

Yüz Görüntülerinden Duygu Durum Tahmini

1st Bengi Yurdusever
Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
bengi.yurdusever@std.yildiz.edu.tr

2nd Şilan Fidan Vural
Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
fidan.vural@std.yildiz.edu.tr

3rd Furkan Yüksel
Bilgisayar Mühendisliği
Yıldız Teknik Üniversitesi
İstanbul, Türkiye
furkan.yuksel@std.yildiz.edu.tr

Abstract—Dünyanın her yerinde yüz ifadeleri iletişimin önemli araçlarından bir tanesidir. Bu kapsamda projede yüz görüntülerinden duygu durum tahmininin gerçekleştirilmesi hedeflenmiştir. Bu hedef doğrultusunda FER2013 veri seti, derin öğrenme modelleri ile eğitilerek veri seti içerisindeki yüz görüntülerinin sınıflandırılması yapılmıştır. FER2013 veri seti içerisinde duygu durum tahmini için 7 farklı sınıf bulunmaktadır. Çalışmamızda bu sınıfları tahmin edebilmek için kendi eğittiğimiz CNN modeli, transfer öğrenmesi yoluyla eğittiğimiz önceden eğitilen ResNet ve VGG modelleri kullanılmıştır. Çalışmamız sonucunda kendi eğittiğimiz CNN modeli test setinde %67 başarı ile en yüksek skoru vermiştir.

Index Terms—derin öğrenme, yüz, duygu durum, transfer öğrenmesi, CNN, VGG16, ResNet50

I. GİRİŞ

Günlük hayatımızda birçok farklı durumla karşılaşır ve bu durumlara farklı tepkiler veririz. Duygular, insanların karşılaştıkları olaylar hakkında hissettikleri güçlü hisler olarak tanımlanabilir. Duyguların ifade edilmesinde yüz ifadeleri önemli rol oynar.

Duygu durum tahmini, görüntü işleme teknikleri kullanarak insan yüzündeki ifadeleri okumayı sağlayan bir teknolojidir. Yıllar içinde gelişen ve büyüyen birçok yüz tanıma teknolojilerinden biri haline gelmiştir ve birçok kullanım alanı bulunmaktadır. Çevrimçi eğitimde öğrencilerin duygularının algılanarak eğitimin kalitesinin artırılması, gerçek zamanlı oyunlarda kullanıcıların hangi duyguları deneyimlediğini belirleme ve sürücünün araba kullanırken yorgunluk durumunun tespitinin yapılması gibi alanlar örnek verilebilir.

Duygu durum tahmini problemini çözebilmek için evrimsel sinir ağları ile derin öğrenme modelleri oluşturduk. Kendi CNN modelimiz, transfer öğrenmesi yoluyla oluşturduğumuz ResNet ve VGG olmak üzere 3 farklı model ile çalışmalarımızı gerçekleştirdik.

Bu makaledeki amacımız, duygu durum tahmininin nasıl yapıldığının anlaşılabilir olarak kullanılan algoritmaların performanslarının değerlendirilmesidir. Ek olarak en iyi performansla sahip olan modelin gerçek zamanlı olarak nasıl çalıştığı incelenmiştir.

II. VERİ KEŞFİ

A. Veri Seti

FER (Facial Emotion Recognition), birçok veri seti ile üzerinde fazla çalışma yapılmış bir alandır. Projemizde bu radaki veri setleri içinden FER2013 veri setini eğitim ve test amaçlı olarak kullandık.

FER-2013 veri seti 35,887 tane etiketli yüz görselinden oluşmaktadır. Bu verilerin 28709 tanesi eğitim için ayrılmışken 7178 tanesi ise test için ayrılmıştır. FER veri setindeki görseller 48x48 boyutuna sahip gri-seviye görsellerdir. Bu veri setindeki ifadeler 7 duygu durumundan biriyle etiketlenmişlerdir. Bunlar kızgın (angry), iğrenmiş (disgust), korkmuş (fear), mutlu (happy), nötr (neutral), üzgün (sad) ve şaşkın (surprise) şeklindedir. Veri setindeki sınıf örnekleri Figure 1’de gösterilmiştir.



