

## Evrişimli Sinir Ağları Kullanılarak Duygu Durum Tespitinin İyileştirilmesi ve Eğitim Verimliliğinin Analizi

Berkay Çakmak<sup>\*,1</sup>, İbrahim Develi<sup>2</sup>

<sup>\*,1,2</sup>Erciyes Üniversitesi, Mühendislik Fakültesi, Elektrik-Elektronik Mühendisliği, KAYSERİ

(Alınış / Received: 12.02.2023, Kabul / Accepted: 10.04.2023, Online Yayınlanma / Published Online: 02.05.2023)

### Anahtar Kelimeler

Evrişimli Sinir Ağları,  
Yüz duygu ifadeleri  
tanımlama,  
FER2013 veri seti,  
Eğitim verimliliği

**Öz:** Bu çalışmada, yüzdeki duygu ifadelerini tespit etmek için literatürdeki diğer modellerden daha yüksek doğruluk oranına sahip bir evrişimli sinir ağı modeli (CNN) önerilmiştir. Evrişimli sinir ağı modelini eğitmek için yedi duygu kategorisinde insan yüzleri içeren ve 30.000 imge'den oluşan FER2013 veri seti kullanılmıştır. Modelin eğitim doğruluğu %97,83 ve test doğruluğu %83,52 olarak elde edilmiştir. İnternet üzerinden yapılan eğitim ve sunumlarda; dinleyicilerin duygu durumları, geliştirilen CNN modeli ile gerçek zamanlı olarak tespit edilmekte ve tasarlanan algoritma ile eğitim süresince ve eğitimin sonunda katılımcıların duygu yoğunlukları sunucuya zaman bazlı olarak rapor halinde sunulmasını sağlayan bir algoritma geliştirilmiştir. Sunulan rapor sayesinde dinleyicilerin zamana göre duygu durumları analiz edilerek eğitim verimliliği artırılmaktadır.

## Improving Facial Expression Detection Using Convolutional Neural Networks and Analysis of Education Efficiency

### Keywords

Convolutional Neural  
Networks,  
Identifying facial expressions  
of emotions,  
FER2013 dataset,  
Online education efficiency

**Abstract:** In this paper, a convolutional neural network model (CNN) with higher accuracy than other models in the literature is proposed to detect facial emotional expressions. To train the convolutional neural network model, the FER2013 dataset consisting of 30,000 images and human faces in seven emotion categories was used. The training accuracy of the model was 97.83% and the test accuracy was 83.52%. In training and presentations made over the Internet; The emotional states of the listeners are detected in real time with the developed CNN model, and an algorithm has been developed that allows the emotional intensity of the participants to be reported to the presenter on a time basis during the training and at the end of the training with the designed algorithm. Thanks to the presented report, the emotional states of the listeners are analyzed according to time, thereby increasing the educational efficiency.

\*İlgili Yazar, email: brkayckmak@gmail.com

### 1. Giriş

Duygu analizi, insan duygularını görüntü işleme yöntemleri ile tanımayı hedefleyen geniş ve önemli bir araştırma alanıdır. Bir iletişim sırasındaki yüz değişiklikleri, duygusal durumu ileten ilk işaretlerdir [1]. Pek çok sözel olmayan bileşen arasında yüz ifadeleri, duygusal anlam taşımamasından dolayı kişilerarası iletişimdeki temel bilgi kanallarından biridir. Yirminci yüzyıldan beri insanoğlu, içinde büyüdüğü kültürden bağımsız olarak öfke, korku, mutlu, üzgün, küçümseme, iğrenme ve şaşkınlık olmak üzere yedi temel duygu ile duygularını tanımlanmıştır [2].

Geleneksel yüz ifadelerinden duygu analizi yaklaşımları; yüz ve yüz bileşeni algılama, özellik çıkarma ve ifade sınıflandırma olmak üzere üç ana adımdan oluşmaktadır. İlk olarak, bir giriş görüntüsünden bir yüz görüntüsü algılanmaktadır ve yüz bölgesinden yüz bileşenleri (örneğin, gözler ve burun) veya yer işaretleri algılanmaktadır. İkinci olarak, yüz bileşenlerinden çeşitli uzamsal ve zamansal özellikler çıkarılmaktadır. Üçüncüsü, derin öğrenme algoritmaları ile çıkarılan öznitelikleri kullanarak tanıma sonuçları üretilmektedir [3].

CNN, görsel görüntüleri analiz etmek için yaygın olarak uygulanan bir derin sinir ağı sınıfıdır. CNN, ileri beslemeli bir sinir ağıdır ve çok katmanlı algılayıcılar (MLP'ler) olarak da adlandırılmaktadır. CNN'in ana çalışma alanı görüntü tanıma ve görüntü sınıflandırmadır. Evrişim katmanı, bir giriş görüntüsünün özelliklerini eğitim veri kümesinden çıkarmaktadır. CNN modeli; evrişim, birleştirme, düzleştirme ve yoğun katmanlardan oluşmaktadır. Büyük veri kümelerinde daha iyi doğruluk için evrişim katmanları eklenmektedir [4].

