

ÖNSÖZ

Yüksek Lisans öğrenim sürem boyunca bana her zaman yol gösteren saygıdeğer danışmanım Sayın Prof. Dr. Ahmet AKBAŞ' a, zamanını ve bilgilerini benimle paylaşan değerli arkadaşım Sebahattin BABUR' a, her konuda bana destek olan sevgili ablam Elif TURHAL UÇURUM' a ve beni büyütüp bugünlere getiren, maddi ve manevi her türlü desteği sağlayan canım aileme sonsuz teşekkürlerimi sunarım.

Haziran 2016

Uğur TURHAL



İÇİNDEKİLER

Sayfa

ÖNSÖZ.....	vii
İÇİNDEKİLER	ix
KISALTMALAR	xi
ÇİZELGE LİSTESİ.....	xiii
ŞEKİL LİSTESİ.....	xv
ÖZET.....	xvii
SUMMARY	xix
1. GİRİŞ.....	1
1.1 Tezin Amacı.....	2
1.2 Literatür Araştırması.....	3
1.3 Bölümlerin Organizasyonu	4
2. VERİ HAZIRLAMA.....	5
2.1 Yüz Bölgesinin Elde Edilmesi.....	6
2.2 Göz Çifti Bölgesinin Tespiti	7
2.3 Gözlerin Tespiti ve Örüntünün Standardizasyonu.....	9
2.4 Öznitelik Çıkartma.....	10
2.5 Dokusal Öznitelikler	11
2.5.1 Yerel ikili örüntü (LBP).....	11
2.5.2 Yönlü gradyanların histogramı (HOG).....	14
2.6 Öznitelik Seçme	17
2.6.1 İlişki temelli öznitelik altküme seçme (CFS)	18
3. SINIFLANDIRICILAR VE PERFORMANS KRİTERLERİ.....	21
3.1 Sınıflandırma Algoritmaları.....	21
3.1.1 Bayes ağı (BayesNet)	22
3.1.2 Naïve bayes.....	22
3.1.3 LibSVM (polinom çekirdek)	23
3.1.4 Sıralı minimal optimizasyon (SMO - polinom çekirdek).....	24
3.1.5 K-en yakın komşu (k-NN)	25
3.2 Performans Kriterleri	26
3.2.1 Çapraz doğrulama (cross validation)	27
3.2.2 Sınıflandırma süresi	27
3.2.3 Doğruluk (accuracy)	27
3.2.4 Kappa.....	28
3.2.5 Alıcı işletim karakteristiği (ROC)	29
4. SINIFLANDIRMA İŞLEMİ VE SONUÇ ANALİZİ.....	31
5. SONUÇLAR.....	51
KAYNAKLAR	55
ÖZGEÇMİŞ.....	61

KISALTMALAR

AAM	: Active Appearance Model
CFS	: Correlation-based Feature Subset Selection
CV	: Cross Validation
C-SVC	: C-Support Vector Classification
FN	: False Negatives
FP	: False Positives
FPR	: False Positive Rate
HOG	: Histogram of Oriented Gradients
k-NN	: k-Nearest Neighbors
LBP	: Local Binary Pattern
LDA	: Linear Discriminant Analysis
LibSVM	: Library for Support Vector Machines
PCA	: Principal Component Analysis
ROC	: Receiver Operating Characteristic
SFS	: Sequential Feature Selection
SMO	: Sequential Minimal Optimization
SVM	: Support Vector Machine
TN	: True Negatives
TP	: True Positives
TPR	: True Positive Rate



ÇİZELGE LİSTESİ

	<u>Sayfa</u>
Çizelge 3.1: Kappa istatistik değeri yorum tablosu.	28
Çizelge 4.1: Tüm öznitelikler ile yaş sınıflandırma performans sonuçları.	33
Çizelge 4.2: Öznitelik seçimi sonrası elde edilen veri setleri ile yaş sınıflandırma performans sonuçları.	35
Çizelge 4.3: Tüm öznitelikler ile cinsiyet sınıflandırma performans sonuçları.	37
Çizelge 4.4: Öznitelik seçimi sonrası elde edilen veri setleri ile cinsiyet sınıflandırma performans sonuçları.	39
Çizelge 4.5: Yaş ve Cinsiyet tahmininde kullanılan özniteliklerin birleştirilmesi sonucunda elde edilen veri seti ile yaş sınıflandırma ve performans analizi.	41
Çizelge 4.6: Yaş ve Cinsiyet tahmininde kullanılan özniteliklerin birleştirilmesi sonucunda elde edilen veri seti ile cinsiyet sınıflandırma ve performans analizi.	43
Çizelge 4.7: Önerilen yöntem kullanılarak elde edilen veri seti ile yaş sınıflandırma ve performans analizi.	45
Çizelge 4.8: Önerilen yöntem kullanılarak elde edilen veri seti ile cinsiyet sınıflandırma ve performans analizi.	47
Çizelge 4.9: Yaş gruplarının sınıflandırılmasında önerilen yöntem kullanılarak elde edilen sonuçların literatür ile karşılaştırılması.	49
Çizelge 4.10: Cinsiyet sınıflandırmada önerilen yöntem kullanılarak elde edilen sonuçların literatür ile karşılaştırılması.	49



ŞEKİL LİSTESİ

Sayfa

Şekil 2.1: Bireylerin yaş ve cinsiyet gruplarına göre dağılımı.	5
Şekil 2.2: 30-80 yaşları arasında insan yüzündeki değişimin 10 yaş aralıklarla gösterimi.	11
Şekil 2.3: LBP algoritması kullanılarak merkez pikselin etiketlenmesi.	12
Şekil 2.4: Yönelim açısı değerlerine ait örnek histogram.	15
Şekil 2.5: HOG Algoritmasının çalışma diyagramı.	16
Şekil 2.6: CFS algoritması ile alt küme seçimi.	18
Şekil 3.1: SVM algoritması ile destek vektörlerinin tespiti.	24
Şekil 3.2: k-NN algoritması ile örnek sınıflandırma.	25
Şekil 3.3: ROC karakteristiği.	29
Şekil 4.1: Yaş ve cinsiyet özniteliklerini bir arada içeren veri setinin hazırlanmasında önerilen yöntem.	31



YÜZ RESİMLERİNDEN ÇOKLU MAKİNE ÖĞRENMESİ YAKLAŞIMLARI İLE YAŞ VE CİNSİYET BİLGİSİNİN TESPİT EDİLMESİ

ÖZET

Bilgisayarlı sistemlerin yaşantımızın neredeyse her alanına girdiği günümüzde, bu sistemler ile yapılan ve hayatı kolaylaştıran uygulamalar oldukça önem arz etmektedir. Bu uygulamalar arasında güvenlik sistemlerinde oldukça sık kullanılan örüntü tanıma, günümüzde önemi gittikçe artan bir çalışma alanı haline gelmiştir. Bu alan; el yazısından karakter tahmini, ses ve görüntüde tespit edilen bozukluklardan hastalık teşhisi, insansız hava ve kara araçları gibi birçok amaç için kullanılmaktadır. Bu kapsamda yüz örüntülerinden kişi, karakter, yaş ve cinsiyet tespiti gibi birçok çalışma alanı da mevcuttur.

Bu tez çalışmasında yüz resimleri kullanılarak yaş ve cinsiyet tahmini yapılmıştır. Temelde iki farklı problem olan yaş ve cinsiyet tahmini bu çalışmada tek bir problem olarak ele alınmıştır. Bu iki problemi en iyi tanımlayacak minimum sayıda özneliğin belirlenmesi ve elde edilen öznelik kümesinin sınıflandırılması sonucunda minimum hesaplama sürelerinde maksimum doğruluk değerlerinin üretilmesi ana hedef olarak belirlenmiştir.

Çalışmada kullanılan örüntülerdeki yüz bölgelerinin tespitinde, literatürde bulunan birçok çalışmada yaygın olarak tercih edilmiş Viola-Jones algoritması kullanılmıştır. Bu algoritma, yüz bölgesinin düşey ile 15 dereceye kadar açığa sahip olduğu örüntülerde bu bölgeleri tespit etmede başarılıdır, ancak daha yüksek döndürme açılarında başarılı olamamaktadır. Geliştirilen algoritma yardımıyla, kullanıcı tarafından girilecek bir döndürme parametresine bağlı olarak, Viola-Jones algoritmasına düşey ile yapılmış her açıda yüz bölgelerini tespit edebilme becerisi kazandırılmıştır. Bu çalışmada döndürme parametresi 15 derece olarak belirlenmiş ve düşey ile 30 dereceye kadar açığa sahip tüm örüntülerde yüz bölgeleri tespit edilmeye çalışılmıştır. Elde edilen yüz örüntülerinden öznelik çıkartma işleminde iki farklı öznelik çıkartma algoritması hibrit olarak kullanılmış, bu algoritmalarından elde edilen veri setinin boyutları öznelik seçme algoritması kullanılarak düşürülmüştür.

Bu çalışmada kullanılan color FERET veri setine ait 800 örüntü üzerinde önerilen yöntem ile yapılan sınıflandırma sonuçlarına göre; yaş tespiti 2,63 saniyede %93,1250 doğrulukla, cinsiyet tespiti ise 2,86 saniyede %96,8750 doğrulukla gerçekleştirilmiştir.

Elde edilen sonuçlardaki sınıflandırma sürelerinin düşüklüğü ve birbirlerine yakınlığı, önerilen yöntemin bir gömülü sistem uygulamasında kullanılabilirliğini göstermektedir. Ayrıca önerilen yöntemle ait doğruluk değerlerinin, literatürde bulunan çalışmaların birçoğundan daha yüksek olduğu gözlemlenmiştir.



ESTIMATING OF AGE AND GENDER INFORMATION FROM FACE IMAGES WITH MULTIPLE MACHINE LEARNING METHODS

SUMMARY

The applications that facilitating the life with computer systems are quite important. Pattern recognition which is being used commonly in security systems has become an increasingly important field of study. This field of study is being used for many purposes such as character estimation from handwriting, medical diagnosis from disordered voice and physical appearance, controlling of unmanned aerial and ground vehicles. There are also many fields of studies like human, character, age and gender estimation from face images.

Age and genders are estimated by using face images in this thesis. Essentially age and gender estimation are two different problems. In this study both of these problems are considered as a single. Determining the minimum feature count to describes these problems and getting maximum accuracy values in minimum calculation times after classification are defined as the main target.

In this study the Viola-Jones algorithm, which is used commonly in literature, was used for detection of face patterns. This algorithm is successful only when the face area has a vertical angle of 15 degrees or less. In this study, an improved version of the Viola-Jones algorithm was developed which is able to detect face patterns with any vertical angle depending on a rotation parameter given by the user. The rotation angle was presumed as 15 degrees in this study. Thus, face patterns which have a vertical angle of 30 degrees was searched in images. Two different feature extraction method was used as hybrid for feature extraction from face patterns. Acquired dataset dimensions have been reduced by using feature selection algorithm.

According to the classification results that have been acquired by using the proposed method on 800 images from color FERET database; the age estimation was completed in 2,63 seconds with %93,1250 accuracy, the gender estimation was completed in 2,86 seconds with %96,8750 accuracy. All of these results obtained by using 20 fold cross validation method.

High accuracy values and short calculation times show that the proposed method is usable in embedded systems. It is observed that the accuracy results obtained by the proposed method are higher than many results in literature.



1. GİRİŞ

Örüntü; olay ve nesnelerin düzenli bir biçimde birbirini takip ederek gelişmesidir (Url-1). Buna göre insan yüzü bir örüntüdür ve kendisini diğer insanlardan ayırmada etkili bir bileşendir. Bir yüz örüntüsü doğru ve dikkatli incelendiğinde, kişiye ait birçok bilgiyi gösterebilir. Yaş, cinsiyet, ırk gibi bilgilerin yanında iyi eğitilmiş bir sistem ile kişinin yaşadığı bölge, bu bölgenin yaklaşık iklim durumu ve bunun gibi birçok spesifik ve tespiti zor bilginin tahmin edilmesi mümkün olabilmektedir. Ancak bu durum düşünüldüğü kadar kolay değildir. İnsan yüzünden yapılacak yaş ve cinsiyet tahmini insanlar için dahi zor bir problem haline gelebilmektedir.

Yüz örüntülerinden öznitelik çıkartma ve bunların sınıflandırılması işlemi 1964 yılına kadar uzanan bir geçmişe sahiptir (Bledsoe, 1964). Cinsiyet tahmini alanında yapılmış birçok çalışma olmasına rağmen, yaş tespitine yönelik yapılan çalışmaların azlığı dikkat çekmektedir. Bunun sebebi, cinsiyet tahmininin iki, yaş tahmininin ise çok sınıflı bir problem olmasıdır.

Cinsiyet tahmini alanında yapılan en eski çalışma 1991 yılına dayanmaktadır (Jain & Huang, 2004). Bu çalışmada yapay sinir ağları kullanılarak %10' un altında bir hata başarıları ile tahmin yapılmıştır. Aynı alanda yapılan bir başka çalışmada, yüz örüntüsü alanlara bölünmüş ve kullanılan analiz algoritmaları ile %93,7' lik doğruluk elde edilmiştir. Çalışmanın sonucunda, insan yüzünde cinsiyeti tespit etmede etkili alanın çene olduğu sonucuna varılmıştır (Kawano, Kato, & Yamamoto, 2004). Ayrıca Yang ve Moghaddam tarafından yapılan bir çalışmada Destek Vektör Makineleri (SVM) kullanılarak yapılan cinsiyet tahmininde hata oranı %3,4' e kadar düşürülmüştür (Yang & Moghaddam, 2000).

Tüm bu çalışmaların dışında, kişiye ait bir resimden, yıllar sonraki ve yıllar önceki görünümünün tahmin edilmesi de ilgi çeken konulardandır. Kayıpların bulunmasına yönelik yapılmış çalışmalarda, kişinin önceki yıllarda çekilmiş fotoğrafları kullanılarak doğal yaşlandırmalar yapılmış ve yıllar sonraki görünümü elde edilmiştir (Chellappa & Ramanathan, 2006; Wang, Shang, Su, & Lin, 2006).

Yukarıda bahsedilen görünüş tabanlı sınıflandırma algoritmalarının dışında geometrik özellikler kullanılarak cinsiyet sınıflandırma çalışmaları da yapılmıştır. V. Bruce, yüz üzerinde belirlediği 73 nokta ile Brunelli ve Poggio yüzden çıkardığı 16 geometrik öznitelik ile M. Kılınç; 34 geometrik öznitelik ile sınıflandırma yapmış ve cinsiyet tahmininde bulunmuştur (Burton, Bruce, & Dench, 1993; Kılınç, 2012; Poggio, Brunelli, & Poggio, 1992). Geometrik oran parametrelerinin parlaklığa karşı direnç gösterdiği bilinmektedir. Ancak kamera açısı, kafa pozisyonu, kamera parametreleri gibi değerlerden etkileniyor oluşu, kullanıcıları diğer ya da hibrit yöntemlerin kullanılmasına itmektedir.

Cinsiyet tahmininin iki sınıflı bir problem olması, giriş örüntülerinden elde edilen özniteliklerin sınıflandırılması ve bir örüntünün hangi sınıfa ait olduğunun belirlenmesi konusunda kullanıcıya kolaylık sağlamaktadır. Buna karşın yaş tahmini çok sınıflı bir problemdir ve sınıflandırıcının hata yapma olasılığını oldukça arttırmaktadır. Bunun sebebi; iki sınıflı sistemlerin doğrusal ya da parabolik yöntemlerle birbirinden ayrıştırılabilmemesine karşın çok sınıflı sistemlerin (4,5,6 vb.) ayrıştırılmasında yaşanan güçlülüdür. İki sınıflı verileri ayrıştırmada yüksek başarı gösteren birçok sınıflandırma algoritması, çok sınıflı verileri ayrıştırmada performans kaybı yaşamaktadır.

Bu tez çalışmasında; yaş ve cinsiyet tahmin sistemlerinde sıklıkla kullanılan color FERET veri seti kullanılmıştır. Giriş örüntülerinin işlenmesi, öznitelik çıkartma ve öznitelik seçme işlemleri MATLAB uygulamasında, sınıflandırma işlemleri ise WEKA uygulamasında gerçekleştirilmiştir. Yapılan işlemler sonucunda bireylerin yaş grupları ve cinsiyetlerinin yüksek başarı oranları ile tespit edilmesi sağlanmıştır.

1.1 Tezin Amacı

Bu tez, giriş örüntülerinden elde edilen yüz bölgelerinin işlenmesi ve sınıflandırılması konusunu ele almıştır. Sınıflandırma sonucunda ulaşılan doğruluk değerleri ve hesaplama sürelerinde performans artışı sağlamak için bir yöntem önerilmiştir. Elde edilen sonuçlar, özellikle gerçek zamanlı bir sistem tasarlanırken, önerilen yöntemin geleneksel yöntemlere oranla daha kullanışlı olduğunu göstermektedir.

1.2 Literatür Araştırması

Yüz örüntülerinden yaş ve cinsiyet tahmini amacıyla birçok çalışma yapılmıştır. Bu çalışmalarda genellikle geometrik öznitelikler cinsiyet tahmininde, dokusal öznitelikler ise yaş tahmininde kullanılmıştır. Önceki çalışmalarda farklı yöntemler ve olaylar denenerek insan yüzünün tanınması, şekillendirilmesi ve sınıflandırılması problemlerine çözümler aranmıştır. Bir çalışmada, insan yüzünde zamana bağlı olarak meydana gelen değişimler modellenmiş ve yüz tanıma sistemleriyle sınıflandırılmıştır (Heikkila & Pietikainen, 2006). Diğer bir çalışmada ise insan yüzlerinin harmanlanması sonucunda yeni yüzlerin oluşturulması, böylece gelecek nesillere ait yüz kestiriminin yapılması amaçlanmıştır (Burt & Perrett, 1995). Bir başka çalışmada Aktif Görünüm Modeli (AAM) kullanılarak yüz tanıma algoritması geliştirilmiş (Cootes, Edwards, & Taylor, 1998), bir diğerinde ise, bu çalışmada da kullanılan yerel ikili örnekler algoritmasıyla yüz güzelliği kararı üzerinde durulmuştur (V. Nabyev & Koçak, 2011).

