "80/20" accuracies were consistently near or above 80% (Brown et al. (2013)).

Aguilar et al. (2015) tried not only to map greenhouses but also to identify the crops growing inside them using satellite remote sensing. This was done in southern Spain, where the use of greenhouses is abundant and crucial for the economy. Some problems right from the start were the plastics used in greenhouses since these can vary in thickness, transparency, ultraviolet and infrared reflection and transmission properties, additives, age and colours. Besides these, they are also usually painted white during summer months to protect the plants from excessive sun and inhibit the temperature from getting too high.

Eight images from Landsat 8 OLI were used in conjunction with a single image from WorldView-2 satellite. A large amount of vegetation indices were used, over 25, but the authors found GVI, GNDVI and Normalised Differential Senescent Vegetation Index (NDSVI), proposed by Qi et al. (2002), to be the most significant ones (Aguilar et al. (2015)). Decision Trees were also used in this study, however, boosting was not applied

like in the previously described one. Decision Trees were chosen due to being easy to understand and interpret, being computationally fast and making no assumption about the distribution of data. For the node-splitting rule impurity was measured using Gini's index and a 10-fold cross-validation procedure was used to evaluate the resulting classifier. This was done using an object-based approach, rather than pixel-based, and the best overall accuracy achieved using Landsat was 81,3% with some crops resulting in better outcomes than others. Tomatoes and peppers achieved F_1-scores of 87,7% and 87,0%, respectively, while cucumber and aubergine ranged from 50% to 63%. These results were achieved using Landsat imagery only, as it was found that using WV-2 along with Landsat only slightly improved the results, with the best overall accuracy using both datasets being 82,3%. It was found that although high spatial resolution is important, the use of multitemporal images along growing season was critical to crop discrimination.

Karakizi et al. (2016) also used object-based classification to map and discriminate vine varieties. This was done in four different viticulture regions in Greece with images taken by the satellite WorldView-2 accompanied by ground measurements using a portable spectroradiometer to update and verify boundaries in existing maps. In this study fuzzy logic and knowledge-based rule sets were used along with a Nearest Neighbour classifier. The vegetation indices used in this study were NDVI and modified Soil-Adjusted Vegetation Index (MSAVI), created by) Qi et al. (1994). The first step of segmentation into objects used a multi-resolution segmentation algorithm, based on a region-merging procedure. For classification, data was divided into two classes, vegetation and non-vegetation, and then the class vegetation was further divided into vineyard and other vegetation. Membership functions were developed based on this division. Another segmentation was then performed on the vineyard class to extract the vine canopy, proceeded by a nearest neighbour classifier. For vine variety discrimination nearest neighbour classifier was also used followed by a post-classification procedure based on a majority voting process. The results obtained for vineyard detection had an overall quality above 80% and for vine canopy extraction overall accuracies were above 96%.

Solano-Correa et al. (2017) analysed the evolution of crop fields using Sentinel-2 data. Mapping and tracking every stage of the vegetative cycle is important, however the different varieties of crops will present different maturation in the vegetative cycle and might not always be detected. NDVI and NDWI are used to make a spectral and spatial analysis, creating a multitemporal spectral mask and a multitemporal edge mask, respectively, to separate crop fields. These masks are then fused together and afterwards a connected component labelling algorithm is applied so each crop can be analysed separately.

2.4 Urban Area Detection

Urbanisation has been rapidly increasing all over the world, especially in developing countries with high economic growth. Since 1950 urban areas have expanded more than 25%, increasing from 30% to 54% in 2014 and still expanding to this day (Goldblatt

et al. (2016)). These changes are drastic and in a short period of time, and will usually replace of natural vegetation cover with buildings and paved surfaces (Xu (2008)). This impacts the environment in many negative ways such as increases in temperature, reduces in water quality, water pollution, high surface runoff (As-syakur et al. (2012)), less precipitation, more dryness (Xu (2008)), danger to biodiversity, among other. For this reason, it is important to monitor these changes. Some other reasons to map these changes include better management, planning and decision making when expanding urban areas.

Zha et al. (2003) proposed an index for mapping urban areas, Normalised Difference Built-up Index (NDBI). This was done with Landsat TM imagery from Nanjing, the capital of Jiangsu province in China. Surrounding the city there is farmland and mountainous areas where the natural vegetation is mainly coniferous forests. To filter out vegetation NDVI was used and its result image was recoded into a binary image, considering all positive pixels as vegetation and the remaining as non-vegetation. NDBI was then used to extract built-up land. It resulted in a value close to 0 for vegetation, negative for water, and a positive value for urban areas. It was then, also recoded into a binary image using the same method. This does not ensure that all vegetation will be coded as non-built-up since their value can be positive, although close to 0. To resolve this issue the recoded NDVI image was subtracted to the recoded NDBI image leading to only built-up pixels being considered. This resulted in an accuracy of 92,6% in this area. However, this result is dependant on the assumption that vegetation will always have a NDVI value greater than 0, which is not always the case. Vegetation reflectance can vary depending on the location, species and soil moisture. So, although NDBI is effective, it is not foolproof and should be used bearing in mind the surrounding vegetation.

Xu (2008) proposed Index-based Built-up Index (IBI), a new index to highlight built-up land in satellite imagery using the previous index, NDBI. Similarly to the previous study, other indices were also used to suppress the noise created by other elements such as vegetation. In this study SAVI was used instead of NDVI to filter out vegetation. This choice was made due to the fact that NDVI will only work when vegetation cover is above 30%, while SAVI can work with percentages of plant cover as low as 15%, which are more common in urban areas. Xu also used MNDWI to remove water bodies from urban area detection. NDBI is then calculated and the three images resulting from the indices are used as if they were reflectance bands to calculate IBI.

The index was then verified using images from Fuzhou City, also in China, collected by the satellite Landsat ETM+. IBI was calculated and then a manually determined threshold was used to extract built-up areas. The non-urban areas were then removed using a vector polygon defining the detected urban outline (Xu (2008)) and the final overall accuracy achieved 96,77%.

In 2012 another new index was proposed by As-syakur et al. (2012) named Enhanced Built-up and Bareness Index (EBBI). As the name suggests, this index is capable of mapping and distinguish both urban areas and bare land. Landsat ETM+ was used for this study and EBBI was compared with five other indices which are IBI; NDBI; Urban

Index (UI), created by) Kawamura (1996); Normalised Difference Bareness Index (NDBaI), proposed by Zhao and Chen (2005) and NDVI. IKONOS imagery was also used for comparison to determine the levels of accuracy and effectiveness (As-syakur et al. (2012)). When comparing the similarity of built-up areas detected with EBBI and IKONOS imagery the results were 69.65%, and for bare land they were 62.82%. When comparing to the other indices EBBI was the one achieving the highest accuracy. EBBI had an average accuracy of 66,24% while IBI had 54,25% and NDBI had 51,87%. However it is important to notice that EBBI is a general index for both built-up and bare land and when compared to specific indices for either of those classes, namely IBI for built-up land and NDBaI for bare areas, it performs worse for those individual classes even though its average accuracy is higher. One of the reasons EBBI performs better than the other indices used is the usage of the Thermal Infrared (TIR) band since this band has high emissivity for both built-up and bare land.

Zhou et al. (2014) tested the applicability of NDBI with Landsat 8 OLI, the latest sensor available during the time of this study. They also proposed a new index based on the characteristics of the study area Zhengzhou, the capital of Henan province in China. Landsat OLI differed from Landsat TM (the one used in 2003 NDBI study (Zha et al. (2003))) in band boundaries hence why it is important to test if the index is still valid with these new bands. Similarly to the reference study both NDBI and NDVI were calculated, with equivalent bands from Landsat OLI, then recoded into binary images and the NDVI image was subtracted from the NDBI one.

Then they started developing the new index. They analysed the spectral profiles of the area and found regular patterns. The reflectance of band 2 (Blue) was greater than band 3 (Green), and so was band 4 (Red), so they created two versions of NDBI that used these bands, $NDBI_{OLI2\text{-}OLI3}$ and $NDBI_{OLI4\text{-}OLI3}$ that would help to identify blue-roofed and red/grey-roofed built-up areas, respectively. The images resulting from these two indices were then binarised and the new index, denoted BBI_{OLI}, was calculated by adding both binary images. BBI_{OLI} denotes a binary image with only built-up and bare land area pixels having positive sum values (Zhou et al. (2014)).

The results were clear, NDBI only achieved an overall accuracy of 57,4%, while BBI_{OLI} achieved 90,8%, mapping built-up and bare areas, proving to be superior for this study. The traditional NDBI approach was thus considered to not be applicable to OLI data. Nevertheless, it is important to emphasise that BBI_{OLI} was developed with this particular region, and this particular satellite, in mind and its applicability might not be valid for different cities with different spectral profile patterns or other satellites. BBI_{OLI} also had trouble separating large bare land areas with no relationship to urbanisation such as farmland after harvesting (Zhou et al. (2014)).

Rasul et al. (2018) also used Landsat 8 OLI to test already existing indices and propose two new ones, Dry Built-up Index (DBI) and Dry Bare-Soil Index (DBSI). This study was performed in Erbil, Iraq where the climate is much dryer, resulting in more bare areas and less vegetation, which can have an impact on the indices' performance since

most of them were developed in China where the climate is a lot more humid. DBI was compared to NDBI, UI and EBBI, while DBSI was compared to NDBaI, Bare soil Index (BI), proposed by Chen et al. (2004) and Normalised Difference Soil Index (NDSI), by Rogers and Kearney (2004). Both DBI and DBSI were found to produce better results than the remaining indices with an overall accuracy of 93% and 92%, respectively.

Tian et al. (2018) did not propose a new index for urban mapping but instead employed a new method of automatic classification. They used Deep Convolution Neural Networks (DCNN) to detect urban areas, a rather new and uncommon (but gaining popularity) method in remote sensing. Training samples were divided into $n \times n$ non-overlapping normalised patches and their features were then extracted using a pre-trained DCNN followed by a k-means clustering algorithm to produce visual words and construct a dictionary with them, that would then be used for detection of urban land. The urban words are then learned using a frequency of occurrence histogram, followed by Bayes' rule and a threshold judgement. Detection is a similar process to what was already described. The test image is segmented into patches that will be normalised and have its features extracted and then each of these patches is assigned to the nearest dictionary word. The probability of it belonging to an urban area is then calculated also with Bayes' rule and the decision is made based on a pre-determined threshold. The results were good, achieving an overall accuracy of 87,65% on Google Earth dataset and 92,12% on DigitalGlobe dataset.

2.5 Bare Land Detection

The issues with bare land and its identification are very closely related to the ones with urban areas. As it could be seen in the previous sections many studies developed indices that would map both urban and bare land. Nevertheless, there are also indices and problems specific to bare areas that are worth mentioning in a separate specific section.

Bare lands are important for the ecosystem when they are of natural origins, such as rock areas in mountainous regions or sand areas close to the shore. However, when that is not the case they can have serious negative impacts on both the ecosystem and human lives. Their origins are usually either fallow cropland or construction sites preceding new buildings (Li et al. (2017)) and for that reason, they can be considered an indicator of urban expansion. Although this expansion can be beneficial for citizens, providing new job opportunities and better quality of life, it can be very detrimental to the ecosystem inducing air pollution, such as smog, water pollution and soil loss to the city (Li et al. (2017)). Bare areas can also be a source of sandstorms which can be very damaging in an economic, social and environmental way(Chen et al. (2004)).

For these reasons, it is important to map both urban areas and bare land. Some of the studies leading to new indices used for this purpose were already described in the previous section, being them EBBI, $\text{BBI}_{\text{OLI}}$and DBSI.

Chen et al. (2004) tried to map bare land close to Beijing that was causing harmful dust storms. Multitemporal imagery was collected from Landsat TM and some geometric and atmospheric corrections were performed. They also proposed a new index, the previously mentioned Bare soil Index (BI) and used it in conjunction with NDVI and Shadow Index (SI). The images were then classified using the maximum likelihood method and masking techniques.

Normalised Difference Bareness Index (NDBaI) was previously mentioned, although its study was not described since its focus is on bare land and not urban areas. It was proposed by Zhao and Chen (2005) where the use of TIR band was introduced. The study focused on a segment of the yellow river delta, in China, and Landsat ETM+, coupled with a TIR sensor, was used to gather imagery of the river. The TIR band had a lower spatial resolution of 60 m and had to be rectified to match the spatial resolution of 30 m of the remaining bands using a Nearest Neighbour re-sampling method. These images then went through a process of geometric correction and noise reduction. Then NDBaI was computed and an appropriate threshold was set. The results were very good in inland areas, however, it had some problems alongshore. For this reason, Normalised Difference Soil Index (NDSI) was used to correct this problem which resulted in a combined accuracy of 92,08%

Li et al. (2017) proposed another new index Difference Bare Land Index (NBLI). This study was also performed in China, in the Wuhan city, the largest megacity in central China. The authors studied the indices NDBI, IBI and UI that, although being able to detect bare areas they were not able to clearly distinguish them from urban areas. They also studied NDBaI and EBBI which can better distinguish between bare land and urban land, however, an appropriate threshold had to be specified. Finding a suitable threshold is not always an easy task and can be very time consuming (Li et al. (2017)) so an automatic method is desired.

Landsat TM was used in this study and some atmospheric correction was performed on its imagery. The images were then divided into five classes (built-up, bare land, water, forest and agriculture) and NBLI was calculated as well as the other mentioned indices as a means of comparison. All indices were able to highlight bare areas, however, as expected NDBI, IBI and UI produced values too close to built-up areas. NBLI showed the best results to distinguish bare land from the remaining classes. MNDWI was further computed to remove water bodies. Then they created an unsupervised classifier, using the k-means algorithm, to automatically find an appropriate threshold to extract water bodies. These water bodies were then used as a mask to remove water bodies from NBLI image (Li et al. (2017)).

Then a k-means classifier was again applied to divide the image into several classes (Li et al. (2017)). The results showed IBI, NDBI, and UI images to confuse bare lands with built-up areas excessively, not producing any meaningful results. EBBI was also unable to distinguish them but NDBaI showed better separation between bare and built-up areas but had difficulties distinguishing bare land from agriculture. NBLI had a clear

distinction between bare lands and other classes however some parts of the river had a large amount of suspended soil that could be mistaken for bare land, hence why MNDWI was used to remove water bodies. By removing water bodies NBLI reaches an overall accuracy of 95%, higher than all the remaining indices.

Chapter 3

Methodology

In this chapter the methodology employed is described, firstly presenting some details about the collected data and then describing the preprocessing performed to be later used for both training and testing the developed classifiers. It goes into detail about the indices used and COS classes that were aggregated explaining the reasons behind their choice for each class. Similarly, it presents the reasoning behind the choice of type of classifier (decision trees) explaining some of their advantages when compared to other classifiers. Lastly, the hierarchical structure of the developed classifiers is presented and explained.

3.1 Available Data

The data collected for this study consisted of satellite multi-spectral images from the Sado estuary region all year round. These images were taken by the satellite Sentinel-2 and possess thirteen different spectral bands, at 60 m, 20 m and 10 m spatial resolutions, which are listed on table 3.1.

These samples range from March 2018 until January 2020, choosing only one day per month and using its set of multi-spectral images. Each month had its set of images chosen based on the percentage of cloud coverage, as this would hinder the performance of the classifier due to the lack of visibility.

Figure 3.1 presents the date of each sample and its respective cloud coverage, highlighting higher percentages with reddish tones while using blue for the opposite.

		2018-03-31	2018-04-30	2018-05-10	2018-06-19	2018-07-29	2018-08-18	2018-09-27	2018-10-07	2018-11-16	2018-12-06
		22,74%	30,38%	0,28%	0,20%	0,16%	0,12%	5,91%	0,14%	2,54%	3,84%
2019-01-25	2019-02-14	2019-03-16	2019-04-30	2019-05-15	2019-06-09	2019-07-24	2019-08-23	2019-09-12	2019-10-22	2019-11-16	2019-12-06
0,19%	0,12%	5,02%	8,69%	0,13%	0,33%	0,07%	0,07%	0,08%	4,63%	1,27%	3,03%
2020-01-10											
0,29%											

Figure 3.1: Cloud coverage of each individual sample.

Table 3.1: Sentinel-2 spectral bands.

Bands	Sentinel-2A		Sentinel-2B		
	Central wavelength (nm)	Bandwidth (nm)	Central wavelength (nm)	Bandwidth (nm)	Spatial resolution (m)
Band 1 – Coastal aerosol	442.7	21	442.2	21	60
Band 2 – Blue	492.4	66	492.1	66	10
Band 3 – Green	559.8	36	559.0	36	10
Band 4 – Red	664.6	31	664.9	31	10
Band 5 – Vegetation red edge	704.1	15	703.8	16	20
Band 6 – Vegetation red edge	740.5	15	739.1	15	20
Band 7 – Vegetation red edge	782.8	20	779.7	20	20
Band 8 – NIR	832.8	106	832.9	106	10
Band 8A – Narrow NIR	864.7	21	864.0	22	20
Band 9 – Water vapour	945.1	20	943.2	21	60
Band 10 – SWIR – Cirrus	1373.5	31	1376.9	30	60
Band 11 – SWIR	1613.7	91	1610.4	94	20
Band 12 – SWIR	2202.4	175	2185.7	185	20

As it can be seen, even though efforts were made to select cloudless samples, this is not always possible due to the seasonality of these meteorologic events. March, April, November and December appear cloudy on both years, while September and October vary between the two. The remaining months (May, June, July, August, January and February) present almost nonexistent clouds, having great visibility.

3.1.1 Data Preprocessing

The data mentioned in the previous section were then prepared, using Matlab R2018b, to create a suitable dataset for the training and testing of the several classifiers.

For each month, the first step was resizing all the images to the same size, since there are different spatial resolutions across spectral bands. This could be done by either upscaling all the images to a spatial resolution of 60 m or downscaling them to the resolution of 10 m. The latter was chosen using Nearest-Neighbour interpolation due to higher resolutions holding more detailed spatial information, leading to higher accuracy, (Du et al. (2016); Huang et al. (2018)) that would otherwise be lost if the former was used instead.

The resulting images' size was 10980 by 10980 pixels, which were then reshaped into one column dimension, totalling 120560400 pixels/rows. This was done to every spectral band available excluding band 8A - Narrow NIR, since band 8 - NIR also includes its range of frequencies and has a higher spatial resolution of 10 m instead of 20 m, and also band 10 - SWIR (Cirrus) due to being mainly used for cirrus detection, a type of cloud, which falls out of the scope of the present work and the fact that other SWIR bands (11 and 12) are readily available at better spatial resolutions of 20 m against 60 m. This resulted in a base dataset common to all classifiers, possessing 11 columns and 120560400 rows as it can be seen in the first section of figure 3.2.

To further increase the success of the developed classifiers some commonly used indices were also calculated based on the spectral bands available as shown in the second section of figure 3.2. Each classifier had a different set of indices that were chosen based

on the class to be detected, so they could improve the performance of the classifier. The specific indices used in each classifier will be described in detail in subsection 3.1.2.

The last column missing in these datasets should identify which class each pixel/row belongs to. For this purpose, a section of COS, of the same territory, was used. As it was done previously, this image was reshaped into one single column containing 10 different classes. This process is shown in the last section of figure 3.2 and the classes can be seen in table 3.2 along with their total amount of pixels and the percentage they occupy in the image.

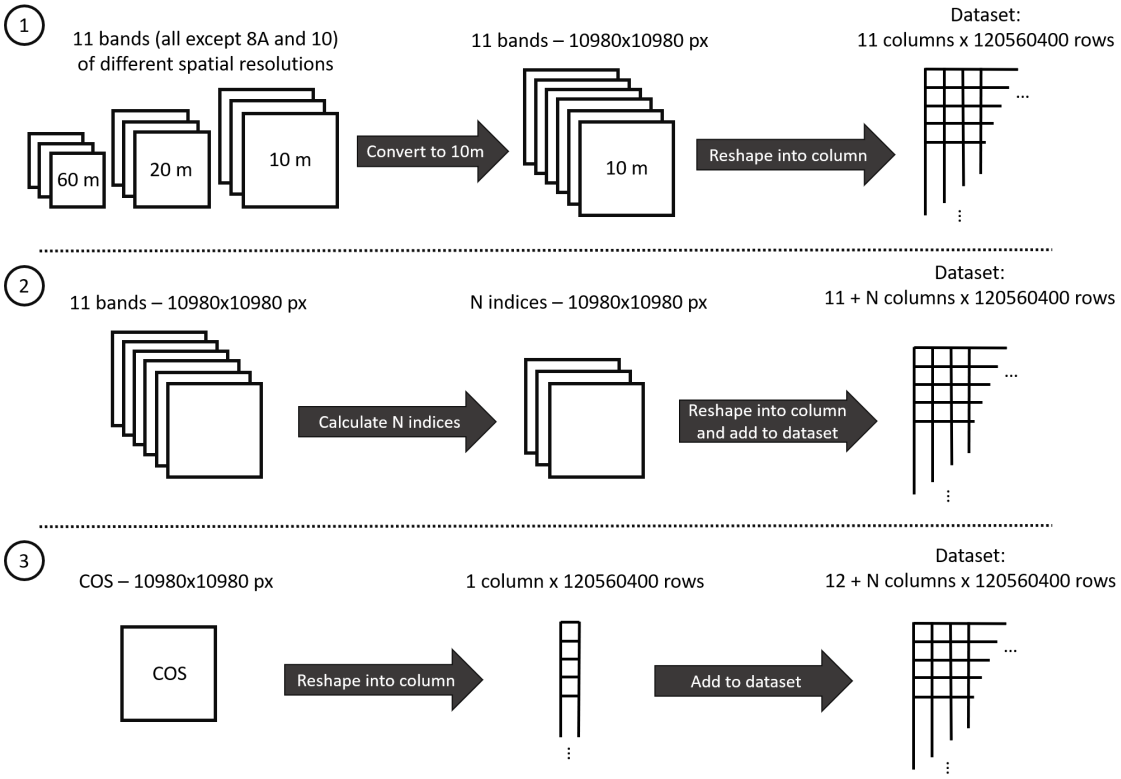


Figure 3.2: Steps to build the dataset of a single month for binary classifiers.

In COS the label 0 is used to identify land that does not fit in the remaining mega-classes, so it could be something else besides ocean (since these are excluded from class 9). However it was confirmed that in the territory portion used in this study all the pixels with the assigned label of 0 were indeed always ocean so they will be considered as such. These 10 classes will then be aggregated and converted into 5 new classes, as it can be seen in figure 3.3 so the classifiers can make binary predictions of a specific class, or non-class (for example, water bodies and non-water bodies).

This process was repeated through all the 23 samples acquired, for each of the 5 classifiers.

Table 3.2: List of COS classes and their predominance in the sample region.

Label	Type of land	Number of pixels	Percentage of pixels
0	Ocean	11.523.931	9,6%
1	Artificial territories	2.961.625	2,6%
2	Cropland	30.584.736	25,4%
3	Grassland	12.280.154	10,2%
4	Agroforestry surfaces	25.348.402	21,0%
5	Forest	32.730.095	27,1%
6	Shrubland	769.446	0,6%
7	Bare areas	245.703	0,2%
8	Humid areas	617.522	0,5%
9	Superficial water bodies	3.498.786	2,9%
	Total:	**120.560.400**	**100%**

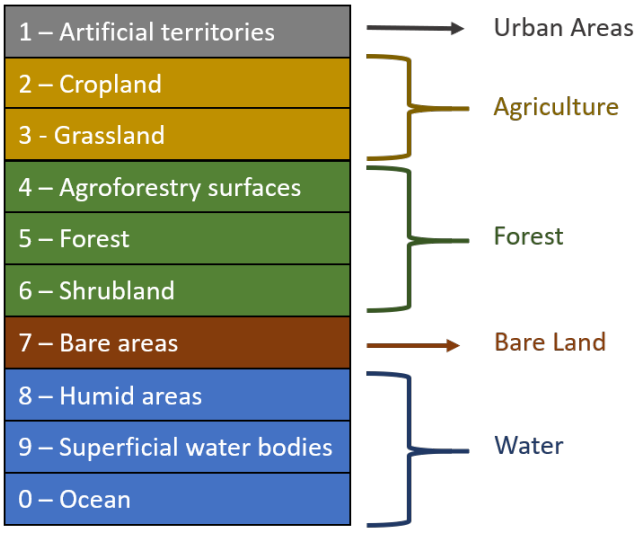


Figure 3.3: Conversion of COS classes to the ones used in this study.

3.1.2 Indices used and classes aggregated

3.1.2.1 Water classifier

For the water classifier three different indices were used, with one of them having two different versions, making it four in total.

The first index used in the water classifier was Normalised Difference Vegetation Index (NDVI) which will help to differentiate and recognise vegetation. It is calculated from the spectral reflectance measured in the red part of the spectrum, which corresponds to Sentinel-2's band 4, and in the Near Infrared (NIR) portion, which is band 8, and its formula can be seen in equation (3.1).

$$NDVI = \frac{NIR - Red}{NIR + Red} = \frac{Band\ 8 - Band\ 4}{Band\ 8 + Band\ 4} \tag{3.1}$$

A second one is Normalised Difference Water Index (NDWI) (McFeeters (1996)) that

enhances water detection. Similarly to NDVI, this index also uses the spectral reflectance in the NIR section of the spectrum (band 8), and, instead of the red portion, uses the green portion of the spectrum (band 3). Following below, it can be calculated using equation (3.2).

$$NDWI = \frac{Green - NIR}{Green + NIR} = \frac{Band\ 3 - Band\ 8}{Band\ 3 + Band\ 8} \tag{3.2}$$

Lastly, since NDWI has shown poor results in some situations due to signal noise and inseparability of built-up areas(Yang et al. (2017)), Modified Normalised Difference Water Index (MNDWI) (Xu (2006)) was also calculated since it has shown better results in such situations. MNDWI is similar to NDWI but it is calculated using the Shortwave Infrared (SWIR) band instead of NIR since water bodies have better absorbability in the SWIR band than NIR band (Du et al. (2016)). Due to the fact that Sentinel-2 has two SWIR bands (excluding cirrus band) two MNDWI were therefore calculated, MNDWI1 with band 11 and MNDWI2 with band 12. They can be calculated as shown in equations (3.3) and (3.4).

$$MNDWI = \frac{Green - SWIR}{Green + SWIR}; \tag{3.3}$$

$$MNDWI1 = \frac{Band\ 3 - Band\ 11}{Band\ 3 + Band\ 11};\ MNDWI2 = \frac{Band\ 3 - Band\ 12}{Band\ 3 + Band\ 12} \tag{3.4}$$

Similarly to the initial images, all these indices had their bands resized to a spatial resolution of 10 m, also resulting in a 10980 by 10980 matrix that was then reshaped into a column of 120560400 rows. They were then added to the base dataset which now counts 15 columns.