Bu çalışmada, görüntü işleme teknolojisi kullanılarak insanların duygularını yüz ifadeleri vasıtasıyla tespit edebilen bir CNN modeli tasarlanmıştır [5]. Tasarlanan CNN modeli, literatürdeki mevcut FER2013 veri seti ile eğitilmiş diğer algoritmalarından daha verimli bir şekilde çalışmaktadır. İnternet üzerinden gerçekleştirilen canlı eğitimlerde gerçek zamanlı olarak elde edilen duygu durumları ile eğitim alan kişilerin zamana göre yoğunluklu duygu durumları analiz edilmektedir. Analiz sonucunda eğitime anlık duygu yoğunluğundaki değişimin bildirimi ve eğitim sonunda rapor verilerek eğitim verimliliğinin artırılmasına yönelik bir çalışma gerçekleştirilmiştir.

Bu çalışmada, literatürde bulunan FER2013 veri seti ile eğitilmiş duygu tanıma algoritmalarına göre daha yüksek doğruluk oranına sahip bir makine öğrenmesi modeli eğitilmiştir. Ayrıca çalışma, eğitilen modelin kullanım amacı ile de literatürde bulunan diğer çalışmalardan yüksek oranda ayrılmaktadır. Yüksek doğruluk oranına sahip CNN modeli ve gerçek zamanlı veri işlemeye dayalı raporlar veren algoritma ile literatürdeki çalışmalardan ayrılan bir çalışma gerçekleştirilmiştir.

## 2. Literatür Değerlendirmesi

Yüz ifadeleri, duyguları belirtmek için tüm insanlar için ortak sinyaldir. Robotik, tıp, sürüş destek sistemleri gibi birçok alanda uygulama alanı mevcut olduğu için, görüntü işleme ile yüz ifadesi analiz araçları yapmak adına birçok girişimde bulunmaktadır [6].

Literatürde, FER2013 veri seti kullanılarak duygu analizi gerçekleştiren çalışmalar mevcuttur. [7]'de FER2013 veri seti ile eğitilen Xception algoritmasının performansı %61,7 olarak elde edilmiş, ardından yüz görüntüsü harmanlama işlemi ile doğruluk %63'e çıkarılmıştır. [8]'de yapılan çalışmada duygu analizi için bir CNN modeli tasarlanmış ve %57,1 doğruluğa ulaşılmıştır. Başka bir duygu analizi çalışmasında ise FER2013 veri seti ile mini Xception mimarisi eğitilerek %65,97 doğruluğa ulaşılmıştır. Bu çalışmada gerçek zamanlı CNN modeli doğruluğunu da göstermişlerdir [9]. [10]'da hazırlanan 11 katmanlı CNN modeli, FER2013 veri seti ile 106 tekrar boyunca eğitilerek %70 doğruluk oranına ulaşmayı başarmıştır. Bu çalışmada ayrıca yüz ifadelerin hangi ifadeye yüzdesel bağlamda ne kadar benzediği çalışması da yapılmıştır. [11]'de birkaç CNN modeli, önceden eğitilmiş model ve eğitim prosedürleri ile ilgili çalışmalar yapılmış, yapılan çalışmalar karşılaştırılmıştır. [11]'de bir giriş görüntüsü için en yakın eğitim örneklerini seçmek amacıyla bir k-en yakın komşu modeli uygulanmış, sonrasında seçilen eğitim örnekleri üzerinde destek vektör makineleri (SVM) sınıflandırıcısı eğitilmiş ve son olarak SVM sınıflandırıcısı, eğitildiği test görüntüsü için sınıf etiketini tahmin etmek için kullanılmıştır. Bu yöntem ile FER2013 veri seti üzerinde %75,42 doğruluğa ulaşılmıştır. Başka bir çalışmada ise 11 katmanlı bir CNN modeli tasarlanmış olup ölçek değişmeyen özellik dönüşümü (SIFT) metodu bu modele uygulanmıştır. SIFT metodunun amacı düşük hacimli veri kümesi ile yüksek doğruluk yakalamaktır ancak doğruluk oranı %73,4 olarak belirlenmiştir [12]. Bahsedilen çalışmaların başarımları Tablo 1'de özetlenmiştir.

