MULTITEMPORAL CO-POLAR X-BAND SAR DATA CLASSIFICATION AS A TOOL FOR PADDY-RICE PHENOLOGY ESTIMATION

M.Sc. THESIS

Çağlar KÜÇÜK

Department of Applied Informatics

Geographical Information Technologies Masters Program

JUNE 2016

MULTITEMPORAL CO-POLAR X-BAND SAR DATA CLASSIFICATION AS A TOOL FOR PADDY-RICE PHENOLOGY ESTIMATION

M.Sc. THESIS

Çağlar KÜÇÜK
(706131028)

Department of Applied Informatics

Geographical Information Technologies Masters Program

Thesis Advisor: Asst. Prof. Dr. Gülşen TAŞKIN KAYA
Co-advisor: Assoc. Prof. Dr. Esra ERTEN

JUNE 2016

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ BİLİŞİM ENSTİTÜSÜ

**ÇELTİK TARLASI FENOLOJİ KESTİRİMİ
İÇİN ÇOK ZAMANLI CO-POLAR X-BANT SAR
VERİSİ ÜZERİNDE SINIFLANDIRMA YÖNTEMİ**

YÜKSEK LİSANS TEZİ

**Çağlar KÜÇÜK
(706131028)**

Bilişim UygulamalarıAnabilim Dalı

Coğrafi Bilgi Teknolojileri Yüksek Lisans Programı

**Tez Danışmanı: Yar. Doç. Dr. Gülşen TAŞKIN KAYA
Eş Danışman: Doç. Dr. Esra ERTEN**

HAZİRAN 2016

Çağlar KÜÇÜK, a M.Sc. student of ITU Informatics Institute, 706131028 successfully defended the thesis entitled "MULTITEMPORAL CO-POLAR X-BAND SAR DATA CLASSIFICATION AS A TOOL FOR PADDY-RICE PHENOLOGY ESTIMATION", which he prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :   **Asst. Prof. Dr. Gülşen TAŞKIN KAYA**   ....
Istanbul Technical University

Co-advisor :   **Assoc. Prof. Dr. Esra ERTEN**   ....
Istanbul Technical University

Jury Members :   **Asst. Prof. Dr. Gülşen TAŞKIN KAYA**   ....
Istanbul Technical University

Assoc. Prof. Dr. Esra ERTEN   ....
Istanbul Technical University

Assoc. Prof. Dr. Mustafa Ersel KAMAŞAK   ....
Istanbul Technical University

Asst. Prof. Dr. Oğuzhan CEYLAN   ....
Istanbul Kemerburgaz University

Assoc. Prof. Dr. Füsun ŞANLI   ....
Yıldız Technical University

Date of Submission :   **29 April 2016**
Date of Defense :   **9 June 2016**

FOREWORD

First of all, I would like to thank to my advisors Asst. Prof. Dr. Gülşen Taşkın KAYA and Assoc. Prof. Dr. Esra ERTEN for their assistance throughout my thesis with a perfect combination of professional and friendly relations.

I would like to thank to Prof. Dr. Juan M. LOPEZ-SANCHEZ and Onur YÜZÜGÜLLÜ for sharing and pre-processing the TerraSAR-X images. I also give my grateful acknowledges to the Scientific and Technological Research Council of Turkey (TÜBİTAK) for the financial support to the research under the project *Sulak ve Çeltik Alanların X-band ile İzlenmesi* with the project number TÜBİTAK ÇAYDAG 113Y446.

I also give my best regards to Assoc. Prof. Dr. Cecilio ANGULO-BAHON for giving me the chance to complete my studies in Spain and giving valuable comments to the research.

I have to thank to my family for growing an independent person and supporting my decisions with respect. I should also thank to some of my friends that shaped my personality as much as my family; starting from Narin FANDOĞLU, Asaf Burak ŞAFAK, Hakan SİPAHİOĞLU, Betül CELEP, Cem ÇERİBAŞI and Çağan SAZAK. Another thing to be mentioned is that I feel too lucky is having a friend such trustworthy like Yücel TORUN. Also I have to declare my special thanks to Uğur Mikail GÖKŞEN and Eleni TIMOGIANNAKIS for listening and calming me down and Selçuk YALÇIN for sharing enjoyable times in my most stressed periods. I also have to give my deepest gratitude to Orkun AYDIN and Altuğ YILDIRIM for accompanying me in the long laboratory nights. I can't miss to give my thanks to Hazal ÖZTETİKLER who helped me many times that I felt like I can't do more during the research.

For my last period of research in Barcelona; I have to thank for my biggest mate Ertan KUNTMAN, the lovely twins Oriol and Joel POCOSTALES-MERCE, Joshua Adonai PÉREZ-CAWKILL and many friends in Mas Franch and Mas Les Vinyes for their supports in my latest periods of writing the thesis.

June 2016   Çağlar KÜÇÜK

TABLE OF CONTENTS

ABBREVIATIONS

BBCH   **:** Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie
CDT   **:** Compact Decision Tree
DR   **:** Dimensionality Reduction
FS   **:** Feature Selection
HH   **:** Horizontal-Horizontal
kNN   **:** K-Nearest Neighbours
KW   **:** Kruskall-Wallis
RAR   **:** Real Aperture Radar
RBF   **:** Radial Basis Function
SAR   **:** Synthetic Aperture Radar
SBMLR   **:** Sparse Bayesian Multinomial Logistic Regression
SVM   **:** Support Vector Machines
SVM-RFE   **:** Support Vector Machines - Recursive Feature Elimination
VV   **:** Vertical-Vertical

LIST OF TABLES

LIST OF FIGURES

**MULTITEMPORAL CO-POLAR X-BAND SAR
DATA CLASSIFICATION AS A TOOL FOR
PADDY-RICE PHENOLOGY ESTIMATION**

SUMMARY

Crop monitoring with remote sensing techniques become more important as the demand of crops increases. Monitoring crop development, controlling their phenology, checking for problems and yield prediction are some of the most popular applications of remote sensing technology in precision agriculture. In those remote sensing techniques, it is also possible to use the space based sensors thanks to the new sensors' high spatial and temporal resolution. Usage of space based sensors also enables the user to make global applications, which increases the importance of those technologies even more.

Active remote sensing systems are used in this study which enables observation without affecting from daylight or weather conditions. In this study, SAR images are used which have been taken by TerraSAR-X, operating in X-band. Generally electromagnetic modelling has been applied for phenology estimation with SAR systems. Those modellings require too complex mathematical operations with complex Physics background. In order to find alternative approaches, this study is focused on determining and monitoring the phenological information of rice fields with machine learning algorithms.

As classification algorithms, Support Vector Machines (SVM), k-Nearest Neighbours (kNN) and Compact Decision Trees (CDT) were used in the experiments. In order to control the parameters effecting the classification accuracy, different cases were appended in the experiments with different number of classes, different features sets and different clustering approaches. Two different datasets were used in the experiments. One of them is images covering the rice fields around Seville, Spain with ground information covering whole cultivation period. The second dataset used in the experiments covers rice fields around Ipsala, Turkey but with limited ground information, covering two thirds of the cultivation period with less amount of fields. Therefore, all the experiments were done for Seville dataset and the best method was applied for Ipsala dataset to control the applicability of the method.

First parameter to be controlled was class labelling, which also corresponds to the precision of information obtained by classification. Four different class labelling cases were appended with 3, 5, 6 and 10 classes, having different growth intervals. Without making any supplementary achievement, the classification accuracies were higher than 70 %, which encouraged to keep the research.

Second parameter to be used was usage of feature sets. Using the two polarimetric channels that the sensor has, 9 polarimetric and 8 gray level co-occurrence features were calculated for the two channels. Then, classification performances with different features sets were compared and observed that textural features have a contribution

to the classification accuracy. With this positive information, research was extended to feature selection algorithms and the most important features were defined. It is observed that the same classification performance could be achieved with 4 features in stead of 25 features; which reduces the computational cost a lot.

Third parameter to be inspected was the evaluation of features using the pixels within the fields. In order to deal with the non-homogeneity of fields and lack of sampling data; different clustering approaches were produced and tested. In the end, classification performance was increased by removing outlier pixels and increasing the number of samples for the same ground information. To sum up, this paper explores the performance of classification algorithms for multitemporal SAR images on rice crops.

ÇELTİK TARLASI FENOLOJİ KESTİRİMİ İÇİN ÇOK ZAMANLI CO-POLAR X-BANT SAR VERİSİ ÜZERİNDE SINIFLANDIRMA YÖNTEMİ

ÖZET

Tahıl ürünlerine olan talebin global olarak artışına paralel olarak tahıl ürünlerinin uzaktan algılama yöntemleriyle ekin takibinin yapılması da popüler bir hale gelmiştir. Ekin gelişiminin takibi, fenoloji hakkında bilgi çıkartımı, hasat tahmini gibi çalışmalar uzaktan algılama ile hassas tarımın birleştirildiği önemli çalışma konularındandır. Bu konular üzerine hem yerküre tabanlı hem de uydu tabanlı çalışmalar yapılmaktadır. Büyük tarım organizasyonları drone kullanmak, traktör üzerine sensör yerleştirmek gibi yerel çözümler üzerine yoğunlaşırken global uygulamalar yapmak isteyen uluslararası kuruluşlar veya devletlere bağlı araştırma merkezleri de uydu tabanlı sensörler aracılığıyla uygulamalar geliştirmeye çalışmaktadır. Yeni nesil uydu tabanlı sensörlerin mekansal ve zamansal çözünürlüğünün yüksekliği sayesinde tarım uygulamalarında kullanımı gittikçe daha kolay hale gelmiştir.

Uydu tabanlı uzaktan algılama sistemlerinin global ve çok daha etkin bir biçimde kullanılabilir olduğunu biliyoruz. Ayrıca, aktif uzaktan algılama sistemlerinin pasif sistemlere göre çok daha kullanışlı, çevresel koşullardan etkilenmeyen ve daha detaylı bilgi toplayabilen sistemler olduğunu da göz önünde bulundurarak daha fazla önem taşıdığını rahatlıkla söyleyebiliriz. Bu gerekçeler göz önünde bulundurularak bu çalışmada uydu tabanlı bir aktif uzaktan algılama sensörü olan TerraSAR-X kullanılmıştır. TerraSAR-X, ikili polarizasyona sahip, co-polar, bir X-bant aktif Yapay Açıklıklı Radar sensörüdür. SAR sistemlerinde genellikle çok karmaşık matematiksel operasyonlar gerektiren, kompleks fizik altyapısı içeren gerisaçılım teorileri kullanılmaktadır. Bu sistemler karmaşık olmasından dolayı veri işleme yükünün de oldukça ağır olduğu çözümlerdir.

Bu karmaşık sistemlerden kaçınmak amacıyla bu çalışmada SAR verisi üzerinde sınıflandırma algoritmaları kullanılmıştır. Sınıflandırma için kullanılan algoritmalar Destek Vektör Makineleri , k-En Yakın Komşuluk ve Kompakt Karar Ağaçları kullanılmıştır. Destek Vektör Makineleri kullanılırken çekirdek fonksiyonları aracılığıyla hem doğrusal hem de doğrusal olmayan vektörler üretilmiş; sınıflandırma hem doğrusal hem de doğrusal olmayan koşullara göre yapılmıştır. Sınıflandırma başarısını etkileyen parametreleri kontrol etmek amacıyla değişik vakalar oluşturulmuş ve tek tek incelenmiştir. Bu vakalar farklı sınıf sayıları ve sınıf aralıkları, farklı öznitelik verileri ve verinin farklı kümelenme yöntemleri üzerine yoğunlaşmıştır.

Ayrıca, yapılan deneylerin global ölçekte kullanılabilir olup olmadığının kontrolü için deneyler birbirinden tamamen bağımsız iki farklı veri seti üzerinde gerçekleştirilmiştir. İlk veri seti İspanya'nın Seville şehrinin güneyinde bulunan çeltik tarlalarının görüntülerini içermektedir. Bu veri seti tüm hasat dönemini kapsayan, aynı zamanda 12 adet tarlanın çeltiklerin gelişimine ait yersel ölçümleri de içeren 12 adet görüntüden oluşmaktadır. İkinci veri seti ise Türkiye - Yunanistan sınırında bulunan İpsala

kasabasının batısında bulunan çeltik tarlalarını kapsayan 6 adet görüntü ile 8 tarlanın yersel ölçümden oluşmaktadır. İpsala veri setinin görüntülerinin tüm ekim dönemini kapsamamasından ve yeterli sayıda yersel ölçümün de bulunmamasından dolayı tüm deneyler Seville veri setinde yapılmış ve en uygun yöntem İpsala veri setinde denenip işlevsel olduğu kanıtlanmıştır.

İncelenen ilk parametre sınıf sayısı ve sınıf aralıkları olmuştur. Sınıf sayısı aynı zamanda ne kadar detaylı bir sınıflandırma yapılacağını da belirlediği için kilit niteliğindedir. Çeltiğin gelişim evresini 3, 5, 6 ve 10 sınıfa bölerek dört farklı değerlendirme yapılmıştır. Her ne kadar sınıf sayısı arttıkça sınıflandırma performansında düşüş gözleniyor olsa da herhangi bir ek bilgi veya algoritmaları geliştirici yöntem kullanmadan tüm vakalar için % 70'in üzerinde sınıflandırma performansı elde edilmiştir. Bu sonuç araştırmanın geliştirilmesi açısından cesaret vericidir.

İkincil olarak incelenen parametre ise sınıflandırmada kullanılan öznitelik setleridir. Sensörün iki kanalından (yatay-yatay ve dikey-dikey) alınan verilerle toplamda 9 adet polarimetrik öznitelik verisi üretilmiştir. Ayrıca, tarlaların doku özelliklerini incelemek amacıyla iki kanalın da ayrı ayrı Haralick öznitelik verileri çıkartılmış ve doku öznitelik verileri olarak kullanılmıştır. Her ne kadar SAR verisi çeltik tarlası gibi rastgele medyalarda doku özelliği taşımasa da problemin çokzamanlı yapısından kaynaklı doku öznitelik verileri sınıflandırma performansını yükseltmiş; bu da araştırmanın ilerletilmesi için motivasyon sağlamıştır. Sınıflandırma performansına etkisi en fazla olan öznitelik verilerini belirlemek için farklı yaklaşımlara sahip üç adet öznitelik seçme algoritması kullanılmıştır. Öznitelik seçme algoritmalarında filtreleme , sarma ve gömülü yöntemler kullanan Kruskal-Wallis, Ayrık Bayesçi Multinomial Biçimsel Regresyon ve Destek Vektör Makineleri Gereksiz Öznitelik Elemesi yöntemleri kullanılmıştır. Deneyler sonucunda beklendiği üzere en önemli öznitelik verileri polarimetrik veriler olarak belirlenmiştir. Ancak, şaşırtıcı biçimde doku verilerinin çokzamanlı problemlerde sınıflandırma başarısını arttırabildiği gözlenmiştir. Öznitelik seçme deneyleri aynı zamanda 9 polarimetrik ve her iki kanal için 8 adet üretilmiş doku verileriyle toplamda 25 öznitelik verisiyle elde edilen sınıflandırma performansının aynısını yalnızca 4 öznitelik verisiyle de elde edebilmiş; işlem maliyetini ciddi bir biçimde azaltılabildiğini göstermiştir.

