T.C.
ERCİYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASINDA HIZLANDIRMA (BOOSTING), DESTEK VEKTÖR MAKİNELERİ, RASTGELE ORMAN (RANDOM FOREST) VE REGRESYON AĞAÇLARI YÖNTEMLERİNİN KULLANILMASI

TEZİ HAZIRLAYAN
Ümit Haluk ATASEVER

TEZİ YÖNETEN
Doç. Dr. Coşkun ÖZKAN

Harita Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

OCAK 2011
KAYSERİ

T.C.
ERCİYES ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASINDA HIZLANDIRMA (BOOSTING), DESTEK VEKTÖR MAKİNELERİ, RASTGELE ORMAN (RANDOM FOREST) VE REGRESYON AĞAÇLARI YÖNTEMLERİNİN KULLANILMASI

TEZİ HAZIRLAYAN
Ümit Haluk ATASEVER

TEZİ YÖNETEN
Doç. Dr. Coşkun ÖZKAN

Harita Mühendisliği Anabilim Dalı
Yüksek Lisans Tezi

Bu çalışma Erciyes Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından FBY-9-1133 kodlu proje ile desteklenmiştir.

OCAK 2011
KAYSERİ

Doç. Dr. Coşkun Özkan danışmanlığında Ümit Haluk ATASEVER tarafından hazırlanan "**Uydu Görüntülerinin Sınıflandırılmasında Boosting, DVM, Random Forest ve Regresyon Ağaçları Yöntemlerinin Kullanması**" adlı bu çalışma jürimiz tarafından Erciyes Üniversitesi Fen Bilimleri Enstitüsü Harita Mühendisliği Anabilim Dalında Yüksek Lisans tezi olarak kabul edilmiştir.

25/1/2011

JÜRİ:

Başkan: Doç. Dr. Erkan BEŞDOK

Üye: Doç. Dr. Coşkun ÖZKAN

Üye: Yrd. Doç. Dr. Bahriye AKAY

ONAY:

Bu tezin kabulü Enstitü Yönetim Kurulunun 03/02/2011 Tarih ve 2011/06-05 sayılı kararı ile onaylanmıştır.

03/02/2011

Prof. Dr. Necmettin MARAŞLI
Enstitü Müdürü

TEŞEKKÜR

Yüksek lisans tez çalışmamda, tez konusunun seçiminden hazırlanmasına kadar tüm aşamalarında değerli görüşlerinden ve bilgisinden istifade ettiğim tez danışmanım değerli hocam Sayın Doç. Dr. Coşkun ÖZKAN'a, değerli bilgilerini benden esirgemeyen değerli hocam Doç. Dr. Erkan BEŞDOK'a ve tezim için jüri üyeliğini kabul eden sayın hocam Yrd. Doç. Dr. Bahriye AKAY'a ve son olarak bana her konuda ilgisini ve desteğini hiçbir zaman eksik etmeyen değerli aileme teşekkürlerimi sunarım.

UYDU GÖRÜNTÜLERİNİN SINIFLANDIRILMASINDA HIZLANDIRMA (BOOSTING), DESTEK VEKTÖR MAKİNELERİ, RASTGELE ORMAN (RANDOM FOREST) VE REGRESYON AĞAÇLARI YÖNTEMLERİNİN KULLANILMASI

Ümit Haluk ATASEVER
Erciyes Üniversitesi Fen Bilimleri Enstitüsü
Yüksek Lisans Tezi, Ocak 2011
Tez Danışmanı: Doç. Dr. Coşkun ÖZKAN

ÖZET

Dünyada birçok uygulamada yeryüzüne ait doğruluğu yüksek, hassas bilgi ve verilere gereksinim duyulmaktadır. Bu bilgiler kentsel gelişim, çevresel değişimlerin izlenmesi gibi çok sayıda uygulama için altlık oluşturmaktadır. Uydular tarafından elde edilen görüntülerin çok fazla veri içermesinden dolayı anlamlı bilginin çıkarılarak yorumlanabilmesi için görüntülerin ön işlemden geçirilmesi gerekir. Sınıflandırma bu aşamada kullanılan en yaygın tekniktir.

Sınıflandırma işleminin doğruluğu, kullanılacak yöntem ve parametrelerle doğrudan ilişkilidir. Bundan dolayı problemin niteliğine göre uygun yöntemin seçilmesi çok önemlidir. Literatürde En Çok Olabilirlik, En Kısa Mesafe, Yapay Sinir Ağları gibi klasik yöntemler sık olarak kullanılmaktadır.

Bu çalışmada klasik yöntemlere göre daha başarılı sonuçlar elde eden Destek Vektör Makinesi, Rastgele Orman (RF) , Sınıflandırma ve Regresyon Ağaçları yöntemleri ile sınıflandırma yöntemlerinin başarısını arttıran Boosting (Adaboost.M1) kullanılmış ve doğrulukları sınanmıştır.

Anahtar Kelimeler: Destek vektör makinesi, rastgele orman, yapay sinir ağları, sınıflandırma ve regresyon ağaçları.

THE USE OF BOOSTING, SUPPORT VECTOR MACHINES, RANDOM FOREST AND REGRESSION TREE METHODS IN SATELLITE IMAGES CLASSIFICATION

Ümit Haluk ATASEVER
Erciyes University, Graduate School of Natural and Applied Science
M.Sc. Thesis, January 2010
Thesis Supervisor: Assoc. Prof. Dr. Coşkun ÖZKAN

ABSTRACT

High accuracy and precision information and data which belong earth are needed for many applications in world. This information is base for many applications such as monitoring of environmental changes and urban development. Most important step is classification process for interpreting image obtained by satellites.

Accuracy of classification process is directly related to method and parameters. For this reason, according to the nature of the problem preferring appropriate method is very important. In literature, classical methods like Minimum Distance, Maximum Likelihood and Artificial Neural Networks are used frequently.

In this study, Support Vector Machine, Random Forest (RF) and Classification and Regression Tree Methods and Boosting (Adaboost.M1) Algorithm which increase the success of classification methods are used and tested that more successful than classical methods.

Keywords: Support vector machine, random forest, artifical neural networks, classification and regression trees.

İÇİNDEKİLER

KABUL VE ONAY ..... **i**

TEŞEKKÜR ..... **ii**

ÖZET ..... **iii**

ABSTRACT ..... **iv**

İÇİNDEKİLER ..... **v**

TABLOLAR LİSTESİ ..... **viii**

ŞEKİLLER LİSTESİ ..... **xi**

1.BÖLÜM **1**

GİRİŞ **1**

2. BÖLÜM **3**

GENEL BİLGİLER **3**

2.1. Uzaktan Algılama 3

2.2. Uzaktan Algılamanın Temel Bileşenleri 4

2.3. Sınıflandırma 5

2.3.1. Kontrollü Sınıflandırma 6

2.3.2. Kontrolsüz Sınıflandırma 7

3. BÖLÜM **8**

YÖNTEMLER **8**

3.1. Yapay Sinir Ağları 8

3.1.1. Biyolojik ve Yapay Sinir Ağları 9

3.1.2. Yapılarına Göre Yapay Sinir Ağları 11

3.1.2.1. İleri Beslemeli Yapay Sinir Ağları 11

3.1.2.2. Geri Beslemeli Yapay Sinir Ağları 12

3.1.3. Öğrenme Algoritmalarına Göre Yapay Sinir Ağları 12

3.1.3.1. Kontrollü Öğrenme 12

3.1.3.2. Kontrolsüz Öğrenme 12

3.1.3.3. Takviyeli Öğrenme 13

3.1.4. Çok Katmanlı Algılayıcı 13

3.2. Destek Vektör Makineleri 13
3.2.1. Lineer Destek Vektör Makineleri 15
3.2.1.1. Lineer Ayrılabilme Durumu 15
3.2.1.2. Lineer Ayrılamama Durumu 17
3.2.2. Lineer Olmayan Destek Vektör Makineleri 19
3.2.3. Çok Sınıflı Destek Vektör Makineleri 20
3.3. Karar Ağaçları 21
3.3.1 Karar Ağaçları Oluşturma 23
3.3.2. Böl ve Elde Et 23
3.3.3. Karar Ağaçlarında Dallanma Kriterleri 24
3.3.4. Karar Ağaçlarının Budanması 26
3.3.5. Sınıflandırma ve Regresyon Ağaçları (CART) 27
3.3.5.1. Gini İndeksi Algoritması 28
3.3.5.2. Twoing Kuralı 29
3.3.6. Ağaç Tabanlı Topluluk Yöntemleri 30
3.3.6.1. Rastgele Orman (Random Forest) 31
3.3.6.1.1. Rastgele Orman Modelinin Kurulması 32
3.3.6.1.2. Hata Oranının Tahmin Edilmesi 33
3.3.6.1.3. Hold Out Yöntemi 33
3.3.6.1.4. Çapraz Doğrulama Yöntemi 33
3.3.7. En Çok Olabilirlik Yöntemi 34
3.3.8. Hızlandırma (Boosting) Yöntemi 35
3.3.8.1. Adaboost Algoritması 36
4. BÖLÜM 38
UYGULAMA VE BULGULAR 38
4.1. Kullanılan Veri ve Yazılımlar 38
4.2. Yapay Sinir Ağları İle Görüntü Sınıflandırma 39
4.3. Rastgele Orman Yöntemi ile Görüntü Sınıflandırma 46
4.4. Sınıflandırma ve Regresyon Ağaçları (CART) İle Görüntü Sınıflandırma 52
4.5. Destek Vektör Makineleri İle Görüntü Sınıflandırma 58
5. BÖLÜM 63
SONUÇLAR 63

ÖZGEÇMİŞ.....69

TABLOLAR LİSTESİ

Tablo 3.1. Biyolojik sinir sistemi ve YSA sistemi arasındaki benzerlikler 10

Tablo 3.2. DVM Kernel Fonksiyonlar 20

Tablo 4.1. Kullanılan uydu görüntüleri ve özellikleri 38

Tablo 4.2. YSA ile sınıflandırmada kullanılan parametre değerleri 40

Tablo 4.3 Terra ASTER görüntüsünün YSA ile sınıflandırılma doğrulukları (Epoch:2000) 40

Tablo 4.4 Terra ASTER görüntüsünün YSA ile sınıflandırılma doğrulukları (Epoch:1000) 41

Tablo 4.5. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 41

Tablo 4.6. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 42

Tablo 4.7. Landsat görüntüsünün YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 43

Tablo 4.8. Landsat görüntüsünün YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 43

Tablo 4.9. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırma parametreleri(1) 44

Tablo 4.10. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 45

Tablo 4.11. Terra ASTER görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 46

Tablo 4.12. Terra ASTER görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 47

Tablo 4.13. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1). 47

Tablo 4.14. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 48

Tablo 4.15. Landsat görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 49

Tablo 4.16. Landsat görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 50

Tablo 4.17. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1). 50

Tablo 4.18. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2). 51

Tablo 4.19. Terra ASTER görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 52

Tablo 4.20. Terra ASTER görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 53

Tablo 4.21. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1). 54

Tablo 4.22. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1). 54

Tablo 4.23. Landsat görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1) 55

Tablo 4.24. Landsat görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2) 56

Tablo 4.25. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1).....56

Tablo 4.27. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(1) 58

Tablo 4.28. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(2) 59

Tablo 4.29. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(3) 60

Tablo 4.30. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(1) 60

Tablo 4.31. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(2) 61

Tablo 4.32. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(3) 62

ŞEKİLLER LİSTESİ

Şekil 2.1. Uzaktan algılamanın temel bileşenleri. 4

Şekil 3.1. Biyolojik nöronun yapısı 9

Şekil 3.2. Yapay nöronun genel yapısı 10

Şekil 3.3. Basit bir ileri beslemeli yapay sinir ağı modeli (İBYSA) 11

Şekil 3.4. Çok katmanlı perceptron genel yapısı 13

Şekil 3.5. Lineer olarak ayrılabilen ve ayrılamayan nitelik uzayı 14

Şekil 3.6. Optimum Ayırıcı Yüzey 16

Şekil 3.7. Lineer ayrılamama durumunda optimum ayırıcı yüzey 18

Şekil 3.8 Örnek bir karar ağacı yapısı 21

Şekil 3.9. Basit bir karar ağacı modeli 22

Şekil 3.10. Hunt Algortiması 23

Şekil 3.11. Topluluk (Ensemble) Yöntemi 30

Şekil 4.1. Sınıflandırmada kullanılan Terra ASTER görüntüsü 38

Şekil 4.2. Sınıflandırmada kullanılan Landsat görüntüsü 39

Şekil 4.3. Terra ASTER görüntüsünün YSA ile sınıflandırılması (Epoch:2000) 40

Şekil 4.4. Terra ASTER görüntüsünün YSA ile sınıflandırılması (Epoch:1000) 41

Şekil 4.5. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:2000,İterasyon Sayısı:20,Örnekleme Oranı:0.5) 42

Şekil 4.6. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:1000,İterasyon Sayısı:20,Örnekleme Oranı:0.5) 42

Şekil 4.7. Landsat görüntüsünün YSA ile sınıflandırılması (Epoch:1000).....43

Şekil 4.8. Landsat görüntüsünün YSA ile sınıflandırılması (Epoch:2000).....44

Şekil 4.9. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:1000,İterasyon Sayısı:20,Örnekleme Oranı:0.5).....44

Şekil 4.10. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılması …… (Epoch:2000, İterasyon Sayısı:20, Örnekleme Oranı:0.5).....45

Şekil 4.11. Terra ASTER görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:50, Özellik Sayısı:5).....46

Şekil 4.12. Terra ASTER görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:100, Özellik Sayısı:5).....47

Şekil 4.13. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:50, İterasyon Sayısı:20,Örnekleme Oranı:0.3).....48

Şekil 4.14. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:100 ,İterasyon Sayısı:27,Örnekleme Oranı:0.2).....48

Şekil 4.15. Landsat görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:50, Özellik Sayısı:5) 49

Şekil 4.16. Landsat görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:500, Özellik Sayısı:5) 50

Şekil 4.17. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:50,Özellik Sayısı:5,İterasyon Sayısı:10,Örnekleme Oranı:0.2).....51

Şekil 4.18. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:500,Özellik Sayısı:5,İterasyon Sayısı:10,Örnekleme Oranı:0.2).....51

Şekil 4.19. Terra ASTER görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:10-Gini).....53

Şekil 4.20. Terra ASTER görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:20-Gini).....53

Şekil 4.21. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılması 54

Şekil 4.22. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:20-Gini, İterasyon Sayısı:5, Örnekleme Oranı:0.2) 55

Şekil 4.23. Gini Algoritması ile oluşan ağaç yapısı 55

Şekil 4.24. Landsat görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:20-Gini) 56

Şekil 4.25. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:10-Gini,İterasyon Sayısı:11, Örnekleme Oranı:0.05) 57

Şekil 4.26. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:20-Gini,İterasyon Sayısı:15, Örnekleme Oranı:0.05) 57

Şekil 4.27. Terra ASTER görüntüsünün DVM ile sınıflandırılması (Lineer Kernel, C:900) 59

Şekil 4.28. Terra ASTER görüntüsünün DVM ile sınıflandırılması (Polinom Kerneli,C:900,d:2).....59

Şekil 4.29. Terra ASTER görüntüsünün DVM ile sınıflandırılması ..... (RTF Kerneli,C:100 ,Gama:2) 60

Şekil 4.30. Landsat görüntüsünün DVM ile sınıflandırılması (Lineer Kernel,C:900)...61

Şekil 4.31. Landsat görüntüsünün DVM ile sınıflandırılması (Polinom Kerneli, C: 20000, d:2).....61

Şekil 4.32. Landsat görüntüsünün DVM ile sınıflandırılması (RTF Kerneli,C:200,Gama:1) 62

Şekil 5.1. Terra ASTER görüntüsünün sınıflandırılmasında kullanılan yöntemler ve sınıflandırma doğrulukları.....64

Şekil 5.2. Landsat görüntüsünün sınıflandırılmasında kullanılan yöntemler ve sınıflandırma doğrulukları.....64

1.BÖLÜM

GİRİŞ

Günümüzde birçok uygulamada arazi örtüsünün kullanımına ilişkin hassas bilgilere ihtiyaç duyulmaktadır. Bu bilgiler çevresel değişimlerin saptanması, yerel yönetimlerin faaliyetleri, kaynak yönetimi gibi birçok konuda kaynak teşkil etmektedir. Bu tür değişimler ekonomik, demografik ve çevresel farklılıkları gösterdiği için ulusların en büyük bilgi kaynaklarından olmuşlardır.

Arazi kullanımındaki değişimleri anlamak ancak görüntü sınıflandırmayla mümkün olabilmektedir. Görüntü sınıflandırmayı, arazi örtüsünü tanımlayan spektral yansıtım değerlerinin ya da objelerin geometrik şekilleri, büyüklükleri ve topolojik ilişkileri gibi uzaysal özelliklerinin bir karar kuralına göre analiz edilerek, benzer özelliğe sahip piksel veya objelerin (bina, yol vs.) aynı sınıflara atanması olarak tanımlayabiliriz. Sınıflandırma işleminde kontrollü ve kontrolsüz olmak üzere başlıca iki yaklaşım vardır. Kontrollü sınıflandırmada sınıfları temsil eden eğitim verileri toplanır ve belirlenen karar-kural fonksiyonuna göre sınıflandırma işlemi yapılır. Kontrolsüz sınıflandırmada ise sınıfları temsil eden herhangi bir eğitim verisi kullanılmamakta, karar fonksiyonuna göre spektral kümeler oluşturulmaktadır.

Modern teknolojinin en fazla etkilediği alanlardan biri olan Uzaktan Algılama disiplininde değişim analizlerinde kullanmak üzere birçok sınıflandırma yöntemi geliştirilmiştir. Bu yöntemlerin başında istatistiksel bir temele gereksinim duymayan Yapay Sinir Ağları yöntemi gelmektedir. Bu yöntem parametrik bir sınıflandırma yöntemi değildir ve sınıfların normal dağılıma uyduğu varsayımına gereksinim duymaz Bu yönteme ek olarak günümüzde Destek Vektör Makineleri (DVM), Boosting (Hızlandırma), Sınıflandırma Ağaçları (CART), Rastgele Orman (RF) gibi başarılı sonuçlar üretebilen teknikler de kullanılmaya başlanmıştır.

Özellikle, ilk zamanlarda iki sınıflı problemlerin çözümünde kullanılan Boosting (Hızlandırma) ve Destek Vektör Makineleri yöntemleri son zamanların çok sık kullanılan yöntemler olmuşlardır. Boosting tekniğinde, zayıf sınıflandırıcı olarak isimlendirilen yöntemin ürettiği sonuçta yanlış sınıflandırılan her bir elemanın ağırlığı arttırılarak doğruluk arttırılmaya çalışılır.

Destek Vektör Makineleri (DVM) yönteminin temelinde ise sınıfları en iyi ayıran hiperdüzlem hesaplanmaya çalışılır. DVM, uydu görüntüleri gibi ikiden çok sınıfa sahip ve doğrusal olmayan verilerin sınıflandırılmasında, veriyi daha yüksek boyutlu bir uzayda işleme tabi tutar ve ayrımı gerçekleştirir [1].

Karar ağaçları uydu görüntüleri gibi boyutu çok büyük ve karmaşık olan verilerin sınıflandırmasında, problemi çok aşamalı bir hale getirerek karar verme işlemini gerçekleştiren çok hızlı bir yöntemdir.

Yapay zekanın başka bir uygulama alanı olan yapay sinir ağları insan beyin hücrelerindeki nöronların çalışma prensibini örnek alan bir yöntemdir. YSA farklı veri tiplerine uygulanabilmeleri, En Çok Olabilirlik (EÇO) yönteminde olduğu gibi herhangi bir istatistiksel kabul gerektirmemeleri ve yüksek öğrenebilme kabiliyetlerinden dolayı uzaktan algılanmış görüntüleri sınıflandırılmasında çok sıklıkla kullanılan bir yöntem olmuştur [2].

Bu tez çalışmasında Kayseri iline ait farklı tarihlerdeki Terra ASTER ve Landsat uydu görüntüleri sınıflandırma işlemine tabi tutulmuştur. Görüntüler Yapay Sinir Ağları ile Destek Vektör Makineleri (DVM), Boosting (Hızlandırma), Sınıflandırma Ağaçları (CART), Rastgele Orman (RF) ve Boosting yöntemleri kullanılmış ve performansları değerlendirilmiştir.

2. BÖLÜM

GENEL BİLGİLER

2.1. Uzaktan Algılama

Günümüzde uzaktan algılama ile elde edilmiş uydu görüntüleri, arazi değişiminin belirlenmesi, çevresel ve kentsel değişim analizlerinin yapılması için en temel gereksinimlerden biridir.

Uzaktan Algılama'nın günümüzde çok sayıda tanımı bulunmaktadır. Uzaktan Algılama, arada fiziksel bir temas olmaksızın bir cisimden yayılan ışınımın nitelik ve nicelik yönünden değerlendirilmesi ile cismin özelliklerinin uzaktan ortaya konması ve ölçülmesi şeklinde tanımlanabilir. Bu tanımın içine; Elektromanyetik ışımanın, akustik enerjinin, nükleer ışımanın, ortamdaki kuvvet alanlarının, algılanması girse de, uzaktan algılamada asıl olan Elektromanyetik Işıma'dır (EMR) [3].