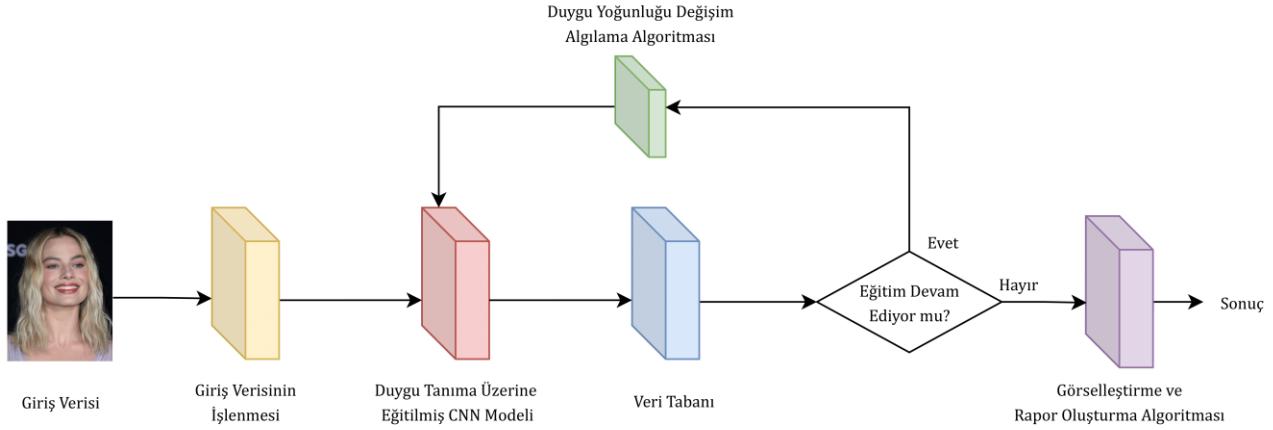
**Tablo 1.** Literatürdeki çalışmalar ve önerilen yaklaşımın doğruluk başarımları.

Literatürdeki Çalışmalar	Algoritma	Doğruluk Başarımı
[7]	Xception (CNN)	63,00%
[8]	CNN	57,10%
[9]	mini Xception (CNN)	65,97%
[10]	11 katmanlı CNN	70,00%
[11]	KNN + SVM	75,42%
[12]	CNN + SIFT	73,40%
Önerilen	CNN	97,83%

Yapılan başka bir çalışmada görüntü işleme algoritmaları ile eğitimlerdeki öğrencilerin duygularının analizi eğitim sonunda tek seferlik bir görsel alınarak analiz edilmiştir [13]. Yaptığımız çalışma, dinleyicilerin duygularını her saniyede 4 defa analiz ederek anlatıcıya anlık duygu değişimi bildirimleri, eğitim sonunda zamana bağlı duygu değişim grafiği sunması ve belirtildiği şekilde yüksek doğruluk oranı ile [13]'den ayrılmaktadır.

### 3. Materyal ve Metot

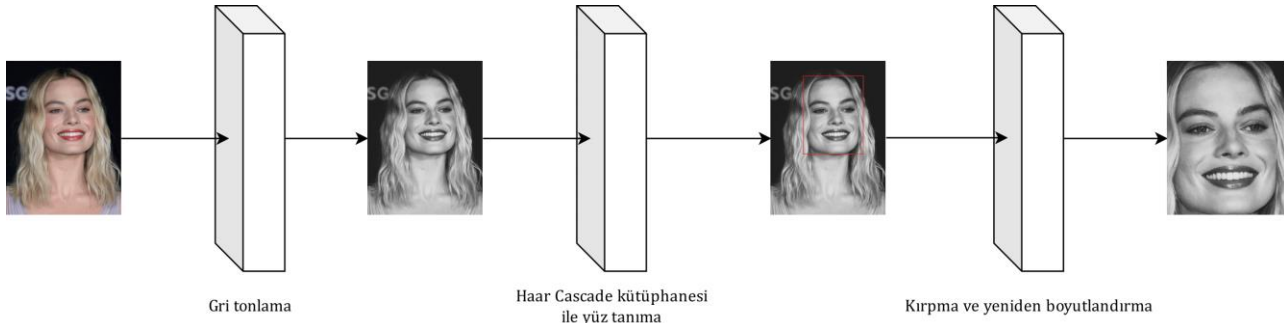
Giriş verisi olarak dinleyicilerin yüz görsellerini zamana bağlı her saniyede 4 defa alan ve duyguların zamana bağlı grafiğini oluşturan bir algoritma tasarlanmıştır. Algoritma temel olarak 4 ana bölümden oluşmaktadır. Bu bölümler sırasıyla; giriş görüntüsünü işleme, duygu tanıma, duygu verilerini veri tabanına aktarma ve son olarak verilerin görselleştirilmesi ve rapor oluşturma olarak tanımlanabilmektedir. Şekil 1'de algoritma diyagramı görülmektedir.



Şekil 1. Algoritma diyagramı.