From the 10 mega-classes considered, three (0 - Ocean, 8 - Humid Areas, 9 - Superficial water bodies) were merged into one class representing water bodies (identified as 0), while the remaining were joined to identify non-water bodies (identified with the number 1). Thus this dataset is complete, having a total of 16 columns (15 predictors and 1 response column).

Water classifier	
Indices used	
NDVI	NDWI
MNDWI1	MNDWI2
COS classes	
0 (Ocean), **8** (Humid areas), **9** (Superficial water bodies)	

Figure 3.4: Water classifier summary table, indicating the indices used and COS classes considered as water.

3.1.2.2 Forest classifier

For the forest, and following classifiers six different indices were used.

The first one was the most commonly used NDVI to enhance vegetation. This index is incredibly popular due to its good results and can be seen in several studies surrounding vegetation and remote sensing (Arroyo-Mora et al. (2005); Etteieb et al. (2013); Huete (1997); Oliveira et al. (2021)). For this reason, it was unavoidable to use it for forest detection and its formula can be seen in equation (3.1).

A second index is Soil Adjusted Vegetation Index (SAVI) (Huete (1988)) which was calculated using equation (3.5). This index was developed to remove canopy background noise produced by the soil underneath it. It was developed based on NDVI and PVI and similarly uses the NIR part of the spectrum (Sentinel-2's band 8) along with the red band (band 4). The formula has a variable L which is the soil adjustment factor and its standard value for better adjustment is 0,5 (when this value is 0 SAVI equals NDVI). However, this value may not always be the most optimal and can be tweaked to specific scenarios since soil emissivity will vary depending on its composition and water content. Since the study area is too vast and samples are taken all throughout the seasons the water content and composition of the land will vary immensely and the standard value of 0,5 will consequently be used.

$$SAVI = \frac{NIR - Red}{NIR + Red + L} \times (1 + L) = \frac{Band\ 8 - Band\ 4}{Band\ 8 + Band\ 4 + 0,5} \times (1 + 0,5) \tag{3.5}$$

Another index used is Normalised Difference Moisture Index (NDMI) (Wilson and Sader (2002)) which measures vegetation water stress levels and can detect forest harvesting, which is good for monitoring changes in forests' overall health and cover. It is also similar to NDVI but uses the SWIR band (band 11) instead of the red band, due to its ability to better absorb water, and has shown better results in certain situations (Wilson and Sader (2002)). Its formula can be seen in equation (3.6).

$$NDMI = \frac{NIR - SWIR}{NIR + SWIR} = \frac{Band\ 8 - Band\ 11}{Band\ 8 + Band\ 11} \tag{3.6}$$

Next index is Soil Composition Index (SCI), also known as Normalised Difference Built-up Index (NDBI) (Zha et al. (2003)), which is useful to detect both built-up and bare land and thus differentiate them from vegetation so it is easier to identify forests and non-forests land. It uses Sentinel-2's band 11 (SWIR) and band 8 (NIR) and can be computed through equation (3.7).

$$NDBI = \frac{SWIR - NIR}{SWIR + NIR} = \frac{Band\ 11 - Band\ 8}{Band\ 11 + Band\ 8} \tag{3.7}$$

The fifth index is Green Chlorophyll Index (CI_{green}) (Gitelson et al. (2005)) which will calculate chlorophyll content of leaves. Chlorophyll is a pigment that is present in the vast majority of plants and so it can be used to highlight vegetation from surrounding land. This index uses band 8 (NIR) and band 3 (Green) as shown in equation (3.8).

$$CI_{green} = \frac{NIR}{Green} - 1 = \frac{Band\ 8}{Band\ 3} - 1 \tag{3.8}$$

The last index is Green Normalised Difference Vegetation Index (GNDVI) (Gitelson et al. (1996)) which is very similar to NDVI but, as the name implies, uses the green portion of the spectrum (band 3) instead of the red one, as seen in equation (3.9). The use of the green band makes this index at least five times more sensitive to chlorophyll content than NDVI while also having a wider dynamic range of chlorophyll variations (Gitelson et al. (1996)).

$$GNDVI = \frac{NIR - Green}{NIR + Green} = \frac{Band\ 8 - Band\ 3}{Band\ 8 + Band\ 3} \tag{3.9}$$

Like previously mentioned, to calculate these indices all the bands used had to be resized into a 10 m spatial resolution and then the resulting index was reshaped into a single column that could then be added to the base dataset. Along with the response column, this resulted in a final dataset size of 120560400 rows and 18 columns.

Then the values of the response column had to be converted so this classifier would identify land as either forest or non-forest. Both class 4 (agroforestry surfaces) and class 5 (forest) were considered as forest land, due to their similarity, but class 6 (shrubland) was also included since it is also wild vegetation, and the three were identified as the value of 0 while the remaining were all considered as non-forest being attributed the value of 1.

Forest classifier	
Indices used	
NDVI	SAVI
NDMI	NDBI
CI green	GNDVI
COS classes	
4 (Agroforestry surfaces), **5** (Forest), **6** (Shrubland)	

Figure 3.5: Forest classifier summary table, indicating the indices used and COS classes considered as forest.

3.1.2.3 Agriculture classifier

The agriculture classifier is very similar to the forest classifier, as they both identify vegetation, and for this reason, some of the indices used will be common to both (NDVI, GNDVI, SAVI). However they also have some differences, for example, the agriculture classifier will have to identify vegetation all throughout the year and the plants will be in very different growth stages, while the forest one will usually identify plants that are typically slower in growth and do not present as many changes; agricultural crops also tend to be somewhat geometrically organised with bare soil in between plants while on mature forests the crown of the trees might conceal the soil beneath, and the soil itself is usually not completely bare but being occupied instead by small and low vegetation.

NDVI was used as it is the staple index for detection of vegetation and its formula can be seen in equation (3.1).

SAVI was also used as it can help to differentiate plants from the soil surrounding it, which, as stated before, is very useful in the typical grid-like cropland where plants are separated by bare soil. It can be calculated through equation (3.5).

The last index common to both agriculture and forest classifier is GNDVI. This index is far more sensitive to chlorophyll content than NDVI which is a crucial indicator of the plant's health and future yield being very important to the agricultural sector. It can be seen in equation (3.9).,

The next vegetation index used was Greenness Vegetation Index (GVI), as seen in equation (3.10). This is a simple index that uses the green and red reflectance portion of the spectrum, Sentinel-2's band 3 and 4, respectively. It has been proven to effectively identify various different types of vegetation covers, which is really useful in agriculture due to the diversity of species cultivated (Bannari et al. (1995); Wiegand et al. (1991)).

$$GVI = \frac{Green - Red}{Green + Red} = \frac{Band\ 3 - Band\ 4}{Band\ 3 + Band\ 4} \tag{3.10}$$

Next, we have the index Optimised Soil Adjusted Vegetation Index (OSAVI) (Rondeaux et al. (1996)), which belongs to the family of SAVI indexes, like SAVI, MSAVI and TSAVI, thus serving the similar purpose of contrasting vegetation from the underlying soil. It is a simplified version of TSAVI that does not require prior knowledge of the soil optical properties which makes it more accessible to studies where these are not able to be determined. It has also shown to be better suited for agricultural application (Rondeaux et al. (1996)). This index can be calculated through equation (3.11) which uses NIR and red bands (bands 8 and 4, respectively), along with a factor L, for soil adjustment, which standard value is 0,16.

$$OSAVI = \frac{NIR - R}{NIR + R + L} = \frac{Band\ 8 - Band\ 4}{Band\ 8 + Band\ 4 + 0,16} \tag{3.11}$$

The last index used is Transformed Chlorophyll Absorption in Reflectance Index (TCARI) (Haboudane et al. (2002)) which is also a chlorophyll sensitive index. Chlorophyll is an indicator of photosynthetic activity which is strongly correlated to nitrogen concentrations. Monitoring chlorophyll concentrations can help farmers dictate the correct amount of nitrogen to be applied at a given moment to later increase yield. This index has shown to be very sensitive to chlorophyll concentrations at a wide range, including low ones which make it good to detect vegetation at different growth stages. The formula for this index can be seen in equation (3.12) where the number associated with each R variable indicates the wavelength, in nanometers, that should be used. Sentinel-2's bands were then chosen based on these numbers so that their central wavelength would be as close as possible to the value specified so the results would not differ too much.

$$TCARI = 3 \cdot \left((R_{700} - R_{670}) - 0,2 \cdot (R_{700} - R_{550}) \times \frac{R_{700}}{R_{670}} \right) \tag{3.12}$$
$$= 3 \cdot \left((Band\ 5 - Band\ 4) - 0,2 \cdot (Band\ 5 - Band\ 3) \times \frac{Band\ 5}{Band\ 4} \right)$$

As for the classes considered for this classifier, both classes 2 (Cropland) and 3 (Grasslands) were merged to symbolise agricultural land, with the remaining being non agricultural land. The reason for grasslands to be included in this classifier is due to both plant cultivation and livestock production being part of agriculture, and grasslands being used a lot for livestock. Besides this, some grasslands are not of natural origin but instead cultivated by man, therefore, being different from other kinds of spontaneous wild low vegetation.

Agriculture classifier	
Indices used	
NDVI	SAVI
GVI	OSAVI
TCARI	GNDVI
COS classes	
2 (Cropland), **3** (Grassland)	

Figure 3.6: Agriculture classifier summary table, indicating the indices used and COS classes considered as agriculture.

3.1.2.4 Urban Areas classifier

The indices used in the urban areas classifier will not only focus on detecting built-up land but also the typical surrounding land like vegetation and water so they can be more easily discarded as non-built-up areas. For this reason both NDVI (equation (3.1)) and SAVI (equation (3.5)) were used to detect vegetation and MNDWI (equation (3.3)) was used to identify water. The reason behind the use of both NDVI and SAVI resides on their sensitivity on plant cover. NDVI is known to work effectively when the plant cover is above 30%, while SAVI can achieve this in situations with a plant cover as low as 15% (Xu (2008)). This is relevant due to vegetation inside urban areas typically having lower coverage.

Another index previously mentioned that was also used in this classifier is NDBI (equation (3.7)). NDBI was developed with built-up land identification in mind, however, it can also be used to detect soil and bare land, hence why it was used in the forest classifier.

Next we have Urban Index (UI) (Kawamura (1996)) which uses the reflectance of the SWIR and NIR part of the spectrum which correspond to Sentinel-2's band 12 and 8, respectively. This index was developed for mapping built-up land, although it occasionally mistakenly identifies some bare land as built-up land due to their spectral similarities. It can be seen in equation (3.13).

$$UI = \frac{SWIR - NIR}{SWIR + NIR} = \frac{Band\ 12 - Band\ 8}{Band\ 12 + Band\ 8} \tag{3.13}$$

Lastly, there is the Index-based Built-up Index (IBI) (Xu (2008)) which is a slightly different index from the ones previously presented. Instead of directly using the spectral band available this index uses the result of other indices to identify urban areas as it can be seen in equation (3.14). These indices are NDBI, SAVI and MNDWI. NDBI is used to highlight urban areas, however, this index often mixes built-up land with vegetation (Zha et al. (2003)). Therefore SAVI is used to counteract this limitation of NDBI, identifying plant cover to be then removed. NDVI can also be used in place of SAVI, however, it is not as effective in situations where the plant cover is low, which is fairly common in urban scenarios, hence SAVI being preferred, as it can highlight vegetation with its cover is as low as 15%. MNDWI is used to identify water bodies and remove them from the picture. So, by combining the results of both SAVI and MNDWI and subtracting them from NDBI, only built-up land is remaining. The reason why the sum of SAVI and MNDWI is divided by two is to avoid IBI from getting too small values (Xu (2008)).

$$IBI = \left(\frac{NDBI - \frac{SAVI+MNDWI}{2}}{NDBI + \frac{SAVI+MNDWI}{2}} \right) \tag{3.14}$$

Finally, for the COS classes considered in this classifier, only class 1 (Artificial territories) was treated as urban areas, identified by the number 0, while the remaining nine classes were aggregated to identify non-urban areas, identified by the number 1.

Urban Areas classifier	
Indices used	
NDVI	SAVI
MNDWI	NDBI
UI	IBI
COS classes	
1 (Artificial territories)	

Figure 3.7: Urban Areas classifier summary table, indicating the indices used and COS classes considered as urban areas.

3.1.2.5 Bare Land classifier

Similarly to the Urban area classifier, the indices used in this case will also not only focus on bare land but also the neighbouring land such as vegetation and water. Thus, both NDVI (equation (3.1)) and NDWI (equation (3.2) were used to detect each of these cases, respectively.

The index NDBI (equation (3.7)) is also used because as formerly mentioned, although its development focused on built-up land it was found to also be effective to map bare land.

Next index is Bare soil Index (BI) (Chen et al. (2004)) which is an index designed to identify bare land that uses four different spectral bands, SWIR (band 11), NIR (band 8), red (band 4) and blue (band 2). It can be calculated through the formula seen in equation (3.15).

$$BI = \frac{(SWIR + Red) - (NIR + Blue)}{(SWIR + Red) + (NIR + Blue)} = \frac{(Band\ 11 + Band\ 4) - (Band\ 8 + Band\ 2)}{(Band\ 11 + Band\ 4) + (Band\ 8 + Band\ 2)} \tag{3.15}$$

Another bare land index is Dry Bare-Soil Index (DBSI) (Rasul et al. (2018)) which focuses on bare land in dry climates. This index was included due to the Alentejo region in Portugal having very dry summers which result in dry vegetation and soil, which this index aims to detect. Its formula is also uncommon, since it uses both spectral bands and an index, and can be seen in equation (3.16). It uses SWIR and green reflectance bands which are Sentinel-2's band 11 and 3, respectively. The index used is NDVI which is subtracted so vegetation is removed.

$$DBSI = \frac{SWIR - Green}{SWIR + Green} - NDVI = \frac{Band\ 11 - Band\ 3}{Band\ 11 + Band\ 3} - NDVI \tag{3.16}$$

Finally, an index simply denoted by B (equation (3.17)) (Zhou et al. (2014)) is calculated to detect bare land. This index also works in a different way than most traditional indices by using two other indices instead of spectral bands. Besides that, these indices, NDBI and NDVI are first binarised instead of being used as-is. For both indices, any negative value is represented by 0, and any positive value represented by 1. The resulting indices, $NDBI_b$ and $NDVI_b$ are then subtracted so all vegetation is removed and only bare land has a positive value. The use of NDBI has the downside of also detecting urban areas along with bare land which can not always give the best results.

$$B = NDBI_b - NDVI_b \tag{3.17}$$

As for the classes used, COS mega-class 7 (Bare areas) was used to identify bare land, with the value 0, while the rest of the mega-classes were joined to identify non-bare land, with the value of 1.

Bare Land classifier	
Indices used	
NDVI	NDWI
NDBI	BI
DBSI	B
COS classes	
7 (Bare areas)	

Figure 3.8: Bare Land classifier summary table, indicating the indices used and COS classes considered as bare land.

3.2 Classifiers

As stated before, five different classifiers were trained to detect five different classes: water, forest, agriculture, urban areas and bare land, plus one final classifier that uses the result of the latter to further improve performance and solve any conflicts between the binary classifiers. These five classifiers are binary decision trees that can only predict if a pixel is either the class or not the class that it was trained to detect, while the final decision classifier is multi-class and will predict which of the five classes is more appropriate for a given pixel. This process can be seen in figure 3.9.

Decision trees were used because they (and their variants such as boosted decision trees and random forests) have been commonly used, with good results, in the field of remote sensing for a long time (some examples include (Aguilar et al. (2015); Brown et al. (2013); Goldblatt et al. (2016); Huang et al. (2018))). This classification procedure will recursively divide the initial dataset into smaller, homogeneous and mutually exclusive subsets using an impurity function such as information gain index, Gini index, chi-square measure, and others (Brown et al. (2013); Oliveira et al. (2021); Pal and Mather (2003)). This results in a set of rules arranged as an inverted tree-like structure, hence the name, where each node represents a single rule that decides which of its descendants is more appropriate (has less impurity) for the given input data. (Oliveira et al. (2021)).

Decision trees are non-parametric classifiers, meaning they do not make assumptions regarding the distribution of the data (Aguilar et al. (2015); Brodley and Friedl (1997)) that could limit its learning ability. A lot of different datasets can also be used with these classifiers with little preprocessing since they can handle noisy and non-linear relationships between classes, data measured in different scales, both categorical and numerical data and missing values (Brodley and Friedl (1997); Pal and Mather (2003)).

In addition, decision trees are computationally fast for both training and testing, especially when compared to other popular classifiers nowadays such as neural networks and its variants. They also contrast with neural networks when it comes to interpretation, after all, neural networks are viewed as a black box, where the inside cannot be interpreted,

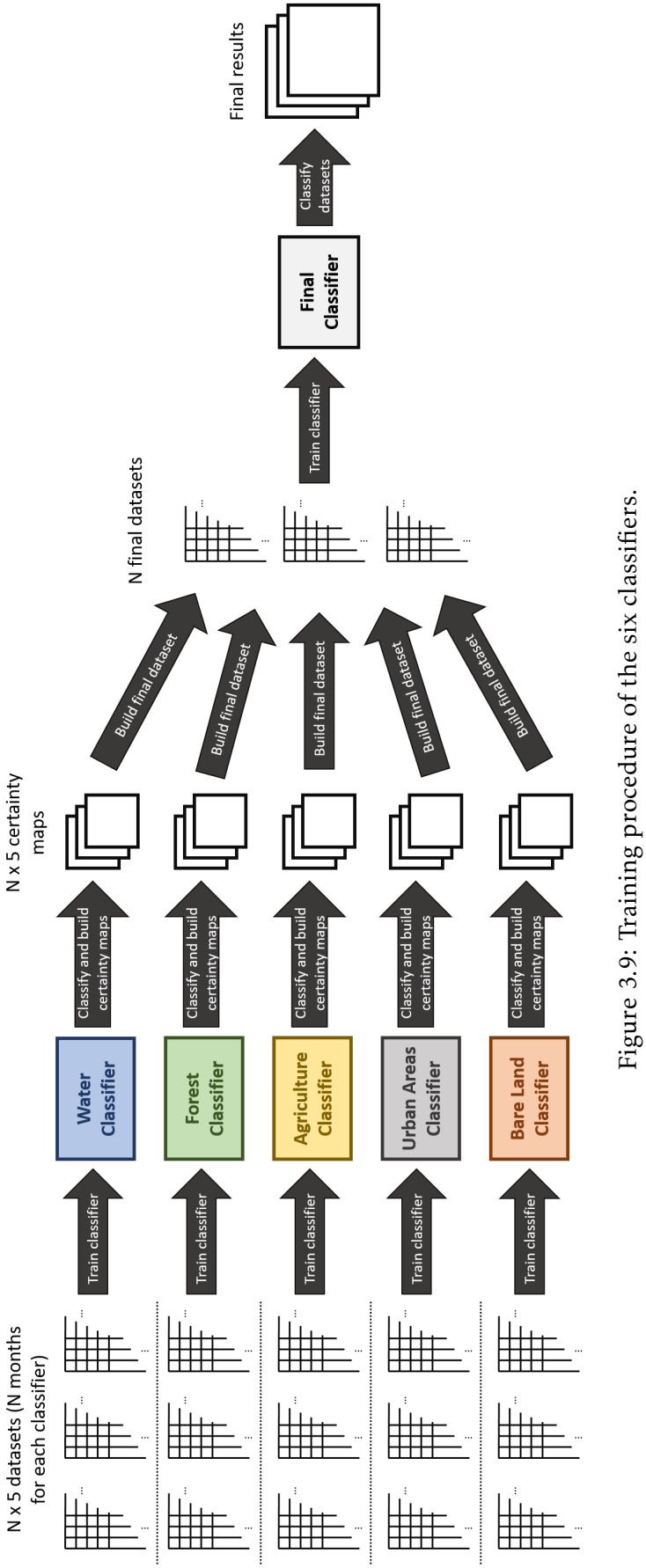


Figure 3.9: Training procedure of the six classifiers.

while decision trees are considered white boxes because their classification structure is made from simple and explicit rules which can easily be read and interpreted by humans.

However, decision trees also have some disadvantages. Once a decision tree is trained its structure and rules are set, not being able to learn and adapt to new data, unlike neural networks. Therefore the decision trees specifically built for this study case may not produce the same results if applied to other areas, especially in different climates.

The development of the classifiers used in this study was also performed using Matlab, version R2018b, and its app "Classification Learner". Given the training data already preprocessed a decision tree of type 'Fine Tree' was used, with a maximum number of splits of 10, using the split criterion of Gini's diversity index and with surrogate decision splits turned off. Due to the very extensive dataset used for training (30.000.000 rows) the holdout method was chosen for validation, with 25% held out, instead of the more commonly used cross-validation method. These parameters were used for all the six classifiers developed in this thesis.

3.2.1 Binary Classifiers

The first part of this work focused on the development of binary classifiers for each of the five defined classes: water, forest, agriculture, urban areas and bare land. For each of these cases there are two versions, one that will use the prepared dataset as it is, and other that will first filter pixels obscured by clouds, by applying an appropriate mask to the satellite images. This is done so the classifier will not create rules based on fallacious values. These were called and will be referred to as, respectively, "Clouds" and "Cloudless".

Then, for each of these two versions, there are other four different variants by the names of "allMonths", "MonthVar", "2019" and "2019_MonthVar". From these variants, "allMonths" and "MonthVar" are very similar to each other, and so are "2019" and "2019_MonthVar", with the only difference being the presence, indicated by the name MonthVar, or absence of an extra column in the training dataset specifying the month (using a number between 1 and 12) at which each sample (each line of the dataset) was taken. The purpose of this extra information is to help identification of land cover that change a lot throughout seasons, namely: water, that is affected by temperature and climate events, suffering periods of drought or shallower water and floods and deeper water which change its spectral reflectance; forest, that similarly go through periods of dry and humid weather which in turn change the spectral reflectance of tree leaves, and in the case of deciduous tree species they completely lose their foliage which also significantly alters the spectral readings of the satellites; and agriculture, that rapidly changes throughout the year to optimise and maximise yield.

The variants "allMonths" and "MonthVar" used all the 23 months for both training and testing the classifier. A sample of 30 million random pixels was taken from the original dataset to create the training dataset, maintaining similar class proportions as to avoid unwanted biases. The numbers of rows taken from each class can be seen in

table 3.3, along with the total number of pixels of each class and their percentage, which can be compared to the original percentages in table 3.2. Some rows indicate values with a decimal part (shown with an asterisk (*)) which is counter-intuitive when pixels are concerned, however, it is just the result of the division between the total numbers of pixels and the number of months concerned. This was corrected by taking an extra pixel from some months until the desired total number of pixels for that class was achieved. It is important to point that in all the 23 months available there are always two distinct observations of each month except for February, and due to that, the number of pixels extracted from February was doubled to correct this imbalance.

Table 3.3: Number of pixels used for the training dataset of classifiers "allMonths" and "MonthVar".

Label	Type of land	Number of pixels per month	Total number of pixels	Percentage of pixels
0	Ocean	* 119.583,(3)	2.870.000	9,6%
1	Artificial territories	37.500	900.000	3,0%
2	Agriculture	* 316.666,(6)	7.600.000	25,3%
3	Grassland	125.000	3.000.000	10,0%
4	Agroforestry surfaces	262.500	6.300.000	21,0%
5	Forest	* 330.416,(6)	7.930.000	26,4%
6	Shrubland	* 8.333,(3)	200.000	0,7%
7	Bare areas	* 4.166,(6)	100.000	0,3%
8	Humid areas	* 8.333,(3)	200.000	0,7%
9	Water bodies	37.500	900.000	3,0%
	Total:		**30.000.000**	**100%**

However, training and testing with the same dataset can lead to optimistic results and for this reasons the variants "2019" and "2019_MonthVar" were created. These variants will train with the 12 months from 2019 and test with the remaining. The training dataset was built similarly to the previous one, selecting 30 million random pixels from each class, while maintaining a similar percentage to the original dataset, as it can be seen in table 3.4 (the value with a decimal part, indicated by asterisk, are treated as it was described for table 3.3, taking an extra pixel from some months until the desired total number of pixels for that class is achieved). Unlike the previous one, since we only use one month of each, it is not necessary to double the value of pixels taken from February.

A flowchart illustrating the process of creating a training dataset depending on the intended version and variant can be seen in figure 3.10. It can be seen how variants "allMonths" and "MonthVar" differ from variants "2019" and "2019_MonthVar" in the number of datasets (months) used. Then it is possible to observe that clouds are immediately filtered out before any further processing when the version "Cloudless" is chosen, and after that a column for the month variable may be added depending on the variant. Afterwards a sample from each class if taken from each dataset (month) according to

Table 3.4: Number of pixels used for the training dataset of classifiers "2019" and "2019_MonthVar".

