

T. C.
BALIKESİR ÜNİVERSİTESİ
MÜHENDİSLİK FAKÜLTESİ
BİLGİSAYAR BİLİMLERİ ANABİLİM DALI



GERÇEK ZAMANLI EMOJİ OLUŞTURUCU

LİSANS BİTİRME PROJESİ

İBRAHİM ETHEM BİLDİRİCİ

BALIKESİR, HAZİRAN-2021

ÖZET

GERÇEK ZAMANLI EMOJİ OLUŞTURUCU

İBRAHİM ETHEM BİLDİRİCİ

BALIKESİR ÜNİVERSİTESİ

MÜHENDİSLİK FAKÜLTESİ

BİLGİSAYAR BİLİMLERİ ANABİLİM DALI

(DANIŞMAN: Dr. Öğr Üyesi FATİH AYDIN)

BALIKESİR, 2021

Günümüzde derin öğrenme, bilgisayarla görme uygulamaları ve araştırmalarında sıklıkla kullanılan bir tekniktir. Her ne kadar sıklıkla içerik tabanlı görüntü elde edimi uygulamalarında kullanılsa da farklı bilgisayar görmesi alanlarında da kullanımına imkân vardır. Bu çalışmada, FER2013 veri setinde bulunan yüz ifadelerini otomatik olarak sınıflandırmak üzere Konvolüsyonel Sinir Ağları (CNN) tabanlı bir yüz ifadesi tanıma sistemi geliştirilmiştir. Geliştirilen Sinir Ağı bir arayüzde kullanılarak Gerçek Zamanlı Emoji Oluşturan bir program tasarlanmıştır. Geliştirilen CNN, FER2013 veri seti üzerinde % 73.8 başarımla sonuçlandırmıştır.

ANAHTAR KELİMELER: Yüz İfadesi Tespiti, Derin Öğrenme, Duygu Analizi, Konvolüsyonel Sinir Ağları

İÇİNDEKİLER

	<u>Sayfa</u>
ÖZET.....	i
İÇİNDEKİLER.....	ii
1. GİRİŞ.....	1
2. DERİN ÖĞRENME.....	2
2.1 KONVOLÜSYONEL SİNİR AĞLARI(CNN).....	2
2.1.1 KONVOLÜSYONEL KATMANLAR.....	2
2.1.2 POOLING.....	3
2.1.3 AKTİVASYON FONKSİYONU KATMANI.....	3
2.1.4 NORMALİZASYON.....	4
2.1.5 SOFTMAX4	
3. DENEYLER.....	5
3.1 KONVOLÜSYON SİNİR AĞI YAPISI(CNN).....	6
3.2 KULLANILAN YAZILIM ve DONANIMLAR.....	7
3.3 VERİ TABANI.....	7
3.4 UYGULAMA SONUÇLARI.....	7
4. SONUÇ.....	10

1. GİRİŞ

Görüntü işleme, hareketli ve hareketsiz imgeler üzerinde düzenleme, çıkarım veya görüntü tanıma yapmak için çalışılan bir alandır.

Günümüzde hızla bir yenisi eklenen görüntü işleme teknolojisi sayesinde gerçek görüntülerin işlenmesi ile görüntülerden anlam çıkarılarak doğruya çok yakın sonuçlar elde edilebilmektedir. Sistemlerin gelişimi ile kısa sürede daha kolay, düşük hata oranı ve zaman kaybına uğramadan sonuçlara ulaşılabilmektedir. Derin öğrenme sınıfına ait teknolojiler gidererek artmakta ve makine öğrenme teknolojisi, farklı uygulama alanlarında hayatlarımızı kolaylaştırmaktadır.

Örneğin; e-ticaret yaparken ilgi ve kişi ihtiyaçlarımıza göre ürün önermesi, sosyal medyada otomatik resim etiketleme, daha önce yapılan aramalardan yeni aramalar önerme ve göstermesi, mobil cihazlarda ses kontrolü ve konuşma tanıma fırsatı tanımaktadır.