#### 3.1. Giriş görüntüsünün işlenmesi

Bu çalışmada giriş görüntüsü, 3 adımdan oluşan bir ön işleme aşamasından geçmektedir. İlk adımda kırmızı, yeşil ve mavi olmak üzere 3 boyuta sahip olan giriş görüntüsü gri tonlama işleminden geçirilerek tek boyuta indirgenmektedir. Bu işlemin ardından ikinci adımda yüz bölgesi haricinde kalan piksellerin geliştirilen CNN modelinin tahmin başarısını etkilemesi ihtimalini ortadan kaldırmak amacıyla yüzün olduğu bölge Haar Cascade kütüphanesi kullanılarak tespit edilmektedir [14]. Üçüncü adımda ise giriş görüntüsünden yüzün olduğu bölge kırılmakta ve 48x48 boyutuna dönüştürülmektedir. Şekil 2'de görüntü ön işleme aşamaları görselleştirilerek sunulmuştur. Bu işlemler sonucunda elde edilen görüntü, eğitilmiş olan CNN modeli ile işlenerek duygu tespiti yapılmaktadır.

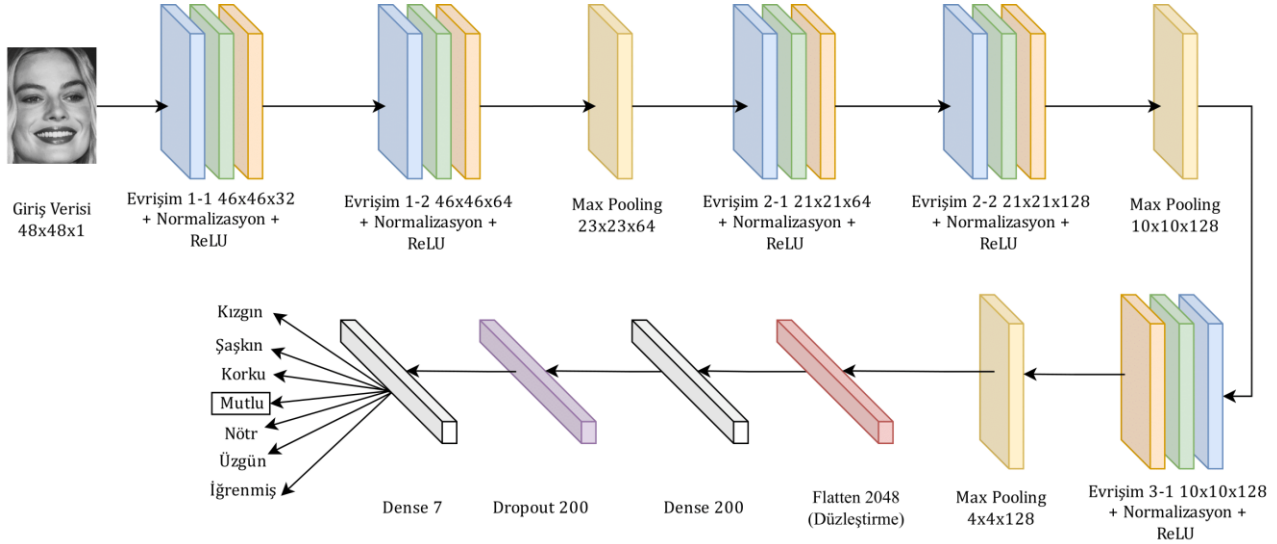


Şekil 2. Giriş görüntüsünün ön işleme aşamaları.

#### 3.2. Duygu tanıma üzerine eğitilen CNN modeli

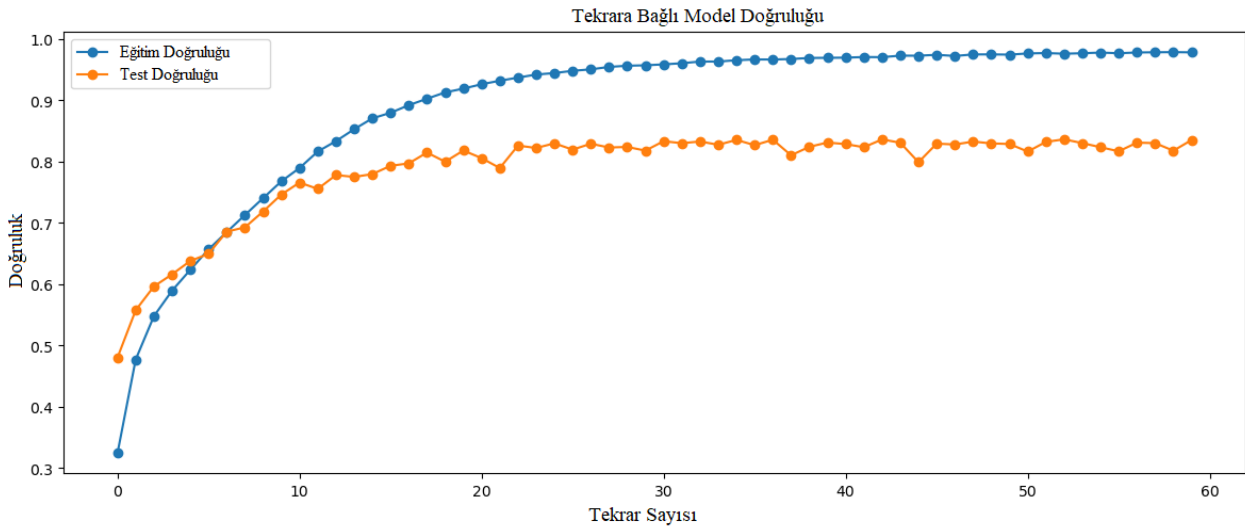
İlk katman 48x48 boyutlarında bir giriş görselinden oluşmaktadır. Bu katmanı 32 özellik haritasından oluşan evrişim, normalizasyon ve aktivasyon (ReLU) katmanlarından oluşan bir blok takip etmektedir. Bu bloğun ardından 64 özellik haritasından oluşan, normalizasyon ve aktivasyon katmanları bulunduran başka bir blok ile model yapısı devam etmektedir. Bu bloğun devamında 2x2 boyutunda bir havuzlama (max pooling) katmanı