Fig. 1. FER2013 Veri Seti Sınıf Örnekleri

B. Eğitim ve Test Verileri Dağılımları

Bu kısımda eğitim ve test veri setlerinin dağılımları incelenmiştir. Bu dağılımlar Figure 2 ve Figure 3’te bar grafiği ile görselleştirilmiştir.

Grafikler incelendiğinde veri setinin dengesiz dağıldığı görülmektedir. En fazla veri sayısının mutlu (happy) sınıfına ve en az veri sayısının ise iğrenmiş (disgust) sınıfına aittir.

Eğitim seti 3995 tane kızgın, 3171 tane şaşkın, 4830 tane üzgün, 436 tane iğrenmiş, 4097 tane korkmuş, 7215 tane mutlu ve 4965 tane nötr duygu durumundan oluşmaktadır. Test seti ise 958 tane kızgın, 831 tane şaşkın, 1247 tane üzgün, 111 tane

iğrenmiş, 1024 tane korkmuş, 1774 tane mutlu ve 1233 tane nötr duygu durumundan oluşmaktadır. Bakıldığında hem veri setinin dengesiz olduğu görülmektedir hem de görece küçük bir veri setidir.

Ayrıca, test setinin eğitim setindeki dağılımı iyi temsil edecek şekilde ayrıldığı görülmektedir.

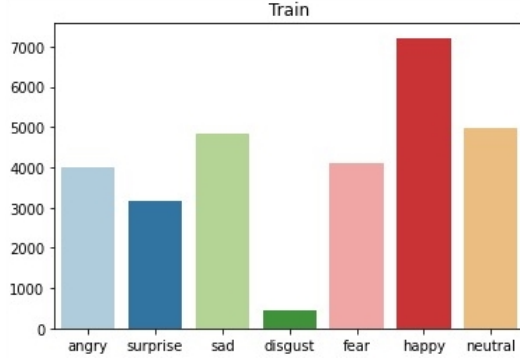


Fig. 2. Eğitim Seti Dağılımı

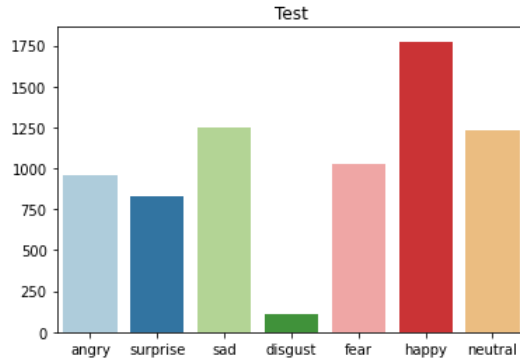


Fig. 3. Test Seti Dağılımı

III. MODELLER

A. CNN Modeli

İlk aşamada bir CNN modeli kullanılarak duygu durum tahmini yapılmaya çalışılmıştır. CNN modeli 3 adet 2 aşamalı Convolution katmanlarından ve bunu takip eden MaxPool katmanından oluşmaktadır. [3] Bu yapıyı 2 aşamalı Fully Connected yani Dense katmanı takip etmektedir ve son olarak ise 7 sınıflı bir çıkış katmanı bulunmaktadır. Convolution katmanlarının filtre sayıları sırasıyla 32, 64, 128, 256, 512 ve 1024'tür. Kernel boyutu ise (3,3) olarak belirlenmiştir. Aktivasyon fonksiyonu olarak da ReLu kullanılmıştır. Maxpool katmanı kernel büyüklüğü ise default olarak (2,2) şeklindedir. Fully Connected katmanların büyüklükleri sırasıyla 256 ve 512'dir. Aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Çıkış katmanının büyüklüğü ise 7'dir ve aktivasyon fonksiyonu olarak Softmax belirlenmiştir. Katmanlardan sonra hem hesaplama karmaşıklığını düşürmek hem de aşırı öğrenmeyi engellemek için 0.3 oranında bir Dropout uygulanmıştır. Bu

CNN yapısının 50 epoch kadar eğitilmiştir ve bunun sonucunda elde edilen modelde %67 civarında bir test başarıları ortaya çıkmıştır.

