

Boyutu Yüksek Görüntülerin Öznitelik Dönüşüm Yöntemiyle Analizi

Halit Çetiner¹, Ömer Kuşcu², Muzaffer Tatlı³

¹ Süleyman Demirel Üniversitesi, Araştırma ve Uygulama Has., Bilgi İşlem Merkezi, Isparta

² Süleyman Demirel Üniversitesi, Bilgi İşlem, Isparta

³ Kahramanmaraş Sütçü İmam Üniversitesi, Andırın MYO,

Kahramanmaraş

halitcetiner@sdu.edu.tr,

omerkuscu@sdu.edu.tr,

muzaffertatli@gmail.com

Özet: Endüstriyel ve kamusal ortamda insanların gündelik yapmış olduğu işlemler posta sıralama, farklı kamusal formlara el ile giriş yapmak gibi onlarca görev işgücü, zaman ve enerji kaybına neden olmaktadır. Bu tekdüze olan ve gündelik olarak periyodikleşen işlerde çalışanların zamanla dikkati dağılmakta ve ürün kalitesi düşmektedir. Ürün kalitesini artırabilmek için örüntülerin yani görüntüler içerisindeki rakamların yüksek oranlarda doğru ve hızlı sınıflandırılması gerekmektedir. Yüksek oranda doğru ve hızlı bir sınıflandırma gerçekleştirebilmek için boyut dönüşümü gerçekleştirmek gerekmektedir. Bu anlamda yapılan çalışmada temel bileşen analizi kullanılarak ortalama, kovaryans matrisi, özdeğer ve özvektör hesabı, skor değerlerinin belirlenmesi gibi hesapsal maliyeti yüksek işlemlerden geçirerek görüntü boyut azaltımı gerçekleştirilmiştir. Karşılaştırma yapmadan aynı işlem temel bileşen ve hesaplama maliyeti yüksek adımları kullanmadan gerçekleştirilmiştir. Yapılan karşılaştırmalar sonucunda elde edilen sonuçlar verilmiştir.

Anahtar Sözcükler: Temel Bileşen Analizi, Doğrusal Sınıflandırma, Kovaryans Matris, Öznitelik Dönüşümü, Özgün öznitelik.

Analysis by Feature Transformation of High Dimensional Images

Abstract: People's daily have made operations that dozens of labor tasks which are lead time and energy consuming such as mail sorting make to input of different public forms by hand. Employees are distracted and reduced product quality in these uniform and everyday jobs with time. Patterns that are figures into images should be classified as highly accurate and quickly of in order to improve product quality. Dimension transformation carried out requirement to perform highly accurate and fast a classification. In this sense, by passing from the high cost of computational processes such as mean, covariance matrices, accounting of eigenvalues and eigenvectors, the determination of score values using principal component analysis carried out image dimension reduction. The same process was carried out without using principal component and high cost of computational processes without comparison. As a result of the comparison, the results obtained are given.

Keywords: Principal Component Analysis, Linear Classifier, Covariance Matrix, Feature Transformation, Original Feature

1. Giriş

Asal bileşen analizi veya Karhunen-Loeve dönüşümü olarak da adlandırılan Temel

Bileşen Analizi, örüntü tanıma çalışmalarında geniş kullanım alanı bulunan bir altuzay izdüşüm yöntemidir [19]. Görünüşe dayalı

birçok uygulamada kullanılan boyut dönüşüm tekniğini temel alan uygulamalarının büyük bir çoğunluğu bu boyut indirgeme yöntemine dayanmaktadır. Bunun nedeni olarak karmaşık hesaplama algoritmaları gerektiren yeni tekniklere temel çözüm modeli oluşturması ve ayrıca vektör, matris, özdeğer, özvektör gibi matematik elemanlarının anlaşılmasındaki kolaylık gösterilebilir [20]. Temel bileşen analizi; değişkenlerden oluşan bir veri kümesinin varyans-kovaryans yapısını, bu değişkenlerin doğrusal birleşimleri yardımıyla açıklayarak, boyut indirgemesi ve yorumlanmasını sağlayan bir çok değişkenli istatistiksel yöntemdir [21]. Temel bileşen analizi vektör tabanlı, büyük boyutlu verileri ve birbiriyle ilişkili vektörleri, küçük boyutlu birbiriyle ilişkisiz vektörlere çevirirken özgün verinin dönüşümünü sağlayan bir analizdir. Bu dönüşüm sonrasında özgün veri temel bileşen şeklinde farklı bir boyutta temsil

