



AFLATOKSİNLİ BİBERLERİN HİPERSPEKTRAL GÖRÜNTÜLERİNİN SINIFLANDIRILMASI İÇİN YENİ YAKLAŞIMLAR

*(NEW APPROACHES FOR CLASSIFICATION OF HYPERSPECTRAL
IMAGES OF CHILI PEPPERS WITH AFLATOXINS)*

Musa ATAŞ*, Yasemin YARDIMCI*, Alptekin TEMİZEL*

ÖZET/ABSTRACT

Birçok gıda (fındık, fıstık, badem, mısır, buğday, kuru incir ve pul biber gibi), hasat, üretim ve depolama dönemlerinde kanserojen etkisi bilinen aflatoksin üreten küfler ile karşılaşmaktadır. Aflatoksin tespiti için kullanılan kimyasal yöntemler, daha doğru sonuçlar vermesine karşın uzun zaman alan, pahalı ve tahribatlı süreçlerdir. Bu çalışmada aflatoksinli pul biberlerin sınıflandırması için halojen ve morötesi (UV) ışık altında edinilen hiperspektral görüntüler kullanılmış ve bu görüntülerden parlaklık histogramı öznitelik vektörleri çıkarılmıştır. En önemli öznitelikler, yapay sinir ağları (YSA) bağlantı ağırlıkları ve asgari artıklık/azami ilişki yöntemleri kullanılarak elde edilmiş, değişik topolojilerdeki YSA ile sınıflandırma yapılarak veri tümleştirmenin sınıflandırma başarısına etkisi araştırılmıştır.

Many foods (such as hazelnut, pistachio nut, almond, corn, wheat, dried fig, and chili pepper) are prone to carcinogenic aflatoxin formation during harvesting, production and storage periods. Chemical methods are used for detection of aflatoxins give accurate results, but they are slow, expensive and destructive. In this study, intensity histogram features of hyperspectral images of chili peppers are extracted under halogen and ultraviolet (UV) illumination source. Salient features are selected by using connection weights of artificial neural networks and minimum redundancy maximum relevance techniques. With various topologies of artificial neural networks, effect of data fusion on classification performance is investigated.

ANAHTAR KELİMELELER/KEYWORDS

Veri tümleştirme, Hiperspektral görüntüleme, Gıda güvenliği, Makine öğrenmesi, Aflatoksin

Data fusion, Hyperspectral imaging, Food safety, Machine learning, Aflatoxins

* Orta Doğu Teknik Ün., Enformatik Enstitüsü, ANKARA

1. GİRİŞ

Birçok gıda (findık, fıstık badem, mısır, buğday, kuru incir ve pul biber gibi), kanserojen etkisi bilinen aflatoksin üreten küflere maruz kalmaktadır. Bu küfler toprakta yaşayan mikroorganizmalar olup gıda ürünlerinin toprakla temas etmeleri sonucu gıdalara geçmektedir. Aflatoksin ile bulaşma genelde hasat, üretim ve depolama dönemlerinde olmaktadır. Mevsimsel etkiler, ortamın nem ve sıcaklık koşulları, küf oluşumu ve böcek tahribatı gıdada aflatoksinin gelişmesine neden olurlar. Aflatoksin, çeşitli *Aspergillus* küf türlerinin, özellikle de *Aspergillus Flavus* ve *Aspergillus Parasiticus* küflerinin oluşturduğu toksik bileşiklerdir (Zeringue ve Shih, 1998). Uluslararası kanser araştırma ajansının verilerine göre aflatoksin, insanlarda karaciğer kanserine sebep olmaktadır. Bu yüzden birçok ülkede, tüketilen gıdalardaki aflatoksin miktarını kontrol etmek için sert önlemler alınmıştır. Genel olarak kabul edilen aflatoksin seviyesi ABD ve Türkiye’de 20 ppb (milyarda 20) ve Avrupa ülkelerinde 5ppb (milyarda 5)’dir (Çizelge 1). UV ışığına verdikleri tepkiye göre dört aflatoksin çeşidi vardır. Mavi ışıklı olanlar B1 ve B2, yeşil ışık verenler ise G1 ve G2 olarak adlandırılırlar. Bu türlerin arasında sağlık açısından en tehlikeli olan B1 türüdür (Şekil 1).

Çizelge 1. Avrupa birliği ülkeleri ve ABD’de çeşitli gıdalar için kabul edilen aflatoksin seviyeleri

Ülke	İzin verilen seviye	Ürünler	Açıklama
Avusturya	B1<1ppb	Mekanik olarak hazırlanan tahıllar hariç bütün gıda ürünleri	
Belçika	<5 ppb	Yer fıstığı	
Almanya	B1+B2+G1+G2<4ppb	Bütün gıda ürünlerinde	
Danimarka	B1<2ppb	Bütün gıda ürünlerinde	
Hollanda, İsveç, Finlandiya	B1<5ppb	Bütün gıda ürünlerinde	
İsviçre	B1<1ppb	Bütün gıda ürünlerinde (mısır hariç)	
	B2+G1+G2<5ppb	Bütün gıda ürünlerinde	
İtalya, Fransa	< 10 ppb for B1		Herhangi bir düzenleme henüz yok
İngiltere	<50ppb	Baharatla ilgili bir düzenleme yok, ancak, findık, kuru incir için <4 ppb olmak zorunda.	
ABD	<20 ppb	Bütün gıda ürünlerinde	

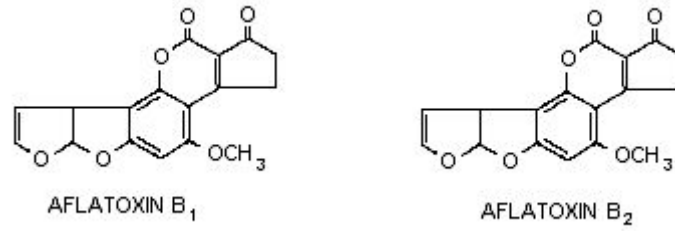
Aflatoksin miktarını belirlemek için çeşitli kimyasal yöntemler vardır. Kitle spektroskopisi (Mass spectroscopy), ince film kromatografi (TLC), yüksek başarımlı sıvı kromatografi (HPLC), ve enzim ilintili immün test (ELISA) en çok bilinen kimyasal yöntemlerdendir. Bu yöntemlerin arasında en hassas ve doğru sonuç üreten yöntem HPLC'dir (Lipps ve Mills, 2010). Kimyasal yöntemlerin daha doğru sonuçlar vermesine rağmen bu yöntemlerin pahalı, yavaş ve tahribatçı oluşları yüzünden, bilgisayarla görü tabanlı sistemler üzerinde çalışmalar yapılmaktadır. Bu sistemler temelde aflatoksin'den meydana gelen bileşiğin 365 nm UV ışık altında, parlak yeşil ve sarımsı floresan (BGYF, Bright Green Yellowish Fluorescent) ışık yayması prensibinden hareketle, belirli bir eşik değerini kullanarak sınıflandırma yapmaya çalışırlar. Diğer taraftan, ışımaya sebep olan bileşik gerçekte kojik asit ve gıdadaki peroksidaz enzimlerinden meydana geldiğinden dolayı BGYF ışması doğrudan aflatoksin ile ilintili olmayabilir (Jardine ve Peter, 2010). Kojik asit ise küflerden olabileceği gibi (Aspergillus Flavus, Aspergillus Parasiticus ve değişik türlerinden), bakterilerden de ortaya çıkabilir. Dolayısıyla, bazen BGYF ışması görüldüğü halde gerçekte gıdada aflatoksin olmayabilir. Bu durum yanlış pozitif olarak nitelendirilebilir. Diğer taraftan, aflatoksin var olmasına rağmen, gıdada peroksidaz enziminin olmaması, BGYF ışmasını engelleyerek sınıflandırmada en tehlikeli olan yanlış negatif tipindeki hatayı üretecektir (Jardine ve Peter, 2010) (Şekil 2) .

