

KOCAELİ ÜNİVERSİTESİ
Mühendislik Fakültesi

Elektronik ve Haberleşme Mühendisliği

Mühtas -2-



Proje Konusu

**Hiperspektral Verilerin Sığ Yöntemler ile
Sınıflandırılması**

Hazırlayan
Zeynep ÖZÇELİK
190208062

Proje Danışmanı
Alp ERTÜRK

Kocaeli, Mayıs
2023

İÇİNDEKİLER

HAZIRLAYAN	1
İÇİNDEKİLER	2
ŞEKİL LİSTESİ	3
TABLO LİSTESİ	3
KISALTMALAR DİZİNİ	4
Hiperspektral Verilerde Sığ Yöntemler ile Sınıflandırma	5
Anahtar Kelimeler	5
ÖZET	5
1. AMAÇ	5
2. GİRİŞ	6
3. GENEL KISIMLAR	6
3.1 UZAKTAN ALGILAMA	6
3.2 HİPERSPEKTRAL GÖRÜNTÜLEME	7
3.2.1 Hiperspektral Görüntüler	8
Literatür Özeti	12
3.3 BOYUT İNDİRGEME	13
3.3.1 Temel Bileşen Analizi (TBA)	13
3.4 MODEL EĞİTİMİ	14
3.4.1 Aşırı Öğrenme (Overfitting)	14
3.4.2 Eksik Öğrenme (Underfitting)	14
3.5 SINIFLANDIRMA YÖNTEMİ	15
3.5.1 Destek Vektör Makineleri	15
3.5.2 Hard Margin vs Soft Margin	16
3.6 KERNEL TRICK	17
3.6.1 Gauss Çekirdeklemesi (Gaussian Rbf-Radial Basis Function/Kernel)	17
3.7 METRİKLER	17
3.8 SINIFLANDIRMA BAŞARIM ÖLÇÜTLERİ	19
3.8.1 Genel Doğruluk (OA)	19
3.8.2 Ortalama Doğruluk (AA)	19
4. VERİ KÜMELERİ	20
4.1 Indian Pines	20
4.2 Pavia University	21
5. DENEYSEL SONUÇLAR	22
5.1 TBA ile fazlalık analizi	24
5.2 Model Eğitimi ve Testi	27
5.2.1 Veri Hazırlama ve Ön İşleme	27
5.3 DVM Modelimizin Eğitimi	27
5.4 Sınıflandırıcının Değerlendirilmesi	27
6. SONUÇLAR	32
7. PROJE KODU	34
8. KAYNAKÇA	37

ŞEKİL LİSTESİ

Şekil 3.1: Uzaktan algılama genel şeması[3]	6
Şekil 3.2 : Elektromanyetik spektrum [5]	7
Şekil 3.3 : Hiperspektral görüntüleme düzeneği [6]	7
Şekil 3.4: Hiperspektral görüntü küpü[1].....	8
Şekil 3.5: Hiperspektral görüntüdeki bir pikseldeki maddeye ait spektral imza [7].....	8
Şekil 3.6: Pavia University veri kümesindeki her bir sınıfa ait birer piksel için spektral imza örnekleri.....	9
Şekil 3.7 : Bu çalışmada yeni vejetasyon indeksleri ile tahmin edilen yaprak seviyesindeki SPAD, LNC ve ChiP parametreleri haritası [10]	10
Şekil 3.8 : İstilacı ot haritalama [11].....	10
Şekil 3.9 : Aşırı öğrenme (a) ve (b) Eksik öğrenme [30].....	13
Şekil 3.10: SVM'nin Destek Vektörleri ve Margin Görüntülerinin Grafikselleştirilmesi [27]	14
Şekil 3.11 (a) Hard Margin (b) Soft Margin [28]	15
Şekil 3.12 : DVM'deki C Parametresinin Etkisi [29].....	15
Şekil 3.13 : Gamma Hiperparametresinin etkisi [29]	16
Şekil 4.1: Indian Pines a) bant (43,21,11) renkli görüntü b) yer doğrusu.....	20
Şekil 4.2 : Pavia University a) bant (44,24,10) renkli görüntüsü b) yer doğrusu	21
Şekil 5.1 : Kızılderili Çamlarının Bantlarının Görselleştirilmesi (IP) Veri Kümesi.....	21
Şekil 5.2: Pavia University rasgele seçilen 6 bant görüntüsü	22
Şekil 5.3 : (a) Indian Pines renk çubuğu (b) Pavia University renk çubuğu.....	22
Şekil 5.4 : IP veri kümesi bileşenlerinin varyans açıklama oranlarının görselleştirilmesi	23
Şekil 5.5: Pavia University veri kümesi bileşenlerinin varyans açıklama oranlarının görselleştirilmesi	23
Şekil 5.6: IP veri kümesinin TBA işlemi sonrası 8 temel bileşeni.....	25
Şekil 5.7 : Pavia University veri kümesinin TBA işlemi sonrası 8 temel bileşeni....	25
Şekil 5.8: IP verilerinden elde edilen Karışıklık Matrisi.....	27
Şekil 5.9 : Pavia University verilerinden elde edilen Karışıklık Matrisi	27
Şekil 5.10 : (a) Pavia University veri kümesinin (b) IP veri kümesinin DVM ile sınıflandırma sonucu	29
Şekil 5.11 : Indian Pines ve Pavia University C = 10 değerlerindeki sınıflandırma sonuçları	30
Şekil 5.12: Indian Pines ve Pavia University C = 60 değerlerindeki sınıflandırma sonuçları.....	30
Şekil 5.13 Indian Pines ve Pavia University veri kümelerinin görselleştirilmiş sonuçları	7

TABLO LİSTESİ

Tablo 3.1: İki sınıflı sınıflandırma için karışıklık matrisi	18
Tablo 4.1: Indian Pines veri kümesine ait sınıflar ve örnekler sayıları	20
Tablo 4.2: Pavia University veri kümesine ait sınıflar ve örnekler sayıları.....	21
Tablo 5.1 : IP veri kümesi sınıflandırma raporu	29
Tablo 5.2: Pavia University veri kümesi sınıflandırma raporu	30

KISALTMALAR DİZİNİ

- HSG : Hiperspektral Görüntü
- DVM : Destek Vektör Makinaları
- TBA : Temel Bileşen Analizi
- KYM : Kırmızı-Yeşil-Mavi
- HIS : Hyperspectral Image
- IP : Indian Pines
- TP : True-Pozitif
- TN : True Negatif
- FP : False Pozitif
- FN : False Negatif
- AVIRIS : Airborne Visible / Infrared Imaging Spectrometer
- KYM : Kampala-Yale-Munich
- OA : Overall Accuracy
- RBF : Radial Basis Function
- AA : Average Accuracy

Hiperspektral Verilerde Sığ Yöntemler ile Sınıflandırma

Anahtar Kelimeler

Uzaktan Algılama, Hiperspektral Görüntüleme, Hiperspektral Görüntülerde Sınıflandırma, Destek Vektör Makinası, Temel Bileşen Analizi

ÖZET

Hiperspektral görüntüleme birçok dar dalga boyu bandına ait imge verilerini elde etmekte ve dalga boyuna göre bir ayrışım sağlamaktadır. Hiperspektral görüntülerde her piksel için dalga boyuna bağlı bir spektrum bilgisi elde edilmesi ile piksellerin değişimi, benzerlikleri ve farklılıklarının algılanabilmesi sonucu, bölgelerin sınıflandırılması standart görüntü algılayıcılarına oranla çok daha yüksek bir başarımla sağlanabilmektedir [1]. Hiperspektral verilerde sınıflandırma, her bir pikselin belirli sınıflara atanması veya sınıflara ayrılması amacıyla gerçekleştirilmiştir. Bu süreç, piksel tabanlı bir yaklaşımı takip eder ve her piksele bir sınıf etiketi atanır. Sınıflandırma için öncelikle bir eğitim veri kümesi oluşturulur. Bu veri kümesi, hiperspektral görüntülerin pikselleri ve ilgili sınıf etiketleri içerir. Eğitim veri kümesi, farklı nesne veya yüzey sınıflarını temsil eden örneklerden oluşur. Daha sonra, sınıflandırma algoritması kullanılarak özellik çıkartma ve sınıflandırma işlemi gerçekleştirilir. Özellik çıkartma, hiperspektral verinin boyutunu azaltmak veya önemli bilgileri vurgulamak için yapılır. TBA yöntemi kullanılarak verinin boyutu azaltılmıştır. Spektral imza analizi ile özellikler elde edilmiştir. Elde edilen özellikler, bir sınıflandırma algoritması kullanılarak sınıflandırma işlemine tabi tutulur. Bu algoritmalar arasında projede destek vektör makineleri (DVM) yöntemi kullanılmıştır. DVM, özellikleri analiz ederek ve sınıf ayrımı için bir karar kriteri kullanarak pikselleri ilgili sınıflara atar. Sınıflandırma sonucunda, hiperspektral görüntünün her pikseli bir sınıf etiketiyle ilişkilendirilir. Bu şekilde, nesnelerin veya yüzeylerin farklı sınıflara ait olduğu haritalar elde edilebilir. Bu haritalar, tarım alanları, ormanlar, su kaynakları veya kentsel bölgeler gibi farklı yüzey özelliklerini görselleştirmek veya analiz etmek için kullanılmıştır. Hiperspektral verilerde sınıflandırma, piksel tabanlı bir yaklaşımı benimser ve veriyi belirli sınıflara ayırmayı hedefler. Özellik çıkartma ve sınıflandırma algoritmalarının birleşimi, yüksek doğruluk ve ayrıntılı sınıflandırma sonuçları elde etmek için kullanılmıştır.

1. AMAÇ

Bu proje, DVM (Destek Vektör Makineleri) ve TBA (Ana Bileşen Analizi) kullanılarak hiperspektral veriler üzerinde sınıflandırma gerçekleştirmeyi amaçlamaktadır. DVM, veri noktalarını sınıflandırmak için kullanılan bir makine öğrenimi tekniğidir, TBA ise verilerdeki değişkenliği yakalamak için kullanılan bir boyut azaltma tekniğidir. Bu projede hiperspektral verilerin boyutsallığını azaltmak için TBA uygulanmakta ve DVM algoritması kullanılarak sınıflandırma yapılmaktadır.

2. GİRİŞ

Hızlı nüfus artışı, kısıtlı doğal kaynaklar ve çevre kirliliği, yeryüzü hakkında hızlı ve doğru bilgiye olan ihtiyacın artmasına neden olmaktadır. Hava ve uydu görüntüleri kullanılarak bu problemler çözülmeye çalışılmaktadır. Bu görüntüleri elde etmek amacı ile özellikle havadan ve uzaydan cisimlerin özelliklerini kaydetme ve inceleme tekniği olarak tanımlanan uzaktan algılama yöntemleri önem kazanmaktadır [1]. Uzaktan algılama yöntemleri, hava ve uydu görüntülerini kullanarak, doğal kaynakların yönetimi, çevre izleme, tarım verimliliği, iklim değişikliği analizi gibi konularda bilimsel ve veriye dayalı kararlar almayı mümkün kılmaktadır. Bu sayede, çevresel sorunların anlaşılması, takibi ve yönetimi konusunda daha etkili çözümler üretilebilmektedir. Uzaktan algılamada, hiperspektral verilerin büyük bir etkisi vardır. Hiperspektral görüntüler, geniş bir elektromanyetik spektrumda çok sayıda dar bantlı spektral kanalı içeren görüntülerdir. Bu kanallar, nesnelerin spektral imzalarını ayrıntılı bir şekilde yakalar ve piksel değerlerinin dalgaboyuna göre değişimini inceler.

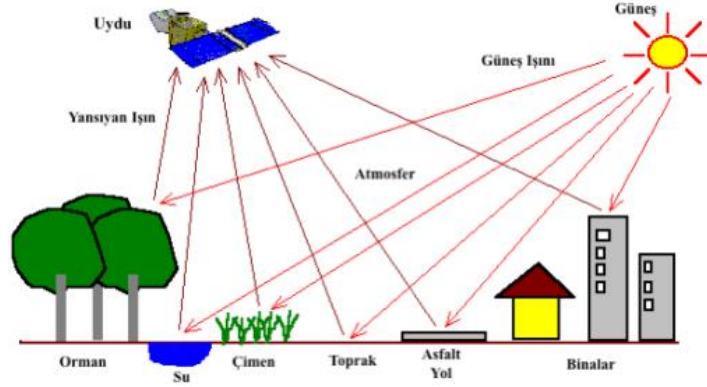
Bu proje kapsamında Hiperspektral veriler, Indian Pines ve Pavia Universitysinin veri kümesi üzerinde DVM kullanarak sınıflandırma görevlerini gerçekleştirmek için yürütülmektedir. Veri kümesi bölüm 4'te detaylı şekilde verilmiştir. DVM (Destek Vektör Makineleri) algoritması, veri noktalarını sınıflandırmak için bir makine öğrenimi tekniği olarak kullanılmıştır. Hiperspektral veriler genellikle çok yüksek boyutlu veri kümeleridir, çünkü her bir piksel spektral bantlar boyunca bir dizi değere sahiptir. Bu yüksek boyutluluk, veri analizini zorlaştırabilir ve sınıflandırma veya analiz algoritmalarının performansını olumsuz etkileyebilir. Bunun önüne geçebilmek adına projede TBA (Temel Bileşen Analizi) kullanılmıştır. TBA boyutsallığını azaltarak verilerdeki değişkenliği yakalamak için kullanılan bir boyutsallık azaltma tekniğidir. Bu kullanılan tekniklerin detaylı anlatımı bölüm 3'te verilmiştir. Bu yöntemler sonucu elde edilen deneysel sonuçlar ve yorumlar bölüm 5 ve 6'da verilmiştir. Rapor, proje sürecinin ve elde edilen sonuçların ayrıntılı bir açıklamasını sağlar.

3. GENEL KISIMLAR

3.1 UZAKTAN ALGILAMA

Uzaktan algılama, fiziksel nesneler arasındaki konumsal, spektral ve zamansal ilişkilerin belirlenmesini sağlar. Nesnelerle temas etmeden uydu veya uçaklar üzerindeki sensörler aracılığıyla yeryüzü ölçümü yaparak veriye ulaşmamıza yardımcı olur [2]. Nesneye ait bu bilgiler elektromanyetik spektrumun belirli bant aralıklarında hava araçları veya uydular aracılığı ile ölçülmesi sonucunda elde edilirler. Bu sayede yeryüzünde fiziksel olarak ulaşılması zor bölgelerin görüntülenmesi veya nesneye ait gözle görülemeyen özelliklerin elde edilmesi sağlanmış olur. Bir nesneye ulaşan elektromanyetik enerjinin bir kısmı nesne tarafından emilmekte bir kısmı ise nesneden yansımaktadır. Uzaktan algılamada nesnelerden bilgi elde etmenin temelinde bu nesnelerden yansıyan elektromanyetik enerjinin hava araçlarında veya uydularda bulunan algılayıcılar ile ölçülmesi

yatmaktadır. Ölçülen bu değerler sayısal verilere dönüştürülerek yer istasyonlarına aktarılır. Şekil 2.1’de uzaktan algılamanın genel şeması gösterilmektedir.