Üçüncü incelenen parametre ise öznitelik verilerinin nasıl üretildiğiydi. Sensörden alınan veri piksel bazında olup sınıflandırma ise tarla bazındaydı. Yüksek mekansal çözünürlük ve tarlaların boyutları göz önüne alındığında piksel ölçeğinden tarla ölçeğine geçmenin incelenmesi gereken bir konu olduğuna karar verildi. Tarımsal uygulamalarda önlenmesi çok zor birçok sorun nedeniyle tarla içindeki düzensizliğin önüne geçmek, aynı zamanda sınırlı miktarda olan eğitim verisini arttırmak için üç farklı kümeleme seçeneği belirlendi ve deneyler yapıldı. Aykırı piksel değerlerinden kurtulmak için tarlaların yarıdan fazlasını içeren tek küme kullanımı olumlu sonuç vermese de tarlaları farklı parçalara ayıran ve bunun yanında aykırı değerleri de çıkartan bir algoritma sınıflandırma performansını oldukça iyileştirdi. Tüm bu analizlerin ardından ise belirlenen en iyi sınıflandırma yöntemi ilk veri grubundan tamamen bağımsız olan İpsala verisi üzerinde denenip yöntemin geçerliliği test edilmiş ve başarılı sonuçlar alınmıştır.

Sonuç olarak; bu çalışmada SAR verisi üzerinde sınıflandırma algoritmaları kullanarak çokzamanlı görüntülerde çeltik tarlalarının gelişim sürecinin takip edilebileceği

gösterilmiş ve sınıflandırma performansının yeterince yüksek olduğu gözlenmiştir. Neredeyse tüm deneylerin sonunda elde edilen sınıflandırma performansı Kappa sayısı biriminde önemli derecede uyuşma olduğunu ortaya koymuştur. Bu sonuçlar ışığında elbette ki geliştirilecek uygulamanın gereksinimlerine bağlı olmak koşuluyla sınıflandırma algoritmalarının pirinç gelişim evrelerinin belirlenmesinde kullanılabilir olduğunu söyleyebiliriz.

1. INTRODUCTION

1.1 Motivation

From the beginning of the written history, people have been trying to control and optimize the yields in farming. It forced people to make calculations, find mathematical relations, even invent calenders. With the 5000 years of scientific progress, we are now capable of controlling agricultural fields with various methods, including space technology. Therefore, precision agriculture, which is a farming management system from space, has become a cornerstone in sustainable agriculture. One of the aspects of this kind of agronomy applications is phenology monitoring. This is the reason that phenology monitoring in global scales with space based remote sensing technologies is quite a hot topic to research and attracts a lot of researchers with its importance and positive motivation for the Earth.

1.2 Objective

Although phenology estimation with SAR images is not a new application area; most researchers have generally focused on complex electromagnetic theories that rely on physical mechanics of the signals. Even though these methods are quite powerful, they require high amount of calculations which could be even difficult for high performance computer systems. The objective of this study is to determine the phenology of rice without dealing with electromagnetic theory. Therefore, abilities of machine learning algorithms were investigated on SAR data to classify the phenological development of rice. Using polarimetric information as features for classification algorithms, encouraging results were obtained, and the research was expanded on analysing the effects of different parameters of the classification methods.

1.3 Literature Review

For many years, there has been many studies focusing on using SAR sensors in agriculture. Thanks to their capability of detecting small changes without being affected from the weather and time conditions, SAR sensors, which are elements of active remote sensing systems, are quite favourable to be used in agriculture. Those studies could be divided into two parts as deterministic and statistical studies.

Deterministic methods model the scattering mechanism of the signal within the target. They depend on electromagnetic theory and the physical/geometric properties of the target. Some examples of those studies could be seen in [6–8]. The main aim of these studies is to build a function of scattering with physical parameters like canopy height, leaf width, leaf area index, angle of leaves, density of leaves, and etc. However, building those kind of functions require complex computational operations to generate Monte Carlo simulations. Another aspect of deterministic methods are interferometric SAR systems. They require two SAR images to control the change of scattering and build interferograms. Some recent studies like [9, 10] show that those methods are also quite successful although there is still not any operational usage in agriculture.

Statistical methods, which are more simple than the deterministic ones, had also been used in agricultural applications with SAR sensors as can be seen in [11]. By combining multiple polarizations, quite successful results can be obtained. In [5, 12, 13], parameters obtained by multiple polarimeters were used by simple thresholding methods for different band widths. However, some studies showed that thresholding methods do not always give successful results [14]. Later, combination of backscattering models and image processing was used in some studies. Phenology estimation problem was tried to be solved by dynamic modellings like Kalman filtering [15] and Particle filtering [16]. Those studies were also successful thanks to the high temporal and spatial resolution of the new sensors.

In terms of using machine learning algorithms in remote sensing, various studies have been done for active and passive sensors. Although the studies have started by using passive remote sensing sensors [17–19], some active sensors are also used in various types of machine learning algorithms [20–22]. Besides, there are various studies about specific algorithms or specific combinations of some type of algorithms which have been used for both active and passive systems [23]. As the dimensionality of the

data to be processed is getting bigger and the complicity of operations are increasing; dependency of machine learning in any field of research is inevitable.

1.4 Structure of Thesis

This study is explained in four chapters in this thesis. In Chapter 1, an introduction is given with the motivation of the study and the objectives to be reached followed by the literature review for the research. Chapter 2 is devoted for the theoretical background of the study and divided into two parts to explain SAR fundamentals and machine learning fundamentals separately. In the part of SAR fundamentals, a brief explanation of electromagnetic scattering mechanism is given. Then the methods used for extraction of information, which is to be used as features in classification, are explained in detail. In the second part of Chapter 2, machine learning fundamentals that need to be known for the study are given under specific titles as classification, feature selection and segmentation with necessary explanations. Chapter 3 gives information about the experimental design and results obtained. Each parameter affected the classification performance is explained with their definition, reasons and results in specific subsections. Lastly, the thesis is concluded with Chapter 4 by making a discussion on the experimental results and by pointing out possible future extension of the study.

2. THEORETICAL BACKGROUND

All the theoretical foundation of the study are given in this chapter, which is divided into two sections as fundamentals of SAR and machine learning. First section starts with a brief theoretical information about SAR basics, then continues with the electromagnetic scattering mechanism and features extracted from SAR images with their physical meanings.

Second section is divided into three subsections according to the type of machine learning algorithms used in this study, which are classification, feature selection and segmentation. First subsection is dedicated for SVM with linear and nonlinear kernels, kNN and CDT are explained as classification algorithms, second subsection is dedicated for KW, SVM-RFE and SBMLR are explained as feature selection algorithms, and the last subsection is dedicated for clustering as a method for segmentation.

2.1 SAR Fundamentals

Remote sensing sensors are divided into two parts as active and passive sensors. Passive sensors measure only the radiated waves of an exterior energy source, while active sensors measure the radiated waves of a specific energy source, which is the electromagnetic wave of the sensor itself. Taking pictures without flash can be given as an example of a passive system while taking a picture with a flash can be regarded as an active system.

One of the most popular applications of active remote sensing systems are RAdio Detection And Ranging (RADAR) systems. Today, it has still been named as radar although different waves are being used including the radio waves. After this acronym became popular and started being used in a lot of places, the capitalization of letters were lost and today a lot of people use this acronym without knowing its meaning.

In radar systems, the physical properties of the system like range, angle or velocity of an object is determined by specific, discrete radio waves that named as signals. Then, imaging can be done when this operation is repeated in multi dimensional space. Another concept to be mentioned is depending on the way to provide the aperture, radar systems are divided into two groups as Real Aperture Radar (RAR) and Synthetic Aperture Radar (SAR).

2.1.1 Electromagnetic scattering mechanism for vegetation

TerraSAR-X is the name of the space-borne sensor that is used in this study. As it can be understood from its name, the signals used to make the imaging are in X-band; which has a wavelength range of 2.4 to 3.8 centimetres. Since the ability to detect the changes is directly related with the range of wavelength; use of X-band sensors is quite suitable to monitor the small changes of the crop.

The theory behind the SAR sensors are one of the most complex topics in remote sensing since it needs a strong knowledge about a wide variety of topics from mathematics to physics. The distribution of the values in a dataset can be analysed mathematically while the mechanics of the scattering can be studied in terms of electromagnetic theory, which extends from physics. It is possible to detect anomalies of an area by using statistical operations on a data even though the physical dynamics of the system is unknown. If physical background of the interactions are to be analysed; the subject shifts to electromagnetic theory. In order to understand the polarimetric dynamics, extensive experience on electromagnetic theory is required. Very briefly, due to the interaction between the signal and the object, propagation of the signal, which is a sinusoidal movement in 2-dimension, changes in terms of phase and direction. Those directions are defined by the names of the channels of the sensors. Depending on the direction of the sent and received signals; channels are named as HH, VV, HV and VH where H and V represents horizontal and vertical directions, respectively. First letter of the channel code gives the direction of the transmitter, and the second gives the direction of the receiver. The information extracted about the object can be increased by using multiple channels. In this study, only HH and VV channels are used.

Another important topic to mention is scattering. Scattering is divided into three types as surface, volume and multiple volume-surface scattering; depending on the physical and geometrical mechanism of the scatter. In surface scattering, it is assumed that scattering can only be seen between the object and the surface; which is a mechanism of reflection depending on the characteristics and geometry of the surface. In volume scattering, signal partially penetrates into a volumetric media with particles having a size around the wavelength. Lastly, in cases of having both of the scattering mechanisms, the concept of multiple scattering occurs. To explain the case in this study in a very brief way, the dominant scattering mechanism was in a change of surface to volume scattering; as the fields were under water, followed by another surface scattering because of the body of the crop as small vertical structures and volumetric scattering as crop gets a volumetric shape. During this progress, not only backscattering values but also polarization properties are effected during penetration; which gives the possibility to detect small changes of the crop.

2.1.2 Features extracted and their physical meanings

Features that used in the experiments can be divided into two groups as Polarimetric and Textural features. While polarimetric features were calculated by the polarimetric responses of each of the two channels, textural features were calculated by spatial variation of some of these polarimetric features.

2.1.2.1 Polarimetric features

Nine polarimetric parameters were calculated and used as feature in the classification by using the two channels of the sensor. As the first step, the noisy data was smoothed by a boxcar filter. Briefly, boxcar filters give an averaged value to each pixel considering its neighbours. Dimension of the box was selected as 11 x 11 pixels since the data were quite noisy.

First feature to be mentioned is backscattering values. Since the sensor used in the study is a dual-pol sensor, there are two different channels to be investigated. Backscattered values from the signals from HH and VV channel are calculated as shown in Equation (2.1). In Equation (2.1), *hht* and *vvt* represents the smoothed complex values received by the HH and VV channels respectively and $\otimes$ represents

conjugation.

$$C_{HH} = 10 * log(hht \otimes hht) \ \ and \ \ C_{VV} = 10 * log(vvt \otimes vvt) \tag{2.1}$$

Another important parameter that used as a feature for classification in the study is the ratio of backscattered values of the two channels, formulated in Equation (2.2).

$$R_{HV} = 10 * \log(C_{HH}) - 10 * \log(C_{VV}) \tag{2.2}$$

In order to obtain backscattering values in a more distinguishable way, those values were also calculated in another space. Pauli channels are highly popular in studies in electromagnetic theory and first Pauli channel is calculated by the difference of backscattering values while second Pauli channel is calculated by the summation of them. Backscattering coefficients in Pauli channels are shown in Equation (2.3).

$$T_{11} = \frac{|C_{HH} - C_{VV}|}{2} \ \ and \ \ T_{22} = \frac{|C_{HH} + C_{VV}|}{2} \tag{2.3}$$

Another parameter that used as a feature is the coherence between the channels. Equation of the parameter is shown in Equation (2.4).

$$Cohr = \frac{(hht \otimes vvt)}{\sqrt{(hht \otimes hht) * (vvt \otimes vvt)}} \tag{2.4}$$

Since attenuation is expected to be different in each of the two channels, the phase difference between the channels is an important parameter to be used in the study. Formulation of the parameter is given in Equation (2.5).

$$P_{dif} = \arctan \left(\frac{hht \otimes vvt}{C_{HH} * C_{VV}} \right) \tag{2.5}$$

Last two polarimetric features used are entropy (H) and alpha angle (α). Those parameters are produced by a modified version of a special decomposition method called entropy/anisotropy/alpha angle $(H/A/\alpha)$ Decomposition. This decomposition is done to the Hermitian averaged coherency matrix, T_3. Since the sensor used in this study has two channel, the decomposition is limited with H/α Decomposition and calculations are done in a Hermitian matrix (T_2) of 2x2 rather than 3x3, as shown in Equation (2.6).

$$T_2 = \begin{bmatrix} (hht + vvt) \otimes (hht + vvt) & (hht + vvt) \otimes (hht - vvt) \\ (hht - vvt) \otimes (hht + vvt) & (hht - vvt) \otimes (hht - vvt) \end{bmatrix} \tag{2.6}$$
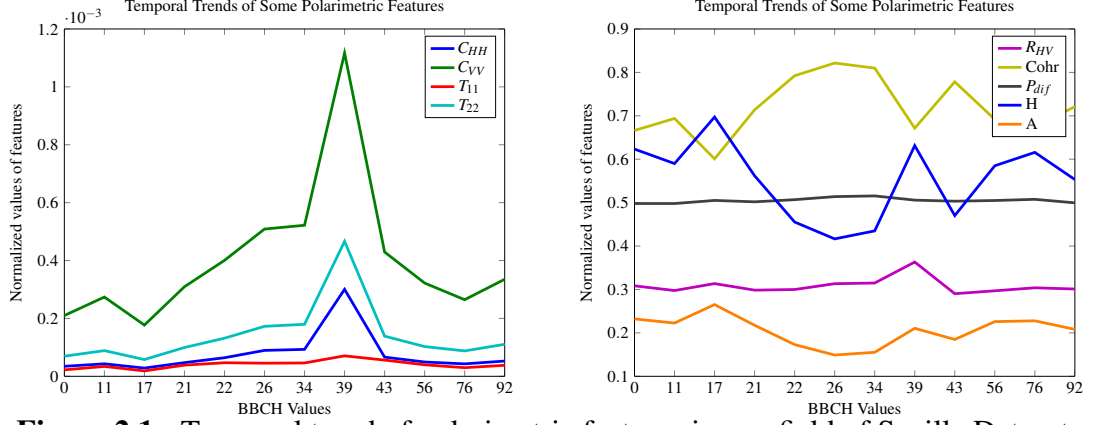
8

Figure 2.1 : Temporal trend of polarimetric features in one field of Sevilla Dataset.