bulunmaktadır. Dolayısıyla ortaya çıkan görüntünün boyutu  $23 \times 23 \times 64$  olmaktadır. Benzer şekilde bu katmanın ardından 64 özellik haritasından oluşan evrşim, normalizasyon ve aktivasyon (ReLU) katmanlarından oluşan bir blok ve 128 özellik haritasından oluşan evrşim, normalizasyon ve aktivasyon (ReLU) katmanlarından oluşan bir blok ve bu blokları takip eden başka bir  $2 \times 2$  boyutunda havuzlama (max pooling) katmanı bulunmaktadır. Bu katmanın ardından ortaya çıkacak görüntünün boyutu  $10 \times 10 \times 128$  olmaktadır. Bu katmanın ardından bir düzleştirme katmanı ile veri tek boyutlu hale getirilmektedir. Sonrasında 200 boyutlu bir dense ardından bir dropout ve son olarak 7 boyutlu bir dense katmanı ile model oluşturulmuştur. Model; kızgın, şaşkın, korku dolu, mutlu, nötr, üzgün ve iğrenmiş olmak üzere 7 farklı duygunun tahmini gerçekleştirilebilmektedir. Şekil 3'te CNN modelinin mimarisi görülmektedir [6].



Şekil 3. Oluşturulan CNN modelinin mimarisi.

FER2013 veri setinin %80'i eğitim, %20'si ise test için kullanılmıştır. Oluşturulan CNN modeli eğitim ve test için ayrılaştırılan FER2013 veri seti ile 60 tekrar boyunca eğitilmiş ve eğitim veri setinde %97,83, test veri setinde %83,52 doğruluğa ulaşmıştır. Şekil 4'te tekrara bağlı model doğruluk grafiği mevcuttur. Şekil 5'te CNN modelinin eğitim veri setindeki hata matrisi görülmektedir. Hata matrisindeki bölümlerde görüntü sayıları yoğunlaştıkça matristeki bölümün rengi koyulaşmaktadır. Hata matrisi incelendiğinde, eğitilmiş olan CNN modeline giriş olarak verilen görüntülerin büyük bölümünün doğru tahmin edildiği görülmektedir.