Yao vd.'ye göre aflatoksinli mısır örneklerinde, BGYF ışması, 500 nm ile 515 nm arasında en yüksek değerine ulaşır (Yao vd., 2006). Ancak, bu ışmanın gözlemlenmesi için aflatoksin kontaminasyonunun oldukça yüksek seviyelerde olması gerekmektedir (>5000 ppb), 10 ppb'den küçük seviyelerde ışma gözlenmeyecektir. Pearson vd., silikon foto-diyot dizili fiber optik spektrometre ve 100 W halojen lamba kullanarak, yüksek seviyede (>100 ppb) kontamine olmuş mısır tanelerini, temiz olanlarından % 96,6 başarı oranı ile sınıflandırabilmiştir (Pearson vd., 2001). Ayrımsallık analizi tekniğini kullanarak, en yüksek ayrımsallığı [735nm/1005nm] spektral yansıma oranında yakalamıştır. Diğer taraftan yine mısırlarda, % 91,2 başarı oranını, en iyi ayrımsallığı veren spektral geçirgenlik oranında [720nm/780nm] elde etmiştir. Bunlara ek olarak, Hirano vd., yer fıstığı üzerinde çalışmalarında, % 93 sınıflandırma başarısını [700nm/1100nm] spektral geçirgenlik oranında elde etmiştir (Hirano vd., 1998). Diğer gıdalarla yapılan çalışmalarda, hedef gıdaların iri taneli olmaları sebebiyle aflatoksin kontaminasyonunun incelenen örneğin belirli bir bölgesinde oluşunun, hiperspektral analizde kolaylık sağladığı görülmüştür (Pearson vd., 2001; Yao vd., 2006; Hirano vd., 1998). Bu makalede ele alınan problemde, pul biber taneciklerini tek tek inceleyebilmek mümkün olmadığı gibi, hastalığın dağılımını tahmin etmek de oldukça zordur. Dağılımın homojen olduğu kabul edildiğinde, incelenen pul biber başına düşen aflatoksin oranı da çok düşük olacağı gibi gruptan ayrılması pratik olmayacaktır. Dolayısıyla diğer gıdalarda kullanılan yöntemlerin bu problem için kullanılması uygulanabilir değildir. Bu çalışmada, hiperspektral görüntü farklarını kullanarak öznelik vektörü elde edilmeye çalışılmıştır.

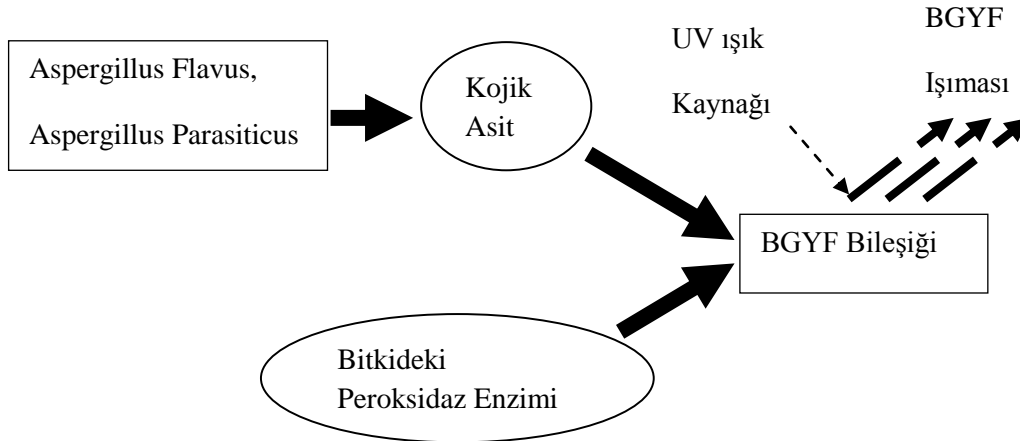
İkinci bölümde, bilgisayarla görü sisteminin yapısı ve önerdiğimiz öznelik vektörü çıkarımı ve seçim yöntemleri ele alınacaktır. Üçüncü bölümde, deney sonuçları verilecektir. Dördüncü bölümde ise kullanılan yöntemlerin sonuçları tartışılacaktır.

2. BİLGİSAYARLA GÖRÜ SİSTEMİ

Bilgisayarla görü sistemi, kimyasal yöntemlere göre hızlı ve tahribatsız oluşundan dolayı tercih edilmiştir. Önceki çalışmalarda, genellikle UV ışık kullanılmıştır (Halojen ışık kaynağının da kullanıldığı çalışmalar vardır) (Pearson vd., 2001; Yao vd., 2006; Hirano vd., 1998). Bu çalışmada UV ve Halojen ışık kaynakları, en iyi ayrımsallık veren bantları bulabilmek için kullanılmıştır.



Şekil 1. Aflatoksin B1 ve B2 nin kimyasal yapısı

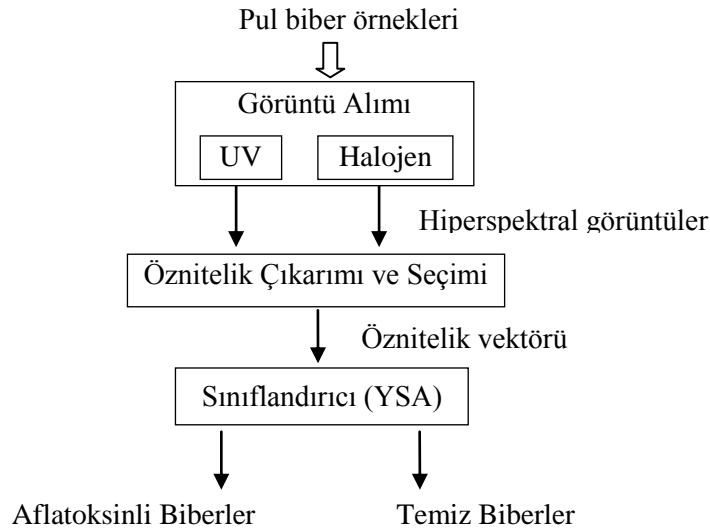


Şekil 2. BGYF ışıması oluşumu

Bilgisayarla görü sistemi, Şekil 3'te görüldüğü gibi, üç temel bileşenden oluşmaktadır. Bunlar; görüntü alımı, öznitelik çıkarımı ve öznitelik seçimi, ve sınıflandırıcı bileşenleridir. Halojen ve UV ışık kaynakları altında, 53 farklı pul biber örneğinin 3 farklı bölgesinden 400 nm den 720 nm'ye kadar hiperspektral görüntüleri alınmıştır. Öznitelik olarak, görüntülerin ortalama parlaklık değerleri ve komşu bantların nicelenmiş fark histogram matrisleri kullanılmıştır. Öznitelik setindeki en önemli öznitelikler, mRmR ve önerilen YSA bağlantı ağırlıklarına göre sıralanmış ve YSA ile danışmanlı öğrenmeye tabi tutulmuştur.