edilmektedir. Farklı boyutta temsil edilme sonrasında elde edilen öznelikler, ilk özneliklerinin temel bileşenleri olarak adlandırılmaktadır. İlk temel bileşen varyans değeri en büyük olanıdır ve diğer temel bileşenler varyans değerleri azalacak şekilde sıralanmaktadır [19]. Gürültüye karşı düşük hassasiyet, bellek ve kapasite ihtiyaçlarının azalması, az boyutlu uzaylarda daha etkin çalışması TBA'nın temel avantajları arasında sıralanabilmektedir [24].

1901 yılında Karl Pearson'un başlattığı temel bileşenler analizi çalışmaları, 1933 yılında Hotelling tarafından geliştirilmiştir [22]. TBA çok sayıda birbiri ile ilişkili değişkenler içeren veri setinin boyutlarını veri içerisinde varolan değişimlerin mümkün olduğunca korunarak daha az boyuta indirgenmesini sağlayan bir dönüşüm tekniğidir [23].

2. Temel Bileşen Algoritması

Temel bileşen analizini gerçekleştirilen algoritmanın adımları:

1. her bir eğitim rakamının ortalama merkezini hesapla: her bir rakam grubu ('0', '1', ..., '9') için hesaplanmış rakam ortalamasını elde et. Sonrasında, her bir rakam için onu çıkartma yap.

$$\bar{x} = \frac{1}{M} \sum_{i=1}^M x_i \quad (1) \quad [26]$$

$$\Phi_i = x_i - \bar{x} \quad (2) \quad [26]$$

2. Kovaryans matrisi oluştur: veri seti matrisinin dış ürününden kovaryans matrisi hesapla. Aynı zamanda kovaryans matrisin transpozunu almak değerler ile ilişkili özdeğerleri normalize etmektedir.

$$A = [\Phi_1 \Phi_2 \Phi_3 \Phi_4 \Phi_5 \Phi_6 \Phi_7 \Phi_8 \Phi_9 \Phi_{10} \Phi_{11} \Phi_{12} \Phi_{13} \Phi_{14} \dots \dots \dots \Phi_M] \quad (3) \quad [26]$$

(NxM matrislik bir normalleştirme vektörü elde edildi), sonrasında

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T = AA^T \quad (4) \quad [26] \text{ veri dağılımı}$$

(NxN matrislik veri dağılımı karakterize edildi),

3. Özdeğer ve Özvektör ayrışımı: vektörlerin sütunları özvektörleri temsil etmektedir. Değerlerin köşegeni ise özdeğerleri temsil etmektedir. Bu temel vektörleri öz rakamları olarak etiketlenmiştir.

$$C: \lambda_1 > \lambda_2 > \lambda_3 > \dots > \lambda_N \quad (5) \quad [26]$$

özdeğerleri hesapla

$$C: u_1 > u_2 > u_3 > \dots > u_N \quad (6) \quad [26]$$

özvektörleri hesapla

C simetrik, $u_1, u_2, u_3, \dots, u_N$ temel oluşturmaktadır. (örneğin; herhangi bir vektör x veya gerçekten $(x - \bar{x})$, özvektörlerin doğrusal birleşimi olarak yazılabilir:

$$x - \bar{x} = b_1 u_1 + b_2 u_2 + \dots + b_N u_N = \sum_{i=1}^N b_i u_i \quad (7) \quad [26]$$

4. Sıralama: özdeğerleriyle özvektörleri sırala ve en önemli K tane özvektörlerin seçimini yap.

$\hat{x} - \bar{x} = \sum_{i=1}^K b_i u_i \quad K \ll N \quad (8) \quad [26]$ en büyük K özdeğerlere göre terimler tutulacaktır. $u_1, u_2, u_3, \dots, u_K$ temelinde, $\hat{x} - \bar{x}$ 'nin temsili;

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} \quad (9) \quad [26].$$

K değeri seçiminde;

$$\frac{\sum_{i=1}^K \lambda_i}{\sum_{i=1}^N \lambda_i} >$$

Eşik değeri (örneğin 0.9 veya 0.95)