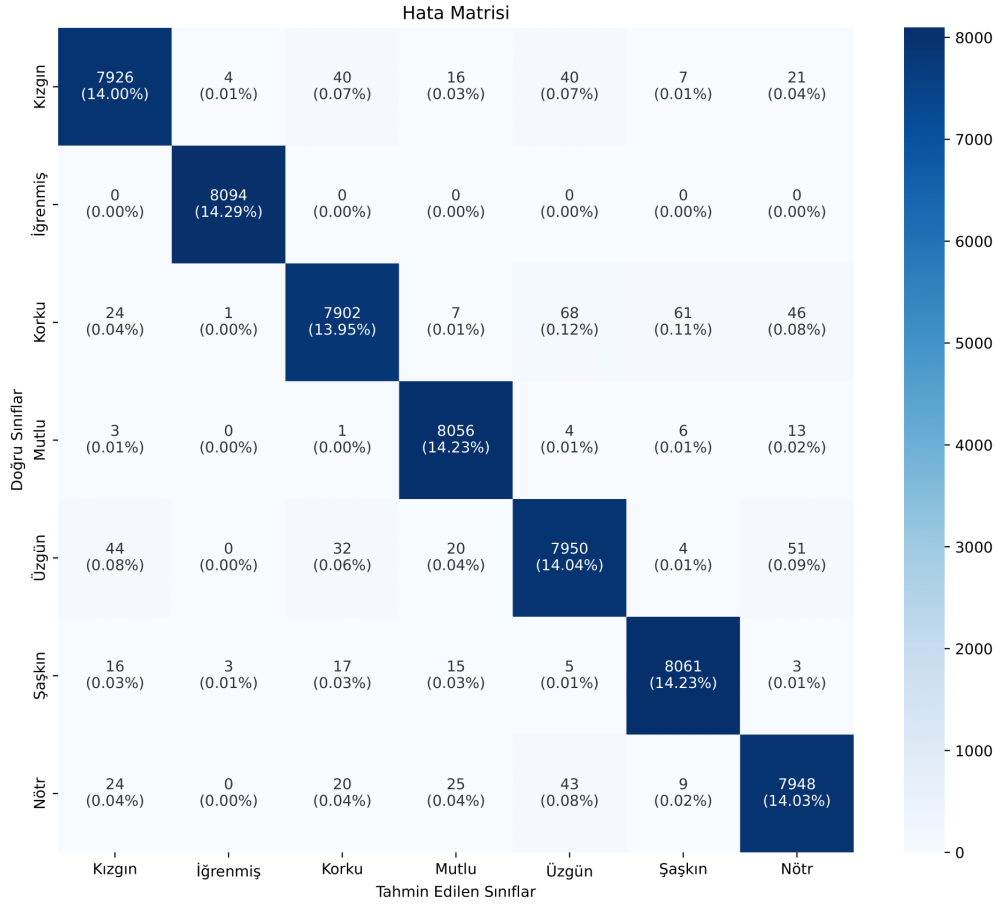


Şekil 4. Tekrara bağlı model doğruluk grafiği.

### 3.3. Elde edilen duygu verilerini veri tabanına aktarmak

Kurulan algoritma dinleyicilerin yüzlerini analiz ederek her saniyede 4 defa duygu tahmini yapmaktadır. CNN modelinden çıkan veriler, eğitim devam ettiği süre boyunca, her bir dinleyici için saniyede 4 duygu verisi MsSQL veri tabanına Python bağlantısı yardımı ile aktarılarak görselleştirilmek ve raporlanmak üzere saklanmaktadır. Veriler eğitim bitiminde raporlandıktan sonra silinmektedir. Anlık olarak gelen veriler veri tabanı bağlantısı ile

eş zamanlı olarak başka bir algoritma tarafından işlenerek duygu yoğunluğundaki belli değişimler anlatıcıya ders devam ederken bildirim şeklinde sunulmaktadır.



Şekil 5. CNN modelinin eğitim verisindeki hata matrisi

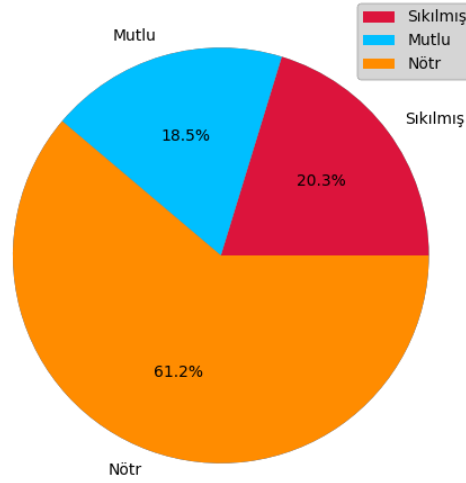
### 3.4. Verilerin görselleştirilmesi ve rapor oluşturmak

Eğitim sona erdiğinde anlatıcının analiz etmesi için bir rapor oluşturulmaktadır. Eğitim esnasında kızgın, şaşkın, korku dolu ve iğrenme duyguları nadir gerçekleşeceği için bu duygular raporda analiz edilmemektedir. Rapor içerisinde yazılı olarak eğitim süresi ve tespit edilen duyguların eğitim boyunca olan toplam duygulara oranı olan yüzdesel veriler yer almaktadır. Duyguların oranı verisinin daire grafiği olarak görselleştirilmiş hali de eğitime rapor üzerinde sunulmaktadır. Eğitmenin dinleyicilerin duygularını zamana bağlı analiz edebilmesi için çizgi grafik olarak bir duygu/yoğunluk grafiği oluşturulmuştur.

Eğitim sonunda eğitime gönderilen raporda; derste kaç öğrenci bulunduğu, öğrencilerin ders süresi boyunca sürenin yüzde kaçında sıkıldığı, mutlu olduğu ve nötr kaldığı ve derste belirli zaman aralıklarında hangi duygu durumunun yoğun olduğu bilgisi yazılı ve görsel olarak verilmektedir.