Bu özelliklere ilaveten, insanın yüz ifadesinden duygu analizi, kişi tanıma, nesne algılama ve tanıma, doğal dil işleme, tıbbi uygulamalarda, sürücüsüz otomobillerde ve daha birçok alanda kullanılmaktadır [1].

Yüz ifadeleri, günlük sosyal etkileşimlerimizde hayati bir role sahiptir. Yüz ifadelerini, duygularımızı ifade etmek ve başkalarının bize karşı olan duygu ve tutumlarını anlamakta kullanmaktayız. Yüz ifadelerinden anlamlar çıkarmak insanoğlunun henüz birkaç aylıkken öğrendiği ve sahip olduğu bir yetenektir. Makinelerin de insanlar kadar olmasa bile buna yakınsayacak bir şekilde yüz ifadelerini tanımaları hedeflenmektedir. Özellikle insan Bilgisayar etkileşimi konusu üzerinde sıklıkla kullanım alanı bulan otomatik yüz ifade analizi için gerçekleştirilen çalışmalar büyük önem arz etmektedir [2].

Tipik bir yüz tanıma sistemi üç aşamadan oluşmaktadır. Bunlar;

Aşama 1: Yüz tespiti ve lokalizasyon,

Aşama 2: Elde edilen yüzlerden özellik çıkarımı,

Aşama 3: Çıkarılan özellikleri kullanarak verilen sınıflara göre yüzleri sınıflandırmak.

Bu çalışmada, çok sınıflı bir yüz ifadesi tespit sistemi için CNN tabanlı bir yaklaşım önerilmiştir. CNN modelini eğitmek ve doğrulamak için FER2013 veri seti tarafından sunulan eğitim ve doğrulama verileri kullanılmıştır. Eğitim ve doğrulama aşamalarından sonra ortaya çıkan CNN, aynı veri setinin test verisiyle doğrulanmıştır. Son aşamada CNN, doğrulama verilerinde % 73.8 başarı sağlamıştır.

2. DERİN ÖĞRENME

Derin öğrenme, birçok katmanlı ileri beslemeli sinir ağlarının eğitim süreçlerinden oluşmaktadır. Hazırlanan modeller çok sayıda farklı nitelikte gizli katman ile oluşturulduğundan derin öğrenme ismi verilmiştir. Akademik ve özel sektör alanlarında çalışan veri bilimciler hareketli hareketsiz imge sınıflandırma, imge işleme-düzenleme, video analizi ve sınıflama, ses tanıma ve işleme ve doğal dil öğrenme süreci olmak üzere çeşitli uygulamalarda kullanılmaktadır. Derin öğrenme özellikle, büyük miktarlarda, etiketlenmemiş eğitim verilerinden öznitelik çıkarım yöntemleri kullanarak özelliklerin saptamasını yapabilen sistemler oluşturmak için ileri teknoloji yapay sinir ağların kullanılması ile oluşmaktadır.

Derin öğrenme mimarisi birçok katman ve saklı değişkenden oluşur. Derin öğrenmenin en sık kullanılan algoritmaları, Derin Sinir Ağları (Deep Neural Networks), Otomatik Kodlayıcılar (Autoencoders) ve Boltzmann Makinelerinin türevleri olan Kısıtlı Boltzmann Makineleridir [4]. Son zamanlarda özellikle görüntü işleme alanlarında görüntüyü bütün olarak işleyen ve veriyi sınıflara veya özelliklerine ayıran, başarı oranı yüksek olan konvolüsyon yöntemi kullanılmaktadır. Bu çalışmamızda da derin öğrenme yöntemlerinden olan Konvolüsyonel Sinir Ağları (Convolutional Neural Networks – CNN) yöntemi kullanılmıştır.