After calculating the eigenvalues (λ) and the eigenvectors υ of T_2, the probabilities (P_1) are calculated as shown in Equation (2.7).

$$P_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{2.7}$$

Following this step, H and α can be calculated like shown on Equations (2.8) and (2.9), respectively.

$$H = -P_1 * log\left(\frac{P_1}{2}\right) + (1 - P_1) * log\left(\frac{1 - P_1}{2}\right) \tag{2.8}$$

$$\alpha = P_1 * \arccos(|\upsilon(1)|) + (1 - P_1) * \arccos(|\upsilon(2)|) \tag{2.9}$$

Normalized versions of the temporal trends of the polarimetric features with respect to the BBCH values for one field in Sevilla dataset are shown in Fig. 2.1. As seen from the figure, separability of features is quite difficult considering the temporal similarity of the features.

2.1.2.2 Textural features

Textures of the received signals are also used as features in classifications. Although vegetative structures completely have random backscattering, means no persistence of texture, those parameters are calculated. Since the target does not reflect much textural behaviour, the operations for extraction of features were done with the smallest size of neighbouring cells, with a window size of 3x3. Eight of the Haralick's features were calculated for each channel. Those parameters are also called as gray-level co-occurrence parameters since all the variables are fit in a distinct gray levels to be used in imaging. Haralick's features are listed with their equations in Table 2.1, and their notation is also explained below.

9

Table 2.1 : Equations of the Haralick's features.

$\text{Mean} = \frac{1}{IJ}\sum_i\sum_j p(i,j)$	$\text{Entropy} = -\sum_i\sum_j p(i,j)log(p(i,j))$
$\text{Variance} = \sum_i\sum_j (i-\mu)^2 p(i,j)$	$\text{Second moment} = \sum_i\sum_j \{p(i,j)\}^2$
$\text{Homogeneity} = \sum_i\sum_j \frac{1}{1+(i-j)^2} p(i,j)$	$\text{Correlation} = \frac{\sum_i\sum_j (ij)p(i,j)-\mu_x\mu_y}{\sigma_x\sigma_y}$
$\text{Contrast} = \sum_{n=0}^{N_g-1} n^2 \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \right\}$	$\text{Dissimilarity} = \sum_{n=0}^{N_g-1} n \left\{ \sum_{i=1}^{N_g}\sum_{j=1}^{N_g} p(i,j) \right\}$

- R : number of neighbouring resolution cell pairs (area of the box).

- $p(i,j)$: the (i,j) value of the normalized gray-level matrix, $P(i,j)/R$.

- N_g : number of discrete gray levels to be used in imaging.

- $n = |i-j|$

- $\{\}$ the operation of calculating angular second moment.

- μ_x, μ_y : mean values of p_x and p_y.

- σ_x, σ_y : standard deviations of p_x and p_y.

All the features were calculated for each channel separately. Therefore, twenty five features were produced to be used as features in machine learning applications that are nine polarimetric, eight HH texture and eight VV texture. Normalized values of the temporal trends of the textural features with respect to BBCH values are given in Fig. 2.2. Note that, the trends of the features are not very distinctive but it might be possible to obtain distinguishable feature sets with a good selection of features.

2.2 Machine Learning Fundamentals

The second half of the study is based on machine learning applications. Since remote sensing applications always cover big issues, researchers always face with big data problems that cannot be handled by manual operations. Moreover, most of the important mathematical operations are too complex to be solved manually. Therefore, it is a must to use a computer technology to perform mathematical expressions in remote sensing applications. Several studies have been done in machine learning to handle problems raised in remote sensing and geographic information systems [17, 19, 21, 24–26]. The algorithms that are used in this study as machine learning
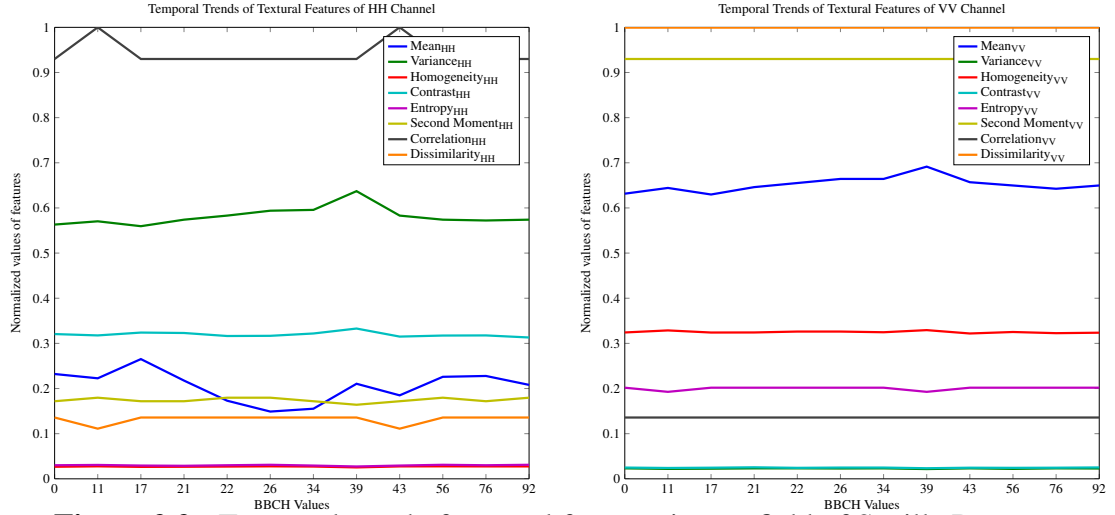
Figure 2.2 : Temporal trend of textural features in one field of Sevilla Dataset.

methods can be grouped as classification, feature selection, feature extraction and segmentation algorithms; which have been explained in the following subsections.

2.2.1 Classification algorithms

Supervised classification methods were used which require class labels of the data in the learning process. The features explained in Sections 2.1.2.1 and 2.1.2.2 are calculated for each field whose ground information about their phenological development has already been known. Afterwards, they were labelled, in other words classified based on their ground information to be used in ML algorithms. Basically, a phenomena is being learned by a machine depending on some properties, features, and their identities, class labels, to be able to classify new data having the same features. Three popular and prospering supervised classification algorithms were used in this study that are Support Vector Machines (SVM), k-Nearest Neighbours (kNN) and Compact Decision Trees (CDT).

2.2.1.1 Support vector machines

SVM is a classifier that builds a separating hyperplane between two classes and makes classification according to that separating hyperplane [1, 27, 28]. To explain this in a simple way; let's assume that we have data with two features and two classes that can be separated linearly. If the hyperplane passes from the origin, the equation of the hyperplane could be written as $\mathbf{w}^T\mathbf{x} = 0$ where $\mathbf{w}$ is the function of the linear line and $\mathbf{x}$ is the location of the samples in Fig. 2.3. In case of shifting the hyperplane from origin,
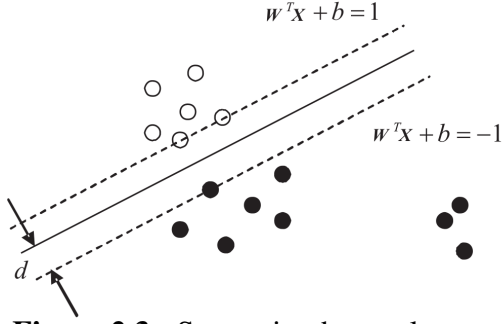
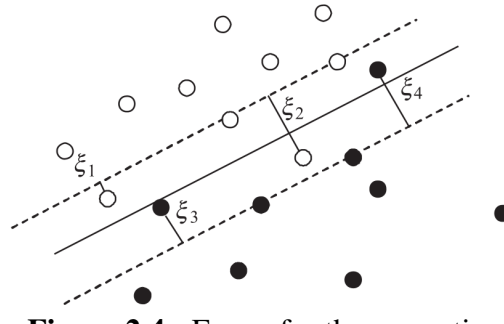**Figure 2.3** : Separating hyperplane and the margins [1].

Figure 2.4 : Errors for the separating hyperplane [1].

a constant, b, should be added into the equation of hyperplane, resulting in $\mathbf{w}^T\mathbf{x} + b = 0$. In order to minimize the misclassification, which is due to misplacement of the separating hyperplane that leaves the samples in wrong sides, the optimum place for the hyperplane should be found by adjusting the parameters $\mathbf{w}$ and b. To minimize the misclassification, the hyperplane should be located as far as possible from the nearest samples, which are named as support vectors, while keeping all the samples in the correct side. We can continue by defining a margin with a distance d and placing it to the both side of the hyperplane, as shown with the dotted lines in Fig. 2.3. Now, the objective is to maximize the distance d, whose equation is $\mathbf{w}^T\mathbf{x} + b = \pm 1$. In order to achieve this, all the samples should satisfy Equation (2.10) where y_i is the label of the sample $\mathbf{x}_i$.

$$y_i[\mathbf{w}^T\mathbf{x}_i + b] \geq 1 \quad \text{where} \quad y_i \in \{+1, -1\} \tag{2.10}$$

Since the margin is expressed as $d = 1/||\mathbf{w}||^2$, the maximization problem is converted to a minimization problem and then solved by using Lagrangian multipliers. Now the new equation is (2.11) as shown below:

$$L_p = ||\mathbf{w}||^2 \quad \text{subject to} \quad y_i[\mathbf{w}^T\mathbf{x}_i + b] \geq 1 \tag{2.11}$$

Generally, Equation (2.11) can not be satisfied because of the irregularity of the data. Therefore, an error term called slack variable is introduced to the Equation (2.11), yielding a new optimization problem:

$$L_p = ||\mathbf{w}||^2 + C\sum_{i=1}^{N} \xi_i \quad \text{subject to} \quad y_i[\mathbf{w}^T\mathbf{x}_i + b] \geq 1 - \xi_i \tag{2.12}$$

12

The setting of the margins is named as hard-margining and soft-margining depending on the permit of error or not. In other words; if small errors is permitted, ξ is greater than 0, the problem is handled as soft-margining. Even it is possible to make classification with hard-margining, without the error term, it is not guaranteed to have a better classification performance than the soft-margining approach. Since ξ corresponds to the errors inside the margin as shown in Fig. 2.4, which is a distance in analytic terms in the two featured example, ξ can not be less than 0. For the samples inside the margin, $0 < \xi < 1$ should be satisfied. As the samples approach outside of the margin, ξ converges to 0 while ξ greater than 1 for the misclassified cases. In order to optimize the problem between maximization of the margin and minimization of the errors, a penalty coefficient named as C is added in Equation (2.12). Since it is a classical constrained optimization problem, Lagrange optimization procedure can be applied by using several Lagrange multipliers, α_i. Equation (2.12) is converted into its dual problem to be maximized as shown below in Equation (2.13).

$$L_d = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j \mathbf{x}_i^T \mathbf{x}_j + \sum_{i=1}^{N}\alpha_i \tag{2.13}$$

In Equation (2.13), $0 \geq \alpha_i \geq C$ and $\sum_i \alpha_i y_i = 0$, should be satisfied. By manipulating the Equation (2.13) based on Karush-Kuhn-Tucker conditions, optimization could be solved by using quadratic optimization programming and some samples having $\alpha_i = 0$ are named as Support Vectors.

Since most of the natural phenomenons can not be separated linearly, before applying the procedure stated above, the data might need to be moved into a higher dimensional space by applying a nonlinear kernel function. To do so, the operations would be done with $\phi(\mathbf{x}_i)$ rather than $\mathbf{x}_i$, where $\phi(\cdot)$ is the mapping function. A variety of kernel functions are available in literature like Radial Basis Function (RBF), Polynomial and Hyperbolic Tangent Functions. In RBF, which is used in this study, the Kernel function for two samples $\mathbf{x}$ and $\mathbf{x}'$ can be calculated by Equation (2.14):

$$K(\mathbf{x},\mathbf{x}') = \exp\left(-\frac{||\mathbf{x}-\mathbf{x}'||^2}{2\sigma^2}\right) \tag{2.14}$$

By using the transformation matrix $K(\mathbf{x},\mathbf{x}')$ in Equation (2.13), the same dual problem can be obtained as shown in Equation (2.15).

$$L_d = -\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j K(\mathbf{x},\mathbf{x}') + \sum_{i=1}^{N}\alpha_i \tag{2.15}$$

The C parameter in Equation (2.12) and the σ parameter, which is the kernel width, in Equation (2.14) should be tuned using cross validation strategy to be able to build an efficient separating hyperplane in terms of classification performance.

SVM is designed as a binary classification method. Therefore, it is worth to note that different classification strategies should be used for the classification of multi class data. The procedure explained above should be done for many times with binary conditions and the majority of the results of the binary classifications is the ultimate value for classification. The common strategies used in the literature are one-against-one and one-against-all approaches. In one-against-one classification, classification is done for all possible binary class cases. However, in one-against-all classification, same procedure is done for each of the classes with an artificial class that covers the rest. After the labelling with all cases, actual class label is found is found with the majority vote.

2.2.1.2 K-nearest neighbours

The second classification algorithm to be used is kNN. The general name for this classification method is nearest neighbour classifier. k-nearest neighbour classifier assigns the most frequent label of the k nearest labelled samples with respect to the proximity of the samples which is defined by distance functions [29, 30]. Those distance functions could be depended on the numerical or categorical attributes of the data, depending on the dataset. Some of the most common distance measures are Euclidean, Hamming, Manhattan or Chebychev measures.

One thing to be optimized in kNN is the number of neighbours, k. In order to find the optimum k, classification is done with kNN classifier having different number of k's based on the validation dataset. The k value providing the highest classification accuracy is selected and the test data, whose labels are not known, are labelled. In terms of computational complexity, kNN is much more simple than SVM. However, this does not mean that kNN is not an accurate classifier. Thus, it should not be forgotten that performance of the classifiers always depend on the data to be classified. A visual example could be shown as Fig. 2.5 for a 2 featured dataset. The sample with the **?** is classified with different number of neighbours as $k = 1, 2$ and 3, respectively.