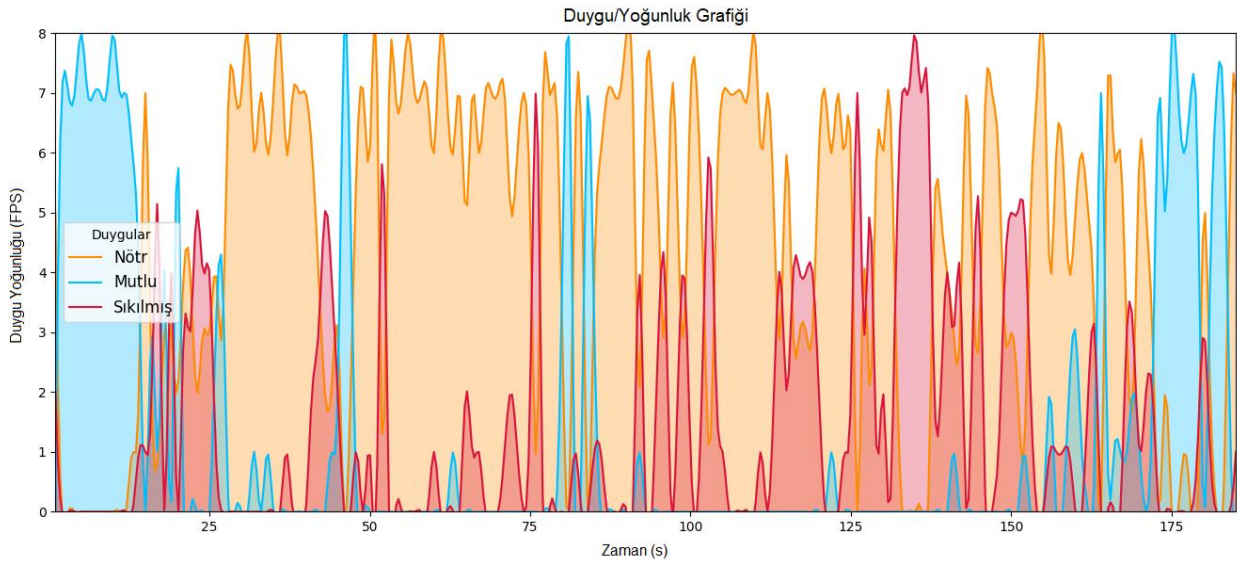
## 4. Bulgular

2 dinleyiciden oluşan ve 3 dakika süren bir ders ortamı oluşturulmuş ve oluşturulan algoritma bu ortamda test edilmiştir. Deney süresince yoğun duygu durum değişimlerinde anlatıcıya bilgi verilmiş, deney sonucunda ise anlatıcıya deney süresince sınıftaki öğrencilerin duygu durumlarını içeren bir rapor gönderilmiştir. Rapor içerisinde ham veriyi anlaşılır hale getirip bu veriden faydalı bilgi elde etmek amacıyla sınıftaki duygu durumlarının yüzdesi, bu verinin daire grafiği ile görselleştirilmesi ve zamana bağlı duygu durumlarının çizgi grafiği bulunmaktadır. Şekil 6'da deney sonucunda raporlanan daire grafiği görülmektedir.



Şekil 6. Eğitim sonucunda raporlanan daire grafiği.

Raporlanan daire grafiğinin açıklaması olarak anlatıcıya rapor üzerinde, “Dinleyiciler; 186 saniye boyunca %61 oranında nötr, %18 oranında mutlu, %20 oranında sıkılmıştır.” bilgisi yazılı olarak verilmektedir. Şekil 7’de deney eğitimi sonucunda raporlanan duygu yoğunluk grafiği görülmektedir. Grafikte X eksenini kamera vasıtasıyla her saniye elde edilen 4 görüntünün işlenmesi ile elde edilen duygu sayısını yani saniye başına düşen kareyi (FPS), Y eksenini ise zamanı saniye türünden ifade etmektedir.



Şekil 7. Eğitim sonucunda raporlanan duygu/yoğunluk grafiği

186 saniye süren deney eğitimi sonucunda çıkan veriler üzerinden dinleyicilerin ders başında mutlu oldukları; 25, 50 ve 100. saniyeler civarlarında sıkıldıklarını ayrıca 130 ile 155. saniyeler arasında da sıkıldıklarını, genel olarak nötr kaldıklarını ve dersin bitiminde mutlu oldukları çıkarılabilmektedir. Anlatıcı 130. saniyede başlayan yoğun ve kalıcı olan duygu değişimini nedeniyle bu değişimin bildirimini “Dinleyicilerin duyguları yoğunluklu olarak ‘Sıkılmış’tır.” şeklinde canlı olarak almıştır. 175. saniyede ise duygu yoğunluğu değişim bildirimini “Dinleyicilerin duyguları yoğunluklu olarak ‘Mutlu’dur.” şeklinde canlı olarak vermiştir.

## 5. Tartışma ve Sonuç

Bu çalışmada yedi farklı duyguyu gerçek zamanlı olarak tespit edebilmek için yüksek sınıflandırma doğruluğuna sahip bir CNN modeli hazırlanmıştır. Hazırlanan CNN modeli, çevrimiçi eğitim platformlarında katılımcıların duygu durumlarının analizi için tasarlanan algoritmanın içerisine entegre edilmiştir. Tasarlanan sistem ile anlatıcı, katılımcıların duygu durumlarını ders esnasında canlı olarak takip edebilmekte, ders sonunda gönderilen raporda bulunan faydalı bilgiler ve grafikler ile eğitimlerinin verimliliğini, kalitesini ve performansını olumlu olarak etkileyecek aksiyonlar alabilmektedir. Bu çalışma için yapılan, 2 öğrenciden oluşan ve 186 saniye süren deney eğitiminden çıkarılan sonuçlar ile çalışmanın başarısı analiz edilebilmektedir.