Figure 4'te CNN modeli kısımında bahsettiğimiz modelin basit bir görselleştirmesi görülmektedir.

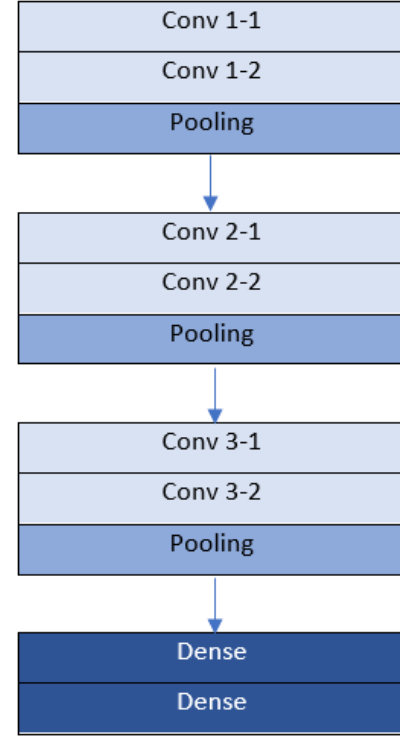


Fig. 4. CNN Modeli

B. Transfer Öğrenmesi

Transfer öğrenmesi (Transfer Learning), bir problemi çözerken kazanılan bilgiyi depolamaya ve depolanan bilgiyi farklı ancak ilgili bir probleme uygulamaya odaklanan bir makine öğrenmesi (ML) problemidir. Proje kapsamında VGG ve ResNet modelleri transfer öğrenmesi yöntemi kullanılarak eğitilmiştir.

Çoğu derin öğrenme ağı, yüksek doğruluk sağlamak için milyonlarca parametreyi öğrenir. Aynı amaç için çalışan birçok ağ benzer özellik ve ağırlıklar öğrenecektir. Transfer öğrenmesi bu zaman ve yer karmaşıklığının azaltılmasını sağlar.

1) *ResNet50*: ResNet50, ImageNet veri seti üzerinde eğitilmiş 50 katmanlı bir ağıdır.

Önceden eğitilmiş ResNet modelinin ağırlıklarını dondurmada ve üzerine 3 tane Fully Connected yani Dense katmanı ekleyerek bir model oluşturulmuştur. Bu Dense katmanlarının büyüklükleri sırasıyla 256, 128 ve 64 şeklindedir. Bu katmanlarda aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Öğrenme parametresi ise 0.00001 olarak belirlenmiştir.

Önceden eğitilmiş olan katmanları dondurmak istediğimizde ise oldukça düşük bir başarı elde etmekle birlikte belli bir iterasyon sayısından sonra da başarı artmamıştır. Bu nedenle

katmanları dondurmak yerine eğitilmelerine izin verilmiştir. ResNet50 modeli 50 epoch kadar eğitilmiştir ve bunun sonucunda %59'luk bir test başarısı elde edilmiştir.

2) *VGG16*: VGG16, ImageNet veri seti üzerinde eğitilmiş 16 katmanlı bir ağıdır.

Önceden eğitilmiş VGG16 modelinin ağırlıklarını dondurmada ve üzerine 3 tane Fully Connected katmanı ekleyerek bir model oluşturulmuştur. Bu katmanların büyüklükleri sırasıyla 256, 128 ve 64 şeklindedir. Aktivasyon fonksiyonu olarak ReLu kullanılmıştır. Öğrenme parametresi 0.00001 olarak belirlenmiştir.

Önceden eğitilmiş olan katmanları dondurmak istediğimizde ise oldukça düşük bir başarı elde edilmiştir. Bu nedenle katmanları dondurmak yerine eğitilmelerine izin verilmiştir. VGG16 modeli 50 epoch kadar eğitilmiştir ve bunun sonucunda %61'lik bir test başarısı elde edilmiştir.

IV. MODEL SONUÇLARI

Aşağıda Figure 5'de eğittiğimiz modeller ve bunların test seti üzerindeki başarıları gösterilmiştir.