Şekil 3.1: Uzaktan algılama genel şeması [3]

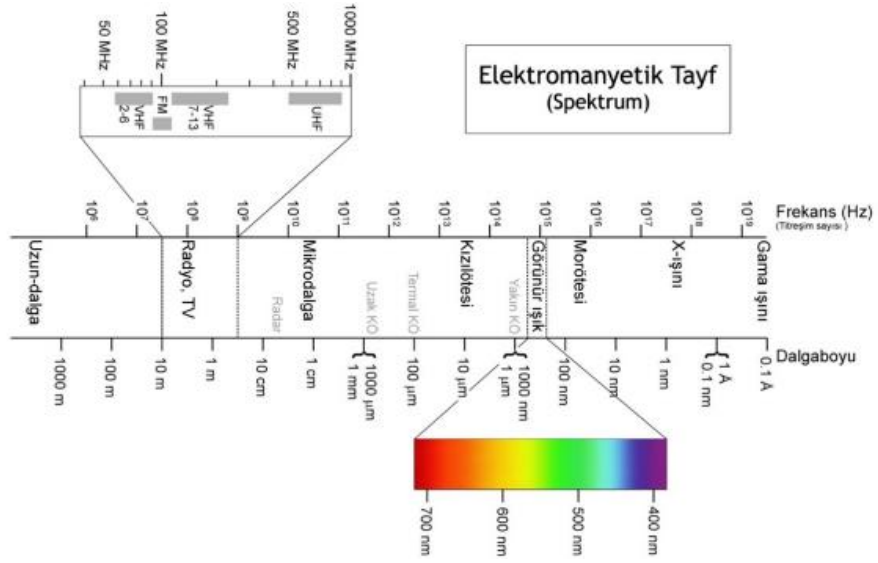
Nesnelerden yansıyan elektromanyetik enerji her bir maddeye özgü olmaktadır. Bu nedenle nesneden yansıyan enerjinin bu kısmı, elektromanyetik spektrumda o nesneye ait spektral imzasını oluşturmaktadır. Spektral imza her bir maddeye özgü olduğundan nesnelerin ayırt edilmesi ve tanınması mümkün hale gelmektedir.

Uzaktan algılama alanındaki gelişmeler arttıkça elde edilen verilerin çeşitliliği artmakta ve dolayısı ile bu verilerin kullanım alanları ve amaçları da çeşitlilik kazanmaktadır. Uzaktan algılama teknolojisi jeoloji, tarım, ormancılık, uzay bilimleri, atmosfer gözlemleri ve askeri uygulamalar gibi birçok alanda aktif olarak kullanılmaktadır.

3.2 HİPERSPEKTRAL GÖRÜNTÜLEME

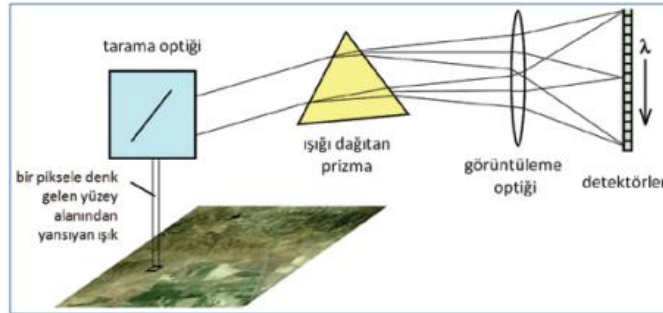
Hiperspektral görüntüleme yüzey materyallerinden yansıyan enerjinin dar ve bitişik çok sayıda dalga boyu bandında ölçümüdür. Hiperspektral alıcıların en belirgin özelliği çok bantlı (multispektral) görüntülerden farklı olarak çok daha fazla banda sahip olmalarıdır. Ancak, alıcıları hiperspektral yapan özellik bant sayısından çok, bantların ne kadar dar ve birbirine ne kadar bitişik olduğudur [4].

Hiperspektral görüntüleme, elektromanyetik spektrumun optik bölgesi olarak adlandırılan 0.4 μ m ile 14 μ m dalga boyunu içeren geniş bir aralık kullanır. Elektromanyetik spektrum dalga boyları Şekil 2.2 'de gösterilmektedir. Ancak hiperspektral görüntüleme için daha sık kullanılan dalga boyu aralığı 0.4 μ m ile 2.5 μ m'dir. Bu aralık, elektromanyetik spektrumun görünür bölgesi, yakın kızılötesi ve orta kızılötesi bölgelerini kapsar. Hiperspektral görüntülemede, genellikle bant genişlikleri 10nm ila 20nm arasında değişen yüzlerce dalga boyunda ölçümler yapılır.



Şekil 3.2: Elektromanyetik spektrum [5]

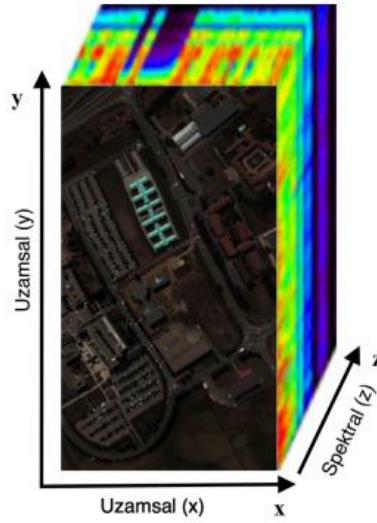
Hiperspektral görüntüleme, spektrometre ya da spektoradyometre adı verilen cihazlarla yapılmaktadır. Bu cihazlar yardımı ile nesnelerden yansıyan ışık nanometreler seviyesindeki dalga boyu aralıkları ile ölçülebilir. Spektrometre içindeki prizmaya benzer optik bileşen, spektrometreye ulaşan ışığı dar ve birbirini takip eden çok sayıda dalga boyu bandına ayırıştırır. Yüzeyden gelen enerji her bir bant için ayrı bir algılayıcı ile ölçülür. Spektrometre, 0,01 mikrometreye kadar dar aralıklarda ve 0,4 – 2,5 μ m gibi geniş bir spektrum için ölçüm yapabilir [6].



Şekil 3.3: Hiperspektral görüntüleme düzeneği [6]

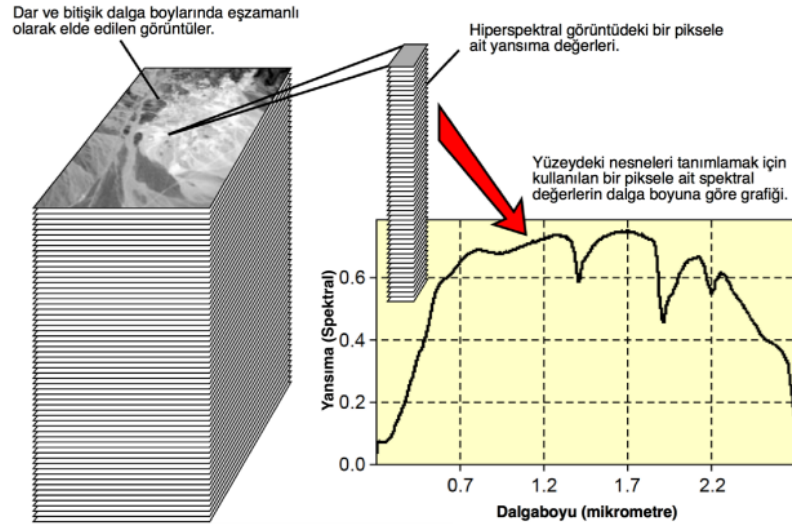
3.2.1 Hiperspektral Görüntüler

Elde edilen hiperspektral görüntüler, nesnelere ait spektral bilgileri içeren bir veri türüdür. Bu görüntüler, hiperspektral görüntü (HSG) olarak adlandırılır ve üç boyutlu bir veri içerir. Bu veri, hiper-küp olarak bilinir çünkü uzamsal çözünürlüğü ifade eden ilk iki boyutun yanı sıra spektral bant sayısını ifade eden üçüncü bir boyut içerir. Şekil 2.4'te bir hiperspektral görüntü küpünün örneğini göstermektedir.



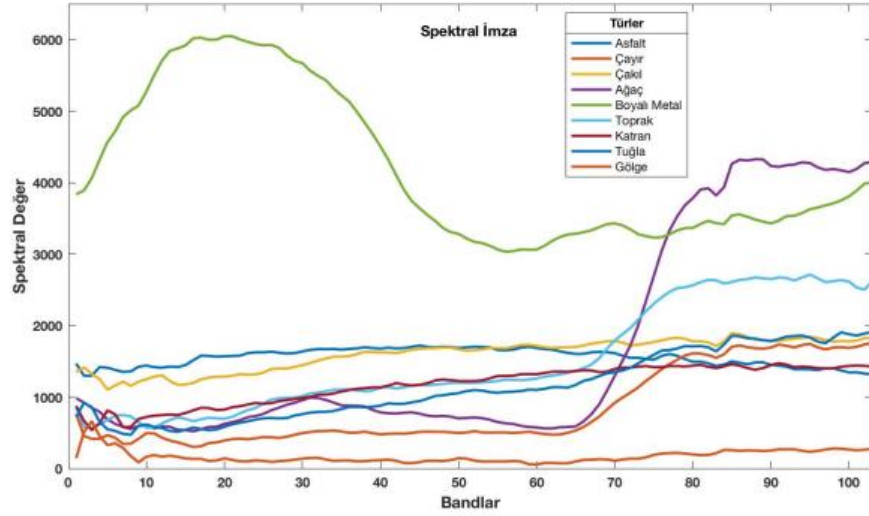
Şekil 3.4 : Hiperspektral görüntü küpü [1]

Hiperspektral görüntüde her piksel için elde edilen spektral değerler, her madde için ayırt edici bir özellik taşır. Aynı türden nesneler veya maddelerin yüzeyinde ölçülen spektral değerler tam olarak aynı olmasa da birbirlerine çok yakın değerler içerir. Bu nedenle farklı türden maddeler için ölçülen spektral değerler, o maddeye özgü ve bir tür parmak izi niteliği taşır. Bu spektral değerler, maddeye ait "spektral imza" olarak adlandırılır. Şekil 2.5'te bir hiperspektral görüntünün bir pikselindeki maddeye ait spektral yansımaya değerlerinden oluşan spektral imza grafiğini göstermektedir.



Şekil 3.5 : Hiperspektral görüntüdeki bir pikseldeki maddeye ait spektral imza [7]

Şekil 2.6'da Pavia University [8] hiperspektral görüntüsündeki farklı her bir sınıfa ait birer piksel için spektral imza grafiği örnekleri gösterilmiştir. Şekilde 2.6' da görüldüğü üzere her bir madde türü için oluşan spektral imza grafiği sınıf türüne özgü oluşmuştur. Hiperspektral görüntüdeki nesnelerin sınıflandırılması gibi birçok uygulamada spektral imzalardan yararlanılmaktadır.



Şekil 3.6 : Pavia University veri kümesindeki her bir sınıfa ait birer piksel için spektral imza örnekleri [7]

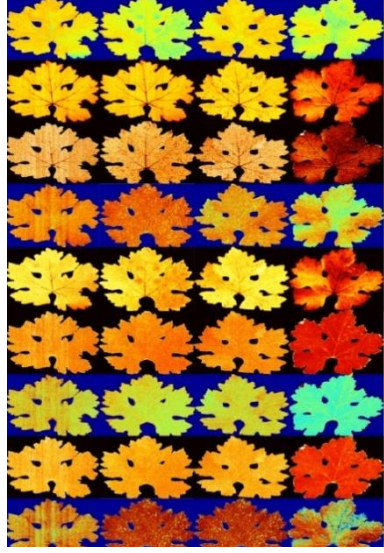
Hiperspektral görüntüler, zengin ve anlamlı bilgiler içerdiği için çeşitli alanlarda aktif olarak kullanılmaktadır. Özellikle nesneleri belirleyici bilgileri içermesi sayesinde sınıflandırma çalışmalarında yaygın bir şekilde kullanılır. Hiperspektral görüntülerin kullanıldığı uygulama alanları şunlardır:

- Uzaktan algılama: Toprak özelliklerinin, bitki sağlığının, su kalitesinin, hava kirliliğinin ve diğer çevresel faktörlerin izlenmesi.
- Tarım ve ormancılık: Bitki türü sınıflandırması, bitki hastalıklarının tespiti, verim tahmini ve tarım alanlarının yönetimi.
- Tıp ve biyoteknoloji: Hastalık teşhisi, tıbbi görüntüleme, biyomedikal araştırmalar ve genetik çalışmalar.
- Savunma ve güvenlik: Hedef tespiti, hedef tanıma, kimyasal madde tespiti ve keşif amaçlı kullanım.
- Jeoloji ve madencilik: Maden arama, mineral tanımlama, jeolojik haritalama ve maden çıkarımı çalışmaları.
- Uzay bilimleri: Uzay araştırmaları, gök cisimlerinin karakterizasyonu ve gezegenlerin incelenmesi.
- Çevresel çalışmalar: Su kalitesi analizi, kirlilik tespiti, çevresel değişimlerin izlenmesi ve doğal kaynakların yönetimi.

Bu alanlar sadece hiperspektral görüntülerin kullanıldığı bazı örnek uygulama alanlarıdır. Hiperspektral görüntüleme teknolojisi, birçok disiplin ve sektörde farklı amaçlarla kullanılmaktadır.

Hiperspektral görüntülere birkaç örnek verecek olursak :

1. Asmada Hiperspektral Floresans Cabernet Sauvignon asmalarında klorofil ve nitrojeni ölçmek için bir Resonon Pika L hiperspektral görüntüleyici kullanıldı [9].



Şekil 3.7 : Bu çalışmada yeni vejetasyon indeksleri ile tahmin edilen yaprak seviyesindeki SPAD, LNC ve ChiP parametreleri haritası [10]

2. Bir istilacı yabancı ot istilasını belirlemek ve bulmak için bir Resonon havadan hiperspektral görüntüleme sistemi kullanıldı [9].



Şekil 3.8 : İstilacı ot haritalama [11]

Literatür Özeti

Hiperspektral görüntülerin sınıflandırılması, hiperspektral görüntülemedeki en popüler çalışma alanlarından biridir. Hiperspektral görüntü (HSG) birçok dar dalga boyu bandından oluştuğu için çok fazla miktarda veri içermektedir. Verinin fazla olması işlem bakımından sınıflandırıcının performansını etkilemektedir. Bu nedenle incelenen önceki çalışmalarda hiperspektral görüntü boyutunun azaltılmasının sınıflandırma başarımı kadar önemli olduğu görülmüştür. Hiperspektral görüntü sınıflandırmada büyük miktardaki spektral bant sayısı, spektral imzanın uzamsal değişebilirliği doğru sınıflandırmanın yüksek maliyeti ve verinin kalitesi gibi konular bu alandaki temel sorunlardır [12]. HSG' den çeşitli yöntemlerle elde edilen uzamsal özniteliklerin kullanılması ile sınıflandırma başarısının artırıldığı çalışmalarda literatürde yer almaktadır. Destek Vektör Makineleri (Support Vector Machines) [13], yüksek boyuttaki veri ve az sayıdaki eğitim kümesi ile spektral bilgileri kullanarak yüksek başarımlı sonuçlar sağlamaktadır [14], [15].