Label	Type of land	Number of pixels per month	Total number of pixels	Percentage of pixels
0	Ocean	* 239.166,6 (6)	2.870.000	9,6%
1	Artificial territories	75.000	900.000	3,0%
2	Agriculture	* 633.333,3 (3)	7.600.000	25,3%
3	Grassland	250.000	3.000.000	10,0%
4	Agroforestry surfaces	525.000	6.300.000	21,0%
5	Forest	* 660.833,3 (3)	7.930.000	26,4%
6	Shrubland	* 16.666,6 (6)	200.000	0,7%
7	Bare areas	* 8.333,3 (3)	100.000	0,3%
8	Humid areas	* 16.666,6 (6)	200.000	0,7%
9	Water bodies	75.000	900.000	3,0%
	Total:		**30.000.000**	**100%**

tables 3.3 or 3.4 depending on the variant chosen. For any version or variant the training dataset will total 30 million rows and the training of decision trees can then begin.

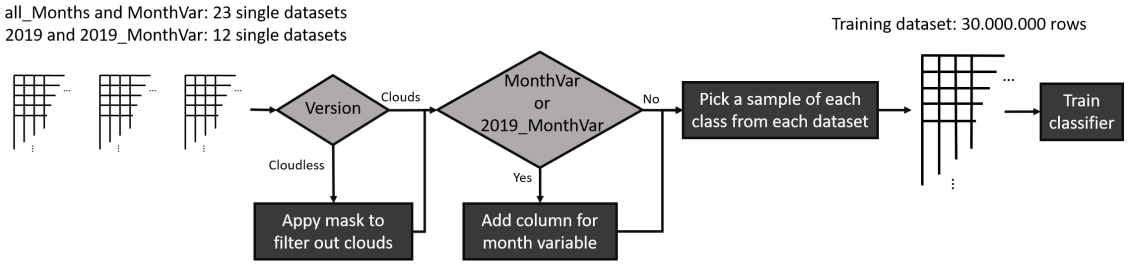


Figure 3.10: Flowchart depicting the process of creating a training dataset depending on the desired version and variant.

3.2.2 Final Decision Classifier

After all the versions and variants from each class were computed the final decision classifier was developed. This classifier takes a different approach, not using the original dataset of satellite imagery but instead the results from the developed binary classifiers, as seen in figure 3.9. Each binary classifier can output if a pixel is either its class or not, and it is also possible to know how certain the classifier is that the pixel in question represents in fact the class it was trained to detect. This percentage of certainty (certainty maps) of each class was obtained through the function *[label,score,node,cnum] = predict(Mdl,X)*, from Matlab, and will form the training dataset for this final classifier as shown in figure 3.11.

Only the "Cloudless" version of binary classifiers was employed to create this training dataset, so rules and results derived from fallacious readings due to cloud cover will be avoided. Then, for each variant, the developed binary decision trees were used to classify the intended months (23 months for variants "allMonths" and "MonthVar", 12 months

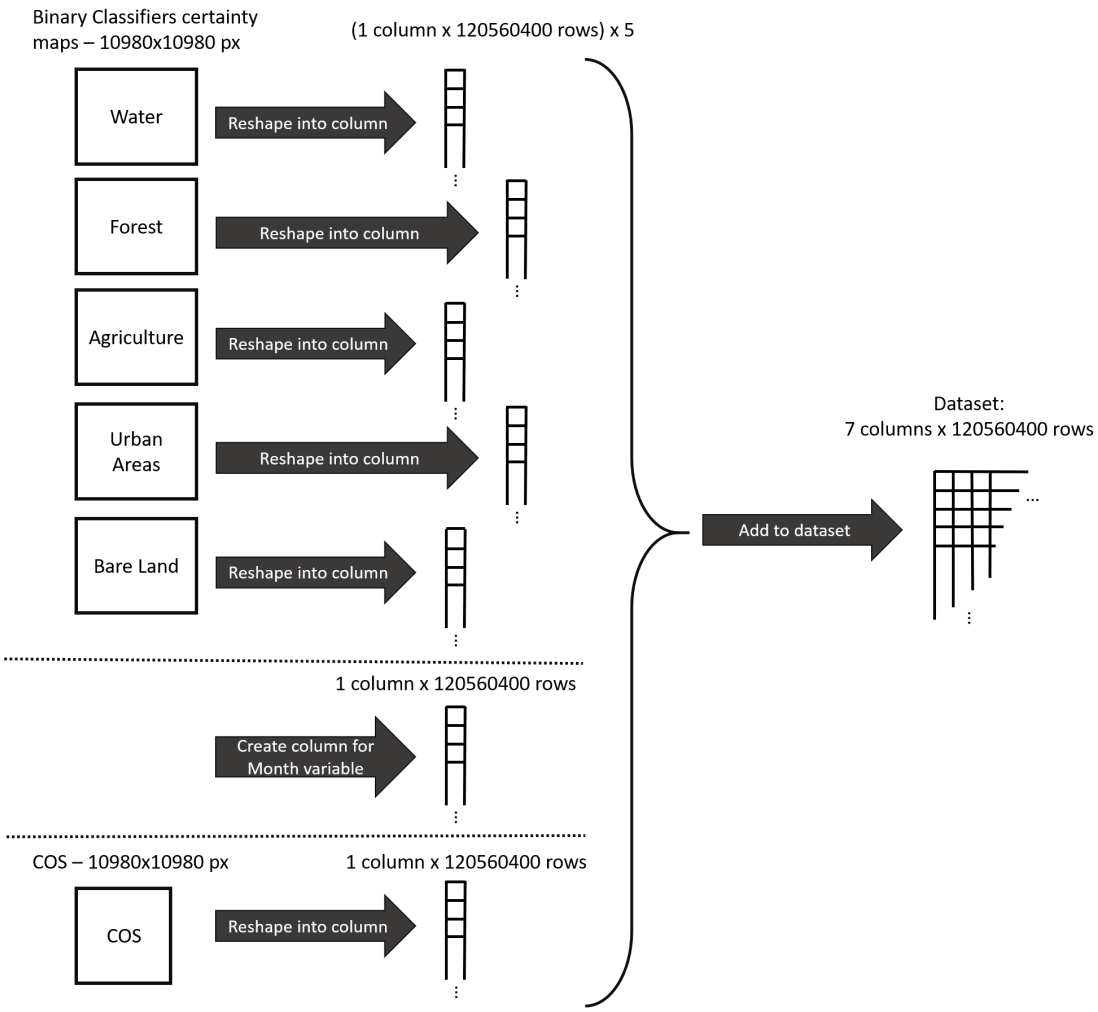


Figure 3.11: Steps to build the dataset of a single month for the final classifier.

for variants "2019" and "2019_MonthVar") and the resulting certainty maps were stored to build each dataset.

Since there are five different classes each single dataset will have five columns, along with an extra column indicating the month at which the input was taken, and a final response column with COS values, totalling seven columns. The construction of a single dataset, for a single month, can be seen in figure 3.11.

Then for the development of the training dataset a sample was taken from each of these datasets. The number of pixels used for each of the four variants is the same as previously stated and can be seen in tables 3.3 for "allMonths" and "MonthVar" where every month is used and the amount of pixels for February is doubled, and 3.4 for "2019" and "2019_MonthVar" where only 2019's months are used.

Unlike in the previous case variants "MonthVar" and "2019_MonthVar" do not indicate that a column for month variable was added, since it is already present in all datasets, instead they specify the variant of the binary classifiers used. This means that a final classifier of specific variant was developed based on the results of the group of

binary classifiers of that same variant. A diagram showing the process of training the final classifier based on the previous classifiers is displayed in figure 3.12.

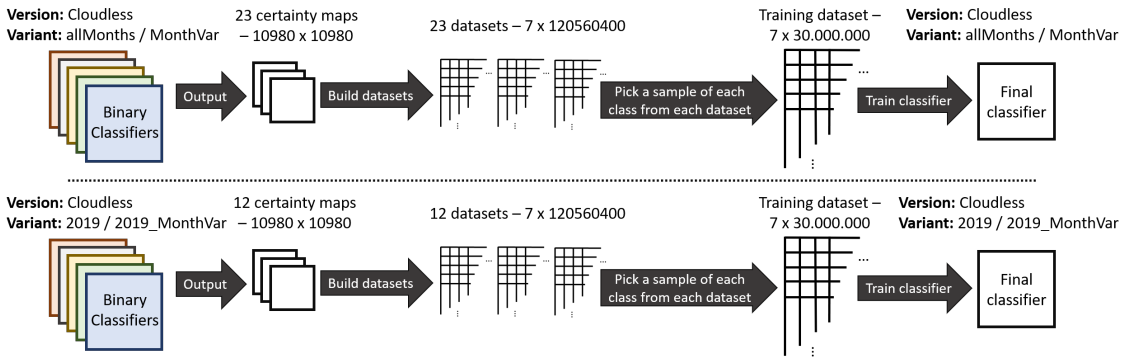


Figure 3.12: Process of building a training dataset from the developed binary classifiers to train a final classifier.

Contrary the previous classifiers, the final decision classifier is multi-class and can have as an output a number between 0 and 4 that corresponds to one of the five considered classes, which are, in order: water (0), forest (1), agriculture (2), urban areas (3) and bare land (4).

Chapter 4

Results

In the present chapter, the results obtained from the various classifiers developed will be presented and analysed. These results will be compared between the various metrics chosen to assess performance, between all versions and variants, and also between the binary classifiers and the final decision classifier. Thus, making it possible to draw some conclusions about each classifier.

4.1 Metrics for Classification Assessment

For a correct assessment of a classifier's performance, the choice of metrics is crucial and factors such as dataset balance should be considered. Metrics commonly found in the literature include accuracy and its variants such as overall accuracy, producer accuracy and user accuracy, F_1-score, sensitivity, specificity, among others.

However some of these metrics, namely accuracy and F_1-score, are sensitive to unbalanced data and their results can be overoptimistic, not always reflecting the true performance of the classifier. This is important since the dataset used for this study is not perfectly balanced between the five classes considered. Urban areas and bare land are notably unbalanced compared to the remaining classes as it can be seen in table 4.1.

Due to this imbalance, an effort was put into finding other metrics that would not be affected by this problem. There is a myriad of different metrics in the field of machine learning and the ones used in the present work were based on the confusion matrix built with the output of each classifier. A general confusion matrix for a binary classification problem is presented in figure 4.1.

At the top the "Actual Class" can be seen which is the ground truth, in this case it is the binarised COS classes, and on the left, there is the "Predicted Class" which is what the developed classifier predicted for that instance. When both the actual class and

Table 4.1: List of final classes considered and their cover percentage.

Label	Type of land	Total number of pixels	Percentage of pixels
0	Water	15.640.239	13,0%
1	Forest	58.847.943	48,8%
2	Agriculture	42.864.890	35,6%
3	Urban Area	2.961.625	2,5%
4	Bare Land	245.703	0,2%
	Total:	**120.560.400**	**100%**

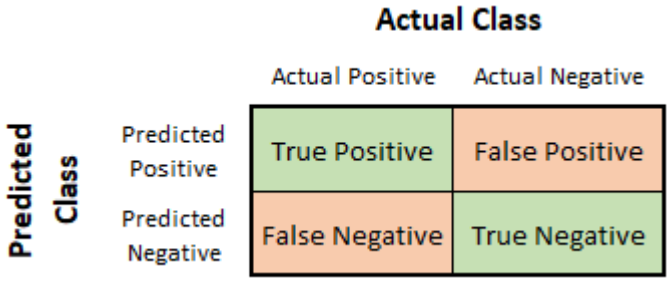


Figure 4.1: Binary confusion matrix.

the prediction from the classifier are positive (value 0 in this case) then it is called a True Positive (TP), similarly, when both the ground truth and the prediction are negative (value 1) then it is a True Negative (TN). When the true class is positive but the classifier mistakenly identifies it as negative it is called a False Negative (FN) and when the opposite happens, the real class is negative and prediction is positive, then it is a False Positive (FP).

To test if a metric is affected by data imbalance, its formula and the confusion matrix must be analysed to understand how this imbalance impacts it. Data imbalance means that there will more actual positives than actual negatives or vice-versa. To represent this a constant α is multiplied by the values of either column. Then the formula is once again analysed with this new constant in place. If both formulas (with and without the constant) lead to the same result then it can be said that the metric is not affected by unbalanced data, however, if that is not the case then the constant (the unbalance between positive and negative classes) had an impact on the result and it should be interpreted with caution (Tharwat (2020)).

For example, by analysing accuracy's formula $Acc = \frac{TP+TN}{TP+TN+FP+FN}$, and applying the constant α to the actual positives (TP and FN) this will result in $Acc' = \frac{\alpha TP+TN}{\alpha TP+TN+FP+\alpha FN}$. These two formulas are different and cannot be solved in any way that would make them equivalent, therefore proving that accuracy is a metric affected by data distribution (Tharwat (2020)). In contrast, if the same process is applied to sensitivity's formula $sensitivity = \frac{TP}{TP+FN}$ the resulting formula $sensitivity' = \frac{\alpha TP}{\alpha TP+\alpha FN} = \frac{\cancel{\alpha} TP}{\cancel{\alpha}(TP+FN)} = \frac{TP}{TP+FN}$ leads to the same result as the original, which means it is not affected by unbalanced data.

When the classifier is not binary, as it is the case of the Final Decision classifier developed, the confusion matrix will be different and metrics cannot be directly calculated through it. For the sake of simplicity, a confusion matrix of 3 variables (the Final Decision classifier has 5) is presented on figure 4.2.

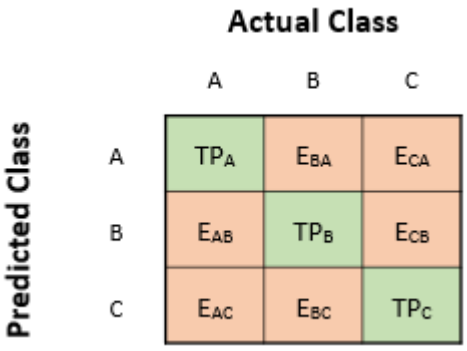


Figure 4.2: Multi-class confusion matrix. Based on (Tharwat (2020)).

The TPs can easily be identified, however, the same cannot be said for TN, FP and FN. To calculate the metrics for this kind of problem each class is analysed separately against the remaining, like if it was a binary classifier. For example, considering A the positive class, all others will be considered negative and FP will be any of the other classes mistakenly identified as A, that is E_{BA} and E_{CA}. Similarly, FN will be any instance of class A that was predicted as any of the others, meaning E_{AB} and E_{AC}. For the TN we considered the remaining cells which are TP_B, TP_C, E_{CB} and E_{BC}. In the end values are grouped and summed as seen in figure 4.3. With these values calculated the metrics used can be computed with ease. This process is repeated for each class.

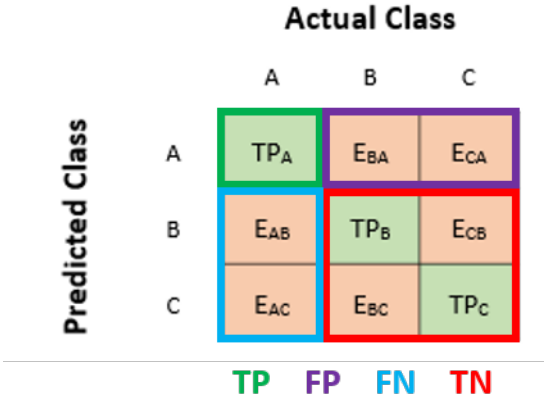


Figure 4.3: By grouping and summing the values inside the coloured rectangles a multi-class problem can be treated as a simpler binary problem.

4.1.1 Metrics Chosen

Eight different metrics were chosen to evaluate the performance of the developed classifiers. They are True Positive Rate (TPR), True Negative Rate (TNR), Positive Predictive Value (PPV), Negative Predictive Value (NPV), accuracy, Balanced Accuracy (BA), F_1-score and Matthews Correlation Coefficient (MCC). Each one will be described, mentioning if they are sensitive to data distribution, which was assessed using the process described previously.

Sensitivity Also called True Positive Rate (TPR), Recall or Probability of Detection. This metric is calculated by dividing the number of true positives by all the actual positives, indicating the percentage of positive pixels correctly identified.

- Formula: $TPR = \frac{TP}{TP+FN}$;
- Domain: $[0, 1]$;
- Data imbalance: insensitive.

Specificity Also known as True Negative Rate (TNR) and Selectivity. It is very similar to TPR but it refers to the negative values instead of positive ones. Indicates the percentage of negative pixels correctly identified.

- Formula: $TNR = \frac{TN}{TN+FP}$;
- Domain: $[0, 1]$;
- Data imbalance: insensitive.

Precision Or Positive Predictive Value (PPV) is calculated by dividing the true positives by all the predicted positives indicating how much the classifier got right when declaring a pixel as positive. This can tell the user how much to trust the classifier when it identifies the result as positive.

- Formula: $PPV = \frac{TP}{TP+FP}$;
- Domain: $[0, 1]$;
- Data imbalance: sensitive.

Negative Predictive Value (NPV) This metric is the negative version of PPV, telling the user how much the classifier can be trusted when deeming a pixel as negative.

- Formula: $NPV = \frac{TN}{TN+FN}$;
- Domain: $[0, 1]$;
- Data imbalance: sensitive.

Accuracy Very popular metric that takes into account both positive and negative samples, summing true positives and true negatives and then dividing by total population. It can estimate how well the classifier performed overall.

- Formula: $Acc = \frac{TP+TN}{TP+TN+FP+FN}$;
- Domain: [0, 1];
- Data imbalance: sensitive.

Balanced Accuracy (BA) This metric aims to give a better estimation of the classifier performance than traditional accuracy (Brodersen et al. (2010)). Because the latter is sensitive to unbalanced data, the former attempts to show a more appropriate estimate for situations where the data distribution is not perfect (Brodersen et al. (2010)).

- Formula: $BA = \frac{TPR+TNR}{2} = \frac{1}{2} \cdot \left(\frac{TP}{TP+FN} + \frac{TN}{TN+FP} \right)$
- Domain: [0, 1];
- Data imbalance: insensitive.

F_1-score Very commonly used metric that represents the harmonic mean between precision and recall. It has been criticised for being independent of TN which makes it not reliable for unbalanced datasets (Chicco and Jurman (2020)).

- Formula: $F_1\text{-}score = 2 \cdot \frac{PPV \times TPR}{PPV+TPR} = \frac{2TP}{2TP+FP+FN}$
- Domain: [0, 1];
- Data imbalance: sensitive.

Matthews Correlation Coefficient (MCC) Considered to be a better alternative to F_1-score, this metric uses all the cells of the confusion matrix (Chicco and Jurman (2020)). Its domain is slightly different from the other metrics, with 0 meaning the classification was no better than a random classifier and -1 and 1 meaning perfect misclassification and perfect classification, respectively (Chicco and Jurman (2020)).

- Formula: $MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}}$
- Domain: [-1, 1];
- Data imbalance: insensitive.

4.2 Classifier Results

Using the monthly images collected to test the already trained classifiers, the results produced were then evaluated by calculating all eight metrics and registering them in a spreadsheet. Since there are 23 different months it would be difficult to analyse so many values and because of that, an auxiliary table was built with the minimum value a metric displayed, its maximum, and the average between all the months. In figure 4.4 and 4.5 the full table is shown along with the auxiliary table on the right. As it can be seen it is illegible, especially the first one that uses all months for testing, and so from now on only the auxiliary table will be displayed. The yellow highlighted metrics are insensible to data imbalance and should be more relied upon than the remaining.

	2018_03_31	2018_04_30	2018_05_10	2018_06_19	2018_07_29	2018_08_18	2018_09_27	2018_10_07	2018_11_16	2018_12_06	2019_01_25	2019_02_14	2019_03_16	2019_04_30	2019_05_15	2019_06_09	2019_07_24	2019_08_23	2019_09_12	2019_10_22	2019_11_16	2019_12_06	2020_01_10	Min	Max	Average
TPR	89,43%	81,64%	94,63%	94,46%	94,16%	94,01%	92,52%	92,60%	92,94%	92,99%	92,05%	93,59%	92,78%	93,72%	93,75%	91,84%	92,42%	92,18%	91,86%	91,27%	89,89%	90,76%	91,75%	81,64%	94,63%	92,05%
TNR	99,18%	98,97%	99,78%	99,41%	99,80%	99,88%	99,83%	99,91%	99,79%	99,68%	99,78%	99,63%	99,78%	99,84%	99,74%	99,56%	99,87%	99,75%	99,88%	99,71%	99,92%	99,87%	99,80%	98,97%	99,92%	99,71%
PPV	94,21%	92,16%	98,48%	95,97%	98,63%	99,15%	98,80%	99,33%	98,49%	97,73%	98,41%	97,42%	98,40%	98,85%	98,18%	96,91%	99,06%	98,19%	99,10%	97,94%	99,40%	99,05%	98,58%	92,16%	99,40%	97,93%
NPV	98,44%	97,31%	99,20%	99,18%	99,14%	99,11%	98,90%	98,91%	98,96%	98,96%	98,83%	99,05%	98,93%	99,07%	99,08%	98,79%	98,88%	98,84%	98,80%	98,71%	98,51%	98,64%	98,78%	97,31%	99,20%	98,83%
Acc	97,92%	96,72%	99,11%	98,77%	99,07%	99,12%	98,88%	98,96%	98,90%	98,81%	98,78%	98,85%	98,87%	99,04%	98,96%	98,56%	98,90%	98,77%	98,84%	98,62%	98,62%	98,69%	98,76%	96,72%	99,12%	98,72%
BA	94,31%	90,30%	97,21%	96,93%	96,98%	96,95%	96,18%	96,25%	96,36%	96,33%	95,92%	96,61%	96,28%	96,78%	96,75%	95,70%	96,14%	95,96%	95,87%	95,49%	94,91%	95,32%	95,77%	90,30%	97,21%	95,88%
F1_score	91,76%	86,59%	96,52%	95,21%	96,34%	96,51%	95,55%	95,85%	95,63%	95,30%	95,13%	95,47%	95,51%	96,22%	95,92%	94,31%	95,63%	95,09%	95,34%	94,48%	94,41%	94,72%	95,04%	86,59%	96,52%	94,89%
MCC	90,60%	84,93%	96,04%	94,50%	95,84%	96,05%	94,98%	95,33%	95,05%	94,66%	94,50%	94,83%	94,92%	95,71%	95,36%	93,53%	95,07%	94,45%	94,77%	93,77%	93,78%	94,09%	94,41%	84,93%	96,05%	94,23%

Figure 4.4: Full spreadsheet from water classifier, "Clouds" version, variant "allMonths".

	2018_03_31	2018_04_30	2018_05_10	2018_06_19	2018_07_29	2018_08_18	2018_09_27	2018_10_07	2018_11_16	2018_12_06	2020_01_10	Min	Max	Average
TPR	90,33%	82,16%	94,91%	94,76%	94,30%	94,24%	93,28%	94,04%	93,83%	93,85%	92,61%	82,16%	94,91%	92,57%
TNR	97,56%	95,44%	99,71%	99,33%	99,85%	99,88%	99,70%	96,85%	99,46%	99,39%	99,67%	95,44%	99,88%	98,80%
PPV	84,68%	72,85%	98,02%	95,47%	98,91%	99,12%	97,90%	81,65%	96,27%	95,85%	97,68%	72,85%	99,12%	92,58%
NPV	98,54%	97,29%	99,25%	99,22%	99,16%	99,15%	99,01%	99,09%	99,08%	99,09%	98,91%	97,29%	99,25%	98,89%
Acc	96,63%	93,71%	99,09%	98,74%	99,13%	99,14%	98,87%	96,49%	98,73%	98,67%	98,76%	93,71%	99,14%	98,00%
BA	93,95%	88,80%	97,31%	97,04%	97,07%	97,06%	96,49%	95,45%	96,64%	96,62%	96,14%	88,80%	97,31%	95,69%
F1_score	87,41%	77,22%	96,44%	95,11%	96,55%	96,62%	95,53%	87,41%	95,03%	94,84%	95,08%	77,22%	96,62%	92,48%
MCC	85,53%	73,77%	95,94%	94,39%	96,09%	96,17%	94,92%	85,67%	94,32%	94,09%	94,41%	73,77%	96,17%	91,39%

Figure 4.5: Full spreadsheet from water classifier, "Clouds" version, variant "2019".

4.2.1 Water Classifier Results

Here the results of the various versions and variants of the water classifier are displayed. Figure 4.6 shows the results of the version "Clouds" where no filter was applied, and figure 4.7 displays the result for the "Cloudless" version. Inside each of these figures, there are four subfigures presenting the results for each of the four variants.

By analysing both figures we can perceive a general decrease in the classifier performance for variants "2019" and "2019_MonthVar" when compared to "allMonths" and "MonthVar". This was expected since the classifier is testing data that it did not train with, therefore causing the results to be worse, but more realistic to a real scenario.

	Min	Max	Average
TPR	81,64%	94,63%	92,05%
TNR	98,97%	99,92%	99,71%
PPV	92,16%	99,40%	97,93%
NPV	97,31%	99,20%	98,83%
Acc	96,72%	99,12%	98,72%
BA	90,30%	97,21%	95,88%
F1_score	86,59%	96,52%	94,89%
MCC	84,93%	96,05%	94,23%

(a) allMonths

	Min	Max	Average
TPR	81,19%	93,66%	91,51%
TNR	99,54%	99,92%	99,81%
PPV	96,78%	99,43%	98,60%
NPV	97,27%	99,06%	98,75%
Acc	97,33%	99,07%	98,73%
BA	90,46%	96,75%	95,66%
F1_score	88,74%	96,30%	94,90%
MCC	87,73%	95,82%	94,28%

(b) MonthVar

	Min	Max	Average
TPR	82,16%	94,91%	92,57%
TNR	95,44%	99,88%	98,80%
PPV	72,85%	99,12%	92,58%
NPV	97,29%	99,25%	98,89%
Acc	93,71%	99,14%	98,00%
BA	88,80%	97,31%	95,69%
F1_score	77,22%	96,62%	92,48%
MCC	73,77%	96,17%	91,39%

(c) 2019

	Min	Max	Average
TPR	82,49%	94,80%	92,72%
TNR	93,85%	99,89%	98,57%
PPV	66,65%	99,25%	91,56%
NPV	97,29%	99,23%	98,90%
Acc	92,37%	99,16%	97,81%
BA	88,17%	97,22%	95,64%
F1_score	73,73%	96,68%	91,96%
MCC	69,87%	96,24%	90,81%

(d) 2019_MonthVar

Figure 4.6: Results of water classifier, Clouds version.