İnsan yüzünden yaş bilgisinin elde edilmesi ile ilgili yapılan çalışmalarda ise iki farklı temel konu ele alınmıştır. Bunlardan biri kişinin yaklaşık yaşının tahmin edilmesi (Gao & Ai, 2009), diğeri ise hangi yaş grubu aralığına ait olduğunun belirlenmesidir (Kwon & Lobo, 1994; V. V. Nabyev & Günay, 2010). Yaş grubu aralıklarının belirlenmesi, yaklaşık yaşın tahmin edilmesine oranla daha yüksek sonuçlarla tespit edilebilmektedir. Bunun sebebi, artan sınıf değerlerinin, sınıflandırıcı kararlarını olumsuz etkilemesidir.

Yüz ve cinsiyet tahmininde yapılan bir başka çalışmada ise yüz tespitinin hassasiyetini arttırmak ve öznitelik uzayını genişletmek için henüz veri setine eklenmemiş ancak eklenmesi muhtemel yüzleri belirlemek hedef alınmıştır. Farklı yüz resimlerinin birleştirilmesi ve bunlara ait özniteliklerin çıkartılmasıyla veri kümesinde çeşitlilik sağlanmıştır (Karungaru, Fukumi, & Akamatsu, 2003).

Yaş gruplarının belirlenmesi ve tespit edilmesi üzerine yapılan bir başka çalışmada, Viola-Jones algoritması kullanılarak giriş örüntülerinden elde edilen yüz bölgesi düzleştirilerek normalize edilmiş, maskelenmiş ve Yerel İkili Örüntü (LPB) algoritması ile histogramları çıkartılmıştır. Elde edilen öznitelik uzayı Temel Bileşen Analizi (PCA) ile küçültülmüş ve yeni öznitelik uzayına dönüştürülmüştür. Veri kümesi sınıflandırılmış ve bireyin ait olduğu yaş grubu %89 doğrulukla tahmin edilmiştir (Günay & Nabiyev, 2008; Viola & Jones, 2004). Çalışmada kullanılan Temel Bileşen Analizinin en büyük avantajı; yüksek boyutlu öznitelik uzaylarını daha düşük boyutlara indirgeyerek hesaplama süresi ve maliyeti açısından kolaylık meydana getirmesidir.

1.3 Bölümlerin Organizasyonu

Literatür araştırması ve sonuçların verildiği giriş bölümünün ardından, ikinci bölümde; verilerin hazırlanması amacıyla örüntülerden öznitelik çıkartmada kullanılan Yerel İkili Örüntü (LBP) ve Yönlü Eğimlerin Histogramı (HOG) yöntemleri açıklanmış, elde edilen öznitelik kümelerinin boyutlarını azaltmak için kullanılan İlişki Temelli Öznitelik Altküme Seçimi (CFS Subset Evaluation) algoritmasından bahsedilmiştir. Üçüncü bölümde; cinsiyet ve yaş sınıflandırması amacıyla kullanılan Naïve Bayes, Bayes Ağı, LibSVM, Sıralı Minimal Optimizasyon (SMO) ve k-En Yakın Komşuluk (k-NN) algoritmalarından ve bu algoritmaların değerlendirilmesinde kullanılan performans kriterlerinden bahsedilmiştir. Dördüncü bölümde; elde edilen sınıflandırma sonuçları tablolarla gösterilmiş ve önerilen yönteme ait sonuçlarla karşılaştırılmıştır. Ayrıca önerilen yöntemin başarımı, literatürde elde edilmiş diğer sonuçlarla karşılaştırılmıştır. Beşinci bölümde ise elde edilen veriler değerlendirilmiş ve yapılabilecek iyileştirme çalışmalarından bahsedilmiştir.

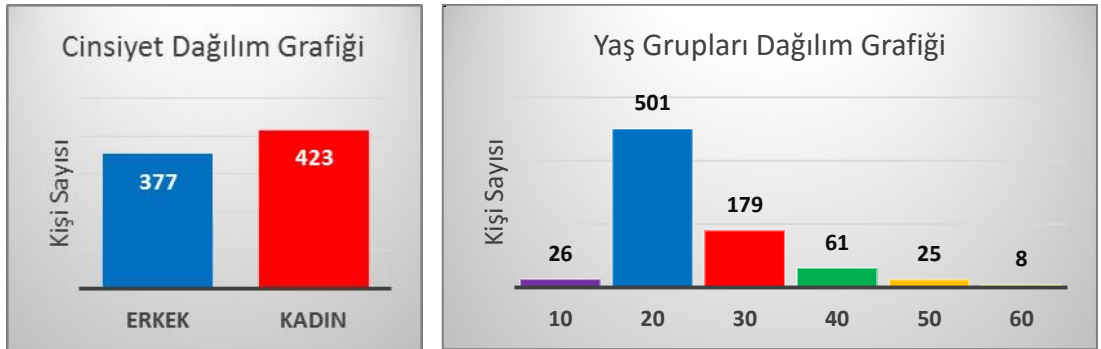
2. VERİ HAZIRLAMA

Yaş ve cinsiyetin belirlenmesinde en önemli etkenlerden biri şüphesiz yüz görüntüsüdür. İyi işlenmiş bir yüz görüntüsünden bireyin yaşı, cinsiyeti, ırkı, yaşadığı yerin iklim şartları ve hatta sosyo-ekonomik durumu bile tahmin edilebilir. Ancak böyle bir tahminin yapılabilmesi için güçlü bir modele ve bu modeli iyi çözümleyecek sınıflandırıcılara ihtiyaç vardır.

Sınıflandırıcılar; daha önceden öğrenilmiş bilgilere dayanarak, sistem girişine karşılık verilecek cevabın tahmin edilmesi mantığına göre çalışırlar. Birbirine benzer örneklerin fazla sayıda kullanılmasıyla sistemin eğitilmesi, sınıflandırıcının sonucunu etkileyen en temel etkenlerdendir. Sistemin eğitileceği parametrelerin doğru olarak belirlenmesi ise bir optimizasyon problemidir.

Bu çalışmada kullanılan color FERET veri seti 1199 bireye ait farklı açı, poz, ışık ve yüz ifadesi ile çekilmiş 14,051 adet gri tonlamalı yüz fotoğrafı içermektedir (Phillips, Moon, Rizvi, & Rauss, 2000; Phillips, Wechsler, Huang, & Rauss, 1998; Url-2). Veri seti oluşturulurken çekilen fotoğraflarda çekim saati, çekim açısı, ışık şiddeti, denek bilgileri gibi tüm detaylar kaydedilmiş ve bu veriler eğitim/test kümesi olarak kullanıma sunulmuştur.

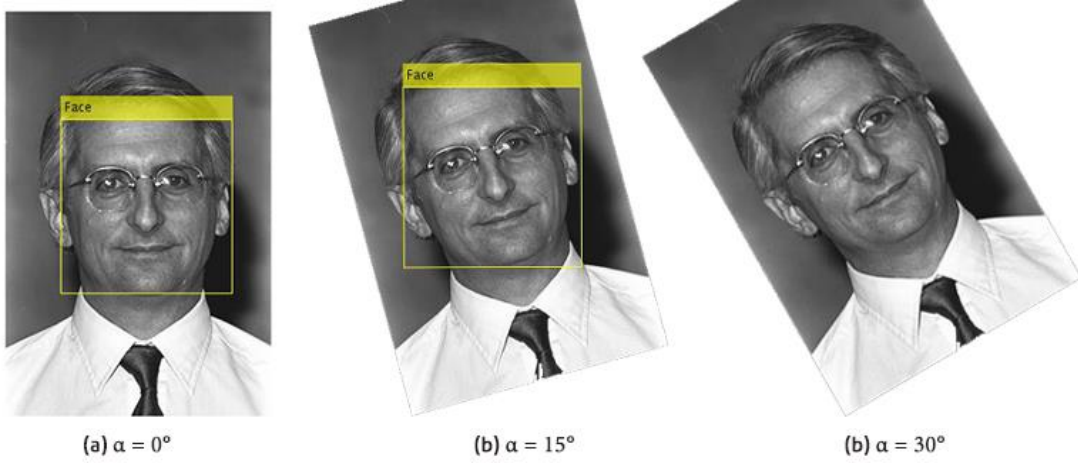
Sınıflandırma işlemi yapılırken tüm veri setini oluşturan 14,051 görüntünün tamamı yerine, düzenlenen algoritma tarafından yüz ve göz bölgeleri tespit edilebilen 800 bireye ait görüntüler kullanılmıştır. Kullanılan verilere ait yaş ve cinsiyet dağılım grafiği Şekil 2.1’ de verilmiştir.



Şekil 2.1: Bireylerin yaş ve cinsiyet gruplarına göre dağılımı.

2.1 Yüz Bölgesinin Elde Edilmesi

Veri setinin hazırlanması aşamasındaki ilk adım giriş örüntüsünden yüz bölgesinin elde edilmesidir. Bu işlem için diğer çalışmalarda da sıkça kullanılan Viola-Jones algoritması kullanılmıştır. Viola-Jones algoritması; Haar filtreleri kullanılarak öznitelikleri çıkartılan örüntüler ile bir eğitim sistemi oluşturur. Bu sistem içerisinde yüz örüntülerini ve yüz örüntüsü olmayan verileri içerir. Ardından test edilecek örüntüden üretilen Haar öznitelikleri, eğitim kümesi kullanılarak sınıflandırılır ve yüz tespiti yapılır (Kılınç, 2012). Uygulama; kısa tepki süresi ile oldukça kullanışlıdır ancak birtakım dezavantajları olduğu bilinmektedir. Örnek olarak giriş örüntüsünde düşey ile 15 dereceden fazla açı yapan yüz bölgelerinin tanınması mümkün olamamaktadır. Bunun sebebi ise kullanılan Haar filtrelerinin sabit yapısıdır. Resim 2.1’ de farklı döndürme açılara sahip örüntülerde yüz bölgelerinin Viola-Jones algoritması tarafından tespiti gösterilmiştir.



Resim 2.1: Viola-Jones algoritması kullanılarak yüz bölgesinin tespiti.

Yukarıda görülen üç resim, düşey ile sırasıyla 0, 15 ve 30 derece açı yapmaktadır. Giriş örüntüsü olarak Viola-Jones algoritmasına verilen bu örüntülerde yüz bölgeleri, düşey ile 0 ve 15 derecelik açı yapanlarda başarıyla tespit edilmiş ancak 30 derece döndürme açısına sahip örüntüde tespit edilememiştir. Bu problemin önüne geçebilmek adına geliştirilen algoritmada örüntü önce 15 derece sola, ardından ilk konumundan 15 derece sağa döndürülmüş ve yüz bölgesi aranmıştır. Algoritmada kullanılan döndürme açısı kullanıcı tarafından belirlenebilir yapıda esnek tutulmuştur. Çalışmada kullanılan veri seti oldukça fazla sayıda örüntü içerdiğinden, zaman kaybı yaşamamak adına döndürme açıları 15 derece olarak belirlenmiştir.

2.2 Göz Çifti Bölgesinin Tespiti

Bir önceki aşamada yüz bölgesine ulaşılan örüntüde sıra göz çifti bölgesinin tespitine gelmiştir. Yine aynı tespit algoritması kullanılarak, örüntü içerisindeki göz çiftlerini içeren bölgeler aranmıştır.

Burada öncelikli olarak gözleri tespit etmek yerine göz çifti bölgesinin tespit edilmesinin amacı; yüz örüntüsünde birbirine benzeyen birçok alanın mevcut olması ve Viola-Jones algoritmasının bu alanların tümünü tespit etmesidir. Örneğin; yüzde bulunan siyah bir benek, o bölgenin algoritma tarafından göz olarak algılanmasına sebep olabilmektedir. Resim 2.2' de Viola-Jones algoritması tarafından göz olarak tespit edilen ancak bir kısmı hatalı olan bölgeler gösterilmiştir.



Resim 2.2: Viola-Jones Algoritması tarafından hatalı tahmin edilen göz bölgeleri.

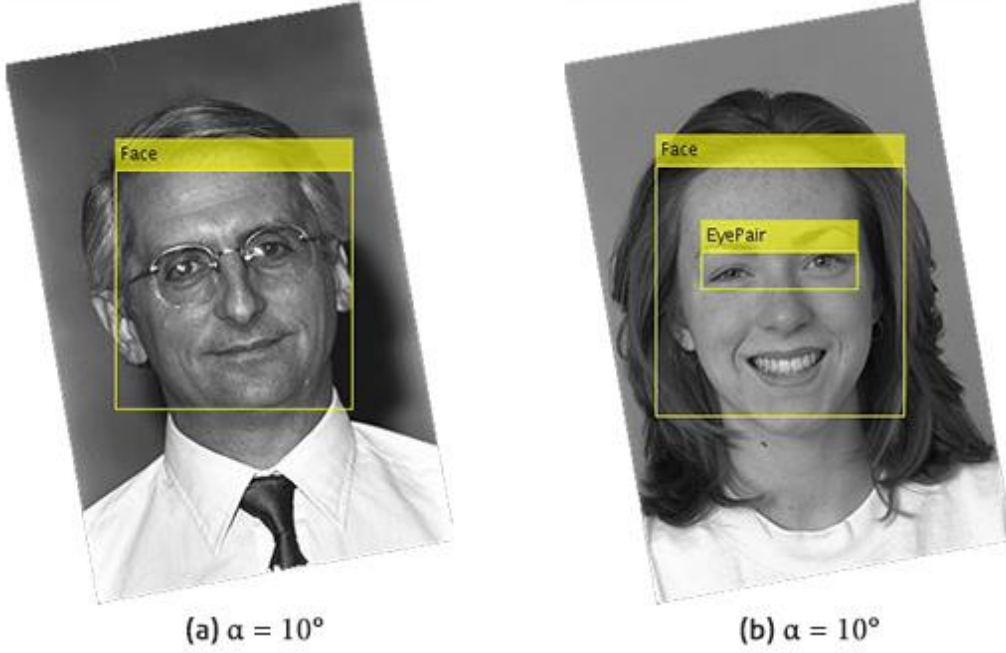
Bu problemin önüne geçebilmek için algoritmada göz bölgeleri aşağıdaki aşamalarla tespit edilmiştir:

- Giriş örüntülerinden yüz bölgeleri tespit edilir.
- Tespit edilen yüz bölgesi içerisinde göz çifti bölgesi aranır.
- Tespit edilen göz çifti bölgesi içerisinde yer alan sağ ve sol gözler tespit edilir.

Bu aşamalar ile algoritma tarafından tespit edilen hatalı bölgelerin minimize edilmesi amaçlanmıştır.

Göz çifti ve göz bölgelerinin tespitini etkileyen birçok bozucu etken mevcuttur. Aşırı gülümseme sonucu gözlerin kapanması, renkli cam bulunduran gözlüklerin kullanılması, kısık ya da kapalı gözle çekilmiş fotoğraflar, bu bölgelerin tespit edilmesini zorlaştırmaktadır.

Resim 2.3' te aynı döndürme açısına sahip (10 derece) ancak gözlük kullanan ve kullanmayan bireylerde, Viola-Jones algoritmasının yüz ve göz çifti bölgelerini tespitindeki başarısı gösterilmiştir.



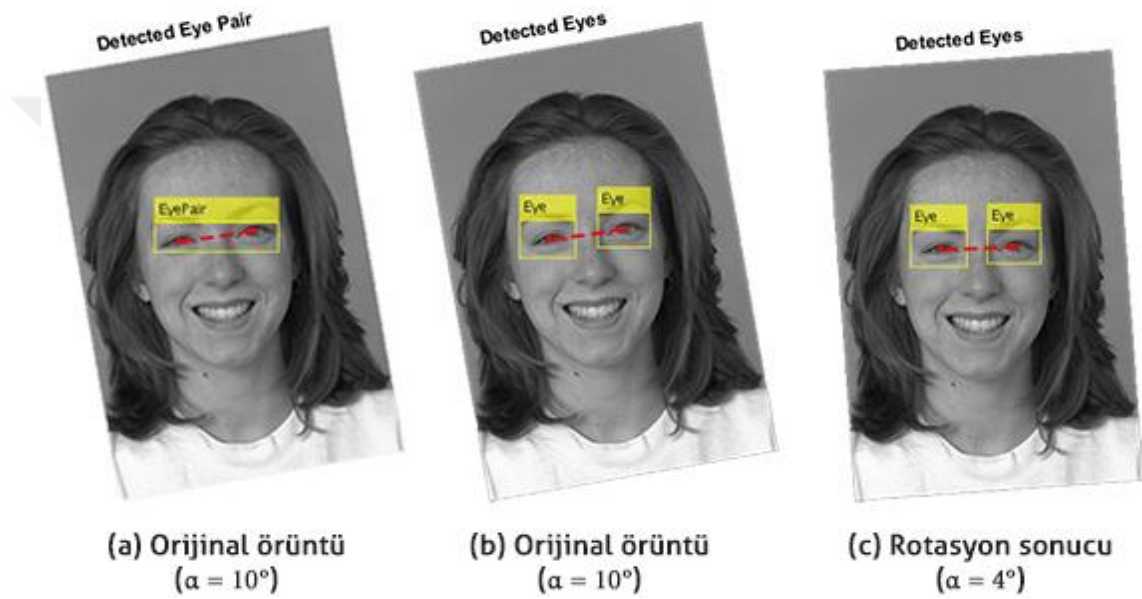
Resim 2.3: Yüz bölgeleri içerisinde bulunan göz bölgelerinin tespiti

Görüldüğü üzere iki giriş örüntüsünde de yüz bölgeleri başarıyla tespit edilmiş ancak gözlük kullanan bireyde göz çifti bölgesi tespit edilememiştir. Buna karşın diğer fotoğraftaki bireyin göz çifti bölgesi, herhangi bir bozucu etken olmadığından başarılı bir şekilde tespit edilebilmiştir.