2.2.1.3 Compact decision trees

14
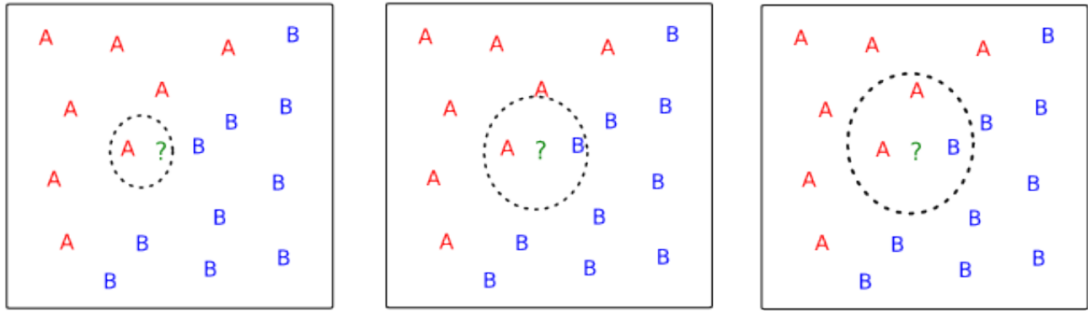
Figure 2.5 : kNN with $k = 1, 2$ and 3, respectively.

The last algorithm used in classification is compact decision tree (CDT). The tree is assisted according to the loss functions of related branches, then decision was made accordingly [31–33]. Depending on the probabilities and distribution of the data, prunes are provided. As shown in Fig. 2.6, algorithm defines prunes step by step depending on the distinctive features with their distinctive values.
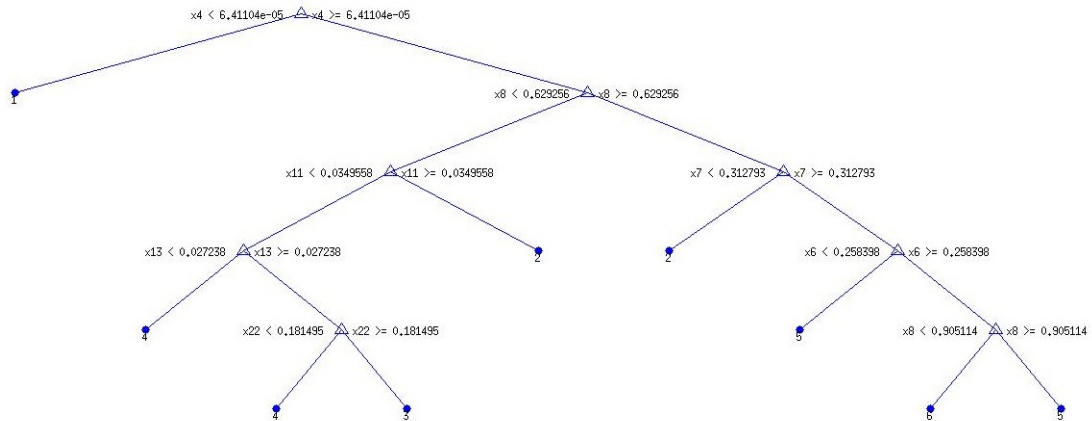
Figure 2.6 : Classification tree with threshold values of the distinctive features.

2.2.2 Feature selection algorithms

It is a very important thing to consider that number of features is not necessarily related with the classification performance. Depending on the properties of the dataset, sometimes it is enough to get the same classification performance with less number of features. Moreover, it is even possible to have less performance with the excessive number of features; which is a phenomena called 'Curse of Dimensionality'.

Dimensionality reduction (DR) is the general name of the operations in order to overcome this problem. DR operations are divided into two categories as feature selection (FS) and feature extraction. In FS, a new subset of the existing features are selected with respect to their contribution to classification accuracy from the most

distinctive features of the dataset to make a better classification. The most important property of FS is that original features are not modified but kept the same. In feature extraction algorithms, new features are developed from the original features in order to have more distinctive ones. Principal component analysis, linear discriminant analysis or canonical correlation analysis are one of the most important examples of feature extraction analysis. However, it should be underlined that data is manipulated with feature extraction methods, which can cause some loss of physical reality in the dataset.

In this study, since the features are derived from real physical interactions; feature selection methods are used in order not to lose the information of physical reality in the dataset. Feature selection algorithms are divided into three types depending on their method as wrapper, filter and embedded methods. One feature selection algorithm was used from each of those methods in this study, which are explained in the coming subsections.

2.2.2.1 Kruskal-wallis

Kruskal-Wallis (KW) is a filtering method developed by Kruskal & Wallis in 1952 to build new feature sets from the existing ones [34]. Filter models evaluate the importance of each feature considering their distinctive characters. As a result, they filter the redundant features as shown in Fig. 2.7. Depending on their simplicity, convergence in terms of classification accuracy of those methods can be slower than the other methods although overfitting problems do not occur frequently. Filter models can be divided into two types as univariate and multivariate models. Univariate models evaluate each feature individually while multivariate models evaluate features with subsets having one or more features. As a result, univariate methods are much faster than multivariate methods but limited in defining redundant features. Although it is not very convenient to make generalization about feature selection methods, it can be said that filtering methods have the most simple computational background but the least efficient ones in terms of selecting the most important features in first orders. As a univariate filtering method, KW ranks the features using medians of each classes and gives a feature importance order accordingly.
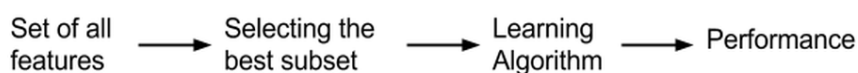
Figure 2.7 : Diagram of a general filtering algorithm.

16

2.2.2.2 Support vector machines recursive feature elimination

Support Vector Machines recursive feature elimination (SVM-RFE) is a wrapper method depended on the **w** vector of the hyperplane in SVM [35]. Wrapper methods select the best subset of features in two consecutive steps as generating subset of features by clustering and measuring their usefulness in a learning algorithm as shown in Fig 2.8. Since this iterative approach does not stop until a good quality is reached, wrapper methods have much more computational cost than filtering methods. However, they are generally more successful than filtering methods but overfitting problems may occur during the process. In SVM-RFE, the best feature subset is found depending on the separating hyperplane **w**, found in SVM. In other types of wrapper methods, maximum likelihood criteria, k-means or similar statistical tests can be used in order to produce a subset of feature.
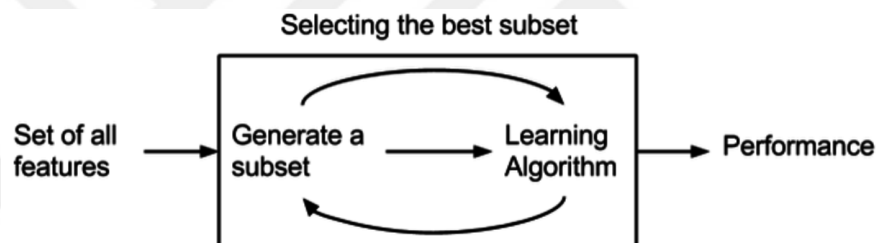
Figure 2.8 : Diagram of a general wrapper algorithm.

2.2.2.3 Sparse bayesian multinomial logistic regression

Sparse Bayesian multinomial logistic regression (SBMLR) is an embedded feature selection method depending on Bayesian regularization with a Laplace prior [36]. Embedded methods build subset of features by clustering methods like wrappers but they do not make a cross validation in order to measure the usefulness of the feature subsets. The search of subset features are guided by the learning process, as seen in Fig. 2.9. Therefore computationally, they are less expensive than wrapper methods and have less tendency for overfitting. SBMLR is an advanced version of Sparse Logistic Regression. Basically, it eliminates some terms of the logistic regression by using Bayesian regularisation using a Laplace prior. Besides the advantages of embedded methods, SBMLR stops iteration after classification accuracy is converged to its maximum value; which gives additional computational efficiency. In other

words, SBMLR stops listing the features after finding the most important features and converge to the maximum accuracy in classification.

Selecting the best subset
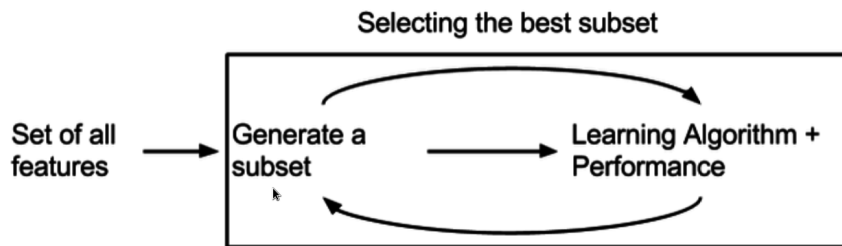
Figure 2.9 : Diagram of a general embedded algorithm.

2.2.3 Segmentation algorithms

Since fields are contained of big numbers of pixels and classification is done in field scale, treatment of the pixels becomes an important point. Signals can vary dramatically, which totally effects the features to be used, due to presence of some man-made object during the acquisition of the images or inhomogeneities in the fields due to agricultural problems or environmental factors. Another problem of the dataset was the lack of samples. Also using all the pixels inside the fields in order to have a representative value for the features is not consistent with the definition of BBCH scale, which is explained in Section 3.1. Therefore three different cases were implemented using k-means clustering in the experiments for the usage of the values in the pixels to calculate samples. The details of the application are given in Section 3.4 while the theory is explained in the coming paragraphs.

k-means clustering, which is also known as Lloyd's algorithm, is basically an iteration in order to divide the data into k clusters [37]. Iteration starts with selecting random samples as centroids of each cluster. Then, for each centroid, the nearest individual sample is found as another member of the cluster and the centroid of each cluster is updated accordingly. In Fig. 2.10, which is taken from [2], a randomly generated dataset and it's clustered version is shown with the cluster centroids.

Figure 2.10 : An example of clustering a random dataset with two clusters [2].

3. EXPERIMENTS AND RESULTS

This chapter is focused on giving information about the experiments conducted with the algorithms explained theoretically in Chapter 2 . Firstly, the study sites, details of the dataset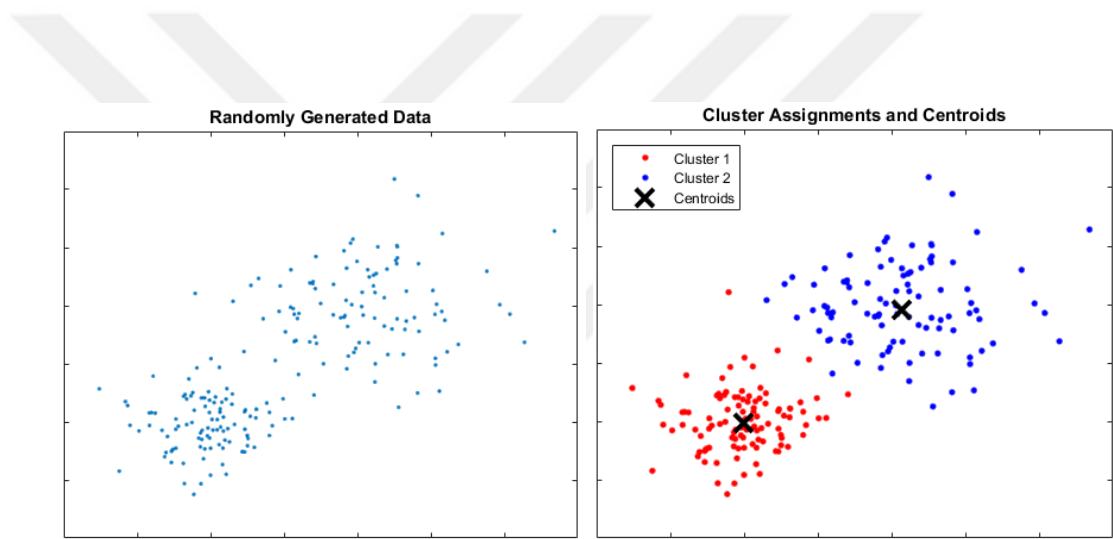s and experimental setup is defined in Section 3.1. Afterwards, each of the specially inspected topic is explained and results are provided. Effects of labelling is mentioned in Section 3.2, effects of features to be used is analysed deeply in Section 3.3, effects of segmentation is explained in Section 3.4 and lastly, large scale implementations depending on the results of the previous experiments is provided in Section 3.5 to show the applicability of the method in big data and different study sites.

3.1 Study Sites, Data and Experimental Setup

In the experiments, two completely independent dataset were used. Those datasets contain multitemporal X-band SAR images acquired by TerraSAR-X on HH and VV channels whose operating frequency is 9.6 GHz. Temporal resolution of the images, which is quite important to be able to detect the changes of the crop, is 11 days. This temporal resolution is suitable enough to detect the phenological changes of rice considering its growth speed. Parallel to the acquisitions, ground measurements for phenological information of some spatially independent fields are also valid.

The scale of the phenology information is an important subject to be mentioned. Phenology information is provided with a widely used growth information scale called *Biologische Bundesanstalt, bundessortenamt und CHemische industrie* (BBCH), which is defined by agronomist in order to define a global scale for mono- and dictyledonous plant species like cereal, rice, maize, sunflower, bean, potato, coffee, cotton and so on [38]. This is a discrete scale from 0 to 99, divided into 10 principal growth stages and each of those growth stages are again divided into 10 parts as secondary growth stages. According to the definition of BBCH scale, the appended BBCH value should be valid for at least half of the field in order to overcome with the inhomogeneity of the fields. The features calculated for the parcels using the acquired

images were labelled according to their BBCH values as explained in Section 3.2 and dataset is produced to be used in machine learning applications.

The independence of the datasets is very important since the differences between different farms are quite big to effect any algorithms. The reason of the differences of the information to be extracted from satellite images can be because of the agricultural practices, climate and soil conditions. In terms of agricultural practices, cultivation method including texture of sowing, sowing date and water level, type of rice species sowed and presence of other plants are some of the important factors that effects the study site. In terms of climate conditions, precipitation rates, temperature and wind regime of the region, irrigation and flooding cases are important factors that change the data. In terms of soil conditions, fertilization and mineral content of the soil can make differences between different agricultural sites.