2.1. Görüntü Alımı

Aydınlatma kaynağı olarak 100 W quartz-tungsten halojen lamba ile 365 nm UV ışın tabancası kullanılmıştır. Görüntüler, Sony firewire CCD kamera ve varispec elektronik süzgeç kullanılarak çekilmiştir. Türkiye'nin farklı bölgelerinden toplanan 53 değişik pul biber örneğinin her birinin üç farklı yerinden görüntüler alınarak örnek kümesi zenginleştirilmiştir. Varispec elektronik süzgeç ile 400 nm'den 720 nm'ye kadar 10'ar nm aralıklarla toplam 33 adet hiperspektral görüntü alınmıştır. Her görüntü 1280x960 piksel boyutundadır. Tüm bantlar için tek bir pozlama süresi kullanımı görüntülerde doymayan (unsaturated) ve fazla doymuş (oversaturated) bölgeler oluşturduğu için, kullanılan ışık kaynaklarının karakteristiğine göre pozlama süreleri belirlenmiştir. UV ışık kaynağının kullanılmasıyla floresan olgusu, halojen ışık kaynağıyla da yansıma olgusu incelenmiştir. Çizelge 2, kamera kazanç (gain) parametresi 850 elektron/CCD algılayıcı hücrelerine ayarlandıktan sonra tespit edilen bölgelerin, sözkonusu ışık kaynağı kipi için pozlama sürelerini göstermektedir.



Şekil 3. Önerilen Sistemin Taslağı

Çizelge 2. Kamera pozlama süresinin kiplere göre değişimi

Kip	Pozlama süresi (s)
Halojen (400-490) nm	4.5
Halojen (500-590) nm	2.4
Halojen (600-720) nm	0.5
UV (400-690) nm	9.6
UV (700-720) nm	3.1

Görüntü alınırken, sistem üzerinde herhangi bir titreşim ve sarsılma olmadığından, varispec elektronik süzgeç ile çekilen spektral görüntülerin çakıştırılmasına ihtiyaç duyulmamıştır. Ancak kameranın yüksek kazanç parametresi ve ortamdaki tozlardan

kaynaklanan gürültüleri temizlemek için 3x3 ortanca süzgeci kullanılmıştır. Görüntü alımından sonra örnekler, TÜBİTAK Ankara Test ve Analiz Laboratuvarı (ATAL)'a gönderilerek, danışmanlı öğrenmede kullanılacak referans aflatoksin değerleri elde edilmiştir. Gelen sonuçlarda dağılımın dengeli olması ve sağlıklı sınıflandırma yapılabilmesi için kontaminasyon eşik değeri 10 ppb alınmıştır. Şekil 4'de halojen ve UV ışık kaynağı kullanılarak elde edilmiş örnek hiperspektral görüntüler görülmektedir.

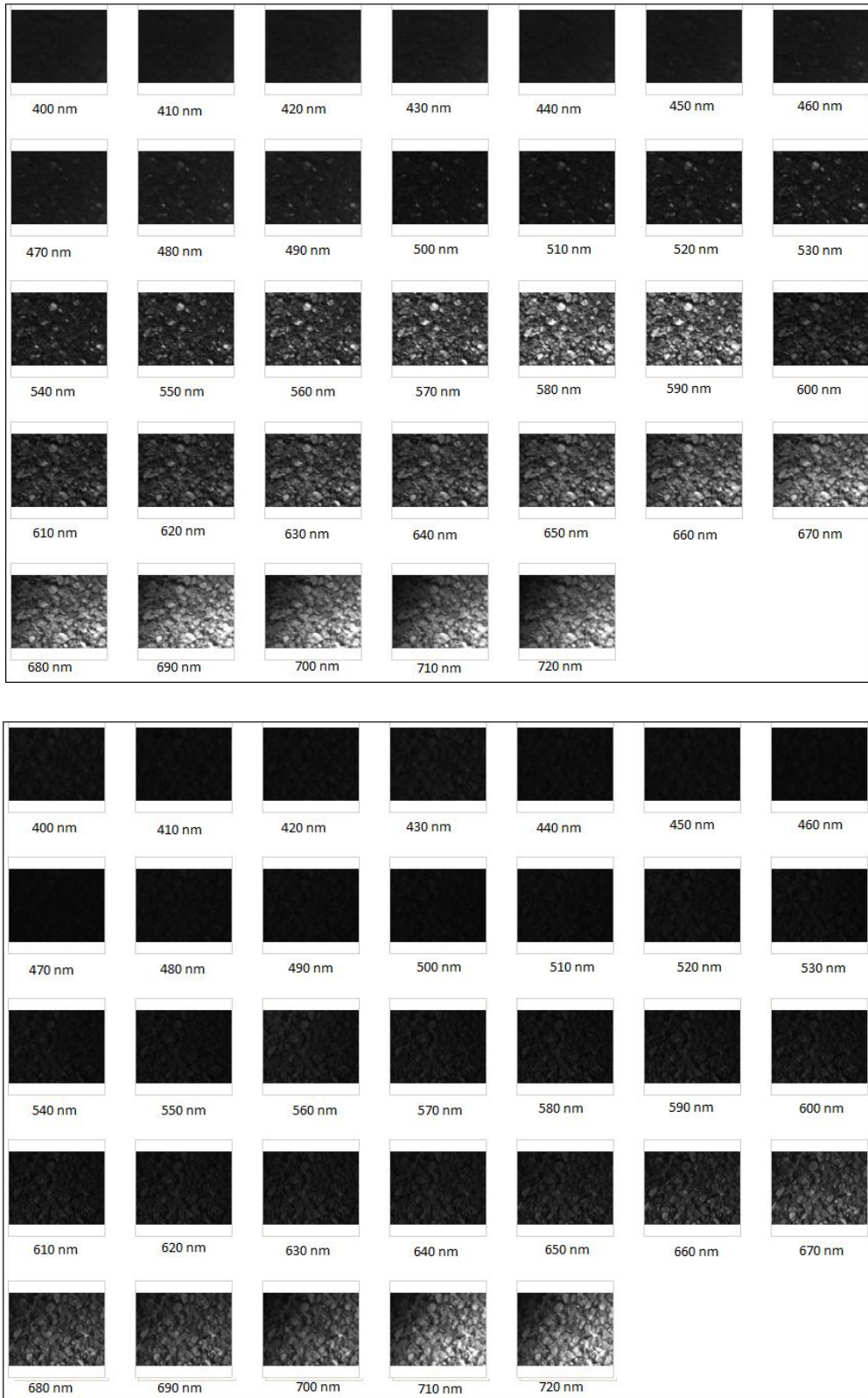
Örnek kümesindeki 53 adet biberin aflatoksin seviyelerinin dağılımı, Şekil 5'te görülmektedir.



Şekil 5. 53 biberin aflatoksin düzeylerinin dağılımı (küçükten büyüğe doğru sıralanmıştır)

2.2. Öznitelik Çıkarımı ve Seçimi

Bir sınıflandırıcının başarımı, öznitelik vektörünün kalitesi ile ilintilidir. Özniteliklerin tanımlayıcı olmayan veriler ve yüksek miktarda gürültü içerdiği durumda, sınıflandırıcı doğru çözüme ulaşamayabilir. Bununla birlikte öznitelik vektörünün boyutunun büyük oluşu veri işlemeyi zorlaştırır. Bu yüzden, öznitelik çıkarımı ve seçim yöntemlerini kullanarak görel katkıları az olan özniteliklerin, öznitelik vektöründen çıkarılması gerekir. Her iki yöntemde de asıl amaç, orijinal öznitelik vektörünün boyutunun küçültülmesidir ki, buna boyutsal indirgeme (dimensionality reduction) denir. Öznitelik seçimi, temel anlamda, öznitelik vektörü elemanlarının önem büyüklüğüne göre sıralanmasıdır. Diğer taraftan, öznitelik çıkarımı, orijinal verilerin başka bir boyuta yansıtılarak taşınması fikrine dayanmaktadır. Öznitelik çıkarımının, seçime göre en büyük avantajı, öznitelik çıkarımı neticesinde elden edilen öznitelik kümesinin daha küçük boyutlu ve daha zengin olmasıdır. Öznitelik seçiminin en büyük faydası ise, gerçekleştirilen sistemin daha az sayıda spektral bantla çalışabilmesi ve basit olmasıdır.