Uzamsal özniteliklerin komşu piksellerin ortalaması veya standart sapması alınarak elde edildiği ve spektral özniteliklerle birlikte kullanıldığı çalışmada DVM için farklı çekirdeklerin birleştirilmesi ile sınıflandırma yapılmıştır [16]. HSG' den bant çıkarması ile elde edilen bant görüntülerine açma-kapama gibi morfolojik işlemler uygulanarak özniteliklerin elde edilerek sınıflandırılması [17] ve [18]'te önerilmiştir. Markov Rastgele Alanlar [19], uzamsal komşuluğa bağlı olarak doku özniteliklerini çıkarmak için kullanılan ayrı bir yöntem olarak HSG sınıflandırmada [20] ve [21] 'de kullanılmıştır. [22] ve [23] çalışmalarında ise bant çıkarma boyut indirgeme yöntemi ile çıkarılan bantlara Gabor Filtresi uygulanarak uzamsal özniteliklerin kullanıldığı sınıflandırma önerilmiştir. Örüntü tanımada sıklıkla kullanılan Yerel İkili Örüntü (Local Binary Patterns) HSG sınıflandırmada da başarılı sonuçlar vermiştir [24].

Temel Bileşen Analizi (TBA) [25], özellikle yüksek boyutlu verilerde kullanılan denetimsiz ve doğrusal bir boyut indirgeme yöntemidir. TBA' inde temel amaç, çok değişkenli verilerin boyutsallığını azaltmak ve mümkün olduğunca ilgili bilgileri yüksek oranda korumaktır. Bunu yaparken en bilgilendirici ve ayırt edici özellikleri bulmaya çalışır. Bunun için ilintisi (kovaryans) yüksek olan özellikleri gruplayıp değişintisi yüksek olan özellikleri çıkarmaktadır. 10 Yüksek boyutlu verideki özellikler doğrusal bir yaklaşım ile farklı bir uzaya taşınır ve bu uzaydaki değişintisi daha yüksek olan özellikler öncelikli temel bileşenler olmaktadır

HSG sınıflandırmadaki literatürde son zamanlarda yapılan çalışmalar başarımın daha da artırılması, az sayıda eğitim kümesi ile yüksek başarımların elde edilmesi, uzamsal bilgilerden yararlanılması ve boyut indirgeme gibi alt başlıklar halinde sıralanabilir. Literatürde son yapılan çalışmalarda genellikle sınıflandırmalarda eğitim kümesindeki örnek sayıları DVM gibi geleneksel sınıflandırma yöntemlerine yer verilmiştir. Ayrıca DVM gibi geleneksel sınıflandırma yöntemlerinin eğitim sürelerinin diğer sınıflandırma örneklerine nazaran çok daha fazla olduğu görülmüştür. Literatürde HSG sınıflandırmada başarımın artırılmasının yanında az sayıda eğitim verisi kullanarak başarımın yüksek olduğu sınıflandırma çalışmaları da yapılmıştır.

3.3 BOYUT İNDİRGEME

Hiperspektral veriler, çok sayıda spektral bant içerdiği için yüksek boyutta verilere sahiptir. Bu durum, sınıflandırma ve analiz çalışmalarında bazı zorluklar ortaya çıkarır. Boyut indirgeme, bu yüksek boyutlu verilerdeki bilgiyi daha anlamlı ve kullanışlı hale getirmek amacıyla uygulanan bir yöntemdir.

Boyut indirgeme işlemi, orijinal veri boyutunu azaltarak, veri setinin temel özelliklerini korumak ve anlamlı bilgileri koruyarak veri analizini daha etkili hale getirmek için kullanılır. Temel amacı, verinin karmaşıklığını azaltmak ve veri setinin içerdiği bilgiyi daha anlaşılır ve kullanılabilir hale getirmektir.

Bant seçme yöntemi, boyut indirgeme için kullanılan bir yöntemdir. Bu yöntemde, orijinal veri setindeki en bilgilendirici ve ayırt edici bantlar seçilir ve diğer bantlar atılır veya ihmal edilir. Bu şekilde, veri boyutu azaltılarak, sınıflandırma ve analiz için yeterli bilgi elde edilir.

Bant çıkarma yöntemi ise farklı bir yaklaşımdır. Bu yöntemde, veri doğrusal veya doğrusal olmayan yöntemlerle farklı bir uzaya taşınır ve bu yeni uzayda boyut indirgeme gerçekleştirilir. Bu yöntemde, elde edilen veri seti, orijinal veriden farklı olabilir, ancak temsil ettiği bilgilerin anlamlı ve kullanışlı olduğu düşünülür.

Her iki yöntemde de, yüksek kovaryansa sahip bantlar gruplanarak en ayırt edici ve bilgilendirici bantlar seçilir. Bu sayede, boyut indirgeme işlemi gerçekleştirilir ve daha az boyutlu bir veri seti elde edilir.

Boyut indirgeme, hiperspektral verilerin sınıflandırma ve analiz süreçlerinde önemli bir adımdır. Doğru boyut indirgeme yöntemi seçilerek, veri setindeki bilgi kaybı minimize edilirken, sınıflandırma performansı artırılabilir ve veri analizi daha etkili hale getirilebilir.

3.3.1 Temel Bileşen Analizi (TBA)

Temel Bileşen Analizi, veri kümesindeki değişkenlik yapısını anlamak ve boyut indirgeme yapmak için kullanılan bir istatistiksel yöntemdir. Veri kümesindeki yüksek boyutlu değişkenler arasındaki ilişkiyi anlamak için kullanılan bir gözetimsiz analiz tekniğidir.

TBA verideki gerekli bilgileri ortaya çıkarmada oldukça etkili bir yöntemdir. TBA'nın arkasında yatan temel mantık çok boyutlu bir veriyi, verideki temel özellikleri yakalayarak daha az sayıda değişkenle göstermektir [26].

TBA'nın başlıca kullanım amaçları şunlardır:

- Boyut indirgeme: Yüksek boyutlu veri kümelerindeki değişken sayısını azaltmak ve veri kümesini daha yönetilebilir hale getirmek.
- Veri keşfi: Veri kümesindeki değişkenler arasındaki ilişkileri anlamak, veri kümesinin yapısını görselleştirmek ve anlam çıkarmak.
- Gürbüzlük (Robustness): Gürültülü veya eksik veri kümelerinde bile değişkenlik yapısını yakalama yeteneği.

Bu projede TBA kullanılma nedeni, hiperspektral verinin boyutunu azaltmak ve veri kümesindeki önemli bileşenleri belirlemektir. Hiperspektral veriler genellikle yüksek boyutlu olabilir ve her bir spektral bant bir değişken olarak temsil edilebilir. Bu nedenle, TBA, veri kümesinin boyutunu azaltarak önemli bilgileri korumayı ve sınıflandırma performansını artırmayı hedefler. TBA'nın uygulanmasıyla, hiperspektral verideki değişkenlik yapısı daha iyi anlaşılabilir ve sınıflandırma için daha iyi bir temsil elde edilebilir.

3.4 MODEL EĞİTİMİ

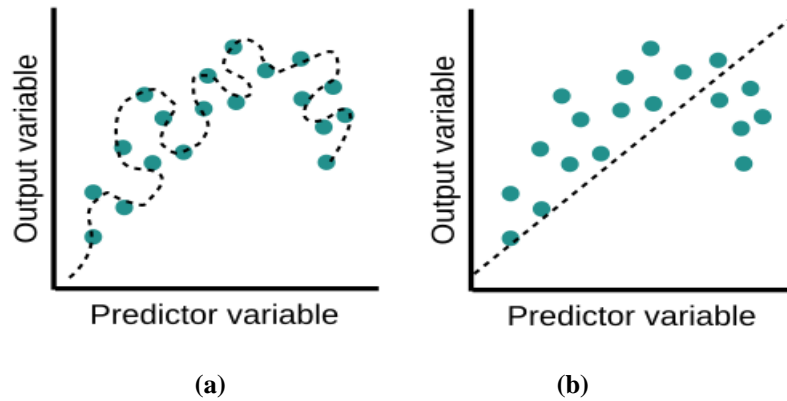
Makine öğreniminde gelecek veriler hakkında tahmin yapabilmek için verilerimizi eğitim verileri (train) ve test verileri olmak üzere iki alt kümeye ayırıyoruz. Modelimizi eğitim verilerinden elde edilen örüntülere göre oluşturuyoruz. Bu işlem sonucunda iki şeyden biri olabilir; modelimiz aşırı öğrenebilir veya eksik öğrenebilir. Bu durumda modelimiz yeterli öngörde bulunamayacak ve tahminlerimizde hata oranı yüksek olacaktır [27] .

3.4.1 Aşırı Öğrenme (Overfitting)

Overfitting, makine öğrenmesi ve istatistiksel modellemede bir modelin eğitim verilerine aşırı derecede uyum sağlaması durumudur. Eğer modelimiz, eğitim için kullandığımız veri setimiz üzerinde gereğinden fazla çalışıp ezber yapmaya başlamışsa ya da eğitim setimiz tek düze ise overfitting olma riski büyük demektir [30]. Bu durumda model, eğitim verilerini çok iyi tahmin edebilirken, yeni ve görmediği verilerde hata yapma eğilimindedir.

3.4.2 Eksik Öğrenme (Underfitting)

Underfitting, makine öğrenmesi ve istatistiksel modellemede bir modelin eğitim verilerine yeterince uyum sağlamaması durumudur. Bu durumda model, hem eğitim verilerinde hem de yeni verilerde düşük performans gösterir. Aşırı öğrenmenin aksine, bir model yetersiz öğrenmeye sahipse, modelin eğitim verilerine uymadığı ve bu nedenle verilerdeki trendleri kaçırdığı anlamına gelir [30]. Bahsedilen iki durum da şekil 3'te örnekleri gösterilmiştir.



Şekil 3.9 : Aşırı Öğrenme (a) ve (b) Eksik Öğrenme [30]

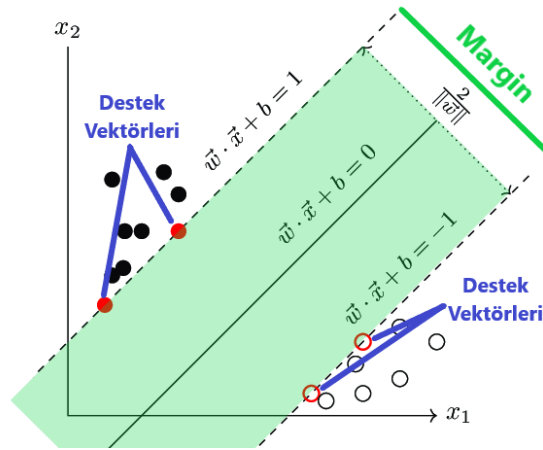
3.5 SINIFLANDIRMA YÖNTEMİ

HSG sınıflandırma için geleneksel sınıflandırma yöntemlerinden olan Destek Vektör Makineleri kullanılmıştır.

3.5.1 Destek Vektör Makineleri

DVM, hiperspektral veri analizi alanında sınıflandırma ve regresyon için kullanılan bir makine öğrenimi algoritmasıdır. DVM, veri noktalarını iki veya daha fazla sınıfa ayırmayı amaçlar ve doğrusal veya doğrusal olmayan sınıflandırma problemlerini çözebilir.

DVM'nin temel amacı, veri noktalarını farklı sınıflara en iyi ayıran bir karar sınırı (hiper düzlem) bulmaktır. DVM, sınıflar arasında maksimum marj ayırımı sağlamaya çalışır. Kenar boşluğu, karar sınırı ile en yakın veri noktaları arasındaki mesafeyi ifade eder. Bu marjı en üst düzeye çıkararak DVM, genelleme yeteneğini geliştirir. DVM'nin çalışma prensibi, İki etiketlenmiş sınıfı ayıran en iyi eğriyi bulmaya çalışır. Algoritma çizilecek doğrunun iki sınıfında elemanlarına en uzak yerden geçecek şekilde optimize eder. Hiçbir parametre almayan bir sınıflayıcıdır [28]. Şekil 3'te gösterilmiştir.



Şekil 3.10 : SVM'in Destek Vektörleri ve Margin Görüntülerinin Grafiksel Görüntüsü [27]

Şekilde de görüldüğü gibi siyahlar ve beyazlar gibi farklı sınıflara ait veriler arasında bir ayırım yapmak için bir doğru çizer. Çizilen doğrunun ± 1 arasında kalan yeşil bölgeye "Margin" denir. Margin, iki sınıf arasındaki ayırımı belirler ve ne kadar geniş olursa, sınıfların daha iyi bir şekilde ayrıştırıldığı anlamına gelir. Yani, geniş bir Margin, daha fazla verinin doğru bir şekilde sınıflandırılmasını sağlar.

Formülle bakacak olursak;

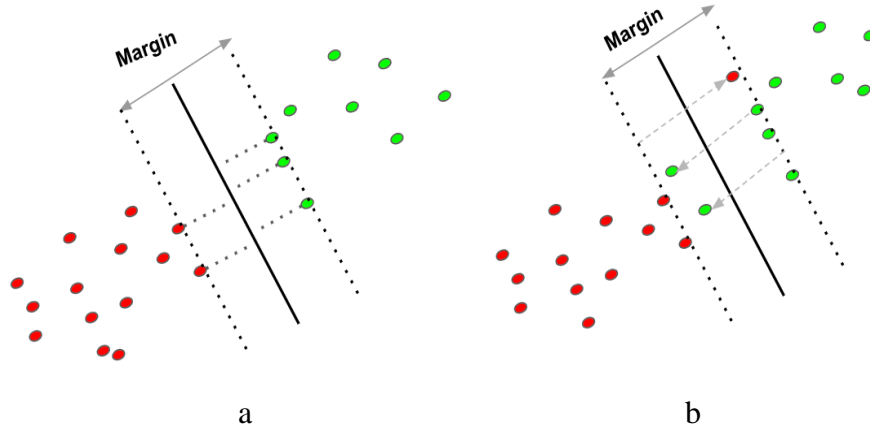
$$\hat{y} = \begin{cases} 0 & \text{if } \mathbf{w}^T \cdot \mathbf{x} + b < 0, \\ 1 & \text{if } \mathbf{w}^T \cdot \mathbf{x} + b \geq 0 \end{cases}$$

\mathbf{w} ; ağırlık vektörü (θ_1), \mathbf{x} ; girdi vektörü, b ; sapmadır (θ_0).

Yeni bir deęer için ıkan sonu 0'dan kkse, beyaz noktalara daha yakın olacaktır. Tam tersi, ıkan sonu 0'a eit veya bkse, bu durumda siyah noktalara daha yakın olacaktır.