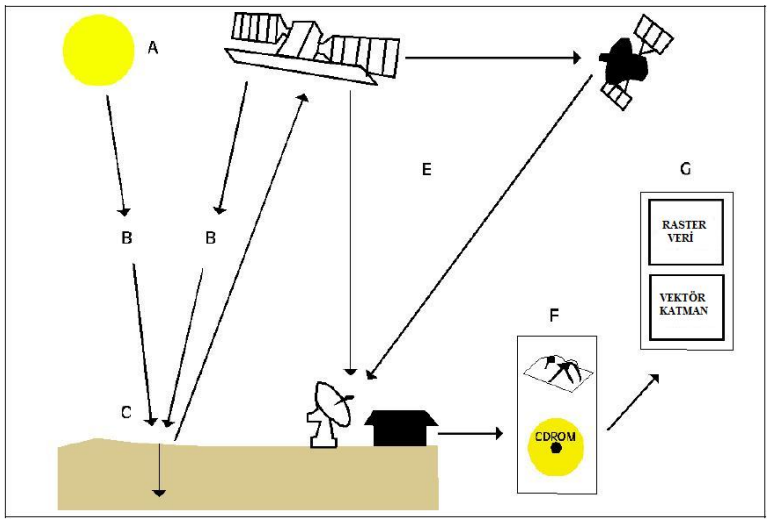
Uzaktan Algılama çok sayıda bilim dalı tarafından yoğun olarak kullanılmaktadır. Uzaktan Algılama'nın kullanım alanlarına şu şekilde sıralayabiliriz:

- Meteoroloji
- Oşinografi
- Buzul bilimi
- Jeoloji, jeomorfoloji ve jeodezi
- Topografya ve kartografya
- Ziraat, ormancılık ve botanik
- Doğal felaket kontrolü
- Planlama uygulamaları
- Askeri uygulamalar

Uzaktan algılamanın sağladığı çok çeşitli sinoptik karakterli bilgiler CBS (Coğrafi Bilgi Sistemleri) ortamında başka bilgilerle bütünleştirilip veya o bilgilere altlık olarak kullanılıp isteyen herkesin ulaşabileceği sistematik bir bilgi bankası oluşturulabilir. Böylece uygulamaların etkinliği süre ve maliyet açısından önemli derecede arttırılırken bilgilerin güncelliği ve buna bağlı olarak da gerçek duruma uygunluğu sağlanmış olur [3].

2.2. Uzaktan Algılamanın Temel Bileşenleri

Uzaktan Algılama; bir enerji kaynağı, enerjinin atmosferdeki etkileşimi, yeryüzüyle etkileşimi, enerjinin bir sensör tarafından veri olarak kaydedilmesi ve verinin görsel-sayısal analiz için gösterilmesi olmak üzere beş temel bileşenden oluşur.



Şekil 2.1. Uzaktan algılamanın temel bileşenleri, A: Enerji kaynağı (EMR), B: Işıma ve atmosfer, C: Hedefle etkileşim, D:Sensör tarafından enerjinin kaydedilmesi, E: İletim ve işleme, F:Yorumlama ve analiz, G:Uygulama.

Yapılan çalışmalar, elektromanyetik enerjinin hem dalga hem de parçacık karakterli olduğunu göstermiştir. Uzaktan algılama için enerjinin bu çift özelliği bir paradoks değil aksine tam bir uygunluk göstermektedir. Dalga karakteriyle elektromanyetik spektrum oluşturulurken, parçacık karakteriyle de elektromanyetik enerjinin atmosfer ve yeryüzüyle etkileşimi açıklanmaktadır. EMR'nin dalga formunda sürekli modeli Maxwell tarafından araştırılmıştır, parçacık karakteri Planck tarafından açıklanmıştır. Parçacık anlamında enerji sürekli dalgalarla aynı doğrultuda ve hızda yayılan küçük

parçacıklar (kuantum) olarak düşünülür. Her bir ışıma parçacığının sahip olduğu enerji foton olarak adlandırılır [3].

Elektromanyetik enerji maddenin atom altı titreşimlerden kaynaklanır ve dalga boyu ile ölçülür. EMR ışık hızında yol alır ve nesneler tarafından yansıtılır ve/veya absorbe edilir ve/veya geçirilir. Bütün EMR'nin orijini foton olarak adlandırılan kütlesiz atom altı parçacıkların vibrasyonuyla başlar. Mutlak sıcaklığın (-273.15 ºC) üstündeki bütün cisimler vibrasyon etkisiyle belirli formda bir ışıma yapar. Sıcak cisimler daha çok ışıma yaparken göreceli olarak soğuk cisimler daha az ışıma yapar. Sıcaklık bir objeden yayılan vibrasyonel enerjinin bir ölçüsüdür.

Elektron yüksek enerji seviyesinden alçak enerji seviyesine geçerken foton yayınlanır. Tersi durumda da atoma foton geldiğinde uygun enerji durumunda elektron yüksek enerji seviyesine çıkar. EMR'nin en büyük kaynağı olan güneşin çekirdeğindeki hidrojenden helyuma olan dönüşümle dış katmanlarından enerji yayımı yapılır. Bu enerjinin atmosferin dış katmanındaki ortalama birim ışınırlık miktarı 1367 Wm^{-2}. Bu değere güneş sabiti denir. Gerçek birim ışınırlık yıllık çevrimde dünya güneş konumuna bağlı olarak %3 oranında değişkenlik gösterir [3].

Ortalama olarak, gelen birim ışınırlığın %36'sı atmosfer ve bulut ortamlarını da içeren yeryüzü tarafından yansıtılır. Gelen enerjinin %17'si atmosfer tarafından, %47'si yeryüzü tarafından yutulur.

EMR bir sinusoidal dalga şeklinde ışık hızında yayılır. EMR yayılma doğrultusuna ve birbirlerine dik iki bileşene sahiptir: Elektrik ve manyetik alanlar. Bu enerji frekans (ν) ve dalga boyu (λ) ile belirlidir. Elektrik ve manyetik alanlar birbiriyle aynı fazda aynı frekans ve dalga boyunda hareket doğrultuna dik doğrultuda salınım hareketi yaparlar.

2.3. Sınıflandırma

Ham halde edilen uydu görüntülerinden herhangi bir işlem yapmadan, konumsal yorum ve analiz yapmak oldukça zordur. Bu verilerden anlamlı bilgiye ulaşmanın temel yolu sınıflandırmadır. Uzaktan algılamada, doğadaki her nesne farklı dalga boylarında farklı yansıtım ve yutulma davranışları gösterdiği için çok kanallı veriler kullanılmaktadır. Kanal sayısı yapılan çalışmanın niteliğine göre farklılıklar gösterebilir.

Uzaktan algılamada sınıflandırma, görüntüdeki tüm piksellerin özellik grubunun belirlenmesi ve doğru gruba atanması işlemi olarak tanımlanabilir.

Spektral tabanlı sınıflandırmada, pikseller ait yansıtım değerleri kullanılır ve tüm pikseller ayrı ayrı sınıflandırılırlar. Uzaysal tabanlı sınıflandırmada ise görüntüdeki objelerin büyüklükleri, geometrik şekilleri gibi uzaysal nitelikleri kullanılır [3].

Sınıflandırma işlemi sonucunda üretilecek olan sınıflar tematik sınıflar ve spektral sınıflar olarak ikiye ayrılırlar. Tematik sınıflar toprak bitki örtüsü türleri ve su gibi arazi örtüsü sınıflarına ek olarak yerleşim, sanayi ve park alanları gibi arazi kullanım sınıflarını içerirler. Spektral sınıflar ise verideki yansıtım değerlerine göre oluşturulan homojen piksel guruplarıdır [3].

Literatürde kontrollü ve kontrolsüz olmak üzere iki tür sınıflandırma kullanılmaktadır. Kontrollü sınıflandırmada, sınıflandırmayı yapacak fonksiyonun eğitimi için eğitim verilerinin seçilmesi gerekmektedir. Kontrolsüz sınıflandırmada ise eğitim verisi gerekmemekte, bunun yerine yöntemin kriterlerine göre kümeleme yapılmaktadır.

2.3.1. Kontrollü Sınıflandırma

Kontrollü sınıflandırmada, görüntünün hangi sınıflara ayrılacağı kullanıcı tarafından belirlenir. Bunun için kullanıcı tarafından eğitim alanları seçilir. Eğitim alanları seçimi sırasında sınıf çakışmalarının ortaya çıkmaması için dikkatli olunması gerekmektedir. Eğitim alanlarının seçimi sınıflandırma doğruluğunu doğrudan etkileyen bir işlemdir. Kontrollü sınıflandırma işlemi şu adımları içerir:

1. Çalışma yapılacak bölgenin özelliklerini ortaya çıkarabilecek sınıf ve kontrol alanlarının belirlenmesi.
2. Uydu görüntüsü ve çalışmanın niteliğine göre spektral bantların seçilmesi.
3. Kontrol alanlarından eğitim ve test alanlarının seçilmesi.
4. Sınıflandırma yönteminin seçilmesi ve uygulanması.
5. Eğitim ve test alanlarını kullanarak sınıflandırılmış görüntünün doğruluk analizlerinin yapılması ve yorumlanması [4].

Eğitim alanlarının seçimi dışında, sınıflandırma doğruluğunu etkileyen en önemli etken uygun sınıflandırma metodunun seçilmesidir. Yakın zamana kadar En Kısa Mesafe, En Çok Olabilirlik ve YSA gibi yöntemler yoğun olarak kullanılmaktayken; daha yüksek doğruluklar sağlayabilen, istatistiksel öğrenme temelli DVM, birçok karar ağacının birleştirilmesi ile geliştirilen RF ve yine karar ağacı temeli CART yöntemlerin yanı sıra mevcut yöntemlerin doğruluğunu arttıran Hızlandırma (Boosting) algoritmaları da kullanılmaya başlanmıştır.

2.3.2. Kontrolsüz Sınıflandırma

Kontrolsüz sınıflandırma, herhangi bir eğitim verisi olmadan belirlenen yöntemin işleyiş biçimine göre benzer piksellerin kümelenmesi olarak tanımlanabilir. Kontrolsüz sınıflandırma genelde çalışma bölgesine ait yeterli bilgi olmadığı durumlarda kullanılan bir yöntemdir. Bu sınıflandırma türünde, kullanılan yöntemin parametreleri ve küme sayısı dışında herhangi bir bilgiye gerek duyulmaz. Benzer özelliklere sahip pikseller aynı sınıf ya da küme etiketine atanırlar.

Kontrolsüz sınıflandırmada, piksellerin küme merkezlerine olan uzaklıkları hesaplanır ve kendisine en yakın olan küme ya da sınıfa atanırlar. Uygulamalarda, K-Means, Fuzzy C-Means ve Isodata gibi klasik yöntemlerin yanı sıra yapay zeka optimizasyon algoritmalarından Diferansiyel Gelişim, Yapay Arı Kolonisi ve Parçacık Sürüsü algoritmaları da kullanılmaya başlanmıştır.

3. BÖLÜM

YÖNTEMLER

3.1. Yapay Sinir Ağları

Yapay sinir ağları diğer istatistiksel sınıflandırma yöntemleri ile karşılaştırıldığında verilerin karmaşık yapıya sahip olduğu ve gürültü içerdiği durumlarda ve veriler arasındaki doğrusal olmayan ilişkilerin öğrenilmesinde başarılı ve güçlü bir sınıflandırma yöntemidir [5,6]. Yapay sinir ağlarının temelinde insan beynine ait üstün karakteristikleri taklit ederek, yazılımları tanıma-öğrenme süreçlerin performanslarını arttırmak vardır. Yapay sinir ağları, mühendislik, finans gibi neredeyse tüm fen ve matematiksel branşlarda yoğun olarak kullanılmaktadır. Yapay sinir ağlarının en önemli avantajlarını şu şekilde sıralayabiliriz [6]:

- Parametrik olmayan bir yapıya sahiptir,
- Kurallar eğitim ile verilerden yapılır,
- Farklı yapıdaki verilerin birleştirilmesine olanak sağlar,
- Genelleştirme yeteneği yüksektir,
- Diğer yöntemlere karşın gürültüye karşı güçlüdür.

Yapay sinir ağlarının en önemli avantajları, yöntemin herhangi bir istatistiksel kabule gereksinim duymaması ve küçük boyutta eğitim seti ile istatistiksel yöntemlere göre daha başarılı sonuçlar üretebilmesidir [5]. Diğer yöntemlere nazaran daha başarılı sonuçlar üretmesine rağmen, YSA bazı dezavantajlara sahiptir. Daha fazla eğitim süresine ihtiyaç duyması, ağ eğitim zamanını ve sınıflandırma başarısını doğrudan etkileyen parametre değişimlerine karşı hassas olması bunların en önemlileri arasındadır [6,7].

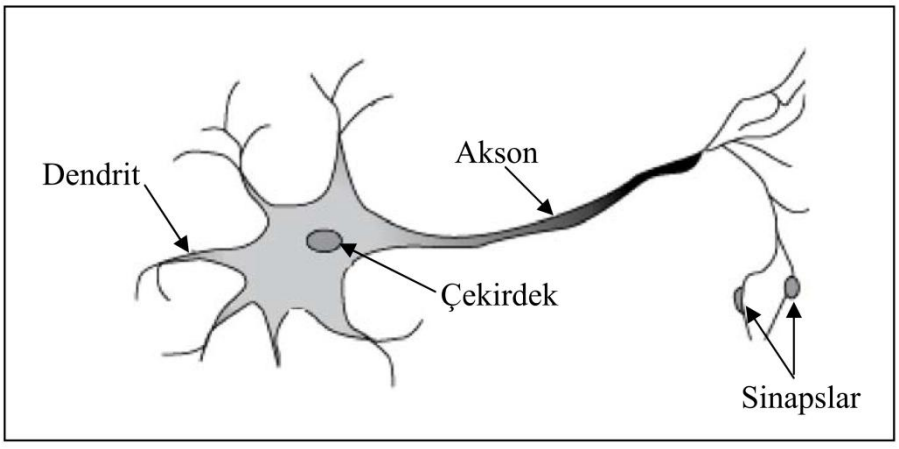
Şimdiye kadar çok sayıda sinir ağı modeli üretilmiş ve başarıları sınanmıştır. Yapay Sinir Ağı modellerini kontrollü öğrenme ve kontrolsüz öğrenme olmak üzere iki

kategoride incelemek mümkündür. Hopfield ağları, SOM ve Grossberg ağları kontrolsüz öğrenme; çok katmanlı perseptron ve RBF gibi modeller ise kontrollü öğrenmeye örnek gösterilebilirler. Başka bir sınıflandırma kriteri ise ağ modelinin öğrenme yönteminin yönü ile ilgili olandır. Bunda da yine iki durum söz konusudur. Bilgi akışı giriş katmanından çıkış katmanına doğru ise bu ağ yapısı ileri beslemeli, tam tersi durumda ise geri beslemeli ağ modeli olarak adlandırılır.

Uzaktan algılama görüntülerinin sınıflandırılmasında en yaygın olarak kullanılan yöntem çok katmanlı perseptron (ÇKP) modelidir.

3.1.1. Biyolojik ve Yapay Sinir Ağları

İnsan beyni sinir sisteminin en önemli elemanıdır. Buna rağmen çalışma şekli çok karmaşık bir yapıya sahip olmasından dolayı henüz açıklanamamış bir konudur. Beyin sürekli olarak vücuttan gelen sinyalleri alır, biyolojik nöronları (sinir hücreleri) ile işler ve ardından tekrar iletir. Biyolojik nöronlar dendrit, çekirdek, akson ve sinaps olmak üzere dört temel bileşenden oluşurlar. Dendritler nöronun giriş katmanıdır. Sinapslar nöronlar arası etkileşimi sağlar. Çekirdek ise dendritten gelen bu bilgiyi işler ve bilgi ilk olarak aksona oradan da sinapslar aracılığıyla başka bir nörona iletilirler. Şekil 3.1'de bir biyolojik sinir yapısı gösterilmektedir.



Şekil 3.1. Biyolojik nöronun yapısı

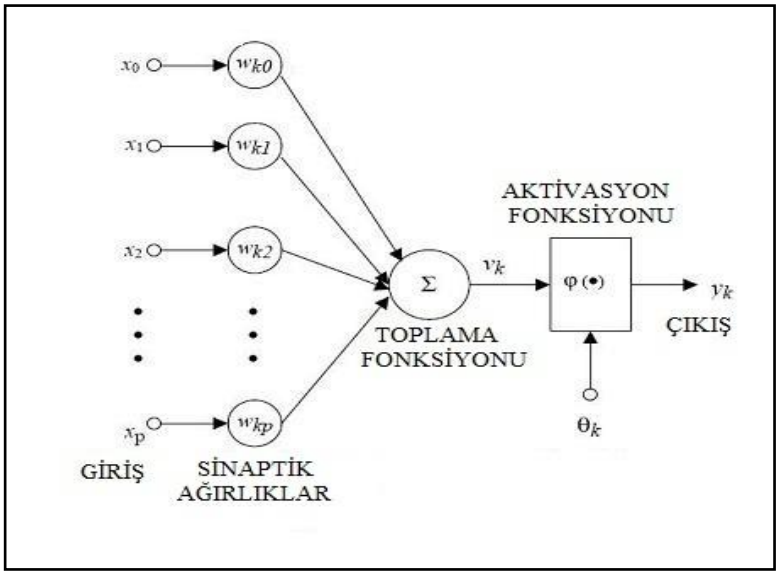
Yapay sinir ağları genel yapısı itibari ile biyolojik beyinden esinlenerek geliştirilen yöntemler olduklarından temel elemanları nöronlardır. Yapay nöronlar birleşerek yapay

sinir ağlarını oluşturmaktadırlar. Doğal ve yapay nöronlara ilişkin benzerlikler Tablo 3.1'de verilmiştir.

Tablo 3.1. Biyolojik sinir sistemi ve YSA sistemi arasındaki benzerlikler

YAPAY SİNİR AĞLARI	DOĞAL SİNİR SİSTEMİ
Ağırlıklar	Sinapslar
Eleman Çıkışı	Aksonlar
Transfer Fonksiyonu	Hücre Gövdesi
Toplama Fonksiyonu	Dendrit
İşlem Elemanı	Nöron

Şekil 3.2'de x_i olarak tanımlanan giriş değerleri w_{ki} ile gösterilen ağırlık değerleri ile çarpılırlar. Basit YSA yapısında bu çarpımlar toplam fonksiyonu ile toplanır ve bir aktivasyon fonksiyonu kullanılarak sonuç üretilir. Uygulamalarda farklı toplam ve aktivasyon fonksiyonları kullanılarak farklı ağ yapıları elde edilebilir [8].



Şekil 3.2. Yapay nöronun genel yapısı

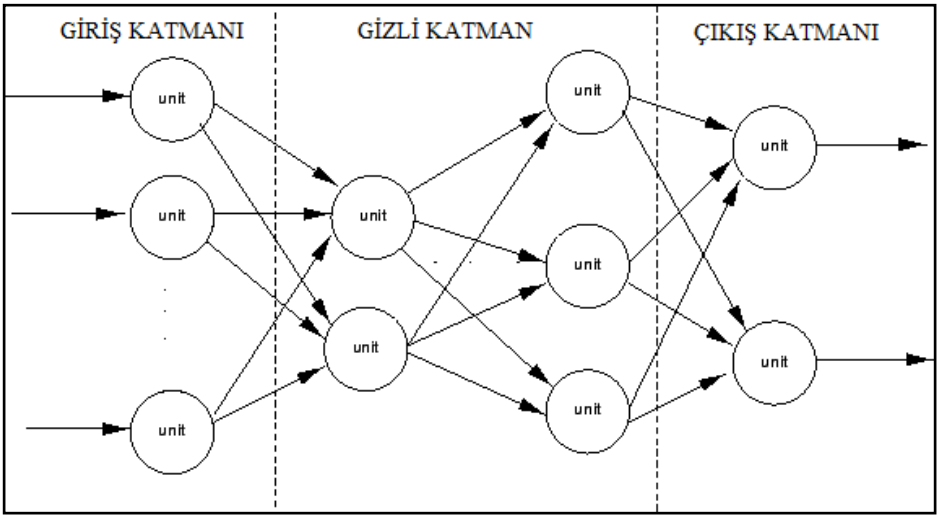
Aktivasyon fonksiyonu, YSA'ya lineer olmayan bir yapı kazandırmak için kullanılan matematiksel bir fonksiyondur. Bu sayede doğrusal olmayan problemlerin çözümünde kullanılabilecek bir yapıya sahip olmaktadır. Çok katmanlı perseptron modelinde aktivasyon fonksiyonu olarak genellikle tanjant hiperbolik ve logaritmik sigmoidal fonksiyonları kullanılır [4,8].

3.1.2. Yapılarına Göre Yapay Sinir Ağları

Yapay sinir ağları nöronların birbirine bağlanma biçimlerine göre ileri beslemeli ve geri beslemeli ağlar olmak üzere ikiye ayrılırlar.