	Min	Max	Average
TPR	80,97%	93,92%	91,23%
TNR	99,34%	99,92%	99,80%
PPV	94,81%	99,45%	98,52%
NPV	97,22%	99,10%	98,71%
Acc	96,96%	99,09%	98,68%
BA	90,15%	96,89%	95,51%
F1_score	87,34%	96,40%	94,71%
MCC	85,97%	95,93%	94,07%

(a) allMonths

	Min	Max	Average
TPR	81,76%	93,91%	91,57%
TNR	99,45%	99,94%	99,81%
PPV	95,65%	99,56%	98,65%
NPV	97,34%	99,10%	98,76%
Acc	97,15%	99,08%	98,74%
BA	90,61%	96,75%	95,69%
F1_score	88,16%	96,33%	94,96%
MCC	86,90%	95,87%	94,34%

(b) MonthVar

	Min	Max	Average
TPR	81,94%	95,94%	93,04%
TNR	60,68%	99,86%	95,34%
PPV	26,67%	99,01%	86,62%
NPV	97,24%	99,30%	98,92%
Acc	65,25%	99,16%	95,04%
BA	78,31%	97,49%	94,19%
F1_score	41,74%	96,71%	87,98%
MCC	38,13%	96,26%	86,66%

(c) 2019

	Min	Max	Average
TPR	82,02%	95,20%	92,90%
TNR	92,13%	99,89%	97,75%
PPV	60,84%	99,21%	88,26%
NPV	97,17%	99,28%	98,92%
Acc	90,82%	99,17%	97,12%
BA	87,08%	97,31%	95,32%
F1_score	69,86%	96,73%	90,06%
MCC	65,59%	96,29%	88,75%

(d) 2019_MonthVar

Figure 4.7: Results of water classifier, Cloudless version.

The addition of the month variable had mixed results. When looking at accuracy, BA, F_1-score and MCC it is possible to observe that in the "Clouds" version (figure 4.6) it had almost no impact on "MonthVar" and a slight decrease on "2019_MonthVar", while on "Cloudless" (figure 4.7) it had a very small increase on "MonthVar"and slightly bigger, but still small, increase on "2019_MonthVar". On any of these cases, the change is almost insignificant but still important to be reported. It is also important to stress that there is always an element of random when creating the training classifier and thus the results will always vary slightly independently of new variables that might be added.

The results of this classifier are considered to be good overall. The averages registered present high values and even the minimum values are still high and reliable for the most part. The exception is figure 4.7(c) where one month registered a very low PPV, bringing F_1-score and MCC down as well. However, it can be noticed that the average of these values is still high, meaning this result might have been just a fluke.

4.2.2 Forest Classifier Results

The results from the forest classifier are displayed in figure 4.8 and 4.9, which correspond to "Clouds" and "Cloudless" version, respectively, and have four subfigures each concerning each of the four variants inside those versions.

	Min	Max	Average
TPR	48,97%	86,83%	72,44%
TNR	58,48%	88,31%	75,29%
PPV	61,21%	81,07%	73,91%
NPV	59,13%	84,60%	74,89%
Acc	59,94%	78,53%	73,90%
BA	59,69%	78,44%	73,87%
F1_score	54,41%	79,16%	72,72%
MCC	19,85%	57,20%	48,26%

(a) allMonths

	Min	Max	Average
TPR	68,52%	90,77%	80,68%
TNR	48,79%	81,96%	69,18%
PPV	61,52%	80,11%	71,91%
NPV	66,33%	84,71%	79,23%
Acc	63,71%	79,15%	74,79%
BA	63,82%	79,08%	74,93%
F1_score	64,83%	79,20%	75,80%
MCC	27,74%	58,29%	50,49%

(b) MonthVar

	Min	Max	Average
TPR	30,63%	85,36%	68,84%
TNR	53,04%	88,03%	73,64%
PPV	63,41%	80,02%	71,60%
NPV	56,26%	82,49%	73,36%
Acc	58,50%	78,47%	71,30%
BA	57,85%	78,50%	71,24%
F1_score	41,88%	78,95%	68,67%
MCC	18,76%	56,98%	43,67%

(c) 2019

	Min	Max	Average
TPR	36,71%	90,65%	70,29%
TNR	54,39%	89,13%	73,38%
PPV	65,46%	83,99%	72,25%
NPV	58,27%	85,92%	74,20%
Acc	61,05%	77,95%	71,87%
BA	60,49%	77,90%	71,84%
F1_score	47,92%	77,35%	69,94%
MCC	23,91%	55,87%	45,02%

(d) 2019_MonthVar

Figure 4.8: Results of forest classifier, Clouds version.

Immediately, it is apparent that the results are lower than the ones obtained for the water classifier. This is mainly due to the spectral similarities to class Agriculture since both concern vegetation and can get easily mistakenly identified as each other. This class,

	Min	Max	Average
TPR	52,89%	86,65%	73,92%
TNR	58,79%	84,02%	73,97%
PPV	59,40%	81,12%	73,26%
NPV	59,32%	84,48%	75,33%
Acc	59,36%	78,62%	73,95%
BA	59,21%	78,53%	73,94%
F1_score	55,95%	79,19%	73,30%
MCC	18,57%	57,34%	48,24%

(a) allMonths

	Min	Max	Average
TPR	49,67%	91,14%	80,21%
TNR	48,29%	80,82%	69,40%
PPV	62,70%	79,11%	71,86%
NPV	61,97%	85,11%	79,38%
Acc	64,28%	78,74%	74,67%
BA	63,95%	78,72%	74,80%
F1_score	57,58%	78,96%	75,41%
MCC	29,16%	57,44%	50,40%

(b) MonthVar

	Min	Max	Average
TPR	27,80%	85,39%	68,62%
TNR	53,24%	88,58%	72,97%
PPV	62,97%	76,72%	71,23%
NPV	56,27%	82,53%	73,23%
Acc	58,91%	78,41%	70,85%
BA	58,19%	78,45%	70,79%
F1_score	39,78%	78,94%	68,09%
MCC	20,70%	56,90%	42,95%

(c) 2019

	Min	Max	Average
TPR	36,82%	90,63%	70,32%
TNR	54,45%	89,08%	73,31%
PPV	65,48%	83,94%	72,20%
NPV	58,29%	85,90%	74,20%
Acc	61,06%	77,93%	71,85%
BA	60,50%	77,89%	71,82%
F1_score	48,00%	77,38%	69,94%
MCC	23,91%	55,83%	44,98%

(d) 2019_MonthVar

Figure 4.9: Results of forest classifier, Cloudless version.

and Agriculture too, also goes through more changes spectrally throughout the year than the water class, which also makes rule creation more difficult and less accurate.

Like the results of the water classifier, these results also suffered a decrease, on both "Clouds" and "Cloudless" version, when using only 2019's months to train instead of all the 23 months available. However, unlike the water classifier, adding the month variable to the training dataset had a positive impact on the results. Albeit still small it helped to correctly identify more forest land, as it can be seen by the increase of TPR by about 7-10% on both versions, that consequently also rose MCC by roughly 2%.

4.2.3 Agriculture Classifier Results

The results for the agriculture classifier can be seen on figures 4.10 (without cloud filtering) and 4.11 (cloudless). These results are very similar to the forest classifier results, although lower in some instances.

This classifier has similar problems already described in the previous subsection. It tries to identify land that is very similar to the forest classifier and the spectral values of each pixel vary a lot throughout the year, especially in this case where crops are more frequently harvested and planted than in the forest case. This classifier also has a smaller sample size than the forest classifier (about 35,6%, for agriculture class, and 48,8%, for forest class, of the total dataset, as seen on table 4.1) which can contribute to slightly worse results.

As expected variants "2019" and "2019_MonthVar" were as good as "allMonths" and "MonthVar" and suffered a decrease of about 1% to 4%, depending on the metric.

	Min	Max	Average
TPR	38,16%	80,11%	57,37%
TNR	71,39%	92,15%	84,36%
PPV	49,55%	76,53%	67,59%
NPV	70,72%	86,68%	78,51%
Acc	64,16%	78,28%	74,77%
BA	59,46%	76,73%	70,87%
F1_score	46,13%	70,08%	61,35%
MCC	19,58%	52,08%	43,79%

(a) allMonths

	Min	Max	Average
TPR	35,70%	78,22%	59,23%
TNR	73,83%	91,27%	84,10%
PPV	52,94%	76,17%	67,41%
NPV	70,83%	86,00%	79,25%
Acc	66,02%	78,26%	75,26%
BA	60,19%	76,69%	71,67%
F1_score	45,55%	70,00%	62,42%
MCC	22,00%	52,43%	44,89%

(b) MonthVar

	Min	Max	Average
TPR	35,52%	71,80%	54,02%
TNR	71,19%	95,08%	83,07%
PPV	46,89%	79,93%	65,60%
NPV	70,54%	83,01%	76,90%
Acc	62,27%	78,08%	72,74%
BA	58,65%	75,74%	68,55%
F1_score	46,50%	68,71%	57,95%
MCC	17,37%	51,67%	39,51%

(c) 2019

	Min	Max	Average
TPR	38,07%	71,48%	59,44%
TNR	70,78%	88,93%	81,50%
PPV	48,99%	74,95%	64,08%
NPV	70,77%	83,78%	78,68%
Acc	63,80%	78,66%	73,66%
BA	59,38%	76,39%	70,47%
F1_score	46,41%	69,61%	61,24%
MCC	19,25%	52,19%	41,80%

(d) 2019_MonthVar

Figure 4.10: Results of agriculture classifier, Clouds version.

	Min	Max	Average
TPR	34,58%	76,10%	54,42%
TNR	75,91%	92,54%	86,09%
PPV	51,26%	77,63%	68,80%
NPV	70,47%	85,20%	77,70%
Acc	65,14%	78,04%	74,83%
BA	59,50%	76,13%	70,25%
F1_score	44,91%	69,33%	60,02%
MCC	20,32%	51,68%	43,30%

(a) allMonths

	Min	Max	Average
TPR	33,34%	76,14%	57,78%
TNR	75,95%	92,57%	84,94%
PPV	54,16%	77,28%	68,07%
NPV	70,57%	85,23%	78,88%
Acc	66,51%	78,44%	75,29%
BA	60,05%	76,12%	71,36%
F1_score	44,47%	69,30%	61,70%
MCC	22,30%	51,96%	44,68%

(b) MonthVar

	Min	Max	Average
TPR	38,14%	74,06%	59,47%
TNR	68,56%	89,50%	79,70%
PPV	45,70%	74,29%	62,44%
NPV	70,33%	83,07%	78,26%
Acc	61,29%	78,49%	72,51%
BA	58,13%	76,39%	69,58%
F1_score	46,44%	69,55%	60,41%
MCC	16,15%	52,93%	39,88%

(c) 2019

	Min	Max	Average
TPR	38,41%	71,55%	59,28%
TNR	70,86%	88,71%	81,73%
PPV	49,10%	74,72%	64,28%
NPV	70,65%	83,74%	78,67%
Acc	63,88%	78,68%	73,75%
BA	59,29%	76,28%	70,50%
F1_score	46,06%	69,45%	61,25%
MCC	19,15%	52,28%	41,92%

(d) 2019_MonthVar

Figure 4.11: Results of agriculture classifier, Cloudless version.

The addition of the month variable did not have as much benefit in this case when compared to the forest classifier. The month variable still helped to increase the result on all variants, however, it was not as significant as in the forest classifier. The reason for this probably stems from having more variation throughout the year than the forest, with some crops reaching its mature stage and being consequently harvested at different times of the year.

4.2.4 Urban Areas Classifier Results

The results for the urban areas classifier are on figures 4.12 and 4.13, for "Clouds" and "Cloudless", respectively.

	Min	Max	Average
TPR	2,97%	21,47%	12,01%
TNR	99,59%	99,96%	99,79%
PPV	24,70%	82,96%	61,13%
NPV	97,61%	98,05%	97,83%
Acc	97,37%	97,75%	97,64%
BA	51,37%	60,57%	55,90%
F1_score	5,31%	31,90%	19,67%
MCC	7,83%	35,64%	25,90%

(a) allMonths

	Min	Max	Average
TPR	1,90%	14,16%	7,77%
TNR	99,69%	99,98%	99,88%
PPV	27,09%	87,67%	64,44%
NPV	97,59%	97,88%	97,73%
Acc	97,46%	97,68%	97,62%
BA	50,89%	56,93%	53,82%
F1_score	3,55%	22,58%	13,53%
MCC	6,61%	27,58%	21,15%

(b) MonthVar

	Min	Max	Average
TPR	6,05%	35,54%	17,21%
TNR	93,87%	99,91%	98,94%
PPV	12,05%	73,24%	44,88%
NPV	97,66%	98,30%	97,94%
Acc	92,43%	97,79%	96,94%
BA	52,47%	64,70%	58,08%
F1_score	8,06%	35,45%	22,19%
MCC	6,93%	36,98%	24,82%

(c) 2019

	Min	Max	Average
TPR	3,70%	32,25%	12,81%
TNR	94,64%	99,96%	99,26%
PPV	13,17%	82,24%	49,31%
NPV	97,63%	98,23%	97,84%
Acc	93,11%	97,71%	97,14%
BA	51,71%	63,45%	56,04%
F1_score	6,45%	26,92%	17,74%
MCC	8,82%	31,32%	22,24%

(d) 2019_MonthVar

Figure 4.12: Results of urban areas classifier, Clouds version.

These results are not so great, especially compared to the other classifiers seen until now. The percentage of urban areas in the original dataset is very small, only 2,5% (table 4.1). This is a tremendous imbalance in the training dataset and the decision tree will have trouble to learn from it, as shown in the results.

Another reason is the fact that Sentinel-2 lacks a Thermal Infrared (TIR) band. This band is very useful for detecting built-up land since it is normal for cities to have a higher temperature than forests, agriculture or water. Some indices could not be used due to the lack of this feature, namely the Enhanced Built-Up and Bareness Index (EBBI) and Dry Built-Up Index (DBI). These indices could enhance the detection of urban areas and increase the classifier's performance.

It is expected that the variants "allMonths" and "MonthVar" will yield better results than "2019" and "2019_MonthVar", respectively, however that is not the case in this

	Min	Max	Average
TPR	5,18%	25,52%	14,38%
TNR	99,31%	99,95%	99,74%
PPV	15,91%	82,72%	62,34%
NPV	97,65%	98,15%	97,88%
Acc	97,00%	97,81%	97,64%
BA	52,25%	62,50%	57,06%
F1_score	7,82%	34,97%	22,77%
MCC	7,81%	36,64%	28,57%

(a) allMonths

	Min	Max	Average
TPR	5,60%	18,34%	12,22%
TNR	99,32%	99,94%	99,79%
PPV	17,23%	78,51%	62,22%
NPV	97,66%	97,98%	97,83%
Acc	97,02%	97,75%	97,63%
BA	52,46%	58,96%	56,00%
F1_score	8,45%	27,85%	20,10%
MCC	8,56%	32,95%	26,47%

(b) MonthVar

	Min	Max	Average
TPR	6,47%	32,53%	17,21%
TNR	93,90%	99,93%	98,93%
PPV	11,80%	78,53%	44,78%
NPV	97,67%	98,22%	97,94%
Acc	92,39%	97,78%	96,92%
BA	52,62%	63,22%	58,07%
F1_score	8,36%	34,79%	22,39%
MCC	7,05%	36,51%	24,91%

(c) 2019

	Min	Max	Average
TPR	4,70%	34,80%	16,52%
TNR	93,88%	99,92%	98,96%
PPV	12,38%	72,54%	43,32%
NPV	97,64%	98,28%	97,92%
Acc	92,43%	97,75%	96,94%
BA	51,93%	64,34%	57,74%
F1_score	6,81%	32,54%	20,96%
MCC	6,22%	35,87%	23,60%

(d) 2019_MonthVar

Figure 4.13: Results of urban areas classifier, Cloudless version.

classifier. The results are unstable with some metrics indicating a better performance and others indicating worse. Since pixels are chosen at random when building the training dataset, and that the sample of urban areas is already so minimal it could be more heavily affecting the results than if it was a bigger sample. The classifier already has trouble learning from such an unbalanced dataset and the randomness could be selecting "bad" pixels that do not reflect as well the spectral signature of built-up land, making it worse.

It is also possible to observe that when adding the month variable the outcomes are worse. Built-up land spectral reflectance typically does not change much along the year, as vegetation does. So this new variable could be adding redundant information that with such imbalanced data distribution would make it even harder to create accurate rules and an acceptable decision tree.

4.2.5 Bare Land Classifier Results

The bare land classifier has similar problems to the urban areas classifier, since these classes are very similar to each other. Its results can be seen in figures 4.14, for "Clouds" version, and 4.15, for "Cloudless" version.

Similarly to the urban areas classifier, this classifier would also benefit from the readings of a TIR band since, likewise, bare areas typically absorb more heat than surrounding classes, in most cases, vegetation or water. Because of this, the presence of a TIR spectral band would by itself benefit the detection of this class, and it would also allow the computation of additional indices such as EBBI and Normalised Difference Bareness Index (NDBaI).

	Min	Max	Average
TPR	7,80%	20,50%	14,88%
TNR	99,96%	99,99%	99,98%
PPV	46,36%	73,92%	61,14%
NPV	99,81%	99,84%	99,83%
Acc	99,79%	99,82%	99,81%
BA	53,90%	60,23%	57,43%
F1_score	13,81%	29,84%	23,64%
MCC	21,60%	33,51%	29,80%

(a) allMonths

	Min	Max	Average
TPR	0,00%	18,24%	12,45%
TNR	99,96%	100,0%	99,98%
PPV	0,00%	78,96%	58,35%
NPV	99,80%	99,83%	99,82%
Acc	99,79%	99,82%	99,80%
BA	50,00%	59,10%	56,22%
F1_score	12,88%	27,15%	21,13%
MCC	0,00%	33,38%	26,55%

(b) MonthVar

	Min	Max	Average
TPR	1,13%	20,28%	14,39%
TNR	99,67%	100,0%	99,95%
PPV	8,70%	70,98%	53,73%
NPV	99,80%	99,84%	99,83%
Acc	99,50%	99,81%	99,78%
BA	50,56%	60,13%	57,17%
F1_score	2,20%	30,59%	21,48%
MCC	7,08%	35,45%	26,59%

(c) 2019

	Min	Max	Average
TPR	0,18%	18,65%	13,45%
TNR	97,02%	100,0%	99,70%
PPV	0,97%	71,91%	46,97%
NPV	99,80%	99,83%	99,82%
Acc	96,85%	99,81%	99,52%
BA	50,09%	59,31%	56,57%
F1_score	0,37%	27,95%	18,54%
MCC	2,99%	32,17%	22,70%

(d) 2019_MonthVar

Figure 4.14: Results of bare land classifier, Clouds version.

	Min	Max	Average
TPR	7,05%	18,95%	14,41%
TNR	99,07%	99,99%	99,94%
PPV	3,34%	79,62%	62,77%
NPV	99,81%	99,83%	99,83%
Acc	98,90%	99,82%	99,77%
BA	53,52%	59,46%	57,18%
F1_score	5,51%	28,74%	22,58%
MCC	6,85%	33,50%	29,36%

(a) allMonths

	Min	Max	Average
TPR	6,39%	19,06%	14,85%
TNR	99,07%	99,99%	99,94%
PPV	3,03%	76,35%	57,35%
NPV	99,81%	99,83%	99,83%
Acc	98,90%	99,82%	99,76%
BA	53,19%	59,51%	57,40%
F1_score	4,99%	28,23%	22,62%
MCC	6,15%	33,07%	28,34%

(b) MonthVar

	Min	Max	Average
TPR	3,85%	19,27%	14,67%
TNR	99,66%	100,0%	99,95%
PPV	9,16%	80,67%	60,16%
NPV	99,80%	99,84%	99,83%
Acc	99,49%	99,82%	99,78%
BA	51,92%	59,63%	57,31%
F1_score	7,34%	29,89%	22,21%
MCC	12,19%	35,74%	28,38%

(c) 2019

	Min	Max	Average
TPR	2,26%	18,13%	12,35%
TNR	99,96%	100,0%	99,98%
PPV	39,91%	74,00%	57,62%
NPV	99,80%	99,83%	99,82%
Acc	99,79%	99,81%	99,80%
BA	51,13%	59,05%	56,17%
F1_score	4,35%	28,04%	19,88%
MCC	11,71%	33,42%	25,92%

(d) 2019_MonthVar

Figure 4.15: Results of bare land classifier, Cloudless version.

The sample for this class is even smaller, being only 0,2% of the total dataset, as seen in table 4.1. As previously explained this represents an extremely data imbalance that affects the performance of the classifier due to being difficult to create accurate rules with so few samples.

However, unlike urban areas classifier, this classifier follows the expected decrease in performance when training with only 2019's months instead of the whole 23 months available.

Looking at how the addition of the month variable affected the results then we see a similarity to the urban areas classifier again. The performance decreases with this extra information, and just like the class of urban areas, bare land does not change a lot along the year and for this reason, an extra variable is adding redundancy which will make it harder to learn from the minute amount of samples present in the training dataset.

4.2.6 Final Decision Classifier Results

The final decision classifier is different from the previous ones since it is multi-class instead of binary and does not use satellite imagery and indices as its input, using instead certainty maps produced by the binary classifiers.

The confusion matrices for these classifiers can be seen in figures 4.16, 4.17, 4.18 and 4.19 where they display the total number of pixels that were correctly identified (diagonal cells with a blueish tone) and those who were mistakenly classified as other classes (the remaining cells painted with a cream/pink colour). These matrices were produced based on the confusion matrices from each test, which means that for variants "allMonths" and "MonthVar", that test with all the 23 months available, the final confusion matrix is built based on the output of each of those 23 months; and for variants "2019" and "2019_MonthVar", that only use 11 of the 23 months to test then the final confusion matrix will be based on the output produced by those 11 months. This can be noticed when comparing the numbers present in both matrices, since variants "2019" and "2019_MonthVar" have approximately half the amount of pixels in each cell when compared to variants "allMonths" and "MonthVar".

By examining them it is possible to observe one of the problems already mentioned, the classifier has trouble distinguishing between the forest and agriculture class as it is evident by the stronger colour (indicating higher values in these cells) of the cells corresponding to instances where forest was mistaken for agriculture and vice-versa.

However, these matrices should be examined with caution, as they may give the impression that the classifier is performing terrifically in forest identification as the colour of that cell is extremely vivid with a very high value. But that is not the case, this classifier presents similar results to the binary classifier previously presented, with its best ability residing in water detection. The high value and bright blue colour is due to the high number of pixels present in the dataset, almost half (table 4.1), so even if its identification is not perfect it will still be a very high and contrasting value compared to the other classes.