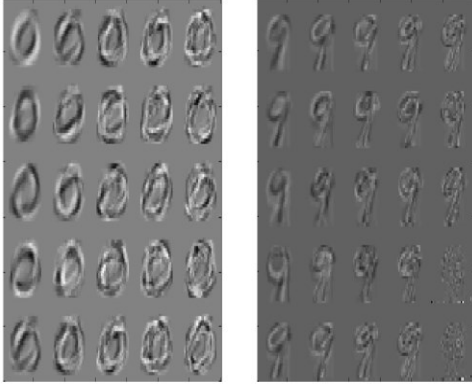
(10) [26]. Literatürde 0.95 değeri daha çok kabul görmektedir.

5. İzdüşüm: k boyutlu özvektör uzayında her bir rakamın izdüşümüyle izdüşüm haritası üret.

$$\begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} u_1^T \\ u_2^T \\ \dots \\ u_K^T \end{bmatrix} (x - \bar{x}) = U^T (x - \bar{x}) \quad (11)$$

[26]

0 ve 9 rakamları için PCA algoritması ile üretilen özrakamların temel bileşen ile temsil görüntüsü verilmiştir [25].



Şekil 2. PCA ile üretilmiş 0 ve 9 rakamlarının örnek temel bileşenleri [25]

3. Kullanılan Verisetleri ve Görüntü

Veritabanı

3.1 İris Veriseti

Bu veri çiçeklerin yapı ve renklerine göre elde edilen bir sınıflandırma verisidir. İris verisi her sınıfta 50şer olmak üzere toplamda 150 örneklı, dört öznitelikli (sepal uzunluğu, sepal genişliği, petal uzunluğu, petal genişliği) ve üç sınıflı (Setosa, Versi color, Virginica) bir veridir [1]. İris verisinde, sınıflardan biri diğerlerinden çok kolay ayrışırken diğer ikisi iç içe girmektedir. Burada her sınıfın yarısı eğitim diğer yarısı test veri kümelerine konulmuştur. İris verisine uygulanan öznitelik seçimi sonucunda elde edilen p dilsel kuvvet değerleri Tablo 1’te göstermektedir. Tablo 1’in son satırında ise her öznitelik için bütün sınıfların dilsel kuvvetlerinden elde edilen P ortak seçim değeri verilmektedir [2].

Sınıf/Öznitelik	1. Ö. Sepal U.	2.Ö. Sepal G.	3.Ö. Petal U.	4.Ö. Petal G.
Setosa	0.5	0.5	0.5	0.5
Versi color	0.0	0.0	1.0	1.0
Virginica	0.0	0.0	1.0	1.0
P değeri	0.0	0.0	0.5	0.5

Tablo 1. İris verisinin öznitelik basamağında elde edilen her sınıf ve öznitelikteki p dilsel

kuvvet değerleri [2]

Tablo 1’e bakıldığı zaman iris veriseti için 3. ve 4. özniteliklerin ayırt edici olduğu belirlenmiştir. Temel bileşen analizinde iris verisetinin öznitelikleri ayrı ayrı kullanılarak gerçekleştirilen doğrusal sınıflandırma sonuçları Tablo 2.’de verilmiştir.

Sınıflandırma Sonucu (%) /Öznitelik	1. Ö. Sepal U.	2.Ö. Sepal G.	3.Ö. Petal U.	4.Ö. Petal G.
Yüzde Oranı	36	44	45	93

Tablo 2. İris verisinin temel bileşen analizleri çıkartıldıktan sonra, her bir öznitelik için tek başına doğrusal sınıflandırmada kullanılmasından sonra elde edilen sınıflandırma sonuçları

Tablo 2.’de çıkartılan temel bileşenlerin ayrı ayrı sınıflandırmaya etkisi tespit edilmiştir. Tablo 3.’te ise çıkartılan temel bileşenlerin birlikte kullanımı sonucunda sınıflandırmaya etkisi tespit edilmiştir.