3.1.2.1. İleri Beslemeli Yapay Sinir Ağları

İleri besleme sinir ağları, geliştirilen ağ yapıları içinde en basit yapıya sahip olanlardan ilki olmuşlardır. Verilerin giriş katmanından çıkış katmanına doğru tek yönlü bağlantılar ile iletildiği bir ağ yapısına sahiptirler. Bu ağ yapısında, yapay nöronlar bir üst katmanın yapay nöronları ile bağlantılı olmalarına rağmen aynı katmanda bulunan yapay nöronlar arasında hiçbir bağlantı bulunmamaktadır. İleri beslemeli ağlarına örnek olarak LVQ (Learning Vector Quantization) ve ÇKP (Çok Katmanlı Perseptron) verilebilir. Şekil 3.3'de basit bir ileri beslemeli yapay sinir ağı modeli görülmektedir.



Şekil 3.3. Basit bir ileri beslemeli yapay sinir ağı modeli (İBYSA)

3.1.2.2. Geri Beslemeli Yapay Sinir Ağları

Geri beslemeli yapay sinir ağlarında nöronların çıkışları aynı nörona ya da önceki katmandaki geri iletilirler. Bu sayede de bilgi hem ileri hem de geri yönde iletilmiş olur. Geri beslemeli yapay sinir ağlarında (GBYSA) herhangi bir çıkış giriş verilerinin bir fonksiyonu olmasına rağmen çıkış değerlerini de yansıtılmaktadır. GBYSA, bilgi iletiminin tek yönlü olmamasından dolayı ileri beslemeli sinir ağlarının aksine ağırlıklar dinamik bir hafıza yapısına sahiptir. Bu ağlarda, bağlantıların bağlanma biçimlerine göre farklı davranışta veya yapıda geri beslemeli sinir ağı elde edebilmek mümkündür. Hopfield, Elman ve Jordan Ağları bu tür yapılara örnek olarak verilebilir [5].

3.1.3. Öğrenme Algoritmalarına Göre Yapay Sinir Ağları

Yapay sinir ağlarının giriş değerlerine göre çıkış değerlerini üretebilmesi ağın öğrenme yetisine bağlıdır. Yapay sinir ağları öğrenme algoritmalarına göre kontrollü, kontrolsüz ve takviyeli olmak üzere üçe ayrılırlar.

3.1.3.1. Kontrollü Öğrenme

Bu tip öğrenme modelinde, öznitelik bilgileri giriş verisi, öznitelik bilgilerine ait sınıf etiket değerleri ise çıkış verisi olarak tanımlanır. Öğrenme algoritması, nöronlar arasındaki bağlantı ağırlıklarını, gerçek etiket değerleri ve elde edilen etiket değerleri arasındaki farka göre güncellemektedir. LVQ ve Delta Kuralı algoritmaları kontrollü öğrenmeye örnek olarak verilebilir.

3.1.3.2. Kontrolsüz Öğrenme

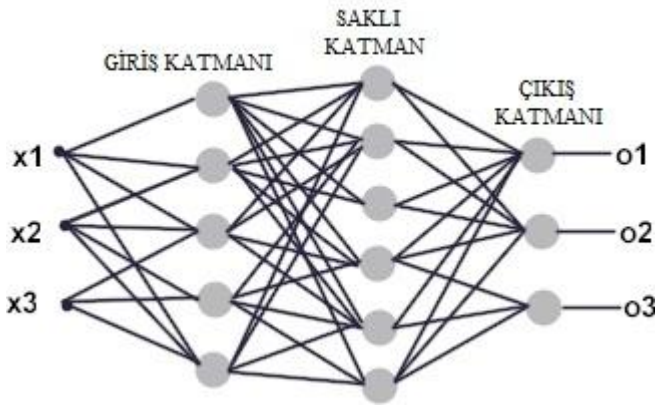
Kontrolsüz öğrenmede ağ çıkış bilgisine ihtiyaç duymaz, bunun yerine ağırlıkları giriş paternlerini gruplara ayırarak ayarlanırlar. Kohonen tarafından geliştirilen SOM ve Grossberg tarafından geliştirilen ART öğrenme algoritmaları kontrolsüz öğrenmeye örnek olarak verilebilir.

3.1.3.3. Takviyeli Öğrenme

Kontrollü öğrenmenin özel bir hali olan bir öğrenme biçimidir. Sinir ağına çıkış değerleri verilmek yerine, üretilen çıkışın giriş değerlerine olan uygunluğu belirli bir kriter kullanılarak değerlendirilir. Boltzman Kuralı ve Genetik Algoritma takviyeli öğrenmeye örnek olarak gösterilebilirler.

3.1.4. Çok Katmanlı Algılayıcı

Günümüzde uydu görüntülerinin sınıflandırılması için en sık kullanılan yapay sinir ağı modeli çok katmanlı algılayıcıdır. Çok katmanlı algılayıcılar esnek ve çok sayıdaki lineer olmayan nöronlardan oluşurlar. Nöronlar genellikle, döngü oluşturmayan ileri iletimli bağlantılara sahiptirler. Çok katmanlı algılayıcı Şekil 3.4'de görüldüğü gibi giriş katmanı, saklı katman ve çıkış katmanı olmak üzere üç bölümden oluşur.



Şekil 3.4. Çok katmanlı perceptron genel yapısı

3.2. Destek Vektör Makineleri

Genel anlamda istatistiksel öğrenme, sonlu sayıda eğitme verisi kullanan tahmin fonksiyonunun öğrenilmesini sağlayan öğrenme makinesi veya algoritmanın oluşturulması olarak tanımlanabilir. Öğrenme işlemi, regresyon, kümeleme ve sınıflandırma gibi işlemler için temel basamak niteliğindedir.

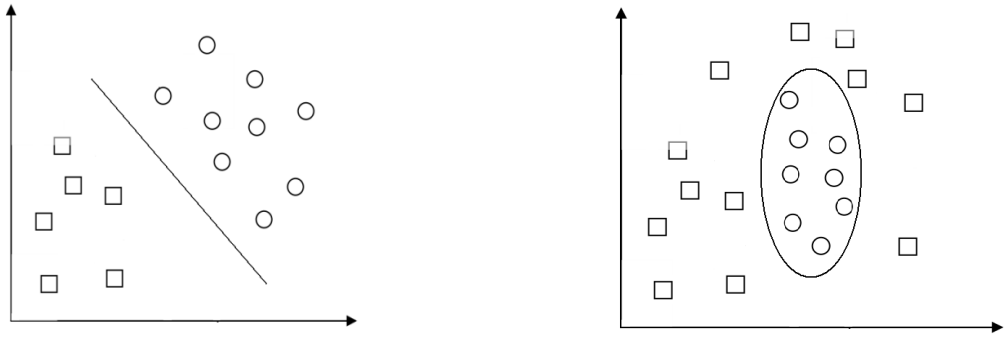
Makine öğrenmesi yöntemleri geçmişteki verileri kullanarak, verinin ve yapılacak işlemin niteliğine göre (sınıflandırma, regresyon) en uygun modeli hesaplamayı amaçlar.

Makine öğrenmesi teknikleri, istatistiksel dağılıma bağımlı ve bağımsız olmak üzere iki ana başlık altında incelenirler. Bu tekniklerde eğitimli ve eğitimsiz olmak üzere ikiye ayrılırlar [9]. Destek Vektör Makineleri yöntemi dağılımdan bağımsız, eğitimli ya da yarı eğitimli olarak sınıflandırma işlemlerinde kullanılabilirler.

Eğiticili öğrenmede, eğitim örneklerine ait çıkış değerleri öğretici veya yine öğretici tarafından modellenmiş sistem tarafından belirlenir. Eğiticisiz öğrenmede ise herhangi bir çıkış temsil edilmez bunun yerine giriş dağılımın olasılığının yani girilen verinin doğal kümeleme yapısının bulunması amaçlanır [10].

DVM istatistiksel öğrenme teorisi alanında ortaya atılan, Cortes ve Vapnik tarafından geliştirilmiş bir öğrenme metodudur [11]. DVM temel olarak, lineer olmayan örnek uzayının, örneklerin lineer olarak ayrılabileceği bir yüksek boyuta aktararak farklı örnekler arasındaki maksimum sınırın bulunması esasına dayanır [12].

Destek Vektör Makinelerinde, verilerin lineer olarak ayrılabilmeleri ve lineer olarak ayrılamamaları olarak iki yapı bulunur. Lineer olarak ayrılmış verilerden maksimum sınırın bulunması oldukça kolay olmasına rağmen, lineer olarak ayrılamayan verilerden maksimum sınırın bulunması için, verilerin lineer olarak ayrılabilecekleri bir uzaya aktarılmaları gerekir [12].



Şekil 3.5. Lineer olarak ayrılabilen ve ayrılamayan nitelik uzayı

3.2.1. Lineer Destek Vektör Makineleri

3.2.1.1. Lineer Ayrılabilme Durumu

Eğitim amacıyla kullanılacak veri setinin N elemandan oluşan $\chi = \{x_i, y_i\}, i = 1,2,....N$ olduğunu varsayalım. Burada $y_i \in \{-1,1\}$ etiket değerlerini ve $x_i \in R^d$ özellik vektörünü ifade etmektedir. Lineer olarak ayrılabilme durumunda ayırıcı aşırıdüzlem ile ayrılabilir. DVM, bu aşırı düzlemin örnek gruplarına eşit uzaklıkta olmasını amaçlayan bir yöntemdir [13]. Veri kümelerini birbirinden ayırmak için birçok aşırıdüzlem kullanılabilir. Ancak, en uygun yol iki aşırıdüzlem arasında en büyük boşluğa sahip olanları seçmektir. Örnek olarak Şekil 3.6 üzerinde yer alan H_1 ve H_2 göz önüne alınabilir. Bu iki aşırıdüzlemin ortasına oluşturulan H_0 ise optimal ayırma düzlemi adı verilen ve iki sınıf veriyi birbirinden ayıran doğrusal aşırıdüzlemdir. Bir aşırıdüzlem üzerindeki noktalar cinsinden H_0 düzlemi şu şekilde tanımlanır:

$$H_0 = W^T X + b = 0 \tag{3.1}$$

Bu ifade şu şekilde de yazılabilir:

$$\sum_{i=1}^{n} w_i x_i + b = 0 \tag{3.2}$$

Burada W ağırlık vektörünü $W = \{w_1, w_2, ..., w_n\}$; n ise niteliklerin sayısını göstermektedir. İfade içinde yer alan b ise sabit bir sayıyı göstermektedir. S_1 ve S_2 olmak üzere iki niteliğimizin olduğunu varsayalım. Eğitim kümesi ise iki boyutlu uzay söz konusu olduğu için $X = (x_1, x_2)$ biçimindedir. Burada x_1 ve x_2 değerleri X için, S_1 ve S_2 'nin değerleri olarak göz önüne alınır. H_1 aşırıdüzlemi şu şekilde ifade edilir:

$$H_1 : W^T X + b = 1 \tag{3.3}$$

Aynı şekilde H_2 aşırıdüzlemi

$$H_2 : W^T X + b = 1 \tag{3.4}$$

biçiminde ifade edilir. Şekil 3.6' da belirtilen aşırıdüzlemin üst tarafında kalan noktalar şu eşitsizliğe uymaktadır:

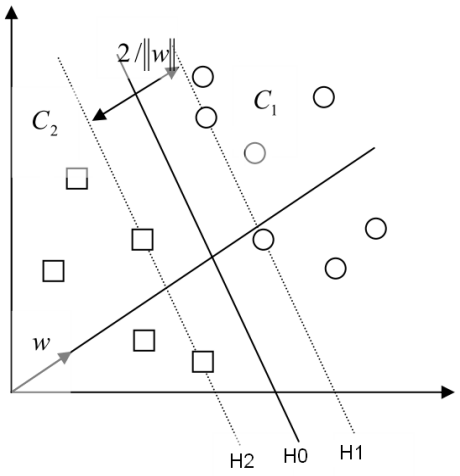
$$W^T X + b > 0, y_1 = +1 \tag{3.5}$$

Benzer şekilde aşırıdüzlemin alt kısmında kalan noktalar da aşağıdaki eşitsizliğe uymaktadır:

$$W^T X + b < 0, y_2 = -1 \tag{3.6}$$

Bu iki eşitsizlikten faydalanarak,

$y_i(w^T x_i + b) \geq +1$ elde edilir. Bu koşulu sağlayan aşırıdüzlemin iki kümenin en yakın örneklerine olan dik uzaklıkları toplamı sınır olarak adlandırılır. Optimum aşırı düzlemi bulmak için uygun w ve b(orijine olan uzaklık) değerleri hesaplanmalıdır [6].



Şekil 3.6. Optimum ayırıcı yüzey

Optimum aşırı düzlem sınırını en yüksek değere ulaşabilmesi için $\|w\|$ değerinin minimum olması gerekir. Bu durumda en uygun aşırı düzlemin bulunması için aşağıda belirtilen kısıtlı en uygun şekle sokma probleminin çözümü gerekmektedir [13].

$$\min\left[\frac{1}{2}\|w\|^2\right] \tag{3.7}$$

Bu optimizasyon probleminin sınırlamaları ise;

$y_1(w \bullet x_i + b) - 1 \geq 0$ ve $y_i \in \{1, -1\}$ şeklindedir [14]. Bu optimizasyon probleminin çözümü için Lagrange formülasyonu yapılır. Problemin Lagrange denklemi ise,

$$L_p = \frac{1}{2}\|w\|^2 - \sum_{i=1}^{N} \alpha_i y_i (w^T x_i + b) + \sum_{i=1}^{N} \alpha_i \tag{3.8}$$

şeklindedir [15]. Bu eşitlikte α_i>0 değerleri pozitif Lagrange çarpanları olarak adlandırılır ve Destek Vektör Makineleri'ni tanımlamaktadır. Denklemin çözümü karmaşık olduğundan dolayı ifade Karush-Kuhn-Tucker koşulları kullanılarak dual problem formuna dönüştürülür. Problemin çözümü için KKT koşulları:

$$\frac{\partial L(w_1, b, \alpha)}{\partial w} = w = \sum_{i=1}^{n} x_i y_i \alpha_i = 0 \tag{3.9}$$

$$\frac{\partial L(w_1, b, \alpha)}{\partial b} = \sum_{i=1}^{n} y_i \alpha_i = 0 \tag{3.10}$$

şeklindedir. Bu koşullar problemin Lagrange formülasyonunda yerine konacak olursa:

$$L_d = \frac{1}{2}(w_T w) - w_T \sum_i y_i \alpha_i x_i - b \sum_i y_i \alpha_i + \sum_i \alpha_i \tag{3.11}$$

$$= \frac{1}{2}(w_T w) + \sum_i \alpha_i \tag{3.12}$$

$$= \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y_i y_j x_i^T x_j \qquad \alpha_i \geq 0, \forall i \tag{3.13}$$

ifadesi elde edilir. Burada tüm eğitim örnekleri için bir Lagrange çarpanının olduğu görülmektedir. Çözümde elde edilen çarpanların büyük bir kısmı sıfır olacaktır. Geriye kalan α_i>0 değerlerine sahip x_i örnekleri destek vektörleridir ve H_1 veya H_2 aşırıdüzlemlerinin üzerinde yer alırlar [5].

3.2.1.2. Lineer Ayrılamama Durumu

Birçok problemde olduğu gibi uydu görüntülerinin de sınıflandırılmasında veri setinin lineer olarak ayrılması mümkün değildir. Bu durumda eğitim verilerinin bir kısmının

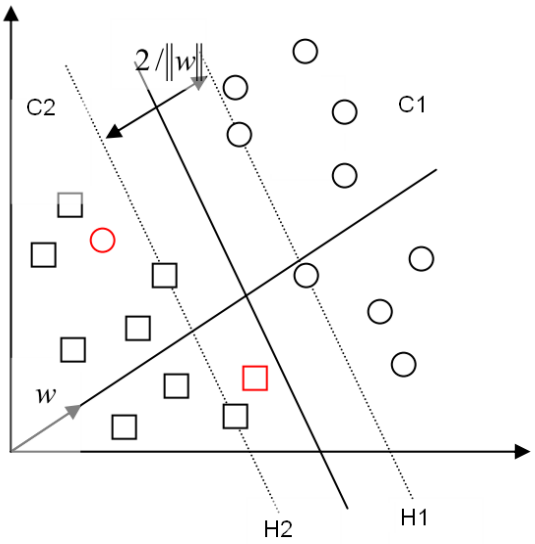
aşırıdüzlemin diğer tarafında kalması problemiyle karşılaşılır ve problem pozitif, hataları ifade eden, yapay bir değişkenin $(\xi_i, i=1,2,...,N)$ optimizasyona eklenmesi ile çözülmeye çalışılır. (3.5) ve (3.6)'daki eşitsizlikler yapay değişken (zayıflık değişkeni) ile çözülmeye çalışılırsa, yeni eşitsizlik:

$$w^T x_i + b \geq 1 - \xi_i, y_i = +1 \text{ için,} \tag{3.14}$$

$$w^T x_i + b \leq -1 + \xi_i, y_i = -1 \text{ için} \tag{3.15}$$

$$\xi_i \geq 0, \forall i \tag{3.16}$$

şeklinde olacaktır. Denklemde $\xi_i = 0$ ise örnek doğru sınıflandırılmış, $0 < \xi_i < 1$ durumunda örnek doğru sınıflandırılmış , H_1 ve H_2 aşırıdüzlemleri arasında yer alıyor; son olarak $\xi_i \geq 1$ ise sınıflandırma yanlış yapılmış demektir [15].



Şekil 3.7. Lineer ayrılamama durumunda optimum ayırıcı yüzey

Lineer olarak ayrılamayan bir sistemde, hiperdüzlemler arasındaki uzaklığı maksimum yaparken, sınıflandırma hatalarını minimum yapacak; Lagrange çarpanlarının alabilecekleri maksimum değeri gösteren bir C üst sınırının eklenmesi gerekir. Bu bilgilerle Lagrange denkleminin yeni formu şu şekilde ifade edilir:

$$L_p = \frac{1}{2}\|w\|^2 + C\sum_i \xi_i - \sum_i \alpha_i\{y_i(w^T x_i + b) - 1 + \xi_i\} - \sum_i \mu_i \xi_i \qquad (3.17)$$

Yukarıdaki formülasyonun çözümü için KKT şartları uygulanırsa;

$$L_d = \sum_i \alpha_i - \frac{1}{2}\sum_{i,j} x_i^T x_j y_i y_j \alpha_i \alpha_j \qquad (3.18)$$

$$0 \le \alpha_i \le C, \forall i \qquad (3.19)$$

elde edilir. Yukarıdaki formülasyonda $0 < \alpha_i < C$ aralığında yer alan Lagrange çarpanlarına karşılık gelen x_i değerleri destek vektörleridir.

3.2.2. Lineer Olmayan Destek Vektör Makineleri

Lineer olarak ayrılabilen sınıflar hipotezi pratik uygulamalar için yetersizdir. Bu sebepten ötürü örneklerin yüksek boyutlu ve doğrusal olarak ayrılabilecekleri uzaya taşınması ve çözümün bu yüksek boyutlu uzayda aranması için, uzaylar arası dönüşüm yapan çekirdek fonksiyonlarının kullanılması gerekir. Destek Vektör Makineleri $K(x_i, x_j) = \varphi(x) \cdot \varphi(x_j)$ şeklinde ifade edilen Kernel fonksiyonu kullanarak doğrusal olmayan dönüşümler yapabilmekte ve bu sayede yüksek boyutta doğrusal olarak ayrımına imkan vermektedir. Burada K çekirdek fonksiyonunu, x_i ve x_j ise uzaydaki verileri temsil etmektedir [5]. Literatürde çok farklı çekirdek fonksiyonları bulunmaktadır. Bu fonksiyonlar içinde en fazla kullanılanlar ise Pearson VII (PUK), Radyal Tabanlı Fonksiyon, Polinom ve Normalleştirilmiş Polinom fonksiyonlarına ilişkin formül ve parametreler aşağıdaki tabloda verilmiştir.

Tablo 3.2. Destek vektör makineleri kernel fonksiyonlar

Kernel Fonksiyonu	Matematiksel İfadesi	Parametre
Radyal Tabanlı Fonksiyon Kerneli	$K(x,y)=e^{-\gamma\|(x-x_i)\|^2}$	Kernel boyutu (γ)
Pearson VII (PUK) Kerneli	$\frac{1}{\left[1+\left(\frac{2\cdot\sqrt{\|x-y\|^2\sqrt{2^{(1/\omega)}}-1}}{\sigma}\right)\right]^{\omega}}$	Pearson genişliği parametreleri (σ , ω)
Normalleştirilmiş Polinom Kerneli	$K(x,y)=\frac{((x\cdot y)+1)^d}{\sqrt{((x^2+1)^d)((y^2+1)^d)}}$	Polinom derecesi (d)
Polinom Kerneli	$K(x,y)=((x\cdot y)+1)^d$	Polinom derecesi (d)
Lineer Kernel	$K(x,y)=(x\cdot y)$	Genel Kernel İfadesi

3.2.3. Çok Sınıflı Destek Vektör Makineleri

Buraya kadar bahsedilen konular, sınıf sayısının iki olması durumunda geçerli olan işlem adımlarıdır. Sınıf sayısının ikiden fazla olması durumunda farklı yaklaşımlar söz konusudur.

Destek Vektör Makineleri'nde çok sınıf probleminin çözümü için iki yöntem kullanılmaktadır. Bunlardan ilki Lagrange fonksiyonunun çok sınıfla işlem yapabilecek hale getirilmesidir. Yalnız bu yöntem, sınıf sayısının artması ile hata miktarının da artması nedeniyle tercih edilen bir yöntem değildir.