The first study site is in Spain, located around Guadalquivir river and Isla Major which is south west of Seville, Spain. The cultivation period starts on May with the random sowing of rice on flooded fields and generally lasts until October. The dataset is the same one that is used in [12] and [16], which contains 12 co-polar TerraSAR-X images with an incidence angel of 30°. The study are is shown in Fig. 3.1 with its location in Spain and a closer look from Google Earth. Those 12 images cover all the cultivation period of 2009, which makes it possible to make an extensive analysis for the whole cultivation process. Those 12 images are supported by ground information in BBCH scale for 12 spatially independent fields. To sum up, 144 samples are suitable to be used with their phenology information in the experiments for the first dataset.
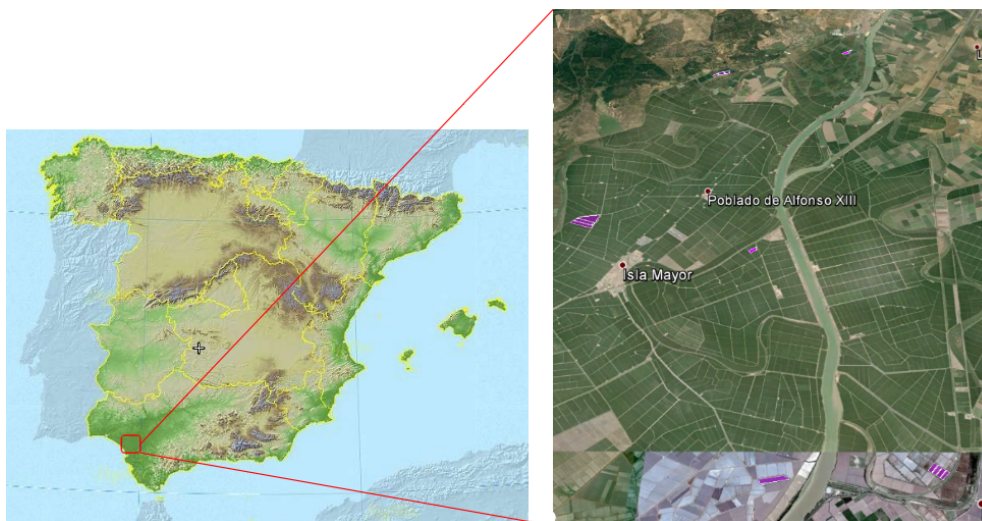
Figure 3.1 : Study area of Seville, Spain.

The second study site is in Turkey, located around Meriç (Evros or Meritsa) river and Ipsala, which is just on the eastern side of the border between Turkey and Greece. The study area is shown in Fig. 3.2 using the map provided in [3]. The cultivation period starts on May with again sowing randomly and lasts until September. The same dataset, which covers some of the cultivation period of 2014, is partially used in [39] as well as some other studies. Unfortunately, the images for Ipsala dataset do not cover all the cultivation period since there is only 6 images with ground information for 8 spatially independent fields. The dataset covers the first two stages of the main three stages in cultivation; vegetation and reproductive stages. To make it clear, 48 samples are available to be used in the statistical tests from the Ipsala dataset covering two thirds of the cultivation period.
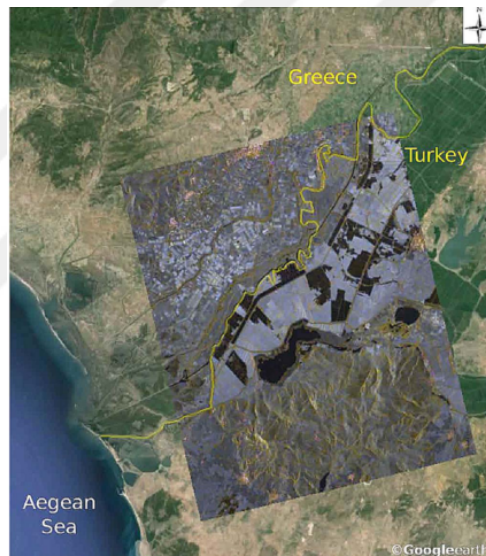
Figure 3.2 : Study area of Ipsala, Turkey [3].

The number of samples makes it quite challenging to make an extensive machine learning analysis. Besides the lack of samples, another thing to make the problem more challenging is the inhomogeneous distribution of the samples. In Fig. 3.3, the distribution of samples in accordance to their BBCH values are shown for both Sevilla and Ipsala dataset. Each sample is shown with red lines with their corresponding BBCH value while BBCH scale from 0 to 99 is shown in the lowest row. In Fig. 3.3(a), the first row corresponds to the most general and widely accepted three vegetative stage of the crop which is defined by agronomists [40], while the row above BBCH scale shows the principal growth stages defined by BBCH [38]. The rows between the 'Three Main Stages' and 'Principal Growth Stages' of Fig. 3.3(a) are going to be explained

23

in Section 3.2. the distribution of data is obviously inhomogeneous considering the principal growth stages of BBCH scale. In Ipsala dataset, not only the lack of data but also the inhomogeneity becomes more dramatic as can be seen in Fig. 3.3(b). Therefore the extensive analysis is done in Sevilla dataset and Ipsala dataset was used as a control for the robustness of the selected framework.
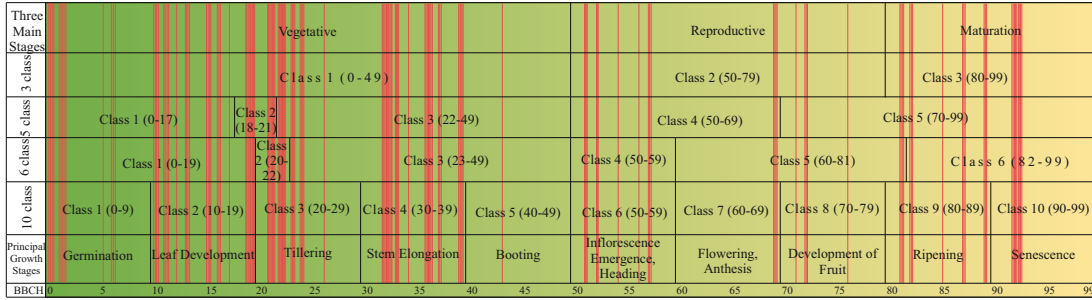
(a) Seville dataset
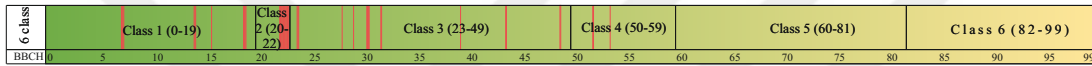
(b) Ipsala dataset

Figure 3.3 : Distribution of the samples having ground information with time [4].

In this study, since the phenology estimation of rice was done with supervised classification techniques on multitemporal images, ground information of the fields should be used as class labels for the samples, where the values coming from electromagnetic scattering are used as features of the samples. Polarimetric features are calculated in IDL while textural features were calculated in ENVI. After calculating each feature explained in Section 2.1.2 for the fields, each of them is labelled according to their phenological information taken from in-situ surveys. After the labelling, datasets become suitable to be used in supervised classification methods and inspected in detail. All those machine learning analysis were done in MatLab environment. Also, LibSVM package is used for SVM [41] and algorithms of KW and SBMLR are built on the codes of [42]. For the statistical tests, each dataset is divided into two equal parts randomly to be used as a training and testing dataset. In order to have robust results from those random generations, each dataset is divided into 100 times and all the experiments were done for 100 times with those random generations. Therefore all the results in the coming sections are given with their standard deviations. Each of the inspected variable is explained in the coming sections separately.

In this study, the optimization of the penalty parameter, C, of SVM for linear and nonlinear cases and the kernel width, σ, of RBF kernel function are optimized by grid search from 2^{-8} to 2^8 in logarithmic scale. Also it should be noted that one-against-one approach is used to adapt the SVM algorithms for multi-class classification in the experiments. In order to calculate the distances between samples for kNN, Euclidean distance measure is used for kNN in the experiments. Lastly, for CDT, 13 to 17 child nodes were developed in the trees with different realizations by the algorithm in order to make the decision.

3.2 Effect of Labelling

As explained in the previous section, phenological evolution, which is a continuous variable of crops is defined in a discrete scale called BBCH. With this scale, phenological evolution of the fields in a multitemporal dataset is defined numerically. The decision of the number of classes and their intervals have a vital importance for the performance of classifiers since the classifiers' performances are dependent on the separability of features between classes. The multitemporal evolution of some of the features can be found in [12]. The division of the class boundaries should be done by considering both the features' separability and the important stages for farming applications.

Four different class number case is selected to be inspected. As the most rough one, 3-class case was used with the BBCH intervals defined in [40]. To make the study more precise, another case with 5 classes is also used. The boundaries of the classes are defined in [12] which also uses the same dataset for Sevilla. To increase the precision more, another case having 6 classes was defined by dividing the reproductive and maturative stages into three classes as defined in [21]. Lastly another case having 10 classes was defined depending the principal growth stages of BBCH [38], which could be one of the best cases for the farming applications. All the intervals of the BBCH values can be seen in Fig. 3.3(a) with respect to classes.

The classification performance of the classifiers are given in Table 3.1. While the overall classification accuracies (Acc.) are given in the upper rows, the kappa values (κ) are given in the lower rows of each case. For each classification case, the best classifier is shown with bold case. Nonlinear SVM and kNN achieved to get a

25

Table 3.1 : Classification performances of the classifiers on different labelling cases.

		Nonlinear SVM	Linear SVM	kNN	CDT
3-class	Acc.	**89.9 ± 3.8**	85.1 ± 3.2	88.2 ± 3.8	82.3 ± 5.6
	κ	**78.7 ± 7.7**	67.6 ± 6.4	75.5 ± 7.3	61.9 ± 11.6
5-class	Acc.	**84.0 ± 4.0**	80.8 ± 3.9	83.4 ± 3.8	79.3 ± 4.9
	κ	**79.1 ± 5.3**	74.5 ± 5.2	78.4 ± 4.9	73.0 ± 6.2
6-class	Acc.	81.1 ± 4.4	77.5 ± 4.4	**81.2 ± 4.1**	73.2 ± 5.0
	κ	75.9 ± 5.5	71.0 ± 5.7	**76.2 ± 5.1**	65.9 ± 6.1
10-class	Acc.	**77.3 ± 4.3**	74.4 ± 3.7	75.4 ± 4.6	67.0 ± 5.6
	κ	**73.3 ± 5.1**	69.5 ± 4.4	71.1 ± 5.4	61.2 ± 6.4

classification accuracy higher than 80% except the 10-class case. Linear SVM and CDT are also compatible with the other two classification methods. The reason of the slightly lower performance of Linear SVM is its disability to adapt nonlinear problems well. In CDT, the reason of the lower performance could be the high number of branches of the decision. However, it has to be underlined that performance of the classifiers strongly depends on the data, which is against generalization about the performance of classification algorithms.

Regardless of the classifier, as the number of classes increase the classification performance decrease as expected. One of the important reasons of this point is the similarity in the SAR signatures in different classes. However, the results are quite encouraging considering that in all cases a substantial agreement is achieved, which requires a kappa value greater than 60%. The biggest reason of the lower accuracy in 10-class case is the limited inhomogeneously distributed dataset. However, even in 10-class case classification accuracy could be obtained higher than 75%. From the four different class labelling cases, 6-class case gives the most successful performance considering the classification performance of the classifiers and the operational capacity of agricultural applications with the information produced by this case. Therefore, the study is extended to further ways by using the 6-class labelling case.

Classification performance of the classifiers for 6-class case can be seen in detail in Table 3.2. In Table 3.2, first class wise accuracies, then overall accuracies and lastly kappa agreements are given for each classification algorithm in (%). As done in Table 3.1, the best value obtained for each class is shown with bold text in the Table 3.2. According to the Table 3.2, the easiest class to be determined is the first class, which is

Table 3.2 : Class wise performances of different classification algorithms for 6-class case.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Overall	κ
Nonl. SVM	**92.2 ± 7.3**	75.0 ± 17.5	87.9 ± 8.1	77.2 ± 22.7	48.1 ± 22.3	68.8 ± 17.4	81.1 ± 4.4	75.9 ± 5
Lin. SVM	87.1 ± 8.6	72.3 ± 19.1	**92.9 ± 6.7**	42.3 ± 19.8	11.4 ± 14.9	**91.6 ± 9.8**	77.5 ± 4.4	71.0 ± 5
kNN	91.1 ± 7.1	**78.4 ± 17.2**	87.6 ± 7.1	**80.6 ± 15.9**	**59.6 ± 19.5**	60.4 ± 14.3	**81.2 ± 4.1**	**76.2 ± 5**
CDT	91.4 ± 5.9	53.8 ± 20.9	80.9 ± 12.1	48.8 ± 24.4	51.8 ± 26.2	59.2 ± 17.8	73.2 ± 5.0	65.9 ± 6

expected. Since the fields are under water in the first class due to agricultural practices, SAR signatures are quite distinguishable than the other ones which are not completely flooded. Following the first class, third class is also an easy class to be determined since crops have a vertically dominated structure. Thanks to that vertical structure, the two channels of the sensor changes in unique ways. While backscattering values of the HH channel are higher than the other stages because of the double bounce effect, signals from VV channel gets effected from the attenuation of the signal in the vertical structure. The other classes are not that easy to detect. SAR signatures of the other classes are quite similar since the crop is in a transition stage [12]. Besides, the variance of the features are quite higher in the last stages compared to the first ones [10].

3.3 Effect of Features

After the number of classes to be estimated is defined; the effects of features and their contribution on classification performance is analysed. It is an important point to be analysed about the contribution of the features to classification. Besides, since all the 25 features were calculated from only 2 channels, it is very likely to have redundant features that might reduce the performance of classifiers. In order to analyse the effects of the features, two different analysis is done. Firstly, classification accuracies using different feature sets were compared. Then some special feature selection algorithms were used in order to see the importance of each feature individually.

In order to compare effects of different feature sets, features were grouped into three parts as Polarimetric, HH texture and VV texture. With that feature set, all possible combinations of feature sets were used in classification algorithms. Classification with only textural features did not give any successful results, which is expected since agricultural targets are so random that they do not have any textural formation. To sum up, four different feature set is used as polarimetric, polarimetric & HH texture,

27

polarimetric & VV and polarimetric with HH & VV texture features. The number of features in those sets are 9, 17, 17 and 25, respectively.

The classification performance of the classifiers using those feature sets are shown in Table 3.3 with overall accuracies (Acc.) and Kappa agreements (κ) and the best feature set for each classifier is written in bold. As can be seen from the table, usage of textural features, especially in VV channel, increases the classification accuracy up to 5% although fields do not have a textural formation. The reason of this contribution is the temporal trend of the textural features even though they do not give any information in a single time, t. However, neither which of the textural features contribute to the classification nor the importance order of features could not be clarified with this analysis. Therefore, feature selection algorithms are used in order to check each of the features individually.

Table 3.3 : Classification accuracies for different feature sets.