Düşük örüntü sayısına sahip uygulamalarda bu problemler çeşitli optimizasyon yöntemleri ile çözülebilir ancak bu çalışmada kullanılan veri setinin barındırdığı yüksek örüntü sayısı, kullanılacak örneklerin seçilmesinde esneklik sağlamaktadır. Veri seti hazırlanırken, tüm bölgelerin algoritma tarafından tespit edilebildiği örüntüler kullanılmıştır. Bu bölgelerden herhangi birinin tespit edilemediği durumda ise kullanıcıya hata mesajı verilmiş ve ilgili örüntü oluşturulan veri setinin dışında tutulmuştur.

2.3 Gözlerin Tespiti ve Örüntünün Standardizasyonu

Giriş örüntüsünden yüz ve göz çifti bölgelerinin tespit edilmesinden sonra sıra gözlerin tespitine gelmiştir. Buradaki temel amaç; düşey ile yapılan açının belirlenmesi ve uygulanacak döndürme işlemi ile bu açının sifıra yaklaştırılmasıdır. Bu işlem sonucunda tüm örüntülerde göz, ağız ve burun gibi ayırt edici noktalar benzer alanlara denk gelir ve verilerde standardizasyon sağlanır. Resim 2.4' te düşey ile 10 derecelik açığa sahip giriş örüntüsünde göz bölgelerinin tespiti ve buna bağlı döndürme açısının hesaplanması gösterilmiştir.



Resim 2.4: Yüz örüntülerinden göz bölgelerinin tespiti edilmesi

Resimde görüldüğü üzere, giriş örüntüsü olarak seçilen (a) resmi, düşey ile 10 derecelik açı yapmış bir yüz örüntüsüdür. Göz çifti bölgesi tespit edilen yüz örüntüsünde, Viola-Jones algoritması kullanılarak, bu bölge içerisinde bulunan gözlerin tespit edilmesiyle (b) resmi oluşmaktadır. Burada gözleri ön plana çıkaran sarı renkli kutular, göz bebeğini merkez almaktadır. Dolayısıyla göz bebeklerini tespit etmek için bu kutuların orta noktalarının tespit edilmesi yeterlidir. İki göz arasındaki açı Eşitlik 2.1 ile hesaplanır;

$$Döndürme_Acisi = \frac{180}{\pi} \left(\tan^{-1} \left(\frac{SolGöz_y - SagGöz_y}{SolGöz_x - SagGöz_x} \right) \right) \quad (2.1)$$

Eşitlikte;

$SolGöz_y$, $SagGöz_y$: Sırasıyla Sol ve Sağ gözlerin Y koordinatlarını,

$SolGöz_x$, $SagGöz_x$: Sırasıyla Sol ve Sağ gözlerin X koordinatlarını,

$Döndürme_Acisi$: İlgili örüntü için döndürme açısını temsil etmektedir.

Giriş örüntüsü (a), hesaplanan döndürme açısı kadar döndürülür ve düşey ile minimum açığa sahip (c) örüntüsü elde edilir. Bununla birlikte, örüntüde bulunan bireyin gözleri arasındaki açı sıfıra yaklaştırılmış ve örüntü düzleştirilmiştir. Tüm giriş örüntüleri bu işlemlerden geçirilmektedir. Böylece kamera, açı, kafa pozisyonu gibi bozucu etkiler ortadan kaldırılmaktadır.

2.4 Öznitelik Çıkartma

Önceki aşamalarda yüz ve gözlerin tespit edildiği örüntüden bu aşamada öznitelik çıkartma işlemi yapılmaktadır. Öznitelik; makine öğrenmesi ve örüntü tanımda bir olgu için tespit edilen bireysel ölçülebilir özelliktir (Bishop, 2006). Problemlerin çözümü için birçok öznitelik çıkartma algoritması mevcuttur ancak yüz tanıma alanında yaygınca kullanılan iki yöntem vardır;

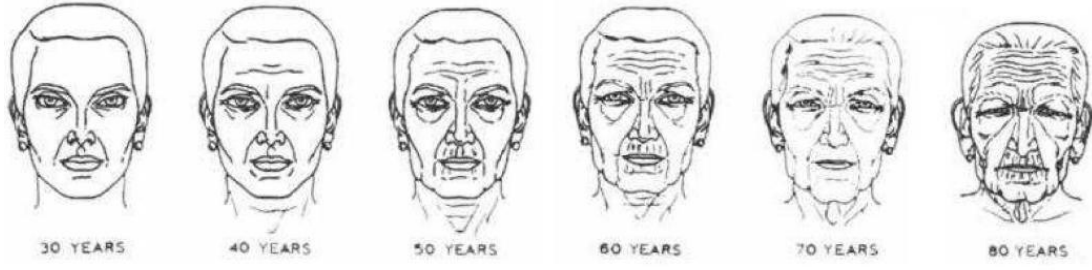
- Geometrik Öznitelikler
- Dokusal Öznitelikler

Bu çalışmada yüz örüntülerinden dokusal öznitelikler çıkartılmış ve sınıflandırma işlemlerinde bu veriler kullanılmıştır. Öznitelikler çıkartılırken giriş örüntüsüne aşağıdaki aşamalar uygulanmıştır:

- Yüz bölgesi tespit edilen örüntü 100x100 piksele yeniden boyutlandırılır.
- Elde edilen örüntü 24x24 piksellik bölgelere ayrılır.
- Bu bölgelere uygulanan öznitelik çıkartma algoritmaları sonucunda elde edilen veriler uç uca eklenerek giriş örüntüsüne ait öznitelik vektörü oluşturulur.
- Bu aşamalar tüm giriş örüntüleri için tekrarlanır ve kullanılacak veri seti elde edilir.

2.5 Dokusal Öznitelikler

Bireyin yüzünde bebeklikten yetişkinlik dönemine kadar geometrik, yetişkinlikten yaşlılık dönemine kadar dokusal değişiklikler meydana gelir. Geometrik değişiklikler kafa boyutlarındaki artışı gösterirken, dokusal değişiklikler kişinin cildinde meydana gelen yaşlanmayı betimler. Kasların esnekliğini kaybetmesi ve yerçekiminin de etkisiyle yüzde sarkmalar görülmektedir. Yüz dokusunda lekeler ve renk değişimleri meydana gelir. Özellikle alın, yanaklar ve göz çevresindeki kırışıklıklar yaşlanmanın en büyük göstergesi sayılmaktadır. Yüzdeki bu kırışıklıkların derinliği, yeri ve oryantasyonu aynı yaş grubundaki kişiler için benzer özellikler göstermektedir (Kılınç, 2012). Şekil 2.2’ de insan yüzüne ait değişim 10’ lu yaş aralıklarıyla gösterilmiştir (Gonzalez-Ulloa & Flores, 1965).



Şekil 2.2: 30-80 yaşları arasında insan yüzündeki değişimin 10 yaş aralıklarla gösterimi.

2.5.1 Yerel İkili örüntü (LBP)

Yerel İkili Örüntü (LBP) dokusal öznitelik çıkartma algoritması, resmin yerel yapısını özetleyen, gri seviyeden bağımsız bir doku ölçü yöntemidir (Ojala, Pietikäinen, & Harwood, 1996). Orijinal LBP algoritması; merkez ve komşu pikseller arasında sıralı ikili kod sistemi karşılaştırmalarına dayalı, ayırsama gücü yüksek bir analiz tekniğidir (Marcel, Rodriguez, & Heusch, 2007). Bu algoritmaya göre her bir piksel, kendisini çevreleyen 3x3’ lük komşu bölgeleriyle arasındaki farkın, $u(x)$ basamak fonksiyona göre ikilileştirilmesiyle etiketlenir (Ojala, Pietikäinen, & Mäenpää, 2002). LBP histogramları Eşitlik 2.2 ile hesaplanır:

$$LBP_{p,R}(x_c, y_c) = \sum_{p=0}^{p-1} s(g_p - g_c) 2^p, \quad s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \text{ ise.} \quad (2.2)$$

Eşitlikte;

(x_c, y_c) : Merkez pikselin LBP algoritmasına göre değerini,

g_p : Merkez pikselin komşularını,

g_c : Merkez pikselin gri değerini,

R : Komşuların merkez piksele olan uzaklığını,

P : İşleme giren komşu sayısını ifade etmektedir (*Computer Vision - ACCV 2010, 2011*).

Komşu etiketlerinin birleştirilmesiyle elde edilen ikili LBP değeri merkez piksel için bir tanımlayıcı olarak kullanılır. Şekil 2.3' te merkez pikselin LBP algoritması tarafından etiketlenmesi gösterilmiştir.



Şekil 2.3: LBP algoritması kullanılarak merkez pikselin etiketlenmesi.

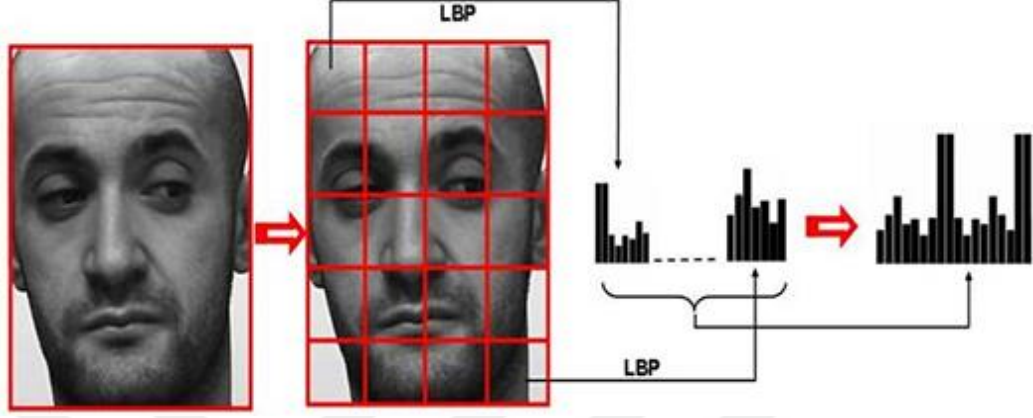
LBP algoritmasından düzgün ve düzgün olmayan olmak üzere iki tür öznitelik elde edilir. Tanımlamada elde edilen LBP histogramlarından sadece düzgün olanlar kullanılır. Düzgün histogramlar; ikili LBP kodunda 0-1 ya da 1-0 geçiş sayısı iki ya da daha az olanları kapsamaktadır. Örneğin; 00000000 sıfır geçişe, 01110000 ve 11001111 iki geçişe sahip olduğundan düzgün, 11001001 dört, 01010010 altı geçişe sahip olduğundan düzgün değildir (Heikkila & Pietikainen, 2006).

Ojala ve arkadaşları yaptıkları çalışmada $LBP_{(8,1)}$ algoritmasına göre ürettikleri düzgün örüntülerin, tüm örüntünün %90' ını oluşturduğunu, $LBP_{(16,2)}$ algoritmasına göre ürettikleri düzgün örüntülerin, tüm örüntünün %70' ini oluşturduğunu gözlemlemişlerdir (Ojala et al., 2002).

LBP ile daha verimli öznitelik çıkartmak için örüntünün bölümlendirilerek kullanılması da mümkündür. Burada giriş örüntüsü I_0, I_1, \dots, I_{n-1} ile ifade edilen n adet bölgeye bölünür ve her bölge için ayrı ayrı LBP histogramları Eşitlik 2.3 ile hesaplanır:

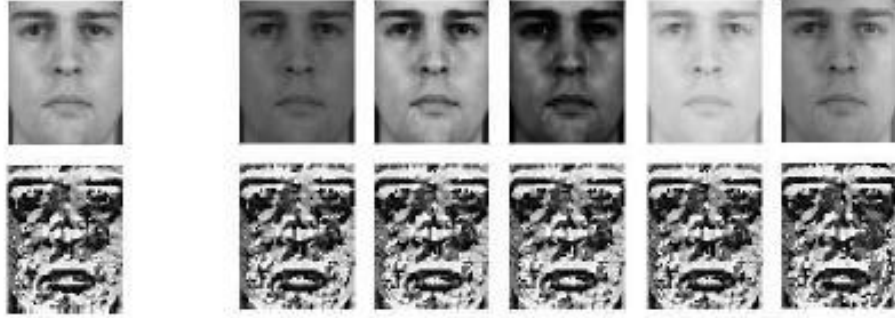
$$H = \sum_{i=0}^n LBP_{8,1}(I_i) \quad (2.3)$$

Eşitlikte H ; I_0, I_1, \dots, I_{n-1} bölgelerine ait $LBP_{8,1}$ değerlerinin toplamından oluşan histogramı temsil etmektedir.



Resim 2.5: 5x4 alanlara bölünmüş görüntüde LBP histogramlarının üretilmesi.

LBP algoritması merkez pikselin komşularıyla olan ilişkisini incelediğinden, aydınlanmaya karşı oldukça dayanıklıdır (Heusch, Rodriguez, & Marcel, 2006). Resim 2.6' da farklı aydınlanma seviyelerine sahip görüntülere LBP algoritması uygulandığında, çıkış görüntülerinin benzerliği görülmektedir.

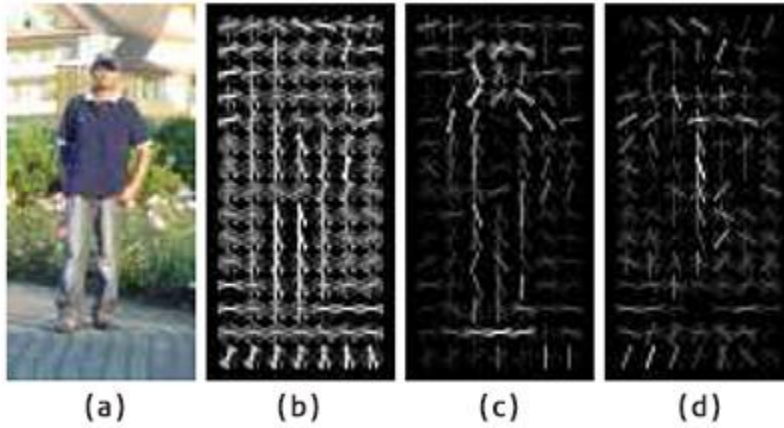


Resim 2.6: Aydınlanmaya dayanıklı LBP görüntüleri.

Daha önce yüz görüntüleri üzerinde yapılmış çalışmalarda LBP algoritmasının sınıflandırma işlemlerinde etkili olduğu görülmüştür (Ahonen, Hadid, & Pietikäinen, 2004; Günay & Nabiye, 2008; Stiefelhagen, Bernardin, Ekenel, & Voit, 2008). Etkin tanımlama performansı sebebiyle algoritmanın bu çalışmada kullanılması uygun görülmüştür.

2.5.2 Yönlü gradyanların histogramı (HOG)

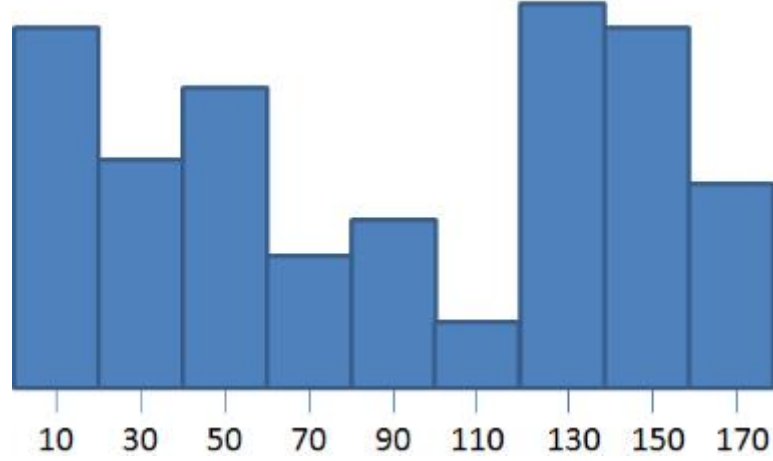
İmgedeki piksellerin yönelim (θ) ve büyüklük değerlerinin karakteristiği olarak da adlandırılan Yönlü Gradyanların Histogramı (HOG) algoritması son yıllarda birçok alanda oldukça ilgi görmektedir (Peker, Altun, & Karakaya, 2009). HOG algoritması ilk olarak insan tespiti yapılan bir çalışmada önerilmiş ve bu tespiti %84 ~%89 doğruluk değerleri ile sağlamıştır (Dalal & Triggs, 2005). Resim 2.7’ de insan resmi içeren giriş örüntüsüne (a) farklı özelliklerle uygulanan yönlü gradyanların oluşturduğu çıktılar gösterilmiştir (b,c,d).



Resim 2.7: Giriş örüntüsüne (a) farklı özelliklerle uygulanan HOG algoritması ve çıktıları.

Birçok araştırmacı tarafından oldukça ilgi gören HOG algoritmasındaki temel amaç imgeyi bir grup yerel histogramlar olarak tanımlamaktır. Bu gruplar, imgenin yerel bir bölgesindeki gradyanların yönelimlerinde, gradyanların büyüklüklerinin toplandığı histogramlardır.

HOG algoritması temelde 64x128 piksellik bir pencere kullanarak, bu bölge içerisinde bulunan nesnelere tanımlanmasını sağlamaktadır. Sobel filtreleri kullanılarak kenar noktaları tespit edilen görüntüde her piksele ait hesaplanan yönelim açısı değeri, $0^\circ - 180^\circ$ arasında 20° açılarla ayrıştırılır ve histograma dönüştürülür. Bu işlem yapılırken elde edilen ara değerler, kendisine en yakın değere yuvarlanır. Örnek olarak 85° gradyan değeri ele alınırsa, bu değer kendisine en yakın diğer iki açı değeri ile arasındaki farklar sırasıyla 90° için 5° , 70° derece için 15° olduğundan, bu gradyan kendisine en yakın olan 90° ile etiketlenir. Örnek histogram grafiği Şekil 2.4’ te verilmiştir.