Diğer bir yaklaşım ise, destek vektör makinesinin ikili sınıflar halinde sınıflandırmalar yapacak biçimde çalıştırılmasını sağlamaktır. Bu yaklaşımda bire karşı bir, bire karşı hepsi ve döngüsel olmayan graf gibi farklı yöntemler kullanılmaktadır [12].

Bire karşı bir yaklaşımında her örnek veri seti, diğer örnek veri setleri ile ayrı ayrı eğitilir. Yani K adet sınıf olması durumunda, K·(K-1)/2 eğitim işlemi yapılır. Bire karşı hepsi yönteminde, her bir örnek veri seti bir sınıf, geriye kalan sınıfların tamamı bir sınıf olarak kabul edilir ve eğitim bu şekilde yapılır. Böylelikle K adet sınıf olması durumunda K adet eğitim yapılmış olur.

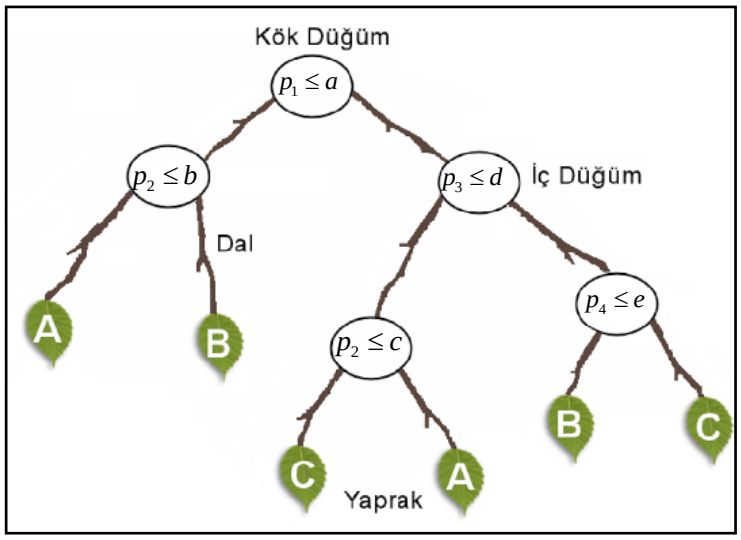
Döngüsel olmayan graf yönteminde, test aşaması dışında bire karşı bir yöntemiyle aynıdır. Eğitim bire karşı bir yöntemine göre yapılır. Test aşamasında ise, eğitim örnekleri kontrol edilmek yerine, sınıflandırılacak elemanın ait olmadığı düşünülen sınıflar elenerek işlem yapılır [12].

3.3. Karar Ağaçları

Karar Ağaçları yöntemi son yıllarda radar sinyallerinin sınıflandırılmasında, karakter tanıma, uzaktan algılama, uzman sistemler gibi çok geniş bir yelpazede başarılı biçimde kullanılan bir algoritmadır [16].Bu yöntemin yaygın olmasının en önemli nedenleri kolay yorumlanabilir ve sade olmasıdır [17].

Karar ağaçları eğitim verilerine bakarak test verilerinin hangi sınıfa ait olduğuna kurallar çıkararak karar vermektedir. Karar ağaçlarının temelinde sorulan sorular ve bunlara dair cevapların birleştirilmesi yatmaktadır. Bu sayede oluşan ağaç "eğer-ise " şartlarından oluşan kurallar bütünü halini alır [18].

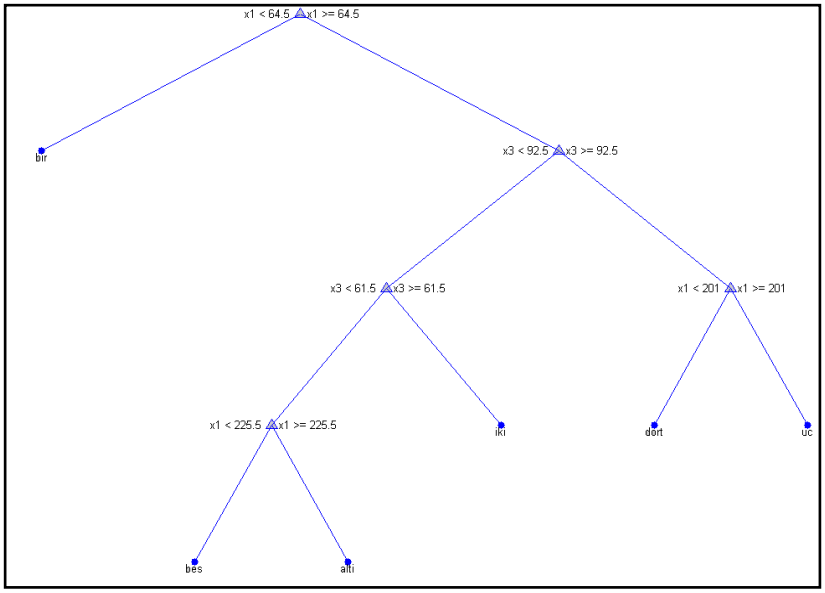
Bir karar ağacı, düğüm, dal ve yaprak olarak adlandırılan üç temel kısımdan oluşur. Her bir nitelik bir düğüm tarafından temsil edilmektedir. Burada düğüm tarafından temsil edilmesi gereken öznitelik uydu görüntüleri için bant değeridir.



Şekil 3.8 Örnek bir karar ağacı yapısı

Veri setinin sınıflandırılması ve ağaç yapısının oluşturulması için sorular sorulmaya kök düğümden başlanır. Eğer yeni sorular sorulamıyorsa dallanma bitmiştir ve bir sınıfı temsil eden yaprağa ulaşılmıştır demektir. Şekil 3.8'de üç sınıf için dört boyutlu öznitelik değerlerinden oluşan bir karar ağacı yapısı gösterilmektedir. Şekilde p_i öznitelik değerlerini yani bant değerlerini; a,b,c,d,e terimleri dallanma için eşik değerlerini, A, B, C ise sınıfları ifade etmektedir. Karar ağaçları oluşturulurken, kullanılan algoritmaya göre ağacın şekli değişebileceğinden, yani farklı sınıflandırma sonuçları elde edilebileceğinden kullanılacak algoritma önem taşımaktadır [19]. Örneğin kullanacağımız yöntem CART ise, dallanma işleminde Gini indeksi ya da Twoing algoritmasının kullanımı farklı ağaç yapıları oluşmasına neden olabilmektedir.

Karar ağaçlarının oluşturulmasının ardından, bir test veri setinin sınıflandırılması işlemine kök düğümden başlayarak koşullar uygulanır ve oluşan tüm sonuçlar için ona ait uygun dal takip edilir. Buradan ya yeni koşulun uygulanacağı bir iç düğüme ya da yaprak düğüme ulaşılır ve böylelikle test verilerinin hangi sınıfa ait olduğu, hangi yaprakta sonlandığına bakılarak belirlenmiş olur. Şekil 3.9'da basit bir görüntü sınıflandırma işlemi için elde edilen karar ağacı modeli görülmektedir.



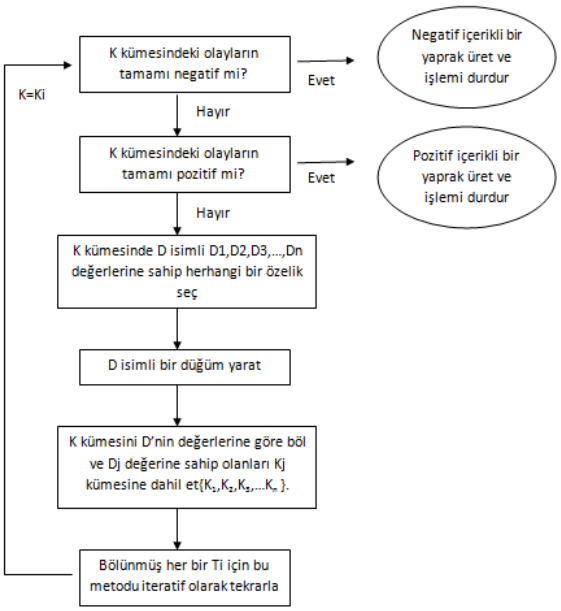
Şekil 3.9. Basit bir karar ağacı modeli

3.3.1 Karar Ağaçları Oluşturma

Sahip olduğumuz verilerden birçok farklı karar ağacı oluşturabilmemize rağmen amaç en optimum karar ağacını oluşturmaktır. Bu her zaman mümkün olmasa da yüksek bir doğruluk derecesine sahip ağacı oluşturacak birçok etkin algoritma geliştirilmiştir. Bu algoritmaların temelinde Greedy yaklaşımı bulunmaktadır. Bu yaklaşıma göre, verinin dallara ayrılabilmesi için hangi değişkenden başlanması gerektiği ve ardından oluşan alt veri grubunu tekrar alt dallara ayırmak için hangi değişkenden başlanması gerektiği ile ilgili lokal kararlar alınır. Bu algoritmaların birçoğu Hunt algortiması olarak bilinen Böl ve Elde Et (divide and conquer) yönteminin farklı biçimlerini kullanmaktadır [20].

3.3.2. Böl ve Elde Et

Bu metot Hunt tarafından kullanılan, en temel ağaç yaratma algoritmasıdır. Bu yöntemin geliştirilmesine yönelik birçok çalışma yapılmıştır. Bunlardan en önemlileri, tek değişkenli karar ağaçları için Quinlan tarafından geliştirilen ID3 ve C4.5 ve çok değişkenli karar verme için Breiman'ın geliştirdiği CART algoritmalarıdır. Bu yöntemde K örnek uzay, sınıflar ise negatif ve pozitif olsun. Bu durumda Karar Ağacının oluşturulması Şekil 3.10'daki gibi olacaktır.



Şekil 3.10 Hunt Algortiması [21].

3.3.3. Karar Ağaçlarında Dallanma Kriterleri

Karar ağacı yönteminin kullanışlı, basit ve yapısı gereği çok hızlı olmasına karşın en önemli problemlerinden biri bir kökten itibaren bölünme ya da başka bir değişle dallanmanın nasıl yapılacağının bilinmemesidir. Algoritmaları şu şekilde sıralayabiliriz:

1. Entropiye Dayalı Algoritmalar,
2. Sınıflandırma ve Regresyon Ağaçları (CART),
3. Bellek Tabanlı Sınıflama Algoritmaları .

3.3.3.1. Entropiye Dayalı Algoritmalar

Entropi, belirsizliğin ölçülmesi için kullanılan bir ölçüt olup, bilgi ile doğrudan ilişkilidir. Entropi değerinin bire yaklaşması belirsizliğin arttığının göstergesidir ve bu durumda daha fazla bilgiye ihtiyaç duyulur. Entropinin sıfıra yaklaşması ise belirsizliğin azaldığının göstergesidir. C4.5, ID3 ve CART entropi temelli yöntemlerdir.

Veri setinin K olduğunu düşünelim. Bu veri seti $\{k_1,k_2,k_{3,..,}k_n\}$ olmak üzere n adet elemandan oluşsun. Burada k_i elemanlarının sınıf olasılıkları p_i'dir. $P=\{p_1,p_2,..,p_n\}$ olasılık dağılıma sahip veri setinin entropisi şu şekildedir [13]:

$$H(K) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3.20}$$

Karar ağaçları kurulurken, kısıtlı bir eğitim kümesinden yararlanarak olası tüm ağaçların yapılarını ortaya çıkarmak ve en uygun olanını seçerek dallanmaya hangi nitelik ile başlanacağının tespiti büyük önem taşımasına rağmen kolay bir işlem değildir. Örneğin 5 nitelik ve 20 elemanlı bir eğitim kümesi için çizilecek karar ağacı sayısı 10^6'dan daha fazladır [22]. Bu problemi aşabilmek için algoritmaların büyük bir kısmının başlangıçta birtakım değerleri hesaplanarak ona göre ağaç oluştururlar. Bu nedenle dallanma işlemi entropi değerine göre gerçekleştirilebilir.

Entropinin, karar ağaçlarında kullanımını irdeleyecek olursak; bir eğitim kümesinde sınıf niteliğinin alacağı değerlere göre $\{S_1,S_2.,..,S_n\}$ olmak üzere n adet sınıfa ayrıldığını düşünelim. Bu sınıflarla ilgili olarak ortalama bilgi miktarına ihtiyaç duyulabilir.

Burada T sınıf değerlerini içeren küme için P_T sınıfların olasılık dağılımları şu şekilde hesaplanır [23]:

$$P_T = \left(\frac{|S_1|}{|T|}, \frac{|S_2|}{|T|}, .., \frac{|S_n|}{|T|} \right) \tag{3.21}$$

Burada, $P_n = \frac{|S_n|}{|T|}$ olasılığını ifade etmektedir. O halde $|S_1|$ terimi, S_1 kümesine ait elemanların sayısını vermektedir. T kümesi için entropi:

$$H(T) = -\sum_{i=1}^{n} p_i \log_2(p_i) \tag{3.22}$$

ile tanımlanır.

3.3.3.2. Bilgi Kazancı ve Kazanç Oranı

Bilgi kazancı entropiye dayalı algoritmalarda(ID3, C4.5, CART) eğitim veri setindeki değişkenin etkinlinin ölçütü olarak tanımlanabilir. Bölünme işlemine, bilgi kazancının en yüksek olduğu değişkenden başlanır. Bilgi kazancının hesaplanmasında ilk olarak veri kümesinin bölünmeden önceki halinin entropisi hesaplanır ardından tüm alt bölümlerin entropileri bulunur. Son işlem olarak hesaplanan bu iki değer arasındaki farkın en yüksek olduğu değişken en iyi dallanma kriteri olarak seçilir. Bilgi kazancı şu şekilde bulunur [13]:

Eğer veriseti, X n adet alt bölüme T değişkeninden bölünecekse bilgi kazancı;

$$H(X,T) = \sum_{k=1}^{n} \frac{|T_i|}{|T|} H(T_i) \tag{3.23}$$

$$Kazanç(X,T) = H(T) - H(X,T) \tag{3.24}$$

Burada $Kazanç(X,T)$ değerinin en yüksek olması hedeflenir ve kazancı maksimum edebilecek X değeri seçilir.

Dallanma işlemi için bilgi kazancının dışında, daha iyi sonuçlar veren Bölünme Bilgisi (Split Information) isimli Quinlan [23] tarafından ortaya atılmış yöntemde

kullanılmaktadır [13]. Bir D kümesinin, X niteliğinin değerini belirlemek, bilgi miktarını ortaya koymak için bu yol tercih edilmektedir. Aşağıdaki eşitlikte bilgi miktarı $H(P_{X,D})$ ile değerlerin olasılık dağılımı ise $P_{X,D}$ ile ifade edilmektedir:

$$P_{X,D} = \left(\frac{|D_1|}{|D|}, \frac{|D_2|}{|D|}, ..., \frac{|D_n|}{|D|} \right) \tag{3.25}$$

Burada $H(P_{X,D})$ terimi D kümesindeki X niteliği için Bilgi Bölünmesi olarak isimlendirilir ve şu şekilde hesaplanır:

$$H(P_{X,D}) = H\left(\frac{|D_1|}{|D|}, \frac{|D_2|}{|D|}, ..., \frac{|D_n|}{|D|} \right) \tag{3.26}$$

Ayırma bilgisi olarak da isimlendirilen terim şu şekilde de bulunabilir:

$$H(P_{X,D}) = \sum_{i=1}^{n} \frac{|D_i|}{|D|} \log_2 \left(\frac{|D_i|}{|D|} \right) \tag{3.27}$$

Kazanç oranı ise ayırma bilgisi ve kazanç ölçütü yardımıyla şu şekilde hesaplanır:

$$KazançOranı(X,D) = \frac{Kazanç(X,D)}{H(P_{X,D})} \tag{3.28}$$

3.3.4. Karar Ağaçlarının Budanması

Karar ağaçlarının üretilmesinin ardından çok geniş ve karmaşık bir görünüm ortaya çıkabilir. Eğitim verisindeki gürültü nedeniyle bir karar ağacının yaprakları tek bir sınıf verisini içerecek biçimde aşırı öğrenme (over fitting) problemiyle karşılaşılabilir. Bu halde, oluşturulan karar ağacı eğitim veri seti için mükemmele yakın sonuçlar üretirken, test veri seti için çok düşük doğruluk değeri üretebilir. Aşırı öğrenme probleminin en önemli sebepleri eğitim verisinin çok gürültülü olması ve modelin çok fazla detayı karakterize etmesidir [18]. Bu problemi çözmek için, sınıflandırma doğruluğunu etkileyemeyen kısımların çıkarılması yani bir alt ağacın çıkarılıp yerine bir yaprak yerleştirilmesi gerekir. Bu işleme karar ağacının budanması denmektedir. Bu sayede öngörülü hata oranı azaltılması ve model kalitesinin arttırılması amaçlanır. Karar

ağaçlarının basitleştirilmesi ve model kalitesinin arttırılması için iki yaklaşım söz konusudur. Bu yöntemler:

1. Eğitim veri setinin daha fazla bölünüp bölünmeyeceğine karar vermek
2. Karar ağacının kurulmasının ardından ağacın bazı bölümlerinin geriye dönük biçimde kaldırılması

İlk yöntem ön budama olarak da adlandırılır. Bölünme işlemine durdurma kriteri olarak χ^2 gibi testler yapılarak karar verilir. Bölünme işleminin sonrası ve öncesinde belirlenen limit kadar bir fark yoksa söz konusu düğüm yaprak olarak gösterilir. Bu yöntemin dezavantajı ise eşik değerin ne olacağıdır. Limit değerin çok yüksek seçilmesi durumunda, bölünme azalacağından, daha başarılı sonuçlar verecek alt bölümlere ayrılmadan işlem tamamlanacaktır [18].

İkinci yöntemde ise, yapraklar sadece bir sınıfa ait veriyi içerdiği ana kadar büyümesine izin verilir ve ağaç bundan sonra budanır. Budama işlemi seçilen bir doğruluk(hata) ölçütüne göre yapılır. Bu yöntemin dezavantajı ise, ağacın sonradan çıkarılan kısımlarının oluşturulmasında çok fazla hesaplama yapılmasının gerekmesidir [18].

3.3.5. Sınıflandırma ve Regresyon Ağaçları (CART)

Breiman tarafından 1984 yılında ortaya atılan bu yöntem veri madenciliğinin sınıflandırma ile ilgili en önemli konularından biridir. İsminden de anlaşıldığı gibi temelinde yine karar ağaçları vardır. CART analizi eğitim verisindeki cevap değişkeninin kategorik olması durumunda sınıflandırma ağacı (Classification Tree), sürekli olması durumunda ise regresyon ağacı (Reggresion Tree) ismini alır ancak uygulamalarda çoğunlukla CART ya da C&RT olarak kullanılmaktadır.

Karar ağaçları temelli bir yöntem olan CART, çok hızlı ve adaptif olmasından dolayı tıp, endüstri ve mühendislik disiplinlerinden birçoğunda sıklıkla kullanılmaktadır. Sınıflandırma ağaçlarında diğer karar ağaçlarından farklı olarak yalnızca ikili dallanma söz konusudur. Yani düğümlerden sadece iki dal elde edilebilir. Bu yöntemde genel olarak iki algoritma kullanılmaktadır:

1. Gini İndeksi Algoritması
2. Twoing Algoritması

3.3.5.1. Gini İndeksi Algoritması

Gini indeksi, sınıflara ilişkin özniteliklerin safsızlığının ölçüldüğü bir yöntemdir. Bu algoritmada sağ ve sol olmak üzere ikili dallanma söz konusudur. Gini indeksi çok hızlı biçimde hesaplanabilir. Yöntemin işlem adımları şu şekildedir:

1. Tüm nitelik değerleri iki gruba ayrılır. Üretilen sağ ve sol bölünmelere denk gelen sınıf değerleri gruplandırılır.
2. Her bir nitelikle ilgili sağ ve sol Gini indeks değerleri hesaplanır.

$$G_{sağ} = 1 - \sum_{i=1}^{n} \left(\frac{N_i}{|D_{sağ}|} \right)^2 \tag{3.29}$$

$$G_{sol} = \sum_{i=1}^{n} \left(\frac{M_i}{|D_{sol}|} \right)^2 \tag{3.30}$$

Formülasyonda yer alan terimler şu şekildedir:

n Sınıfların sayısı

D Bir düğümdeki örnek sayısı

$|D_{sağ}|$ Sağ taraftaki örneklerin sayısı

$|D_{sol}|$ Sol taraftaki örneklerin sayısı

M_i Sol taraftaki i kategorisindeki örneklerin sayısı

N_i Sağ taraftaki i kategorisindeki örneklerin sayısı

3. Her i niteliği için k eğitim veri setindeki satır sayısı olmak üzere şu bağıntı hesaplanır:

$$Gini_i = \frac{1}{k} \left(|D_{sol}| G_{sol} + |D_{sağ}| G_{sağ} \right) \tag{3.31}$$

4. Her bir i niteliği için hesaplanan $Gini_i$ değerleri arasından en küçük olanı seçilir ve bölünme işlemine bu nitelik üzerinden devam edilir [13].

5. İlk adıma tekrar dönülür ve işlemlere devam edilir [13].