		Nonl. SVM	Lin. SVM	kNN	RT
Polarimetric	Acc.	76.8 ± 4.8	73.4 ± 4.2	78.4 ± 4.4	72.7 ± 4.5
	κ	70.5 ± 6.0	65.9 ± 5.2	72.6 ± 5.5	65.3 ± 5.6
Polarimetric + HH texture	Acc.	79.1 ± 4.8	74.9 ± 4.2	79.6 ± 4.3	71.8 ± 4.6
	κ	73.3 ± 6.1	67.7 ± 5.3	74.2 ± 5.4	64.1 ± 5.6
Polarimetric + VV texture	Acc.	$\mathbf{81.3 \pm 4.7}$	76.9 ± 4.3	$\mathbf{81.5 \pm 4.4}$	$\mathbf{73.2 \pm 4.9}$
	κ	$\mathbf{76.2 \pm 6.0}$	70.3 ± 5.5	$\mathbf{76.6 \pm 5.4}$	65.8 ± 6.0
Polarimetric + HH & VV texture	Acc.	81.1 ± 4.4	$\mathbf{77.5 \pm 4.4}$	81.2 ± 4.1	73.2 ± 5.0
	κ	75.9 ± 5.5	$\mathbf{71.0 \pm 5.7}$	76.2 ± 5.1	$\mathbf{65.9 \pm 6.1}$

Three feature selection algorithm were used as feature selection algorithms as KW, SVM-RFE and SBMLR. Each of them are selected from different modelling approaches as explained in Section 2.2.2. By using those feature selection algorithms, feature importance rankings and classification accuracies with those features are obtained.

Overall classification accuracies obtained with kNN for 6-class case are shown in Fig. 3.4 with the increasing number of features. Here it should be underlined that SBMLR stops iteration after the classification performance is converged to its maximum value, after 7 or 8 features depending on the dataset realization. The other classifiers did not needed to be placed since they all have the similar trend. The feature to be used is selected by each of the feature selection algorithm separately. As seen from

Fig. 3.4, a satisfactory classification can be done by using 4 or 5 features that are selected by wrapper or embedded methods. KW converges slower than other methods as expected, since it is a general drawback of the filtering feature selection algorithms. Classification performance could be even increased slightly by discarding the redundant features. Another benefit of using only the necessary features, is the decrease in computational cost. The saving for computation could be higher than expected since producing of the features are also quite costly.
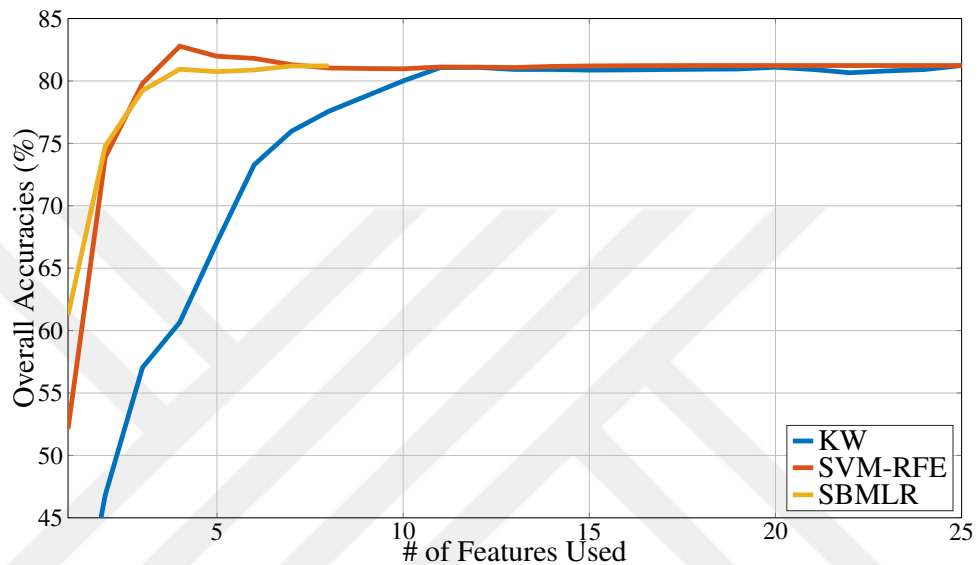
Figure 3.4 : Overall classification accuracies for kNN with increasing number of features using different FS algorithms [4].

The rankings of the features obtained by each feature selection algorithm are shown in Fig. 3.5. In Fig. 3.5, feature ranking for each random generalization is shown with colours in rows of the figures while the mode of each rank is shown in the x-axis of the figures. Therefore, it is expected to see the same colour vertically in case of total robustness. Another thing to be mentioned for the figure is in the cases of quick convergence for SBMLR, the unnecessary pixels left blanc in the figure. As seen from the noise in the figures, feature selection algorithms are very sensitive to data. Although the results are obtained from sub-datasets generated from the same data, the results vary in different generations. However it is still possible to observe a general trend in the experiments.

Blueish tones represent polarimetric features, greenish ones represent HH texture and lastly yellowish ones represent VV textures. The overall ranking of the features can be seen by colours in a rough way while specific ranking can be seen in the x-axis of

the figures. As easily seen from the tones, polarimetric features are the most important ones, followed by VV texture features. For the overall rankings shown in the x-axis of the figures, the values vary within the algorithms since all of them have different modelling approaches. However still the first ones are the polarimetric features. For the most important textural features, it can be said that mean and the entropy values of the Haralick features. Another important point about Fig. 3.5 is that, in KW, entropy of VV texture is selected to be the most important feature. This is because of the modelling dynamics of KW and the consequences of this ranking can be seen in the poorer classification performance obtained by the order of KW in Fig. 3.4.

1) HH Backscat. Coefficient
2) VV Backscat. Coefficient
3) HH/VV Ratio
4) Backscat. coef. 1st Pauli Channel
5) Backscat. coef. 2nd Pauli Channel
6) Coherence
7) Phase Difference
8) Entropy
9) Alpha Angle
10) HH Texture - Mean
11) HH Texture - Variance
12) HH Texture - Homogenity
13) HH Texture - Contrast
14) HH Texture - Dissimilarity
15) HH Texture - Entropy
16) HH Texture - Second Moment
17) HH Texture - Correlation
18) VV Texture - Mean
19) VV Texture - Variance
20) VV Texture - Homogenity
21) VV Texture - Contrast
22) VV Texture - Dissimilarity
23) VV Texture - Entropy
24) VV Texture - Second Moment
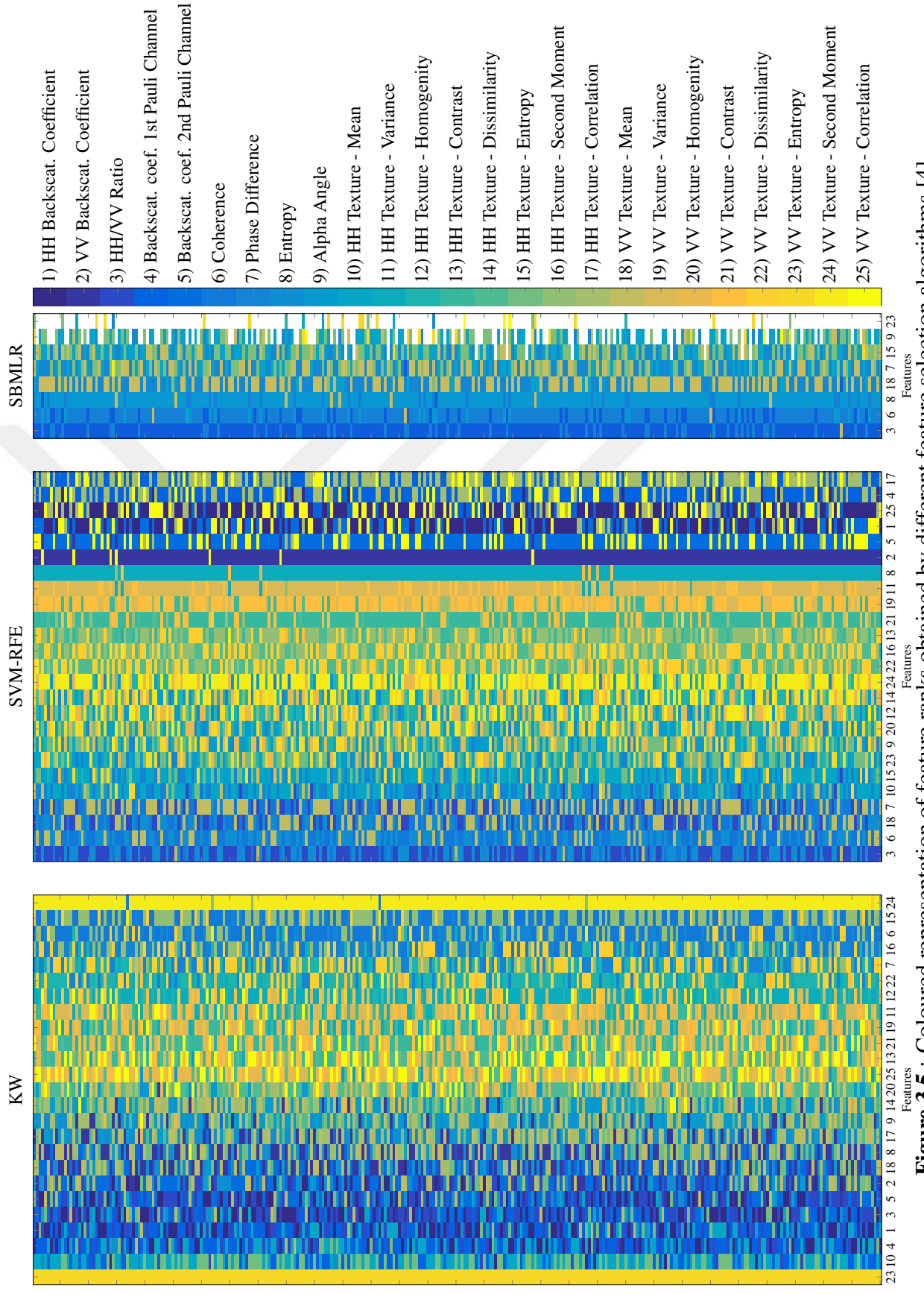25) VV Texture - Correlation

Figure 3.5 : Coloured representation of feature ranks obtained by different feature selection algorithms [4].

3.4 Effect of Segmentation

Since the classification is done in field-scale but it is possible to calculate the features in pixel-scale, the calculation of features for fields becomes another important point to be analysed. In Seville, fields cover 6000 to 25000 pixels and in Ipsala, fields cover 1500 to 10000 pixels; which are big enough to effect the calculations. Moreover, it is a well-known fact that generally fields do not grow homogeneously due to varieties of problems [5]. In Fig. 3.6, which is taken from [5], a field in the same study area having a cultivation problem can be seen. As a result, removal of the outlier pixels have a great importance in the calculation of the representative features in field-scale.

To analyse this issue, three cases were defined in order to calculate representative feature values for fields. Those cases were dependent on segmentation with are done with k-means clustering, which is explained in Section 2.2.3. The clustering is done with respect to backscattering values of HH and VV channels, their ratio and phase difference between channels.

Figure 3.6 : An inhomogeneous field from the study area having a cultivation
problem [5].

In the first case, the easiest and the most classic approach was implemented as taking the median values of all pixels within the fields, which is named as 'No Cluster' in the study. The previous experiments were also done with this approach. In the second case, in order to have more pure samples, the definition of BBCH scale is applied to

feature calculation in field scale [38]. To do so, a cluster that has more than half of the pixels within the field with the maximum number of clusters should be found. An iteration starting from 6 clusters with decreasing number of clusters is programmed and checked the dimension of the biggest cluster. The first biggest cluster that has the majority of the pixels within the field is selected as representative pixels. Then, the median values of the features of those pixels are used as a representative values for field-level data. This case is named as 'One Cluster'.

In the third approach, the aim of making a segmentation was not only have pure samples but also increase the number of them. In order to do that, fields were divided into 10 clusters and the biggest 6 of them were used as six different samples within the field. After calculating median values of the features for each cluster, samples were labelled according to the BBCH value of the whole field by using their median values. By doing so, both outlier pixels of the fields were separated and number of samples to be used in classification algorithms were increased although ground information was the same. This case is named as 'Six Cluster'.

The performance of the classifiers are shown in Table 3.4 with overall classification accuracies (Acc.) and kappa values (κ). The highest performance obtained for each classifier was shown in bold. Although the definition of BBCH was applied in the One Cluster case, classification performance was not better than the other methods. The most likely reason for this problem is the loss of statistical reliability of SAR signals considering the wide distribution of the values of the signals within pixels. However, in Six Cluster case, the classification accuracies are increased up to 5%. Although performance of kNN was still higher in No Cluster case, all other classifiers have higher accuracies than the case of kNN in Six Cluster case. The reason for this increase in performance is the increased number of samples. The need of more samples can also be seen with the standard deviations of the results. As number of samples increased for 6 times, standard deviation of the performances of all classifiers decrease significantly.

3.5 Large Scale Implementations

In order to check the applicability of the method in different study areas, the best framework obtained in Sevilla is implemented into a completely different dataset

Table 3.4 : Classification accuracies for different segmentation cases.

		Nonlinear SVM	Linear SVM	kNN	CDT
No Cluster	Acc.	81.1 ± 4.4	77.5 ± 4.4	$\mathbf{81.2 \pm 4.1}$	73.2 ± 5.0
	κ	75.9 ± 5.5	71.0 ± 5.7	$\mathbf{76.2 \pm 5.1}$	65.9 ± 6.1
One Cluster	Acc.	78.1 ± 4.3	75.3 ± 3.8	77.8 ± 3.8	72.7 ± 5.1
	κ	72.0 ± 5.5	68.1 ± 4.7	71.7 ± 4.7	65.4 ± 6.2
Six Cluster	Acc.	$\mathbf{86.9 \pm 1.5}$	$\mathbf{81.0 \pm 1.4}$	78.2 ± 1.3	$\mathbf{81.4 \pm 2.4}$
	κ	$\mathbf{83.4 \pm 1.8}$	$\mathbf{75.6 \pm 1.8}$	72.1 ± 1.7	$\mathbf{76.3 \pm 3.0}$

which is located in Ipsala. As explained before, the difference between the datasets are coming from various aspects like environmental conditions, agricultural practices and climate. As explained in Section 3.1, Ipsala dataset do not cover all the cultivation period. Therefore only the best method obtained from the Sevilla dataset is applied for the Ipsala dataset, which is using all of the 25 features with Six Cluster segmentation case with a labelling case of 6-classes.

Table 3.5 : Classification performances of different classification algorithms in Ipsala dataset.