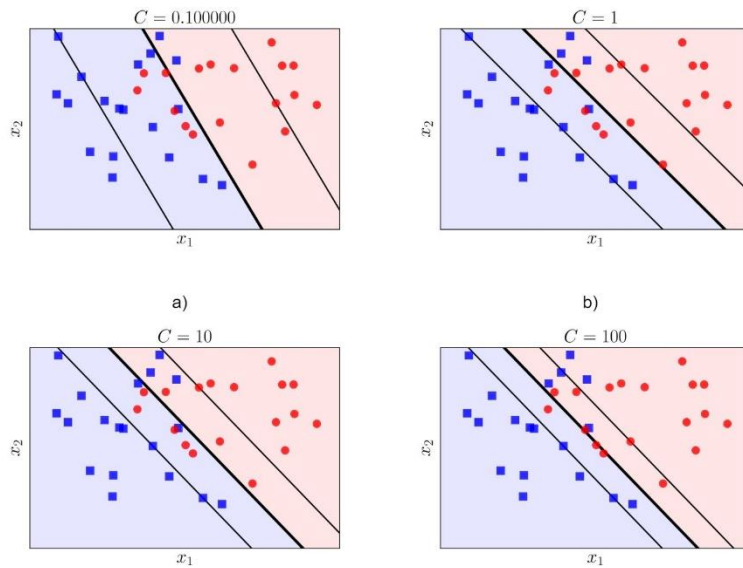
3.5.2 Hard Margin vs Soft Margin

Marginimiz her zaman bu ekilde olmayabilir. Bazen rneklerimiz Margin blgesine girebilir. Buna Soft Margin denir. Hard Margin, verimiz doęrusal olarak ayrılabilirse alıřır ve aykırı deęerlere karřı ok duyarlıdır. Bu yzden bazı durumlarda Soft Margin'i tercih etmemiz gerekebilir.



ekil 3.11 : (a) Hard Margin (b) Soft Margin [28]

İkisi arasındaki dengeyi DVM ierisindeki C hiperparametresi ile kontrol edebiliriz. C ne kadar bkse Margin o kadar dardır. Ayrıca model ařırı ğrenme olursa C'yi azaltmamız gerekir. Biz proje kapsamında C parametresini = 150 olarak tanımladık.



ekil 3.12 : DVM'deki C Parametresinin Etkisi [29]

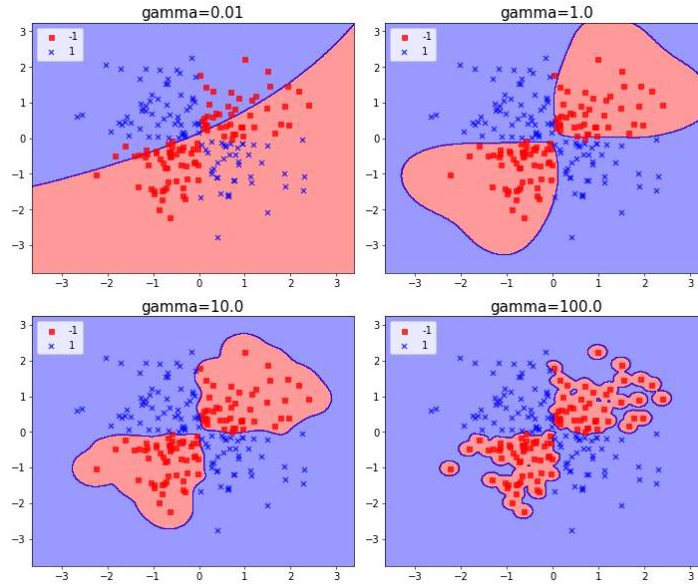
3.6 KERNEL TRICK

DVM'de Kernel Trick, sınıflar arasında doğrusal olarak ayrılabilmesi mümkün olmayan veri setlerini doğrusal olarak ayrılabilir hale getirmek için kullanılan bir yöntemdir. Bu yöntem, verileri daha yüksek boyutlu uzaylara projekte ederek non-lineer karar sınırlarını elde etmeyi sağlar. Kernel Trick, DVM'nin esnekliğini artırarak daha karmaşık veri yapılarının sınıflandırılmasını mümkün kılar. Bu yöntemlerden en çok kullanılan ve bu projede de kullandığımız yöntem Gaussian RBF (Radial Basis Function) Kernel'dir.

3.6.1 Gauss Çekirdekleme (Gaussian Rbf-Radial Basis Function/Kernel)

Gauss Çekirdekleme (Gaussian RBF Kernel); Destek Vektör Makinaları belirlenip, her bir noktanın belirli noktalara ne ölçüde benzediğini normal dağılım yöntemi ile hesaplayıp sınıflandırma işlemini gerçekleştirdiğimiz çözümleme modelidir. Bu modelde dağılım genişliği hiper parametre değeri (Gama – Gamma : γ) ile kontrol edilmektedir.

Çözümleme modele uygulanırken; model Aşırı öğrenme olmasına karşın gama değeri düşürülür, Eksik öğrenme olmasına karşın ise gama değeri artırılır [29]. Buna örnek Şekil 6'da verilmiştir.



Şekil 3.13 : Gamma Hiperparametresinin etkisi [29]

3.7 METRİKLER

Sınıflandırma işleminin başarımını değerlendirebilmek için veya farklı sınıflandırıcıların başarımlarını kıyaslayabilmek için literatürde genellikle karışıklık matrisi (confusion matrix) [30] kullanılmaktadır. Başarımın ölçülmesinde kullanılan karışıklık matrisinde sınıflandırma sonucu her bir sınıf için doğru sınıflandırılmış ve

yanlış sınıflandırılmış örneklerin sayısı veya oranları gösterilmektedir. Bu tablodan yararlanarak doğruluk başarımının değerlendirilmesinde kullanılan doğruluk (accuracy), hata oranı (error rate), kesinlik (precision), duyarlılık veya anma (sensitivity veya recall), ve F-ölçütü (F-score) değerleri hesaplanabilir [30]. İki sınıfa sahip bir verinin sınıflandırılması sonucu oluşturulan karışıklık matrisinin genel yapısı Tablo 2.1’ de gösterilmiştir.

	Pozitif Gerçek Durum	Negatif Gerçek Durum
Pozitif Tahmin	Gerçek Pozitif (TP)	Yanlış Pozitif (FP)
Negatif Tahmin	Yanlış Negatif (FN)	Gerçek Negatif (TN)

Tablo 3.1 : iki sınıflı sınıflandırma için karışıklık matrisi

Temsil edilen değerleri açıklayacak olursak;

Gerçek Pozitif (TP): Gerçekte pozitif bir durumu ifade eden ve sınıflandırıcı tarafından pozitif olarak tahmin edilen örnekleri ifade eder.

Gerçek Negatif (TN): Gerçekte negatif bir durumu ifade eden ve sınıflandırıcı tarafından negatif olarak tahmin edilen örnekleri ifade eder.

Yanlış Pozitif (FP): Gerçekte negatif bir durumu ifade eden ve sınıflandırıcı tarafından pozitif olarak tahmin edilen örnekleri ifade eder.

Yanlış Negatif (FN): Gerçekte pozitif bir durumu ifade eden ve sınıflandırıcı tarafından negatif olarak tahmin edilen örnekleri ifade eder [31].

Karışıklık matrisi, bu dört kategorinin sayısal değerlerini içeren kare bir matristir. Her bir sınıf için gerçek değerler ve tahmin değerlerinin kesişimlerindeki hücrelerde ilgili sayılar yer alır. Karışıklık matrisi, sınıfların doğru ve yanlış sınıflandırma sayılarını görsel olarak sunar ve modelin hangi sınıflarda daha iyi veya daha kötü performans gösterdiğini anlamamıza yardımcı olur.

Bu bilgilerden yararlanarak aşağıdaki Doğruluk, Hata Oranı, Kesinlik, Duyarlılık, başarımlar değerlendirme yöntemleri yazılabilir.

$$\text{Doğruluk (Accuracy)} = \frac{TP + TN}{TP + FN + FP + TN}$$

$$\text{Kesinlik (Precision)} = \frac{TP}{TP + FP}$$

$$\text{Duyarlılık (Sensitivity)} = \frac{TP}{TP + FN}$$

$$\text{Özgüllük (Specificity)} = \frac{TN}{FP + TN}$$

3.8 SINIFLANDIRMA BAŞARIM ÖLÇÜTLERİ :

3.8.1 Genel Doğruluk (OA)

Genel doğruluk (Overall Accuracy), bir sınıflandırma modelinin tüm sınıflar üzerindeki doğruluk oranını temsil eder. Doğru sınıflandırılan örneklerin toplam örnek sayısına oranı olarak hesaplanır. Özellikle dengeli sınıflandırma problemlerinde kullanılan bir ölçüttür. Genel doğruluk, tüm sınıfları eşit derecede önemli olarak kabul eder ve sınıf dengesizliklerini göz ardı eder. Örneğin, 100 örneğin 90'ını doğru sınıflandırırsak, Genel doğruluk değeri %90 olacaktır.

3.8.2 Ortalama Doğruluk (AA)

Ortalama doğruluk (Average Accuracy), sınıflara özgü doğruluk oranlarının ortalamasıdır. Her sınıfın doğruluk oranı hesaplanır ve bu oranların ortalaması alınır. Sınıflar arasındaki dengesizlikleri dikkate alan bir ölçüttür. Sınıf bazında performansı ayrıntılı bir şekilde değerlendirmek için kullanılır. Örneğin, bir sınıfın doğruluk oranı %80, diğer bir sınıfın doğruluk oranı ise %90 ise, Ortalama doğruluk değeri bu iki sınıfın doğruluk oranlarının ortalaması olan %85 olacaktır.

Genel doğruluk genel bir performans ölçütüdür ve tüm sınıfları eşit derecede önemli kabul ederken, Ortalama doğruluk sınıflar arasındaki farkları ortaya koyar. Her iki ölçüt de sınıflandırma modelinin performansını değerlendirmek için kullanılır, ancak problem bağlamına ve sınıf dengesine bağlı olarak hangi ölçütün daha uygun olduğu değişebilir.

4. VERİ KÜMELERİ

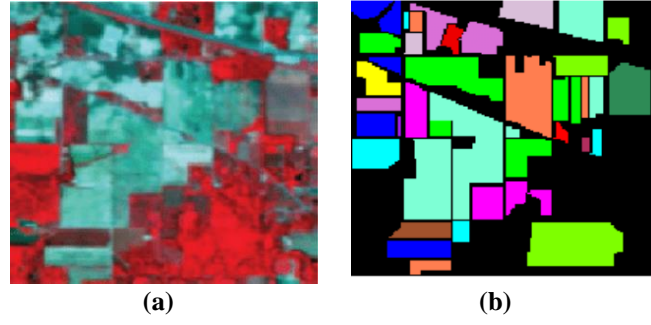
Rapor kapsamındaki yöntemlerin geliştirilmesi aşamasında Indian Pines ve Pavia Centre University hiperspektral veri kümeleri kullanılmıştır

4.1 Indian Pines

Indian Pines veri kümesi [8] AVIRIS sensörü ile 1992 yılında Amerika'nın Indiana eyaletinin kuzey-batı bölgesinden elde edilmiştir. Indian Pines HSG'sü 145x145 piksel ve 0.4 – 2.5 μ dalga boyu aralığına sahip 224 banttan oluşmaktadır ve 20m uzamsal çözünürlüğe sahiptir. Gürültü içeren bantların kaldırılması sonucu 200 bant sınıflandırma amaçlı kullanılmıştır [16]. Bu veri kümesi 16 sınıftan ve toplam 10249 örnekten oluşmaktadır. Tablo 1'de sınıf ve örnek sayıları gösterilmiştir. Indian Pines veri kümesi tarım alanları, orman ve bitki örtüsü gibi birbirine benzer spektral imzalara sahip sınıflardan oluşmaktadır. Spektral imzaların bu yüksek benzerliğinden dolayı sınıflandırılması göreceli olarak daha zor olan bir veri kümesidir [29]. Tabloda görüldüğü gibi, bir kategoriden diğerine örnek sayısı dengeli değil. Bazı kategorilerde çok sayıda örnek varken bazılarında çok az örnek vardır. Bu, veri kümelerinde sık görülen bir sorundur. Herhangi bir kategoriye ait olmayan (Şekildeki siyah alan) açıklamasız pikseller için sınıflandırmaya bir etiket daha ekliyoruz. Şekil 3.1'deki siyah renkli pikseller yer doğrusunda sınıfı olmayan piksellerdir.

Sınıf	Adı	Örnek Sayısı
1	Alfalfa (Yonca)	46
2	Corn-notill (Mısır-işlenmemiş)	1428
3	Corn-mintill (Mısır-az işlenmiş)	830
4	Corn (Mısır)	237
5	Grass-pasture (Çimen-otlak)	483
6	Grass-trees (Çimen-ağaç)	730
7	Grass-pasture-mowed (Çimen-biçilmiş otlak)	28
8	Hay-windrowed (Saman)	478
9	Oats (Yulaf)	20
10	Soybean-notill (Soya fasülyesi- işlenmemiş)	972
11	Soybean-mintill (Soya fasülyesi-az işlenmiş)	2455
12	Soybean-clean (Soya fasülyesi)	593
13	Wheat (Buğday)	205
14	Woods (Orman)	1265
15	Buildings-Grass-Trees-Drives(Bina-Çimen-Ağaç-Araç)	386
16	Stone-Steel-Towers (Kaya-Çelik-Bina)	93
Toplam		10249

Tablo 4.1: Indian Pines veri kümesine ait sınıflar ve örnekler sayıları



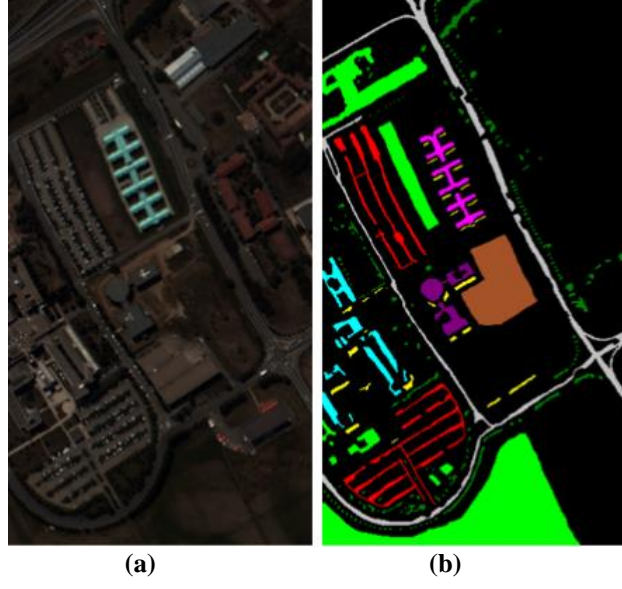
Şekil 4.1: Indian Pines a) 3 bant (43, 21, 11) renkli görüntü b) yer doğrusu

4.2 Pavia University

Pavia University veri kümesi [8] ROSIS sensörü ile 2001 yılında İtalya'nın Pavia şehrinde yer alan Pavia Üniversitesine bir bölgesinden elde edilmiştir. Pavia University HSG'sü 610 x 340 piksel ve 0,43 – 0,86µ dalga boyu aralığına sahip 112 banttan oluşmaktadır ve 1,3m uzamsal çözünürlüğe sahiptir. Gürültü içeren bantların kaldırılması sonucu 103 bant sınıflandırma amaçlı kullanılmıştır [32]. Bu veri kümesi 9 sınıftan ve toplam 42776 örnekten oluşmaktadır. Pavia University 3 bandı ile elde edilmiş renkli görüntüsü ve yer doğrusu Şekil 3.1'de gösterilmiştir. Siyah renkli pikseller yer doğrusunda sınıfı olmayan ve sınıflandırmaya dahil edilmeyen piksellerdir.