3.3.5.2. Twoing Kuralı

Bu kural 1984 yılında Brieman tarafından tarafından kullanılmış ve karar ağacının oluşturulmasında öznitelik seçimi için farklı yaklaşım izlemektedir. Algoritma şu şekilde tanımlanır [13]:

1. Nitelik değerlerine bakılarak eğitim veri seti iki kümeye ayrılır. Herhangi bir T düğümümü T_{sol} ve $T_{sağ}$ olmak üzere iki ayrı dal oluşur.
2. Tüm aday bölünmeler için P_{sol} ve $P(i \mid T_{sol})$ olasılıkları hesaplanır. Burada $P(i \mid T_{sol})$ ifadesi i sınıf değerinin soldaki bölünmede bulunma ihtimalidir.

$$P_{sol=} \frac{T_{sol}\text{'daki herbir nitelik değerinin } i \text{ ilgili sütundaki tekrar sayısı}}{\text{Eğitim verisetindeki kayıtların sayısı}}$$

$$P(i \mid T_{sol}) = \frac{T_{sol}\text{'daki kayıtların } i \text{ sınıfları sayısı}}{T_{sol}\text{'daki herbir nitelik sütunundaki tekrar sayısı}}$$

3. Tüm aday bölünmeler için $P_{sağ}$ ve $P(i \mid T_{sağ})$ olasılıkları hesaplanır. Burada $P(i \mid T_{sağ})$ ifadesi i sınıf değerinin soldaki bölünmede bulunma ihtimalidir.

$$P_{sağ=} \frac{T_{sağ}\text{'daki herbir nitelik değerinin } i \text{ ilgili sütundaki tekrar sayısı}}{\text{Eğitim verisetindeki kayıtların sayısı}}$$

$$P(i \mid T_{sağ}) = \frac{T_{sağ}\text{'daki kayıtların } i \text{ sınıfları sayısı}}{T_{sağ}\text{'daki herbir nitelik sütunundaki tekrar sayısı}}$$

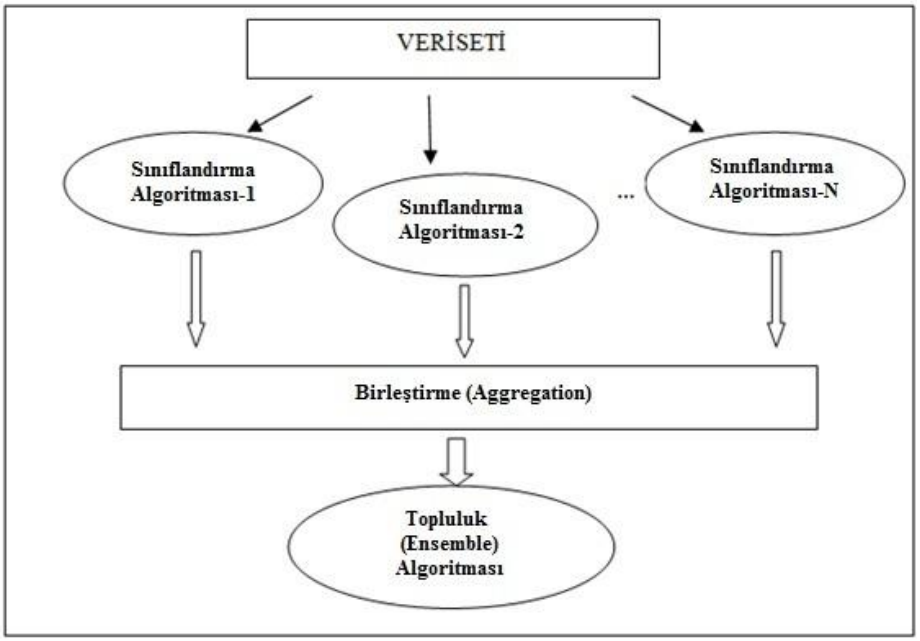
4. $\Phi(s \mid T), T$ düğümündeki s aday bölünmelerinin uygunluk ölçüsü olduğunu düşünürsek, uygunluk ölçütü şu şekilde hesaplanır [13]:

$$\Phi(s \mid T) = 2P_{sol}P_{sağ}\sum_{i=1}^{n}\left|P(i \mid T_{sol}) - P(i \mid T_{sağ})\right| \qquad (3.32)$$

5. $\Phi(s|T)$ değerlerinden en büyüğü seçilir ve bu seçim bize dallanmanın yapılacağı satırı belirleyecektir.
6. Dallanma işlemi bu şekilde yapılır ve bu adıma ait karar ağacı çizilir.
7. İlk adıma dönülür ve ağacın alt kümesine de aynı işlemler yapılır [13,24].

3.3.6. Ağaç Tabanlı Topluluk Yöntemleri

Topluluk öğrenme tekniklerinde (Ensemble Learning) birden fazla sınıflayıcının ürettiği sonuçların bir araya getirilerek, topluluk adına tek bir sonucun üretildiği bir yöntemdir. Bu yöntemlerde birden fazla sayıdaki sınıflayıcının ürettiği tahminler oylamaya tabi tutulur ve en çok oyu alan sınıf topluluğun sınıf tahmini olarak verilir. Topluluk yöntemleri çok sayıdaki sınıflandırıcının sonucunu birleştirdiği için daha güvenilir sonuçlar vermektedir. Yapılan araştırmalarda birleştirme yöntemi ve birleştirmede kullanılan sınıflandırma yöntemlerinin hatalarının da azaldığı gözlemlenmiştir [18].



Şekil 3.11. Topluluk (Ensemble) yöntemi

Topluluk yöntemleri tekil öğrenme algoritmaların doğruluğunu arttırmaktadır. Boosting ve Bagging üzerinde en fazla çalışılan yöntemler olmuşlardır [19]. Breiman tarafından önerilen Random Forest yöntemi Bagging yöntemine göre daha fazla rastgelelik

sağlamaktadır. Bu yöntemde tüm karar ağaçları bootstrap örnekleme ile oluşturulmaktadır.

Tüm öğrenme metotlarında amaç en az örnekle çok fazla sayıda verinin sınıflandırmasını sağlamaktadır. İdeal olan sınıflandırıcılardan beklenen en az zamanda ve en az öğrenme verisi ile tahmin yapabilmektir [18].

3.3.6.1. Rastgele Orman (Random Forest)

Rastgele orman önceki konularda bahsedildiği gibi bir topluluk (ensemble) öğrenme algoritmasıdır. Bu yöntemin temelinde yine karar ağaçları vardır. Tekil olarak oluşturulan karar ağaçları birleşerek karar ormanını oluştururlar ve ağaçların yapmış olduğu tahminler bir araya getirilip nihai tahmin yapılır.

Rastgele orman yöntemi, Torbalama (Bagging) ve Random Subspace tekniklerinin birleştirilmesi ile elde edilmiştir. Torbalama yöntemi ile girdi verilerinden bağımsız bootstrap tekniği ile örneklem oluşturulur. Random Subspace yönteminde ise karar ağacına ait her bir düğümde en iyi dallara ayrılan değişken, tüm değişkenler arasından rastgele seçilen az sayıdaki değişkenin içinden seçilir [18].

Diğer tüm karar ağaçlarında olduğu gibi RF yönteminde de dallanma kriterlerinin belirlenmesi ve uygun bir budama yönteminin seçilmesi önem taşımaktadır. Rastgele orman sınıflandırıcının dallanma kriterlerinin kullanılmasında Gini indeksi yöntemi kullanılmaktadır. Gini indeksi işin temelinde sınıf özniteliklerinin zayıflık derecesini ölçmektedir [26]. Başka sınıflandırma metotlarında olduğu gibi RF yönteminde de kullanıcı tarafından belirlenmesi gereken parametreler vardır. Bu parametre değerleri, ağaç yapısının oluşturulması için her bir düğümde kullanılacak örneklerin sayısı ve oluşturulacak ağaç sayısıdır. Yani bir sınıflandırma işlemi sırasında karar ormanı, kullanıcı tarafından belirlenen K adet ağaçtan oluşturulur. Yeni bir sınıflandırma işleminde ise veri seti bu K adet karar ağacından geçirilir ve Rastgele Orman sınıflandırıcısı ile K adet ağaçtan elde edilen K adet oy arasından en fazla oya sahip olan sınıf seçilir [5,25].

RF yönteminde, üretilen modeli test etmek için ayrı bir veri seti yoksa ya da orijinal veri setinden test veri seti ayrılmamış ise, sınıf dağılımı dikkate alınmak suretiyle orijinal

veri setinin 2/3'ü öğrenme verisi (inBag), 1/3'ü ise test verisi (Out-Of-Bag(OOB)) olarak ayrılır. Eğer test için ayrı bir veri seti varsa ya da orijinal veri setinden test için veri ayrımlı ise eğitim verilerinin 2/3 oranında öğrenme veriseti (inBag), 1/3'ü ise test veriseti (OOB) olarak ayrılır.

Bu yöntemde karar ormanını oluşturulmasında kaç karar ağacı kullanılacaksa, o sayıda bootstrap tekniği kullanılarak örneklem oluşturulur ve her bir örneklem için inBag ve OOB verisi ayrılır. Tüm ağaçlar ayrılan OOB verisi ile test edilerek hata oranı hesaplanır ve ardından bu hata oranlarının ortalaması alınarak karar ormanının OOB hatası hesaplanır [18].

Rastgele orman yönteminde en önemli etkenler bootstrap tekniği ile seçilen örneklem ve her düğümde tüm değişkenlerden belirlenen sayıda değişkenin seçilmesidir. Rastgele seçilen bu değişkenler içinde en iyi bölünme sağlayacak olandan dallara ayrılmaya başlanır. Ek olarak bu şekilde üretilen karar ağacında budama yapılması söz konusu değildir. Rastgele orman sınıflandırıcısı karar ağacı oluşumunda CART yöntemini kullanmaktadır.

Karar ağaçlarının kurulmasının ardından, test verisindeki her bir piksele ait olan veriler satır satır önceden kurulmuş ağaçlar üzerine yerleştirilir ve bu işlem tüm pikseller için uygulanır. Karar ağacı her bir pikseli sınıflardan birine yerleştirir. Hesaplanan OOB hata oranına göre tüm ağaçlara bir ağırlık verilir. Hata oranı en yüksek olan karar ağacı en düşük ağırlığı, hata oranı en düşük olan karar ağacı ise en yüksek ağırlığı alır. Belirlenen ağırlıklara göre tüm ağaçlar sınıflandırma işlemi için bir oylama işlemine tabi tutulurlar. En çok oyu alan sınıf nihai sınıf tahmini olarak belirlenmiş olur. Her bir piksele ait olan değerler satır satır yerleştirildiğinde orman karar vermiş olur. Bu bilgilere ek olarak tüm ağaçlarda kullanılan veri ve değişkenler farklı olduğundan Rastgele Orman yöntemi verideki gürültüye ve aşırı öğrenmeye karşı güçlüdür [18].

3.3.6.1.1. Rastgele Orman Modelinin Kurulması

- Orijinal veri setinden eğitim ve test verileri ayrılmasının ardından bootstrap yöntemi ile örneklemler seçilir.

- Kullanıcı tarafından belirlenen K adet değişkenden, dallara ayrılma işlemine bilgi kazancına bakılarak karar verilir. Dallara ayrılacak değişken belirlendikten sonra, aşağı yönlü iki dal oluşturulur.
- Dallanma işlemine hangi değere göre ayrılacağına Gini indeksi kullanılarak kara verilir ve bu işlem yeni bir dal oluşturulmasına gerek kalmayana kadar devam edilir.
- Modele ait OOB hata oranı hesaplanır. Bu sayede her bir ağaca ağırlık verilmesi sağlanır.
- Sınıflama işlemi sırasında tüm ağaçlar oluşturduğu sınıflardan birine OBB ile hesaplanan ağırlıklı bir oy verir. Karar ormanı verideki her bir piksele ait, tüm ağaçların yaptığı tahminlere göre en çok oy alınan sınıf nihai sınıf olarak seçilir. Örneğin ilk sıradaki piksel için, tüm karar ağaçlarındaki ait olduğu sınıflar belirlenir. Ardından ağaçların ağırlıkları hesaplanır ve her bir sınıf için ağaçların ağırlığına göre verdiği oyların toplamı ayrı ayır belirlenir. Piksel ağırlık toplamı en büyük olan sınıfa atanır [18].

3.3.6.1.2. Hata Oranının Tahmin Edilmesi

Kullanılan yöntem ne olursa olsun tüm sınıflandırma işlemlerinde amaç işlemi yüksek doğrulukla yapabilmektir. Temel hata ölçümü hata matrisinin oluşturulmasıyla hesaplanır. Rastgele orman yönteminde temel olarak iki tür hata tahmini yapılır.

3.3.6.1.3. Hold Out Yöntemi

Bu yöntemde hata oranının hesaplanması orijinal veri seti eğitim ve test verileri olmak üzere ikiye ayrılır. Genelde verinin 2/3'ü öğrenme 1/3'ü test verisi olarak ayrılırlar.

3.3.6.1.4. Çapraz Doğrulama Yöntemi

Bu yöntemde, veri grubu eşit büyüklükte n sayıda alt veri grubuna bölünür. Bu veri gruplarından bir tanesi modeli test etmek için, geriye kalan n-1 sayıdaki veri grubu ise eğitim verisi olarak kullanılır. Her seferinde n adet veri grubundan bir tanesi test verisi

olarak seçilir ve bu işlem n defa tekrarlanır. En son adımın ardından hata oranı n adet hata değerinin ortalamasının hesaplanması ile elde edilir. Çapraz doğrulama tahmini eşitlikte verilmiştir:

$$CVA = \frac{1}{n}\sum_{i=1}^{n} Ai \tag{3.33}$$

Veri madenciliğinde genelde n değeri 10 olarak alınır [18].

3.3.7. En Çok Olabilirlik Yöntemi

Literatürde en çok kullanılan sınıflandırma algoritmalarından olan EÇO yöntemi verilere ait bant histogramlarının normal dağılıma uyduğunu varsayar. Bu algoritmanın temelinde piksellerin sınıflara ait olma olasılıklarının hesaplanması ve olasılığın en yüksek olduğu sınıfa atanması vardır. Bu sebepten ötürü varyans, kovaryans ve ortalama değer gibi istatistiki değerler dikkate alınırlar [5,26].

Sınıfların ilk olasılıkları hakkında herhangi bir bilgi yoksa olasılıkların hepsi eşit olarak kabul edilir ve kullanıcı tarafından bir eşik değer belirlenerek pikselin belirlenen sınıflar dışında bir sınıfa ait olup olmadığı tespit edilir. EÇO yönteminde, olasılıkların hesaplanmasının ardından olasılıklar düzey eksende gösterilerek piksellerin sadece parlaklık değerleri değil aynı zamanda varyans ve kovaryans matris değerleri kullanıldığından, uzaydaki dağılımları da dikkate alınmış olur [26]. Eğer pikselin olasılık değeri kullanıcı tarafından belirlenen eşik değerinin altında ise piksel bilinmeyen etiket olarak tanımlanır. En Çok Olabilirlik Yöntemi'nin formülasyonu aşağıdaki şekildedir:

$$D = \ln(a_c - [0.5 * \ln(|COV_c|)] - [0.5 * (X - M_c)^T * (COV_c^{-1}) * (C - M_c)] \tag{3.34}$$

Bu eşitlikte:

- D: Olasılık değerini
- C: Örnek bir sınıfı
- X: Aday pikselin ölçüm vektörünü
- M_c :C örnek sınıfının ortalama vektörünü
- a_c: Aday pikselin C sınıfına ait olma yüzdesini
- COV_c: C örnek sınıfındaki piksellerin varyans-kovaryans matrisini göstermektedir [26].

3.3.8. Hızlandırma (Boosting) Yöntemi

Boosting herhangi bir öğrenme algoritmasının doğruluğunu arttıran genel bir metottur [28]. Hızlandırma (Boosting) yönteminin temelinde sınıflandırıcı serisinin oluşturulması vardır [27]. Zayıf ya da temel sınıflandırıcı olarak adlandırılan bu bireysel sınıflandırıcılar, Karar Ağaçları, Perceptron Öğrenme Kuralı, Maksimum Olabilirlik Kuralı gibi kurallardan biri olabilir. Her bir iterasyon sırasında, bir zayıf sınıflandırıcı seçilir ve sınıflandırılmamış vektörlere dayanan farklı bir örnek dağılımı kullanılarak eğitilir. Boosting yönteminin diğer yöntemlere göre önemli bir avantajı ise ince ayarlamanın, karmaşık ve lineer olmayan optimizasyon yapmanın gerekli olmamasıdır. Bu yöntemde seriye ait bir önceki sınıflandırıcıların hatalı olarak tahmin ettiği örnekler, bir sonraki sınıflandırıcı tarafından kullanılacak eğitim setindeki doğru tahmin edilen verilere göre daha fazla tekrar edilerek örnekleri daha doğru tahmin edebilen bir sınıflandırıcı oluşturulmaya çalışılır. Hızlandırma yönteminin temelinde her bir gözlem için hata durumuna göre bir ağırlık değerinin atanması vardır [5].

Literatürde hızlandırma yöntemi için kullanılan çok sayıda algoritma bulunmasına rağmen en yaygını Adaboost algoritmasıdır.

3.3.8.1. Adaboost Algoritması

Adaboost algoritması 1995 yılında Freund ve Schapire tarafından geliştirilmiş olan etkin bir topluluk öğrenme algoritmasıdır. Yöntem, bir grup zayıf sınıflandırıcıyı birleştirerek güçlü bir sınıflandırıcı elde etmeyi amaçlar.

Algoritma genel olarak şu şekilde tanımlanabilir [27]:

- $Tn = \{(X_1, Y_1), (X_2, Y_2), ..., (X_n, Y_n),\}$ eğitim seti ve $Y = \{1, 2, ..., k\}$ etiket değerler olmak üzere,
- Başlangıçta 1/n olmak üzere ağırlıklar her bir X_i değerine atanır,
- Aşağıdaki formülde $C_b(i)$ olarak not edilen ağırlıklar her adımda tekrar hesaplanır ve eğitim örneğine uygulanır. Sınıflandırıcı hatası olan ε_b şu şekilde hesaplanır [6]:

$$\varepsilon_b = \sum_{i=1}^{n} w_b(i)\,\xi_b(i) \qquad eğer \quad \xi_b(i) = \begin{cases} 0 & C_b(Xi) = y_i \\ 1 & C_b(Xi) \neq y_i \end{cases} \tag{3.35}$$

- Her b.ci iterasyonda ağırlıkların güncellenmesinde kullanılan α_b sabiti $\alpha_b = 1/2 \ln(1 - \varepsilon_b / \varepsilon_b)$ eşitliği ile hesaplanır. b+1.ci iterasyon sonunda yeni ağırlık şu şekilde hesaplanır:

$$w_{b+1}(i) = w_b(i) \exp(\alpha_b \xi_b(i)) \tag{3.36}$$

- Ardından hesaplanan ağırlıklar normalize edilir ve toplanır. Böylelikle yanlış sınıflandırılan değişkenlerin ağırlığı arttırılarak, sınıflandırıcının bir sonraki iterasyonda başarısız örneklere odaklanması sağlanır.
- Bu işlem iterasyon değeri kadar tekrarlanır. Son olarak topluluk sınıflandırıcı (2.36) nolu eşitlik ile hesaplanır:

$$C(x) = \arg_{y_j} \max \sum_{b=1}^{B} \alpha_b \delta(C_b(x), y_j) \tag{3.37}$$

Adaboost algoritması ilk olarak ikili (binary) sınıflandırma için tasarlanmış olsa da Adaboost.M1 ve Adaboost.M2 çok sınıflı problemlerin çözümü için tasarlanmışlardır.

Adaboost M1 Algoritması şu şekilde tanımlayabiliriz [6]:

- $i = 1, 2, ..., n$ ve $w_i = 1/n$ olmak üzere algoritma başlatılır.
- $b = 1, 2, ..., B$ kadar tekrarla
  a) $C_b(x) \in \{1, 2, ..., k\}$ sınıflandırıcısını w_i^b ağırlıklarını kullanarak yenilenir.
  b) $\varepsilon_b = \sum_{i=1}^{n} w_i^b \xi_b(i)$ ve $\alpha_b = 1/2 \ln(1 - \varepsilon_b / \varepsilon_b)$ hesaplanır.
  c) $w_i^{b+1} = w_i^b \exp(\alpha_b \xi_b(i))$ ağırlıkları hesaplanır ve normalize edilir.
- Son sınıflandırıcı hesaplanır:

$$C(x) = \arg\max_{y_j} \sum_{b=1}^{B} \alpha_b \delta(C_b(x), y_j) \tag{3.38}$$

4. BÖLÜM

UYGULAMA VE BULGULAR

4.1. Kullanılan Veri ve Yazılımlar

Bu tez çalışmasında "FBY-9-1133", numaralı ve "UYDU VERİLERİ İLE KENTSEL GELİŞİM ANALİZİ: KAYSERİ İLİ MELİKGAZİ BELEDİYESİ" isimli proje kapsamında temin edilen 2006 yılına ait Terra ASTER görüntüsünden üretilmiş olan görüntü ile http://glovis.usgs.gov internet adresi üzerinden indirdiğimiz 2004 yılına ait Landsat görüntüsü kullanılmıştır. Görüntülerin özellikleri Tablo 4.1'de gösterildiği şekildedir.