Şekil 4. Halojen (üstte) ve UV (altta) ışık kaynakları kullanılarak elde edilen hiperspektral görüntü serileri (400 nm–720 nm)

Öznitelik çıkarımı için, Temel Bileşenler Analizi (TBA) (Principal Component Analysis, PCA) ve kendi kendine birleşmeli YSA (Auto Associative ANN) iyi bilinen yöntemlerdendir. TBA, boyutsal indirgemede çok sık kullanılan, orijinal veri ve özellikleri iyi tanımlanabilen öznitelikleri çıkarmada kullanılır. Benzer şekilde, kendi kendine birleşmeli YSA'lar da bir gizli katman kullanarak (gizli katmanın nöron sayısının giriş ve çıkışlardan küçük olması gerekir) öznitelik vektörünün boyutunu küçültürler (Bourlard ve Kamp, 1988). TBA ve kendi kendine birleşmeli YSA öznitelik çıkarımı teknikleri, temelde tüm bantları kullanı için bu bantları kullanmak yerine, daha az sayıda hiperspektral bantlarla çalışmaya olanak sağlayan yöntemler kullanılmıştır.

Önceki çalışmaların hemen hemen hepsinde gıda örneklerinin iri taneli yapıya sahip olmaları nedeniyle tek tek incelenebilmeleri mümkündür. Ancak, buradaki problemde pul biberlerin taneli yapısından dolayı bir miktar pul biberin görüntüleri alınıp, aflatoxin dağılımının homojen olduğu varsayılarak araştırma yapılmıştır. Çıkarılacak özniteliklerin, bir şekilde doku ve ortalama piksel yoğunluğundan bağımsız olması gerekmektedir. Bu yüzden spektral bantlardaki görüntüler üzerinde doku tabanlı olan gri seviyesi eş ortaya çıkma matrisi (gray level co-occurrence matrix) özelliklerini kullanmak yerine komşu bantların mutlak değer farkları analiz edilmeye çalışılmıştır. Bununla birlikte, spektral bant farkları kullanılarak sınıflandırıcının doku ve şekle bağımlı olması engellenmiştir. Komşu spektral bantların mutlak değer farkları alındıktan sonra, bu fark görüntülerin histogramları alınmıştır. Histogramlar bir görüntünün 0-255 gri piksel değerlerinden kaç adet olduğunu gösterdiği için, doku ve şekil uzayından uzaklaşmıştır. 0-255 değerleri arasını, daha da küçültmek için 25 değere nicemlenmiştir. Şekil 6, her biber örneği için, doku ve şekilden bağımsız, hiperspektral görüntü farkı histogram matrisini göstermektedir. Şekildeki siyah renkler fark görüntüsünde ilgili renk aralığında piksel olmadığını göstermektedir. Koyu griden, beyaz renge doğru gidildikçe toplam piksel sayısı artmaktadır. Şekil 6'daki matriste görünen en büyük çubuk, 590-600 bantlarının pozlama farkından kaynaklanmaktadır (Çizelge 2).



Şekil 6. Hiperspektral bant çiftleri, yani 33 bantın ikili farkları, toplam 32 adet fark görüntüsü (X eksen) ve 25 değere nicemlenmiş renk bölgeleri (Y eksen)

Her pul biber örneği için 1280x960x33'lük veri 32x25=800'lük bir matrise indirgenmiştir. Önerilen tekniğin başarımını gözlemleyebilmek için, her bant görüntüsünün ortalama parlaklık değerini de ayrı bir öznitelik kümesi olarak ele alıp, ortalama parlaklık değeri ile fark histogram matrisinin sınıflandırma başarımını kıyaslanmıştır. Şekil 7'de ortalama parlaklık değerinin ayrımsallığı görülmektedir. Şekilde görünen artılar ilgili banttaki, aflatoxinli

biberlerin, çemberler ise temiz biberlerin aykırı (outlier) örnekleridir. Sınıfların bükümleri (notch), ilgili dağılımın ortanca noktasını gösterir. Şekilden de görüleceği gibi, genelde orta noktalar birbirlerinden ayrıktırlar. Ancak sınıflar iç içe geçtiklerinden dolayı çok yüksek ayrımsallık beklenmemektedir.

Diğer taraftan 800 özneliğin doğruca sınıflandırıcıya giriş olarak verilmesi, örnek kümesinin (159 adet) sınırlı olmasından dolayı, uygun olmayacağından fark histogram matrisinden en fazla ayrımsallığı yapabilecek en önemli öznelikler, iki farklı öznelik seçimi yöntemi kullanılarak elde edilmiştir.. Bunun için Peng'in mRmR (minimum Redundancy maximum Relevance) yöntemi ile önerilen YSA'nın giriş nöronlarının ağırlıklarının toplam değerlerine göre sıralanması yöntemi kullanılmıştır (Peng, 2005). mRmR yönteminde ilgili fonksiyona bütün öznelikleri parametre olarak girdikten sonra 800, 50, 45, 40, 35, 30, 25, 20, 15, 10 ve 5 boyutlu, ortak bilgi (mutual information) temeline dayanan en önemli öznelik kümesi elde edilmiştir. Daha sonra bu öznelikler, YSA'ya verilerek sınıflandırıcının başarımı incelenmiştir. İkinci öznelik seçimi yöntemi olarak, YSA'nın kendi yapısı ve dinamiği kullanılmıştır. Şöyle ki, YSA, eğitim kümesi üzerinde öğreneceği zaman, nöronlar arasındaki bağlantıların ağırlıklarını devamlı bir şekilde günceller. Sonuçla ilişkisi olmayan nöronların (özneliklerin) bağlantı ağırlıkları, öğrenme süresince değişmez veya ağırlıklarının mutlak değeri toplamı 0'a doğru küçülmeye başlar. Yani YSA, sonuçla ilişkisi az olan veya olmayan giriş nöronlarını cezalandırmak için bağlantı ağırlıklarını -öğrenmeye olan etkileri azalacak şekilde- sifira doğru yakınsatır. Diğer taraftan, çıktıya en uygun ve ilinti içinde olan giriş nöronların bağlantı ağırlıklarını, öğrenme süresince, mümkün oldukça sıfırdan uzaklaştırır. Bu şekilde bu girdilerin öğrenmeye katkısı artırılmış olur. Bu fikirden yola çıkılarak 159 adet hiperspektral fark histogram matrisi (800 adet öznelik barındırmaktadır), iki gizli katmanı olan bir YSA'da eğitilmiştir. Eğitim boyunca bağlantı ağırlıkları toplamı değişen ve artan girdi nöronları, ağırlık toplamı değerlerine göre sıralanıp matris içerisindeki yerleri belirlenmiştir. Seçilen en önemli 800, 50, 45, 40, 35, 30, 25, 20, 15, 10 ve 5 öznelik kümesi, sınıflandırıcıya verilip özellik seçiminin başarımına etkisi ile mRmR sonuçları kıyaslanmıştır .