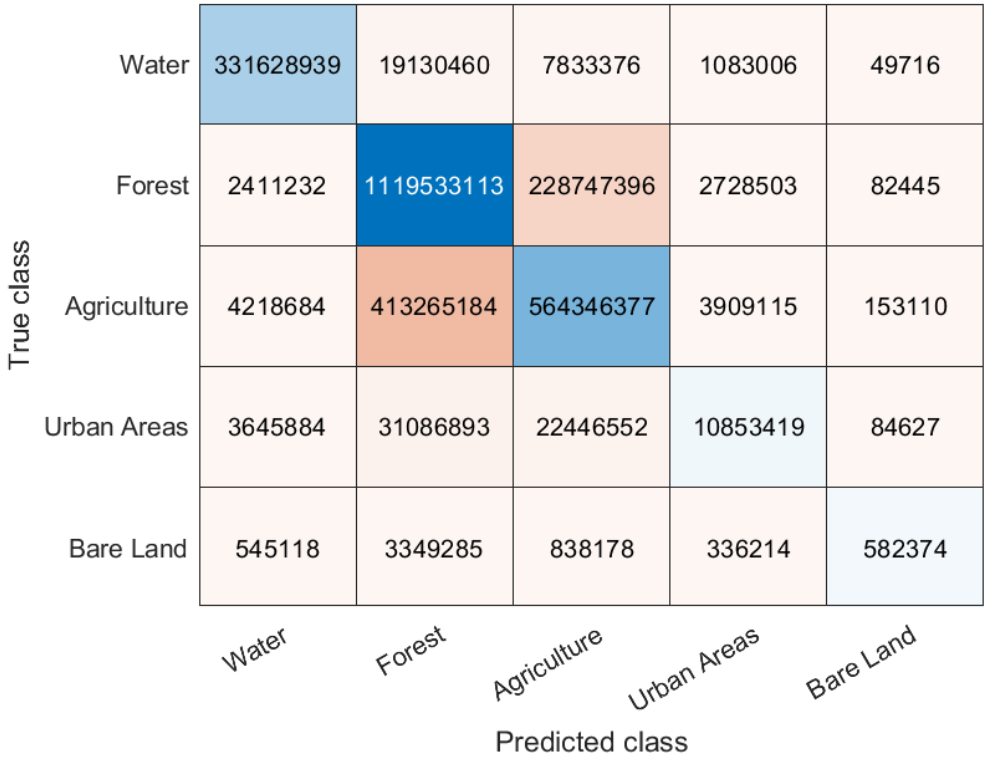


Figure 4.16: Confusion Matrix of all the results from Final classifier, variant "allMonths"

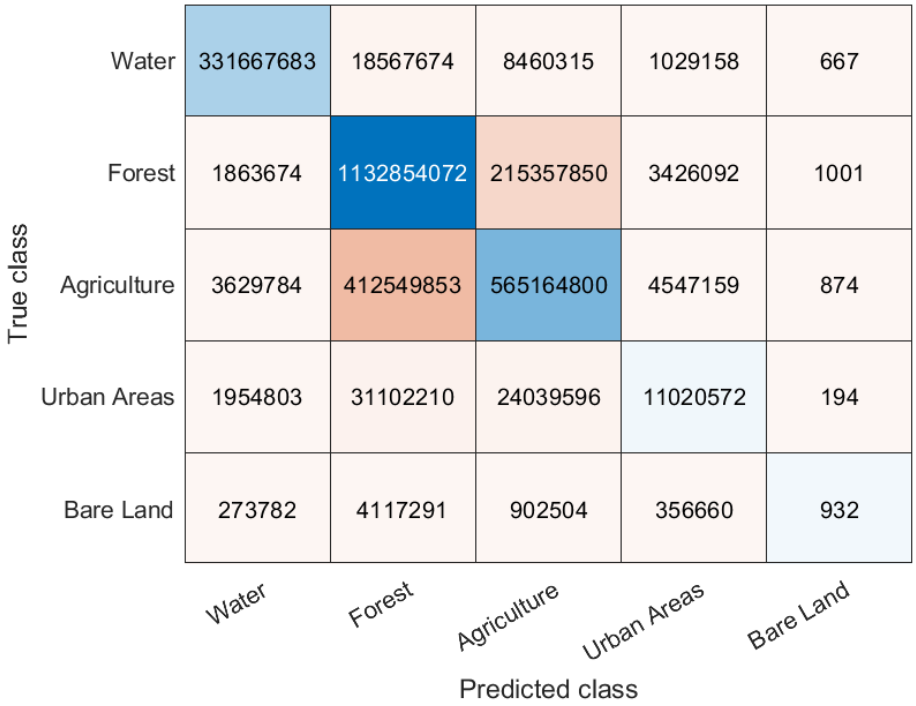


Figure 4.17: Confusion Matrix of all the results from Final classifier, variant "MonthVar"

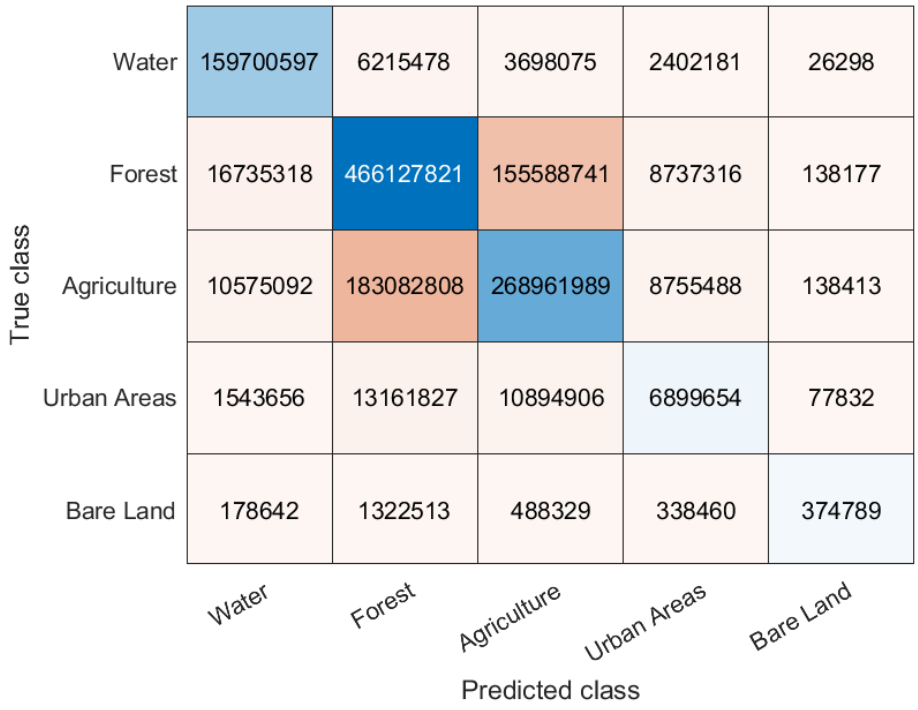


Figure 4.18: Confusion Matrix of all the results from Final classifier, variant "2019"

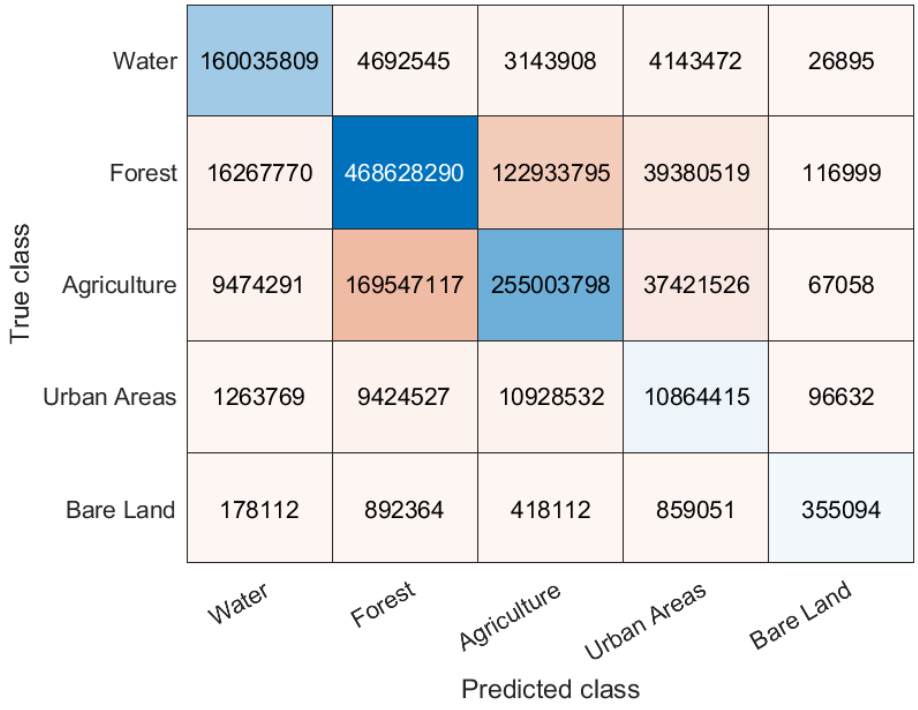


Figure 4.19: Confusion Matrix of all the results from Final classifier, variant "2019_MonthVar"

As stated before, extracting the values from the confusion matrix of a multi-class classifier is not as simple and the approach used is selecting one class against all, which means one of the five classes will be picked and analysed against the remaining four. This is done for each of the five classes. Figure 4.20 presents the results of the class water against all others, then on figure 4.21 we can see the same for the forest class, on figure 4.22 for agriculture, on figure 4.23 for urban areas and finally on figure 4.24 for bare land.

This classifier was developed using the results from the "Cloudless" version of the binary classifiers. The usual four variants were still developed using the respective decision trees. It is important to remember that, in this case, the training dataset for all variants included the month variable. In this classifier variants "MonthVar" and "2019_MonthVar" indicate that the decision trees used to construct the dataset were of said variants; contrary to all the other classifiers, they do not indicate that new variable was added.

The results obtained were mixed, with some classes benefiting from it and other worsening.

In the case of water against all, figure 4.20, when compared to the results of the cloudless binary classifier, figure 4.7, (summary in table 4.2 with the average obtained in all variants) it is possible to observe that in general, it had a negative impact. BA had a better score using this classifier, however the same is not shown looking at F_1-score and MCC. The exception to this is "2019" variant that obtained better results in all these metrics using the final decision classifier.

	Min	Max	Average
TPR	83,26%	94,27%	92,19%
TNR	98,45%	99,78%	99,55%
PPV	88,91%	98,46%	96,85%
NPV	97,53%	99,15%	98,84%
Acc	96,48%	99,01%	98,60%
BA	90,86%	96,99%	95,87%
F1_score	86,00%	96,10%	94,45%
MCC	84,05%	95,55%	93,69%

(a) allMonths (Water)

	Min	Max	Average
TPR	83,33%	94,29%	92,20%
TNR	98,89%	99,90%	99,68%
PPV	91,81%	99,27%	97,73%
NPV	97,55%	99,15%	98,85%
Acc	96,87%	99,10%	98,71%
BA	91,11%	97,02%	95,94%
F1_score	87,37%	96,42%	94,88%
MCC	85,72%	95,95%	94,20%

(b) MonthVar (Water)

	Min	Max	Average
TPR	81,80%	95,01%	92,83%
TNR	85,01%	99,88%	97,48%
PPV	48,53%	99,12%	88,39%
NPV	97,23%	99,26%	98,91%
Acc	86,28%	99,15%	96,88%
BA	88,51%	97,34%	95,16%
F1_score	64,19%	96,62%	89,81%
MCC	61,64%	96,18%	88,58%

(c) 2019 (Water)

	Min	Max	Average
TPR	82,27%	95,12%	93,02%
TNR	92,06%	99,89%	97,64%
PPV	60,69%	99,20%	88,05%
NPV	97,21%	99,27%	98,94%
Acc	90,79%	99,16%	97,04%
BA	87,16%	97,32%	95,33%
F1_score	69,85%	96,68%	89,93%
MCC	65,60%	96,24%	88,63%

(d) 2019_MonthVar (Water)

Figure 4.20: Results of Final Decision classifier, Water against all.

Table 4.2: Average results comparison between Water Binary classifier and Final Decision classifier, Water against all

	Water Binary classifier				**Final Decision classifier**			
	allMonths	MonthVar	2019	2019_MonthVar	allMonths	MonthVar	2019	2019_MonthVar
TPR	91,23%	91,57%	93,04%	92,90%	92,19%	92,20%	92,83%	93,02%
TNR	99,80%	99,81%	95,34%	97,75%	99,55%	99,68%	97,48%	97,64%
PPV	98,52%	98,65%	86,62%	88,26%	96,85%	97,73%	88,39%	88,05%
NPV	98,71%	98,76%	98,92%	98,92%	98,84%	98,85%	98,91%	98,94%
Acc	98,68%	98,74%	95,04%	97,12%	98,60%	98,71%	96,88%	97,04%
BA	95,51%	95,69%	94,19%	95,32%	95,87%	95,94%	95,16%	95,33%
F_1-score	94,71%	94,96%	87,98%	90,06%	94,45%	94,88%	89,81%	89,93%
MCC	94,07%	94,34%	86,66%	88,75%	93,69%	94,20%	88,58%	88,63%

In the case of class forest the opposite can be seen when analysing figures 4.21 (final decision classifier) and 4.9 (cloudless binary classifier). Table 4.3 presents the average results of both classifiers in all metrics and varianst and it can be seen that this final classifier obtained better results in all variants. Only in the variant "2019" the metric MCC obtained a higher score in the binary classifier instead of the final decision classifier but the difference is minimal and could be due to the randomness when building the training dataset for both binary and multi-class classifiers.

	Min	Max	Average
TPR	70,98%	90,91%	82,71%
TNR	51,37%	82,61%	67,11%
PPV	59,56%	80,02%	71,04%
NPV	66,81%	85,92%	80,56%
Acc	62,52%	79,31%	74,73%
BA	62,74%	79,31%	74,91%
F1_score	65,31%	78,89%	76,19%
MCC	25,92%	58,61%	50,69%

(a) allMonths (Forest)

	Min	Max	Average
TPR	70,32%	91,10%	83,70%
TNR	51,48%	81,47%	67,15%
PPV	60,64%	78,35%	71,20%
NPV	69,21%	86,28%	81,36%
Acc	64,04%	78,57%	75,22%
BA	64,29%	78,71%	75,42%
F1_score	67,06%	79,79%	76,77%
MCC	29,21%	57,71%	51,69%

(b) MonthVar (Forest)

	Min	Max	Average
TPR	49,46%	87,27%	72,01%
TNR	51,28%	82,71%	69,98%
PPV	60,40%	79,02%	69,87%
NPV	58,90%	84,66%	73,95%
Acc	59,50%	77,88%	70,97%
BA	59,27%	77,80%	70,99%
F1_score	54,38%	78,91%	70,19%
MCC	18,92%	55,90%	42,89%

(c) 2019 (Forest)

	Min	Max	Average
TPR	34,05%	91,82%	72,39%
TNR	54,68%	87,51%	72,81%
PPV	64,97%	77,73%	72,12%
NPV	58,18%	87,52%	75,28%
Acc	61,41%	78,79%	72,61%
BA	60,78%	78,83%	72,60%
F1_score	46,28%	78,87%	71,09%
MCC	25,60%	57,64%	46,25%

(d) 2019_MonthVar (Forest)

Figure 4.21: Results of Final Decision classifier, Forest against all.

Table 4.3: Average results comparison between Forest Binary classifier and Final Decision classifier, Forest against all

	Forest Binary classifier				Final Decision classifier			
	allMonths	MonthVar	2019	2019_MonthVar	allMonths	MonthVar	2019	2019_MonthVar
TPR	73,92%	80,21%	68,62%	70,32%	82,71%	83,70%	72,01%	72,39%
TNR	73,97%	69,40%	72,97%	73,31%	67,11%	67,15%	69,98%	72,81%
PPV	73,26%	71,86%	71,23%	72,20%	71,04%	71,20%	69,87%	72,12%
NPV	75,33%	79,38%	73,23%	74,20%	80,56%	81,36%	73,95%	75,28%
Acc	73,95%	74,67%	70,85%	71,85%	74,73%	75,22%	70,97%	72,61%
BA	73,94%	74,80%	70,79%	71,82%	74,91%	75,42%	70,99%	72,60%
F_1-score	73,30%	75,41%	68,09%	69,94%	76,19%	76,77%	70,19%	71,09%
MCC	48,24%	50,40%	42,95%	44,98%	50,69%	51,69%	42,89%	46,25%

The results were mixed for the agriculture class. By comparing the outcomes of the final decision classifier, figure 4.22, and the corresponding binary classifier, figure 4.11, summarised in table 4.4, it can be noticed that the performance of the final decision classifier obtained for variants "allMonths" and "MonthVar" is superior to the binary classifier for most of the metrics considered, with an increase between 1% and 3%. However, the same cannot be said for variants "2019" and "2019_MonthVar", where the binary classifier has the advantage and the final classifier performed worse, varying between metrics from approximately 1% to 5%.

	Min	Max	Average
TPR	34,79%	78,17%	57,24%
TNR	75,91%	91,84%	85,46%
PPV	51,37%	75,35%	68,67%
NPV	70,54%	86,43%	78,71%
Acc	65,21%	78,96%	75,43%
BA	59,60%	77,83%	71,35%
F1_score	45,10%	71,42%	61,75%
MCC	20,51%	54,87%	44,90%

(a) allMonths (Agriculture)

	Min	Max	Average
TPR	34,86%	77,41%	57,33%
TNR	75,16%	91,96%	86,08%
PPV	54,23%	75,69%	69,44%
NPV	71,36%	85,78%	78,80%
Acc	66,73%	79,13%	75,86%
BA	61,01%	76,62%	71,70%
F1_score	45,84%	69,85%	62,26%
MCC	23,74%	53,54%	45,69%

(b) MonthVar (Agriculture)

	Min	Max	Average
TPR	32,58%	76,39%	57,04%
TNR	64,31%	89,96%	80,03%
PPV	45,47%	74,88%	62,67%
NPV	70,19%	85,57%	77,52%
Acc	61,12%	77,88%	71,86%
BA	57,94%	77,17%	68,54%
F1_score	42,25%	70,62%	58,58%
MCC	15,77%	52,88%	38,50%

(c) 2019 (Agriculture)

	Min	Max	Average
TPR	38,24%	70,90%	54,08%
TNR	71,67%	92,92%	83,92%
PPV	49,26%	74,96%	65,78%
NPV	70,44%	83,48%	77,13%
Acc	64,00%	78,13%	73,31%
BA	59,10%	76,16%	69,00%
F1_score	45,42%	69,28%	58,44%
MCC	18,93%	52,30%	40,26%

(d) 2019_MonthVar (Agriculture)

Figure 4.22: Results of Final Decision classifier, Agriculture against all.

Table 4.4: Average results comparison between Agriculture Binary classifier and Final Decision classifier, Agriculture against all

	Agriculture Binary classifier				**Final Decision classifier**			
	allMonths	MonthVar	2019	2019_MonthVar	allMonths	MonthVar	2019	2019_MonthVar
TPR	54,42%	57,78%	59,47%	59,28%	57,24%	57,33%	57,04%	54,08%
TNR	86,09%	84,94%	79,70%	81,73%	85,46%	86,08%	80,03%	83,92%
PPV	68,80%	68,07%	62,44%	64,28%	68,67%	69,44%	62,67%	65,78%
NPV	77,70%	78,88%	78,26%	78,67%	78,71%	78,80%	77,52%	77,13%
Acc	74,83%	75,29%	72,51%	73,75%	75,43%	75,86%	71,86%	73,31%
BA	70,25%	71,36%	69,58%	70,50%	71,35%	71,70%	68,54%	69,00%
F_1-score	60,02%	61,70%	60,41%	61,25%	61,75%	62,26%	58,58%	58,44%
MCC	43,30%	44,68%	39,88%	41,92%	44,90%	45,69%	38,50%	40,26%

Next, for urban areas, this final classifier (figure 4.23) had better outcomes than the binary classifier (figure 4.13). By examining and comparing the results in table 4.5 it can be seen that the final decision classifier obtained a better results than the binary classifier in all variants and in most of the metrics considered, with an increase ranging from, approximately, between 1% and 5%, depending on the metric. Therefore this class was the one that most benefited from the development of the final decision classifier as its results were the most affected, positively, when compared to the results of the remaining four classes.

	Min	Max	Average
TPR	5,96%	25,29%	15,93%
TNR	99,21%	99,92%	99,70%
PPV	23,01%	78,40%	60,67%
NPV	97,68%	98,14%	97,92%
Acc	97,10%	97,81%	97,64%
BA	52,73%	62,42%	57,82%
F1_score	9,47%	35,24%	24,78%
MCC	10,63%	37,36%	29,86%

(a) allMonths (Urban Areas)

	Min	Max	Average
TPR	6,71%	23,05%	16,18%
TNR	99,12%	99,88%	99,65%
PPV	19,74%	73,91%	57,64%
NPV	97,69%	98,09%	97,93%
Acc	97,02%	97,80%	97,60%
BA	53,01%	61,28%	57,92%
F1_score	10,02%	33,23%	24,85%
MCC	10,25%	37,01%	29,33%

(b) MonthVar (Urban Areas)

	Min	Max	Average
TPR	14,05%	27,08%	21,18%
TNR	91,89%	99,84%	98,44%
PPV	5,45%	70,29%	42,33%
NPV	97,82%	98,19%	98,02%
Acc	90,09%	97,82%	96,54%
BA	55,23%	63,29%	59,81%
F1_score	8,43%	36,88%	26,32%
MCC	5,85%	39,28%	27,29%

(c) 2019 (Urban Areas)

	Min	Max	Average
TPR	22,59%	48,54%	33,35%
TNR	71,56%	99,50%	93,68%
PPV	4,12%	54,14%	29,78%
NPV	98,08%	98,33%	98,24%
Acc	71,00%	97,66%	92,19%
BA	59,85%	66,10%	63,51%
F1_score	7,60%	39,04%	26,15%
MCC	6,86%	39,57%	25,66%

(d) 2019_MonthVar (Urban Areas)

Figure 4.23: Results of Final Decision classifier, Urban Areas against all.

Table 4.5: Average results comparison between Urban Areas Binary classifier and Final Decision classifier, Urban Areas against all

	Urban Areas Binary classifier				Final Decision classifier			
	allMonths	MonthVar	2019	2019_MonthVar	allMonths	MonthVar	2019	2019_MonthVar
TPR	14,38%	12,22%	17,21%	16,52%	15,93%	16,18%	21,18%	33,35%
TNR	99,74%	99,79%	98,93%	98,96%	99,70%	99,65%	98,44%	93,68%
PPV	62,34%	62,22%	44,78%	43,32%	60,67%	57,64%	42,33%	29,78%
NPV	97,88%	97,83%	97,94%	97,92%	97,92%	97,93%	98,02%	98,24%
Acc	97,64%	97,63%	96,92%	96,94%	97,64%	97,60%	96,54%	92,19%
BA	57,06%	56,00%	58,07%	57,74%	57,82%	57,92%	59,81%	63,51%
F_1-score	22,77%	20,10%	22,39%	20,96%	24,78%	24,85%	26,32%	26,15%
MCC	28,57%	26,47%	24,91%	23,60%	29,86%	29,33%	27,29%	25,66%

Finally, for the bare land class, the results were also mixed. The figures for this class are figure 4.24, for the final classifier, and figure 4.15, for the binary classifier. The results for the variant "allMonths" and "MonthVar" are better in the binary classifier. The results of the final decision classifier for variant "MonthVar" are inexplicably terrible and do not match with the remaining. It is unclear what caused this but such a considerable difference cannot be blamed on the randomness of the training dataset alone however it is difficult to pinpoint a reason.

For the variant "2019" the results between both classifiers are similar, with the binary classifier having higher BA and F_1-score and the final decision classifier having the advantage on MCC metric. For the variant "2019_MonthVar" the results were also similar but slightly higher for the final decision classifier.

	Min	Max	Average
TPR	2,18%	16,06%	10,31%
TNR	99,86%	100,0%	99,99%
PPV	9,53%	88,97%	73,51%
NPV	99,80%	99,83%	99,82%
Acc	99,67%	99,82%	99,80%
BA	51,09%	58,02%	55,15%
F1_score	4,22%	25,78%	17,56%
MCC	8,20%	34,10%	26,62%

(a) allMonths (Bare Land)

	Min	Max	Average
TPR	0,00%	0,36%	0,02%
TNR	100,0%	100,0%	100,0%
PPV	0,00%	66,67%	19,45%
NPV	99,80%	99,80%	99,80%
Acc	99,79%	99,80%	99,80%
BA	50,00%	50,18%	50,01%
F1_score	0,00%	0,71%	0,15%
MCC	0,00%	4,88%	0,83%

(b) MonthVar (Bare Land)

	Min	Max	Average
TPR	3,49%	18,47%	13,87%
TNR	99,82%	100,0%	99,97%
PPV	15,71%	82,33%	63,95%
NPV	99,80%	99,83%	99,82%
Acc	99,65%	99,82%	99,80%
BA	51,74%	59,23%	56,92%
F1_score	6,69%	29,21%	21,76%
MCC	15,88%	35,83%	28,60%

(c) 2019 (Bare Land)

	Min	Max	Average
TPR	2,25%	19,74%	13,14%
TNR	99,96%	100,0%	99,98%
PPV	33,31%	64,55%	54,95%
NPV	99,80%	99,84%	99,82%
Acc	99,78%	99,81%	99,80%
BA	51,12%	59,85%	56,56%
F1_score	4,34%	29,41%	20,72%
MCC	11,77%	33,67%	26,10%

(d) 2019_MonthVar (Bare Land)

Figure 4.24: Results of Final Decision classifier, Bare Land against all.

Table 4.6: Average results comparison between Bare Land Binary classifier and Final Decision classifier, Bare Land against all

	Bare Land Binary classifier				**Final Decision classifier**			
	allMonths	MonthVar	2019	2019_MonthVar	allMonths	MonthVar	2019	2019_MonthVar
TPR	14,41%	14,85%	14,67%	12,35%	10,31%	0,02%	13,87%	13,14%
TNR	99,94%	99,94%	99,95%	99,98%	99,99%	100,0%	99,97%	99,98%
PPV	62,77%	57,35%	60,16%	57,62%	73,51%	19,45%	63,95%	54,95%
NPV	99,83%	99,83%	99,83%	99,82%	99,82%	99,80%	99,82%	99,82%
Acc	99,77%	99,76%	99,78%	99,80%	99,80%	99,80%	99,80%	99,80%
BA	57,18%	57,40%	57,31%	59,17%	55,15%	50,01%	56,92%	56,56%
F_1-score	22,58%	22,62%	22,21%	19,88%	17,56%	0,15%	21,76%	20,72%
MCC	29,36%	28,34%	28,38%	25,92%	26,62%	0,83%	28,60%	26,10%

4.2.6.1 Image results

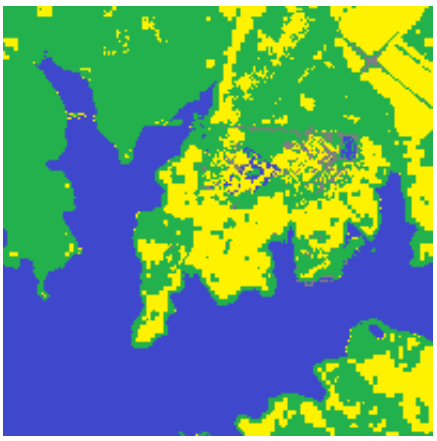
In figure 4.25 it is possible to observe a small section of the region of study. The COS for that section is displayed (this section does not possess the class bare land), along with an RGB satellite image taken at 2018/05/10 and the result of the Final classifier (variant "allMonths") for that day. Full results of this land section can be seen in appendix A. The appendix contains the COS for this region, as seen in figure 4.25(a), satellite RGB images of all the dates used in this study, along with the results of the four variants of the Final classifier developed.

It is interesting to compare the resulting images with COS as it can be noticed how much grainier they are. Classes can be seen having isolated pixels in the middle of other classes and the edges between classes are also much less defined and less smooth than they are in COS.