Sınıf	Adı	Örnek Sayısı
1	Asphalt	6631
2	Meadows	18649
3	Gravel	2099
4	Trees	3064
5	Painted metal sheets	1345
6	Bare Soil	5029
7	Bitumen	1330
8	Self-Blocking Bricks	3682
9	Shadows	947
Toplam		42776

Tablo 4.2: Pavia University veri kümesine ait sınıflar ve örnekler sayıları

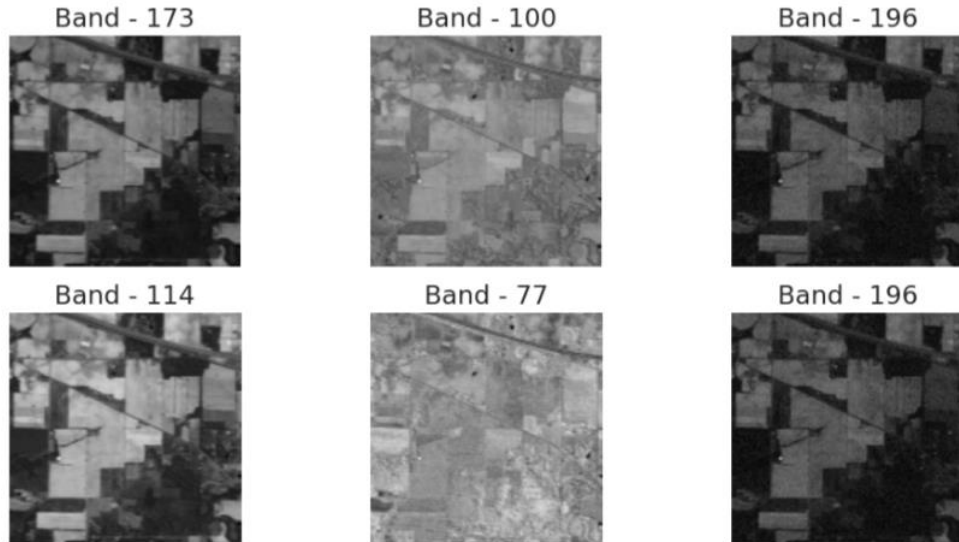


: Şekil 4.2 : Pavia University a) 3 bant (44, 24, 10) renkli görüntü b) yer doğrusu

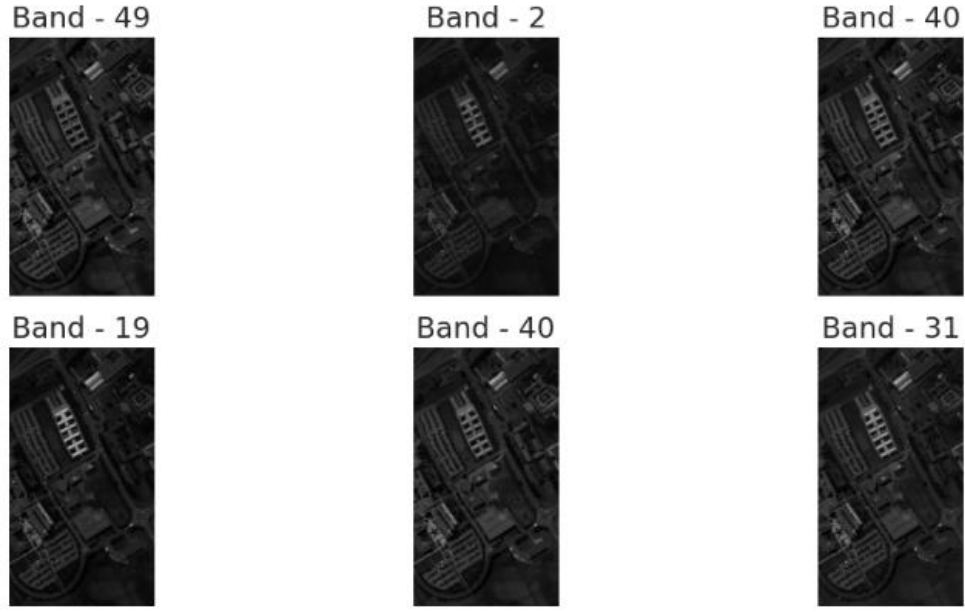
5. DENEYSEL SONUÇLAR

Bu bölümde Indian Pines ve Pavia University verilerinin sonuçları incelenecektir.

Hiperpektral görüntüden rastgele bantlar, bu bantların görüntüleri ve başlıkları şelik 7 ve 8’de gösterilmiştir.

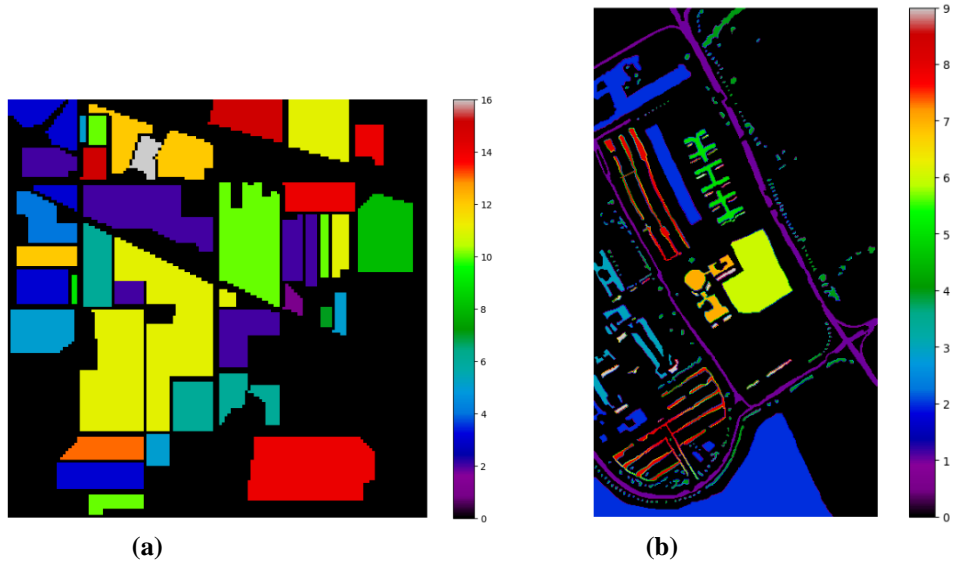


Şekil 5.1 : Kızılderili Çamlarının Bantlarının Görselleştirilmesi (IP) Veri Kümesi



Şekil 5.2 : Pavia University rasgele seçilen 6 bant görüntüsü

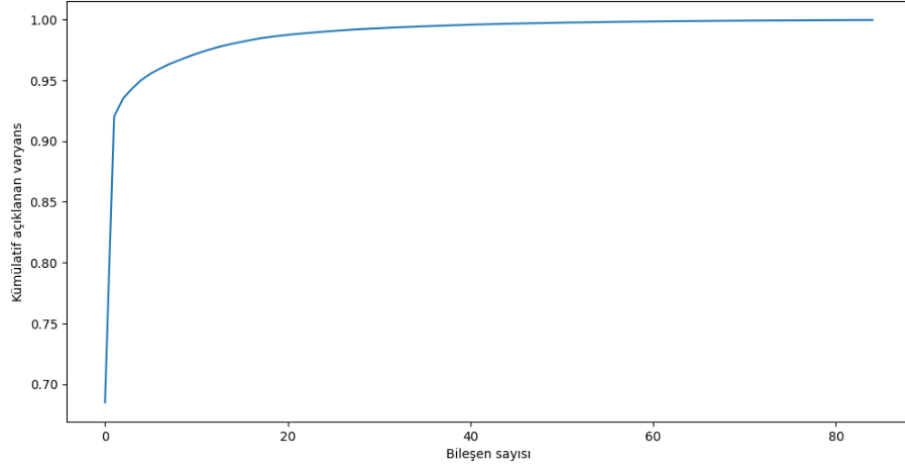
Sınıf etiketlerini bir görüntü olarak görselleştirecek olursak her pikseldeki sınıf etiketi, renk haritasındaki belirli bir renge karşılık gelir. Renk çubuğu, renklerin hangi sınıfları temsil ettiğini gösterir ve dosyaya kaydedilen görüntü, sınıf etiketlerinin görsel bir temsilini sağlar.



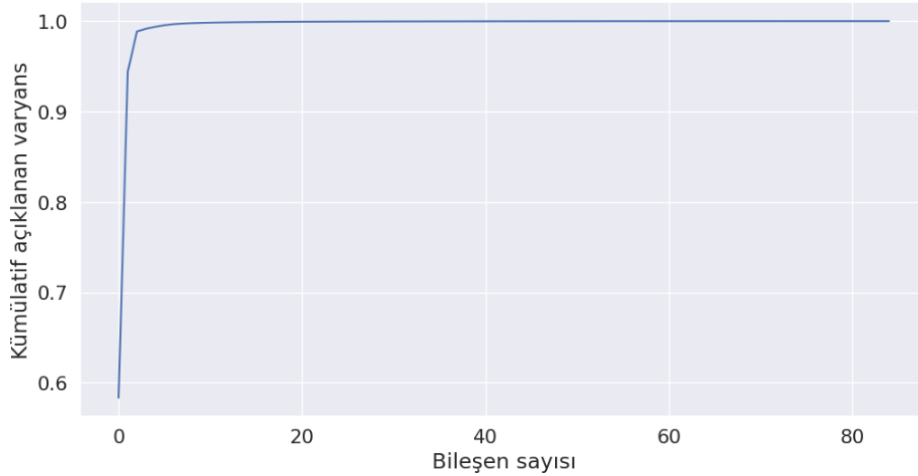
Şekil 5.3 : (a) Indian Pines renk çubuğu (b) Pavia University renk çubuğu

5.1 TBA ile fazlalık analizi

Veri kümesinin boyutunu azaltmak için TBA analizini kullanır ve bileşenlerin varyans açıklama oranlarını görselleştiririz. Böylece, veriyi daha az sayıda bileşenle temsil ederek, gereksiz bilgi kaybı olmadan verinin önemli özelliklerini yakalamayı hedefledik. TBA analizini gerçekleştirmek için 85 bileşen kullanacağımızı belirttik. Bu, veri kümesini 85 boyutlu bir forma indirgeyeceğimiz anlamına gelir.



Şekil 5.4 : IP veri kümesi bileşenlerinin varyans açıklama oranlarını görselleştirilmesi



Şekil 5.5 : Pavia University veri kümesi bileşenlerinin varyans açıklama oranlarını görselleştirilmesi

Eğriler, pratik olarak veri kümelerinin varyansının %98-%99'unu 85 bileşen ile sağlayabildiğini göstermektedir. 85 veri bileşeni, TBA işlemi sırasında verinin en fazla varyansı açıklamasını sağlayan bileşenlerdir. TBA, verinin kovaryans matrisini hesaplayarak, verinin orijinal boyutlarında (örneğin, piksel değerleri) en fazla varyansı içeren bileşenleri belirler. Bu bileşenler, veri kümesindeki en önemli özellikleri temsil eder. TBA işlemi sırasında, verinin boyutu azaltılarak, veri kümesindeki varyansın büyük bir kısmı korunurken, gereksiz veya yetersiz bilgiler ortadan kaldırılır. Bu nedenle, TBA kullanılarak elde edilen bileşenler, verinin

önemli özelliklerini içerirken, gürültü veya gereksiz bilgileri filtreleyerek veri kümesini daha temiz hale getirir. TBA'nın hangi bileşenlerin seçileceği, varyansın açıklama oranına bağlı olarak belirlenir. Örneğin, toplam varyansın %98'i belirlenen bileşen sayısı ile açıklanabiliyorsa, bu bileşenler verinin önemli özelliklerini temsil ederken, geri kalan bileşenlerde genellikle gürültü veya gereksiz bilgiler bulunur.

Indian Pines ve Pavia University veri kümeleri farklı özelliklere ve varyansa sahip oldukları için TBA sonuçları arasında farklılıklar oluşmuştur. TBA'nın sonuçları, veri kümesinin varyansının dağılımına bağlı olarak şekillenir. Bunun sebebi Indian Pines ve Pavia University veri kümeleri arasındaki doğal farklılıklardır. Her bir veri kümesi farklı spektral özelliklere ve dağılımlara sahiptir. Bu farklılıklar, bileşenlerin değişkenliği açıklama yeteneklerini etkilemiştir ve bu da TBA sonuçlarında farklı geçişlerin ortaya çıkmasına neden olmuştur.

Projede TBA işlemini 85 indirgedikten sonra tekrardan 40'a düşürme işlemi uyguluyoruz. İlk olarak, 85 bileşene indirgeme yapıldığı durumda daha fazla bilgi korunur ve verinin orijinal yapısını daha iyi temsil edebiliriz. Ancak, bu durumda boyut azaltma oranı daha düşük olur ve hesaplama ve depolama maliyeti artabilir. Daha sonra, 85 bileşene indirgenmiş veriyi tekrar 40 bileşene indirgeme nedenini açıklayacak olursak :

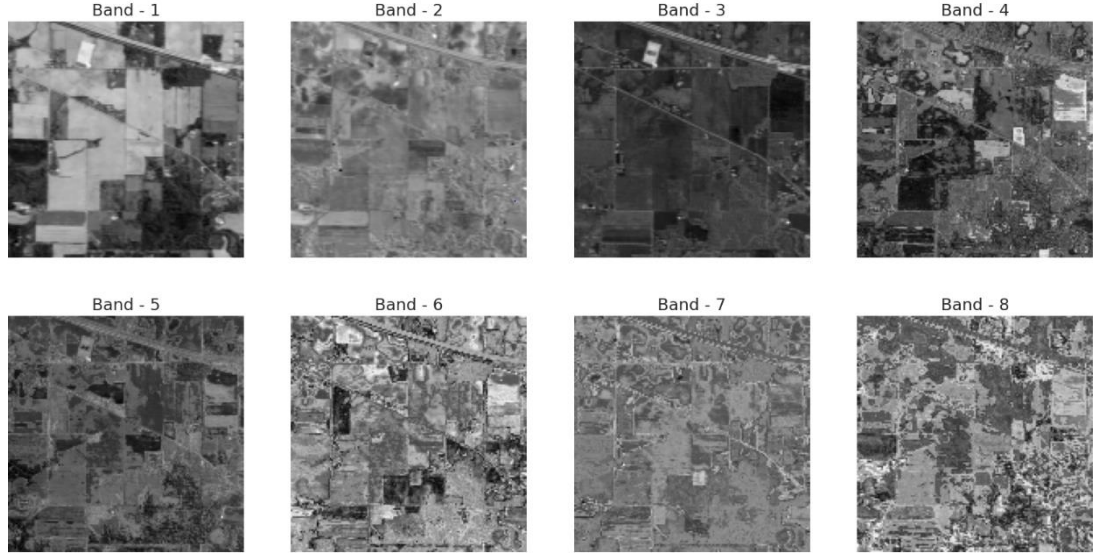
Bilgi koruma: İndirgeme işlemi sırasında, daha fazla bileşenin kullanılması daha fazla varyansın korunmasını sağlar. Ancak, bu fazla varyansın gerçekten gerekli olup olmadığı değerlendirilmelidir. 85 bileşen kullanıldığında önemli bir bilgi kaybı olmadığından, 40 bileşene indirgeme yapılarak daha düşük boyutlu bir temsil elde edilmiştir.