2.3. Sınıflandırıcı

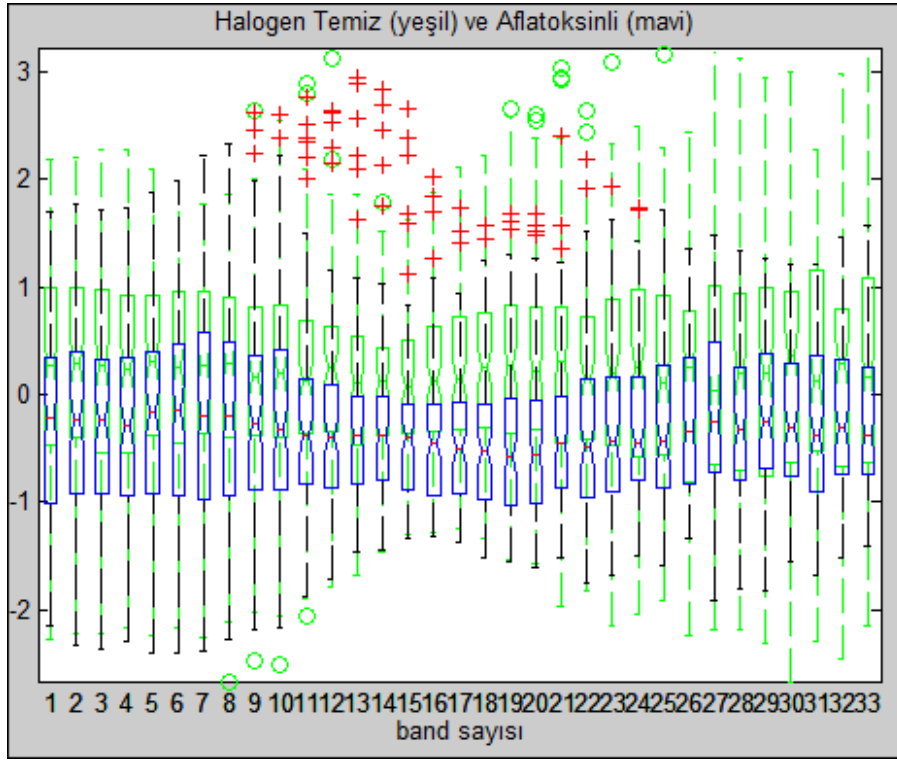
Malek vd.'ye göre, YSA'lar diğer istatistiksel desen algılama yöntemlerine kıyasla, örnek sayısının küçük olması probleminde en az etkilenen yöntemlerdendir (Malek vd., 2000). Bununla birlikte Bishop'un da belirttiğine göre, YSA'lar 'curse of dimensionality' yani boyutun fazla oluşundan kaynaklanan probleme karşı en sağlam ve toleranslı olan sınıflandırıcılardandır (Bishop, 1995). Dahası, YSA'lar öznelik vektörünün fazla olduğu durumlarda, boyut fazlalığı probleminde, k-NN, Parzen ve Quadratic sınıflandırıcılara göre daha az etkilenirler (Hamamoto vd., 1996). Gıda ürünlerinde ve gerçek dünya problemlerinde görünen uyuşmazlık ve doğrusal olmayan özelliklere karşı YSA, kendi içerisinde doğrusal olmayan aktivasyon fonksiyonu barındırdığı için, daha uygun bir sınıflandırıcı olarak bilinmektedir (ElMasrya ve Wangb, 2008). Bu nedenlerden ötürü, bu çalışmada ileri

beslemeli geri yayımlı YSA sınıflandırıcısının kullanımı uygun bulunmuştur. YSA'lar ile ilgili bir dezavantaj, kullanılan YSA topolojisinin ve parametrelerinin probleme uygun seçilmesi gereğidir. Karmaşıklık derecesi düşük bir YSA topolojisi, karmaşık bir problem için kullanıldığında, yeterli düzeyde öğrenme gerçekleşmeyecektir (underfitting). Aksi durumunda ise örnek kümesini ezberlemeye başlayacaktır (overfitting). Her iki durumdan da kurtulabilmek için sınıflandırıcının topolojisinin iyi ayarlanması gerekir. Bu çalışmada, iki gizli katmanlı YSA için her bir gizli katmandaki nöron sayılarını değiştirerek (3:1, 5:3, 7:5, 11:7, 17:11, 21:17, 29:19, 33:21) problemleri için en iyi yapı elde edilmeye çalışılmıştır.