2.1 KONVOLÜSYONEL SİNİR AĞLARI(CNN)

1988 yılında Yann LeCun tarafında geliştirilen CNN, çok katmanlı sinir ağlarının özel geliştirilmiş bir türüdür. Farklı mimarilere sahip olmaları yanı sıra klasik yapay sinir ağları gibi ileri yayılım algoritmasına sahiptirler [5]. CNN, ön işlem hacmini minimum tutularak piksel görüntülerinden doğrudan görsel kalıpları tanımak için geliştirilmiştir. Ani değişkenliğe sahip desenler ve geometrik dönüşümlere karşı iyi sonuç verir.

CNN, öznitelik çoğaltma ve özetleme katmanlarını içererek, diğer algoritmalarından farklı bir öznitelik çıkarma işlevini barındırmaktadır. CNN temel olarak 4 katmandan oluşur. Bu katmanlar; konvolüsyon katmanı, aktivasyon fonksiyonu katmanı, pooling katmanı ve normalizasyon katmanları olarak adlandırılmaktadır [6]. Bu katmanlara ek olarak farklı özellikte katmanlar geliştirilmekle beraber son katmanda çok sınıflı sınıflandırma için genellikle softmax katmanı tercih edilmektedir.

2.1.1 KONVOLÜSYONEL KATMANLAR

Bir görüntüdeki tüm alt bölge için, çıktı özneliği haritasında tek bir değer üretmek için bir dizi matematiksel işlem gerçekleştirir. Bir imge/videolardaki gerçek görüntüler değişimeme özelliğine sahiptir yani imgeler sabittir. Bu nedenle görüntünün bir bölümünün

sayısal değerleri diğer bölümler ile aynı olduğu anlamına gelir. Bu, görüntünün bir bölümünde öğrendiğimiz özelliklerin aynı zamanda resmin diğer bölümlerine de uygulanabileceğini ve aynı özellikleri tüm konumlarda kullanabileceğimizi göstermektedir.

Gerçek ya da ölçeklendirilmiş bir imgeden rastgele seçilen küçük ölçekli parçalardan daha fazla özellik öğrendikten sonra, öğrenilen bu öznelik parçasını görüntünün diğer bölgelerine uygulayabiliriz. Bu yöntem ile daha büyük imgeler ile konvolüsyona sokarak görüntüdeki her konumda farklı bir özellik etkinleştirme değeri elde edilebilir.

2.1.2 POOLING

Konvolüsyon katmanını kullandıktan sonra elde edilen özellikler sınıflandırılmak istenmektedir [7]. Bu çalışmada kullandığımız FER2013 veri setinde bulunan 48x48 boyutundaki oldukça küçük görüntülere hazırlamış olduğumuz 9x9 girişli ve 1000 özellik öğrenimli bir konvolüsyon katmanını ele aldığında; Her konvolüsyon, $(48-9 + 1) * (48-9 + 1) = 1600$ boyutlarında bir çıktı ile sonuçlanır ve 300 özellik öğrenimli bu ağda örnek başına $1600*1000 = 1.6*10^6$ özellik vektörü oluşturur. Oldukça küçük seçilen görüntü üzerinde bile milyondan fazla özelliği olan girdileri olan bir sınıflandırıcı öğrenmek çok yavaş olur ve aynı zamanda aşırı uyumluluk gösterebilir. Bunun önlenmesi için geliştirilen havuzlama teknikleri ile bu oran azaltılır. Sık kullanılan pooling yöntemi maxPooling ve meanPooling yöntemleridir. maxPooling bir grup piksellerde olan en yüksek piksel değerini alırken meanPooling ise ortalama değeri alarak hesaplama yapar.

2.1.3 AKTİVASYON FONKSİYONU KATMANI

Belirli bir eşik değerine göre nöronların aktif olup olmamalarını sağlayan fonksiyonlardan oluşur. Aktivasyon işlemi, bir özellik haritasının her bileşenine (diğer bir deyişle noktasal olarak) uygulanan lineer olmayan bir aktivasyon fonksiyonuyla doğrusal bir filtre izlenerek elde edilir. ELU yöntemi, sigmoid ve tanjant fonksiyonu ile karşılaştırıldığında daha hızlı sonuca ulaşmaktadır. Bu durum özellikle işlem kapasitesi bakımından kısıtlı olan bilgisayarlar için belirgin bir hız artışına olanak sağlamaktadır [8].