Sınıflandırma Sonucu (%) /Öznitelik	1. Ö. + 2.Ö.	1.Ö. . +2. Ö +3. Ö.	1.Ö. +2.Ö . +3.Ö . +4.Ö .	3.Ö. +4.Ö.
Yüzde Oranı	54	60	98	96

Tablo 3. İris verisinin temel bileşen analizleri çıkartıldıktan sonra, özniteliklerin birlikte doğrusal sınıflandırmada kullanılmasından sonra elde edilen sınıflandırma sonuçları

3.2 Pima Yerlileri Diyabet Veriseti

Bu veri kümesinde 768 örnek olup bunlardan ilk 576’sı eğitimde kalan 192’si ise test kümesinde kullanılmaktadır [1]. Sekiz öznitelikli (geçirdiği hamilelik sayısı, oral glikoz tolerans testinde iki saatlik plazma glikoz yoğunluğu, diyastolik kan basıncı, deri altı yağ dokusu, iki saatlik serum insülini, vücut ağırlık indeksi, diyabet soy ağacı

fonksiyonu, yaş) olan veri diyabet olan ve olmayan şeklinde iki sınıfa ayrılmaktadır. Sınıflandırma başarısı genelde %76 dolaylarındadır [1]. Tablo 4’de Pima Amerikan yerlileri diyabet verisinin öznitelik seçim sonuçları verilmektedir.

Öznitelik/Sınıf	1.S. (diyabet değil)	2.S. (diyabet)	P değeri
1.Ö.	0.40	0.58	0.240
2.Ö.	0.96	0.86	0.839
3.Ö.	0.49	0.43	0.215
4.Ö.	0.21	0.41	0.088
5.Ö.	0.20	0.68	0.137
6.Ö.	0.88	0.89	0.785
7.Ö.	0.48	0.65	0.320
8.Ö.	0.62	1.00	0.625

Tablo 4 Pima Amerikan yerlileri diyabet verisinin öznitelik seçim sonuçları [2].

Tablo 4’deki P değerleri sıralanır ve en büyük değerli olanları seçilirse sırasıyla ikinci (oral glikoz tolerans testinde iki saatlik plazma glikoz yoğunluğu), altıncı (vücut ağırlık indeksi), sekizinci (yaş) ve yedinci (diyabet soy ağacı fonksiyonu) öznitelikler ortak en iyi ayırt edici özniteliklerdir [2]. Bu ortak özniteliklerin arasında ikinci ve sekizinci öznitelikler aynı zamanda en iyi bireysel ayırt edici özniteliklerdir.

Temel bileşen analizinde Pima verisetinin öznitelikleri ayrı ayrı kullanılarak gerçekleştirilen doğrusal sınıflandırma sonuçları Tablo 5.’de verilmiştir.

Öznitelik/Sınıflandırma Sonucu (%)	Yüzde Oranı
1. Ö.	57
2.Ö.	59
3.Ö.	57
4.Ö.	61
5.Ö.	52
6.Ö.	52
7.Ö.	72
8.Ö.	60

Tablo 5. Pima verisinin temel bileşen analizleri çıkartıldıktan sonra, her bir özniteliğin tek başına doğrusal sınıflandırmada kullanılmasından sonra elde edilen sınıflandırma sonuçları

Tablo 5.’de çıkartılan temel bileşenlerin ayrı ayrı sınıflandırmaya etkisi tespit edilmiştir. Tablo 6.’da ise çıkartılan temel bileşenlerin birlikte kullanımı sonucunda sınıflandırmaya etkisi tespit edilmiştir.

Öznitelik/Sınıflandırma Sonucu (%)	Yüzde Oranı
2. Ö. + 7.Ö.	73
2.Ö.+8.Ö.	60
2.Ö.+6.Ö.	58
2.Ö.+6.Ö.+ 7.Ö.+ 8.Ö.	73
Tamamı	76

Tablo 6. Pima verisinin temel bileşen analizleri çıkartıldıktan sonra, özniteliklerin birlikte doğrusal sınıflandırmada kullanılmasından sonra elde edilen sınıflandırma sonuçları

Tablo 4., 5. ve Tablo 6. sonuçlarına bakıldığı zaman, temel bileşenleri çıkartılmış 2. ve 7. öznitelikler kullanılarak elde edilen sonuçlar tüm öznitelikler kullanılarak elde edilen sonuçlara yakın çıkmıştır.