Görüldüğü gibi en yüksek test başarısı %67 performans ile CNN modelinde elde edilmiştir. Bununla birlikte VGG ve ResNet modellerinin 50 epoch yerine 100 epoch kadar eğitildiğinde test başarısında artış olduğu bilinmektedir. [1]

Model	Derinlik	Test Başarısı
CNN	8	%67
ResNet50	50	%59
VGG16	16	%61

Fig. 5. Modeller ve Performansları

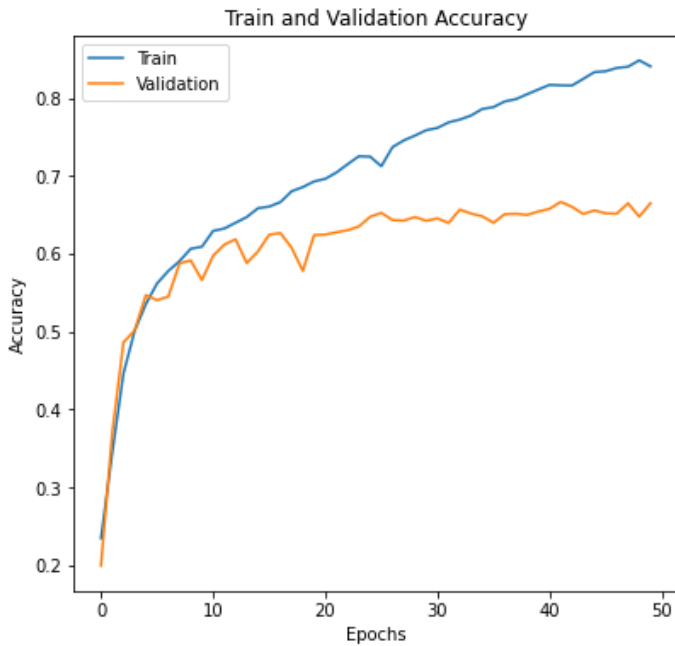


Fig. 6. CNN Modeli Train ve Validation Performansı

Figure 6'da CNN modeli için train ve validation performansları gösterilmiştir. Burada dikkat edilmesi gereken nokta yaklaşık 25-30. epoch'tan sonra train setinin başarısı artmaya devam ederken validation setinin başarısında değişim gözlenmemesidir. Aslında bu durum modelin aşırı öğrendiğine bir işarettir. Aşırı öğrenmeyi önlemek için Dropout değeri artırılabilir, eğitim 25-30. epoch arasında kesilebilir ya da düzenlemeler(regularization) eklenebilir.

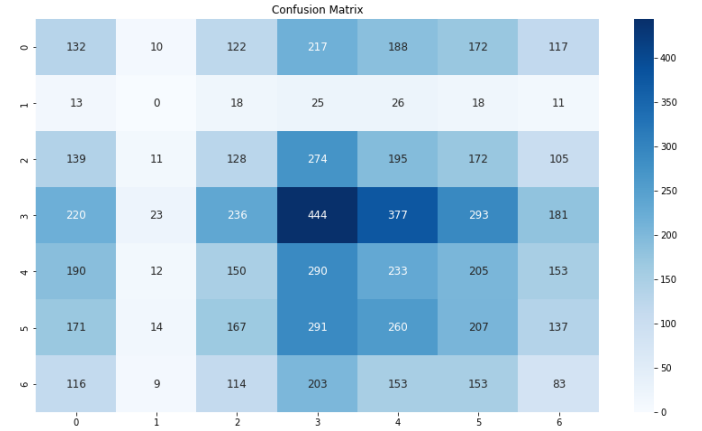


Fig. 7. Confusion Matrix

Figure 7'daki karmaşıklık matrisinde sınıflar sırasıyla (0-Kızgın(Angry), 1-İğrenmiş(Disgust), 2-Korkmuş(Fear), 3-Mutlu(Happy), 4-Üzgün(Sad), 5-Şaşırmış(Surprised), 6-Nötr(Neutral)) olarak gösterilmiştir.