ELU kullanımının en olumsuz yanı, eğitim sırasında bu ünitelerin kırılabilir olması ve bunun da veriye göre olumsuz sonuçlar üretebilmesidir [8]. Örneğin, öğrenme oranı çok yüksek ayarlandığı durumlarda, ağırlık eğitim veri setinin tamamında etkinleştirilmeyen nöronlar bulunabilir ve bu durumda ELU üniteleri eğitim sırasında geri döndürülemez. Bu durumun önlenmesi için öğrenme oranının uygun bir şekilde ayarlanması gerekmektedir.

2.1.4 NORMALİZASYON

ELU katmanları sonucunda oluşabilecek kırılmaları ya da güçlü tepkileri önlemek için normalleştirmeye ihtiyaç duyulmaktadır. Normalleştirme katmanı giriş haritasındaki bütün mekânsal konumdaki özellik kanallarının vektörünü normalleştirmek için kullanılır.

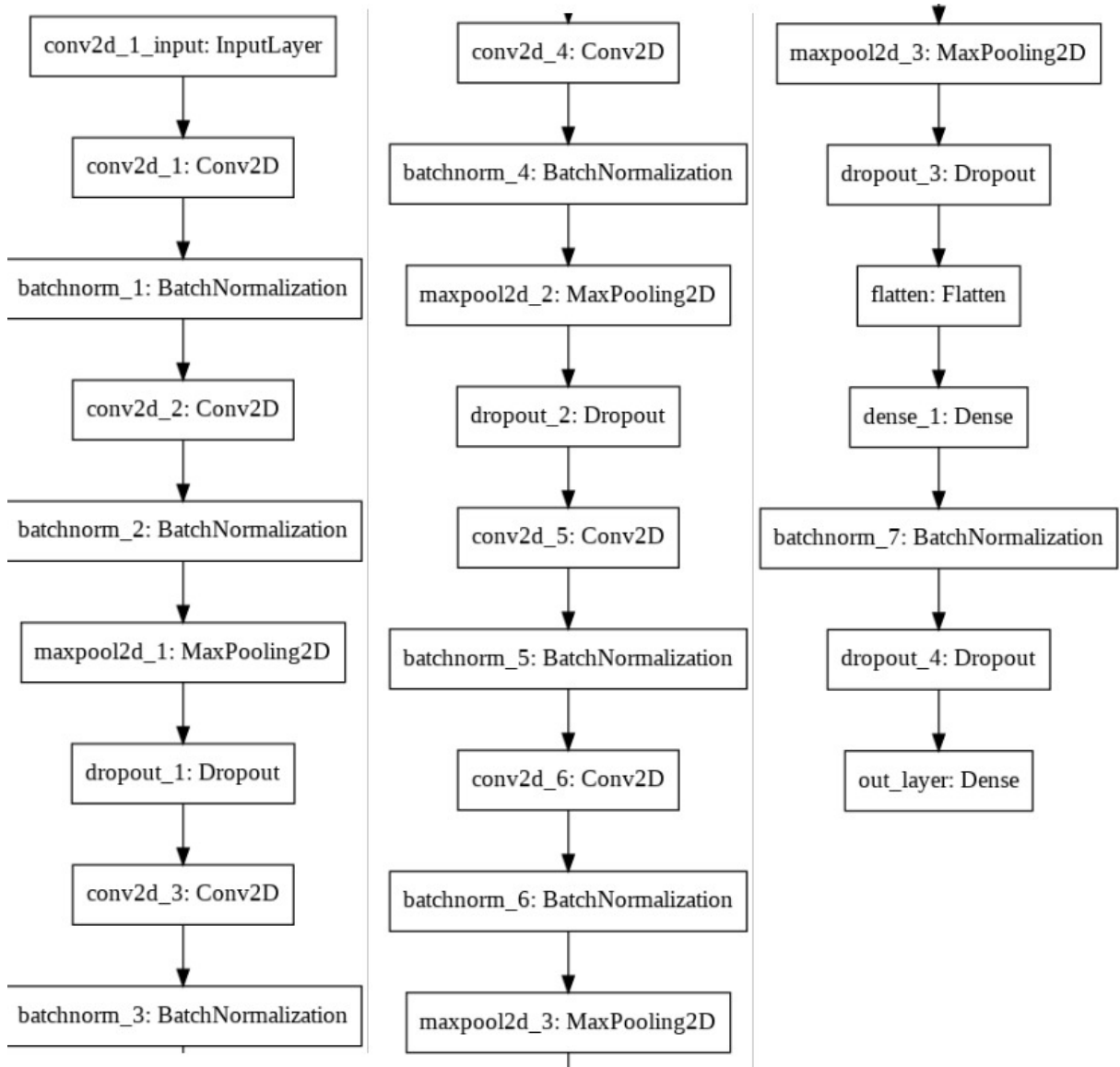
Normalizasyonda imgenin düzenlenmesi için 4 parametre ile ayarlama yapılmalıdır. Bu çalışmamızda $\beta_1 = 0.9$, $\beta_2 = 0.999$, $\epsilon = 10^{-7}$, öğrenme oranı 0.001 olarak ayarlanmıştır.