Şekil 2.4: Yönelim açısı değerlerine ait örnek histogram.

Bu çalışmada giriş örüntüleri 24x24 piksellik alanlara bölünmüştür. Her piksel için bir gradyan değeri üretildiği göz önüne alındığında, hücre başına 576 gradyan değeri Şekil 2.4' te verilen 9'lu histogramda etiketlenmektedir.

Bir HOG tanımlayıcı penceresi (64x128) 3x6 adet 24x24 piksellik hücre içerir. 100x100 piksellik bir örüntünün 2 adet HOG tanımlayıcı penceresi barındıracağı göz önüne alındığında ve tüm bu değerlerin 9'lu histogram ile temsil edileceği düşünüldüğünde, üretilecek toplam öznelik sayısı tüm bu değerlerin çarpımını olan 324' e eşit olacaktır ($2 \times 3 \times 6 \times 9 = 324$).

Bununla birlikte bir imgeye ait HOG değerleri Eşitlik 2.4 - 2.7 ile hesaplanır (Kobayashi, Hidaka, & Kurita, 2007);

$$D_x = [-1 \ 0 \ 1] \quad \text{ve} \quad D_y = \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} \quad (2.4)$$

$$I_x = I * D_x \quad \text{ve} \quad I_y = I * D_y \quad (2.5)$$

$$|G| = \sqrt{I_x^2 + I_y^2} \quad (2.6)$$

$$\theta = \tan^{-1} \frac{I_x}{I_y} \quad (2.7)$$

Eşitlik 2.4' te yatay ve dikey Sobel filtreleri tanımlanmıştır. Eşitlik 2.5' te bu filtreler giriş örüntüsüne uygulanarak örüntünün yatay ve dikey kenarları (I_x ve I_y) tespit edilir. Eşitlik 2.6 ve 2.7' de, Sobel filtresi uygulanmış I_x ve I_y örüntüleri kullanılarak, bu örüntülerin gradyanları (G) ve bu gradyanların yönelim açıları (θ) hesaplanır. Bir giriş örüntüsüne ait Sobel çıktıları Resim 2.8' de gösterilmiştir.



Resim 2.8: Orijinal giriş örüntüsü ve sobel filtrelerine ait çıktılar.

HOG algoritmasının çalışma diyagramı Şekil 2.5' te gösterilmiştir.



Şekil 2.5: HOG Algoritmasının çalışma diyagramı.

2.6 Öznitelik Seçme

Öznitelik seçme yöntemleri; temel ya da bağımsız bileşenlere en az katkı yapan özniteliklerin düşük ayırt edici özelliğe sahip olduğu fikrini esas alarak, daha anlamlı olan özniteliklerin yeni bir sınıflandırma uzayı altında toplanmaları esasına dayanır (Babur, Turhal, & Akbaş, 2012). Bu algoritmalar, yüksek öznitelik vektörlerine sahip veri setlerinde sınıflandırma süresi ve maliyetlerini azaltmak, sınıflandırıcı performansı arttırmak ve çok boyutluluk probleminin (curse of dimensionality) önüne geçmek için geliştirilmiştir (Bellman, 1957). Örneğin; metin sınıflandırma uzay ve zaman karmaşıklığı yüksek olan bir problemdir (Bunte, Hammer, Wismüller, & Biehl, 2010). Bu problemin çözümü için ise kategorize edilecek verinin boyutlarını indirgeme yöntemleri ortaya atılmıştır.

Boyut indirgemedede ilk yaklaşım öznitelik seçimidir. Bu yöntemde başarıyı arttıracak ya da en az düşürecek ve efektif çalışmayı sağlayacak minimum sayıda özneliğin seçilmesi hedeflenir (Biricik, Diri, & Sönmez, 2012). Öznitelik seçmenin avantajları şu şekilde açıklanabilir:

- Yüksek boyutlu verilerde hesaplama süresini ve depolama kapasitelerini azaltmaya yardımcı olur.
- Lineer düzlemde olan noktalar, makine öğrenmesi yöntemlerinde yüksek performans verileri üretir.
- Boyut indirgeme yapılmış verilerin, kaynak verilere göre gösterimi daha kolaydır. Örneğin 2D-3D gibi düşük boyutlu verilerin gösterimi, 10 boyutlu veri uzayının temsiline göre çok daha kolaydır.

Literatürde sıklıkla kullanılan birçok öznitelik seçme/boyut indirgeme algoritması bulunmaktadır. Temel Bileşen Analizi (PCA), Lineer Ayırıştırma Analizi (LDA), Sıralı Öznitelik Seçme (SFS) ve İlişki Temelli Öznitelik Altküme Seçme (CFS) algoritmaları, öznitelik seçme/boyut indirgeme problemlerinde sık kullanılan yöntemlerdendir. Her algoritmanın bir diğerine karşı avantaj ve dezavantajları mevcuttur. Yapılan bu çalışmada, deneysel sonuçlar göz önüne alınarak, İlişki Temelli Öznitelik Altküme Seçme algoritmasının kullanılması uygun görülmüştür.

2.6.1 İlişki temelli öznitelik altküme seçme (CFS)

İlişki Temelli Öznitelik Altküme Seçme (CFS) algoritması, isminden de anlaşılacağı üzere ilişki tabanlı bir filtreleme algoritmasıdır. Özniteliklere ait alt kümeler oluşturup bunların sınıflandırma ile olan ilişkisini inceler. Alt kümeleri oluştururken şu hipoteze göre çalışır: İyi öznitelik alt kümeleri, sınıflandırma ile ilişkisi olmayanlardan ziyade yakından ilişkili olan öznitelikleri barındırır (Hall, 1999). Öznitelik alt kümelerine verilen puanlar Eşitlik 2.8 ile hesaplanır;

$$r_{zc} = \frac{k \bar{r}_{zi}}{\sqrt{k + k(k-1)r_{ii}}} \quad (2.8)$$

Eşitlikte;

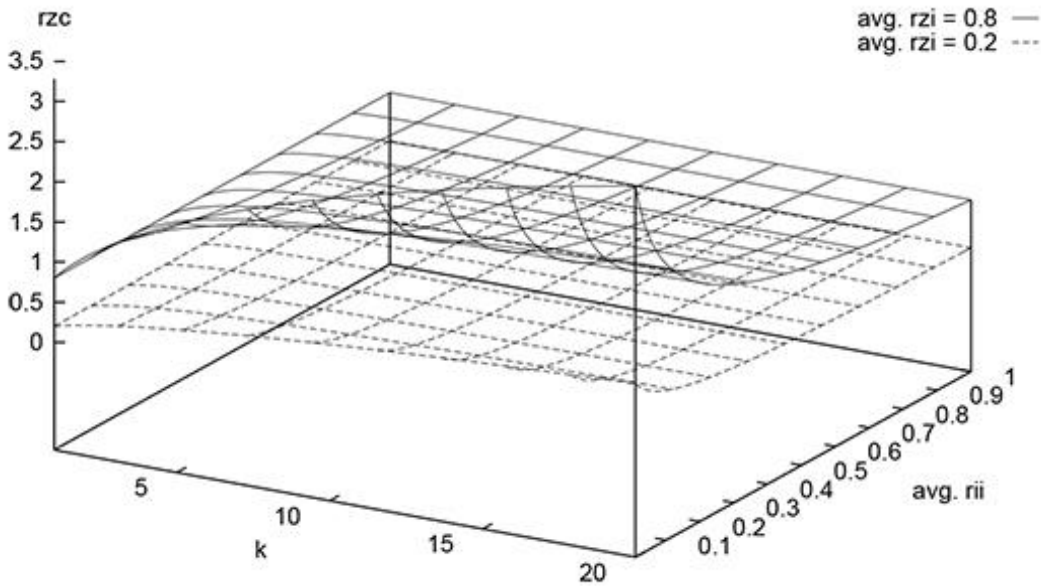
r_{zc} : Bir araya getirilen bileşenler ile diğer değişkenlerin ilişkisini,

k : Bileşen sayısını,

\bar{r}_{zi} : Bir araya getirilen bileşenler ile diğer değişkenlerin ortalama ilişkisini,

\bar{r}_{ii} : Bir araya getirilen bileşenlerin ortalama iç ilişkisini temsil etmektedir.

Şekil 2.6' da CFS algoritması ile alt kümelerin seçimi ve bunlara bağlı değişkenlerin değişimleri gösterilmiştir (Hall, 1999):



Şekil 2.6: CFS algoritması ile alt küme seçimi.

Şekil 2.6' ya göre; alt kümeye eklenecek bileşenlerin sayısındaki artış (k), bileşenler ile diğer verilerin arasındaki ilişkiyi büyük ölçüde arttırmaktadır (r_{zc}). Bununla birlikte, dış değişkenler ile yüksek oranda ilişkili bir grup bileşenin birbirleri arasında düşük ilişkiye sahip olması da muhtemeldir. Ayrıca, ek bir bileşenin eklenme ihtimali göz önünde bulundurulduğunda, hali hazırda seçilmiş düşük iç ilişkiye sahip bileşenler, yüksek ilişkiye sahip dış değişkenlere baskın çıkabilir (Hogarth, 1977).

Bu çalışmada, yaş sınıflarını içeren veri setleri ile cinsiyet sınıflarını içeren veri setlerine uygulanan öznitelik seçme algoritması sonucunda elde edilen özniteliklerin sayıları birbirinden farklıdır. Bu durum, CFS algoritmasının, özniteliklerin sınıflarla olan ilişkilerine göre seçim yapmasından kaynaklanmaktadır.





3. SINIFLANDIRICILAR VE PERFORMANS KRİTERLERİ

Makine öğrenmesi (sınıflandırma algoritmaları) bilgisayar bilimlerinin bir alt alanıdır ve veriler üzerinde tahmin yapabilen algoritmaların yapılarını inceler (Kohavi & Provost, 1998). Arthur Samuel makine öğrenmesini: “Bilgisayarlara açıkça programlanmadan öğrenme yeteneği katan bir çalışma alanıdır.” şeklinde tanımlamıştır (Simon, 2013).

Makine öğrenmesi; bilgisayarlı tahmin tabanlı bir alan olan hesaplamalı istatistikle oldukça ilişkilidir ve hatta çoğu durumda bu alanla örtüşür. İstenmeyen posta filtrelemesi, optik karakter tanıma, arama motorları, bilgisayarlı görme, örüntü tanıma, bilgisayarlı hastalık teşhisi gibi birçok alanda uygulamaları mevcuttur (Wernick, Yang, Brankov, Yourganov, & Strother, 2010). Ayrıca makine öğrenmesi bazı durumlarda veri madenciliği ile bir noktada buluşmaktadır (Mannila, 1996).

Makine öğrenmesi temel olarak iki ana dalda incelenir. Bunlar;

- Danışmanlı (supervised) öğrenme,
- Danışmansız (unsupervised) öğrenmedir.

Makine öğrenmesi terminolojisine göre sınıflandırma, danışmanlı öğrenme tekniğinin bir örneği olarak kabul edilebilir. Danışmanlı öğrenmede girdiler için hedef çıktılara karşılık bir sonuç üretilir. Danışmansız algoritmalarda ise amaç; herhangi bir sınıf verisine ait olmayan girdilerden, benzer doğal özellik ya da mesafeleri içerenleri bir araya getirmektir. Kümeleme, danışmansız öğrenme tekniğine ait bir örnektir (Alpaydin, 2010). Bu çalışmada verilerin sınıf bilgileri mevcut olduğundan, danışmanlı öğrenme teknikleri kullanılmıştır.

3.1 Sınıflandırma Algoritmaları

Bu çalışmada beş adet sınıflandırma algoritması kullanılmış olup, diğer performans kriterleri göz önünde bulundurularak elde edilen sonuçlar, sınıflandırıcılar arası sonuçlar ile karşılaştırılmış ve sınıflandırıcı performansları analiz edilmiştir. Çalışmada kullanılan sınıflandırıcı algoritmaları aşağıda verilmiştir:

- Bayes Ağları (BayesNet)
- Naïve Bayes
- LibSVM (Polinom Çekirdek)
- Sıralı Minimal Optimizasyon (SMO)
- k-En Yakın Komşu (k-NN)

3.1.1 Bayes ağı (BayesNet)

Bayes ağı, çizgesel olasılık modelleri ailesinin bir parçasıdır. Bu çizgesel yapılar, belirsiz bir etki alanı hakkındaki bilgileri temsil etmek için kullanılırlar. Çizgelerdeki her düğüm rastgele bir değişkeni, düğümler arasındaki kenarlar ise rastgele değişkenler arasındaki olasılık bağımlılıklarını temsil eder. Bu koşullu bağımlılıklar genellikle bilinen istatistiksel ve hesaplama yöntemleriyle belirlenir. Bu nedenle Bayes ağı; çizge kuramının, olasılık teorisinin, bilgisayar bilimlerinin ve istatistiğin temel ilkelerini bir arada kullanmaktadır (Ben-Gal, 2007). Bayes ağı Eşitlik 3.1 ile hesaplanır;

$$P(U) = \prod_{u \in U} p(u|pa(u)) \quad (3.1)$$

Eşitlikte;

$U = \{x_1, x_2, \dots, x_n\}$, $n \geq 1$ değişkenlerini içeren kümedir.

$pa(u)$, U kümesinin bir kısmını içeren alt kümedir.

3.1.2 Naïve bayes

Naïve Bayes sınıflandırıcısı, Bayes ağının basitleştirilmiş bir formudur. Tüm öznitelikler verilen sınıf verilerinden bağımsızdır. Bu yöntemle koşullu bağımsızlık adı verilir (Zhang, 2005). Diğer danışmanlı makine öğrenmesi yöntemleriyle kıyaslandığında Naïve Bayes sınıflandırıcısı, sınıflı eğitim verileri için en basit ve güçlü tahmin tekniğine sahip olan sınıflandırıcılardan biridir. Güçlü tahmin yeteneğinin yanında eğitim setinde öznitelik ve sınıflar arasındaki ilişki hakkında iyi bilgiler verebilir (Možina, Demšar, Kattan, & Zupan, 2004). Naïve Bayes fonksiyonu Eşitlik 3.2 ile hesaplanır;

$$f_{nb}(E) = \frac{p(C = +)}{p(C = -)} \prod_{i=1}^n \frac{p(x_i | C = +)}{p(x_i | C = -)} \quad (3.2)$$

Eşitlikte;

E : X öznitelik kümesinin i indisli öznitelik vektöründeki (x_1, x_2, \dots, x_n) değerlerin bir kısmından oluşan veri kümesini,

C : Sınıflandırıcı değerleri. (+ pozitif sınıfları, - negatif sınıfları ifade eder).

$f_{nb}(E)$: E giriş değerleri için Naïve Bayes fonksiyonunu temsil etmektedir.

3.1.3 LibSVM (polinom çekirdek)

LibSVM, Destek Vektör Makinesi (SVM) için geliştirilmiş ve en yaygın kullanılan SVM kütüphanelerinden biridir. İki sınıflı verilerin tahmininde güçlü bir makine öğrenme tekniği olan geleneksel SVM' nin aksine çok sınıflı verilerin kullanılmasına olanak sağlamaktadır. LibSVM çalışması iki aşamadan oluşmaktadır: İlk aşamada model oluşturmak için veri seti eğitilir, ikinci aşamada oluşturulan model test veri setine ait bilgilerin tahmini için kullanılır (Chang & Lin, 2011). LibSVM; sınıflandırma, regresyon ve dağılım tahmini için SVM fonksiyonlarını kullanır. Bu çalışmada LibSVM fonksiyonu, SVM sınıflandırıcısının Polinom çekirdeği ve C - Destek Vektör Sınıflandırma (C -SVC) yöntemi kullanılarak oluşturulmuştur. Algoritma Eşitlik 3.3 ile hesaplanır;

$$\text{sgn}(w^T \phi(x) + b) = \text{sgn}\left(\sum_{i=1}^l y_i \alpha_i K(x_i, x) + b\right) \quad (3.3)$$

Eşitlikte;

sgn : İşaret fonksiyonunu,

w^T : SVM fonksiyonuna ait ağırlık vektörünü,

$\phi(x)$: Ana vektör uzayından ϕ fonksiyonu ile elde edilen giriş vektörünü,

b : Bias değerini,

y_i : İşlem gören vektörün sınıf değerini,

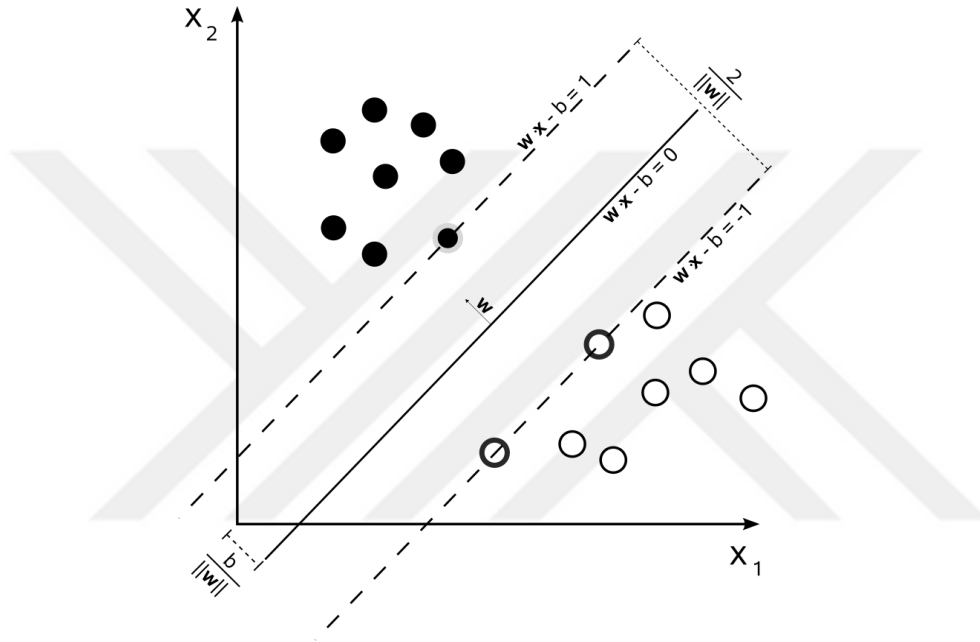
α_i : İşlem gören vektöre ait ağırlık vektörünü,

K : Çekirdek fonksiyonunu,

x_i : Destek vektörlerini,

x : Tahmin edilen sınıf verisini temsil etmektedir.