When examining appendix A it can also be seen how the classifier handled the presence of clouds, especially in date 2018/04/30 where the land in this section was almost entirely obscured by a cloud and its shadow. Since the readings of the satellite cannot penetrate thick clouds like the one present in this instance it is expected that the classifier will not be able to correctly identify the land. In most cases, it classifies the cloud as forest, except on "2019_MonthVar" where it is classified as urban area. It is also possible to observe that even with the shadow the cloud is projecting onto the land the classifiers are still able to identify some water, on the left lower corner, and some agriculture, on the right lower corner.

It is also interesting to observe how the percentage of forest and agriculture identified will change along the seasons. Typically dry seasons, where the land has more of a brown-yellow tone, will have more land identified as forest than agriculture. On the contrary, humid months, where the scenery is much greener, present a bigger percentage of land identified as agriculture than forest. However, this is not always the case with some month presenting opposite results of what was described. It can safely be said that the task of distinguishing both these classes is not trivial, needing some improvements.

(a) COS of the sample section



(b) Final classifier result of sample section

(c) RGB image of the sample section

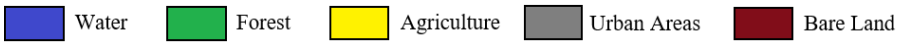


Figure 4.25: COS, satellite RGB image, and Final classifier result of a sample section dating 2018/05/10

4.3 Factors affecting performance

There are some factors that could be leading to lower performance. One of them, as discussed previously, is the fact that the satellites used, Sentinel-2 constellation, lack a TIR sensor. As explained before, this band is very useful for the identification of urban areas and bare land, since these types of terrain tend to absorb and reflect a lot more heat when compared to the remaining classes. This would lead to distinctively higher values of TIR in these regions, making them stand out from the others. This band is also used to calculate some of the indices found in the literature that could not be computed in this study.

Another factor has to do with the type of classifier chosen. Decision trees, although simple to understand and fast to train and test, are static once the training is completed, which means they cannot adapt to new information unless all the training is repeated with these new examples. This limits the scalability of the developed classifiers which

are expected to perform poorly in regions with a very contrasting climate from the one in Portugal. Other types of classifiers, such as artificial neural networks, could be used instead to tackle this limitation, although time is a concerning factor too.

The analysis and training of the classifier are also pixel-based, so each pixel is classified based only on its values, not taking into account the information of its neighbour pixels. This lack of spatial awareness can hinder the ability to detect certain man-made structures. Cities and crops are typically built in a very geometric manner, with straight lines, perfect circles, and similar shapes that do not typically occur in nature. These characteristics usually make them clearly distinguishable from natural elements, which would be especially useful to differentiate forest from agriculture since the spectral values of a pixel of either class are very similar. An object-based approach could allow for an improvement in detection performance of some of these elements.

Lastly, it is important to discuss the veracity of the ground truth. As mention before COS was used as the ground truth for this study. However, it has some important flaws for this use. COS is only developed every few years, using only certain months instead of the whole year for its development which will cause some information, from the left out months, to not be registered.

COS also only registers the land use and not the land cover at the moment. For example, considering a crop field, after harvesting, the land will be mainly bare soil (class bare land), however, COS only registers what the land is used for, in this case, agriculture. So, although the classifier might correctly identify it as bare land, it will be marked as a wrong prediction since the ground truth identifies it as agriculture. Another example is water bodies, which change the amount of water throughout the year. This means that in dryer months the portion of water identified will be smaller, nonetheless COS always considers the maximum amount of water registered, which again would lead to the classifier identifying the dried water body as another class (mostly bare land) and COS stating otherwise.

Besides this, the minimum unit of land is 1 ha, and a class has to occupy at least 75% of that unit of land to be registered in COS (DGTerritório (2019)). Because of this, some smaller portions of a certain class might not be identified in COS, once again leading to good predictions being identified as wrong.

This leads to lower performances than they might be in reality since the ground truth does not adapt to these variations over seasons.

Chapter 5

Conclusions and Future work

This dissertation aimed to develop a better way to perform land cover classification. Currently, COS is mostly used for this purpose but although it has a high spatial resolution, its temporal resolution is very low, only being produced every few years. This makes it very hard to accompany land change that could be of interest for various sectors such as forest monitoring and fire prevention. For this reason, an automatic approach based on satellite images was used, and several decision tree classifiers were developed.

The images obtained, from Sentinel-2 constellation, have a somewhat worse spatial resolution than COS but compensate on temporal resolution. One set of images was taken from each month, from March 2018 to January 2020, however, the resulting classifiers can be used more than once a month, providing a great temporal resolution.

The development of these classifiers used not only the various spectral bands available, just as they are, but also to calculate indices that would help to identify various types of land. Five different classes were considered, contrasting from the nine mega-classes COS possesses, and decision trees classifiers were built for each one. Decision trees were used due to their computational speed and simplicity. Some different approaches were performed in order to correctly assess the true performance of the classifiers and also try to improve it.

The binary classifiers performed differently for each class and it was possible to see how certain factors affected the performance, such as the spectral signature of a type of land and the imbalance of the dataset.

Water was the best performing classifier, having a very unique spectral signature and a good sample size, it was easy to identify and distinguish from surrounding land.

Forest and agriculture both had a good sample size but their spectral signature were very similar which resulted in some difficulty distinguishing these two classes. Nevertheless, the performance for both was considered to be good.

Urban areas and bare land had a very small sample size, and although their spectral signature was distinguishable from the remaining classes that was not enough to produce good decision trees. The results for both of these classes were very weak compared to the other classes. Another factor worth mentioning is the lack of a TIR band which is not available in the satellites used but could have had a positive impact on the detection of both theses classes, as seen in the literature.

These binary classifiers were then used to build a final classifier, so it would be able to better distinguish similar classes and improve performance. The results were mixed with some classes benefiting from it, and others not so much.

These classifiers are simpler and less accurate than COS but they are still considered to be of added-value due to their temporal resolution. Sentinel-2 satellites have a revisit frequency of 10 days for each satellite, or 5 days when considering both. The classifiers can be used with that frequency, which is a good temporal resolution. Although it is important to remember that some images might not produce good results due to cloud cover, it is still a much better approach to monitoring of the land than COS alone.

5.1 Future work

There are still some aspects that would be interesting to further develop and experiment. Some of the factors that could be hindering the performance of the classifiers described in section 4.3 should be looked upon and corrected when possible.

The first different approach would be adopting an object-based approach instead of a pixel-based one as done in this study. An object-based approach could more easily identify land geometrically arranged, such as crops and cities, also helping to distinguish them from natural classes such as forests. As previously discussed the spectral values obtained from forest and agriculture can be extremely similar so using only a pixel-based approach makes it difficult to differentiate them.

Another different approach that could be done would be trying different types of classifiers besides the decision trees used. Other kinds of classifiers could potentially lead to better results, and better adaptability in the long run, which decision trees lack. However, it is important to point out that with images of this dimension this would be a very time-consuming task since most classifiers are much more computationally heavy than decision trees.

Using another satellite with similar characteristics but possessing a TIR sensor would also improve the detection of urban areas and bare land, since these usually accumulate more heat than the surrounding land, besides being used to calculate other indices to further help the detection of these classes.

However, without changing anything in the methodology used, the current classifiers could still be improved with different techniques without needing to re-train them. Image processing could be applied to the images output by the classifier so that isolated pixels could be corrected and rough edges between classes could be smoothed. This would

certainly help to improve performance because as it could be seen in appendix A the images created by the classifier as very different from COS with a lot of isolated pixels and ragged edges.

Bibliography

Aguilar, M. A., A. Vallario, F. J. Aguilar, A. G. Lorca, and C. Parente (2015). "Object-based greenhouse horticultural crop identification from multi-temporal satellite imagery: A case study in Almeria, Spain." In: *Remote Sensing* 7(6), pp. 7378–7401. ISSN: 20724292. DOI: 10.3390/rs70607378.

Arroyo-Mora, J. P., G. A. Sánchez-Azofeifa, M. E. Kalacska, B. Rivard, J. C. Calvo-Alvarado, and D. H. Janzen (2005). "Secondary forest detection in a neotropical dry forest landscape using Landsat 7 ETM+ and IKONOS imagery." In: *Biotropica* 37(4), pp. 497–507. ISSN: 00063606. DOI: 10.1111/j.1744-7429.2005.00068.x.

As-syakur, A. R., I. W. S. Adnyana, I. W. Arthana, and I. W. Nuarsa (2012). "Enhanced built-UP and bareness index (EBBI) for mapping built-UP and bare land in an urban area." In: *Remote Sensing* 4(10), pp. 2957–2970. ISSN: 20724292. DOI: 10.3390/rs4102957.

Bannari, A., D. Morin, F. Bonn, and A. R. Huete (1995). "A review of vegetation indices." In: *Remote Sensing Reviews* 13(1-2), pp. 95–120. ISSN: 02757257. DOI: 10.1080/02757259509532298.

Baret, F., G. Guyot, and D. Major (1989). "TSAVI: A Vegetation Index Which Minimizes Soil Brightness Effects On LAI And APAR Estimation." In: *12th Canadian Symposium on Remote Sensing Geoscience and Remote Sensing Symposium*, vol. 3. IEEE, pp. 1355–1358. DOI: 10.1109/IGARSS.1989.576128. URL: http://ieeexplore.ieee.org/document/576128/.

Brodersen, K. H., C. S. Ong, K. E. Stephan, and J. M. Buhmann (2010). "The balanced accuracy and its posterior distribution." In: *Proceedings - International Conference on Pattern Recognition*, pp. 3121–3124. ISSN: 10514651. DOI: 10.1109/ICPR.2010.764.

Brodley, C. E. and M. A. Friedl (1997). "Decision tree classification of land cover from remotely sensed data." In: *Remote Sensing of Environment* 61(3), pp. 399–409. ISSN: 00344257. DOI: 10.1016/S0034-4257(97)00049-7.

Brown, J. C., J. H. Kastens, A. C. Coutinho, D. d. C. Victoria, and C. R. Bishop (2013). "Classifying multiyear agricultural land use data from Mato Grosso using time-series MODIS vegetation index data." In: *Remote Sensing of Environment* 130, pp. 39–50. ISSN: 00344257. DOI: 10.1016/j.rse.2012.11.009. URL: http://dx.doi.org/10.1016/j.rse.2012.11.009.

Calvao, T. and J. M. Palmeirim (2004). "Mapping Mediterranean scrub with satellite imagery: Biomass estimation and spectral behaviour." In: *International Journal of Remote Sensing* 25(16), pp. 3113–3126. ISSN: 01431161. DOI: 10.1080/01431160310001654978.

Chen, W., L. Liu, C. Zhang, J. Wang, J. Wang, and Y. Pan (2004). "Monitoring the seasonal bare soil areas in Beijing using multi-temporal TM images." In: *International Geoscience and Remote Sensing Symposium (IGARSS)* 5(3), pp. 3379–3382. DOI: 10.1109/igarss.2004.1370429.

Chicco, D. and G. Jurman (2020). "The advantages of the Matthews correlation coefficient (MCC) over F1 score and accuracy in binary classification evaluation." In: *BMC Genomics* 21(1), pp. 1–13. ISSN: 14712164. DOI: 10.1186/s12864-019-6413-7.

DGTerritório (2019). "Especificações técnicas da Carta de uso e ocupação do solo (COS) de Portugal Continental para 2018." In: p. 60.

Du, Y., Y. Zhang, F. Ling, Q. Wang, W. Li, and X. Li (2016). "Water bodies' mapping from Sentinel-2 imagery with Modified Normalized Difference Water Index at 10-m spatial resolution produced by sharpening the swir band." In: *Remote Sensing* 8(4). ISSN: 20724292. DOI: 10.3390/rs8040354.

Etteieb, S., M. Louhaichi, C. Kalaitzidis, and I. Z. Gitas (2013). "Mediterranean forest mapping using hyper-spectral satellite imagery." In: *Arabian Journal of Geosciences* 6(12), pp. 5017–5032. ISSN: 18667511. DOI: 10.1007/s12517-012-0748-6.

Feyisa, G. L., H. Meilby, R. Fensholt, and S. R. Proud (2014). "Automated Water Extraction Index: A new technique for surface water mapping using Landsat imagery." In: *Remote Sensing of Environment* 140, pp. 23–35. ISSN: 00344257. DOI: 10.1016/j.rse.2013.08.029. URL: http://dx.doi.org/10.1016/j.rse.2013.08.029.

Garcia, C., R. S. Tavares, A. D. Mora, J. Fonseca, H. Oliveira, and L. B. Oliveira (2020). "FPGA-based satellite image classification for water bodies detection." In: *Proceedings - 2020 International Young Engineers Forum, YEF-ECE 2020*, pp. 44–48. DOI: 10.1109/YEF-ECE49388.2020.9171811.

Gitelson, A. A., Y. J. Kaufman, and M. N. Merzlyak (1996). "Use of a green channel in remote sensing of global vegetation from EOS-MODIS." In: *Remote sensing of Environment* 58(3), pp. 289–298.

Gitelson, A. A., A. Viña, V. Ciganda, D. C. Rundquist, and T. J. Arkebauer (2005). "Remote estimation of canopy chlorophyll content in crops." In: *Geophysical Research Letters* 32(8), pp. 1–4. ISSN: 00948276. DOI: 10.1029/2005GL022688.

Goldblatt, R., W. You, G. Hanson, and A. K. Khandelwal (2016). "Detecting the boundaries of urban areas in India: A dataset for pixel-based image classification in google earth engine." In: *Remote Sensing* 8(8). ISSN: 20724292. DOI: 10.3390/rs8080634.

Haboudane, D., J. R. Miller, N. Tremblay, P. J. Zarco-Tejada, and L. Dextraze (2002). "Integrated narrow-band vegetation indices for prediction of crop chlorophyll content for application to precision agriculture." In: *Remote Sensing of Environment* 81(2-3), pp. 416–426. ISSN: 00344257. DOI: 10.1016/S0034-4257(02)00018-4.

Huang, C., Y. Chen, S. Zhang, and J. Wu (2018). "Detecting, Extracting, and Monitoring Surface Water From Space Using Optical Sensors: A Review." In: *Reviews of Geophysics* 56(2), pp. 333–360. ISSN: 19449208. DOI: 10.1029/2018RG000598.

Huete, A (Mar. 1997). "A comparison of vegetation indices over a global set of TM images for EOS-MODIS." In: *Remote Sensing of Environment* 59(3), pp. 440–451. ISSN: 00344257. DOI: 10.1016/S0034-4257(96)00112-5. URL: https://linkinghub.elsevier.com/retrieve/pii/S0034425796001125.

Huete, A. R. (1988). "A soil-adjusted vegetation index (SAVI)." In: *Remote Sensing of Environment* 25(3), pp. 295–309. ISSN: 00344257. DOI: 10.1016/0034-4257(88)90106-X.

Huete, A. R. (July 2012). "Vegetation Indices, Remote Sensing and Forest Monitoring." In: *Geography Compass* 6(9), pp. 513–532. ISSN: 17498198. DOI: 10.1111/j.1749-8198.2012.00507.x. URL: http://doi.wiley.com/10.1111/j.1749-8198.2012.00507.x.

Huete, A., C Justice, and W Van Leeuwen (1996). "MODIS Vegetation index (MOD 13), EOS MODIS algorithm–theoretical basis document." In: *Greenbelt, MD: NASA Goddard Space Flight Center*.

Karakizi, C., M. Oikonomou, and K. Karantzalos (2016). "Vineyard detection and vine variety discrimination from very high resolution satellite data." In: *Remote Sensing* 8(3), pp. 1–25. ISSN: 20724292. DOI: 10.3390/rs8030235.

Kaufman, Y and D Tanre. (1992). "Atmospherically resistant vegetation index." In: *IEEE Transactions on Geoscience and Remote Sensing* 30(2), pp. 260–271.

Kawamura, M. (1996). "Relation between social and environmental conditions in Colombo Sri Lanka and the urban index estimated by satellite remote sensing data." In: *Proc. 51st Annual Conference of the Japan Society of Civil Engineers*, pp. 190–191.

Li, H., C. Wang, C. Zhong, A. Su, C. Xiong, J. Wang, and J. Liu (2017). "Mapping urban bare land automatically from Landsat imagery with a simple index." In: *Remote Sensing* 9(3). ISSN: 20724292. DOI: 10.3390/rs9030249.

Liu, H. Q. and A. Huete (1995). "Feedback based modification of the NDVI to minimize canopy background and atmospheric noise." In: *IEEE Transactions on Geoscience and Remote Sensing* 33(2), pp. 457–465. ISSN: 01962892. DOI: 10.1109/36.377946.

McFeeters, S. K. (1996). "The use of the Normalized Difference Water Index (NDWI) in the delineation of open water features." In: *International Journal of Remote Sensing* 17(7), pp. 1425–1432. ISSN: 13665901. DOI: 10.1080/01431169608948714.

McMorrow, J. (2001). "Linear regression modelling for the estimation of oil palm age from Landsat TM." In: *International Journal of Remote Sensing* 22(12), pp. 2243–2264. ISSN: 13665901. DOI: 10.1080/01431160117188.

Mora, A., T. M. Santos, S. Lukasik, J. M. Silva, A. J. Falcão, J. M. Fonseca, and R. A. Ribeiro (2017). "Land cover classification from multispectral data using computational intelligence tools: A comparative study." In: *Information (Switzerland)* 8(4). ISSN: 20782489. DOI: 10.3390/info8040147.

Oliveira, D., L. Martins, A. Mora, C. Damásio, M. Caetano, J. Fonseca, and R. A. Ribeiro (2021). "Data fusion approach for eucalyptus trees identification." In: *International Journal of Remote Sensing* 42(11), pp. 4087–4109. ISSN: 13665901. DOI: 10.1080/01431161.2021.1883198. URL: https://doi.org/10.1080/01431161.2021.1883198.

Pal, M. and P. M. Mather (2003). "An assessment of the effectiveness of decision tree methods for land cover classification." In: *Remote Sensing of Environment* 86(4), pp. 554–565. ISSN: 00344257. DOI: 10.1016/S0034-4257(03)00132-9.

Qi, J, A Chehbouni, A. Huete, Y. Keer, and S Sorooshian (1994). "A modified soil vegetation adjusted index." In: *Remote Sensing of Environment* 48(2), pp. 119–126.

Qi, J., R. Marsett, P. Heilman, S. Biedenbender, S. Moran, D. Goodrich, and M. Weltz (2002). "RANGES improves satellite-based information and land cover assessments in Southwest United States." In: *Eos* 83(51). ISSN: 00963941. DOI: 10.1029/2002E0000411.

Rasul, A., H. Balzter, G. R. Ibrahim, H. M. Hameed, J. Wheeler, B. Adamu, S. Ibrahim, and P. M. Najmaddin (2018). "Applying built-up and bare-soil indices from Landsat 8 to cities in dry climates." In: *Land* 7(3). ISSN: 2073445X. DOI: 10.3390/land7030081.

Ribeiro, R. A., A. Falcão, A. Mora, and J. M. Fonseca (2014). "FIF: A fuzzy information fusion algorithm based on multi-criteria decision making." In: *Knowledge-Based Systems* 58, pp. 23–32. ISSN: 09507051. DOI: 10.1016/j.knosys.2013.08.032. URL: http://dx.doi.org/10.1016/j.knosys.2013.08.032.

Richardson, A. J. and C. L. Wiegand (1977). "Distinguishing vegetation from soil background information." In: *Photogrammetric Engineering and Remote Sensing* 43(12), pp. 1541–1552.

Rogers, A. S. and M. S. Kearney (2004). "Reducing signature variability in unmixing coastal marsh Thematic Mapper scenes using spectral indices." In: *International Journal of Remote Sensing* 25(12), pp. 2317–2335. ISSN: 01431161. DOI: 10.1080/01431160310001618103.

Rondeaux, G., M. Steven, and F. Baret (1996). "Optimization of soil-adjusted vegetation indices." In: *Remote Sensing of Environment* 55(2), pp. 95–107. ISSN: 00344257. DOI: 10.1016/0034-4257(95)00186-7.

Solano-Correa, Y. T., F. Bovolo, L. Bruzzone, and D. Fernandez-Prieto (2017). "Spatiotemporal evolution of crop fields in Sentinel-2 Satellite Image Time Series." In: *2017 9th International Workshop on the Analysis of Multitemporal Remote Sensing Images, MultiTemp 2017*(June), pp. 2–6. DOI: 10.1109/Multi-Temp.2017.8035236.

Tharwat, A. (Aug. 2020). "Classification assessment methods." In: *Applied Computing and Informatics* ahead-of-p(ahead-of-print). ISSN: 2634-1964. DOI: 10.1016/j.aci.2018.08.003. URL: https://www.emerald.com/insight/content/doi/10.1016/j.aci.2018.08.003/full/html.

Tian, T., C. Li, J. Xu, and J. Ma (2018). "Urban area detection in very high resolution remote sensing images using deep convolutional neural networks." In: *Sensors (Switzerland)* 18(3), pp. 1–16. ISSN: 14248220. DOI: 10.3390/s18030904.

Wang, L., J. Yan, L. Mu, and L. Huang (2020). "Knowledge discovery from remote sensing images: A review." In: *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*(October 2019), pp. 1–31. ISSN: 19424795. DOI: 10.1002/widm.1371.

Wiegand, C. L., A. J. Richardson, D. E. Escobar, and A. H. Gerbermann (1991). "Vegetation indices in crop assessments." In: *Remote Sensing of Environment* 35(2-3), pp. 105–119. ISSN: 00344257. DOI: 10.1016/0034-4257(91)90004-P.

Wilson, E. H. and S. A. Sader (2002). "Detection of forest harvest type using multiple dates of Landsat TM imagery." In: *Remote Sensing of Environment* 80(3), pp. 385–396. ISSN: 00344257. DOI: 10.1016/S0034-4257(01)00318-2.

Xu, H. (2008). "A new index for delineating built-up land features in satellite imagery." In: *International Journal of Remote Sensing* 29(14), pp. 4269–4276. ISSN: 13665901. DOI: 10.1080/01431160802039957.

Xu, H. (2006). "Modification of normalised difference water index (NDWI) to enhance open water features in remotely sensed imagery." In: *International Journal of Remote Sensing* 27(14), pp. 3025–3033. ISSN: 13665901. DOI: 10.1080/01431160600589179.

Xue, J. and B. Su (2017). "Significant remote sensing vegetation indices: A review of developments and applications." In: *Journal of Sensors* 2017. ISSN: 16877268. DOI: 10.1155/2017/1353691.

Yang, X., S. Zhao, X. Qin, N. Zhao, and L. Liang (2017). "Mapping of urban surface water bodies from sentinel-2 MSI imagery at 10 m resolution via NDWI-based image sharpening." In: *Remote Sensing* 9(6), pp. 1–19. ISSN: 20724292. DOI: 10.3390/rs9060596.

Zha, Y., J. Gao, and S. Ni (2003). "Use of normalized difference built-up index in automatically mapping urban areas from TM imagery." In: *International Journal of Remote Sensing* 24(3), pp. 583–594. ISSN: 01431161. DOI: 10.1080/01431160304987.

Zhao, H. and X. Chen (2005). "Use of normalized difference bareness index in quickly mapping bare areas from TM/ETM+." In: *International Geoscience and Remote Sensing Symposium (IGARSS)* 3(August), pp. 1666–1668. DOI: 10.1109/igarss.2005.1526319.

Zhou, Y., G. Yang, S. Wang, L. Wang, F. Wang, and X. Liu (2014). "A new index for mapping built-up and bare land areas from Landsat-8 OLI data." In: *Remote Sensing Letters* 5(10), pp. 862–871. ISSN: 21507058. DOI: 10.1080/2150704X.2014.973996. URL: http://dx.doi.org/10.1080/2150704X.2014.973996.

Appendix A

Classifier images

A.1 COS and original images

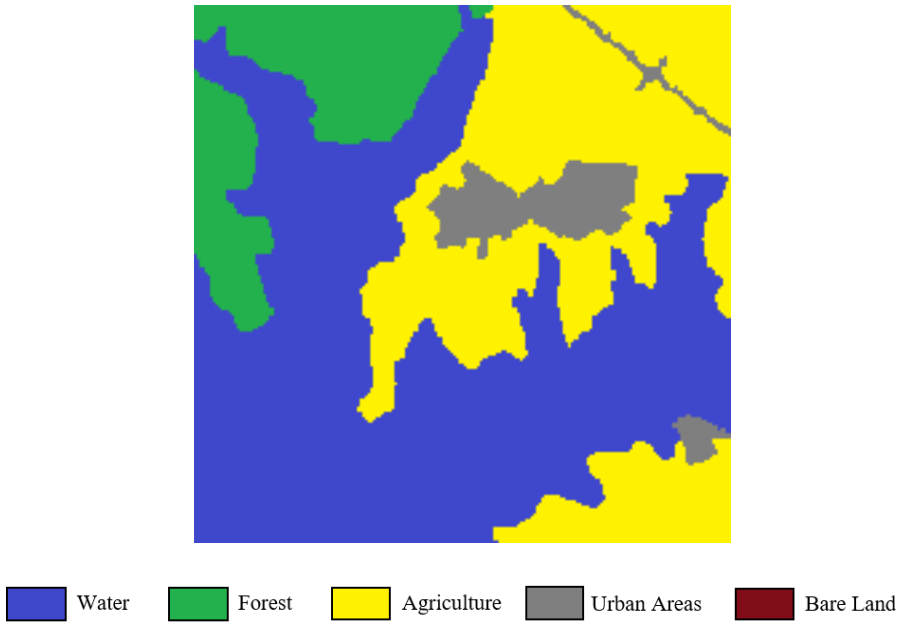
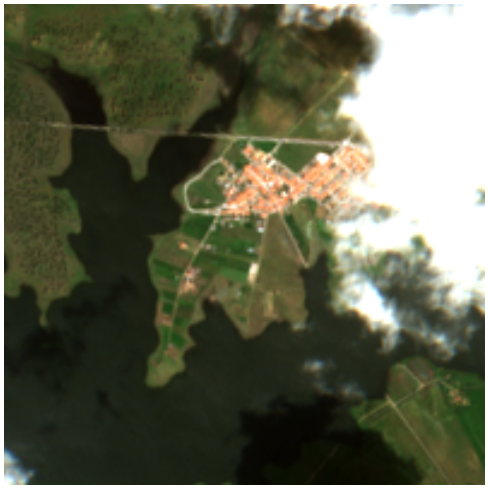
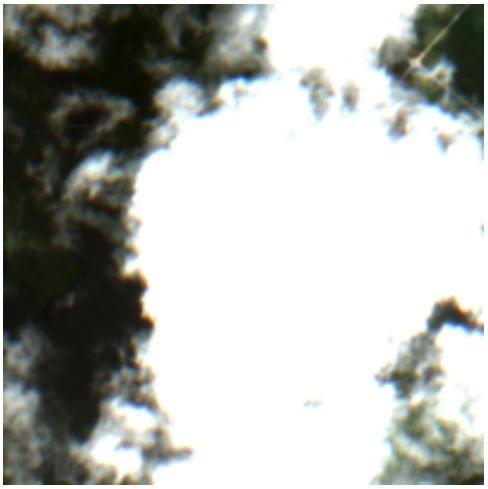


Figure A.1: COS of a sample section in the region of study. Water is identified as blue, forest is identified as green, agriculture is identified as yellow and urban areas are identified as grey. There was no bare land in this sample, but otherwise it would be identified as dark red.