2.1.5 SOFTMAX

Softmax regresyonu çok sınıflı verilerin sınıflandırılması için kullanılan başarılı bir sınıflandırma yöntemidir. Özellikle yüksek sayıda sınıf verileri için kullanılır. Bu çalışmada toplam 7 sınıfta etiketlenmiş veriler bulunmaktadır. Ağımızın son katmanında sınıflandırma için Softmax katmanı kullanılmıştır.

3. DENEYLER

Yüz ifadelerinin tespiti için açık veritabanı olarak paylaşılan toplamda 35887 adet yüz imgeleri [3] kullanılarak veriler 7 yüz ifadesine sınıflandırılmaya çalışılmıştır. FER2013 veri seti üzerinde yapılan literatür taramasında, Liu ve arkadaşlarının [2] dataset üzerinde birçok önışlem yapıldıktan sonra belirlenen CNN modeli üzerinden verilerin tümü eğitilmiş ve doğruluk oranı en yüksek 62.44% olarak tespit edilmiştir. Yine HongWei ve arkadaşlarının yapmış olduđu çalışmasında [9] bir CNN modeli ile aynı veriler üzerinde uygulanmış ve 10% test verisinde 55,6%, doğrulama da ise 48,5% bir başarı oranı elde edilmiştir. Bu çalışmalar incelendiğinde görüntüler model uygulanmadan önce bir dizi ön işlemden geçirildikten sonra hazırlanan modele uygulanmıştır. Bu başarının büyük çoğunluğu verilerin önışlemden geçirilmesinden kaynaklanmaktadır. Bu çalışmamızda [3] verileri üzerinde önışlem yapılmadan gri formatta bulunan 48x48 boyutunda imgeler Şekil 1 de görüldüğü gibi hazırlanan CNN modeline uygulanmış ve başarımları değerlendirilmiştir.



Şekil 1. Hazırlanan CNN modeli.

3.1 KONVOLÜSYON SİNİR AĞI YAPISI(CNN)

CNN modeli Konvolüsyon, ELU, Pooling ve sonuç olarak softmax katmanlarında oluşan model şekil 2’de gösterilmiştir.