Şekil 3.1' de iki sınıfı birbirinden $w x - b = 0$ hiper düzlemi ile ayıran SVM algoritması gösterilmiştir. Hiper düzleme eşit mesafelerde bulunan $w x - b = 1$ ve $w x - b = -1$ düzlemleri üzerinde kalan sınıf verileri destek vektörleri olarak tanımlanır.



Şekil 3.1: SVM algoritması ile destek vektörlerinin tespiti.

3.1.4 Sıralı minimal optimizasyon (SMO - polinom çekirdek)

Sıralı Minimal Optimizasyon (SMO), SVM eğitimleri süresince meydana gelen kuadratik programlama problemlerini çözmek için bir Microsoft çalışanı olan John Platt tarafından 1998 yılında ortaya çıkarılmış bir algoritmadır (Platt, 1998). LibSVM algoritması tarafından da kullanılmakta olan algoritma ile yapılmış çalışmalar, SVM topluluğu tarafından oldukça ilgi görmüş ve önceleri kullanılan yöntemlerden daha az karmaşık ve maliyetsiz olduğu tespit edilmiştir (Rifkin, 2002; Zanni, Serafini, & Zanghirati, 2006). SMO algoritması Eşitlik 3.4 ile hesaplanır;

$$\max_{\alpha} \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n y_i y_j K(x_i, x_j) \alpha_i \alpha_j \quad (3.4)$$

Eşitlikte;

α_i : Lagrange çarpanını, $(0 \leq \alpha_i \leq C)$, $i = 1, 2, \dots, n$

C : SVM hiper parametresini,

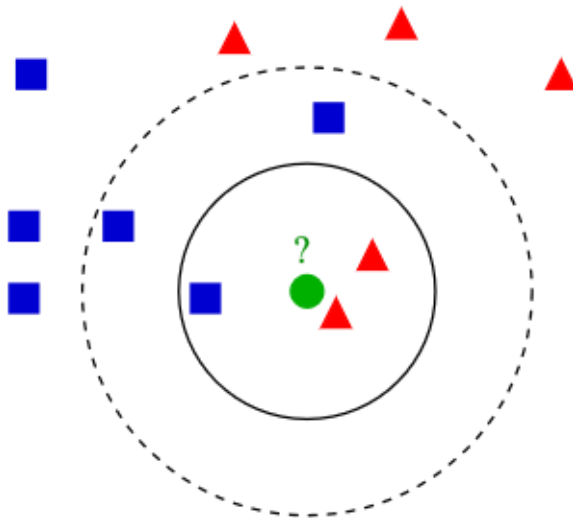
$y_i y_j$: Sınıf verisini,

$K(x_i, x_j)$: Çekirdek fonksiyonunu temsil etmektedir.

3.1.5 K-en yakın komşu (k-NN)

Örüntü tanımada k-En Yakın Komşu (k-NN) algoritması, sınıflandırmada ve regresyonda parametrik olmayan yöntemler kullanır. Her iki olguda da giriş, öznitelik uzayındaki k parametresine en yakın eğitim örneklerinden oluşmaktadır (Altman, 1992). Bu sınıflandırıcıda, komşuları tarafından oy çokluğuyla sınıflanan bir obje, k adet en yakın komşusundan, çevresinde en yaygın olanına atanır. K parametresinin küçük, tek ve pozitif bir sayıdan oluşması genel problemler için idealdir ancak bu durum değişkenlik gösterebilir. Eğer $k=1$ seçilirse, obje kendisine en yakın olan ilk sınıfa atanır.

Şekil 3.2' de, k-NN algoritmasının çalışma şekli gösterilmiştir. Bu örnekte $k=1$ için test örneği kendisine en yakın olan kırmızı sınıfa atanacaktır. $k=3$ için bu örnek, kendisine en yakın üç örnekten çoğunluğa ait olan kırmızı sınıfa atanacaktır (Mavi: 1, Kırmızı: 2). Ancak $k=5$ değeri için durum daha farklıdır. Test örneğine en yakın beş örnekten çoğunluk bu kez mavi gruba geçeceğinden, örnek bu gruba atanacaktır (Mavi:3, Kırmızı:2). Bu örnek, k değerindeki değişimin sınıflandırma sonucundaki etkisine dair en iyi örneklerden biridir.



Şekil 3.2: k-NN algoritması ile örnek sınıflandırma.

k-NN algoritması birçok farklı mesafe fonksiyonu kullanmaktadır. Bunlardan en yaygın olanları aşağıda verilmiştir;

- Chebyshev Uzaklığı
- Euclidean (Öklid) Uzaklığı
- Manhattan Uzaklığı
- Minkowski Uzaklığı.

Bu çalışmada k-NN algoritması Öklid (Euclidean) Uzaklığı mesafe fonksiyonu ile kullanılmıştır. Algoritma Eşitlik 3.5 ile hesaplanır;

$$d(p, q) = d(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2} \quad (3.5)$$

Eşitlikte;

q ve p : Öznitelik vektöründeki iki noktayı,

$d(p, q) = d(q, p)$: Bu noktalar arası hesaplanan Öklid mesafesini gösterir.

3.2 Performans Kriterleri

Sınıflandırma işlemleri sonucunda elde edilen verilerin değerlendirilmesi için bir takım yardımcı yöntemlere ihtiyaç vardır. Bu yöntemler, sınıflandırıcı performansını analiz etmede önemli parametreler içerirler. Ancak bazı durumlarda performans kriterlerinin tek başına kullanılması da yeterli olmamaktadır. Özellikle farklı sınıflara ait verilerin sayısındaki büyük fark, yanlış sınıflandırmaların gözden kaçmasına sebep olabilmektedir. Bir örnekle açıklamak gerekirse; cinsiyet sınıflandırılması yapılacak bir sistemde, erkek bireylerin sayısı ile kadın bireylerin sayısı arasında oldukça fazla fark olduğunu ve erkek bireylerin sayıca üstün olduğunu düşünelim. Sınıflandırma işlemi sonucunda erkek bireylerin tamamını doğru tespit edip, kadın bireyleri tespit edemeyen bir sistem, yapmış olduğu hataya rağmen, erkek bireylerin fazla sayısı sebebiyle yüksek doğruluk değeri üretecektir. Ancak bu sonuca bakılarak uygulanan yöntemin geçerli olduğunun söylenmesi mümkün değildir. Böyle bir problemin önüne geçerek elde edilen sonuçların geçerliliğinin artırılması için performans kriterlerinin bir arada kullanılması gerekmektedir.

Bu çalışmada sınıflandırma işlemlerinden elde edilen Doğruluk (Accuracy), Kappa ve Alıcı İşletim Karakteristiği (ROC) yöntemleri bir arada kullanılmış ve sonuçlar bu verilere göre analiz edilmiştir.

3.2.1 Çapraz doğrulama (cross validation)

Çapraz doğrulama (CV), elde edilen istatistiksel analiz sonuçlarını değerlendirmek için kullanılan bir model doğrulama tekniğidir. Veri setini eşit veriler içeren n adet kümeye bölerek, bu kümelerden $n-1$ adedini eğitim, kalan kümeyi de test için kullanma esasına göre çalışır. Böylece, eğitim yapılan veriler ile test yapılması probleminde kaçınılır. Ayrıca bu işlem n kez tekrarlanarak, tüm verilerin sınıflandırılması sağlanır. Elde edilen n adet performans değerlerinin ortalamaları, ortalama sonuç değerlerini gösterir (Kohavi, 1995). Genel olarak $n=2,5,10,20$ değerleri yaygın olarak kullanılmaktadır.

Bu çalışmada $n=10$ ve $n=20$ değerlerine göre sınıflandırma işlemleri yapılmıştır.

3.2.2 Sınıflandırma süresi

Sınıflandırma Süresi (hesaplama süresi), sınıflama işleminin başlangıcından bitişine kadar geçen süreyi ifade etmektedir. Bu değer tüm platformlarda ve cihazlarda farklı sonuçlar verebilir ancak yapılan çalışmada kullanılan tüm algoritmalarda aynı platform ve cihaz kullanıldığından, bu değer ile sonuçlar arası karşılaştırma yapılmasında herhangi bir engel yoktur.

Özellikle gerçek zamanlı çalışan sistemlere uygulanacak algoritmaların belirlenmesinde bu parametre oldukça önem arz etmektedir.

3.2.3 Doğruluk (accuracy)

Doğruluk, sınıflandırma işlemi sonucunda tahmin edilen sonuçların gerçek sonuçlara olan yakınlığını belirtir. Yüzde 100 üzerinden ifade edilen bir doğruluk skalasında 100'e en yakın olan değerler en iyi sınıflandırma sonuçlarını verir (E. R. Cohen, 1998). Doğruluk değeri Eşitlik 3.6 ile hesaplanır:

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN} \quad (3.6)$$

Eşitlikte;

Accuracy : Elde edilen doğruluk değerini,

- TP* : Doğru tahmin edilen gerçek pozitifleri,
TN : Doğru tahmin edilen gerçek negatifleri,
FP : Yanlış tahmin edilen pozitifleri (gerçek olmayan pozitifler),
FN : Yanlış tahmin edilen negatifleri (gerçek olmayan negatifler) gösterir.

3.2.4 Kappa

Kappa katsayısı, iki değerleyici arasındaki karşılaştırmalı uyuşmanın güvenilirliğini ölçen bir istatistik yöntemidir (J. Cohen, 1960). Kappa ölçüsü ilk olarak sınıflandırma sonuçlarını olasılıksal yönden değerlendirmek amacıyla ortaya atılmış bir yöntemdir. Rastgele elde edilmiş doğrulukları göz önüne almayan Cohen, bu verileri değerlendirmek için yeni bir tekrar-normalizasyon olasılık fonksiyonu belirlemiştir (Powers, 2012). Kappa istatistik değeri Eşitlik 3.7 ile hesaplanır:

$$\kappa = \frac{\Pr(a) - \Pr(e)}{1 - \Pr(e)} \quad (3.7)$$

Eşitlikte;

$\Pr(a)$: İki değerleyici için gözlemlenen uyuşmaların toplamını,

$\Pr(e)$: $\Pr(a)$ 'nın şans eseri ortaya çıkma olasılığını belirtir. Burada $0 < \kappa < 1$ değeri elde edilir. Elde edilen değer 1'e yaklaşma olasılığı, sınıflandırma sonuçlarının birbirleriyle olan uyumunu gösterir.

Landis ve Koch tarafından yapılan bir önermede elde edilen κ Kappa değerlerini yorumlamak için Çizelge 3.1'de verilen değerler kullanılmıştır (Landis & Koch, 1977):

Çizelge 3.1: Kappa istatistik değeri yorum tablosu.

κ	
< 0	Hiç uyuşma olmaması
0.00 – 0.20	Önemsiz uyuşma olması
0.21 – 0.40	Orta derecede uyuşma olması
0.41 – 0.60	Sık derecede uyuşma olması
0.61 – 0.80	Önemli derecede uyuşma olması
0.81 – 1.00	Neredeyse mükemmel uyuşma olması

Ancak bu tabloda verilen yorumlar ve hatta verilen aralıklar hakkında istatistikçiler arasında anlaşmazlık vardır. Bazı istatistikçilere göre bu aralıklar ve yorumlar, Kappa değerinin değişken kategori sayısından etkilendiği gerçeğini unutturabilmesi ihtimaliyle, araştırmacılar için bir yönden zararlı olabilmektedir (Gwet, 2001; Sim & Wright, 2005).

3.2.5 Alıcı İşletim Karakteristiği (ROC)

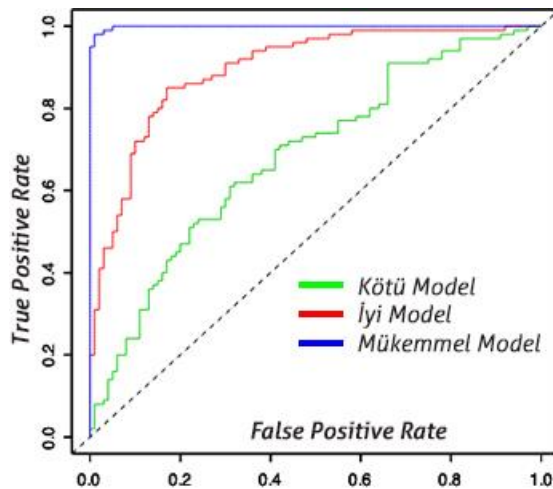
Alıcı İşletim Karakteristiği (Receiver Operating Characteristic - ROC) eğrisi, ikili sınıflandırma sistemlerinde ayırım eşik değerinin farklılık gösterdiği durumlarda, hassasiyetin kesinliğe olan oranıyla ortaya çıkmaktadır. Daha basit anlamda doğru pozitiflerin, yanlış pozitiflere olan kesri olarak da ifade edilebilir. İki işletim karakteristiği arasında bir karşılaştırma yaptığımızdan, bağlı işletim karakteristiği olarak da bilinir (Swets, 2014). ROC eğrisini oluşturan doğru pozitiflerin oranı ve yanlış pozitiflerin oranı Eşitlik 3.8 – 3.9 ile hesaplanır:

$$TPR = \frac{TP}{\text{Tüm Pozitifler}} \quad (3.8)$$

$$FPR = \frac{FP}{\text{Tüm Negatifler}} \quad (3.9)$$

Eşitlikte TPR ; Gerçek pozitiflerin oranını, FPR ; Yanlış pozitiflerin oranını belirtmektedir.

Şekil 3.3' te, örnek sınıflandırma sonuçlarına ait ROC grafiklerinde 1'e en yakın değer, en iyi karakteristiği göstermektedir.



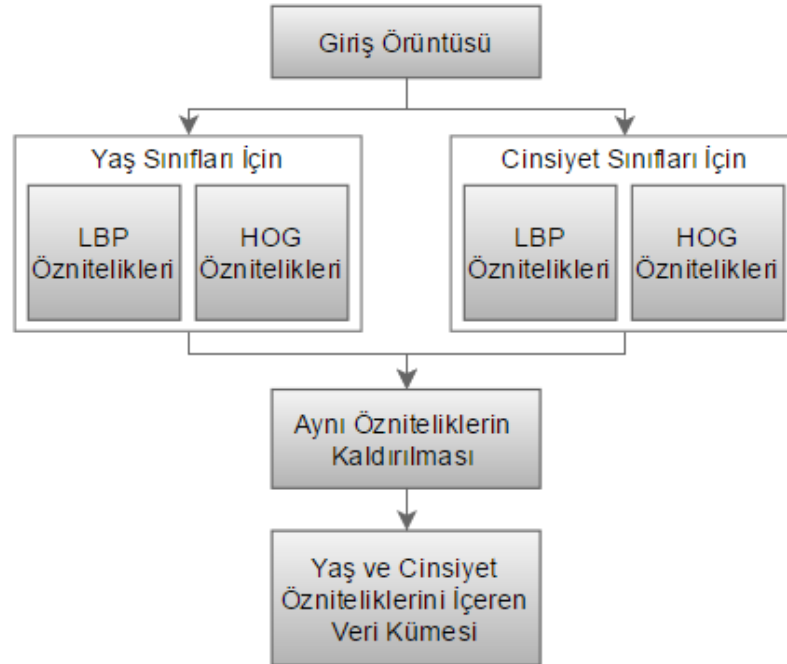
Şekil 3.3: ROC karakteristiği.



4. SINIFLANDIRMA İŞLEMİ VE SONUÇ ANALİZİ

Yüz örüntülerinden yaş ve cinsiyet tahmini konusu temelde birbirinden farklı konulardır. Cinsiyet tahmini iki sınıflı bir model iken, yaş tahmini duruma göre çeşitlilik gösterebilir. Bu yöntemlerin ayrı ayrı değerlendirilmesi, her bir sınıflandırma işlemi için yüksek performans ve doğruluk değerlerine ulaşmayı sağlayabilir. Ancak tasarlanacak sistem içerisinde iki farklı sınıflandırma işleminin oluşu, maliyet ve hesaplama süresi performanslarını da olumsuz etkileyecektir. Bunun yerine iki problemin tek bir problem olarak düşünülüp hareket edilmesinin daha uygun olacaktır.

Veri setinde problemin çözümüne yönelik etkin rol oynayan özniteliklerin belirlenmesi işlemi bir önceki başlıkta anlatılmıştır. Kullanılan öznitelik seçme algoritması, özniteliklerin sınıf verileri ile olan ilişkilerini incelediğinden, bunların belirlenmesi sırasında hem yaş verilerinin hem de cinsiyet verilerinin göz önünde bulundurulması gerekmektedir. Bu çalışmada önerilen yöntemle ait veri seti Şekil 4.1’ de bulunan diyagrama göre oluşturulmuştur.



Şekil 4.1: Yaş ve cinsiyet özniteliklerini bir arada içeren veri setinin hazırlanmasında önerilen yöntem.