Tablo 4.1. Kullanılan uydu görüntüleri ve özellikleri

Görüntü	Boyut	Mekansal Çözünürlük	Spektral Bant Sayısı	Sınıf Sayısı
ASTER	185 x 216	30 m	6	7
LANDSAT	276 x 408	90 m	3	6



Şekil 4.1. Sınıflandırmada kullanılan Landsat görüntüsü



Şekil 4.2. Sınıflandırmada kullanılan Terra ASTER görüntüsü

Çalışmada sınıf örneklerinin seçimi için Doç. Dr. Coşkun Özkan'ın Matlab dilinde yazdığı kod, sınıflandırma işlemi için Java tabanlı açık kaynak kodlu Weka Kodları ve kodların çalıştırılması için ise Matlab programı kullanılmıştır.

4.2. Yapay Sinir Ağları İle Görüntü Sınıflandırma

Yapay sinir ağları, herhangi bir istatistiksel kabule gereksinim duymayan, parametrik olmayan kontrollü bir sınıflandırma yöntemidir. Bu öğrenme türünde, öğrenmeyi temsil eden bağlantı ağırlıkları, hedef değerler ile ağı ürettiği çıkış değerleri karşılaştırılarak üretilen hata değerinin minimizasyonu ile güncellenir. Günümüzde çok sık kullanılan bir yöntemdir.

Genel olarak YSA eğitimi şu aşamaları içermektedir:

1. Giriş ve çıkış değerlerinin belirlenmesi.
2. Yapay sinir ağı topolojisinin belirlenmesi (katman sayıları).
3. Ağın başlangıç değerlerinin rastgele atanması.
4. Momentum ve öğrenme oranı gibi ağ parametrelerinin belirlenmesi.
5. Çıktı kodlamasının belirlenmesi.

Bu tez çalışmasında YSA ile Terra ASTER ve LANDSAT uydularına ait görüntüler hem sadece YSA ve Adaboost.M1-YSA ile sınıflandırma işlemine tabi tutulmuştur. YSA yöntemine ilişkin parametreler aşağıdaki tabloda belirtilmiştir:

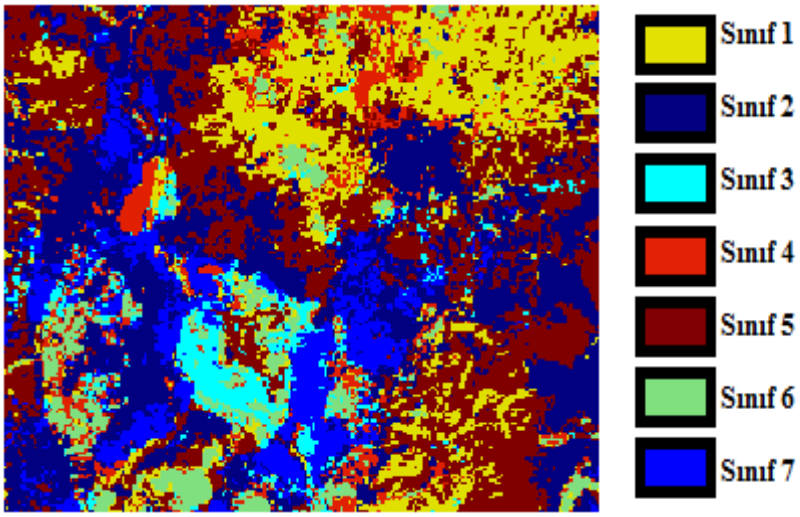
Tablo 4.2. YSA ile sınıflandırmada kullanılan parametre değerleri

Öğrenme Türü	Geri-Yayılım(MLP)
Öğrenme Oranı	0.3
Momentum	0.2
Epoch	2000

Bu parametreler kullanılarak Terra ASTER görüntüsünün YSA ile sınıflandırılmasından elde edilen sonuçlar Tablo 4.3'de verilmiştir.

Tablo 4.3. Terra ASTER görüntüsünün YSA ile sınıflandırılma doğrulukları (Epoch:2000)

GENEL DOĞRULUK	
Eğitim	**Test**
%90.4494	%81.9398

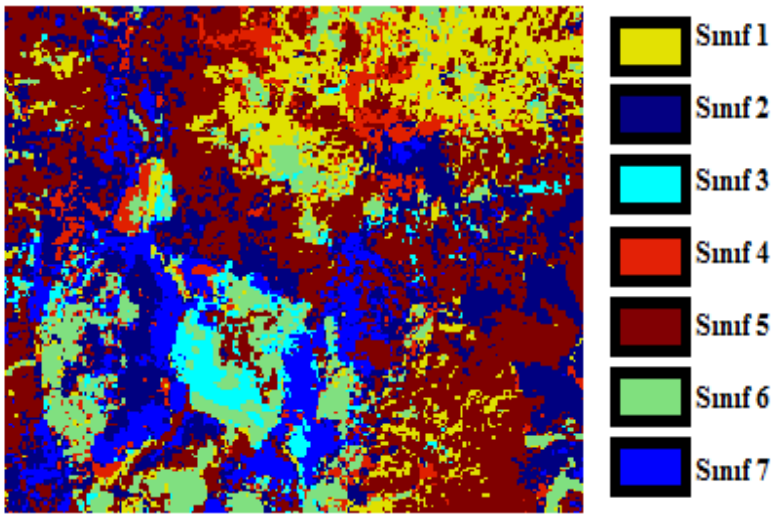


Şekil 4.3. Terra ASTER görüntüsünün YSA ile sınıflandırılması (Epoch:2000)

Momentum ve öğrenme oranları sabit kalmak şartıyla epoch sayısı 1000 olduğunda elde edilen sonuçlar Tablo 4.4'deki gibi olacaktır.

Tablo 4.4. Terra ASTER görüntüsünün YSA ile sınıflandırılma doğrulukları (Epoch:1000)

GENEL DOĞRULUK	
Eğitim	**Test**
%89.4663	%81.4381

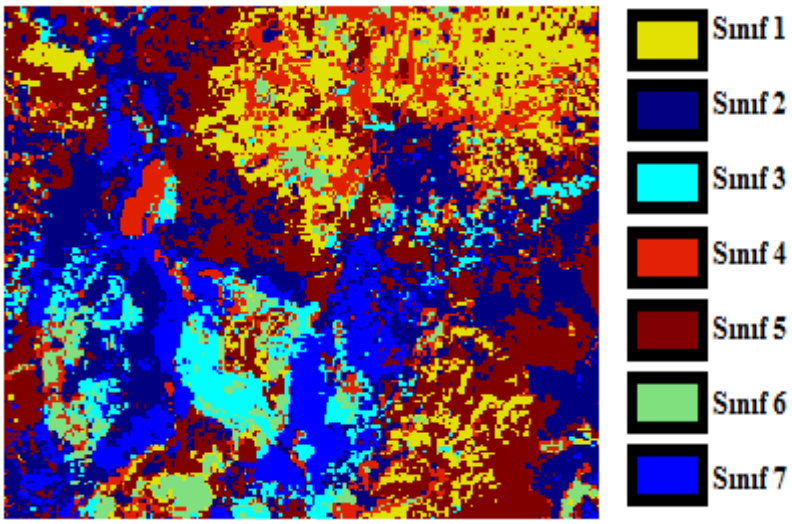


Şekil 4.4. Terra ASTER görüntüsünün YSA ile sınıflandırılması (Epoch:1000)

Aynı verinin çok sınıflı topluluk öğrenme biçimlerinden olan Adaboost.M1 destekli YSA ile sınıflandırılmasından elde edilen sonuç Tablo 4.5 ve tablo Tablo 4.6'da verilmiştir.

Tablo 4.5. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

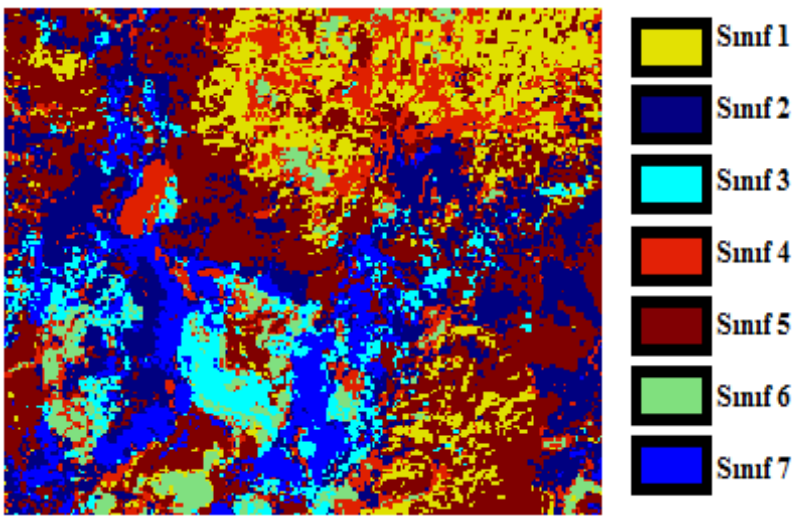
ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	20	**Eğitim**	**Test**
Örnekleme Oranı	0.5	%90.4494	%86.2876
YSA-Epoch Sayısı	2000		



Şekil 4.5. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:2000,İterasyon Sayısı:20,Örnekleme Oranı:0.5)

Tablo 4.6. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	20	**Eğitim**	**Test**
Örnekleme Oranı	0.5	%89.4663	%85.1171
YSA-Epoch Sayısı	1000		

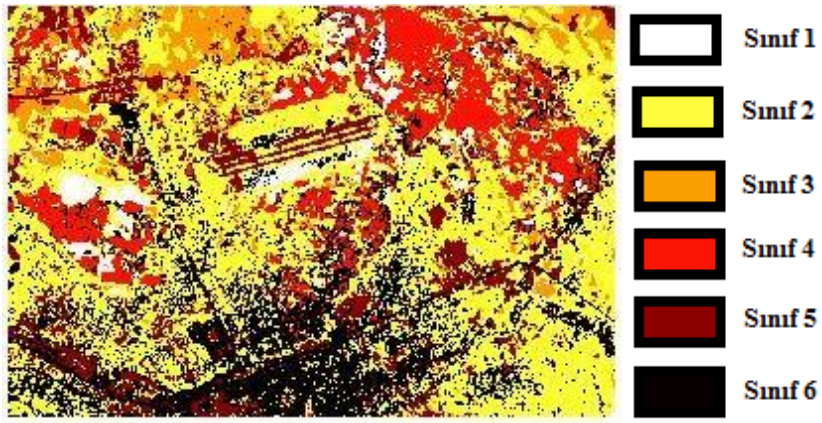


Şekil 4.6. Terra ASTER görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:1000,İterasyon Sayısı:20,Örnekleme Oranı:0.5)

Landsat uydusuna ait görüntüler günümüzde araştırmacılar tarafından sıklıkla kullanılan verilerdendir. Landsat uydusuna ait görüntünün YSA ile farklı epoch sayıları ile sınıflandırılmasına ilişkin sonuçlar Tablo 4.7 ve Tablo 4.8'de verilmiştir.

Tablo 4.7. Landsat görüntüsünün YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

YSA-PARAMETRELER		**GENEL DOĞRULUK**	
Öğrenme Oranı	0.3	**Eğitim**	**Test**
Momentum	0.2	%97.2532	%77.8266
Epoch	1000		



Şekil 4.7. Landsat görüntüsünün YSA ile sınıflandırılması (Epoch:1000)

Tablo 4.8. Landsat görüntüsünün YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

YSA-PARAMETRELER		**GENEL DOĞRULUK**	
Öğrenme Oranı	0.3	**Eğitim**	**Test**
Momentum	0.2	%97.2532	%78.2656
Epoch	2000		



Şekil 4.8. Landsat görüntüsünün YSA ile sınıflandırılması (Epoch:2000)

Landsat verisinin Adaboost.M1 destekli YSA yöntemi ile sınıflandırılmasına ilişkin sonuçlar Tablo 4.9 ve Tablo 4.10 de verilmiştir.

Tablo 4.9. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırma parametreleri(1)

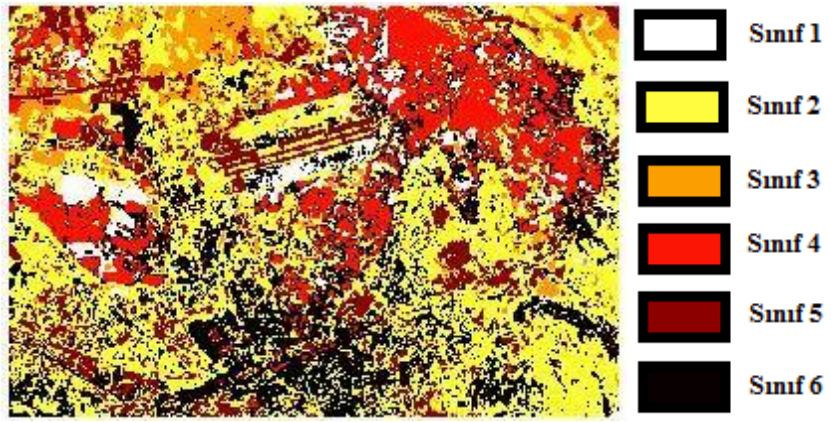
ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	20	**Eğitim**	**Test**
Örnekleme Oranı	0.5	%97.2532	%81.0099
YSA-Epoch Sayısı	1000		
İterasyon 20, beta = 0.928865, alpha = 0.073792			



Şekil 4.9. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:1000,İterasyon Sayısı:20,Örnekleme Oranı:0.5)

Tablo 4.10. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	20	**Eğitim**	**Test**
Örnekleme Oranı	0.5	%97.5966	%79.8024
YSA-Epoch Sayısı	2000		
İterasyon 20, beta = 0.694751, alpha = 0.364201			



Şekil 4.10. Landsat görüntüsünün Adaboost.M1-YSA ile sınıflandırılması (Epoch:2000,İterasyon Sayısı:20,Örnekleme Oranı:0.5)

YSA (ÇKP) yöntemi literatürde geçerliliği kabul gören, verideki gürültülerden çok az oranda etkilenen bir yöntem olmasına rağmen; öğrenme oranı, iterasyon sayısı gibi sınıflandırma başarısını doğrudan etkileyen parametrelere sahiptir. Bu nedenle en uygun parametrelerin belirlenmesi yaşanan en büyük problemdir. Genel olarak En Çok Olabilirlik gibi klasik yöntemlere göre daha iyi sonuçlar verdiği literatürde işlenmiştir. Yapılan YSA ile sınıflandırma işleminde diğer yöntemlerle karşılaştırıldığında orta düzey bir başarı sağlamaktadır. Adaboost.M1 yöntemi işleme dahil edildiğinde sınıflandırma doğruluğunun Aster görüntüsü için % 4.3 ve Landsat görüntüsü için % 1.6 oranında arttığı gözlemlenmiştir.

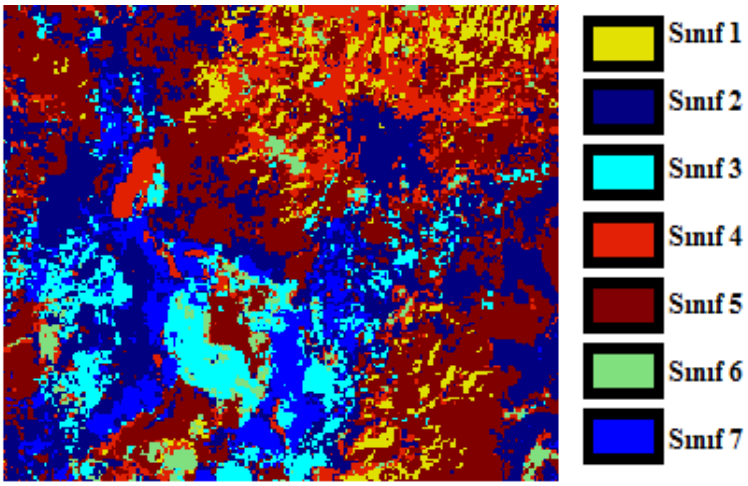
4.3. Rastgele Orman Yöntemi ile Görüntü Sınıflandırma

Rastgele Orman sınıflandırma yöntemi birçok karar ağacından meydana gelen ve bootstrap yöntemi ile rastgele özellik seçimi ile normal karar ağaçlarına göre daha doğru sonuçlar üretebilen bir yöntemdir. Temelinin anlaşılır ve sade olması ve dolayısıyla hızlı çalışması nedeniyle sıklıkla kullanılan bir yöntemdir.

Terra ASTER ve Landsat görüntüleri tez kapsamında Rastgele Orman ve Adaboost.M1 destekli Rastgele Orman yöntemleri ile sınıflandırılmış ve başarımları sunulmuştur. Terra ASTER görüntüsünün Rastgele Orman (RF) yöntemi ile sınıflandırılmış ve en yüksek başarımların ağaç sayısının 100 ve özellik sayısının 5 olduğunda elde edildiği görülmüştür. Tablo 4.11 ve Tablo 4.12'de farklı parametrelerde RF yönteminin doğrulukları görülmektedir.

Tablo 4.11. Terra ASTER görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

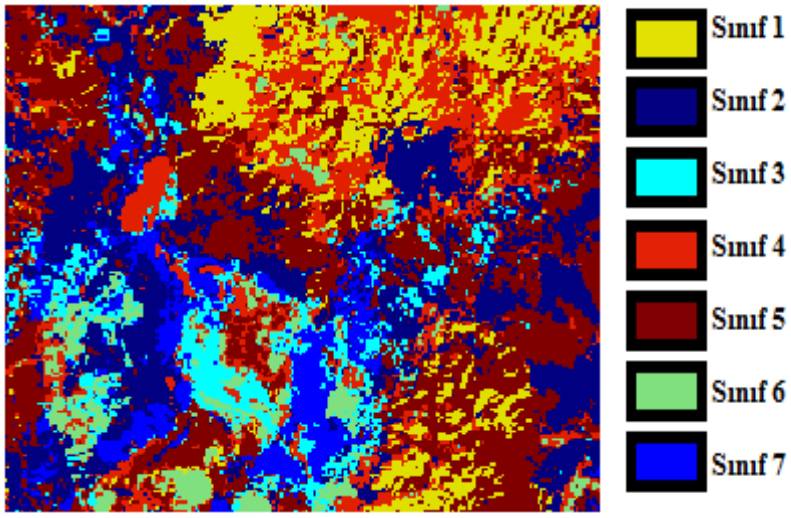
RF PARAMETRELER		**GENEL DOĞRULUK**	
Ağaç Sayısı	50	**Eğitim**	**Test**
Özellik Sayısı	5	%96.92	%76.4214



Şekil 4.11. Terra ASTER görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:50, Özellik Sayısı:5)

Tablo 4.12. Terra ASTER görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

RF PARAMETRELER		**GENEL DOĞRULUK**	
Ağaç Sayısı	100	**Eğitim**	**Test**
Özellik Sayısı	5	%96.92	%76.5886

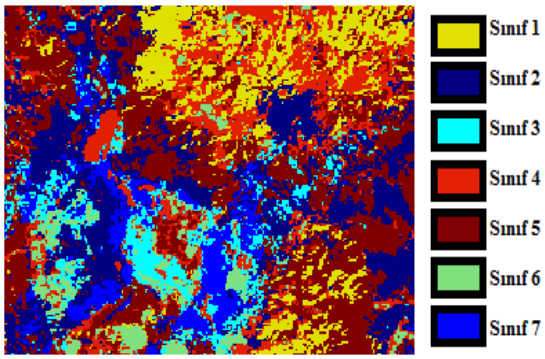


Şekil 4.12. Terra ASTER görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:100, Özellik Sayısı:5)

Adaboost.M1 algoritmasının RF yöntemi ile sınıflandırmaya dahil edilmesi ile elde edilen sonuçlar Tablo 4.13ve Tablo 4.14 da görülmektedir.

Tablo 4.13. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1).