Şekil 1’de görüldüğü gibi ağ 25 katmandan oluşmaktadır. Bu model de 4’er konvolüsyon, ELU, maxPool katmanı ve 3 Full Connect katmanlarından oluşmaktadır. Hazırlanan modelin son katmanında softmax ile imgeler sınıflandırılmaktadır. Bu model kullanılarak elimizdeki kamuya açık gerçek imgeler test edilerek başarımları değerlendirilemeye çalışılmıştır

3.2 KULLANILAN YAZILIM ve DONANIMLAR

İmge veri tabanında belirlenen imgelerin eğitim ve test işlemler FER2013 veri tabanından sağlanmıştır. Burada Google Colab ortamında imgeler sayısal formata dönüştürülmüştür. Şekil 2’deki model kullanılmıştır. Yapılan bu çalışmada, Google Colab’ın sağladığı NVIDIA TESLA T4S, NVIDIA TESLA K80, NVIDIA P4 ve NVIDIA TESLA P100 ekran kartları kullanılmıştır. Arayüz tasarımı Python programlama dili ve Jupyter Notebook IDE’si kullanılarak, i5-3230M işlemcili, Intel HD Graphics 4000 ekran kartı ve 8GB RAM bulunan bilgisayarda yapılmıştır.

3.3 VERİ TABANI

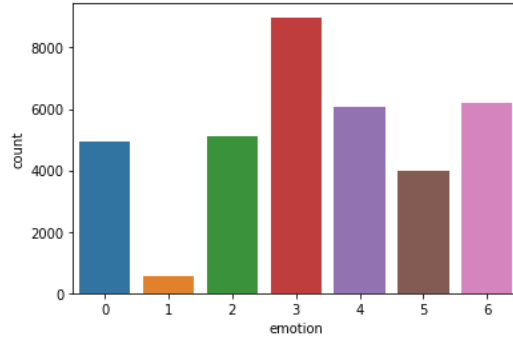
FER2013 dataseti insan yüz ifadeleri barındıran gerçek imgelerden oluşmaktadır. Kamuya açık olarak kullanılan bir veri tabanıdır. FER datasetinde toplamda 35.887 adet rastgele bir araya getirilmiş yüz imgesi ve bu imgelerin etiket kısımları bulunmaktadır. Bu yüz imgeleri incelendiğinde içinde 7 sınıfa ait rastgele dizilmiş imgeler bulunmaktadır. Yüz ifadelerinde kızgın, iğrenme, korkma, sevinç, üzgün, şaşırma ve doğal yüz bulunmaktadır. Buradaki verinin 80%’i eğitim, 10%’u doğrulama ve geriye kalan son %10’luk bölümü ise test edilmek üzere seçilmiştir.

3.4 UYGULAMA SONUÇLARI

Yapılan bu çalışmada yüz imgeleri için toplam da 35887 adet 48x48 boyutlu gri formatta imgeler CNN modelinden geçirildikten sonra elde edilen özellikler ağı son katmanı olan softmax sınıflandırıcı ise sınıflandırılmıştır. Veri setinde farklı yüz ifadelerinden oluşan 7 tür imge aynı zamanda tüm verilerin sınıf bilgileri de bulunmakta ve test işlemi bu değerlere göre yapılmıştır. Aşağıdaki Tablo 1’ de bu veri tabanındaki yüz ifadelerin gerçek sayıları ve tüm verilere oranları görülmektedir.

No	Yüz İfadesi	Sayısı	Oranı
1	Kızgın (Anger)	4.953	0.1380
2	Tiksinme (Disgust)	547	0.0152
3	Korkma (Fear)	5.121	0.1427
4	Mutluluk (Happiness)	8.989	0.2505
5	Üzüntü (Sadness)	6.077	0.1693
6	Şaşkınlık (Surprise)	4.002	0.1115
7	Doğal (Neutral)	6.198	0.1727
Toplam		35.887	1

Tablo 1.

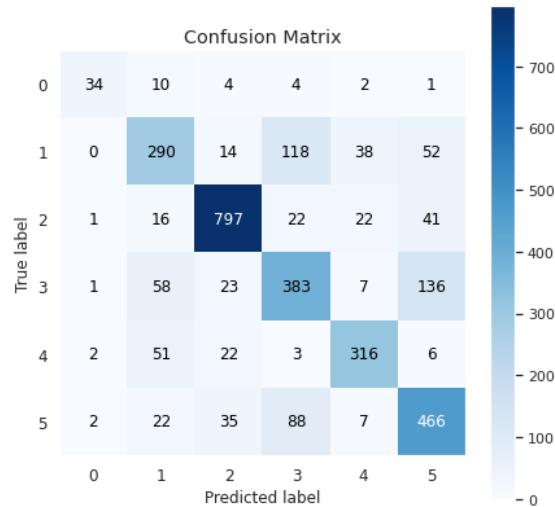


Tablo 2.