Öznitelik/Sınıflandırma Sonucu (%)	Yüzde Oranı
2. Ö. + 7.Ö.	73
2.Ö.+8.Ö.	73
2.Ö.+6.Ö.	73
2.Ö.+6.Ö.+ 7.Ö.+ 8.Ö.	75
Tamamı	76

Tablo 7. Pima verisinin özniteliklerinin temel bileşen analizi çıkartılmadan doğrudan doğrusal sınıflandırmada kullanılmasından sonra elde edilen sınıflandırma sonuçları

Tablo 7. ve Tablo 6. Sonuçları birlikte değerlendirildiğinde Pima verisinde temel bileşen analizi ile öznitelikleri dönüşüme uğratmanın, özgün öznitelikler üzerinde bozulmaya sebep olduğu tespit edilmiştir.

3. 3. MNIST Veritabanı

NIST el yazısı karakter veri tabanı, çok büyük veri tabanı olup, içinde her kişiye ait paragraf, sayı, posta bilgileri yazım şekilleri bulunmaktadır. Bu veri tabanının eğitim kümesi, US nüfus sayım işçileri tarafından, test kümesi ise gönüllü olarak yüksek okul öğrencileri tarafından oluşturulmuştur [3].

Lecun, NIST veri tabanındaki eğitim ve test kümelerinin dağılımlarının örtüşmemesi üzerine bu veri tabanından yeni bir veri tabanı oluşturmuştur [4]. Çünkü birçok makine eğitimi, yapısal riski en küçültme kuralına göre yapılmaktadır. Eğer eğitim ile test kümesi örtüşmezse bu eğitim yapısı başarısız olmaktadır.

MNIST veritabanındaki karakterler, boyut olarak düzleştirilmiş ve karakterler imgenin ağırlık merkezinde yer alacak şekilde taşınmıştır [5]. MNIST veri tabanı NIST'in Özel Veritabanı-3(ÖVT-3) ve ÖVT-1 veri tabanlarından derlenmiştir. ÖVT-3 eğitim, ÖVT-1 ise test olarak NIST tarafından düzenlenmiştir. ÖVT-3 verileri ÖVT-1 verilerine göre daha temizdir ve daha kolay tanınmaktadır. Çünkü ÖVT-3 nüfus sayım çalışanları tarafından oluşturulurken, ÖVT-1 yükseköğretim öğrencilerinden gönüllü olarak oluşturulmuştur. Bu farklılıktan dolayı NIST'in eğitim ve test kümeleri karşılaştırılarak yeniden oluşturulmuştur[4].

MNIST'in eğitim kümesi; 30.000 örnek ÖVT-3'den, 30.000 örnek ÖVT-1'den

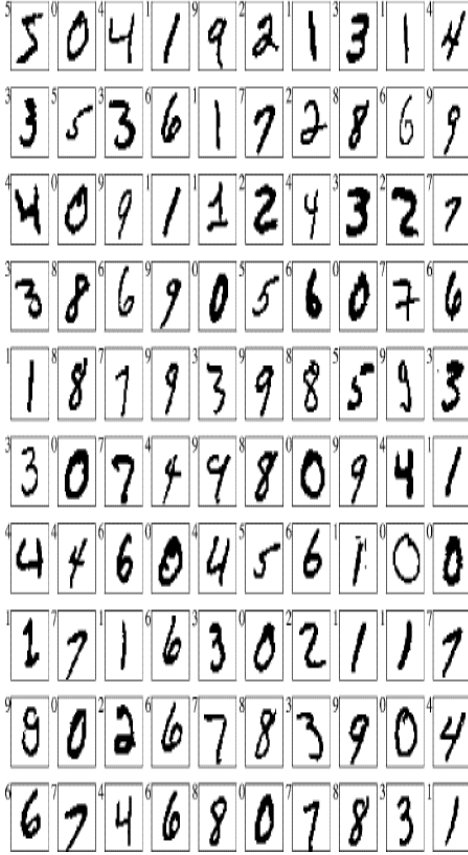
alınarak oluşturulmuştur. Test kümesi ise; 5.000 örnek ÖVT-3'den, 5.000 örnek ÖVT-1'den oluşmaktadır. Eğitim kümesindeki 60.000 örnek yaklaşık 250 yazara aittir. Eğitim ve test kümesindeki yazarlar birbiriyle ilişkisizdir[4].