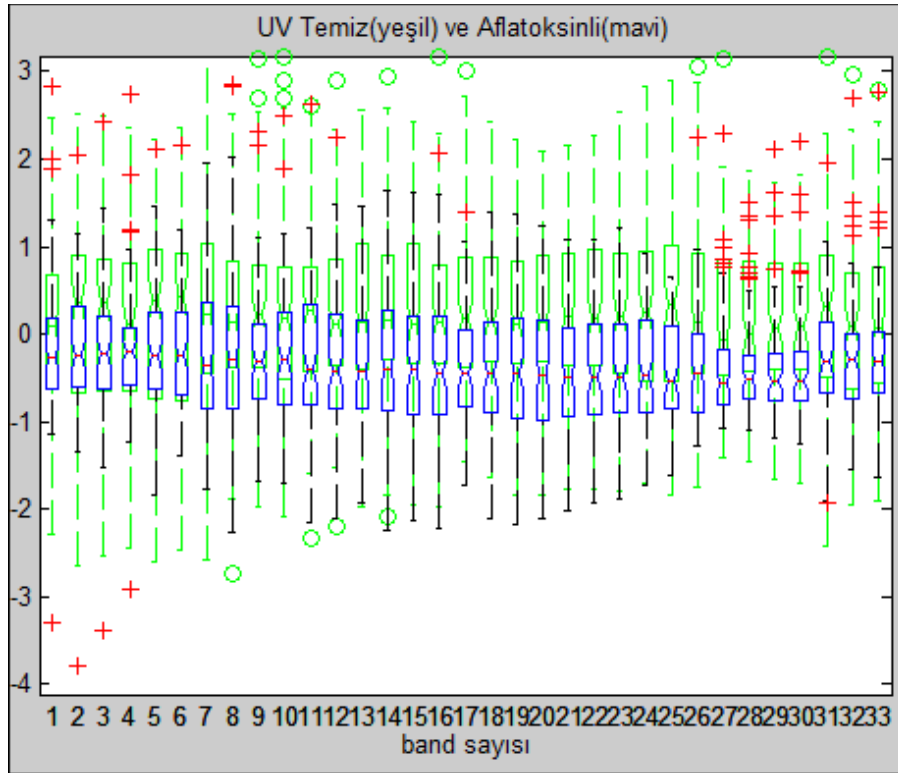
3. DENEY SONUÇLARI

Örnek küme, 53x3x33 görüntüden oluşmaktadır (53 pul biber örneği, pul biberin 3 farklı yerinden alınan görüntüler, 400-720 nm arasındaki 10 nm genişliğinde toplam 33 spektral bant). Öznitelik çıkarımından sonra örnek küme 53x3=159 adet hiperspektral fark histogram matrisinden ve 159 adet 33 boyutundaki ortalama parlaklık değerinden oluşur. Her bir matris 800 öznitelik (32 bant farkıx25 değere nicelenmiş renk bölgesi) içerir. 10 ppb kontaminasyon eşik değeri kabul edildiğinde 53 biberden 29 tanesi sağlıklı, 24 tanesi aflatoxinli olmaktadır. Sınıflandırıcının başarımını arttırmak için tüm veriler, min-max ve z-score normalizasyon teknikleri ile normalize edilmiştir. Yapılan araştırmalarda, z-score yöntemi, min-max normalizasyon yöntemine göre daha iyi sonuçlar verdiği için tüm deneylerde z-score normalizasyon tekniği kullanılmıştır. Öznitelik seçiminin, sınıflandırma başarımına katkısını daha iyi görebilmek için, ilk etapta spektral fark histogram matrisinin tüm elemanları, öznitelik seçimi yapılmadan, 2 gizli katmanlı YSA'ya verilmiştir. YSA'nın parametreleri; öğrenme hızı=0.1, momentum=0.1, çevrim sayısı=500 olarak sabitlenmiştir. Söz konusu YSA'nın çeşitli topolojilerdeki (1. gizli katmandaki nöron sayısı ve 2. gizli katmandaki nöron sayısı) başarımı değişimi Çizelge 3'te görülmektedir. Bu çizelgeden de görüldüğü gibi, validasyon ve test hatası dikkate alındığında 17:11 ve 21:17 topolojileri en düşük değerleri verdiğinden dolayı, öznitelik seçiminin yapısı daha basit olan 17:11 topolojisinde gerçekleştirilip bu öznitelik kümesi tüm deneyler için kullanılmıştır.

Öznitelik seçiminin sınıflandırma başarımına etkisini görebilmek için ilk başta mRmR tekniğini kullanarak 5, 10, 15, 20, 25, 30, 35, 40, 45, 50 ve 800 adet öznitelik için, birini dışarıda tut (leave one out) bağımsız geçerlilik sınaması (cross validation) yöntemi kullanılmıştır. Bu yöntemle göre 159 örnekten aynı bibere ait olan üç tanesini çıkararak geriye kalan 156 örnek ile öğrenme yaptıktan sonra, çıkarılan üç örnek, test kümesi olarak kullanılmıştır. Sonuçlar içerisinde en fazla oyu alma kuralına göre (majority voting) ilgili biber için sınıflandırıcının verdiği karar tespit edilmiştir. Bu şekilde 53 defa, YSA öğretilip test edilerek, ortalama hata oranı bulunmuştur. Bu işlem YSA'nın değişik topolojilerinde denenerek en uygun topoloji bulunmaya çalışılmıştır. Daha sonra, yukarıda anlatılan işlemler, önerilen öznitelik seçim yönteminde kullanılmıştır. Çizelge 4, 5, 6, 7, 8, 9 değişik YSA topolojileri ve öznitelik vektörü büyüklüğüne göre her iki yöntemle beslenmiş sınıflandırıcıların ortalama hata oranını göstermektedir.



(a)



(b)

Şekil 7. a) Halojen ve, b) UV için her bir hiperspektral bantın ayrımsallığı

Çizelge 3. Değişik topolojilerdeki YSA'ların öznelik seçimi yapılmadan önceki hata oranları
(validasyon ve test hatasına göre en iyi topolojiler 17:11 ve 21:17)

YSA topolojisi	3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
Öğrenme hatası	0.25	0.14	0.08	0.10	0.12	0,11	0,09	0,08
Validasyon hatası	0.50	0.35	0.40	0.20	0.22	0,24	0,26	0,32
Test hatası	0.60	0.38	0.39	0.35	0.34	0,32	0,35	0,37

Çizelge 4. Halojen ve her bandın ortalama parlaklık özneliği için değişik senaryolara göre test hatası oranları

Halojen		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,39	0,39	0,41	0,33	0,37	0,43	0,39	0,39
	önerilen	0,33	0,35	0,35	0,37	0,37	0,35	0,37	0,37
10	mRmR	0,33	0,37	0,3	0,33	0,3	0,32	0,33	0,35
	önerilen	0,39	0,41	0,43	0,39	0,3	0,45	0,33	0,35
15	mRmR	0,35	0,35	0,35	0,37	0,37	0,43	0,32	0,37
	önerilen	0,37	0,33	0,35	0,28	0,28	0,33	0,3	0,35
20	mRmR	0,33	0,35	0,35	0,3	0,3	0,3	0,33	0,35
	önerilen	0,35	0,32	0,3	0,35	0,32	0,32	0,35	0,35
25	mRmR	0,28	0,33	0,37	0,26	0,37	0,35	0,33	0,37
	önerilen	0,33	0,33	0,37	0,3	0,41	0,39	0,32	0,37
30	mRmR	0,32	0,35	0,26	0,41	0,35	0,33	0,35	0,35
	önerilen	0,32	0,3	0,33	0,33	0,37	0,37	0,33	0,37
33	mRmR	0,33	0,37	0,33	0,37	0,33	0,32	0,33	0,33
	önerilen	0,33	0,37	0,33	0,37	0,33	0,32	0,33	0,33

Çizelge 5. UV ve her bantın ortalama parlaklık özneliği için değişik senaryolara göre test hatası oranları

UV		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,47	0,47	0,47	0,52	0,47	0,39	0,47	0,47
	önerilen	0,41	0,41	0,45	0,41	0,37	0,37	0,43	0,41
10	mRmR	0,37	0,43	0,41	0,43	0,28	0,35	0,33	0,35
	önerilen	0,37	0,41	0,39	0,41	0,37	0,37	0,41	0,41
15	mRmR	0,41	0,32	0,3	0,28	0,3	0,45	0,39	0,37
	önerilen	0,37	0,32	0,39	0,3	0,3	0,32	0,33	0,35
20	mRmR	0,43	0,41	0,45	0,43	0,26	0,39	0,37	0,39
	önerilen	0,39	0,41	0,39	0,41	0,33	0,43	0,33	0,37
25	mRmR	0,37	0,43	0,41	0,35	0,41	0,32	0,3	0,32
	önerilen	0,35	0,43	0,41	0,37	0,45	0,32	0,3	0,33
30	mRmR	0,37	0,41	0,45	0,43	0,43	0,32	0,32	0,37
	önerilen	0,39	0,43	0,43	0,41	0,47	0,28	0,28	0,33
33	mRmR	0,35	0,35	0,33	0,33	0,3	0,37	0,3	0,33
	önerilen	0,35	0,35	0,33	0,33	0,3	0,37	0,3	0,33

Halojen ışık kaynağı altında alınmış görüntülerden elde edilen ortalama parlaklık değerinin öznitelik olarak kullanıldığı bu deneyde, 11:7 topolojisi hem mRmR da hem de önerilen yöntemde en düşük test hatasına ulaşmıştır (Çizelge 4).

UV ve 17:11 yapısı kullanıldığında mRmR yönteminin önerilen yöntemle göre daha iyi öznitelikleri seçtiği görülmektedir (Çizelge 5).

Çizelge 6. Halojen+UV veri tümleşimi ve her bandın ortalama parlaklık özniteliği için değişik senaryolara göre test hatası oranları

Halojen+UV		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,37	0,41	0,37	0,39	0,35	0,35	0,35	0,35
	önerilen	0,43	0,43	0,37	0,37	0,32	0,3	0,32	0,33
10	mRmR	0,39	0,41	0,41	0,26	0,24	0,26	0,22	0,26
	önerilen	0,35	0,43	0,33	0,33	0,2	0,28	0,28	0,28
15	mRmR	0,45	0,3	0,37	0,22	0,24	0,24	0,22	0,26
	önerilen	0,5	0,2	0,2	0,2	0,2	0,2	0,2	0,22
20	mRmR	0,41	0,54	0,3	0,32	0,22	0,32	0,2	0,28
	önerilen	0,26	0,39	0,28	0,26	0,22	0,33	0,2	0,22
25	mRmR	0,39	0,41	0,33	0,32	0,32	0,24	0,24	0,33
	önerilen	0,28	0,58	0,26	0,33	0,33	0,24	0,28	0,3
30	mRmR	0,3	0,3	0,3	0,33	0,2	0,2	0,22	0,24
	önerilen	0,28	0,3	0,5	0,3	0,24	0,17	0,18	0,2
33	mRmR	0,54	0,3	0,28	0,2	0,26	0,17	0,22	0,26
	önerilen	0,54	0,3	0,28	0,2	0,26	0,17	0,22	0,26

Veri tümleştirmeden sonra 21:17 topolojisi hem mRmR hem de önerilen yöntem, yaklaşık aynı sayıda öznitelik için en düşük test hatası oranı olan % 17'yi elde etmiştir. Veri tümleştirme sonucunda problem zorluğu arttığından dolayı YSA'nın ağ topolojisi de artma eğilimi göstermiştir (Çizelge 6).