Hesaplama ve depolama maliyeti: TBA analizi, bileşenlerin hesaplanması ve saklanması için kaynak gerektirebilir. 85 bileşen kullanıldığında bu maliyet daha yüksek olabilir. Ancak, 40 bileşene indirgeme yapıldığında daha düşük hesaplama ve depolama maliyeti elde edilmiştir.

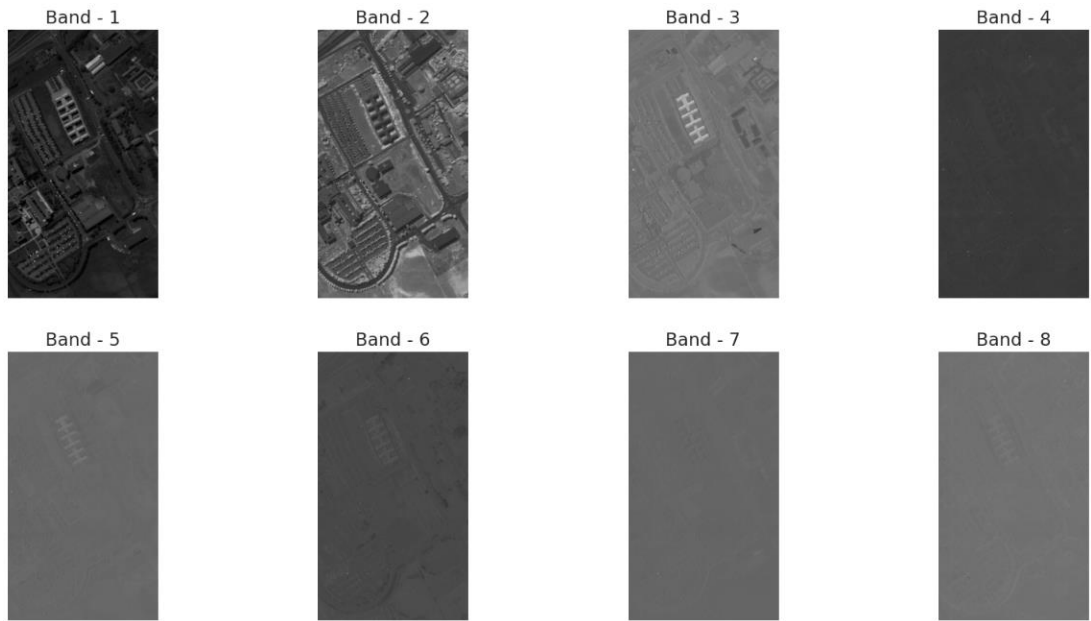
Model performansı: Veri setinin boyutunu azaltırken, model performansının korunması önemlidir. 40 bileşene indirgeme yapıldığında, modelin hala yeterli performansı sağlayıp sağlamadığı değerlendirilmiştir.

Bu nedenlerden dolayı, önce 85 bileşene indirgeme yapılarak daha fazla bilgi korunur ve daha sonra 40 bileşene indirgeme yapılarak daha düşük boyutlu bir temsil elde edilebilir. Bu şekilde, hem bilgi kaybı en aza indirgenir hem de hesaplama ve depolama maliyeti düşürülür. Ancak, bileşen sayısı seçimi veri kümesine ve kullanım senaryosuna bağlı olarak değişebilir ve dikkatlice değerlendirilmelidir

TBA sonucunda elde edilen ilk 8 temel bileşeni (PC-1'den PC-8'e kadar) görselleştiririz. Bu bileşenler, veri kümesindeki değişkenliği en çok açıklayan bileşenlerdir.



Şekil 5.6 : IP veri kümesinin TBA işlemi sonrası 8 temel bileşeni



Şekil 5.7 : Pavia University veri kümesinin TBA işlemi sonrası 8 temel bileşeni

Bu bileşenler TBA sonuçlarını inceleme imkanı sağlar ve bu bileşenlerin veri kümesindeki bilgi içeriğini anlamamıza yardımcı olur.

Beklediğimiz gibi, daha düşük ve daha yüksek varyans değerlerini açıklayan bileşenler giderek daha gürültülü görünmektedir.

TBA sonucunda elde edilen bileşenlerin sınıflandırma işleminde kullanılması amacıyla veri kümesinin uygun bir şekilde hazırlanması ve bir destek vektör makinesi modelinin eğitilmesi işlemleri gerçekleşir.

5.2 Model Eğitimi ve Testi

5.2.1 Veri Hazırlama ve Ön İşleme

İlk olarak veri kümesinde sınıfı 0 olmayan örnekler seçilir. Bu işlem, sınıfı 0 (etiketsiz) olan verileri filtreleyerek sadece etiketli örnekleri içeren bir alt veri kümesi oluşturmayı amaçlar. Sınıf etiketi 0 olan örneklerin çoğunlukla analiz dışında bırakılması tercih edilir. Bu şekilde, modele yalnızca belirli sınıflara ait örnekler sunulur ve sınıflandırma performansını daha iyi elde etmemizi sağlar.

Daha sonra, veri kümesinden bağımsız değişkenler ayrıştırılır. Ayrıştırılan örnekler sınıfı 0 olmayan örneklerin bağımsız değişkenlerini içeren bir dizi olarak tanımlanırken, bu örneklerin sınıf etiketlerini içeren ek bir dizi de tanımlanır. Daha sonra, veri kümesi eğitim ve test alt kümelerine ayrılır. Veri kümesi rastgele şekilde bölünür ve %5'lik bir oranda test alt küme oluşturulur. Sınıf dengesini korunarak bölme işlemi gerçekleştirilir. Burada eğitim kümesinin düşük tutulmasındaki ki amaç modelin ezber yapma (overfitting) eğilimini önlemek yani genelleştirmeyi arttırmaktır. Aynı zamanda veri kümesi çok büyükse, tüm veriyi kullanarak modelin eğitilmesi zaman alabilir bunu engellemektir. Bu durumda, daha küçük bir bölümünü test için ayırmak, hızlı bir şekilde modelin başarısını değerlendirmek için yeterli olabilir.

5.3 DVM Modelimizin Eğitimi

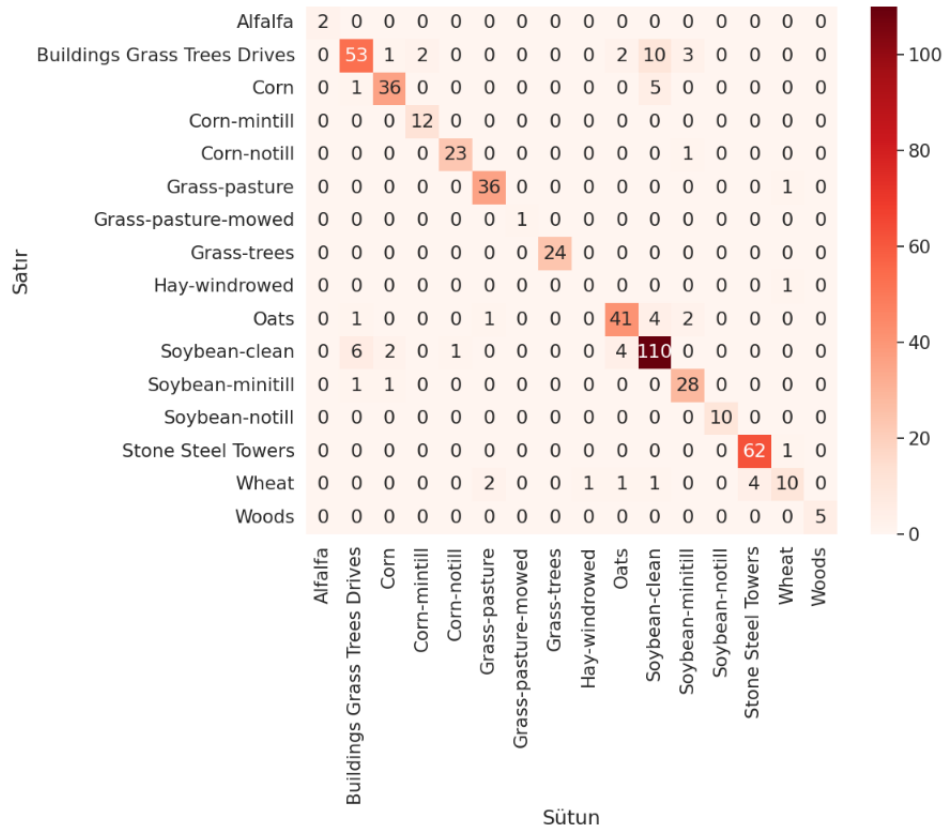
Sonraki adımda, bir destek vektör makinesi modeli oluşturuldu. SVC sınıfı kullanılarak bir DVM modeli tanımlanır ve C değeri 100, çekirdek fonksiyonu RBF (Radyal Taban Fonksiyonu) ve önbellek boyutu 10*1024 olarak ayarlanır. DVM modeli, eğitim veri kümesi üzerinde eğitilir.

Son olarak, eğitilen model kullanılarak test veri kümesi üzerinde tahminler yapılır.

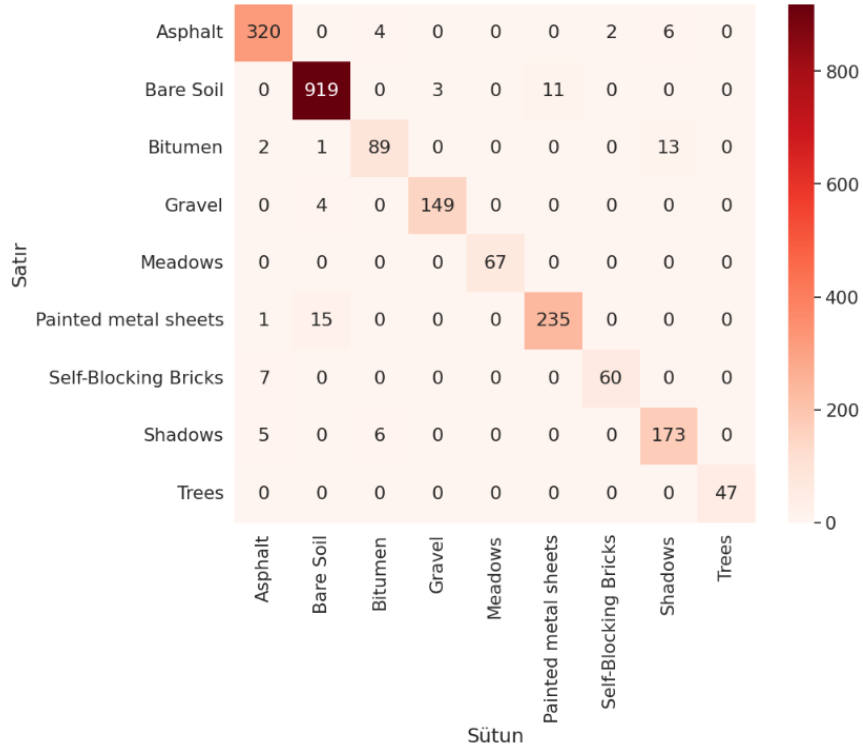
5.4 Sınıflandırıcının Değerlendirilmesi

Sınıflandırma modeli DVM üzerinde gerçekleştirilen test verileri için tahminleri gerçek sınıf etiketleri ile karşılaştırılarak bir karışıklık matrisi oluşturulur. Bu matris, sınıf etiketlerinin doğru ve yanlış tahminlerini gösterir. Sonrasında bir DataFrame oluşturulur. Bu DataFrame, karışıklık matrisini sınıf etiketleriyle ilişkilendirir. Ardından karışıklık matrisi bir ısı haritası olarak görselleştirilir. Bu ısı haritası, sınıf etiketlerinin doğru ve yanlış tahminlerini renklerle gösterir.

Bu işlemin amacı, sınıflandırma modelinin gerçek ve tahmin edilen sınıf etiketlerinin ne kadar doğru eşleştiğini görsel olarak göstermektir. Karışıklık matrisi, sınıflar arasındaki yanlış sınıflandırma oranlarını ve modelin performansını değerlendirmek için kullanılır.



Şekil 5.8 : IP verilerinden elde edilen Karışıklık Matrisi



Şekil 5.9 : Pavia University verilerinden elde edilen Karışıklık Matrisi

Ayrıca, modelin nasıl performans gösterdiğini bize söyleyen sayısal bir ölçütten de memnun oluruz. Bu bağlamda, kesinlik, geri çağırma ve f-skoru gerçekten kullanışlıdır. Kesinlik, bize belirli bir kategori altında doğru şekilde sınıflandırılan numunelerin oranını anlatır. Geri çağırma, modelimizin duyarlılığını, yani belirli bir kategorinin tespit edilen örneklerinin oranını araştırır. Son olarak, f-puanı, hem kesinlik hem de hassasiyet için dengeli bir tahmin sağlar.

Kategorilerin her biri için hesaplanan bu metriklerin yanı sıra tüm kümedeki ortalama değerleri ile birlikte bir sınıflandırma raporu alabiliriz.

Sınıfsal Doğruluk, Doğruluk Kesinliği, Geri Çağırma, F1 Puanı ve Destekten oluşan Sınıflandırma Raporları IP ve Pavia University veri kümeleri için aşağıda gösterilmiştir.