Bu çalışmada el yazısı karakterleri olarak, MNIST veri tabanı kullanılmıştır. Bu veri tabanı NIST 19 veri tabanından derlenmiş ve eğitim kümesi 60000 adet rakamdan, test kümesi ise 10000 adet rakamdan oluşmaktadır [3,4]. NIST 19 veri tabanının yanı sıra, karakter tanıma çalışmalarında kullanmak için farklı ülkelere ait veri tabanları da bulunmaktadır(İngiliz, Kore, Arap, Çin, Kanji karakterleri...) [6-11]. Bu veri tabanının seçilmesinin nedeni ise NIST veri tabanının çok büyük olması ve bu nedenle de üzerinde çok fazla çalışma yapılmamasıdır [5]. NIST verisinin tamamını kullanmak kişisel bilgisayarların sınırlarını zorlamakta ve kilitlenmesine neden olmaktadır. Bu nedenle NIST verisini kullanan çalışmalarda verinin sadece belli bir bölümü kullanılmış ve bu bölümler için belli bir standart karşılaştırmaya gidilmemiştir. MNIST veri tabanı ile yapılmış çok fazla çalışma bulunmaktadır [12-18]. MNIST veri tabanında her rakam 28x28 boyutlarındadır. MNIST veri tabanında eğitim ve test kümeleri farklı sayıda rakam örneklerinden oluşturulmuştur. Veritabanında rakamların kimisi kalın, kimisi ince veya gürültü içermektedir [5].

Rakam	Eğitim Kümesi	Test Kümesi
1	6598	1135
2	5958	1032
3	6275	1010
4	5842	982
5	5421	892
6	5918	958
7	6265	1028
8	5851	974
9	5949	1009

0	5923	980
---	------	-----

Tablo 8. MNIST veri tabanındaki eğitim ve test kümesi için her rakama düşen örnek sayıları [5].



Şekil 9 MNIST veri tabanı eğitim kümesindeki ilk 100 örnek.

3.4 MNIST Veritabanı Görüntülerinde PCA Analizi

MATLAB ortamında gerçekleştirilen uygulamada iki yaklaşım değerlendirilmiştir. Birinci yaklaşım;

1. Adımda; MNIST veritabanı indirilmiştir. Aynı dizine kaydedilmiştir.
2. Adımda; MNIST veritabanı okunmuştur. 10000x1 test etiket seti, 784x10000 test görüntü veri seti, 60000x1 eğitim etiket veri seti, 784x10000 eğitim görüntü veriseti

alınmıştır. Görüntüler sonrasında $28 \times 28 = 784$ olarak boyutlandırılarak okunmuştur.

3. Adımda; özgün veri boyutu olan 50000 örnek görüntü üzerinde işleme yapabilmek çok zordur. Bu yüzden her bir rakam için 500 adet rastgele görüntü örneği alınmıştır ve böylece 5000 örnek görüntüden oluşan bir eğitim kümesi oluşturulmuştur.

4. Adımda; rasgele indirgenmiş değerlere temel bileşen analizi adımları uygulanmıştır. Literatürde genel olarak özdeğerlerin birikimsel olarak %95'ini alacak şekilde boyut sayısı seçilmesi bahsedildiğinden dolayı bu oran dışında kalanlar atılmıştır [26]. Belirtilen adımda temel bileşenlerin çıkartılmasında kullanılan işlemler temel bileşen analizi bölümünde ayrıntılı anlatılmıştır.

5. Adımda; Eğitim kümesindeki her örneğin indirgenmiş boyutlarını hesaplanmıştır.

6. Adımda; Test Örnekleri rastgele değerlerle 5000'e indirgenen test dizileri oluşturulmuştur. Bu örneklerin her biri için temel bileşenler adımları uygulayarak temel bileşenler ile çarparak boyut indirgemesi gerçekleştirilmiştir. Her örnek için Öklit uzaklığı hesap edilmiştir. Bu uzaklık değerlerine göre en yakın vektör tespit edilmiştir. Bu test örneğinin etiketini o vektörün etiketine ataması yapılarak işlem tamamlanıyor.

İkinci yaklaşım;

Birinci adımdaki 5 adım aynen geçerlidir. Farklılık bundan sonraki adımlarda geçerli olacaktır.