Şekilde görüldüğü üzere, giriş örüntüsünden yaş ve cinsiyet sınıflandırmaları için ayrı ayrı çıkartılan öznitelikler benzerlik filtresinden geçirildikten sonra bir araya getirilip, sınıflandırıcı modelleri için eğitim kümesi oluşturulmuştur. Böylece veri kümesinin her iki problem için gerekli öznitelikleri barındırması sağlanmıştır. Bu yöntem kullanılarak, minimum maliyet ile maksimum performans kazanımı hedeflenmiştir. Sistemin tek dezavantajı; bir probleme özgü seçilen özniteliklerin diğer bir problem için yeterli ölçüde önemli olmaması ve buna bağlı olarak sınıflandırma işlemi sonucunda elde edilen doğruluk değerlerinde kayıpların yaşanmasıdır. Ancak her sınıflandırma işlemi için ayrı model oluşturma, hesaplama ve buna bağlı maliyet artışı problemlerinin azaltıldığı göz önüne alındığında, doğruluk değerlerinde meydana gelen azalmaların kabul edilebilir bir seviyede olduğu görülmektedir.

Aşağıdaki çizelgelerde (Çizelge 4.1 - 4.4), color FERET veri setinde bulunan 800 yüz örüntüsünden, farklı yöntemler ile elde edilen özniteliklerin BayesNet, Naïve Bayes, LibSVM, SMO ve k-NN algoritmaları ile yaş ve cinsiyet için ayrı ayrı sınıflandırılmasına ait doğruluk, Kappa, ROC ve hesaplama süresi değerleri verilmiştir. Diğer çizelgelerde ise (Çizelge 4.5 - 4.8) yaş ve cinsiyet tahmini tek bir problem olarak düşünülmüştür. Bu problemler için yöntemlerin ürettiği öznitelikler bir araya getirilmiş, öznitelik seçme algoritmasından geçirilmiş ve sınıflandırma işlemleri yapılmıştır. Ayrıca bu sonuçlar, önerilen yönteme ait veri setinden elde edilen sonuçlar ile karşılaştırılmıştır. Son olarak önerilen yönteme ait doğruluk sonuçları, literatürde kullanılan diğer yöntemlerden elde edilen sonuçlar ile karşılaştırılmıştır (Çizelge 4.9 - 4.10).

Ayrıca elde edilen doğruluk sonuçları, ROC ve Kappa değerleri ile birlikte yorumlanarak güçlendirilmiştir. Bu üç parametre bir arada değerlendirildiğinde, istenilen probleme özgü en iyi sınıflandırıcı modelinin tespiti kolaylaşmaktadır.

Verilen çizelgelerde ilk sütun kullanılan çapraz doğrulama metodunu, ikinci sütun kullanılan öznitelik çıkartma yöntemini ve veri kümesinin içerdiği öznitelik sayısını, diğer sütunlar ise sınıflandırma işlemine ait sonuçları içermektedir.

Çizelge 4.1' de, LBP ve HOG yöntemleri kullanılarak öznitelikleri çıkartılan örüntülerden elde edilen yaş sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.1: Tüm öznitelikler ile yaş sınıflandırma performans sonuçları.

n Küme Çapraz Doğrulama	Yöntem (Öznitelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)
10	LBP (944)	59,5000	0,3441	0,801	5,98	61,8750	0,3739	0,823	<u>4,01</u>	85,0000	0,7219	0,860	179,59	75,1250	0,5309	0,826	14,35	<u>92,5000</u>	<u>0,8641</u>	<u>0,935</u>	8,30
	HOG (324)	53,6250	0,2460	0,702	2,86	51,8750	0,2652	0,735	<u>2,31</u>	62,6250	0	0,500	40,00	75,0000	0,5194	0,798	9,55	<u>84,0000</u>	<u>0,7083</u>	<u>0,856</u>	2,68
20	LBP (944)	57,5000	0,3159	0,796	7,75	62,5000	0,3842	0,824	<u>5,06</u>	85,6250	0,7318	0,862	342,11	76,5000	0,5523	0,833	28,05	<u>94,2500</u>	<u>0,8961</u>	<u>0,952</u>	8,73
	HOG (324)	53,8750	0,2569	0,720	3,25	52,7500	0,2820	0,737	<u>2,10</u>	62,6250	0	0,500	76,00	74,5000	0,5073	0,788	18,95	<u>85,0000</u>	<u>0,7276</u>	<u>0,866</u>	2,53

Çizelge 4.1 incelendiğinde, bir ikili sınıflandırma yöntemi olan SMO' nun, yaş problemi için düşük sonuçlar verdiği görülmektedir. Yaş tahmini çok sınıflı bir problem olduğundan, böylesine düşük değerlerin elde edilmesi normaldir. Bununla birlikte Naïve Bayes ve BayesNet gibi olasılıksal çözümler yapan sınıflandırıcıların da bu problemi tanımlamak için yeterince etkili olmadıkları çizelgede görülmektedir. Elde edilen sonuçların incelendiğinde, LibSVM ve k-NN sınıflandırıcılarının diğerlerinden üstün olduğu görülmektedir. Ancak bu sınıflandırıcıların ürettiği doğruluk, Kappa, ROC değerleri ve hesaplama süreleri arasında da oldukça farklar mevcuttur.

LibSVM model tabanlı bir sınıflandırıcı olduğundan, sınıflandırma işlemi için herhangi bir modele ihtiyaç duymayan k-NN' e göre oldukça yavaş kalmaktadır. Öyle ki, k-NN yönteminde yaş sınıflandırma işlemi 8,73 saniyede %94,25 doğrulukla tamamlanmaktayken, LibSVM' de bu değerler 342,11 saniyede %85,625 doğrulukta kalmaktadır. Bu problem için k-NN algoritması %8,625 daha yüksek doğruluk değeri ve %97,45 daha hızlı hesaplama süresi sunmaktadır.

k-NN algoritmasıyla elde edilen ROC ve Kappa değerleri incelendiğinde, sınıflandırma sonucunun şans ile olan ilişkisinin oldukça az olduğu görülmektedir. Bu, k-NN algoritmasını, problemin çözümünde diğer algoritmalarından ayıran en etkili özelliklerdendir.

Tüm hesaplama süreleri karşılaştırıldığında, en kısa sürelerin Naïve Bayes algoritmasına ait olduğu görülmektedir. Ancak üretilen düşük doğruluk değerleri, bu sınıflandırıcının kullanılabilirliğini azaltmaktadır.

Ayrıca elde edilen sonuçlar yöntem tabanlı karşılaştırıldığında, LBP yönteminin HOG yöntemine oranla üstün olduğu görülmüştür. Özellikle güçlü tanımlama kabiliyeti, LBP yöntemini diğerlerinden ayıran en önemli özelliktir.

Çizelge 4.2' de, veri setlerinin CFS algoritmasından geçirilmelerinin ardından elde edilen yaş sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.2: Öznitelik seçimi sonrası elde edilen veri setleri ile yaş sınıflandırma performans sonuçları.

n Küme Çapraz Doğrulama	Yöntem (Öznitelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)
10	LBP (134)	64,8750	0,4130	0,836	1,76	59,2500	0,3518	0,833	<u>1,50</u>	75,8750	0,5615	0,787	25,12	71,3750	0,4418	0,785	3,86	<u>88,1250</u>	<u>0,7822</u>	<u>0,890</u>	1,65
	HOG (54)	59,1250	0,3022	0,754	1,63	57,8750	0,3218	0,771	<u>0,55</u>	62,6250	0	0,500	8,15	66,7500	0,2814	0,680	3,38	<u>82,6250</u>	<u>0,6831</u>	<u>0,846</u>	0,70
20	LBP (134)	64,1250	0,4024	0,837	1,83	58,7500	0,3456	0,833	<u>1,51</u>	75,8750	0,5538	0,777	46,92	72,2500	0,4554	0,788	5,75	<u>89,0000</u>	<u>0,7996</u>	<u>0,901</u>	1,56
	HOG (54)	61,8750	0,3538	0,763	1,31	59,1250	0,3404	0,771	<u>0,81</u>	62,6250	0	0,500	13,10	68,0000	0,3081	0,685	4,10	<u>83,1250</u>	<u>0,6925</u>	<u>0,850</u>	0,86

Öznitelik seçme işlemleri sonucunda 944 özniteliğe sahip LBP veri seti 134' e, 324 özniteliğe sahip HOG veri seti ise 54' e düşürülmüştür.

Çizelge 4.2 incelendiğinde, öznitelik azaltma işlemlerinin, özellikle sınıflandırma sürelerine oldukça etki ettiği görülmektedir. Ancak bu işlem sonucunda doğruluk değerlerinde yüksek kayıplar meydana gelmiştir. Bir önceki aşamada olduğu gibi yine bu aşamada en düşük sınıflandırma süreleri Naïve Bayes algoritması tarafından, en yüksek doğruluk, Kappa ve ROC değerleri ise k-NN algoritması tarafından üretilmiştir.

Çizelge 4.2' de bulunan sonuçlar incelendiğinde, öznitelik azaltma sonrasında yaş sınıflandırma işlemi için en iyi sonuçlar LBP yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak k-NN algoritması tarafından 1,56 saniyede %89,00 doğrulukla tamamlanmıştır. Bu yöntemle elde edilen sonuçlarda ham veri setine oranla doğrulukta %5,25 azalma, hesaplama süresinde ise %73,86 gibi bir kazanç meydana gelmiştir. Diğer yöntemden daha hızlı çalışması bir avantaj olarak gözüktüğü de, düşük doğruluk değerleri, bu yöntemin kullanılabilirliğini azaltmaktadır.

Çizelge 4.1' de olduğu gibi burada da k-NN algoritmasının başarısı dikkat çekmektedir. Çok sınıfları çözmede, sınıf verisine minimum bağımlılıkla hesaplama yapan bu yöntemin, kullanılan diğer yöntemlere göre daha doğru tahminler yaptığı görülmektedir. Elde edilen doğruluk sonuçları karşılaştırıldığında k-NN, kendisine en yakın yöntemden %16,75 daha yüksek doğruluk değeri üretmiştir.

Çizelge 4.1' de olduğu gibi LBP yöntemi, yaş tespitinde HOG yöntemine oranla yine daha başarılı olmuştur.

Çizelge 4.3' te, LBP ve HOG yöntemleri kullanılarak öznitelikleri çıkartılan örüntülerden elde edilen cinsiyet sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.3: Tüm öznelikler ile cinsiyet sınıflandırma performans sonuçları.

n Küme Çapraz Doğrulama	Yöntem (Öznelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)
10	LBP (944)	87,2500	0,7447	0,936	5,35	86,3750	0,7271	0,919	2,96	96,8750	0,9374	0,969	73,42	96,0000	0,9199	0,961	6,76	96,0000	0,9199	0,961	8,35
	HOG (324)	83,8750	0,6771	0,922	2,61	81,2500	0,6255	0,891	1,73	52,8750	0	0,500	47,57	91,8750	0,8368	0,918	4,50	91,1250	0,8218	0,911	2,48
20	LBP (944)	87,6250	0,7524	0,939	6,95	85,7500	0,7146	0,921	3,76	97,2500	0,9449	0,973	139,06	96,7500	0,9349	0,968	12,63	97,3750	0,9474	0,975	8,63
	HOG (324)	84,5000	0,6899	0,925	2,96	81,1250	0,6229	0,892	1,58	52,8750	0	0,500	91,10	91,5000	0,8293	0,914	7,65	92,5000	0,8495	0,925	2,53

Çizelge 4.3 incelendiğinde, tüm özniteliklerin kullanıldığı bu veri setlerinde, iki sınıflı bir problem olan cinsiyet tahmininden elde edilen sonuçların, çok sınıflı bir problem olan yaş tahmininden elde edilen sonuçlara oranla çok daha yüksek olduğu görülmektedir. Ayrıca LibSVM algoritması yaş tahmini problemine göre daha efektif çalışmış ve yüksek sonuçlar vermiştir. Yaş sınıflandırma probleminin aksine, bu problemde en yüksek doğruluk değerleri, kullanılan öznitelik çıkartma yöntemlerine bağlı olarak farklı algoritmalarda üretilmiştir. 10 kümeli çapraz doğrulama yöntemi kullanıldığında en yüksek doğruluk değerleri LBP öznitelikleri için LibSVM algoritmasıyla, HOG öznitelikleri içinse SMO algoritmasıyla üretilmiştir. Bununla birlikte 20 kümeli çapraz doğrulama yönteminde en yüksek sonuçlar, her iki öznitelik çıkartma algoritması için de k-NN algoritmasıyla üretilmiştir.

Sınıflandırma sürelerinde, daha önceki sonuçlarda olduğu gibi Naïve Bayes algoritması ön plana çıkmıştır ancak bu algoritma rakiplerine oranla yine daha düşük değerler üretmiştir.

Farklı algoritmalarla elde edilen yüksek doğruluk değerleri; hesaplama süreleri, ROC ve Kappa değerleri ile birlikte değerlendirildiğinde, tek bir algoritmayı ön plana çıkarmaktadır. Öyle ki; 10 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak LibSVM algoritması tarafından 73,42 saniyede %96,8750 doğrulukla tamamlanan cinsiyet sınıflandırma işlemi, k-NN algoritması tarafından 8,35 saniyede %96,00 doğrulukla tamamlanmıştır. Bu sonuçlara göre k-NN algoritması LibSVM algoritmasından %0,875 daha düşük ancak %88,62 daha hızlıdır. Aynı şekilde cinsiyet sınıflandırma işlemi; 20 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak k-NN algoritması tarafından 8,63 saniyede %97,375 doğrulukla tamamlanmıştır. Bu yöntemde elde edilen en yüksek ROC ve Kappa değerleri yine k-NN algoritmasına aittir.

Yine burada LBP yönteminin HOG yöntemine karşı üstünlüğü dikkat çekmekte ancak aradaki farkın yaş tespitinde elde edilen sonuçlar kadar olmadığı da görülmektedir. Bu sonuca bakılarak, HOG yöntemi ile üretilen özniteliklerin cinsiyet ayrıştırmada etkili olduğu söylenebilir.

Çizelge 4.4' te, veri setlerinin CFS algoritmasından geçirilmelerinin ardından elde edilen cinsiyet sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.4: Öznitelik seçimi sonrası elde edilen veri setleri ile cinsiyet sınıflandırma performans sonuçları.

n Küme Çapraz Doğrulama	Yöntem (Öznitelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)
10	LBP (101)	90,3750	0,8073	<u>0,965</u>	1,66	89,0000	0,7800	0,956	<u>0,89</u>	92,1250	0,8420	0,921	9,97	90,3750	0,8072	0,904	1,15	<u>94,2500</u>	<u>0,8850</u>	0,944	1,30
	HOG (48)	87,5000	0,7500	<u>0,943</u>	1,33	84,7500	0,6952	0,922	<u>0,90</u>	52,8750	0	0,500	8,90	85,0000	0,6990	0,850	1,20	<u>90,2500</u>	<u>0,8044</u>	0,902	0,75
20	LBP (101)	90,2500	0,8049	0,806	1,39	89,1250	0,7825	<u>0,956</u>	<u>0,86</u>	92,3750	0,8471	0,924	18,08	90,5000	0,8098	0,906	1,86	<u>94,8750</u>	<u>0,8975</u>	0,950	1,21
	HOG (48)	87,6250	0,7522	<u>0,945</u>	1,00	84,7500	0,6952	0,922	<u>0,63</u>	52,8750	0	0,500	14,68	86,1250	0,7215	0,861	1,61	<u>90,8750</u>	<u>0,8171</u>	0,909	0,79

Öznitelik seçme işlemleri sonucunda 944 özniteliğe sahip LBP veri seti 101' e, 324 özniteliğe sahip HOG veri seti ise 48' e düşürülmüştür. Yaş sınıfları cinsiyet sınıflarından fazla olduğundan, elde edilen yeni özniteliklerin sayısı, Çizelge 4.2' de CFS algoritması uygulanmış veri setinden elde edilen özniteliklerin sayısından farklıdır. Burada öznitelik sayılarının az oluşu sınıflandırma sürelerine yansımıştır. Ancak düşük hesaplama süreleri, doğruluk değerlerinde meydana gelen düşüşleri beraberinde getirmiştir. Fakat yaşanan bu düşüşler yaş sınıflandırmada olduğu kadar fazla değildir.

Yine en düşük sınıflandırma süreleri Naïve Bayes algoritmasına aittir. Burada en yüksek ROC değerleri BayesNet algoritması tarafından üretilmiştir ancak sonuçlar birlikte ele alındığında, k-NN algoritmasının bariz üstünlüğü görülmektedir.

Çizelge 4.4' te bulunan sonuçlar incelendiğinde, öznitelik azaltma sonrasında cinsiyet sınıflandırma işlemi için en iyi sonuçlar LBP yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak k-NN algoritması tarafından 1,21 saniyede %94,8750 doğrulukla tamamlanmıştır. Bu değerler, Çizelge 4.3' te bulunan sonuçlarla karşılaştırıldığında, doğrulukta %2,5' lik bir azalmanın yanında hesaplama süresinde %86,08' lik bir kazanç meydana gelmiştir. Özellikle düşük işlem süresi gerektiren ve sadece cinsiyet tahmininin yapılacağı uygulamalarda bu yöntem kullanılabilir.

Önceki sonuçların birçoğunda olduğu gibi tüm yöntemler için en yüksek ROC ve Kappa değerleri yine k-NN algoritmasıyla üretilmiştir.

Yine burada LBP yönteminin HOG yöntemine göre başarısı görülmektedir.

Bu aşamaya kadar yapılan uygulamalarda yaş ve cinsiyet tahmini ayrı problemler olarak ele alınmış ve veri setleri buna göre hazırlanmıştır. Ayrıca elde edilen sonuçlar problemlere özgü olarak değerlendirilmiştir. Bu aşamadan sonra ise iki problem tek bir problem gibi ele alınacak ve işlemler buna göre yapılacaktır.