	Kesinlik	Geri çağırma	F1 puanı	Destek
Alfalfa	0.78	0.16	0.26	45
Corn-notill	0.59	0.65	0.62	1352
Corn-mintill	0.56	0.55	0.55	793
Corn	0.47	0.34	0.39	227
Grass-pasture	0.83	0.83	0.83	457
Grass-trees	0.85	0.98	0.91	695
Grass-pasture-mowed	0.00	0.00	0.00	28
Hay-windrowed	0.89	0.99	0.94	450
Oats	0.00	0.00	0.00	20
Soybean-notill	0.62	0.54	0.58	927
Soybean-minitill	0.72	0.76	0.74	2323
Soybean-clean	0.62	0.51	0.56	561
Wheat	0.84	0.95	0.89	191
Woods	0.88	0.91	0.89	1212
Buildings Grass Trees Drives	0.58	0.40	0.47	367
Stone Steel Towers	0.99	0.76	0.86	89
Kesinlik			0.72	9737
Makro Ortalama	0.64	0.58	0.59	9737
Ağırlıklı Ortalama	0.71	0.72	0.71	9737

Tablo 5.1 : IP veri kümesi sınıflandırma raporu

Bu ölçümler bize, modelimizin en düşük desteğe sahip kategoriler için en kötü ve en iyi sınıflandırma puanlarını verdiğini ve doğrulama için neredeyse hiç örnek kalmadığını söylüyor. Bu, bu vakalar için istatistiksel anlamlılığın bulunmadığına işaret etmektedir. 20'nin üzerinde doğrulama destek örneğine sahip kategoriler arasında, binalar-çimen-ağaçlar-sürücüler sınıflandırma performansının en düşük, Wheat (buğday) kategorisinin ise en yüksek olduğunu görüyoruz. Bu, karışıklık matrisinin (confusion_matrix) incelenmesinden bekleyebileceğimiz şeyle uyumludur. Muhtemel bir açıklama, dikkate alınan ölçekte, buğday yüzeylerinin daha düzenli ve spektral olarak farklı olması, dolayısıyla sınıflandırıcı için öğrenmelerinin daha kolay olmasıdır. Elde ettiğimiz doğruluk oranı (OA) %72'dir.

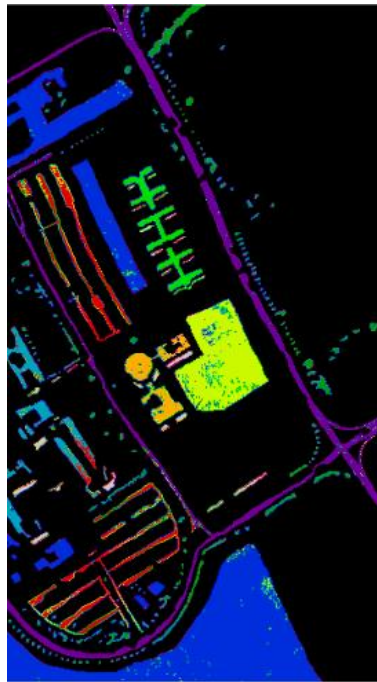
	Kesinlik	Geri çağırma	F1 puanı	Destek
Asphalt	0.95	0.93	0.94	6284
Meadows	0.95	0.98	0.97	17738
Gravel	0.84	0.80	0.82	1991
Trees	0.93	0.94	0.95	2928
Painted metal sheets	1.00	0.99	0.99	1271
Bare Soil	0.93	0.85	0.89	4801
Bitumen	0.90	0.85	0.87	1250
Self-Blocking Bricks	0.83	0.90	0.86	3467
Shadows	1.00	1.00	1.00	988
Kesinlik			0.93	40638
Makro Ortalama	0.93	0.92	0.92	40638
Ağırlıklı Ortalama	0.94	0.93	0.93	40638

Tablo 5.2 : Pavia University veri kümesi sınıflandırma raporu

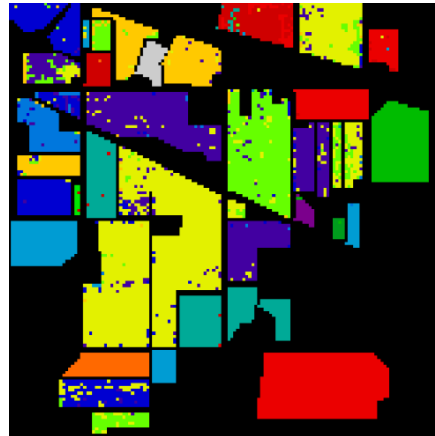
Elde edilen doğruluk oranı %93'dır.

Son olarak, eğitilmiş DVM modeli kullanılarak hiperspektral veri kümesi üzerinde sınıflandırma tahmini yapıldı ve sonuçlar görselleştirildi. Görselleştirme, orijinal etiketlerin (Ground truth) ve DVM tarafından yapılan sınıflandırmanın (Classification SVM) karşılaştırılmasını amacıyla yapılır. Bu karşılaştırma sınıf tahminlerini bir görüntü olarak görselleştirerek, sınıfların dağılımını ve desenlerini anlamaya yardımcı olur.

Bu şekilde, DVM modeli hiperspektral veri kümesi üzerinde eğitildi ve performansı değerlendirildi.



(a)



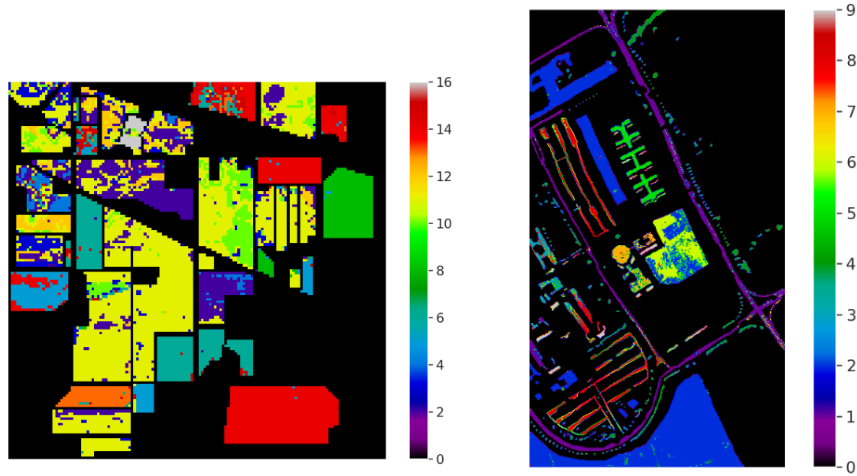
(b)

Şekil 5.10 : (a) Pavia University veri kümesinin (b) IP veri kümesinin DVM ile sınıflandırma sonucu

Gördüğümüz gibi, her bir pikselin spektral bileşimi ile eğitilmiş bir DVM sınıflandırıcı kullanan basit bir makine öğrenme yaklaşımı, önerilen kategoriler için kabul edilebilir bir öngörü yeteneği sağlar. Bu ilk sonuçlara baktıktan sonra, bu hiperspektral görüntüleme kullanım durumunun otomatik sınıflandırma için olgunlaştığını gözlemlemiş olduk.

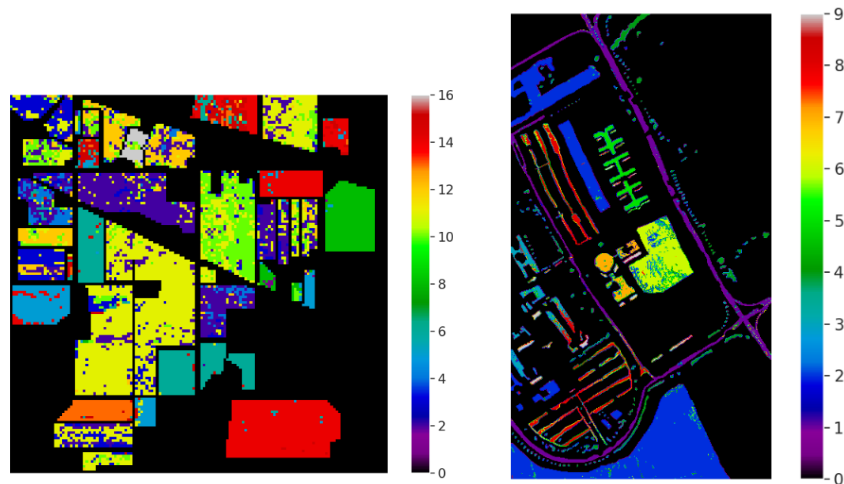
Proje kapsamında veri kümesi rastgele şekilde bölünür ve %5'lik bir oranda test alt küme oluşturulup C parametresi 150 alınmıştır. Şimdi C parametresini değiştirerek bu parametrenin doğruluğa etkisini gözlemleyelim.

C = 10 iken Pavia University doğruluk oranı %65, Indian Pines doğruluk oranı ise %89'dur.



Şekil 5.11 : Indian Pines ve Pavia University C = 10 değerindeki sınıflandırma sonuçları

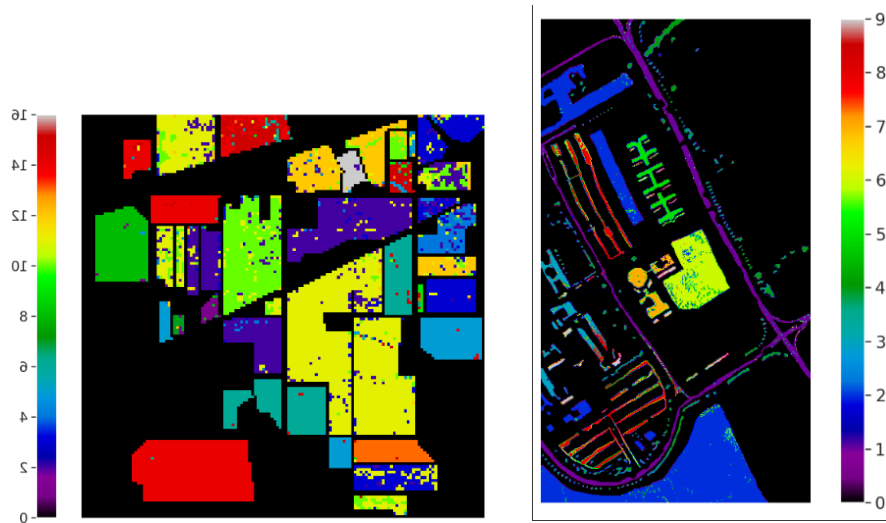
C = 60 iken Pavia University doğruluk oranı %72, Indian Pines doğruluk oranı ise %93'tür.



Şekil 5.12 : Indian Pines ve Pavia University C = 60 değerindeki sınıflandırma sonuçları

Train oranlarını deęiřtirerek başarı üzerine etkisini gözlemleyecek olursak proje kapsamında %5 almıřtık. řimdi bunu %30'a çıkararak başarı üzerindeki etkisini inceleyecek olursak iki veri kümesi için de başarı durumu artacaktır. Bunun sebebi ise train_size parametresini artırdığımızda daha fazla veriyi eğitim için kullanmış oluruz. Bu durumda model daha fazla veriyle eğitildięi için daha iyi öğrenmiş ve daha yüksek bir başarı oranı elde edebilmiştir. Daha fazla veri, modelin genelleme yeteneğini artırır ve yeni, görmedięi veriler üzerinde daha iyi performans göstermesini sağlar.

Indian Pines veri kümesi için başarı oranı %72'den %85'e, Pavia University için ise %93'ten %95'e artmıştır. Ařaęıda görselleřtirilmiştir.



řekil 5.13 : Indian Pines ve Pavia University veri kümelerinin görselleřtirilmiş sonuçları

6. SONUÇLAR

Bu projede, hiperspektral verilerde sıę yöntemler kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. DVM (Destek Vektör Makineleri) ve TBA (Temel Bileřen Analizi) yöntemleri, projenin temel bileřenlerini oluşturmuřtur. Veri kümeleri olarak Indian Pines ve Pavia University hiperspektral veri kümeleri kullanılmış bu veri kümeleri üzerinde DVM ve TBA yöntemleri uygulanmıştır.

Elde edilen sonuçlara göre, Indian Pines veri kümesinde %72, Pavia University veri kümesinde ise %93 doğruluk oranı elde edilmiştir. Bu yüksek doğruluk oranları, hem DVM hem de TBA'nın projeye önemli katkılar sağladığını göstermektedir.

DVM, sınıflandırma işlemi için etkili bir yöntemdir. Veri kümesindeki farklı sınıfları ayırt etme yeteneęi sayesinde, DVM doğruluk oranını artırmış ve hiperspektral verilerin başarılı bir şekilde sınıflandırılmasını sağlamıştır. DVM, veri kümesindeki örnekleri doğru sınıflara atamak için optimize edilmiş bir ayrıştırıcı hiper düzlem bulma algoritması kullanır. Bu sayede, DVM projede sınıflandırma performansını artırmış ve yüksek doğruluk oranlarını elde edilmesini sağlamıştır.

TBA ise boyut indirgeme yöntemi olarak kullanılmıştır. Hiperspektral veriler genellikle çok yüksek boyutlu olabilir ve bu durum işlem süresini ve hesaplama maliyetini artırabilir. TBA, veri kümesinin boyutunu azaltırken, önemli bilgiyi korumayı hedefler. Bu şekilde, daha düşük boyutlu bir temsili elde edilir ve sınıflandırma işlemi daha hızlı ve verimli hale gelir. TBA'nın uygulanmasıyla, projedeki veri kümesinin boyutunu azaltarak hem işlem süresini kısaltmış hem de doğruluk oranını artırmış olduk.

C paramteresini değiştirerek bu parametrenin doğruluğa etkisini de gözlemledik. C ne kadar büyükse Margin o kadar dardır. C' ye çok yüksek değerler verilirse Aşırı öğrenme durumu yaşanabilir. Proje kapsamında verilen değerle başarı oranı istediğimiz yükseklikte yakalamış olduk.

Train_size paramteresini artırdığınızda, daha fazla veriyle modelin eğitilmesini sağladık. Bu durumda, model daha fazla örnek üzerinde eğitildiği için daha iyi öğrenme yapabilir ve daha yüksek bir başarı oranı elde edildi. Ancak, eğer train_size paramteresini aşırı bir şekilde artırırsanız, yani büyük bir oranda eğitim verisini ayırırsanız, bu durum overfitting'e neden olabilir. Overfitting, modelin eğitim verilerine aşırı derecede uyması ve genelleme yapma yeteneğinin düşmesi durumudur. Yani, model eğitim verilerini ezberleyebilir, ancak yeni, görmediği veriler üzerinde düşük bir başarı oranı elde edebilir.

Overfitting genellikle daha karmaşık modellerle ve daha fazla veriyle çalışırken ortaya çıkabilir. Daha fazla veri kullanmanın genelde overfitting'i azaltıcı etkisi vardır. Ancak, eğer train_size paramteresini aşırı büyük seçerseniz, yani eğitim verisinin büyük bir kısmını kullanırsanız, modelin veriyi ezberlemesi ve genelleme yapma yeteneğinin düşmesi olasılığı artar.