6. Adımda; Test Örnekleri rastgele değerlerle 5000'e indirgenen test dizileri oluşturulmuştur. Bu örneklerin her biri için temel bileşenler adımları uygulayarak temel bileşenler ile çarparak boyut indirgemesi gerçekleştirilmiştir. Her örnek için Öklit uzaklığı hesap edilmemiştir. Öklit hesaplaması temel bileşenlerde yapılmadığı için daha hızlı olacaktır. Bu uzaklık değerlerine göre en yakın vektör birden fazla tespit edilmiştir. Bu test örneğinin etiketini o vektörün etiketine ataması yapılarak işlem tamamlanıyor. 10 kez tekrar edilen işlem adımlarında ortalama olarak elde edilen

sonuçlar;

- PCA kullanılarak öklid uzaklığı hesaplama 30.737979 saniye sürmüştür ve 17 adet en yakın vektör hesaplanmıştır.
- PCA kullanılmadan öklid uzaklığı hesaplama 6.995797 saniye sürmüştür ve 82 adet en yakın vektör hesaplanmıştır
-

4. Sonuç ve Öneriler

Yüksek oranda doğru ve hızlı bir sınıflandırma gerçekleştirebilmek için boyut dönüşümü ile boyut dönüşümü gerçekleştirmeden yapılan analizlerin sonuçları karşılaştırılmıştır. Farklı veriseti ve veritabanlarında farklı sonuçlar elde edilmiştir. Literatürde yapılan çalışmalarda ayırt edici özneteliği tespit edilmiş verisetleri tercih edilerek bu verisetleri üzerinde temel bileşen analizinin etkisi incelenmiştir. Çalışmalar sonucunda iris veriseti üzerinde temel bileşen analizinin başarılı olduğu tespit edilmiştir. İris verisetinin sadece iki önemli özneteliği kullanılarak yani veri boyutunun sadece yarısı kullanılarak %96'lık bir başarı oranı elde edilmiştir. Pima verisetinde ise temel bileşen analizi ile öznetelikleri dönüşüme uğratmanın, özgün öznetelikler üzerinde bozulmaya sebep olduğu tespit edilmiştir. Eğitim ve Test Örnekleri rastgele değerlerle 5000'e indirgenen dizileri üzerinde deneysel çalışmalar sonucunda

Bu anlamda yapılan çalışmada temel bileşen analizi kullanılarak ortalama, kovaryans matrisi, özdeğer ve özvektör hesabı, skor değerlerinin belirlenmesi gibi hesapsal maliyeti yüksek işlemlerden geçirek görüntü boyut azaltımı gerçekleştirilmiştir. Karşılaştırma yapmadan aynı işlem temel bileşen ve hesaplama maliyeti yüksek adımları kullanmadan gerçekleştirilmiştir. Yapılan karşılaştırmalar sonucunda elde edilen sonuçlar verilmiştir.

5. Kaynaklar

[1] UCI Machine Learning Group, www.ics.uci.edu/~mlearn/, (2013).

[2] Çetişli, B., “ Öznetelik Seçiminde Dilsel Kuvvetli Sinir Bulanık Sınıflayıcı Kullanımı”, **Eskişehir Osmangazi Üniversitesi Müh.Mim.Fak.Dergisi**, 19(2), (2006).

[3] NIST handprinted forms and characters database(Special No. 19), **National Institute of Standarts and Technology U.S. Department of Commerce**, <http://www.nist.gov/srd/>, (2013).

[4] LeCun, Y., **MNIST OCR data**, <http://yann.lecun.com/exdb/mnist/>, (2013).

[5] Cetişli, B., “El Yazısı Karakter Tanıma: Dalgacık Moment Özneteliklerinin Yenilenen ANFIS ile Sınıflandırılması”, **Eskişehir Osmangazi Üniversitesi, Fen Bilimleri Enstitüsü, Elektrik-Elektronik Mühendisliği A.B.D., Doktora Tezi**, (2005).

[6] Kim, D., Bang, S.-Y., “A handwritten numeral character classification using tolerant Rough set”, **IEEE PAMI**, 22(9), 923-937, (2000).

[7] Alshebeili, S. A., Nabawib, A. A.-F., Mahmoud, S. A., “Arabic character recognition using 1-D slices of the character spectrum”, **Signal Processing, Elsevier Science B.V.**, 56, 59-75, (1997).