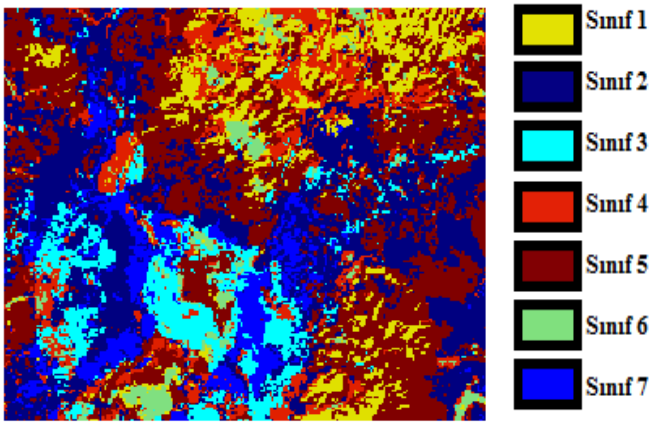
ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	20	**Eğitim**	**Test**
Örnekleme Oranı	0.3	%96.92	%76.5886
RF PARAMETRELER			
Ağaç Sayısı	50		
Özellik Sayısı	5		



Şekil 4.13. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:50,Özellik Sayısı:5,İterasyon Sayısı:20,Örnekleme Oranı:0.3)

Tablo 4.14. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	27	**Eğitim**	**Test**
Örnekleme Oranı	0.2	%96.92	%76.7559
RF PARAMETRELER			
Ağaç Sayısı	100		
Özellik Sayısı	5		
İterasyon 27: beta = 0.840935, alpha = 0.173241			



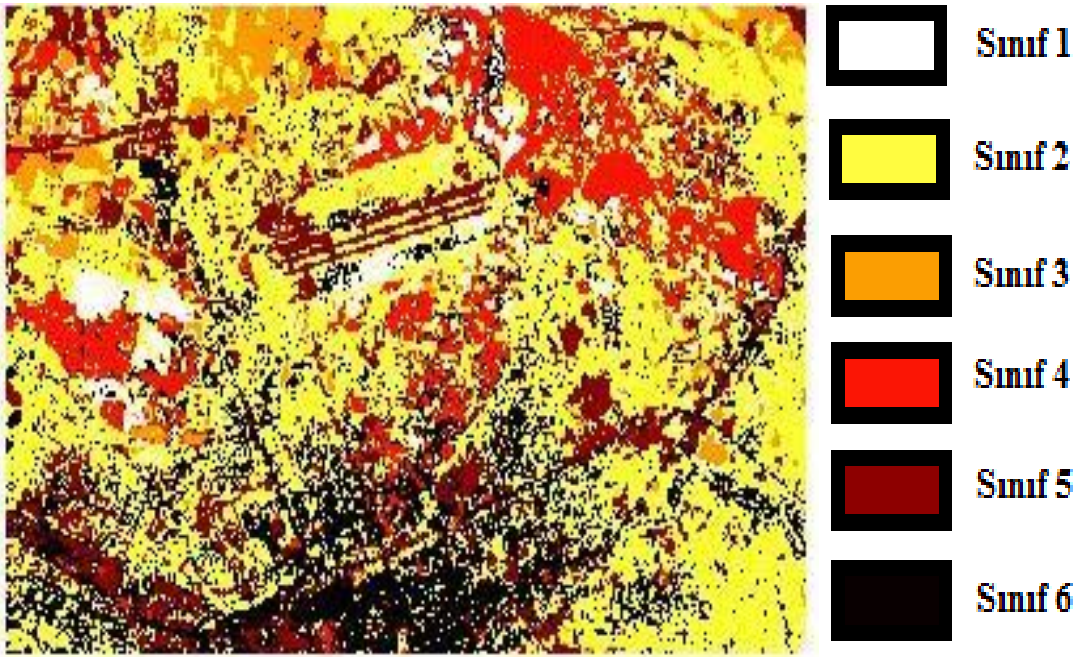
Şekil 4.14. Terra ASTER görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:100,Özellik Sayısı:5,İterasyon Sayısı:27, Örnekleme Oranı:0.2)

Sonuçlardan da anlaşıldığı üzere Terra ASTER verisinin kontrollü sınıflandırılmasında Adaboost.M1 yöntemi işleme dahil edildiğinde RF yönteminin başarımı arttırdığı gözlemlenmiştir.

Landsat verisi yine hem RF hem de Adaboost.M1-RF yöntemleri ile sınıflandırılmıştır. Landsat verisinin RF yöntemi ile sınıflandırılması sonucu elde edilen bulgular Tablo 4.15 ve Tablo 4.16'da verilmiştir.

Tablo 4.15. Landsat görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

RF PARAMETRELER		**GENEL DOĞRULUK**	
Ağaç Sayısı	50	**Eğitim**	**Test**
Özellik Sayısı	5	%99.4850	%78.4852



Şekil 4.15. Landsat görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:50,Özellik Sayısı:5)

Tablo 4.16. Landsat görüntüsünün RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

RF PARAMETRELER		**GENEL DOĞRULUK**	
Ağaç Sayısı	500	**Eğitim**	**Test**
Özellik Sayısı	5	%99.4850	%78.7047



Şekil 4.16. Landsat görüntüsünün RF ile sınıflandırılması (Ağaç Sayısı:500,Özellik Sayısı:5)

Landsat görüntüsünün RF ile sınıflandırılmasına Adaboost.M1 algoritması dahil edildiğinde ortaya çıkan sonuçlar Tablo 4.17 ve Tablo 4.18'de verilmiştir.

Tablo 4.17. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1).

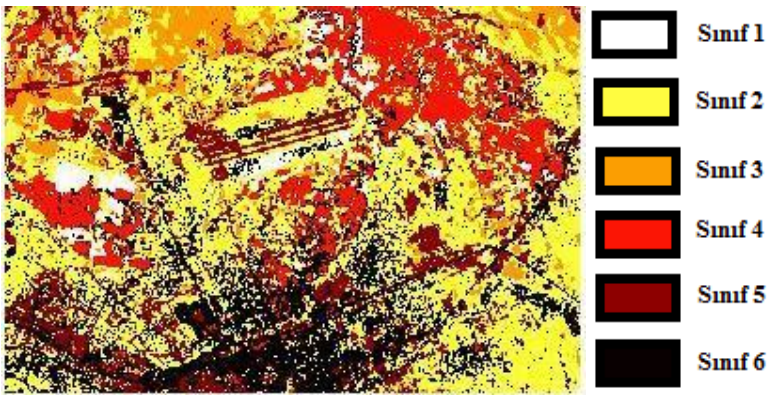
ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	10	**Eğitim**	**Test**
Örnekleme Oranı	0.2	%99.4850	%78.5950
RF PARAMETRELER			
Ağaç Sayısı	50		
Özellik Sayısı	5		
İterasyon 10: beta = 0.625880, alpha = 0.468597			



Şekil 4.17. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:50,Özellik Sayısı:5,İterasyon Sayısı:10,Örnekleme Oranı:0.2)

Tablo 4.18. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2).

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İtersayon Sayısı	10	**Eğitim**	**Test**
Örnekleme Oranı	0.2	%99.4850	%79.1438
RF PARAMETRELER			
Ağaç Sayısı	500		
Özellik Sayısı	5		
İterasyon 10: beta = 0.655872, alpha = 0.421790			



Şekil 4.18. Landsat görüntüsünün Adaboost.M1-RF ile sınıflandırılması (Ağaç Sayısı:500,Özellik Sayısı:5, İterasyon Sayısı:10, Örnekleme Oranı:0.2)

Rastgele orman yöntemi Bagging (Torbalama) ve Random Subspace tekniklerinin birleşiminden oluşan, temelinde karar ağaçları bulunan bir topluluk öğrenme yöntemidir. Bu yöntemde kullanıcı tarafından ağaç sayısı ve özellik sayınsının belirlenmesi gerekmektedir. Parametrelerin ne olacağı, kullanılan veriye göre değişiklik gösterebileceğinden deneme-yanılma yoluyla bulunur. Aster görüntünsün RF ile sınıflandırması sonucu en düşük doğruluk değeri, Landsat görüntüsü ile ise orta düzey bir doğruluk elde ettiği görülmüştür.

RF yönteminde Adaboost.M1 Algoritması'nın kullanılması ile Aster görüntüsünde % 0.2 ve Landsat görüntüsünde ise % 0.4 oranında doğruluğun arttığı gözlemlenmiştir. İki görüntüde de başarıyı arttırması, Adaboost.M1 algoritmasının RF yönteminin başarısını arttırmak için kullanılabileceğini göstermektedir.

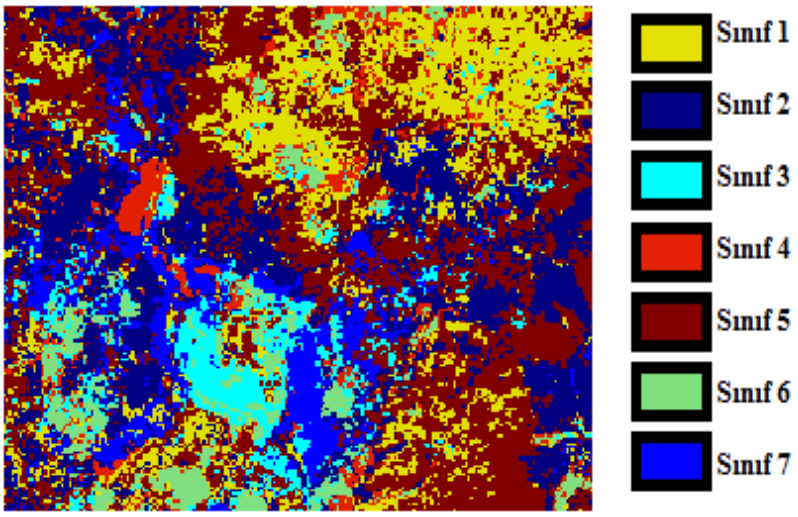
4.4. Sınıflandırma ve Regresyon Ağaçları (CART) İle Görüntü Sınıflandırma

1984 yılında Breiman tarafından ortaya atılan bu yöntem veri madenciliğinin sınıflandırma ile ilgili en önemli en önemli konularından biri olmuştur. Karar ağaçları temelli bir yöntem olan CART, çok hızlı ve adaptif olmasından dolayı birçok farklı disiplinde yoğun olarak kullanılmaktadır. Bu yöntemi diğer karar ağaçlarından ayıran en önemli özellik sadece ikili dallanmalara izin vermesidir. Bu yöntemde genel olarak Twoing Algoritması ve Gini Algoritması kullanılmaktadır.

Terra ASTER verisine CART yöntemi farklı parametreler kullanılarak uygulandığında Tablo 4.19 ve Tablo 4.20'deki sonuçlar elde edilmiştir.

Tablo 4.19. Terra ASTER görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

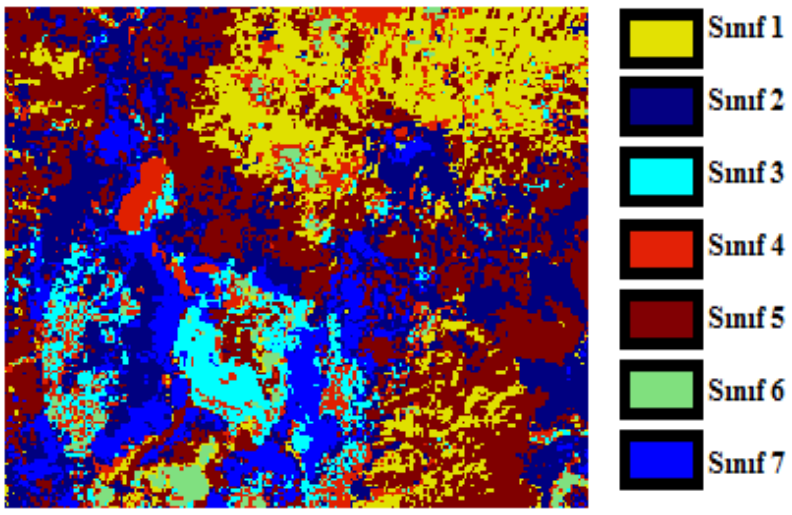
CART PARAMETRELER		**GENEL DOĞRULUK**	
Gözlem Sayısı	10	**Eğitim**	**Test**
Algoritma	GİNİ	%88.7640	%86.6221



Şekil 4.19. Terra ASTER görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:10-Gini)

Tablo 4.20. Terra ASTER görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

CART PARAMETRELER		**GENEL DOĞRULUK**	
Gözlem Sayısı	20	**Eğitim**	**Test**
Algoritma	GİNİ	%94.3820	%81.6054

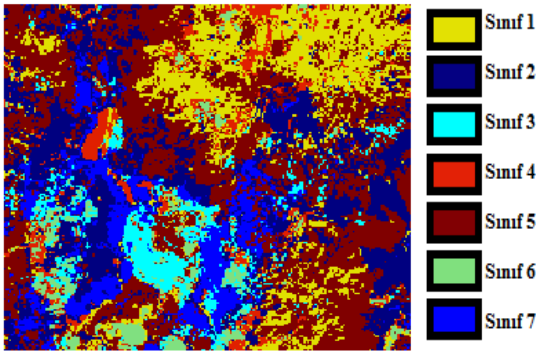


Şekil 4.20. Terra ASTER görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:20-Gini)

Adaboost.M1 Algoritması CART algoritması ile sınıflandırma işlemine dahil edildiğinde ise elde edilen doğruluk değerleri Tablo 4.21 ve Tablo 4.22'de verilmiştir.

Tablo 4.21. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1).

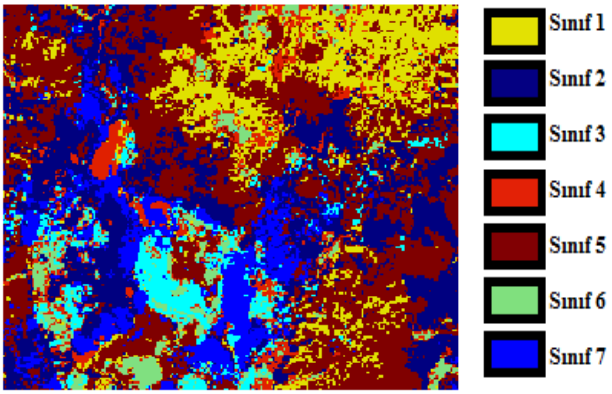
ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İterasyon Sayısı	5	**Eğitim**	**Test**
Örnekleme Oranı	0.2	%88.7640	%87.2910
CART PARAMETRELER			
Gözlem Sayısı	10		
Algoritma	GİNİ		
İterasyon 5: beta = 0.722612, alpha = 0.324882			



Şekil 4.21. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:10-Gini, İterasyon Sayısı:5, Örnekleme Oranı:0.2)

Tablo 4.22. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1).

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İterasyon Sayısı	5	**Eğitim**	**Test**
Örnekleme Oranı	0.2	%94.3820	%83.4448
CART PARAMETRELER			
Gözlem Sayısı	20		
Algoritma	GİNİ		
İterasyon 5: beta = 0.340640, alpha = 1.076930			

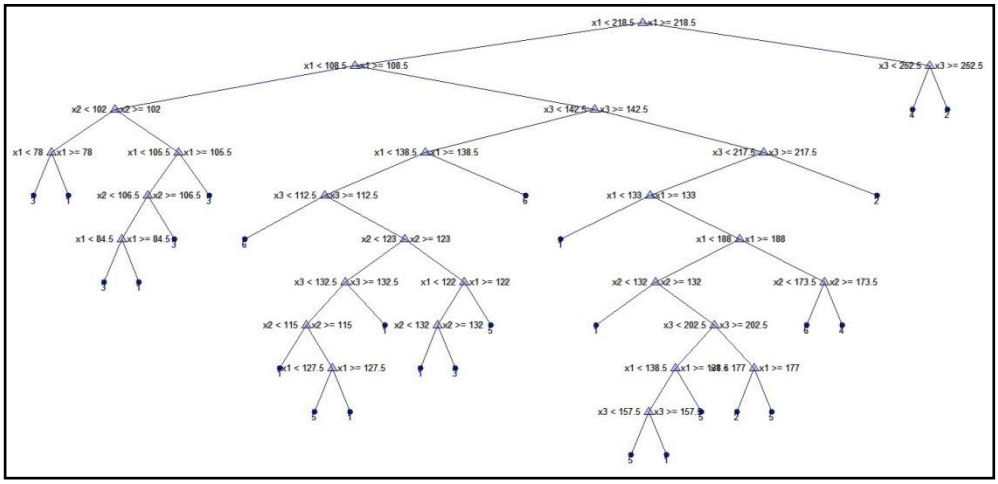


Şekil 4.22. Terra ASTER görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:20-Gini, İterasyon Sayısı:5, Örnekleme Oranı:0.2)

Landsat uydu görüntüsü CART Algoritması ile farklı parametreler kullanılarak sınıflandırıldıklarında elde edilen sonuçlar Tablo 4.23 ve Tablo 4.24'de verilmiştir.

Tablo 4.23. Landsat görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

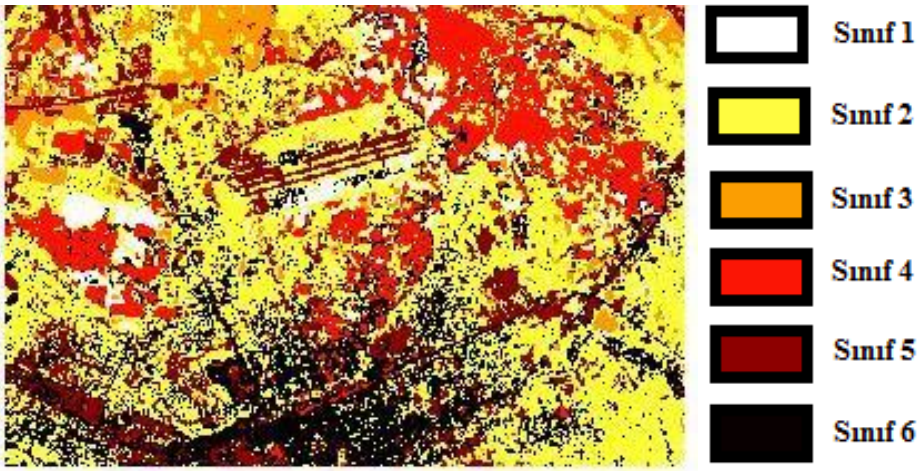
CART PARAMETRELER		**GENEL DOĞRULUK**	
Gözlem Sayısı	10	**Eğitim**	**Test**
Algoritma	GİNİ	%96.5665	%81.4490



Şekil 4.23. Gini Algoritması ile oluşan ağaç yapısı

Tablo 4.24. Landsat görüntüsünün CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

CART PARAMETRELER		**GENEL DOĞRULUK**	
Gözlem Sayısı	20	**Eğitim**	**Test**
Algoritma	GİNİ	%97.2532	%80.1317



Şekil 4.24. Landsat görüntüsünün CART ile sınıflandırılması (Gözlem Sayısı:20-Gini)

Landsat görüntüsünün CART ile sınıflandırılmasına Adaboost.M1 algoritması dahil edildiğinde ortaya çıkan sonuçlar Tablo 4.25 ve Tablo 4.26'da verilmiştir.

Tablo 4.25. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (1)

ADABOOST.M1 PARAMETRELER		**GENEL DOĞRULUK**	
İterasyon Sayısı	11	**Eğitim**	**Test**
Örnekleme Oranı	0.05	%96.5665	%81.6685
CART PARAMETRELER			
Gözlem Sayısı	10		
Algoritma	GİNİ		
İterasyon 11: beta = 0.674019, alpha = 0.394497			



Şekil 4.25. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:10-Gini,İterasyon Sayısı:11, Örnekleme Oranı:0.05)

Tablo 4.26. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılmasına ait parametre ve doğruluk değerleri (2)

<table>
<tr><th colspan="2">ADABOOST.M1 PARAMETRELER</th><th colspan="2">GENEL DOĞRULUK</th></tr>
<tr><td>İterasyon Sayısı</td><td>15</td><td>Eğitim</td><td>Test</td></tr>
<tr><td>Örnekleme Oranı</td><td>0.05</td><td rowspan="4">%97.2532</td><td rowspan="4">%80.2415</td></tr>
<tr><td colspan="2">CART PARAMETRELER</td></tr>
<tr><td>Gözlem Sayısı</td><td>20</td></tr>
<tr><td>Algoritma</td><td>GİNİ</td></tr>
<tr><td colspan="4">İterasyon 11: beta = 0.778684, alpha = 0.250149</td></tr>
</table>



Şekil 4.26. Landsat görüntüsünün Adaboost.M1-CART ile sınıflandırılması (Gözlem Sayısı:20-Gini,İterasyon Sayısı:15, Örnekleme Oranı:0.05)

Adaboost.M1 Algoritması YSA, RF yöntemlerinde olduğu gibi CART yönteminin de sınıflandırma doğruluğunu arttırmıştır. Sınıflandırma ve Regresyon Ağaçları Yöntemi (CART) ikili dallanmalara izin veren ve veri madenciliğinde hem sınıflandırma hem de regresyon amaçlı, yoğun olarak kullanılan karar ağaçları temelli bir yöntemdir. Parametre olarak kullanıcı tarafından ağaç sayısı ve dallanma için kullanılacak algoritmanın seçilmesi gerekmektedir. Bu çalışmada CART yöntemi için Gini Algoritması kullanılmıştır. Yöntem Adaboost.M1 Algoritması ile beraber kullanıldığında ise Aster görüntüsünde % 0.6, Landsat görüntüsünde ise % 0.2 doğruluk artırımı sağlamıştır.

4.5. Destek Vektör Makineleri İle Görüntü Sınıflandırma

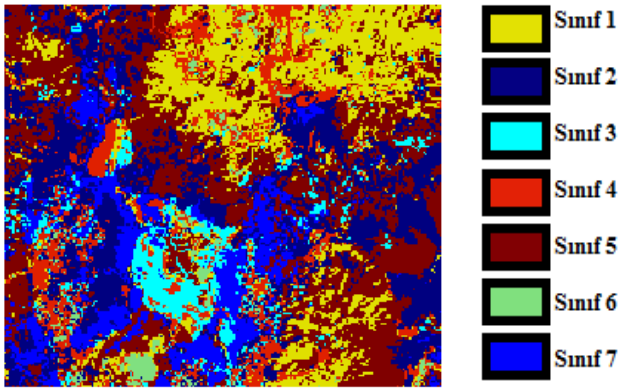
DVM, Cortes ve Vapnik tarafından ortaya atılan [11], günümüzde veri madenciliğinde sıklıkla kullanılan ve istatistiksel öğrenmeye dayalı bir yöntemdir. İlk olarak iki sınıflı sınıflandırma işlemleri için geliştirilmiş olan DVM daha sonra çok sınıflı verilerin sınıflandırılması için kullanılmaya başlanmıştır.

DVM, sınıflandırılacak verilerin lineer olarak ayrılabilme özelliklerine göre farklı stratejileri izlemek gerekmektedir. Günümüzdeki pratik uygulamalarda verilerin lineer olarak ayrılabilmeleri mümkün olmadığı için Kernel fonksiyonları kullanılması gerekmektedir.