Baktığımızda en iyi tahminin Mutlu(Happy) sınıfına ait olduğunu görmekteyiz. Ancak yine görmekteyiz ki modelimiz birçok yanlış tahmin de yapmaktadır ve İğrenmiş(Disgust) sınıfı için oldukça kötü çalışmaktadır.

V. İLGİLİ ÇALIŞMALAR

FER2013, araştırmacıları daha iyi FER sistemleri geliştirmeye teşvik etmek için bir Kaggle yarışması olarak Goodfellow ve ekibi tarafından tasarlanmıştır [4]. Yarışmadaki ilk üç takımın tümü, görüntü dönüşümleri ile ayırtıcı bir şekilde eğitilmiş CNN'leri kullandı. Kazanan Yichuan Tang, eğitim için kayıp fonksiyonu olarak bir SVM'nin temel amacını kullanarak %71.2 doğruluk elde etti.

FER alanında çok sayıda mevcut araştırma bulunmaktadır. Özellikle, S. Li ve W. Deng tarafından FER hakkında yakın zamanda yayınlanan makale, FER'e yönelik derin öğrenmeye dayalı yaklaşımların mevcut durumuna ışık tutmaktadır [5]. Pramerdorfer ve Kampel'in başka bir makalesi [6], FER2013'te %75.2 test doğruluğu elde etmek için mevcut altı tane son teknoloji makalede yer alan yaklaşımları açıklar ve onlardaki ağırları birleştirir; bu, bir dergide yayınlanmış en yüksek performans değeri olarak bilinmektedir.

Altı makale arasında, Zhang ve arkadaşları yardımcı veriler ve ek özellikler kullanarak %75.1'lik yüksek doğruluk elde etti [7]. Yüz görüntüsü parçalarından bir HoG öznelik vektörü hesaplandı ve CNN'in ilk Fully Connected katmanı tarafından işlendi. Ayrıca, zorlu koşullarda bile iyi çalıştığı öne sürüldü.

(Yüz ana hatları çıkarma için FER veri setindeki görüntülerin yaklaşık %15'i hatalı olarak çalışmaktadır). En yüksek ikinci doğruluğa sahip araştırma olarak bilinmektedir.

VI. GERÇEK ZAMANLI UYGULAMA

Elde ettiğimiz modellerden CNN modeli seçilerek gerçek zamanlı olarak nasıl çalışacağı incelenmiştir. Bu kısımda ilk olarak HaarCascade algoritması ile yüz tespiti yapılmıştır [8]. Tespit edilen yüz bölgesine ise CNN modelimiz uygulanarak duygu durumları tespit edilmiştir. Bunun için internet üzerinden bulduğumuz 3 adet fotoğraf denenmiştir. Bu fotoğraflar sırasıyla mutlu, üzgün ve kızgın ifadeleri içermektedir. Bu fotoğraflar kameraya gösterilerek elde edilen duygu durumları Figure 8, Figure 9 ve Figure 10'da belirtilmiştir. Ancak gerçek zamanlı uygulama her ne kadar duygu durumlarını tespit etse de dezavantajı yavaş çalışmasıdır.



Fig. 8. Mutlu Yüzün Algılanması

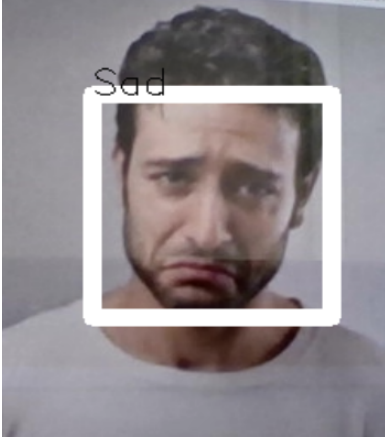


Fig. 9. Üzgün Yüzün Algılanması

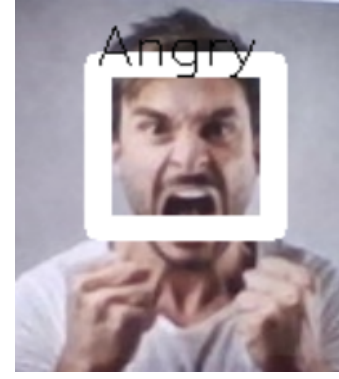


Fig. 10. Kızgın Yüzün Algılanması

VII. ZORLUKLAR

Veri setindeki görseller insanlar tarafından etiketlendiklerinden modeller ile insanların sınıflandırması farklılık göstermektedir. Modellerden bazıları insan gözünün algıladığı doğruluğu aşabilmektedir. Bu durum yanlış sınıflandırılmaların anlaşılmasını zorlaştırmaktadır.