Çizelge 4.5' te, veri setlerine yaş ve cinsiyet problemleri için ayrı ayrı uygulanan CFS algoritması sonucunda elde edilen özniteliklerin, benzerlik filtresinden geçirilmesinin ardından elde edilen yaş sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.5: Yaş ve Cinsiyet tahmininde kullanılan özneliklerin birleştirilmesi sonucunda elde edilen veri seti ile yaş sınıflandırma ve performans analizi.

n Küme Çapraz Doğrulama	Yöntem (Öznelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)
10	LBP (218)	64,1250	0,3974	0,828	2,35	60,3750	0,3723	0,834	1,48	80,6250	0,6425	0,824	40,28	71,7500	0,4615	0,802	5,26	90,0000	0,8178	0,911	2,55
	HOG (90)	57,1250	0,2765	0,742	1,58	55,1250	0,2859	0,757	0,93	62,6250	0	0,500	12,26	68,5000	0,3415	0,704	3,33	82,8750	0,6938	0,855	0,96
20	LBP (218)	63,3750	0,3833	0,828	2,34	60,1250	0,3698	0,834	1,50	81,7500	0,6631	0,833	81,35	71,6250	0,4554	0,797	9,48	91,3750	0,8425	0,922	2,25
	HOG (90)	60,5000	0,3297	0,753	1,28	53,5000	0,2695	0,759	0,86	62,6250	0	0,500	22,16	67,5000	0,3025	0,690	5,93	84,6250	0,7240	0,870	1,01

Çizelge 4.5' te kullanılan veri setleri aşağıdaki adımlarla oluşturulmuştur;

- LBP öznitelikleriyle yaş sınıflarını içeren veri setine CFS algoritması uygulanır.
- LBP öznitelikleriyle cinsiyet sınıflarını içeren veri setine CFS algoritması uygulanır.
- Elde edilen özniteliklere benzerlik filtresi uygulanır ve aynı özniteliklerden sadece bir adet kullanılarak yeni LBP veri kümesi oluşturulur.
- Aynı işlemler HOG özniteliklerini içeren veri kümelerine de uygulanır ve HOG veri kümesi oluşturulur.

Yukarıdaki adımlar uygulanarak elde edilen veri kümeleri hem yaş hem de cinsiyet sınıflandırmada kullanılabilir. İki sınıf verisi için gerekli tüm öznitelikleri barındırıyor oluşu bu veri setlerinin iki problem için kullanılabilmesini mümkün kılmaktadır.

Çizelge 4.5' te bulunan sonuçlar incelendiğinde, yaş sınıflandırma işlemi için en iyi sonuçlar LBP yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak k-NN algoritması tarafından 2,25 saniyede %91,375 doğrulukla tamamlanmıştır. Bu sonuç, yaş sınıflandırması için elde edilen en yüksek doğruluk değerinden %2,875 daha düşük, %74,22 daha hızlıdır. Ayrıca öznitelik seçme işlemi sonrası oluşturulan ve Çizelge 4.2' de LBP veri setinden elde edilen doğruluk değerinden %2,375 daha yüksektir. Dolayısıyla ürettiği daha yüksek doğruluk, ROC ve Kappa değerleri ile düşük hesaplama süresi dikkate alındığında, yaş sınıflandırmasında kullanılan bu yöntemin, ham veri setinden de CFS uygulanmış veri setinden de daha kullanışlı olduğu görülmektedir.

Önceki sonuçlarda olduğu gibi en düşük hesaplama süreleri yine Naïve Bayes algoritması tarafından üretilmiştir.

Çizelge 4.6' da, aynı veri setlerinin kullanılmasıyla elde edilen cinsiyet sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.6: Yaş ve Cinsiyet tahmininde kullanılan özneliklerin birleştirilmesi sonucunda elde edilen veri seti ile cinsiyet sınıflandırma ve performans analizi.

n Küme Çapraz Doğrulama	Yöntem (Öznelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polynomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süre. (sn)
10	LBP (218)	87,0000	0,7396	0,943	2,15	86,7500	0,7346	0,934	1,11	94,6250	0,8923	0,947	21,35	92,5000	0,8496	0,925	2,55	94,3750	0,8875	0,945	2,13
	HOG (90)	86,8750	0,7373	0,933	1,80	84,8750	0,6978	0,913	1,06	52,8750	0	0,500	13,76	88,2500	0,7640	0,882	1,93	92,1250	0,8423	0,922	1,05
20	LBP (218)	87,7500	0,7549	0,944	2,20	86,6250	0,7320	0,934	1,43	95,6250	0,9123	0,957	35,20	92,6250	0,8522	0,927	4,03	96,0000	0,92	0,961	1,93
	HOG (90)	87,0000	0,7399	0,936	1,23	84,8750	0,6978	0,914	0,88	52,8750	0	0,500	24,83	89,0000	0,7792	0,889	2,39	93,1250	0,8623	0,932	1,08

Çizelge 4.6' da bulunan sonuçlar incelendiğinde, cinsiyet sınıflandırma işlemi için en iyi sonuçların LBP yöntemiyle elde edildiği görülmektedir. Genel olarak burada da k-NN algoritmasının üstünlüğü göze çarpmaktadır. 10 kümeli çapraz doğrulama yönteminde LibSVM algoritmasının doğrulukta çok küçük bir farkla k-NN algoritmasını geçtiği görülmektedir ancak k-NN algoritmasının %90,02 daha hızlı hesaplama süresi, yine bu algoritmayı en kullanışlı hale getirmektedir.

Çizelge 4.6' da bulunan sonuçlar incelendiğinde, cinsiyet sınıflandırma işlemi için en iyi sonuçlar LBP yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP öznitelikleri kullanılarak k-NN algoritması tarafından 1,93 saniyede %96,00 doğrulukla tamamlanmıştır. Bu sonuç, cinsiyet sınıflandırma için elde edilen en yüksek doğruluk değerinden %1,375 daha düşük, %77,63 daha hızlıdır. Ayrıca öznitelik seçme işlemi sonrası oluşturulan ve Çizelge 4.4' te LBP veri setinden elde edilen en yüksek doğruluk değerinden %1,125 daha yüksektir. Dolayısıyla ürettiği daha yüksek doğruluk, ROC ve Kappa değerleri ile düşük hesaplama süresi dikkate alındığında, cinsiyet sınıflandırmada kullanılan bu yöntemin, ham veri setinden de CFS uygulanmış veri setinden de daha kullanışlı olduğu görülmektedir.

Çizelge 4.7' de ise, önerilen yöntem kullanılarak oluşturulan veri setinden elde edilen yaş sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.7: Önerilen yöntem kullanılarak elde edilen veri seti ile yaş sınıflandırma ve performans analizi.

n Küme Çapraz Doğrulama	Yöntem (Öznitelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polinomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)	Doğruluk (%)	Kappa	ROC	Hes. Süř. (sn)
10	CFS _(LBP) + CFS _(HOG) (308)	65,1250	0,4066	0,826	2,65	60,8750	0,3785	0,834	<u>2,38</u>	80,6250	0,6425	0,824	53,93	77,1250	0,5746	0,833	6,91	<u>91,6250</u>	<u>0,8466</u>	<u>0,922</u>	2,98
20	CFS _(LBP) + CFS _(HOG) (308)	65,2500	0,4128	0,829	3,10	60,8750	0,3791	0,835	<u>1,83</u>	81,7500	0,6631	0,833	125,67	77,3750	0,5743	0,836	11,60	<u>93,1250</u>	<u>0,8741</u>	<u>0,937</u>	2,63

Çizelge 4.7’ de verilen sonuçların elde edildiği önerilen yönteme ait veri seti aşağıdaki adımlarla oluşturulmuştur;

- LBP öznitelikleriyle yaş sınıflarını içeren veri setine CFS algoritması uygulanır.
- LBP öznitelikleriyle cinsiyet sınıflarını içeren veri setine CFS algoritması uygulanır.
- İki yöntem sonucunda elde edilen öznitelikler birleştirilir.
- Aynı işlemler HOG öznitelikleri için tekrar edilir.
- Veri setleri benzerlik filtresinden geçirilir.
- Son olarak elde edilen iki veri seti bir araya getirilir ve yeni veri seti oluşturulur.

Elde edilen veri setinde iki farklı öznitelik seçme algoritması tarafından üretilen verilerin bir arada kullanılması, daha geniş ve anlamlı bir öznitelik uzayı meydana getirmektedir. Ayrıca iki farklı sınıf verisi için gerekli tüm öznitelikleri barındırıyor oluşu, bu veri setinin iki problem için de kullanılmasını mümkün kılmaktadır.

Çizelge 4.7 incelendiğinde, önerilen yöntem ile yaş sınıflandırma işlemi için elde edilen en iyi sonuçlar, 20 kümeli çapraz doğrulama yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP ve HOG öznitelikleri bir arada kullanılarak k-NN algoritması tarafından 2,63 saniyede %93,125 doğrulukla tamamlanmıştır. Bu değer, yaş sınıflandırmada elde edilen diğer sonuçlar ile karşılaştırıldığında, en yüksek doğruluk değerinden %1,125 daha az, %69,87 daha hızlıdır. Bununla birlikte, Çizelge 4.2’ de verilen CFS algoritması uygulanmış en yüksek sonuçtan % 4,125 daha fazla, Çizelge 4.5’ te verilen, yaş ve cinsiyet verilerini bir arada içeren veri setinden elde edilen en yüksek sonuçtan ise %1,85 daha fazla olmasına karşın sadece %15,45 daha yavaştır.

Elde edilen sonuçlar değerlendirildiğinde, hız ve yüksek doğruluk değerlerinin ön planda olduğu bir yaş sınıflandırması uygulamasında önerilen yöntemin diğer yöntemlere oranla daha başarılı olduğu görülmektedir.

Çizelge 4.8’ de, önerilen veri seti kullanılarak elde edilen cinsiyet sınıflandırma performans sonuçları bulunmaktadır.

Çizelge 4.8: Önerilen yöntem kullanılarak elde edilen veri seti ile cinsiyet sınıflandırma ve performans analizi.

n Küme Çapraz Doğrulama	Yöntem (Öznitelik Sayısı)	BayesNet (Simple Estimator)				Naïve Bayes				LibSVM (Polynomial Kernel)				SMO (Poly Kernel)				k-NN (Euclidean Distance)			
		<i>Doğruluk (%)</i>	<i>Kappa</i>	<i>ROC</i>	<i>Hes. Süř. (sn)</i>	<i>Doğruluk (%)</i>	<i>Kappa</i>	<i>ROC</i>	<i>Hes. Süř. (sn)</i>	<i>Doğruluk (%)</i>	<i>Kappa</i>	<i>ROC</i>	<i>Hes. Süř. (sn)</i>	<i>Doğruluk (%)</i>	<i>Kappa</i>	<i>ROC</i>	<i>Hes. Süř. (sn)</i>	<i>Doğruluk (%)</i>	<i>Kappa</i>	<i>ROC</i>	<i>Hes. Süř. (sn)</i>
10	CFS _(LBP) + CFS _(HOG) (308)	89,1250	0,7824	0,955	2,78	87,8750	0,7569	0,945	<u>1,63</u>	94,6250	0,8923	0,947	20,68	92,8750	0,8571	0,929	2,90	<u>95,3750</u>	<u>0,9074</u>	<u>0,955</u>	2,66
20	CFS _(LBP) + CFS _(HOG) (308)	88,8750	0,7774	0,957	3,05	88,1250	0,7619	0,946	<u>1,55</u>	95,6250	0,9123	0,957	40,08	92,8750	0,8572	0,929	4,41	<u>96,8750</u>	<u>0,9374</u>	<u>0,970</u>	2,86

Çizelge 4.8 incelendiğinde, önerilen yöntem ile cinsiyet sınıflandırma işlemi için elde edilen en iyi sonuçlar, 20 kümeli çapraz doğrulama yöntemiyle elde edilmiştir. Sınıflandırma işlemi, 20 kümeli çapraz doğrulama yönteminde LBP ve HOG öznitelikleri bir arada kullanılarak k-NN algoritması tarafından 2,86 saniyede %96,875 doğrulukla tamamlanmıştır. Bu değer, cinsiyet sınıflandırmada elde edilen diğer sonuçlar ile karşılaştırıldığında, en yüksek doğruluk değerinden %0,5 daha az, %66,85 daha hızlıdır. Bununla birlikte, Çizelge 4.4' te verilen CFS algoritması uygulanmış en yüksek sonuçtan %2,00 daha fazla, Çizelge 4.6' da verilen, yaş ve cinsiyet verilerini bir arada içeren veri setinden elde edilen en yüksek sonuçtan ise %0,875 daha fazla olmasına karşın %32,51 daha yavaştır.

Elde edilen sonuçlar incelendiğinde, hız ve yüksek doğruluk değerlerinin ön planda olduğu bir cinsiyet sınıflandırması uygulamasında önerilen yöntemin diğer yöntemlere oranla daha başarılı olduğu görülmektedir.

Farklı öznitelik seçme algoritmalarının bir arada kullanılmasıyla elde edilen veri seti, çeşitlilik içerdiğinden, diğer yöntemlerden elde edilen sonuçlara oranla daha iyi değerler üretmiştir. Bununla birlikte, önerilen yöntemin içerdiği öznitelik sayısının fazlalığı, hesaplama sürelerinde az da olsa artışa sebep olmuştur. Ancak bu artışın, yüksek doğruluk değerlerine ulaşmada kabul edilebilir bir seviyede olduğu düşünülmektedir.

Önerilen yönteme ait en iyi sınıflandırma sonuçları, 2008-2011 yılları arasında yapılmış çalışmalar ile karşılaştırılarak Çizelge 4.9 ve Çizelge 4.10' da verilmiştir (Berbar, 2014).

Çizelge 4.9: Yaş gruplarının sınıflandırılmasında önerilen yöntem kullanılarak elde edilen sonuçların literatür ile karşılaştırılması.

Yaş Grubu	Önerilen Yönteme Ait Sonuçlar	Benzer Çalışmaya Ait Sonuçlar (V. V. Nabiyev & Günay, 2010)
10	<u>84,6000</u>	?
20	<u>96,4000</u>	93,9153
30	<u>87,7000</u>	87,3239
40	<u>88,5000</u>	79,8780
50	<u>88,0000</u>	87,8048
60	<u>87,5000</u>	<u>89,7435</u>
ORTALAMA	<u>93,1250</u>	89,0070

Çizelge 4.10: Cinsiyet sınıflandırmada önerilen yöntem kullanılarak elde edilen sonuçların literatür ile karşılaştırılması.

Önerilen Yöntem	(Jabid, Kabir, & Chae, 2010)	(Andreu & Mollineda, 2008)	(Özbudak, Kirci, Çakir, & Güneş, 2010)	(Mayo & Zhang, 2008)	(Guo, Lin, & Nguyen, 2010)	
<i>Doğruluk (%)</i>	<u>96,8750</u>	95,0500	95,0000	93,0000	92,5000	88,8900



5. SONUÇLAR

Bu çalışmada yüz örüntülerinden yaş ve cinsiyet tespitinin bir arada gerçekleştirilmesi için bir yöntem önerilmiştir. Önerilen yöntemde, örüntü tanıma sistemlerinde öznelik çıkartma algoritması olan LBP ve HOG yöntemleri bir arada kullanılmış ve hibrit bir sistem tasarlanmıştır. İki farklı yöntemin bir araya getirilmesi öznelik uzayında çeşitliliği sağlamış, bu da sınıflandırma sonuçlarına pozitif olarak yansımıştır. Ayrıca kullanılan öznelik seçme algoritması ile veri setinin boyutu küçültülmüş, böylece sınıflandırma süreleri, elde edilebilecek maksimum doğruluk değerleri için minimize edilmeye çalışılmıştır.

Tasarlanan sistem için karşılaştırılan beş sınıflandırıcıdan LibSVM ve k-NN sınıflandırıcıları farklı konularda birbirlerine üstün gelmiştir. Özellikle düşük hesaplama süresi ve maksimuma en yakın doğruluk değerlerini üreten k-NN algoritması, birçok problemde kullanılan tüm algoritmaların önüne geçmiştir.

Bu çalışmada yaş ve cinsiyet tahmini tek bir problem olarak ele alındığından, bu problemi çözmek için üretilecek veri setini oluşturmak temel hedef olarak belirlenmiştir. Çizelge 4.7 ve 4.8' de, önerilen yöntemin problemi tanımlama başarısının diğer yöntemlerden daha yüksek olduğu görülmektedir. Ayrıca önerilen yöntem ile elde edilen sonuçların, literatürde bulunan diğer çalışmalarda elde edilen sonuçlardan üstün olduğu Çizelge 4.9 ve 4.10' da görülmektedir.

Çizelge 4.1 – 4.8’ de bulunan 20 kümeli çapraz doğrulama yöntemine ait doğruluk ve hesaplama süreleri dikkate alındığında; ham veri seti kullanılarak yapılan yaş tespiti işleminde en yüksek doğruluk değeri k-NN algoritması tarafından 8,73 saniyede %94,250 olarak belirlenmiştir. Bununla birlikte CFS öznitelik seçme algoritması uygulanan veri setinde ise aynı algoritma 1,56 saniyede %89,00 doğrulukla sınıflandırma işlemini gerçekleştirmiştir. Bu sonuç ham veri setinden elde edilen doğruluk değerinden %5,250 düşük, hesaplama süresinden ise 7,17 saniye (%83,13) daha hızlıdır. Aynı veri seti için önerilen yöntem kullanıldığında yaş tespiti 2,63 saniyede %93,125 doğrulukla gerçekleşmiştir. Bu sonuç; ham veri setinden elde edilen doğruluk değerinden %1,125 düşük, hesaplama süresinden ise 6,1 saniye (%69,87) daha hızlıdır. Önerilen yönteme ait sonuçlar, öznitelik seçme işlemi uygulanmış veri setinden elde edilen sonuçlarla karşılaştırıldığında, önerilen yöntem ile üretilen doğruluk değerinin %4,125 fazla, hesaplama süresinin ise 1,07 saniye (%40,68) daha yavaş olduğu görülmektedir. Bu sonuçlara göre hem doğruluk hem de hesaplama süreleri dikkate alındığında önerilen yöntemin daha kullanışlı olduğu görülmektedir.