Buraya kadar her bir görüntünün ortalama parlaklık değerlerini kullanarak yüzde 17 test hatası elde edilmiştir. Komşu bantların fark histogramları kullanılarak daha düşük test hatasına ulaşması hedeflenmiştir.

Halojen görüntülerinden elde edilen fark histogram matrisi özniteliklerinde bir önceki ortalama parlaklık değerine göre bir miktar iyileşme gözlenmiştir. En düşük test hatası % 18 (önerilen) ve % 20 (mRmR) elde edilmiştir (Çizelge 7).

Genelde UV görüntülerden elde edilen sonuçlarda mRmR yönteminin önerilen yöntemden daha başarılı olduğu görülmektedir. Elde edilen en düşük hata oranları ortalama parlaklık değeri özniteliğine göre, öznitelik sayısının artmasından dolayı mRmR için daha yüksek topolojide (29:19, 35 öznitelik) gözlenmiştir (Çizelge 8).

Çizelgelerden de görüldüğü gibi önerilen yöntem, özellikle de veri tümleştirmesinden sonra birçok öznitelik boyutlarında ve topolojilerinde daha az genelleme hatası vermektedir. Genel olarak çok küçük topolojilerde ve yüksek topolojilerde hata oranlarının arttığı

gözlemlenmiştir. Bunun nedeninin Bölüm 2.3’de de belirtildiği gibi, küçük topolojilerde genelleme yapamama ve yüksek topolojilerde ezberlemeden kaynaklandığı düşünülmektedir.

Çizelge 7. Halojen ve fark histogram matris özneliği için değişik senaryolara göre test hatası oranları

Halojen		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,28	0,26	0,26	0,24	0,24	0,24	0,24	0,26
	önerilen	0,52	0,58	0,54	0,54	0,49	0,43	0,41	0,43
10	mRmR	0,24	0,26	0,24	0,2	0,26	0,2	0,22	0,24
	önerilen	0,58	0,58	0,37	0,35	0,41	0,41	0,39	0,41
15	mRmR	0,5	0,3	0,28	0,3	0,32	0,26	0,28	0,28
	önerilen	0,52	0,37	0,35	0,39	0,39	0,32	0,26	0,32
20	mRmR	0,35	0,28	0,28	0,24	0,28	0,26	0,26	0,26
	önerilen	0,54	0,3	0,32	0,35	0,3	0,3	0,3	0,32
25	mRmR	0,32	0,28	0,32	0,28	0,28	0,26	0,28	0,28
	önerilen	0,33	0,28	0,26	0,24	0,22	0,26	0,28	0,3
30	mRmR	0,32	0,3	0,35	0,26	0,3	0,33	0,32	0,33
	önerilen	0,26	0,26	0,18	0,2	0,22	0,22	0,24	0,26
35	mRmR	0,3	0,26	0,22	0,3	0,22	0,3	0,22	0,26
	önerilen	0,26	0,28	0,24	0,24	0,2	0,22	0,22	0,24
40	mRmR	0,32	0,22	0,28	0,26	0,26	0,24	0,24	0,26
	önerilen	0,26	0,2	0,2	0,24	0,2	0,24	0,2	0,24
45	mRmR	0,28	0,28	0,28	0,32	0,26	0,26	0,28	0,28
	önerilen	0,22	0,22	0,24	0,2	0,2	0,22	0,22	0,24
50	mRmR	0,3	0,32	0,3	0,26	0,32	0,26	0,24	0,26
	önerilen	0,32	0,28	0,28	0,26	0,22	0,22	0,24	0,24
800	mRmR	0,41	0,35	0,35	0,43	0,41	0,32	0,32	0,35
	önerilen	0,41	0,35	0,35	0,43	0,41	0,32	0,32	0,35

Veri tümleştirmesinden önce her bandın ortalama parlaklık öznelikleri için halojen için mRmR yöntemi minimum hatayı (% 26) 11:7 topolojisi ve 15 öznelik sayısında vermekte iken, önerilen yöntem % 28 hatayı 11:7 topolojisi ve 25 öznelikte vermektedir. UV için, yaklaşık aynı hata oranları ve topolojiler gözlenmektedir. Veri tümleştirmesinden sonra hata oranı her iki yöntemde de % 17’ye düşmektedir. Diğer yandan fark histogram matrisi kullanıldığında, veri tümleştirmeden önce, Halojen için mRmR yöntemi %20’lik hatayı 11:7 topoloji ve 10 öznelik sayısında yakalamasına karşılık önerilen yöntem kullanıldığında hata oranının 7:5 topoloji ve 30 öznelikte % 18’e düştüğü gözlenmektedir. UV için, mRmR yönteminin daha iyi olduğunu görülmektedir. Şöyle ki, 29:19 topoloji ve 35 öznelikte % 22 olan test hatası, önerilen yöntemde %26’ya yükselmiştir. Ancak, veri tümleştirmesinden sonra mRmR yöntemi ile % 22 olan hata oranı, 29:19 topolojisi ve 35 öznelikte % 15’e düşmüştür. 400, 410 ve 420 nm bantları en iyi ayrışması vermektedir. Önerilen yöntem en küçük test hatasını % 15’de verdiği için bu problem için daha başarılıdır denilebilir.

Çizelge 8. UV ve fark histogram matris özneliği için değişik senaryolara göre test hatası oranları

UV		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,49	0,54	0,35	0,37	0,35	0,39	0,39	0,39
	önerilen	0,49	0,54	0,52	0,47	0,49	0,37	0,41	0,41
10	mRmR	0,5	0,33	0,32	0,35	0,37	0,39	0,35	0,39
	önerilen	0,5	0,41	0,3	0,28	0,33	0,26	0,26	0,33
15	mRmR	0,52	0,3	0,35	0,37	0,33	0,39	0,33	0,3
	önerilen	0,5	0,35	0,35	0,32	0,37	0,3	0,28	0,33
20	mRmR	0,33	0,33	0,35	0,28	0,33	0,26	0,28	0,28
	önerilen	0,43	0,35	0,33	0,3	0,3	0,32	0,3	0,32
25	mRmR	0,37	0,49	0,35	0,3	0,32	0,28	0,24	0,3
	önerilen	0,35	0,39	0,35	0,33	0,33	0,35	0,39	0,39
30	mRmR	0,3	0,26	0,32	0,26	0,24	0,28	0,24	0,28
	önerilen	0,35	0,35	0,37	0,37	0,35	0,37	0,39	0,39
35	mRmR	0,26	0,28	0,24	0,3	0,24	0,24	0,22	0,28
	önerilen	0,39	0,35	0,39	0,41	0,33	0,37	0,35	0,37
40	mRmR	0,28	0,22	0,3	0,28	0,28	0,33	0,3	0,3
	önerilen	0,45	0,32	0,33	0,35	0,37	0,39	0,37	0,39
45	mRmR	0,5	0,3	0,33	0,3	0,28	0,3	0,28	0,3
	önerilen	0,39	0,35	0,33	0,33	0,35	0,37	0,43	0,39
50	mRmR	0,52	0,28	0,28	0,28	0,35	0,26	0,28	0,28
	önerilen	0,52	0,35	0,37	0,37	0,33	0,39	0,37	0,39
800	mRmR	0,39	0,35	0,47	0,43	0,49	0,47	0,43	0,47
	önerilen	0,39	0,35	0,47	0,43	0,49	0,47	0,43	0,47