	Class 1	Class 2	Class 3	Class 4	Class 5	Class 6	Overall	κ
Nonl. SVM	80.3 ± 10.0	78.6 ± 7.6	75.5 ± 7.2	$\mathbf{45.8 \pm 18.7}$	-	-	$\mathbf{75.9 \pm 3.6}$	$\mathbf{65.4 \pm 5.2}$
Lin. SVM	70.6 ± 11.8	75.9 ± 8.8	$\mathbf{79.0 \pm 6.9}$	16.6 ± 19.7	-	-	71.9 ± 3.4	58.8 ± 5.1
kNN	72.6 ± 9.6	$\mathbf{80.9 \pm 8.7}$	68.2 ± 6.6	22.3 ± 17.2	-	-	70.9 ± 2.5	57.6 ± 3.5
CDT	$\mathbf{85.1 \pm 8.4}$	73.7 ± 9.4	73.1 ± 8.6	45.8 ± 19.6	-	-	74.6 ± 4.7	63.6 ± 6.7

Classification performance of Ipsala dataset with this setup is shown in Table 3.5. Even the accuracies obtained for Ipsala dataset is a bit lower than the Sevilla dataset, the results are quite satisfactory. It is possible to have an overall classification accuracy of more than 75% with a substantial kappa agreement. Since the scattering mechanism is the same, the same explanations done in Section 3.2 is still valid and it is expected to have the same trend in Tables 3.2 and Table 3.5. Although there is not any data available for the fifth and sixth class for Ipsala, the lowest performance is obtained for Class 4; which is the most difficult class for Ipsala dataset. Another reason for this poor result is the uneven distribution of the dataset, as shown in Fig. 3.3(b).

After being sure about the applicability of classification algorithms on phenology retrieval, the proposed framework was applied to the whole study area for both Seville and Ipsala datasets. 250 randomly selected fields within the study area were classified for the whole cultivation period with nonlinear SVM for 6-class case using
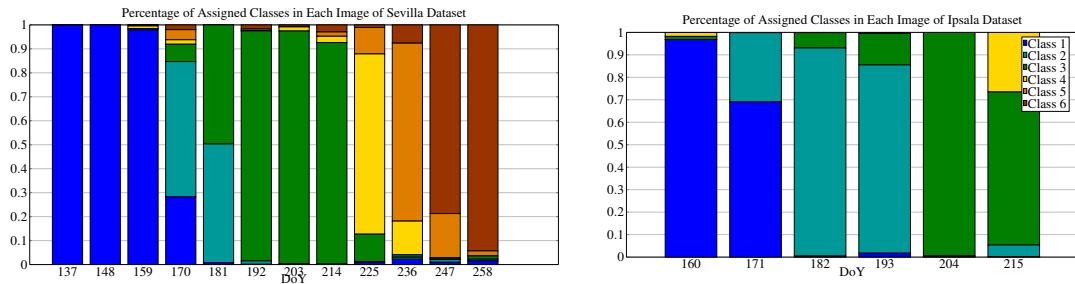
Figure 3.7 : Histogram of appended classes for the thematic applications on both datasets [4].

25 features calculated by Six Cluster segmentation approach. The classification model built using all the labelled data. Since there is not any ground information about those randomly selected fields, statistical performance analysis was impossible for those results. However, it is still possible to have some idea about the accuracy of classification by inspecting the temporal trends of the fields, which should be parallel to the general phenological development of rice.

The results are shown in thematic maps in Fig. 3.8 and Fig. 3.9 for Seville and Ipsala datasets, respectively. Each class is assigned with a specific colour and the outlier pixels chosen by the Six Cluster algorithm are shown in gray in order to show them separately. In order to summarise the information in the thematic maps, histograms of appended classes on each image for both of the datasets are provided in Fig. 3.7. The results are in agreement with the classification accuracies found with the statistical tests. Although there are some obvious misclassification like assigning the forth stage in the early season or assigning the first stages in the late season, especially after seeing third classes; the results are quite encouraging. Since those obvious mistakes could be overcome by applying simple modifications on the classification framework, the framework could be seen satisfactory. In this study, mathematical calculations are presented without any modification since the main idea is to prove the applicability of the framework.
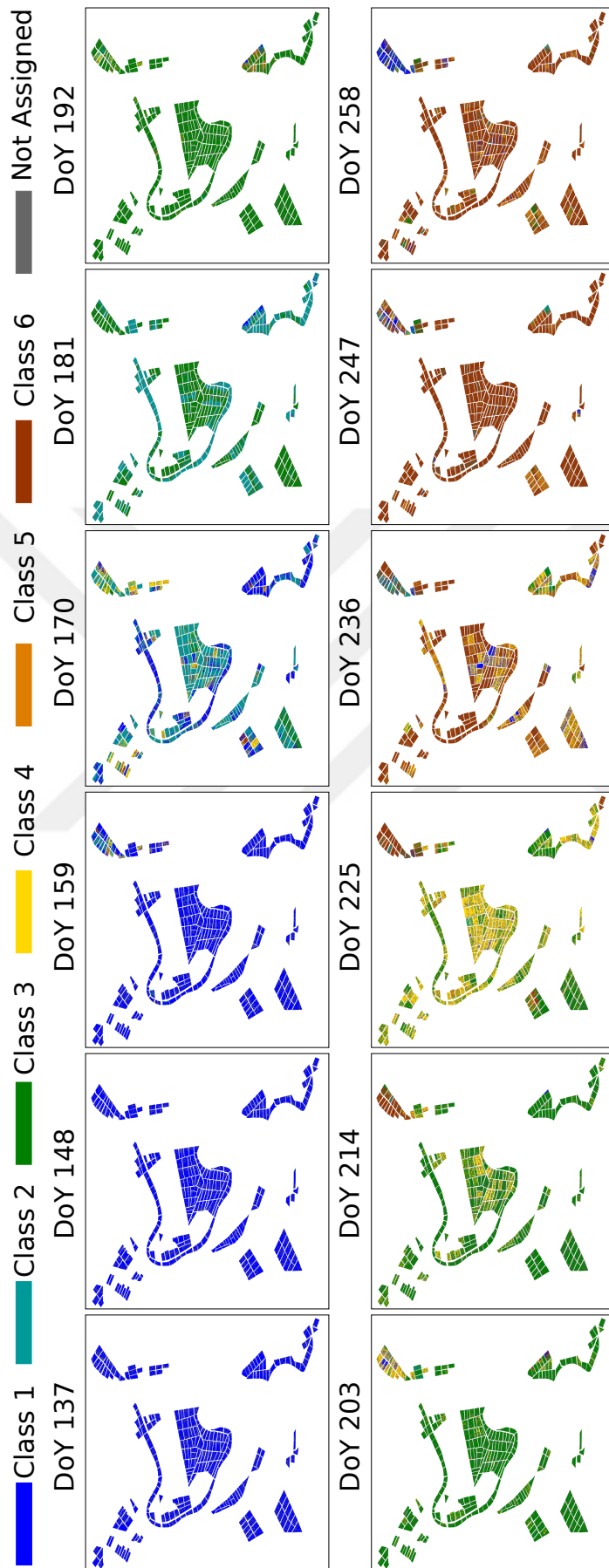
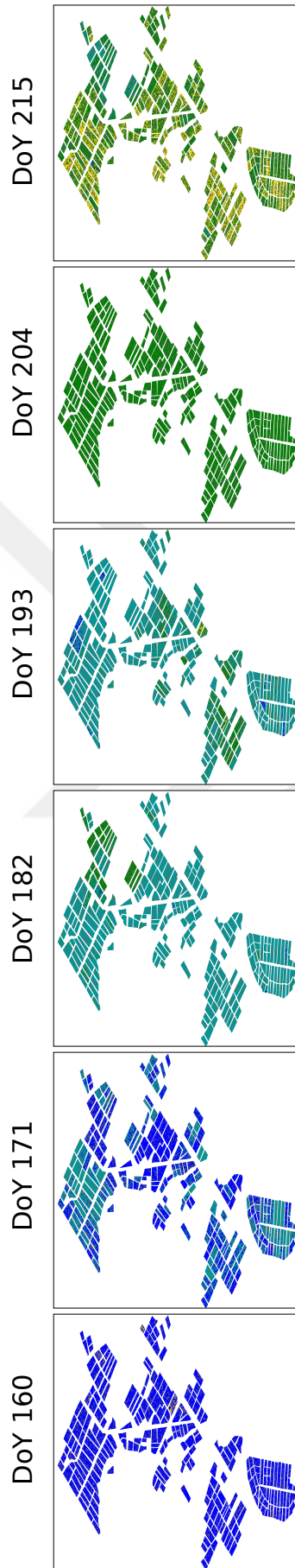**Figure 3.8** : Thematic maps obtained for Seville region [4].

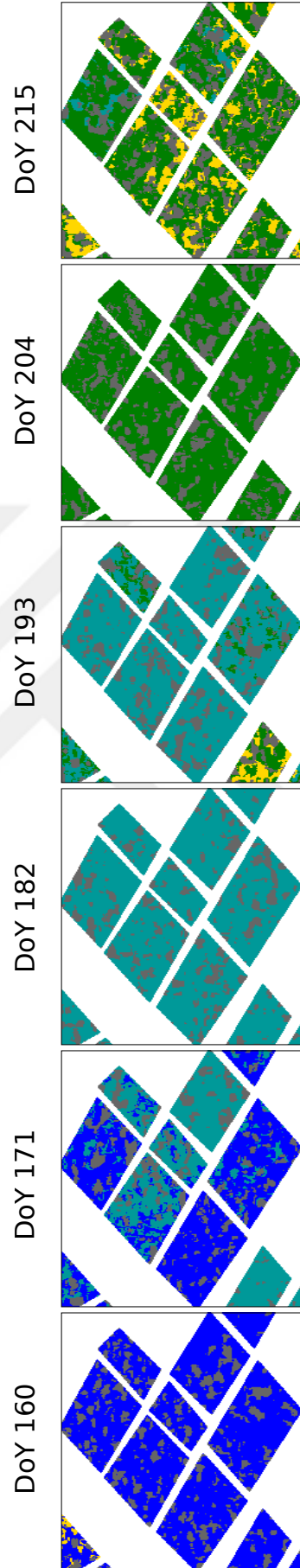**Figure 3.9** : Thematic maps obtained for Ipsala region [4].

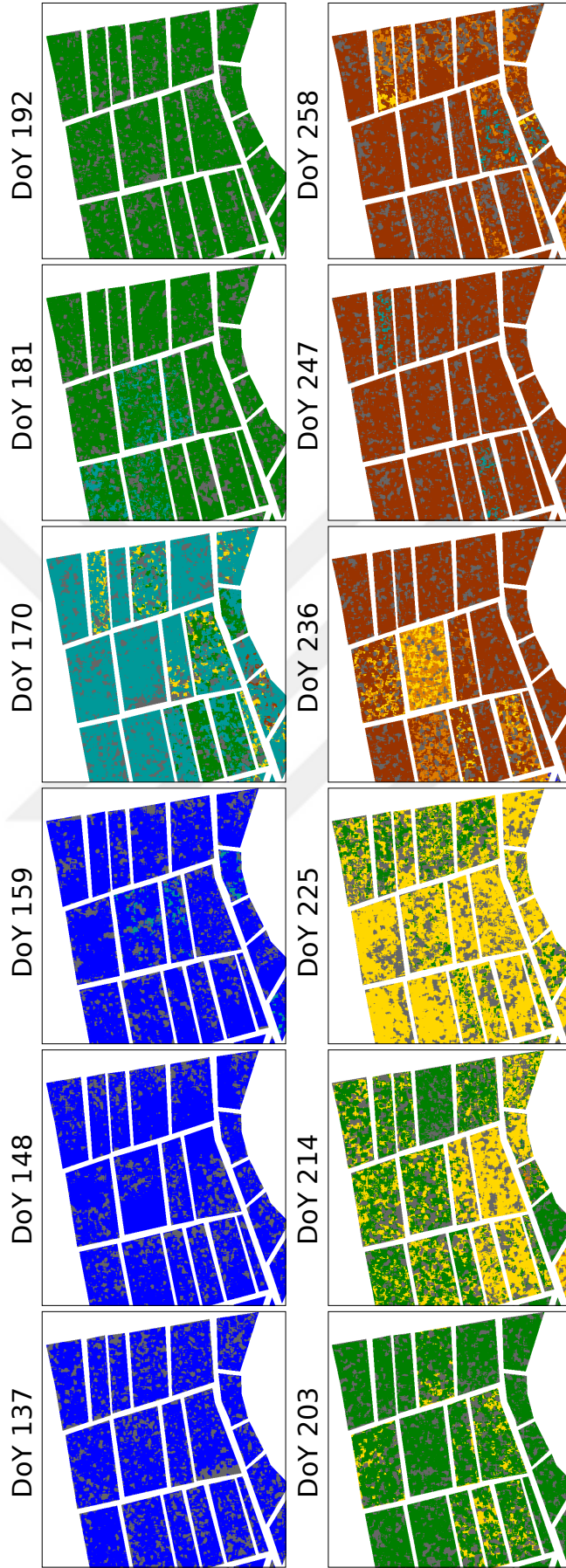**Figure 3.10** : Detailed view of some fields of Ipsala region obtained for the thematic maps [4].

Figure 3.11 : Detailed view of some fields of **Seville** region obtained for the thematic maps [4].

The outcomes of Six Cluster case can be seen in the zoomed version of the thematic maps. In Fig. 3.10 and Fig. 3.11, smaller areas of the study sites are shown, which allow to see the differences within fields. It can be seen from the figures that, some parts of the fields were appended outlier and didn't classified, which are shown in gray. Also it is possible to see the different labelled parts within the fields. However, each of the fields has a dominant class having majority of pixels which makes the same definition with BBCH scale.

4. DISCUSSIONS AND RECOMMENDATIONS FOR FUTURE STUDIES

In this study, machine learning methods were implemented to map the growth stages of the rice crop using the multitemporal dual-pol SAR images. The aim of this paper is to built a convenient classification schema providing a high classification accuracy and more reliable thematic map. Through this process, there are several important points encountered in implementing ML methods, and they need to be discussed in detail.

In the following sections, a brief discussion of the study will be given in Section 4.1 and will be concluded with some feature research recommendations in Section 4.2.

4.1 Discussions

The main aim of this study was to monitor the temporal evolution of rice fields using SAR images in machine learning algorithms. The purpose of the study is to detect the phenology of the crops without dealing with electromagnetic radiation theory, but using statistical methods of machine learning. In order to analyse the parameters and propose a concrete framework with this study, different evaluation methods were inspected separately in one dataset, Seville dataset, and the best approach was conducted on another, totally independent dataset, Ipsala dataset. Since machine learning methods are quite dependent on dataset, a generalization for a concrete framework is quite difficult. However, it is shown that using machine learning in temporal image classification is applicable in monitoring the evolution of rice fields.

The first point to be underlined is class labelling. Since classification algorithms are used in the study, class labels have a vital impact on the algorithm. Besides, those class labels correspond to the phenological stages of the crops which means that the important phenological stages of the crop and needs of cultivators should be considered. In terms of capability of detecting changes, with their wavelength of 3 centimetres, X-band SAR sensors are quite capable to detect the changes of the crop since SAR signals can detect the changes within the range of their wavelengths. In

terms of machine learning, it is also certain that classification algorithms could be successful with sufficient amount of learning data using representative features.