Yapılan bu çalışmada yüz imgeleri için toplam da 35887 adet 48x48 boyutlu gri formatta imgeler CNN modelinden geçirildikten sonra elde edilen özellikler ağı son katmanı olan softmax sınıflandırıcı ile sınıflandırılmıştır. Veri setinde farklı yüz ifadelerinden oluşan 7 tür imge aynı zamanda tüm verilerin sınıf bilgileri de bulunmakta ve test işlemi bu değerlere göre yapılmıştır. Aşağıdaki Tablo 1’ de bu veri tabanındaki yüz ifadelerin gerçek sayıları ve tüm verilere oranları görülmektedir.

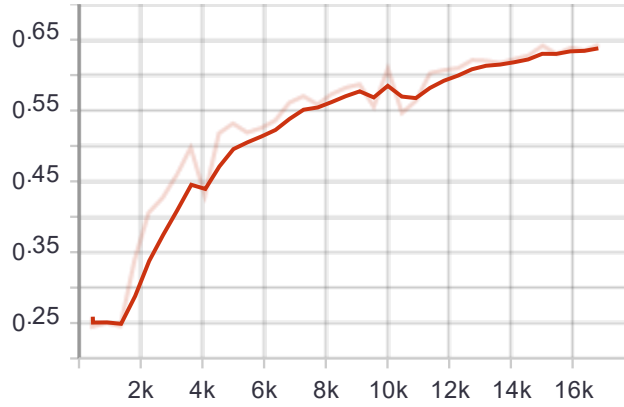
Tablo 1 incelendiğinde veri setinde bulunan yüz ifadesi türlerinin eşit oranda dağılmadığı tespit edilmiştir. Tiksinme türündeki imgelerin sayısı 547 ve tüm verilere oranı 0.0152 iken diğer taraftan en çok bulunan imge türü 8989 sayısı ile mutluluk olarak görülmekte ve bunun da veri tabanına oranı 0.2505 olarak ölçülmüştür.

Önerilen model imgelerin komşular arası ilişkilere bakarak tanımayla çalışmaktadır. Model kendini eğitirken ilk önce bu ilişkileri çıkarmakta daha sonra test aşamasında bu belirlemiş olduğu öğrenmeleri kullanmaktadır. Test işlemi sonucunda elde ettiğimiz sonuç verileri şekil 2’deki karşılaştırma matrisine dönüştürülerek sonuçlar daha anlaşılır yapılmaya çalışılmıştır.

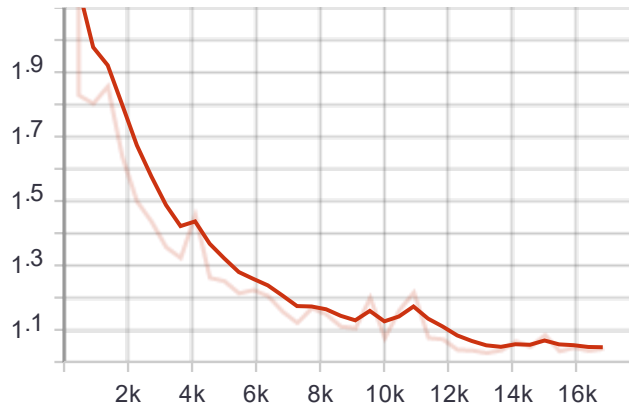


Şekil 2.