Bu tez kapsamında DVM ile sınıflandırma için literatürde en çok kullanılan yöntemlerden Lineer Fonksiyon Kerneli, Polinom Kerneli ve Radyal Tabanlı Fonksiyon Kerneli kullanılmıştır. Terra ASTER görüntüsünün belirtilen fonksiyonlar ile sınıflandırılması ile elde edilen sonuç Tablo 4.27, Tablo 4.28, ve Tablo 4.29'da verilmiştir.

Tablo 4.27. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(1)

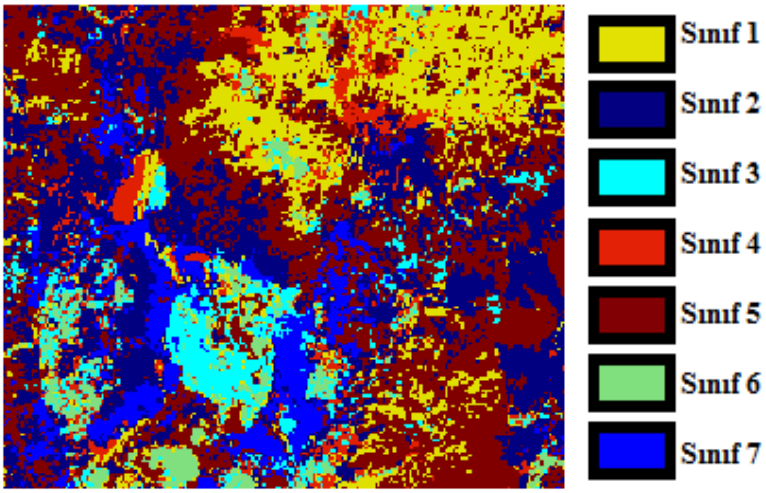
DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	Lineer	**Eğitim**	**Test**
C Değeri	900	%85.3933	%80.1003



Şekil 4.27. Terra ASTER görüntüsünün DVM ile sınıflandırılması (Lineer Kernel, C:900)

Tablo 4.28. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(2)

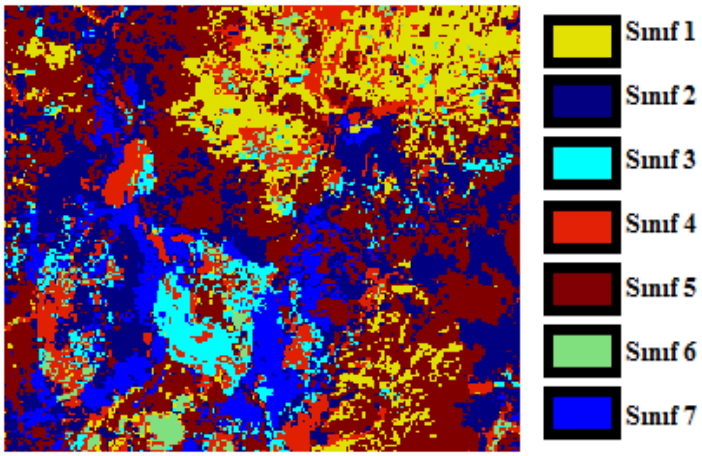
DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	Polinom	**Eğitim**	**Test**
C Değeri	900	%86.2360	%82.2742
Derecesi(d)	2		



Şekil 4.28. Terra ASTER görüntüsünün DVM ile sınıflandırılması (Polinom Kerneli,C:900,d:2)

Tablo 4.29. Terra ASTER görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(3)

DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	RTF	**Eğitim**	**Test**
C Değeri	100	%88.0618	%84.7826
Gama	2		



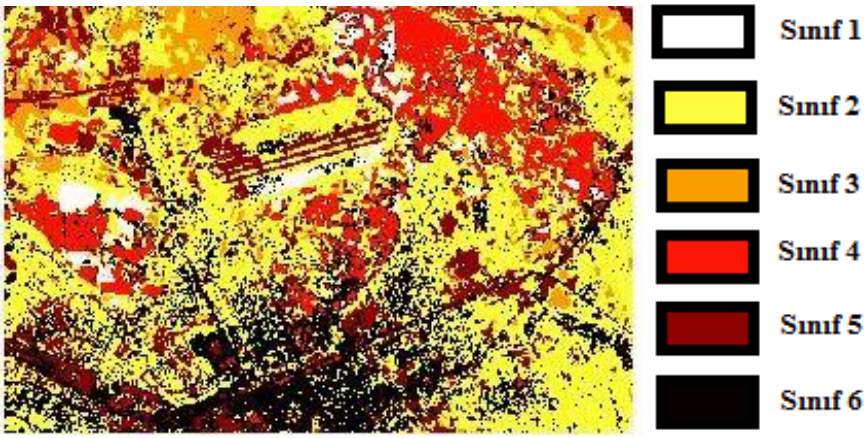
Şekil 4.29. Terra ASTER görüntüsünün DVM ile sınıflandırılması (RTF Kerneli,C:100 ,Gama:2)

Terra ASTER görüntüsünün DVM ile farklı Kernel'ler kullanılarak sınıflandırılması sonucu en yüksek sınıflandırma doğruluğunu RTF Kerneli sağladığı gözlemlenmiştir.

Diğer test verimiz olan Landsat görüntüsü yine RTF, Polinom ve Linner Kernel'ler ile sınıflandırılmış ve sonuçları Tablo 4.30, Tablo 4.31 ve Tablo 4.32'de verilmiştir.

Tablo 4.30. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(1)

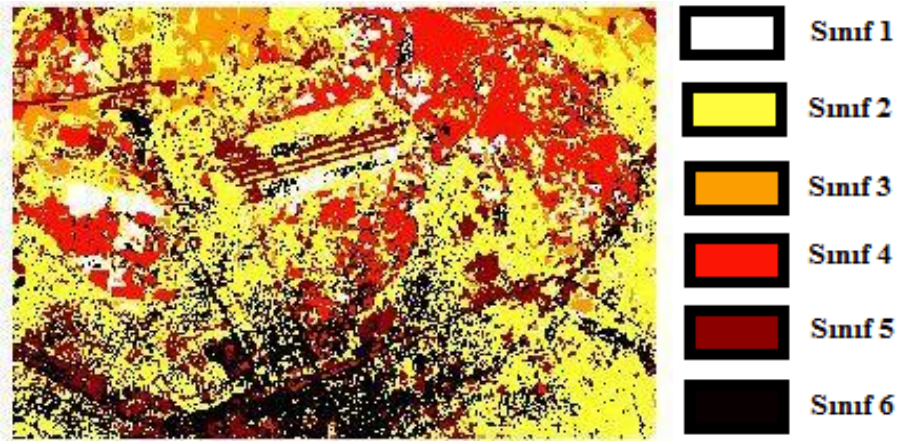
DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	Lineer	**Eğitim**	**Test**
C Değeri	900	%95.2790	%76.5093



Şekil 4.30. Landsat görüntüsünün DVM ile sınıflandırılması
(Lineer Kernel,C:900)

Tablo 4.31. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(2)

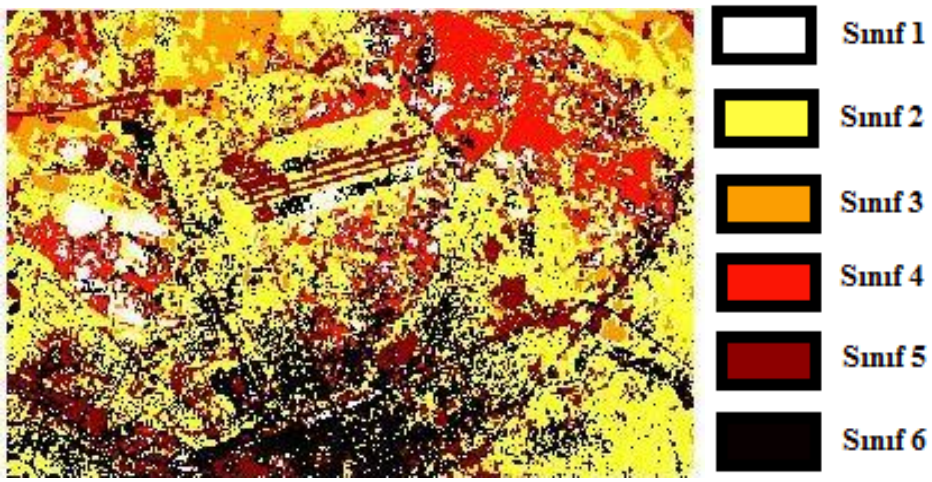
DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	Polinom	**Eğitim**	**Test**
C Değeri	20000	%95.6223	%83.5346
Derecesi(d)	2		



Şekil 4.31. Landsat görüntüsünün DVM ile sınıflandırılması
(Polinom Kerneli, C: 20000, d:2)

Tablo 4.32. Landsat görüntüsünün DVM ile sınıflandırılmasına ait parametre ve doğruluk değerleri(3)

DVM PARAMETRELER		**GENEL DOĞRULUK**	
Kernel	RTF	**Eğitim**	**Test**
C Değeri	200	%95.5365	%83.7541
Gama	1		



Şekil 4.32. Landsat görüntüsünün DVM ile sınıflandırılması (RTF Kerneli,C:200,Gama:1)

Landsat görüntüsünün DVM ile sınıflandırılması sonucu en iyi sonuç RTF Kerneli ile elde edilmiştir. Çalışmalarda kullanılan C, gama ve d (derece) değerleri deneme yanılma yoluyla bulunmuştur.

4.6. En Çok Olabilirlik Yöntemi İle Görüntü Sınıflandırma

Terra ASTER ve Landsat verileri EÇO yöntemi ile sınıflandırılmış ve Aster verisi için % 80.002, Landsat verisi için %77.5406 doğruluk elde edilmiştir.

5. BÖLÜM

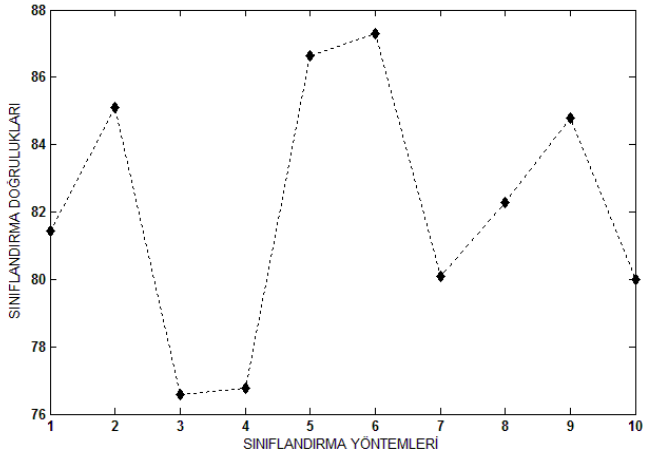
SONUÇLAR

Uzaktan Algılama, çok geniş alanları görüntüleyebilme yeteneğinden dolayı farklı bilim dalları tarafından yoğun olarak kullanılmaktadır. Uydular tarafından elde edilen görüntülerin yorumlanabilmesi için en önemli aşama sınıflandırma işlemidir. Sınıflandırma kontrollü ve kontrolsüz olmak üzere iki biçimde yapılmaktadır. Araştırmacıların ihtiyaçlarına göre eğitim alanlarını kendilerinin belirleyebilmelerine izin vermesi ve daha gerçekçi sonuçlar üretebilmesinden dolayı kontrollü sınıflandırma daha sık kullanılmaktadır. Sınıflandırma işlemi sonucu ortaya çıkan sınıflara ait bilgilerin doğruluğu, sınıflandırma işlemi için seçilecek yöntemle doğrudan ilişkilidir. Dolayısıyla yapılan çalışmanın niteliğine ve veri durumuna göre kullanılacak yöntemin doğru olarak seçimi büyük önem taşımaktadır. Bu tez bu çalışması kapsamında Terra ASTER ve Landsat görüntüleri, YSA ve EÇO yöntemlerinin yanı sıra, Rastgele Orman (RF), CART, DVM ve Boosting yöntemi ile sınıflandırılmıştır.

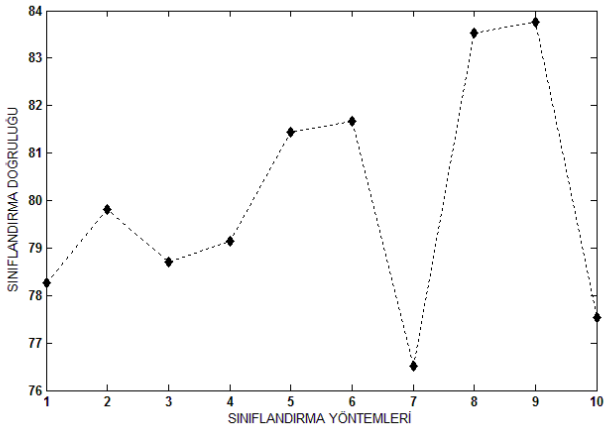
Hızlandırma (Boosting) Yöntemi sınıflandırmanın söz konusu olduğu her alanda kullanılabilmektedir. İlk olarak iki sınıflı verilerin sınıflandırılma doğruluğunu arttırma için tasarlanan yöntemler daha sonra çok sınıflı verilerin sınıflandırılmasına uygun hale getirilmiştir. Günümüzde Adaboost, Logitboost, Brownboost gibi birçok algoritma hızlandırma amaçlı olarak kullanılmaktadır. Bu çalışmada Adaboost Algoritması'nın çok sınıflı veriler için uygun hale getirilmiş olan Adaboost.M1 algoritmasının YSA, RF ve CART yöntemleriyle olan uyumu ve doğruluk artırımı araştırılmıştır. Yapılan işlemler sonucunda Adaboost.M1 Algoritması'nın birlikte kullanıldığı tüm yöntemlerde doğruluğu arttırdığı gözlemlenmiştir.

Aster görüntüsü ile yapılan sınıflandırma işleminde, CART yönteminin en yüksek doğruluğu, RF yönteminin en düşük doğruluğu ve diğer yöntemlerin ise orta derecede doğruluk elde ettiği gözlemlenmiştir (Şekil 5.1).

Landsat görüntüsünün sınıflandırılmasında ise DVM-Lineer ve EÇO yöntemlerinin en düşük doğruluğu, DVM-Polinom ile DVM-RTK yöntemlerinin ise en yüksek doğruluğu elde ettiği görülmüştür (Şekil 5.2).



Şekil 5.1. Terra ASTER görüntüsünün sınıflandırılmasında kullanılan yöntemler ve sınıflandırma doğrulukları. 1:YSA, 2: Adaboost.M1 YSA, 3:RF,4: Adaboost.M1-RF, 5:CART, 6: Adaboost.M1-CART, 7:DVM-Lineer, 8: DVM-Polinom, 9: DVM RTK, 10:EÇO



Şekil 5. 2. Landsat görüntüsünün sınıflandırılmasında kullanılan yöntemler ve sınıflandırma doğrulukları. 1:YSA, 2: Adaboost.M1-YSA, 3:RF,4: Adaboost.M1-RF, 5:CART, 6: Adaboost.M1-CART, 7:DVM-Lineer, 8: DVM-Polinom, 9: DVM-RTK, 10:EÇO

Çalışma boyunca karşılaşılan en büyük sıkıntı yöntemlere ait başlangıç parametrelerinin tahmin edilmesi olmuştur. Literatürde, tez kapsamında kullandığımız yöntemlerin kullanıldığı çok sayıda çalışma mevcuttur. Ama kullanılan verilerin farklı olmasından dolayı, kullanılan parametrelerin başarıları değişim gösterebilmektedir.

Gelecekteki çalışmalarda yöntemlere ait başlangıç parametrelerinin doğru olarak tespit edilebilmesi halinde birlikte kullanıldığı yöntemlerin doğruluğunu şu anda elde edilen değerlerden daha fazla arttırabileceği düşünülmektedir.

KAYNAKLAR

1. Huang, C., Davis, L. S. And Townshed, Jrg., An Assessment Of Support Vector Machines For Land Cover Classification, International Journal Of Remote Sensing, 23, 725–749, 2002.

2. Atkinson, P. M. And Tatnall, A. R. L., Neural Networks In Remote Sensing. International Journal Of Remote Sensing, 18, 699–709, 1997.

3. Özkan C., Uzaktan Algilama Ders Notları (Yayınlanmamış), Lisans Programı, Kayseri, 2008.

4. Kavzoglu, T., An Investigation Of The Design And Use Of Feed-Forwad Artificial Neural Networks İn The Classification Of Remotely Sensed İmages. Phd Thesis, Nottingham University, Uk. 2001.

5. Çölkesen, İ., Uzaktan Algılamada İleri Sınıflandırma Tekniklerinin Karşılaştırılması ve Analizi, Yüksek Lisans Tezi, Gebze, 2009.

6. Cortés, E.,ve ark., Multiclass Corporate Failure Prediction By Adaboost.M1, International Advances İn Economic Research, 13, Issue 3, P. 301-312. 2007.

7. Mather, P.M., Computer Processing Of Remote-Sensed Images. John Wiley And Sons Ltd., 1987.

8. Kansu,O., Uzaktan Algılamada Görüntü Sınıflandırma Yöntemleri Analizi,Yüksek Lisans Tezi, Trabzon, 2006.

9. Çomak, E., Destek Vektör Makinelerinin Eğitimi İçin Yeni Yaklaşımlar, Doktora Tezi, Selçuk Üniversitesi, Konya, 2008.

10. Çomak, E., Destek Vektör Makineleri Çoklu Sınıf Problemleri İçin Çözüm Önerileri, Yüksek Lisans Tezi, Selçuk Üniversitesi, Konya, 2004.

11. Cortes, C. and Vapnik, V., Support-vector network. Machine Learning, 20, 273–297., 1995.

12. Demirci, D., Destek Vektör Makineleri İle Karakter Tanıma, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, İstanbul, 2007.

13. Özkan, Y., Veri Madenciliği Yöntemleri,S.187-188, Papatya Yayıncılık, 2008.

14. Vapnik, V.N., The Nature Of Statistical Learning Theory, Springer-Verlag, New York, 1995.

15. Kavzoğlu, T., Çölkesen, İ., Destek Vektör Makineleri İle Uydu Görüntülerinin Sınıflandırılmasında Kernel Fonksiyonlarının Etkilerinin İncelenmesi, Harita Dergisi, 2010.

16. Kavzoğlu, T., Çölkesen, İ., Karar Ağaçları İle Uydu Görüntülerinin Sınıflandırılması: Kocaeli Örneği, Harita Teknolojileri Dergisi, 36-45, Cilt:2, No:1 2010

17.Safavian S.R., Landgrebe D., A Survey Of Decision Tree Classifier Methodology, IEEE Transactions On Systems Man And Cybernetics, 21, 660-674, 1991.

18. Akman, M., Veri Madenciliğine Genel Bakış ve Random Forests Yönteminin İncelenmesi: Sağlık Alanında Bir Uygulama, Yüksek Lisans Tezi, Ankara Üniversitesi, Ankara, 2010.

19. Silahtaroğlu,G., Kavram ve Algoritmalarıyla Temel Veri Madenciliği,Papatya Yayıncılık Eğitim.S.29-60. 2008.

20. Tan, P., Steınbach, M., Kumar,V. Introduction To Data Mining, Addison Wesley., 2005.

21. Yıldırım, S.,Tümevarım Öğrenme Tekniklerinden C4.5'in İncelenmesi,Yüksek Lisans Tezi, İstanbul Teknik Üniversitesi, İstanbul, 2003.

22. Kantardzic, M., Data Mining: Concepts, Methods And Algorithms, Wiley,2003.

23. Quinlan, J.R., C4.5:Programs For Machine Learning, Morgan Kaufman,1993.

24. Larose, D.T.,Discovering Knowledge In Data: An Introduction To Data Mining, Wiley, 2005.

25. Pal, M., Random Forest Classifier For Remote Sensing Classification. International Journal Of Remote Sensing, 26, 217-222, 2005.

26. Mather, P.M., Computer Processing Of Remote-Sensed Images. John Wiley And Sons Ltd., 1987.

27. Schapire, R. E., Theoretical Views Of Boosting And Applications, Proceedings Of The Tenth International Conference On Algorithmic Learning Theory, 13 – 25, 1999.

28. R. E. Schapire, A Brief Introduction To Boosting, In Proceedings Of The Sixteenth International Joint Conference On Artificial Intelligence, 1999.

ÖZGEÇMİŞ

Ümit Haluk ATASEVER, 1985 yılında Adana'da doğdu. İlk ve orta öğrenimini Adana'da tamamladı. 2003 yılında Danişment Gazi Anadolu Lisesi'nden mezun oldu. 2004 yılında kazandığı Erciyes Üniversitesi Mühendislik Fakültesi Harita Mühendisliği Bölümü'nden 2008 yılında mezun oldu ve aynı yıl Harita Mühendisliği Ana Bilim Dalında yüksek lisansa başladı. 2009 yılında Erciyes Üniversitesi Mühendislik Fakültesi Harita Mühendisliği Bölümü'nde araştırma görevlisi olarak göreve başladı ve halen bu görevine devam etmektedir.

Adres: Erciyes Üniversitesi Mühendislik Fakültesi Harita Mühendisliği Bölümü

Telefon: 0 352 437 49 01-32678

E-posta: uhatasever@erciyes.edu.tr