Cinsiyet tespitinde de en yüksek değerlerin aynı algoritma tarafından üretildiği görülmektedir. Ham veri seti kullanılarak yapılan cinsiyet sınıflandırma işlemi 8,63 saniyede %97,37 doğruluk değeri ile tamamlanmıştır. CFS öznitelik seçme algoritması uygulanan veri setinde ise aynı algoritma 1,21 saniyede %94,87 doğruluk değeri üretmiştir. Bu değer, ham veri setinden elde edilen doğruluk değerinden %2,5 düşük, hesaplama süresinden ise 7,42 saniye (%86) daha hızlıdır. Aynı veri seti için önerilen yöntem kullanıldığında, cinsiyet tespiti 2,86 saniyede %96,875 doğrulukla gerçekleştirilmiştir. Bu sonuç; ham veri setinden elde edilen doğruluk değerinden %0,5 düşük, hesaplama süresinden ise 5,77 saniye (%65,86) daha hızlıdır. Önerilen yönteme ait sonuçlar, öznitelik seçme işlemi uygulanmış veri setinden elde edilen sonuçlarla karşılaştırıldığında, önerilen yöntem ile üretilen doğruluk değerinin %2,0 fazla, hesaplama süresinin ise 1,65 saniye (%61,70) daha yavaş olduğu görülmektedir.

Yapılan deneyler sonucunda, farklı dokusal öznitelik yöntemlerinin bir arada kullanılmasının başarımlarını arttırdığı görülmüştür. Önerilen yöntemin yaş ve cinsiyet sınıflandırma için ürettiği benzer hesaplama süreleri, yöntemin problemlerin çözümündeki tutarlılığını ve kullanılabilirliğini arttırmaktadır. Aynı anda yüksek doğruluk ve ROC değerlerinin üretiliyor oluşu, elde edilen sonuçların geçerliliğini arttırmaktadır. Bununla birlikte herhangi bir sınıflandırma modeline ihtiyaç duymadan çalışıyor oluşu, önerilen yöntemin özellikle düşük işlem gücüne sahip gömülü sistemler için uygunluğunu göstermektedir.

Çalışmada ele alınan konu oldukça önemli ve gerçek hayatta uygulanabilirliği yüksektir. Bu nedenle sonuçlar üzerinde optimizasyon çalışmaları yapılması, uygulamanın gerçek hayatta kullanılabilirliğini olumlu yönde etkileyecektir. Sınıflandırıcı modelinde kullanılan verilerin sayısının artırılması, modelin daha geniş bir kitleye hitap etmesini sağlayabilir. Öznitelik çıkartma algoritmalarında yapılacak optimizasyon çalışmaları, problemleri tanımlamada daha etkili sonuçlar üretebilir. Böylece sınıflandırma sonuçlarında yüksek doğruluk değerlerine ulaşılması sağlanabilir. Ayrıca daha etkin öznitelik seçme algoritmalarının kullanımı, daha düşük sayıda öznitelikle, kısa sürede daha yüksek doğruluk değerlerine ulaşmayı sağlayabilir.

Bundan sonraki çalışmalarda yaş verilerinin gruplar halinde değil, bireysel olarak ele alınması planlanmaktadır. Böylece daha fazla sınıflı bir model üzerinde çalışılacak ve sınıflandırıcıların daha karmaşık işlemler üzerinde ayrıştırma yapması sağlanacaktır. Bununla birlikte hibrit sınıflandırıcılar ve çeşitli öznitelik çıkartma algoritmaları kullanılarak yüksek doğruluk değerlerine ulaşılması hedeflenmektedir.



KAYNAKLAR

- Ahonen, T., Hadid, A., & Pietikäinen, M.** (2004). Face recognition with local binary patterns *Computer vision-eccv 2004* (pp. 469-481): Springer.
- Alpaydin, E.** (2010). Introduction to Machine Learning, MIT Press. *Cambridge, MA, USA*.
- Altman, N. S.** (1992). An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3), 175-185.
- Andreu, Y., & Mollineda, R. A.** (2008). On the complementarity of face parts for gender recognition *Progress in Pattern Recognition, Image Analysis and Applications* (pp. 252-260): Springer.
- Babur, S., Turhal, U., & Akbaş, A.** (2012). DVM Tabanlı Kalın Bağırsak Kanseri Tanısı İçin Performans Geliştirme Performance Improvement for SVM Based Colon Cancer Recognition.
- Bellman, R.** (1957). Dynamic Programming Princeton University Press. *Princeton, NJ*.
- Ben-Gal, I.** (2007). Bayesian networks. *Encyclopedia of statistics in quality and reliability*.
- Berber, M. A.** (2014). Three robust features extraction approaches for facial gender classification. *The Visual Computer*, 30(1), 19-31.
- Biricik, G., Diri, B., & Sönmez, A. C.** (2012). Abstract feature extraction for text classification. *Turkish Journal of Electrical Engineering & Computer Sciences*, 20(Sup. 1), 1137-1159.
- Bishop, C. M.** (2006). Pattern Recognition. *Machine Learning*.
- Bledsoe, W.** (1964). The model method in facial recognition, Panoramic Research Inc., Palo Alto: CA, Technical Report, Technical Report PRI: 15.
- Bunte, K., Hammer, B., Wismüller, A., & Biehl, M.** (2010). Adaptive local dissimilarity measures for discriminative dimension reduction of labeled data. *Neurocomputing*, 73(7), 1074-1092.
- Burt, D. M., & Perrett, D. I.** (1995). Perception of age in adult Caucasian male faces: Computer graphic manipulation of shape and colour information. *Proceedings of the Royal Society of London B: Biological Sciences*, 259(1355), 137-143.
- Burton, A. M., Bruce, V., & Dench, N.** (1993). What's the difference between men and women? Evidence from facial measurement. *Perception*, 22(2), 153-176.
- Chang, C.-C., & Lin, C.-J.** (2011). LIBSVM: a library for support vector machines. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(3), 27.

- Chellappa, R., & Ramanathan, N.** (2006). *Face Recognition across Age Progression*. Paper presented at the Proc. Face Recognition Grand Challenge and ICE Workshop.
- Cohen, E. R.** (1998). *An Introduction to Error Analysis: The Study of Uncertainties in Physical Measurements*: IOP Publishing.
- Cohen, J.** (1960). A coefficient of agreement for nominal scales. *Educational and Psychosocial Measurement*, 20, 37-46.
- Computer Vision - ACCV 2010.** (2011). (R. Kimmel, R. Klette, & A. Sugimoto Eds. 1 ed.): Springer-Verlag Berlin Heidelberg.
- Cootes, T. F., Edwards, G. J., & Taylor, C. J.** (1998). Active appearance models *Computer Vision—ECCV'98* (pp. 484-498): Springer.
- Dalal, N., & Triggs, B.** (2005). *Histograms of oriented gradients for human detection*. Paper presented at the Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on.
- Gao, F., & Ai, H.** (2009). Face age classification on consumer images with gabor feature and fuzzy lda method *Advances in biometrics* (pp. 132-141): Springer.
- Gonzalez-Ulloa, M., & Flores, E. S.** (1965). senility of the face-basic study to understand its causes and effects. *Plastic and Reconstructive Surgery*, 36(2), 239-246.
- Guo, J.-M., Lin, C.-C., & Nguyen, H.-S.** (2010). *Face gender recognition using improved appearance-based average face difference and support vector machine*. Paper presented at the System Science and Engineering (ICSSE), 2010 International Conference on.
- Günay, A., & Nabiyeu, V. V.** (2008). *Automatic age classification with LBP*. Paper presented at the Computer and Information Sciences, 2008. ISCIS'08. 23rd International Symposium on.
- Gwet, K.** (2001). Statistical tables for inter-rater agreement. *StatAxis, Gaithersburg*.
- Hall, M. A.** (1999). *Correlation-based feature selection for machine learning*. The University of Waikato.
- Heikkila, M., & Pietikainen, M.** (2006). A texture-based method for modeling the background and detecting moving objects. *IEEE transactions on pattern analysis and machine intelligence*, 28(4), 657-662.
- Heusch, G., Rodriguez, Y., & Marcel, S.** (2006). *Local binary patterns as an image preprocessing for face authentication*. Paper presented at the Automatic Face and Gesture Recognition, 2006. FGR 2006. 7th International Conference on.
- Hogarth, R. M.** (1977). *Methods for aggregating opinions*: Springer.
- Jabid, T., Kabir, M. H., & Chae, O.** (2010). *Gender classification using local directional pattern (LDP)*. Paper presented at the Pattern Recognition (ICPR), 2010 20th International Conference on.
- Jain, A., & Huang, J.** (2004). *Integrating independent components and linear discriminant analysis for gender classification*. Paper presented at the Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on.

- Karungaru, S., Fukumi, M., & Akamatsu, N.** (2003). *Morphing face images using automatically specified features*. Paper presented at the Circuits and Systems, 2003 IEEE 46th Midwest Symposium on.
- Kawano, T., Kato, K., & Yamamoto, K.** (2004). *A comparison of the gender differentiation capability between facial parts*. Paper presented at the Pattern Recognition, 2004. ICPR 2004. Proceedings of the 17th International Conference on.
- Kılınç, M.** (2012). *Yüz Resimlerinden Yaş Bilgisinin Tespit Edilmesi*. (Yüksek Lisans), Gebze Yüksek Teknoloji Enstitüsü.
- Kobayashi, T., Hidaka, A., & Kurita, T.** (2007). *Selection of histograms of oriented gradients features for pedestrian detection*. Paper presented at the Neural Information Processing.
- Kohavi, R.** (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Paper presented at the Ijcai.
- Kohavi, R., & Provost, F.** (1998). Glossary of terms. *Machine Learning*, 30(2-3), 271-274.
- Kwon, Y. H., & Lobo, N. D. V.** (1994). *Age classification from facial images*. Paper presented at the Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on.
- Landis, J. R., & Koch, G. G.** (1977). The measurement of observer agreement for categorical data. *biometrics*, 159-174.
- Mannila, H.** (1996). *Data mining: machine learning, statistics, and databases*. Paper presented at the ssdbm.
- Marcel, S., Rodriguez, Y., & Heusch, G.** (2007). On the recent use of local binary patterns for face authentication.
- Mayo, M., & Zhang, E.** (2008). *Improving face gender classification by adding deliberately misaligned faces to the training data*. Paper presented at the Image and Vision Computing New Zealand, 2008. IVCNZ 2008. 23rd International Conference.
- Možina, M., Demšar, J., Kattan, M., & Zupan, B.** (2004). *Nomograms for visualization of naive Bayesian classifier*: Springer.
- Nabiyev, V., & Koçak, F.** (2011). Yerel İkili Örneklerle Yüz Güzelliği Kararı.
- Nabiyev, V. V., & Günay, A.** (2010). LBP yardımıyla görüntüdeki kişinin yaşının bulunması. *Çankaya University Journal of Science and Engineering*, 8(1).
- Ojala, T., Pietikäinen, M., & Harwood, D.** (1996). A comparative study of texture measures with classification based on featured distributions. *Pattern recognition*, 29(1), 51-59.
- Ojala, T., Pietikäinen, M., & Mäenpää, T.** (2002). Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 24(7), 971-987.

- Özbudak, Ö., Kirci, M., Çakir, Y., & Güneş, E. O.** (2010). *Effects of the facial and racial features on gender classification*. Paper presented at the MELECON 2010-2010 15th IEEE Mediterranean Electrotechnical Conference.
- Peker, M., Altun, H., & Karakaya, F.** (2009). HOG Temelli Bir Yöntem ile Ölçek ve Yönden Bağımsız Gerçek Zamanlı Nesne Tanıma.
- Phillips, P. J., Moon, H., Rizvi, S. A., & Rauss, P. J.** (2000). The FERET evaluation methodology for face-recognition algorithms. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(10), 1090-1104.
- Phillips, P. J., Wechsler, H., Huang, J., & Rauss, P. J.** (1998). The FERET database and evaluation procedure for face-recognition algorithms. *Image and vision computing*, 16(5), 295-306.
- Platt, J.** (1998). Sequential minimal optimization: A fast algorithm for training support vector machines.
- Poggio, B., Brunelli, R., & Poggio, T.** (1992). HyperBF networks for gender classification.
- Powers, D. M.** (2012). *The problem with kappa*. Paper presented at the Proceedings of the 13th Conference of the European Chapter of the Association for Computational Linguistics.
- Rifkin, R. M.** (2002). *Everything old is new again: a fresh look at historical approaches in machine learning*. Massachusetts Institute of Technology.
- Sim, J., & Wright, C. C.** (2005). The kappa statistic in reliability studies: use, interpretation, and sample size requirements. *Physical therapy*, 85(3), 257-268.
- Simon, P.** (2013). *Too Big to Ignore: The Business Case for Big Data* (Vol. 72): John Wiley & Sons.
- Stiefelhagen, R., Bernardin, K., Ekenel, H. K., & Voit, M.** (2008). *Tracking identities and attention in smart environments-contributions and progress in the CHIL project*. Paper presented at the Automatic Face & Gesture Recognition, 2008. FG'08. 8th IEEE International Conference on.
- Swets, J. A.** (2014). *Signal detection theory and ROC analysis in psychology and diagnostics: Collected papers*: Psychology Press.
- Url-1.** <<http://www.tdk.gov.tr/>>, alındığı tarih: 12.04.2016.
- Url-2.** <<http://www.nist.gov/itl/iad/ig/colorferet.cfm>>, alındığı tarih: 20.03.2016.
- Viola, P., & Jones, M. J.** (2004). Robust real-time face detection. *International journal of computer vision*, 57(2), 137-154.
- Wang, J., Shang, Y., Su, G., & Lin, X.** (2006). *Age simulation for face recognition*. Paper presented at the Pattern Recognition, 2006. ICPR 2006. 18th International Conference on.
- Wernick, M. N., Yang, Y., Brankov, J. G., Yourganov, G., & Strother, S. C.** (2010). Machine learning in medical imaging. *Signal Processing Magazine, IEEE*, 27(4), 25-38.

- Yang, M.-H., & Moghaddam, B.** (2000). *Support vector machines for visual gender classification*. Paper presented at the Pattern Recognition, 2000. Proceedings. 15th International Conference on.
- Zanni, L., Serafini, T., & Zanghirati, G.** (2006). Parallel software for training large scale support vector machines on multiprocessor systems. *The Journal of Machine Learning Research*, 7, 1467-1492.
- Zhang, H.** (2005). Exploring conditions for the optimality of naive Bayes. *International Journal of Pattern Recognition and Artificial Intelligence*, 19(02), 183-198.





ÖZGEÇMİŞ



Ad Soyad: Uğur TURHAL

Doğum Yeri ve Tarihi: Yomra, 13.01.1988

Adres: Balıkesir Üniversitesi, Mühendislik Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü

E-Posta: ugurturhal@hotmail.com

Lisans: Marmara Üniversitesi, Elektronik Bilgisayar Eğitimi Bölümü, Bilgisayar ve Kontrol Öğretmenliği Programı, 2011

Mesleki Deneyim ve Ödüller:

2013 - Halen Uzman,
Balıkesir Üniversitesi, Mühendislik-Mimarlık Fakültesi, Bilgisayar Mühendisliği Bölümü

Yayın ve Patent Listesi:

- Yörük Y.E., **Turhal U.**, İstanbullu A., "Classification of Pima Indians Diabetes Data Set and Effect of Missing Data". *9. International Symposium on Health Informatics and Bioinformatics*, Muğla – Turkey, 16-17 November 2015.
- **Turhal U.**, Gök M, "YUFES Web Server: Building Feature Vectors of Proteins". *International Conference on Applied Informatics for Health and Life Sciences (AIHLS 2014)*, Kuşadası - Turkey, 19-22 October 2014.
- **Turhal U.**, Gök M., Durgut A., "Comparison Among Feature Encoding Techniques For HIV-1 Protease Cleavage Specificity". *The International Conference on Advanced Technology & Sciences (ICAT'14)*, Antalya - Turkey, 12-15 Aug 2014.
- **Turhal, U.**, Gök M., Onur S., Babur S., "Performance Analysis of Feature Ranking Algorithms on Microarray Datasets". *The 5th International Symposium on Sustainable Development (ISSD 2014)*, Sarajevo - Bosnia & Herzegovina, 15-18 May 2014.

- Cömert B., İstanbullu A., **Turhal, U.**, "Low Cost And Portable Heartbeat Rate Measurement From The Finger". *The 5th International Symposium on Sustainable Development (ISSD 2014)*, Sarajevo - Bosnia & Herzegovina, 15-18 May 2014.
- Akbaş A., **Turhal U.**, Babur S. and Avcı C., "Performance Improvement with Combining Multiple Approaches to Diagnosis of Thyroid Cancer". *The 7th International Conference on Bioinformatics and Biomedical Engineering (iCBBE 2013)*, Beijing - China, 26-28 Sep 2013.
- **Turhal U.**, Babur S., Avcı C., Akbas A., "Performance Improvement for Diagnosis of Colon Cancer by Using Ensemble Classification Methods". *The International Conference on Technological Advances in Electrical, Electronics and Computer Engineering (TAECE 2013)*, Konya - Turkey, 9-11 May 2013.
- Babur S., **Turhal U.**, Akbaş A., "DVM Tabanlı Kalın Bağırsak Kanseri Tanısı İçin Performans Geliştirme". *Elektrik-Elektronik ve Bilgisayar Mühendisliği Sempozyumu (ELECO 2012)*, Bursa - Türkiye, 29 Kasım-1 Aralık 2012.

TEZDEN TÜRETİLEN YAYINLAR/SUNUMLAR

- **Turhal U.**, Akbas A., "Yüz Örüntülerinden Cinsiyet Tespitinde Hibrit Özniteliklerle Performans Geliştirme". *Elektrik-Elektronik ve Bilgisayar Sempozyumu (EEB2016)*, Tokat – Türkiye, 11-13 Mayıs 2016.