Ayrıca, duygu analizinin yoruma açık olması sebebiyle verilen görselin birden fazla sınıfla eşleştirilmesi olasıdır. Figure 11'da görsellerin farklı sınıflar ile eşleştirilmesi gösterilmiştir.

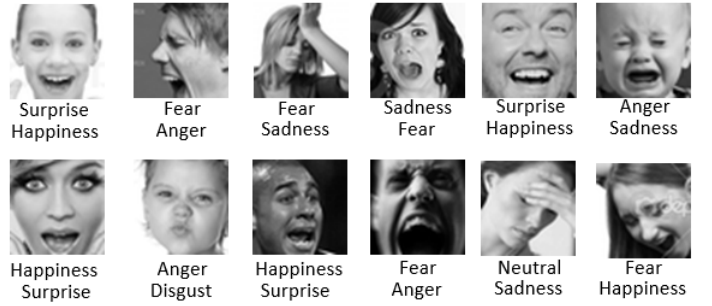


Fig. 11. FER2013 Görsellerinin Olası Farklı Etiketleri

VIII. SONUÇ

Bu projeye başlarkenki amacımız farklı duygu durum tahmininde hangi modellerin nasıl kullanıldığını anlamak ve bu modellerin FER2013 veri seti üzerindeki başarılarını ölçmektir. Bu amaç doğrultusunda çalışma süresince 3 farklı model kullandık. Bu modeller sırasıyla kendi eğittiğimiz CNN modeli ile önceden eğitilmiş ResNet50 ve VGG16 modelleri oldu. Kullanılan modeller içerisinde en yüksek başarı %67 ile CNN modelinde gözlemlendi. Ayrıca bu modeller arasından başarısı diğerlerine göre daha yüksek olduğundan CNN modeli seçilerek gerçek zamanlı olarak nasıl çalıştığı incelendi.

IX. KOD

Proje kapsamında yapılan çalışmaların kodlarına aşağıdaki GitHub linkinden erişilebilir.

<https://github.com/FidanVural/Facial-Emotion-Recognition>

REFERENCES

- [1] Amil Khanzada, Charles Bai, Ferhat Turker Celepcikay, "Facial Expression Recognition with Deep Learning", Stanford University, 2019.
- [2] International Conference on Computational Science, ICCS 2017, 12-14 June 2017, Zurich, Switzerland
- [3] <https://towardsdatascience.com/emotion-recognition-4fba48dabb6e>
- [4] I. J. Goodfellow, D. Erhan, P. L. Carrier, A. Courville, M. Mirza, B. Hamner, W. Cukierski, Y. Tang, D. Thaler, D.-H. Lee et al., "Challenges in representation learning: A report on three machine learning contests," in International Conference on Neural Information Processing. Springer, 2013, pp. 117–124.
- [5] S. Li and W. Deng, "Deep facial expression recognition: A survey," arXiv preprint arXiv:1804.08348, 2018.
- [6] Pramerdorfer, C., Kampel, M.: Facial expression recognition using convolutional neural networks: state of the art. Preprint arXiv:1612.02903v1, 2016.
- [7] Z. Zhang, P. Luo, C.-C. Loy, and X. Tang, "Learning Social Relation Traits from Face Images," in Proc. IEEE Int. Conference on Computer Vision (ICCV), 2015, pp. 3631–3639.
- [8] M. J. Paul Viola. "Rapid object detection using a boosted cascade of simple features." (2001), [Online]. Available: <https://www.cs.cmu.edu/efros/courses/LBMV07/Papers/viola-cvpr-01.pdf> (visited on 08/08/2001).