(a) 2018/03/31



(b) 2018/04/30

(c) 2018/05/10

(d) 2018/06/19

(e) 2018/07/29

(f) 2018/08/18

Figure A.2: RGB images of the sample section, part 1.

(a) 2018/09/27

(b) 2018/10/07

(c) 2018/11/16

(d) 2018/12/06

(e) 2019/01/25

(f) 2019/02/14

Figure A.3: RGB images of the sample section, part 2.

(a) 2019/03/16

(b) 2019/04/30

(c) 2019/05/15

(d) 2019/06/09

(e) 2019/07/24

(f) 2019/08/23

Figure A.4: RGB images of the sample section, part 3.

(a) 2019/09/12

(b) 2019/10/22

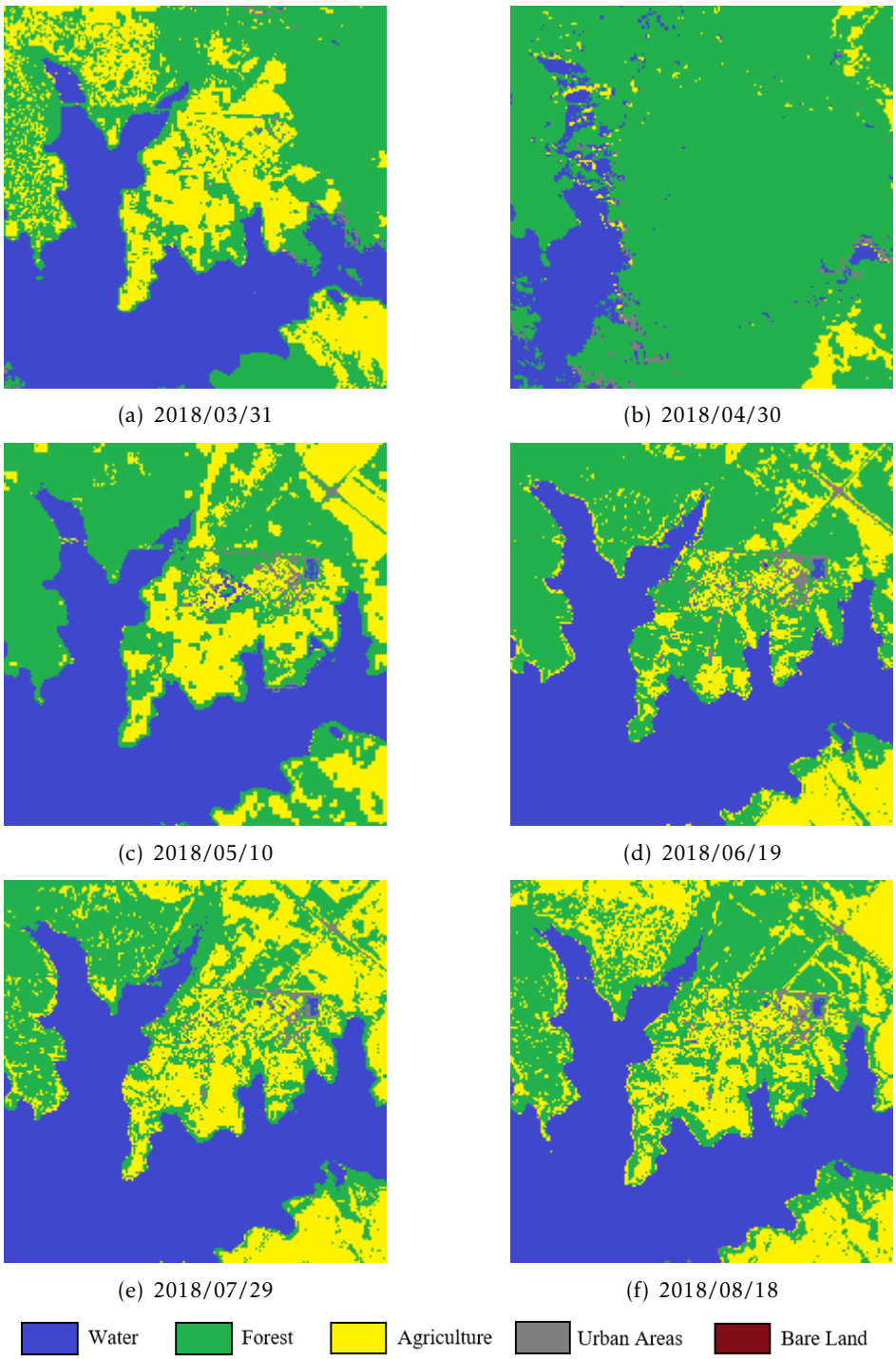
(c) 2019/11/16

(d) 2019/12/06

(e) 2020/01/10

Figure A.5: RGB images of the sample section, part 4.

A.2 Image results from final decision classifier, variant "allMonths"

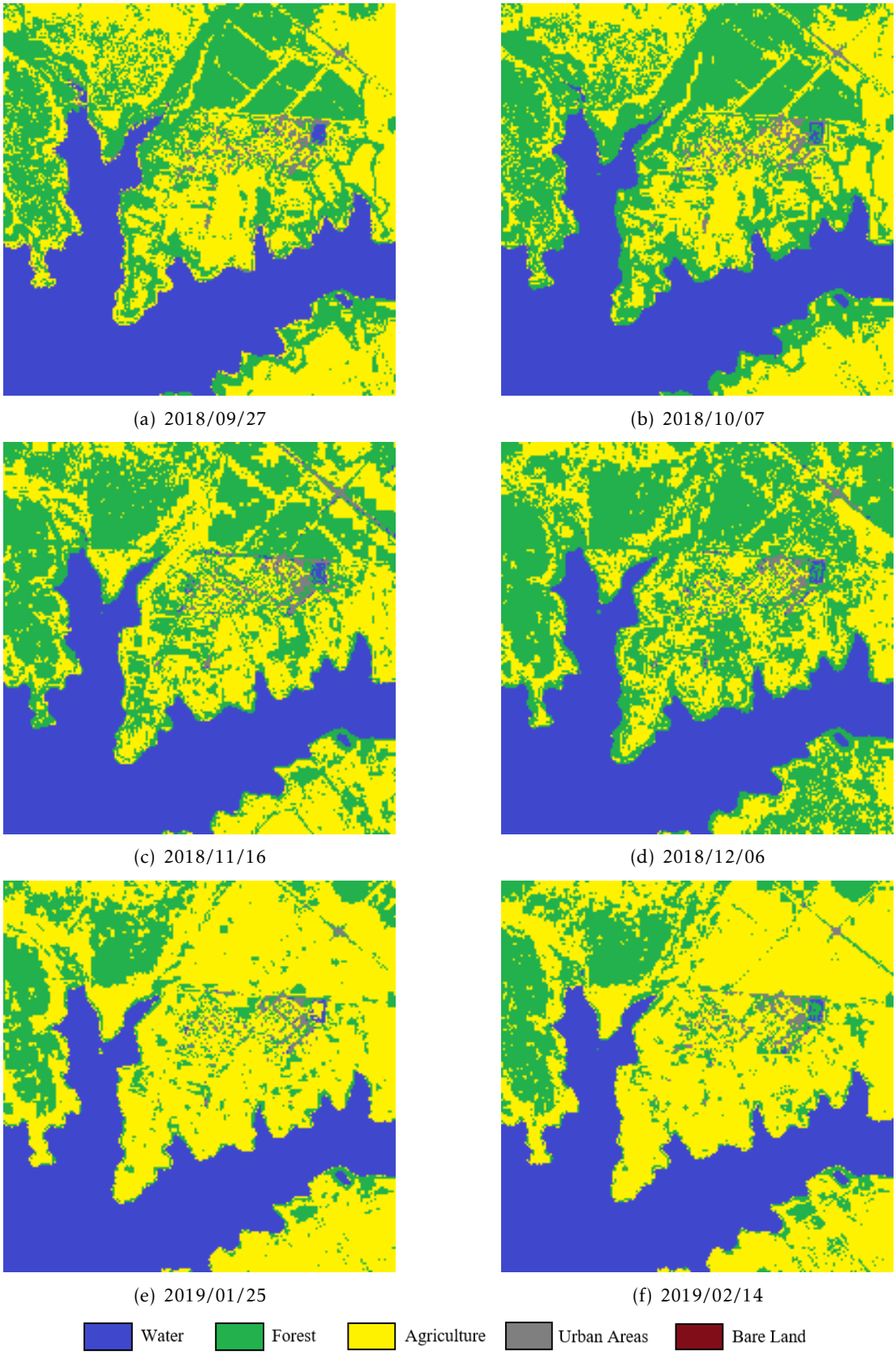


(a) 2018/03/31 (b) 2018/04/30

(c) 2018/05/10 (d) 2018/06/19

(e) 2018/07/29 (f) 2018/08/18

Figure A.6: Resulting images from final decision classifier, variant "allMonths", of the sample section, part 1.

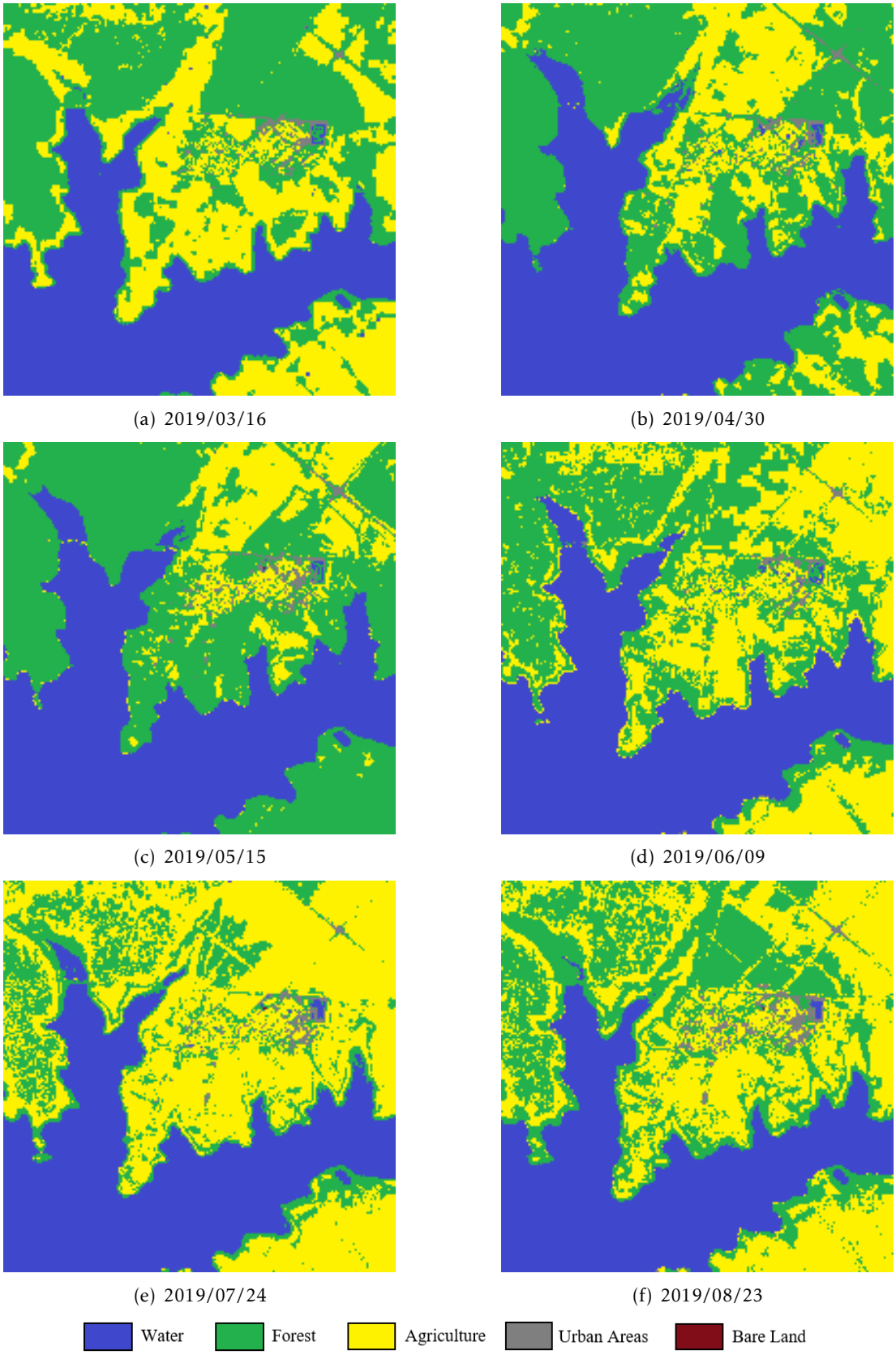


(a) 2018/09/27 (b) 2018/10/07

(c) 2018/11/16 (d) 2018/12/06

(e) 2019/01/25 (f) 2019/02/14

Figure A.7: Resulting images from final decision classifier, variant "allMonths", of the sample section, part 2.



(a) 2019/03/16

(b) 2019/04/30

(c) 2019/05/15

(d) 2019/06/09

(e) 2019/07/24

(f) 2019/08/23

Figure A.8: Resulting images from final decision classifier, variant "allMonths", of the sample section, part 3.

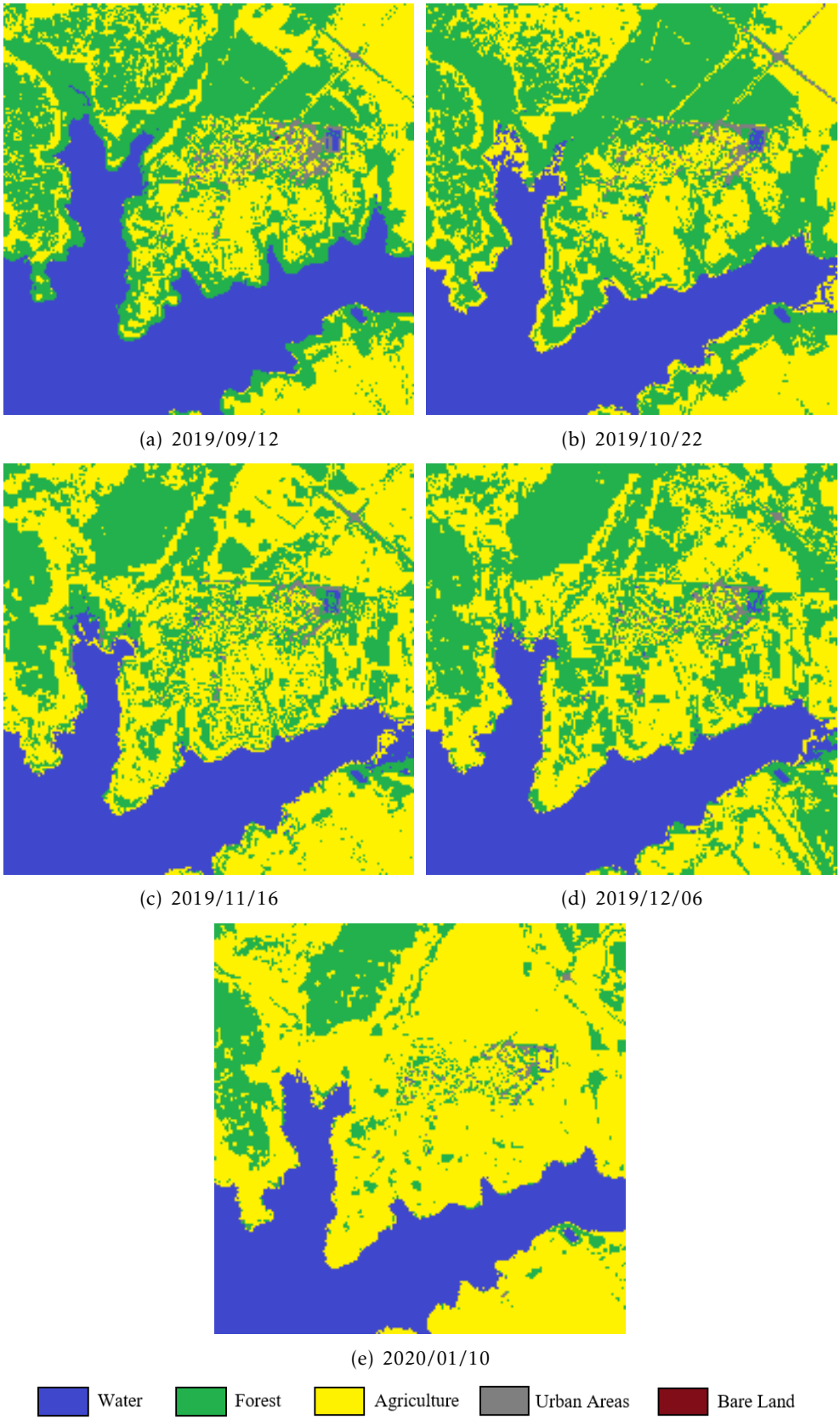
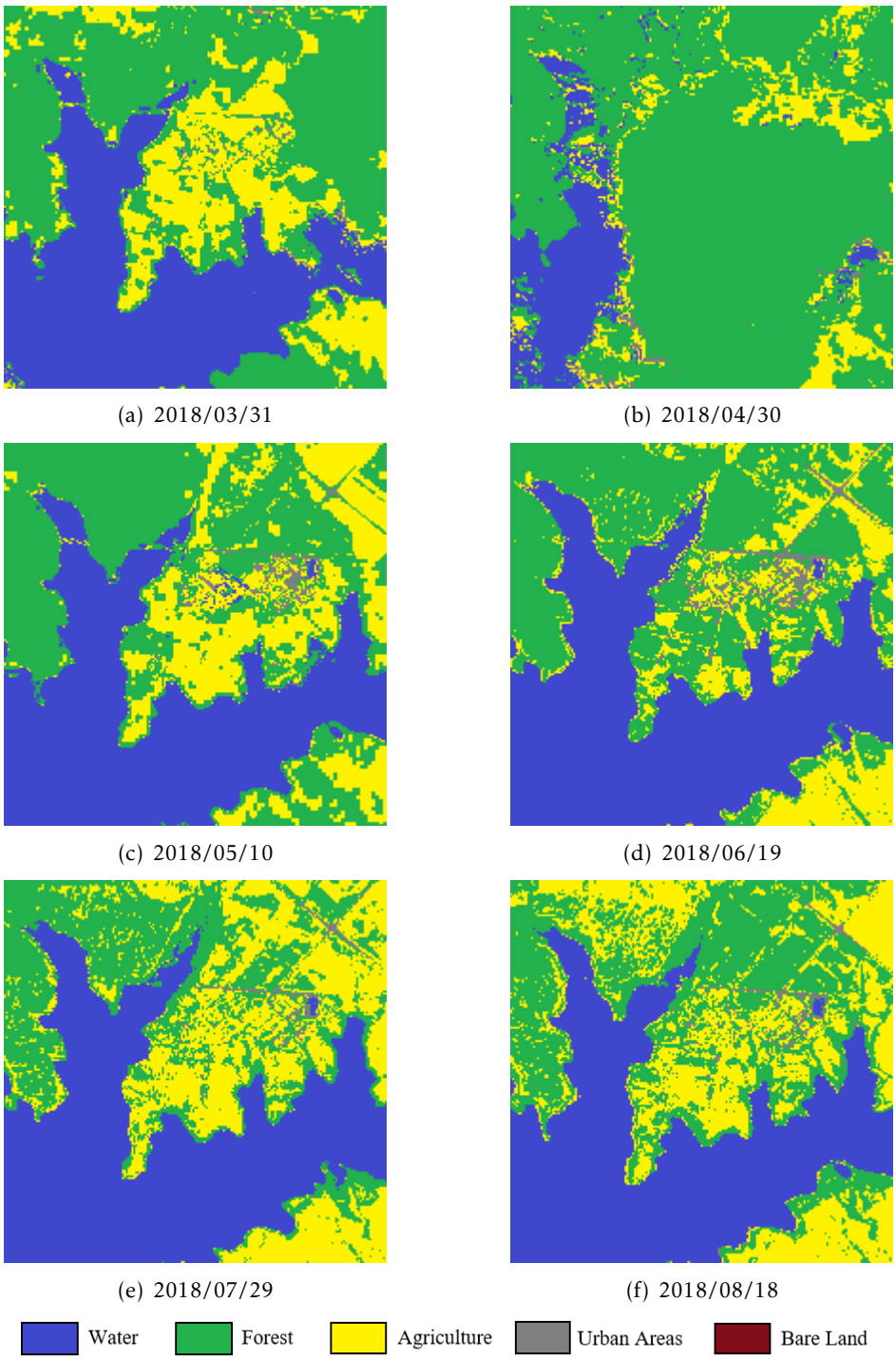


Figure A.9: Resulting images from final decision classifier, variant "allMonths", of the sample section, part 4.

A.3 Image results from final decision classifier, variant "MonthVar"

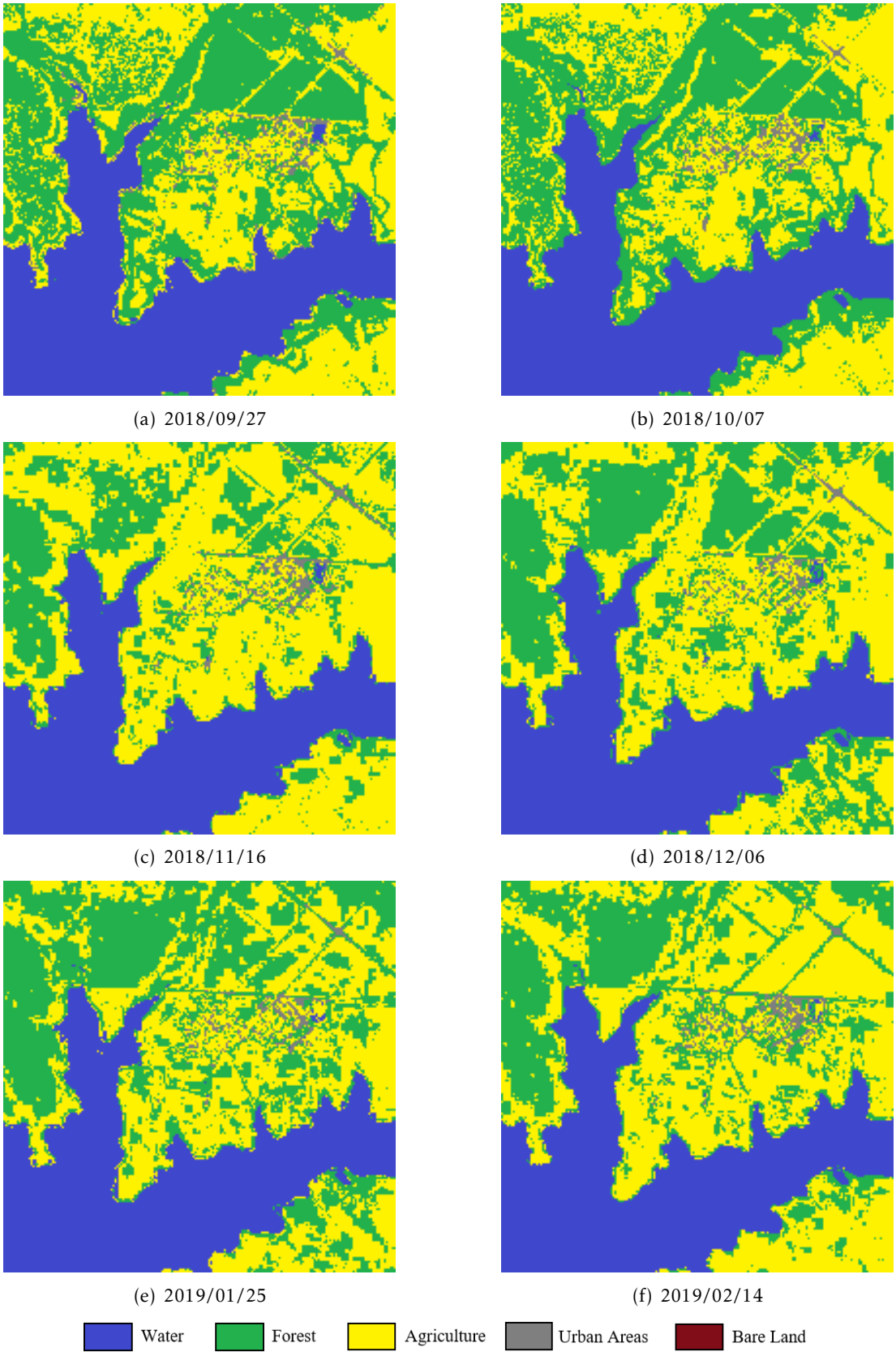


(a) 2018/03/31 (b) 2018/04/30

(c) 2018/05/10 (d) 2018/06/19

(e) 2018/07/29 (f) 2018/08/18

Figure A.10: Resulting images from final decision classifier, variant "MonthVar", of the sample section, part 1.