Yapılan çalışma, CNN modelinin doğruluk başarısının yüksekliği ve zaman bazlı olarak veri işleme ile literatürdeki diğer çalışmalardan ayrılmaktadır. Literatürdeki diğer çalışmalar incelendiğinde FER2013 veri seti ile eğitilmiş derin sinir ağları arasında %97,83 eğitim başarısı ve %83,52 test başarısına ulaşabilen bir çalışma bulunmamaktadır. Model doğruluğunun yüksekliği, tahmin edilen duyguların doğruluğu ile doğrudan bağlantılı olduğu için model doğruluğu derin sinir ağları ile duygu tespiti çalışmalarında önemli rol oynamaktadır. Literatürde bulunan derin sinir ağları ile duygu analizinin eğitim verimliliğinin yükseltilmesinde kullanılmasına yönelik diğer çalışmalarda öğrencilerin duygularının zamana bağlı olarak işlenmemesinden kaynaklanan analiz problemleri mevcuttur. Sadece ders sonunda alınan görüntüden duygu analizinin yapılması ders boyunca öğrencilerin hangi duygular içerisinde olduğu verisini tutmaması nedeniyle çalışmamıza göre büyük boyutta bir faydalı veri sunmamaktadır. Çalışmamızda hazırlanan algoritma ile her bir öğrenci için saniyede 4 duygu tespiti yapıldığı için hem ders süresi içerisinde anlatıcıya bildirim verilerek ders bitmeden verimlilik artırılmakta, hem de ders sonunda paylaşılan detaylı rapor sayesinde öğretmenin gelecek derslerinin verimliliği artırılmaktadır.

## Kaynakça

- [1] Mellouk, W., Wahida H. 2020. Facial emotion recognition using deep learning: review and insights. *Procedia Computer Science* 175 (2020): 689-694.
- [2] Ekman, P., Wallace V. F. 1971. Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology* 17.2 (1971): 124.
- [3] Ko, B. C. 2018. A brief review of facial emotion recognition based on visual information. *Sensors* 18.2 (2018): 401.
- [4] Altekin, F., Demir, H. 2021. Emotion Detection from Facial Expression Using Different Feature Descriptor Methods with Convolutional Neural Networks. *European Journal of Engineering and Applied Sciences* 4.1 (2021): 14-17.
- [5] Cakmak, B., Develi, I. 2023. Convolutional Neural Network-Based Classification of Facial Emotional Expressions and Computational Complexity Analysis. *International Conference on Frontiers in Academic Research*. Vol. 1. (2023) 168-173.
- [6] Mehendale, N. 2020. Facial emotion recognition using convolutional neural networks (FERC). *SN Applied Sciences* 2.3 (2020): 1-8.
- [7] Kim, J., Hwan, A. P., Dong, S. H. 2021. The extensive usage of the facial image threshing machine for facial emotion recognition performance. *Sensors* 21.6 (2021): 2026.
- [8] Tümen, V., Söylemez, Ö. F., Ergen, B. 2017. Facial emotion recognition on a dataset using convolutional neural network. *2017 International Artificial Intelligence and Data Processing Symposium (IDAP)*. IEEE, (2017).
- [9] Zahara, L., et al. 2020. The facial emotion recognition (FER-2013) dataset for prediction system of micro-expressions face using the convolutional neural network (CNN) algorithm based Raspberry Pi. *2020 Fifth international conference on informatics and computing (ICIC)*. IEEE, (2020).
- [10] Lasri, I., Solh, A. R., El Belkacemi, M. 2019. Facial emotion recognition of students using convolutional neural network. *2019 third international conference on intelligent computing in data sciences (ICDS)*. IEEE, (2019).
- [11] Georgescu, M. I., Ionescu, R. T., Popescu, M. 2019. Local learning with deep and handcrafted features for facial expression recognition. *IEEE Access* 7 (2019): 64827-64836.
- [12] Connie, T., et al. 2017. Facial expression recognition using a hybrid CNN-SIFT aggregator. *International Workshop on Multi-disciplinary Trends in Artificial Intelligence*. Springer, Cham, (2017).
- [13] Wang, W., et al. 2020. Emotion recognition of students based on facial expressions in online education based on the perspective of computer simulation. *Complexity* (2020).
- [14] Viola, P., Jones M. 2001. Rapid object detection using a boosted cascade of simple features. *Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition*. CVPR (2001). Vol. 1. IEEE, 2001.