Pavia University veri kümesinin daha yüksek doğruluk oranının nedeni, veri kümesinin özelliklerinden kaynaklanır. Pavia University veri kümesi, Indian Pines veri kümesinden farklı karakteristiklere sahiptir. Aşağıda, bu farklılıkların yüksek doğruluk oranına etkisini açıklamak için bazı faktörler bulunmaktadır:

Örnek Sayısı: Indian Pines veri kümesi 145x145 piksel boyutunda bir görüntüden oluşurken, Pavia University veri kümesi 610x340 piksel boyutunda bir görüntüden oluşur. Bu nedenle, Pavia University veri kümesi daha yüksek çözünürlüğe sahiptir ve veri kümesinde daha fazla örnek bulunur. Daha geniş bir veri kümesi, DVM ve TBA gibi algoritmaların daha iyi bir şekilde eğitilmesini sağlar ve daha yüksek doğruluk oranları elde etmelerine olanak tanımıştır.

Sınıf Sayısı ve Dağılımı : Indian Pines veri kümesinde 16 farklı sınıf bulunurken, Pavia University veri kümesinde 9 farklı sınıf bulunur. Sınıf sayısının az olması, sınıflandırma doğruluğunu etkileyebilir. Daha az sınıf olduğunda, sınıflar arasındaki ayırım daha belirgin olabilir ve sınıflandırma algoritması için karar verme süreci daha kolay hale gelebilir. Bu durumda, sınıfların ayırt edilmesi ve doğru sınıflara atama yapılması daha yüksek bir doğruluk oranı elde etmeyi desteklemiş olabilir.

Spektral Özellikler: Indian Pines veri kümesi 200 spektral bant içerirken, Pavia

University veri kümesi 103 spektral bant içerir. Yani, Pavia University veri kümesi daha az spektral bant içerir. Pavia University veri kümesi, Indian Pines veri kümesinden farklı spektral özelliklere sahip olabilir. Bu farklılık, sınıflar arasındaki spektral farklılıkların daha belirgin olmasını sağlayabilir ve sınıflandırma performansını artırabilir.

Veri Kümesinin Karmaşıklığı: Pavia University veri kümesi, Indian Pines veri kümesinden daha az karmaşıktır. Daha az karmaşık bir veri kümesi, DVM ve TBA gibi sınıflandırma ve boyut indirgeme yöntemlerinin daha etkili çalışmasını sağlamıştır.

Sonuç olarak bakıldığında elde edilen doğruluk oranları sıg yöntemler açısından yüksek verilerdir.

7. PROJE KODU

```
import plotly.express as px
import matplotlib.pyplot as plt
import numpy as np
import seaborn as sns
sns.axes_style('whitegrid')
from sklearn.preprocessing import StandardScaler
from sklearn.model_selection import train_test_split
from sklearn.svm import SVC
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix
from scipy.io import loadmat
def read_HSI():
    data = loadmat('PaviaU.mat')
    X = data['paviaU']
    data_gt = loadmat('PaviaU_gt.mat')
    y = data_gt['paviaU_gt']
    print(f'X shape: {X.shape}\ny shape: {y.shape}')
    return X, y
X, y = read_HSI()
```

#Indian Pines için kod bloğu

```
# from scipy.io import loadmat
#def read_HSI():
# X = loadmat('Indian_pines_corrected.mat')['indian_pines_corrected']
# y = loadmat('Indian_pines_gt.mat')['indian_pines_gt']
# print(f'X shape: {X.shape}\ny shape: {y.shape}')
# return X, y
# X, y = read_HSI()
```

```
fig = plt.figure(figsize=(12, 6))
for i in range(1, 1 + 6):
    fig.add_subplot(2, 3, i)
    q = np.random.randint(X.shape[2])
```

```

plt.imshow(X[:, :, q], cmap='gray')
plt.axis('off')
plt.title(f'Band - {q}')
plt.savefig('IP_Bands.png')
plt.figure(figsize = (10, 8))
plt.imshow(y, cmap = 'nipy_spectral')
plt.colorbar()
plt.axis('off')
plt.savefig('IP_GT.png')
plt.show()
import pandas as pd
import numpy as np
def extract_pixels(X,y):
    q = X.reshape(-1, X.shape[2])
    df = pd.DataFrame(data = q)
    df = pd.concat([df, pd.DataFrame(data = y.ravel())], axis = 1)
    df.columns = [f'bans{i}' for i in range(1, 1+X.shape[2])] + ['class']
    df.to_csv('Dataset.csv')
    return df
df = extract_pixels(X,y)
df.head()
df.info()
df.iloc[:, :-1].describe()
from sklearn.decomposition import TBA
TBA = TBA(n_components = 85)
principalComponents = TBA.fit_transform(df.iloc[:, :-1].values)
ev = TBA.explained_variance_ratio_
plt.figure(figsize = (12, 6))
plt.plot(np.cumsum(ev))
plt.xlabel('Bileşen sayısı')
plt.ylabel('Kümülatif açıklanan varyans')
plt.show()
TBA = TBA(n_components = 40)
dt = TBA.fit_transform(df.iloc[:, :-1].values)
q = pd.concat([pd.DataFrame(data = dt), pd.DataFrame(data = y.ravel())], axis = 1)
q.columns = [f'PC-{i}' for i in range(1,41)] + ['class']
q.head()
fig = plt.figure(figsize = (20,10))
for i in range(1, 1+8):
    fig.add_subplot(2,4, i)
    plt.imshow(q.loc[:, f'PC-{i}'].values.reshape(610,340), cmap = 'gray')
    plt.axis('off')
    plt.title(f'Band - {i}')

plt.savefig('IP_TBA_Bands.png')
q.to_csv('IP_40_TBA.csv', index = False)
x = q[q['class'] != 0]
X = x.iloc[:, :-1].values
y = x.loc[:, 'class'].values

```

```
names = ['Asphalt', 'Meadows', 'Gravel', 'Trees', 'Painted metal sheets', 'Bare Soil',  
'Bitumen',
```

```
'Self-Blocking Bricks', 'Shadows']
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y, train_size= 0.05,  
random_state = 11, stratify=y)
```

```
# X ve y veri kümesi train_test_split() fonksiyonuna verilirken, stratify=y  
#parametresi kullanılarak sınıf dengesi korunur. Böylece, X_train ve X_test alt  
#kümeleri, orijinal veri kümesindeki sınıf dağılımının benzer bir dağılımını korur.  
# Bu, modelinizin sınıf dengesine bağlı olarak daha iyi performans göstermesine  
#yardımcı olmuştur.
```

```
svm = SVC(C = 100, kernel = 'rbf', cache_size = 10*1024)  
svm.fit(X_train, y_train)  
ypred = svm.predict(X_test)  
import seaborn as sn  
data = confusion_matrix(y_test, ypred)  
df_cm = pd.DataFrame(data, columns = np.unique(names), index =  
np.unique(names))  
df_cm.index.name = 'Satır'  
df_cm.columns.name = 'Sütun'  
plt.figure(figsize = (10, 8))  
sn.set(font_scale = 1.4) #for label size  
sn.heatmap(df_cm, cmap = "Reds", annot = True, annot_kws = {"size" : 16}, fmt =  
'd')  
plt.savefig('cmap.png', dpi = 300)  
print(classification_report(y_test, ypred, target_names = names))  
l = []  
for i in range(q.shape[0]):  
    if q.iloc[i, -1] == 0:  
        l.append(0)  
    else:  
        l.append(svm.predict(q.iloc[i, :-1].values.reshape(1, -1)))  
clmap = np.array(l).reshape(610,340).astype('float')  
plt.figure(figsize = (10, 8))  
plt.imshow(clmap, cmap = 'nipy_spectral')  
plt.colorbar()  
plt.axis('off')  
plt.savefig('IP_cmap.png')  
plt.show()
```

8. KAYNAKÇA

- [1] DEMİR, B. (2007, Temmuz 13). Hiperspektral görüntülerde akıllı sınıflandırma yöntemleri. *Yüksek lisans tezi*. Türkiye.
- [2] *Uzaktan Algılama Nedir ve Nereelerde Kullanılır?* (2022, Mayıs 27). MAPTRIKS: <https://maptriks.com/uzaktan-algilama-nedir-ve-nerelerde-kullanilir/>
- [3] YILMAZ, M. (2017, Ocak). UZAKTAN ALGILAMA SİSTEMLERİNİN.
- [4] Görmüş, P. D. (tarih yok). *Hiperspektral Görüntüler ve Analizi*. *Hyperspectral Applications*. (2023). Resonon: <https://resonon.com/applications>
- [5]. Wikipedia, 2018, Elektromanyetik tayf, http://tr.wikipedia.org/wiki/Elektromanyetik_tayf, [Ziyaret Tarihi: 12 Nisan 2018].
- [6]. Gerçek, D., Çeşmeci, D., Güllü, M. K., Ertürk, A., Ertürk, S., 2013, Hiperspektral Görüntüleme ve Görüntü İşlemenin Bugünü, Yakın Geleceği ve Mevcut Araştırma Eğilimleri, XYZ.
- [7] Aydın, G. (2021, Ocak 5). *Makine Öğreniminde Overfitting ve Underfitting*. LinkedIn: <https://www.linkedin.com/pulse/makine-%C3%B6%C4%9Freniminde-overfitting-ve-underfitting-gizem-ayd%C4%B1n/?originalSubdomain=tr>
- [8] *Hyperspectral Remote Sensing Scenes*. (2021, July 12). ehu.eus: https://www.ehu.eus/ccwintco/index.php/Hyperspectral_Remote_Sensing_Scenes
- [9]. (Hyperspectral Applications, 2023).
- [10]. (Hyperspectral Applications, 2023). Z. Yang et al., Plant Phys. and Biochem. 166, 723 (2021).
- [11]. S.C. Jay et al., Geoscience and Remote Sensing Symposium (IGARSS) (2010)
- [12] Demir, B., 2010, Hiperspektral görüntülerin yüksek doğruluklu sınıflandırılması, Doktora, Kocaeli Üniversitesi.
- [13]. Boser, B. E., Guyon, I.M. and Vapnik, V., 1992, A training algorithm for optimal margin classifiers, Proc. 5th Annu. ACM Workshop Comput. Learn. Theory, pp. 144–152.
- [14]. Melgani, F. and Lorenzo, B., 2004, Classification of hyperspectral remote sensing images with support vector machines, IEEE Trans. Geosci. Remote Sens., vol. 42, no. 8, pp. 1778–1790

- [15]. Gualtieri, J. A. and Chettri, S., 2000, Support vector machines for classification of hyperspectral data, in Proc. IEEE Geosci. Remote Sens. Symp. (IGARSS), Honolulu, HI, USA, pp. 813–815
- [16]. Camps-Valls, G., Gomez-Chova, L., Munoz-Mari, J., Vila-Frances, J. and CalpeMaravilla, J., 2006, Composite kernels for hyperspectral image classification, IEEE Geoscience and Remote Sensing Letters, vol. 3, no. 1, pp. 93-97.
- [17]. Fauvel, M., Benediktsson, J. A., Chanussot, J. and Sveinsson, J. R., 2008, Spectral and spatial classification of hyperspectral data using SVMs and morphological profiles, IEEE Trans. Geosci. Remote Sens., vol. 46, no. 11, pp. 3804–3814.
- [18]. Benediktsson, J.A., Palmason J.A., Sveinsson, J.R., 2005, Classification of hyperspectral data from urban areas based on extended morphological profiles, IEEE Transactions on Geoscience and Remote Sensing, 43, 480-491.
- [19]. Li, S., 2009, Markov Random Field Modeling in Image Analysis. Berlin, Germany: Springer-Verlag.
- [20]. Eches, O., Benediktsson, J. A., Dobigeon, N. and Tournieret, J. Y., 2013, Adaptive Markov Random Fields for Joint Unmixing and Segmentation of Hyperspectral Images, IEEE Transactions on Image Processing, vol. 22, no. 1, pp. 5-16
- [21]. Tarabalka, Y., Fauvel, M., Chanussot, J. and Benediktsson, J. A., 2010, SVM and MRFbased method of accurate classification of hyperspectral images, IEEE Geosci. Remote Sens. Lett., vol. 7, no. 4, pp. 736–740.
- [22]. Li, W., and Du, Q., 2014, Gabor-Filtering Based Nearest Regularized Subspace for Hyperspectral Image Classification, IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing, vol. 7, no. 4, pp. 1012-1022.
- [23]. Zhang, L., Zhang, L., Tao, D., Huang, X., 2012, On combining multiple features for hyperspectral remote sensing image classification. IEEE Trans. Geosci. Remote Sens., 50, 879–893.
- [24]. Li, W., Chen, C., Su, H.J. and Du, Q., 2015, Local binary patterns and extreme learning machine for hyperspectral image classification, IEEE Trans. Geosci. Remote Sens., vol. 30, pp. 3681–3693.
- [25]. Rodarmel, C. and Shan, J., 2002, Principal Component Analysis for Hyperspectral image classification, Surveying and Land Information Systems, vol. 62, No. 2, pp.115-122
- [26]. Ögündür, G. (2020, Ocak 14). *TBA (Principal Component Analysis) Temel Bileşenler Analizi*. medium: <https://medium.com/@gulcanogundur/TBA-principal-component-analysis-temel-bile%C5%9Fenler-analizi->

bf9098751c62

[27] Ögündür, G. (2020, Ocak 9). *Overfitting (Aşırı Öğrenme), Underfitting (Eksik Öğrenme) ve Bias-Variance Çelişkisi*. medium:

<https://medium.com/@gulcanogundur/overfitting-a%C5%9F%C4%B1r%C4%B1-%C3%B6%C4%9Frenme-underfitting-eksik-%C3%B6%C4%9Frenme-ve-bias-variance-%C3%A7eli%C5%9Fkisi-b92bef2f770d>

[28] Uğur, B. (2020, Eylül 3). *Medium*. Destek Vektör Makinesi (SVM):

<https://burakugur.medium.com/destek-vekt%C3%B6r-makinesi-svm-f07f57f0a641> adresinden alındı

[29] DINH, A.-T. (2019). *dinhanhthi*. Support Vector Machine (SVM):

<https://dinhanhthi.com/support-vector-machine/> adresinden alındı

[30]. Sokolova, M., Lapalme, G., 2009, A systematic analysis of performance measures for classification tasks, *Information Processing & Management*, Volume 45, Issue 4, Pages 427-437, ISSN 0306-4573.

[31] Torun, H. (2022, Ekim 01). *Karışıklık Matrisi (Confusion Matrix)*. hakan.io:

[https://hakan.io/karisiklik-matrиси-confusion-matrix/#:~:text=Kar%C4%B1%C5%9F%C4%B1kl%C4%B1k%20matrisi%20\(Confusion%20Matrix\)%20bir,t%C3%BCrleri%20hakk%C4%B1nda%20da%20fikir%20verir.](https://hakan.io/karisiklik-matrиси-confusion-matrix/#:~:text=Kar%C4%B1%C5%9F%C4%B1kl%C4%B1k%20matrisi%20(Confusion%20Matrix)%20bir,t%C3%BCrleri%20hakk%C4%B1nda%20da%20fikir%20verir.) adresinden alındı

[32]. Zhou, Y., Peng, J. and Chen, C. L. P., 2015, Extreme Learning Machine With Composite Kernels for Hyperspectral Image Classification, *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, vol. 8, no. 6, pp. 2351-2360.