In order to check the performance of classifiers with different number of classes; four different labelling cases were appended, starting from 3-classes to 10-classes. In those cases not only number of classes but also the ranges of the classes were important. Since classification is dependent on the separability of features in each class; intervals of classes are very important. Temporal trends of important features could be seen in [12]. Another thing to be considered depending on the class labelling is distribution of data. Learning algorithms need sufficient amount of training data to understand the pattern of classification. However, as can be seen in Fig. 3.3; even for the Seville dataset there is an obvious lack of data for training and testing the performances of classification algorithms. For the Ipsala dataset, as mentioned before, there is not ground information for whole cultivation period; which restricts the research for further analyses.

When the performances of classifiers are inspected in detail; as shown in Table 3.1; the results are satisfactory considering that more than 80 % of classification accuracy is obtained for 3-, 5- and 6-class cases. Another measure of the performance is Kappa agreements; which is between 60 to 80 % for all cases that indicated moderate agreement. Although it is not very convenient to propose classifiers since they are strongly dependent on the data to be classified; it can be said that classifiers that can work on nonlinear problems perform better than the linear ones. In this study, classification with 6-classes is chosen considering the maximum number of classes with a classification accuracy higher than 80 % and research was extended with this class labelling case. As shown in Table 3.2, class 2, 5 and 6 are more difficult to distinguish due to the features nonseparable nature and limited data.

The second point to be underlined in this study is the effects of features. Even though vegetation is regarded as random media in SAR systems and textural formation is not expected; using the temporal trends of the textural features with the other polarimetric features can contribute the classification accuracy. As shown in Table 3.3, usage of textural features contribute to the classification accuracy. This is an important information to be used in the future studies of multitemporal image classification for agricultural applications. Although textural information is not sufficient to

make a successful classification, usage of textural information with polarimetric information helps to increase classification accuracy thanks to the temporal trend of these informations. To extend the study in inspecting the feature sets, feature selection algorithms were used. As shown in Fig. 3.4, the classification accuracy could be maximized by using 4 features instead of 25 features depending on the feature selection algorithm. The most important features were also shown in Fig. 3.5. It is obvious that polarimetric features have higher importance but at the same time some textural features like mean values of gray level co-occurrence values of VV channel.

The third point to be underlined is the contribution of samples in classification. With clustering approaches, different ways for calculating representative features for the samples and number of samples were analysed. According to the results that are summarized in Table 3.4, it is possible to increase the classification accuracy by producing more samples within the pixels of the fields. By using the 'Six Cluster' approach which calculates 6 different samples for one field while excluding the outlier pixels of the fields, classification performance could be increased by up to 5 %. This can be a powerful solution for the cases that do not have sufficient amount of data.

4.2 Recommendations

With this study, it is obviously shown that classification could be used in order to monitor the phenological development of crops on SAR images. However, there are certain points to extend the study. First of all, with more data available with ground information, the study could be extended for different years and different study areas. The more type of data to be used, the more global the study is. Moreover, it could be interesting to model the development of the crops in one year and classify the development of another year for the same region. Also the same classification could be done for different type of crops and different type of sensors.

Secondly, the research about the classification algorithms could be extended. More different classification techniques could also be appended like semi supervised classification techniques. With semi supervised classification, algorithms might be more accurate with limited number of training samples. Another point about the classification algorithms is the fuzziness of ground information. Since BBCH values are dependent on the majority of the crops within the field and appended by visual

inspection, a fuzziness in those values are expected. Considering the samples on the class boundaries; this fuzziness could lead dramatic changes for the classification of the samples. To deal with this problem, soft classification techniques might be used although it is not that easy considering that the sensor used has only two channels to obtain information.

REFERENCES

[1] **Martínez-Ramón, M. and Christodoulou, C.** (2006). *Support Vector Machines for Antenna Array Processing and Electromagnetics*, volume 1.

[2] **MATLAB** (2015). *version 8.5 (R2015a)*, The MathWorks Inc., Natick, Massachusetts.

[3] **Rossi, C. and Erten, E.** (2015). Paddy-rice monitoring using TanDEM-X, *IEEE Trans. Geosci. Remote Sens.*, *53*(2), 900–910.

[4] **Küçük, Ç., Taşkın, G. and Erten, E.** (2016). Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images, *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, *PP*(99), 1–11.

[5] **Lopez-Sanchez, J.M., Ballester-Berman, J.D. and Hajnsek, I.** (2011). First results of rice monitoring practices in Spain by means of time series of TerraSAR-X dual-pol images, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, *4*(2), 412–422.

[6] **Le Toan, T., Laur, H., Mougin, E. and Lopes, A.** (1989). Multitemporal and dual-polarization observations of agricultural vegetation covers by X-band SAR images, *IEEE Trans. Geosci. Remote Sens.*, *27*(6), 709 – 718.

[7] **Inoue, Y. and Sakaiya, E.** (2013). Relationship between X-band backscattering coefficients from high-resolution satellite SAR and biophysical variables in paddy rice, *Remote Sens. Lett.*, *4*(3), 288–295.

[8] **Inoue, Y., Sakaiya, E. and Wang, C.** (2014). Capability of C-band backscattering coefficients from high-resolution satellite SAR sensors to assess biophysical variables in paddy rice, *Remote Sens. of Environ.*, *140*, 257 – 266.

[9] **Rossi, C. and Erten, E.** (2015). Paddy-Rice Monitoring Using TanDEM-X, *IEEE Trans. Geosci. Remote Sens.*, *53*(2), 900–910.

[10] **Erten, E., Rossi, C. and Yüzügüllü, O.** (2015). Polarization impact in TanDEM-X data over vertical-oriented vegetation: the paddy-rice case study, *IEEE Geosci. Remote Sens. Lett.*, *12*(7), 1501–1505.

[11] **Bouvet, A., Le Toan, T. and Lam-Dao, N.** (2009). Monitoring of the rice cropping system in the Mekong Delta using ENVISAT/ASAR dual polarization data, *IEEE Trans. Geosci. Remote Sens.*, *47*(2), 517–526.

[12] **Lopez-Sanchez, J.M.**, **Cloude, S.R. and Ballester-Berman, J.D.** (2012). Rice Phenology Monitoring by Means of SAR Polarimetry at X-Band, *IEEE Trans. Geosci. Remote Sens.*, *50*(7), 2695–2709.

[13] **Lopez-Sanchez, J.M.**, **Vicente-Guijalba, F.**, **Ballester-Berman, J.D. and Cloude, S.R.** (2014). Polarimetric Response of Rice Fields at C-Band: Analysis and Phenology Retrieval, *IEEE Trans. Geosci. Remote Sens.*, *52*(5), 2977–2993.

[14] **Yüzügüllü, O.**, **Erten, E. and Hajnsek, I.** (2015). Rice Growth Monitoring by Means of X-Band Co-polar SAR: Feature Clustering and BBCH Scale, *IEEE Geosci. Remote Sens. Lett.*, *12*(6), 1218–1222.

[15] **Vicente-Guijalba, F.**, **Martinez-Marin, T. and Lopez-Sanchez, J.M.** (2014). Crop Phenology Estimation Using a Multitemporal Model and a Kalman Filtering Strategy, *IEEE Geosci. Remote Sens. Lett.*, *11*(6), 1081–1085.

[16] **De Bernardis, C.G.**, **Vicente-Guijalba, F.**, **Martinez-Marin, T. and Lopez-Sanchez, J.M.** (2015). Estimation of Key Dates and Stages in Rice Crops Using Dual-Polarization SAR Time Series and a Particle Filtering Approach, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, *8*(3), 1008–1018.

[17] **Blanchart, P.**, **Ferecatu, M.**, **S., C. and Datcu, M.** (2014). Pattern Retrieval in Large Image Databases Using Multiscale Coarse-to-Fine Cascaded Active Learning, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, *7*(4), 1127–1141.

[18] **Kaya, G.T.** (2013). A Hybrid Model for Classification of Remote Sensing Images With Linear SVM and Support Vector Selection and Adaptation, *IEEE J. Sel. Topics Appl. Earth Observ. in Remote Sens.*, *6*(4), 1988–1997.

[19] **Küçük, Ç.**, **Torun, Y. and Kaya, G.T.** (2014). Remote sensing image classification by non-parallel SVMs, *IEEE Geosci. Remote Sens. Symp. (IGARSS)*, pp.1269–1272.

[20] **Giacinto, G.**, **Roli, F. and Bruzzone, L.** (2000). Combination of neural and statistical algorithms for supervised classification of remote-sensing images, *Pattern Recognition Letters*, *21*(5), 385–397.

[21] **Küçük, Ç.**, **Kaya, G.T. and Erten, E.** (2015). Co-polar SAR data classification as a tool for real time paddy-rice monitoring, *IEEE Geosci. Remote Sens. Symp. (IGARSS)*, pp.4141–4144.

[22] **Kaya, G.T.**, **Torun, Y. and Küçük, Ç.** (2014). Recursive feature selection based on non-parallel SVMs and its application to hyperspectral image classification, *IEEE Geosci. Remote Sens. Symp. (IGARSS)*, pp.3558–3561.

[23] **Kaya, G.T.**, **Ersoy, O.K. and Kamaşak, M.E.** (2011). Support Vector Selection and Adaptation for Remote Sensing Classification, *IEEE Trans. Geosci. Remote Sens.*, *49*(6), 2071–2079.

[24] **Cui, M.**, **Prasad, S.**, **Mahrooghy, M.**, **Aanstoos, J.V.**, **Lee, M.A. and Bruce, L.M.** (2012). Decision Fusion of Textural Features Derived From Polarimetric Data for Levee Assessment, *IEEE J. Sel. Topics Appl. Earth Observ. Remote Sens.*, *5*(3), 970–976.

[25] **Hänsch, R.** (2010). Complex-valued multi-layer perceptrons - An application to polarimetric SAR data, *Photogram. Eng. & Remote Sens.*, *76*(9), 1081–1088.

[26] **Ferro-Famil, L.**, **Pottier, E. and Lee, J.S.** (2001). Unsupervised classification of multifrequency and fully polarimetric SAR images based on the H/A/Alpha-Wishart classifier, *IEEE Trans. Geosci. Remote Sens.*, *39*(11), 2332–2342.

[27] **Vapnik, V.N.** (1998). *Statistical Learning Theory*, Wiley.

[28] **Zhang, L.**, **Zou, B.**, **Zhang, J. and Zhang, Y.** (2010). Classification of polarimetric SAR image based on support vector machine using multiple-component scattering model and texture features, *EURASIP J. on Adv. in Signal Proces.*

[29] **Cover, T. and Hart, P.** (1967). Nearest Neighbor Pattern Classification, *IEEE Trans. on Inf. Theory*, *13*(1).

[30] **Altman, N.S.** (1992). An introduction to kernel and nearest-neighbor nonparametric regression, *The American Statistician*, *46*(3), 175–185.

[31] **Breiman, L.**, **Friedman, J.**, **Stone, C.J. and Olshen, R.A.** (1984). *Classification and Regression Trees*, Chapman and Hall/CRC.

[32] **Quinlan, J.R.** (1986). Induction of decision trees, *Machine Learning*, 81–106.

[33] **Coppersmith, D.**, **Hong, S.J. and Hosking, J.R.M.** (1999). Partitioning Nominal Attributes in Decision Trees, *J. of Data Min. and Knowl. Discov.*, *3*, 197–217.

[34] **Kruskal, W.H. and Wallis, W.A.** (1952). Use of Ranks in One-Criterion Variance Analysis, *J. of the American Statistical Association*, *47*(260), 583–621.

[35] **Guyon, I.**, **Weston, J.**, **Barnhill, S. and Vapnik, V.N.** (2002). Gene Selection for Cancer Classification using Support Vector Machines, *Machine Learning*, *46*(1-3), 389–422.

[36] **Cawley, G.C.**, **Talbot, N.L.C. and Girolami, M.** (2007). Sparse Multinomial Logistic Regression via Bayesian L1 Regularisation, *Adv. in Neural Inf. Process. Syst.*, *19*.

[37] **Lloyd, S.P.** (1982). Least squares quantization in PCM, *IEEE Trans. Inf. Theory*, *28*(2), 129–137.

[38] (2001). Growth stages of mono-and dicotyledonous plants, **Technical Report**, Federal Biologial Research Centre for Agriculture and Foresty - Germany.

[39] **Yüzügüllü, O.**, **Marelli, S.**, **Erten, E.**, **Sudret, B. and Hajnsek, I.** (2015). Global Sensitivity Analysis of a Morphology Based Electromagnetic Scatering Model, *IEEE Geosci. Remote Sens. Symp. (IGARSS)*, pp.2743–2746.

[40] **McLean, J.**, **Hardy, B. and Hettel, G.** (2013). *Rice Almanac, 4th edition*.

[41] **Chang, C.C. and Lin, C.J.** (2011). LIBSVM: A library for support vector machines, *ACM Trans. on Intel. Sys. and Tech.*, 2, 27:1–27:27, software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[42] **Zhao, Z.**, **Morstatter, F.**, **Sharma, S.**, **Alelyani, S.**, **Anand, A. and Liu, H.** (2010). Advancing Feature Selection Research, `http://featureselection.asu.edu/featureselection\ _techreport.pdf`.

CURRICULUM VITAE
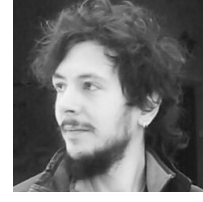
ÇAĞLAR KÜÇÜK:

Place and Date of Birth: Mersin 05/03/1990

Adress: İstanbul Teknik Üniversitesi Uzaktan Algılama Laboratuvarı, Geomatik Mühendisliği Bölümü, İnşaat Fakültesi, Ayazağa Kampüsü 34469 Maslak/İstanbul

E-Mail: cagkucuk@gmail.com

B.Sc.: Civil Engineering, Middle East Technical University

M.Sc.: Geographical Information Technologies, Istanbul Technical University

Professional Experience: 04/2014-09/2015 - Research Assistant in TUBITAK Project ID: 113Y446

List of Publications:
▪ "Paddy-Rice Phenology Classification Based on Machine-Learning Methods Using Multitemporal Co-Polar X-Band SAR Images," in *IEEE Journal of Selected Topics in Applied Earth Observation and Remote Sensing*, PP(99), 1–11
▪ "Co-polar SAR data classification as a tool for real time paddy-rice monitoring," in *Geoscience and Remote Sensing Symposium (IGARSS), 2015 IEEE International, 2015*, pp. 4141–4144
▪ "Remote sensing image classification by non-parallel SVMs," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014*, pp.1269-1272
▪ "Recursive feature selection based on non-parallel SVMs and its application to hyperspectral image classification," in *Geoscience and Remote Sensing Symposium (IGARSS), 2014 IEEE International, 2014*, pp.3558-3561