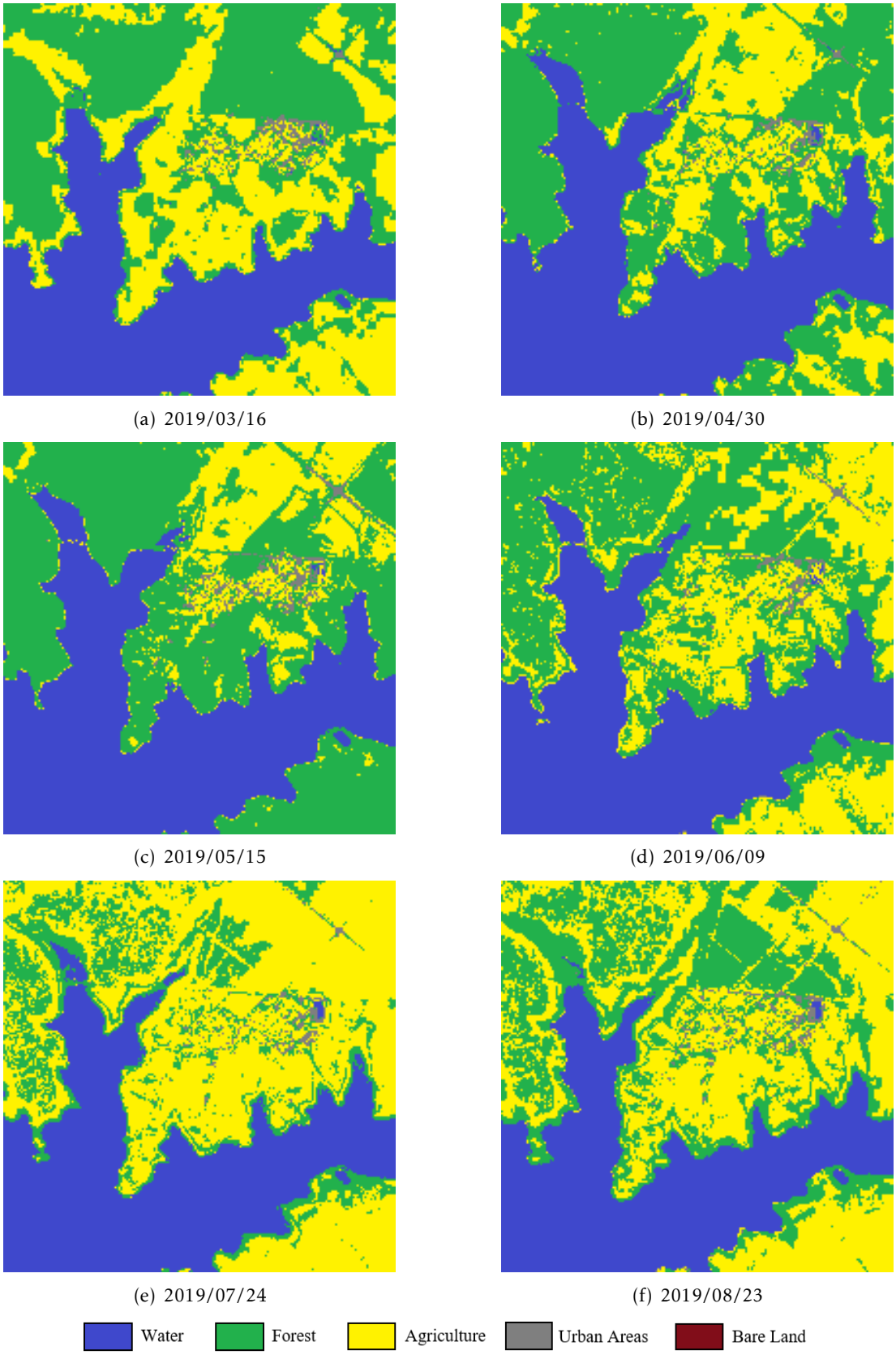


(a) 2018/09/27 (b) 2018/10/07

(c) 2018/11/16 (d) 2018/12/06

(e) 2019/01/25 (f) 2019/02/14

Figure A.11: Resulting images from final decision classifier, variant "MonthVar", of the sample section, part 2.



(a) 2019/03/16

(b) 2019/04/30

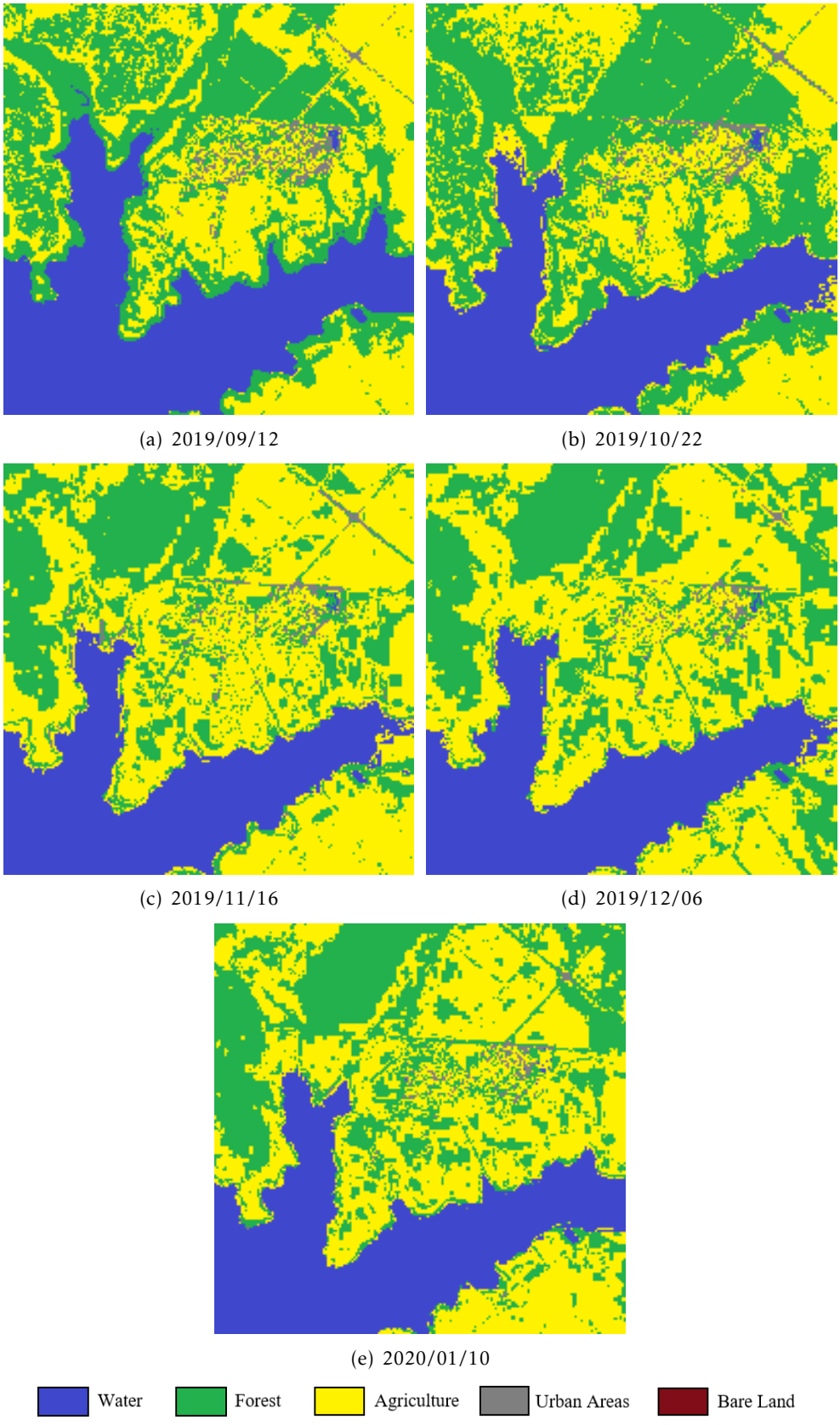
(c) 2019/05/15

(d) 2019/06/09

(e) 2019/07/24

(f) 2019/08/23

Figure A.12: Resulting images from final decision classifier, variant "MonthVar", of the sample section, part 3.



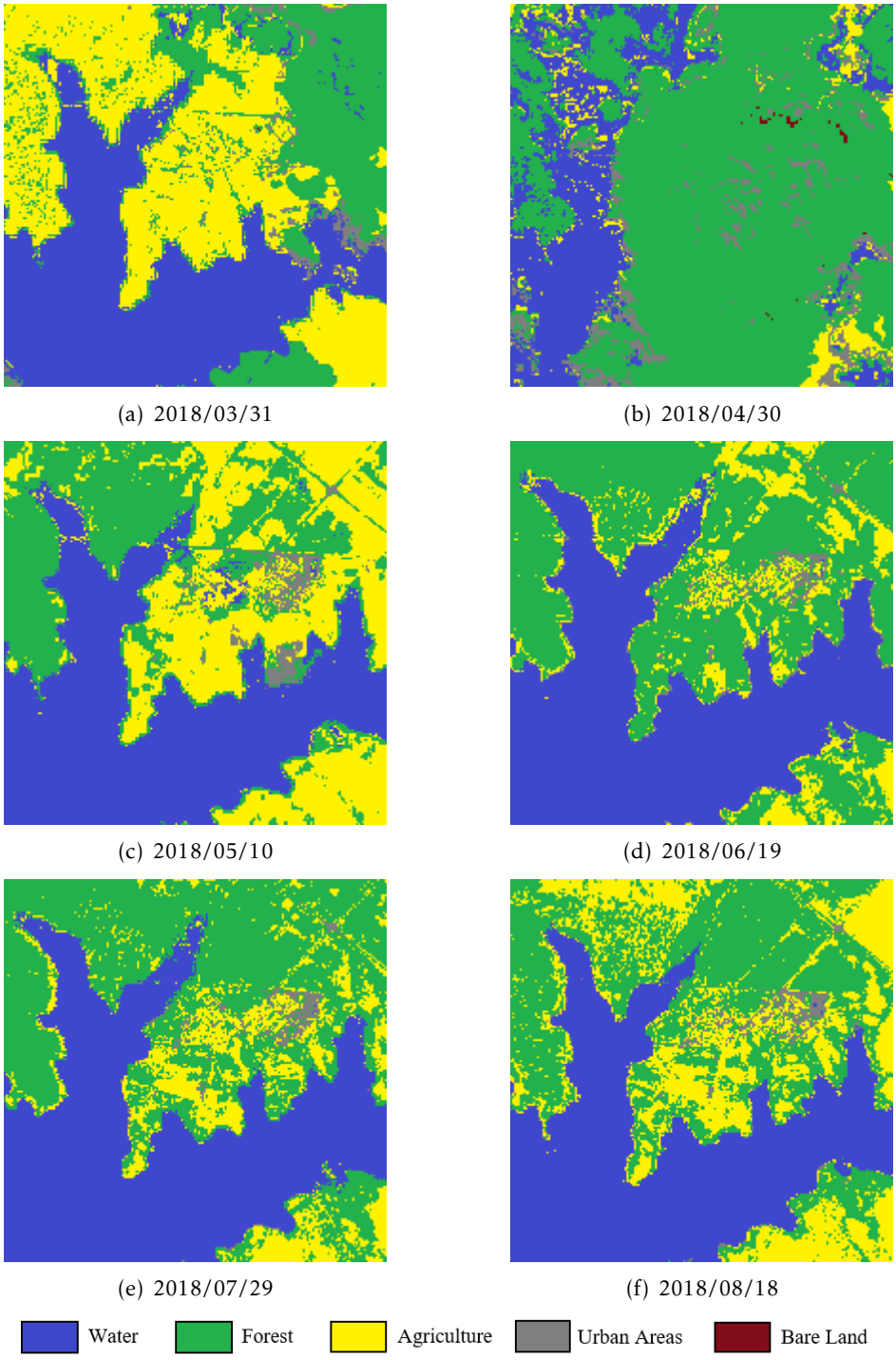
(a) 2019/09/12 (b) 2019/10/22

(c) 2019/11/16 (d) 2019/12/06

(e) 2020/01/10

Figure A.13: Resulting images from final decision classifier, variant "MonthVar", of the sample section, part 4.

A.4 Image results from final decision classifier, variant “2019”

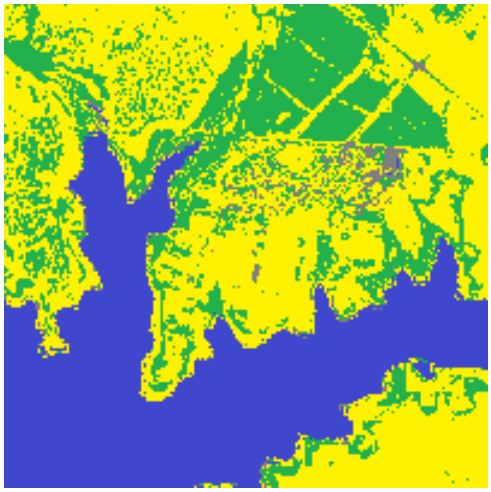


(a) 2018/03/31 (b) 2018/04/30

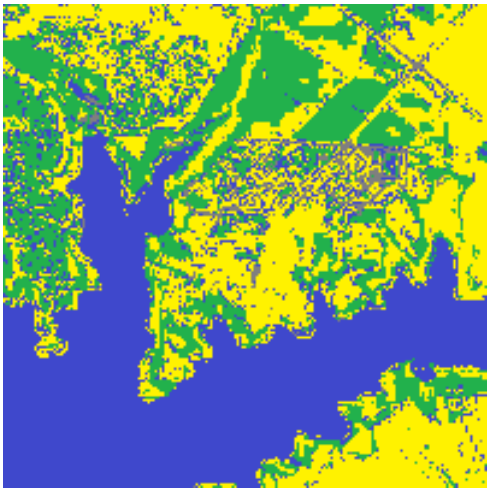
(c) 2018/05/10 (d) 2018/06/19

(e) 2018/07/29 (f) 2018/08/18

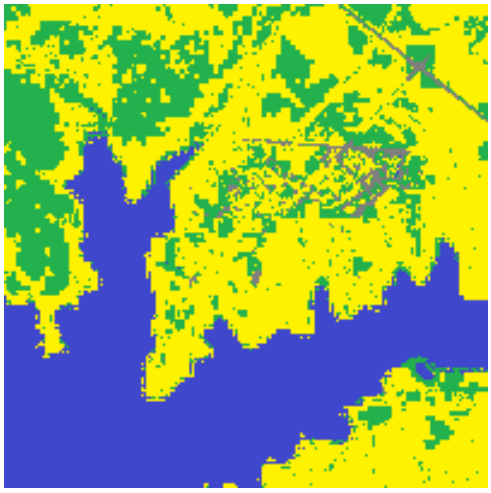
Figure A.14: Resulting images from final decision classifier, variant “2019”, of the sample section, part 1.



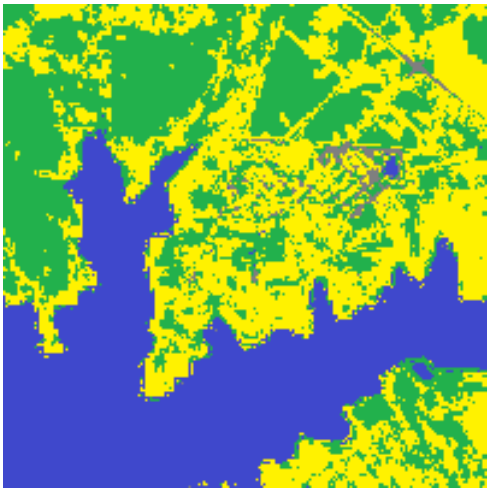
(a) 2018/09/27



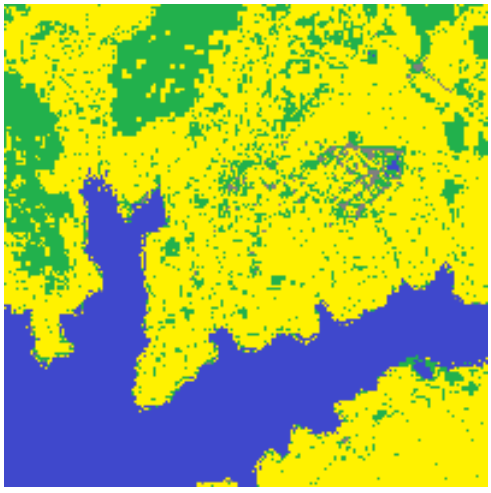
(b) 2018/10/07



(c) 2018/11/16



(d) 2018/12/06



(e) 2020/01/10

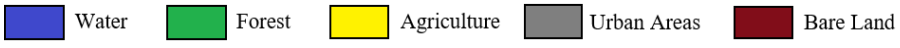
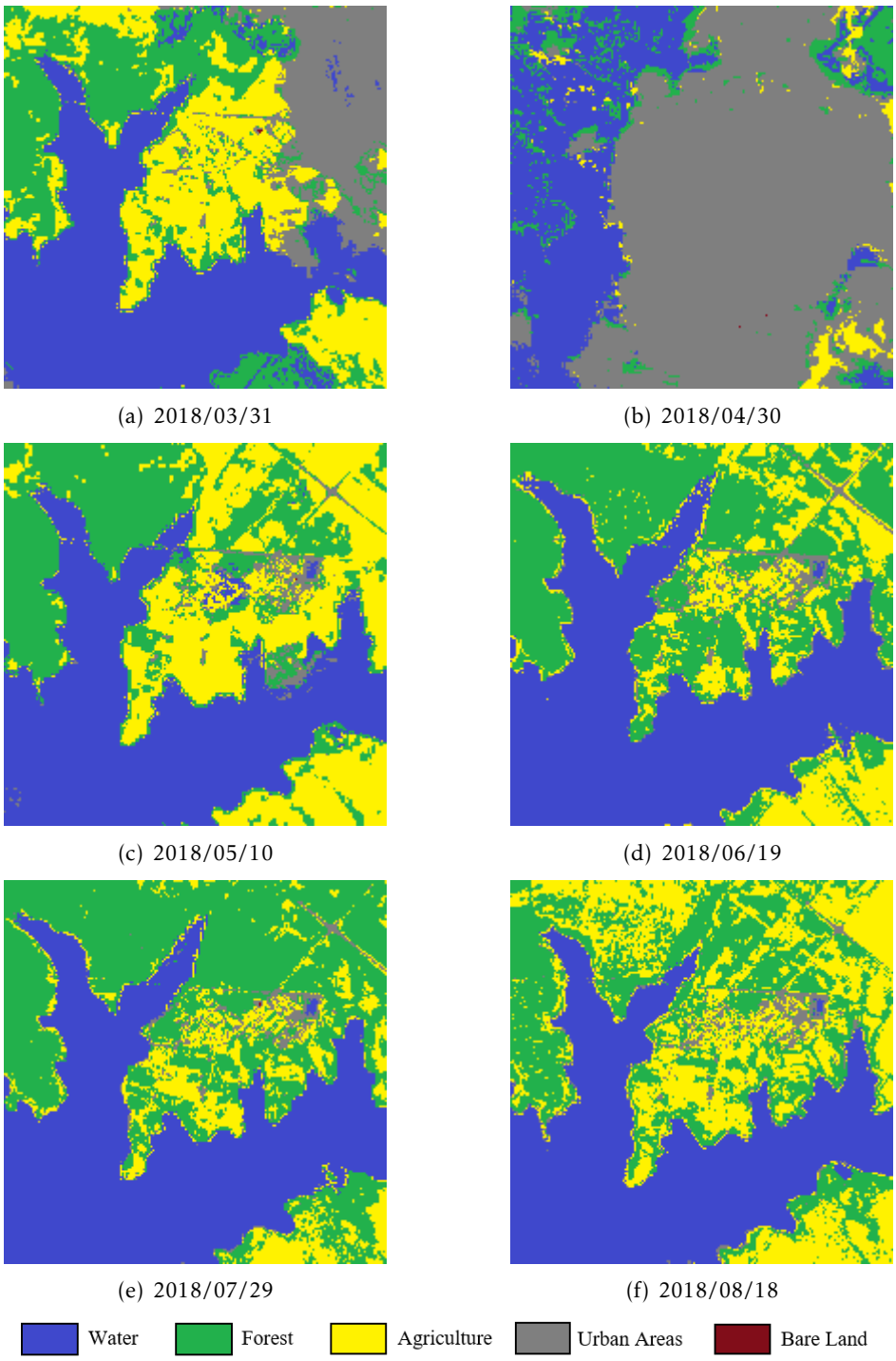


Figure A.15: Resulting images from final decision classifier, variant "2019", of the sample section, part 2.

A.5 Image results from final decision classifier, variant "2019_MonthVar"

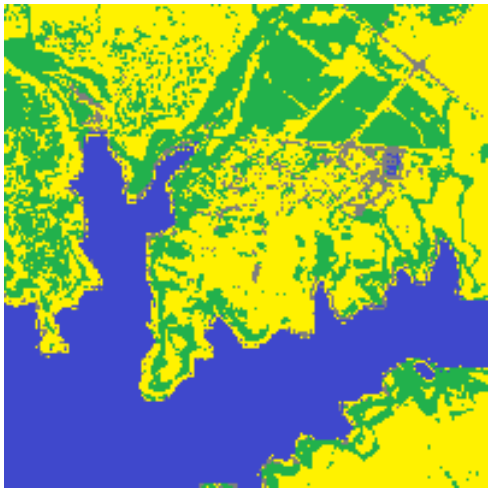


(a) 2018/03/31 (b) 2018/04/30

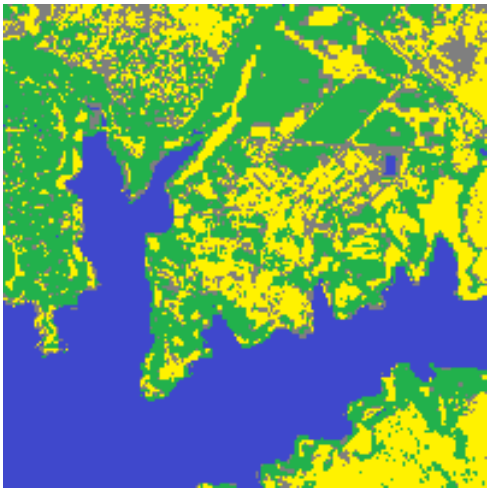
(c) 2018/05/10 (d) 2018/06/19

(e) 2018/07/29 (f) 2018/08/18

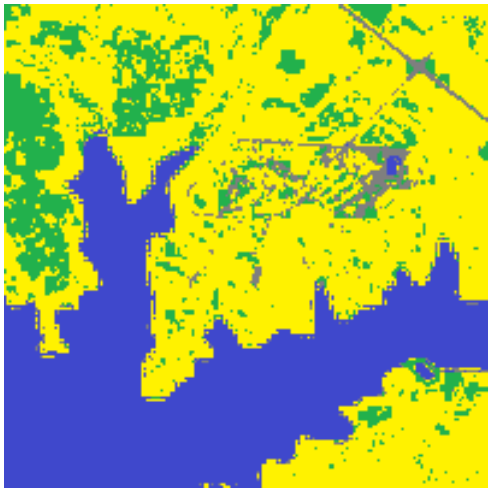
Figure A.16: Resulting images from final decision classifier, variant "2019_MonthVar", of the sample section, part 1.



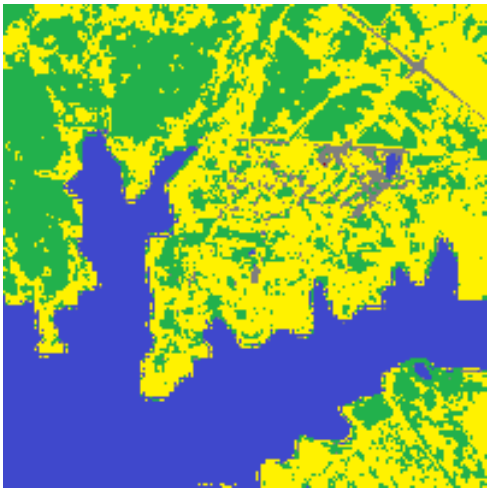
(a) 2018/09/27



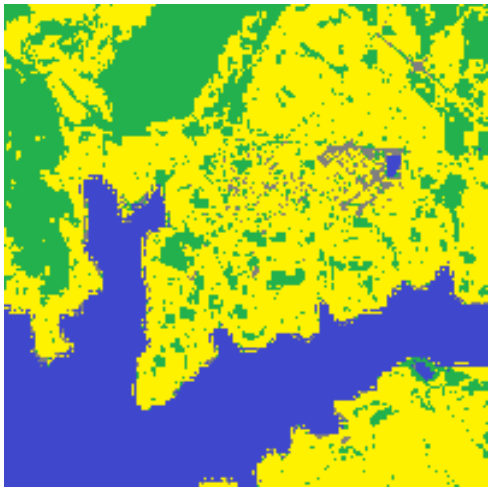
(b) 2018/10/07



(c) 2018/11/16



(d) 2018/12/06



(e) 2020/01/10

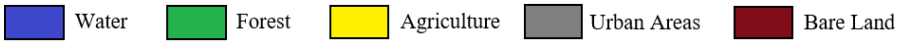


Figure A.17: Resulting images from final decision classifier, variant "2019_MonthVar", of the sample section, part 2.