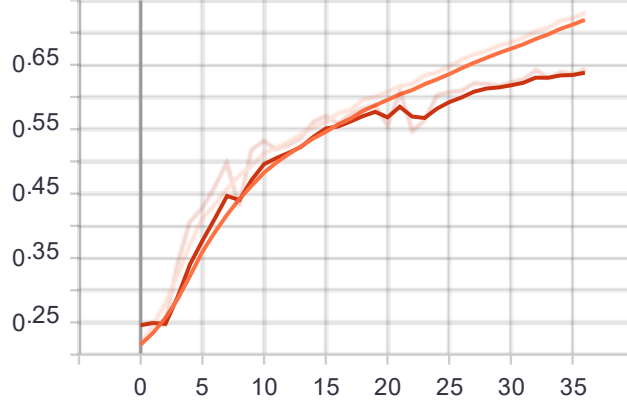
Şekil 2’deki karşılaştırma matrisini incelediğimizde en yüksek başarı oranı 2 yani Korku türünde imgelerin olduğu tespit edilmiştir. Yine 3 nolu mutluluk imgelerinde ve 5 nolu şaşkınlık imgelerinde oldukça yüksek başarı oranı tespit edilmiştir. Doğal görünüm verileri baktığımızda bu görüntüler insan yüz ifadesine göre değişebilmektedir. Bazı insanların normal yüz ifadesi üzgün ya da mutlu olarak ta görünebilir. Bu yüz ifadesi kişiye özgü bir durumdur.



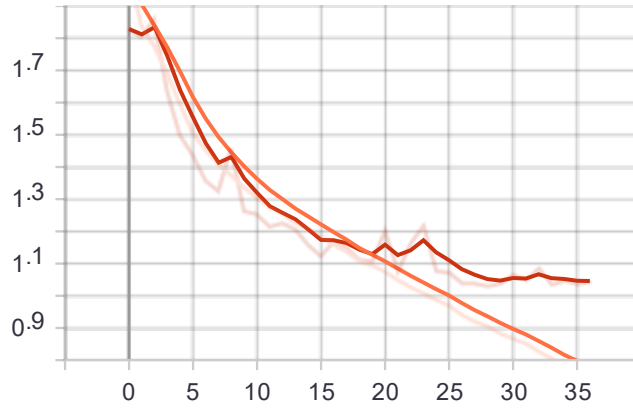
Şekil 3. İterasyon sayısına göre doğruluk oranı



Şekil 4. İterasyon sayısına göre kayıp oranı



Şekil 4. Epoch sayısına göre doğruluk oranı



Şekil 5. Epoch sayısına göre kayıp oranı

Her Epoch için 870 iterasyon yapılmıştır. 36. Epoch'da overfitting'i engellemek için erken durdurma uygulanarak model eğitimi sonlandırılmıştır.

4. SONUÇ

Bu çalışmamızda gri formatta bulunan kamuya açık gerçek görüntüler üzerinde yüz ifadelerine ait imgeler tasarlanan CNN yöntemine göre sınıflandırması yapılmaya çalışılmıştır. Yapılan çalışmada geliştirilen CNN yönteminin yüz ifadelerini herhangi bir ön işlemten geçirmeden tespit etmesinde iyi sonuçlar verdiği görülmüştür. Özellikle mutluluk, şaşkınlık ve tikslenme yüz ifadelerinde oldukça yüksek başarımlar elde edilmiştir. Tüm bu

başarımlar incelendiğinde % 10'luk verilerin test edilmesi ile %73.8 oranında bir başarı elde edilmiştir.

	precision	recall	f1-score	support
0	0.85	0.62	0.72	55
1	0.65	0.57	0.60	512
2	0.89	0.89	0.89	899
3	0.62	0.63	0.62	608
4	0.81	0.79	0.80	400
5	0.66	0.75	0.70	620
accuracy			0.74	3094
macro avg	0.75	0.71	0.72	3094
weighted avg	0.74	0.74	0.74	3094

Tablo 2.

Her sınıf için ayrı ve ortalama precision, recall, f1-score ve support değerleri Tablo 2'de gösterilmiştir.