[8] Chim, Y. C., Kassim, A. A., Ibrahim, Y., “Character recognition using statistical moments”, **Image and Vision Computing, Elsevier Science B.V.**, 17, 299-307, (1999).

[9] Cheng, D., Yan, H., “Recognition of handwritten digits based on contour information”, **Elsevier Science Ltd., Pattern Recognition**, 31(3), 235-255, (1998).

[10] Cho, S.-J., Kim, J. H., “Bayesian network modeling of strokes and their relationships for on-line handwriting recognition”, **Elsevier Science Ltd., Pattern Recognition**, 37, 253- 264, (2004).

- [11] Chen, G.Y., Bui, T.D., Krzyzak, A., “Contour-based handwritten numeral recognition using multiwavelets and neural networks”, **Elsevier Science Ltd., Pattern Recognition**, 36, 1597–1604, (2003).
- [12] Burges, C.J.C., Schölkopf, B., “Improving the accuracy and speed of support vector learning machines”, **Advances in Neural Information Processing Systems** 9, MIT Press, Cambridge, MA, 375–381, (1997).
- [13] Krebel, U., “Pairwise classification and support vector machines”, **Advances in Kernel Methods: Support Vector Learning**, MIT Press, Cambridge, MA, 255–268, (1999).
- [14] Dong, J.X., Krzyzak, A., Suen, C.Y., “A multi-net learning framework for pattern recognition”, **Proceedings of the Sixth International Conference on Document Analysis and Recognition**, Seattle, 328–332, (2001).
- [15] Teow, L.-N., Loe, K.-F., “Robust vision-based features and classification schemes for off-line handwritten digit recognition”, **Pattern Recognition**, 35 (11), 2355–2364, (2002).
- [16] Belongie, S., Malik, J., Puzicha, J., “Shape matching and object recognition using shape contexts”, **IEEE Trans. Pattern Anal. Mach. Intell.**, 24 (4), 509–522, (2002).
- [17] Mayraz, G., Hinton, G.E., “Recognizing handwritten digits using hierarchical products of experts”, **IEEE Trans. Pattern Anal. Mach. Intell.**, 24(2), 189–197, (2002).
- [18] Liu, C.-L., Nakashima, K., Sako, H., Fujisawa, H., “Handwritten digit recognition: benchmarking of state-of-the-art techniques”, **Pattern Recognition Elsevier Ltd.**, 36, 2271 – 2285, (2003).
- [19] Yazar, I., Yavuz, H. S., Çay, M. A., “Temel Bileşen Analizi Yönteminin ve Bazı Klasik ve Robust Uyarlamalarının Yüz Tanıma Uygulamaları”, **Eskişehir Osmangazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi**, XXII(1), (2009).
- [20] Durucasu, H., “Asal Bileşen Analizi ve Bir Uygulama Denemesi”, **Yüksek Lisans Tezi, Anadolu Üniversitesi Fen Bilimleri Enstitüsü**, 89, (1991).
- [21] Yayıncı, A. Ö., “Temel Bileşenler Analizi için Robust Algoritmalar”, **Yüksek Lisans Tezi, Gazi Üniversitesi Fen Bilimleri Enstitüsü**, 56, (2006).
- [22] Filiz, Z., “Güvenilirlik Çözümlemesi, Temel Bileşenler ve Faktör Çözümlemesi”, **Anadolu Üniversitesi Bilim ve Teknoloji Dergisi**, 4(2), 211–222, (2003).
- [23] Çilli, M., “İnsan Hareketlerinin Modellenmesi ve Benzeşiminde Temel Bileşenler Analizi Yönteminin Kullanılması”, **Doktora Tezi, Hacettepe Üniversitesi Sağlık Bilimleri Enstitüsü**, 240, (2007).
- [24] Sütçüler, E., “Gerçek Zamanlı Video Görüntülerinden Yüz Bulma ve Tanıma Sistemi”, **Yüksek Lisans tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü**, 90, (2006).
- [25] Jain, G., Ko, J., “Handwritten Digits Recognition”, **Multimedia Systems, Project Report University of Toronto**, (2008).
- [26] Temel Bileşen Analizi, <http://www.iro.umontreal.ca/~pift6266/A08/ours/pca.pdf>, (2013).