4. TARTIŞMA

Bu çalışmada, hiperspektral görüntüleme ve bilgisayarla görü sistemi kullanarak, aflatoksinli pul biberleri sağlıklı olanlarından ayırmak için bilgisayarla görü sistemi geliştirilmiştir. Bunun için geleneksel öznelik çıkarımı teknikleri yerine, hiperspektral bant farkı histogram matrisi yöntemiyle öznelikler çıkarılmış, bu özneliklerde boyutsal indirgeme için mRmR ve önerilen toplam nöron bağlantı ağırlığı yöntemi ile öznelik seçimi yapılmış ve değişik topolojilerdeki YSA'lar eğitilmiştir. Veri tümleştirmenin sınıflandırma başarımına etkisi izlenmiş ve önerilen yöntem ile % 85 doğru sınıflandırma oranı elde edilmiştir. Önerilen yöntemin, bu problem için mRmR yönteminden daha iyi sonuç verdiği gözlenmiştir. İleride, önerilen yöntemleri farklı sınıflandırıcılarda (Destekçi Vektör Makinesi (SVM), Lineer Ayrimsallık Analizi (LDA) ve Genetik Programlama) da deneyerek geçerliliğini inceleyip, sınıflandırıcılar arasında başarım kıyaslaması yapılacaktır.

Çizelge 9. Halojen+UV veri tümleşimi ve fark histogram matris öz niteliği için değişik senaryolara göre test hatası oranları

Halojen+UV		3:1	5:3	7:5	11:7	17:11	21:17	29:19	33:21
5	mRmR	0,39	0,52	0,39	0,39	0,35	0,32	0,32	0,35
	önerilen	0,39	0,49	0,22	0,22	0,2	0,24	0,24	0,26
10	mRmR	0,49	0,41	0,39	0,33	0,33	0,26	0,33	0,33
	önerilen	0,5	0,2	0,22	0,18	0,2	0,2	0,24	0,24
15	mRmR	0,5	0,32	0,33	0,32	0,3	0,32	0,28	0,32
	önerilen	0,5	0,2	0,26	0,24	0,28	0,22	0,2	0,22
20	mRmR	0,37	0,33	0,35	0,32	0,33	0,28	0,26	0,28
	önerilen	0,28	0,24	0,26	0,26	0,24	0,2	0,22	0,22
25	mRmR	0,35	0,33	0,33	0,32	0,33	0,32	0,28	0,32
	önerilen	0,3	0,52	0,26	0,2	0,24	0,2	0,2	0,24
30	mRmR	0,35	0,33	0,35	0,33	0,32	0,32	0,28	0,32
	önerilen	0,35	0,32	0,35	0,35	0,28	0,22	0,22	0,22
35	mRmR	0,3	0,37	0,35	0,35	0,33	0,32	0,28	0,35
	önerilen	0,3	0,3	0,28	0,26	0,3	0,28	0,15	0,24
40	mRmR	0,32	0,3	0,33	0,3	0,3	0,26	0,3	0,3
	önerilen	0,33	0,35	0,32	0,28	0,28	0,28	0,24	0,28
45	mRmR	0,28	0,33	0,32	0,28	0,30	0,22	0,33	0,32
	önerilen	0,47	0,26	0,26	0,24	0,2	0,18	0,17	0,2
50	mRmR	0,28	0,28	0,3	0,3	0,26	0,24	0,22	0,24
	önerilen	0,3	0,26	0,28	0,26	0,15	0,2	0,2	0,22
800	mRmR	0,28	0,28	0,26	0,22	0,22	0,26	0,3	0,26
	önerilen	0,28	0,28	0,26	0,22	0,22	0,26	0,3	0,26

TEŞEKKÜR

Dr. Habil Kalkan'a laboratuardaki ekipmanların teminindeki katkılarından dolayı, Dr. Tuğba Taşkaya Temizel ve Rahime Belen'e YSA'larla ilgili değerli yardımları ve mRmR algoritmasının kodlarını sağladıkları, İsa Ataş, Tahsin Duran ve Siraceddin Musaboğlu'na değişik bölgelerden biber örnekleri temin ettikleri için teşekkür ederiz.

KAYNAKLAR

- Bishop C.M. (1995): "Neural Network for Pattern Recognition", Oxford University Press Inc., USA.
- Bourlard H., Kamp Y. (1988): "Auto-Association by Multilayer Perceptrons and Singular Value Decomposition", Bio. Cybernetics, Cilt 59, s. 291-294.
- ElMasrya G., Ning Wangb C. V. (2008): "Detecting Chilling Injury In Red Delicious Apple Using Hyperspectral Imaging And Neural Networks", Postharvest Biology and Technology, Cilt 52, s. 1-8.

- Hamamoto Y., Uchimura S., Tomita S. (1996): "On The Behavior of Artificial Neural Network Classifiers In High-Dimensional Spaces", IEEE Trans. on Pattern Anal. and Machine Intel, Cilt 18, No. 5, s. 571-575.
- Hirano S., Okawara N., Narazaki S. (1998): "Near Infra Red Detection of Internally Moldy Nuts", Bioscience, Biotechnology, and Biochemistry., cilt 62, s. 102-107.
- Jardine D., Peter M. L. (2010): "Black Light Test for Aflatoxin Is QuestionableProcess", http://www.ksre.ksu.edu/news/sty/2003/blacklight_test082803.htm, 17 Ocak 2010'da kaydedilmiştir.
- Lipps P.E., Mills D. (2010): "Where to Send Grain Samples for mycotoxin Analysis", <http://www.oardc.ohio-state.edu/ohiofieldcropdisease/wheat/mycotoxin%20text2.htm>, 20 Ocak 2010'da kaydedilmiştir.
- Malek J. E., Alimi A. M., Tourki R. (2000): "Effect of the Feature Vector Size on the Generalization Error: The Case of MLPNN and RBFNN Classifiers", ICPR.
- Pearson T., Wicklow D., Maghirang E., Xie F., Dowell F. (2001): "Detecting Aflatoxin in Single Corn Kernels by Using Transmittance and Reflectance Spectroscopy", Transactions of the ASAE, Cilt 44, No. 5, s. 1247-1254.
- Peng H. (2005): "Feature Selection Based on Mutual Information", IEEE Trans. on Pattern Analysis and Machine Intel, s. 1226-1238.
- Yao H., Hruska Z., Brown R. L., Cleveland T.E. (2006): "Hyperspectral BGYF Imaging of Aflatoxin Contaminated Corn Kernels", Proc. Of SPIE, Cilt 6381, s. 63810B.
- Zeringue H. J., Shih B. Y. (1998) "Extraction and Separation of the BGYF Material from Aflatoxigenic Aspergillus spp. Infected Cotton Lint by HPLC-UV/FL", J. Agric. Food Chemistry, Cilt 46, s. 1071-1075.