

# **Ham Piksel Tabanlı Yüz İfadesi Tanıma:**

## **CNN ile 8 Sınıflı Duygu Sınıflandırması**

**Konvolüsyonel Sinir Ağları Mimarisi Kullanarak FER2013 Veri Seti  
Üzerinde Duygusal İfade Tanımı**

**Öğrenci Adı: YASİR ÖMER ALPARSLAN**  
**Öğrenci no: 244225708**

**Danışman: Öğretim Üyesi Dr. SELİM YILMAZ**

**MUĞLA ÜNİVERSİTESİ / FEN BİLİMLERİ ENSTİTÜSÜ / Bölüm: YZ**

**| Tarih: 21-12-2025**



# Motivasyon ve Problem Tanımı



## İnsan-Bilgisayar Etkileşimi

İnsan-bilgisayar arayüzlerinde yüz ifadesi tanımı kullanıcı deneyimini iyileştirmektedir. HCI



## Sağlık ve Psikiyatri

Duygusal bozuklıkların teşhisinde ve terapötik müdahalelerin izlenmesinde önem taşıyor.



## Güvenlik Sistemleri

Biyometrik kimlik doğrulama ve davranış analizi uygulamalarında kritik rol oynuyor.



## Klasik Yöntemlerin Sınırlamaları

HOG, LBP ve SIFT gibi elde yazılan özellikler + SVM, taşınabilir olmayan özellikler çıkarma gerektirir.

# FER2013 Veri Seti



**48×48 piksellik gri tonlamalı yüz görüntülerinden oluşan kapsamlı duygusal ifade veri seti. Veri seti eğitim, doğrulama ve test alt kümelerine ayrılmıştır. Sekiz duygusal sınıf tanımlanmış olup, sınıf dengesizliği sorunu dikkat çekici bir zorluktur.**

Özellik	Değer
Görüntü Çözünürlüğü	<b>48×48 piksel</b>
Görüntü Formatı	<b>Gri Tonlama (Grayscale)</b>
Toplam Görüntü Sayısı	<b>35.887</b>
Eğitim Seti	<b>28.709</b>
Test Seti	<b>7.178</b>
Sınıf Sayısı	<b>8 (Anger, Contempt, Disgust, Fear, Happiness, Neutral, Sadness, Surprise)</b>
Sınıf Dengesizliği	<b>Evet - Özellikle Disgust sınıfında düşük temsil</b>

# FER2013 Veri Seti

<https://www.kaggle.com/code/pedroadorighello/gradcam-fer2013-test/input>



# Ön İşleme (Preprocessing)

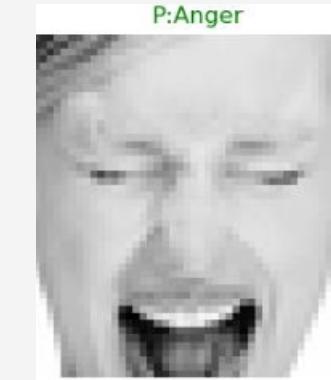
## Normalizasyon

Piksel değerleri 0-255 aralığından 0-1 aralığına normalize edilmiştir. Düşük çözünürlük görüntülerde bu işlem, Özellikle ince detayların korunması açısından kritik öneme sahiptir.

## One-Hot Kodlama

Sekiz sınıf one-hot vektörleri olarak kodlanmıştır (örneğin [0,0,1,0,0,0,0,0] Anger sınıfı temsil eder). Bu kodlama yöntemi, kategorik verilerin makine öğrenmesi modelleri tarafından işlenebilmesini sağlar.

Önce



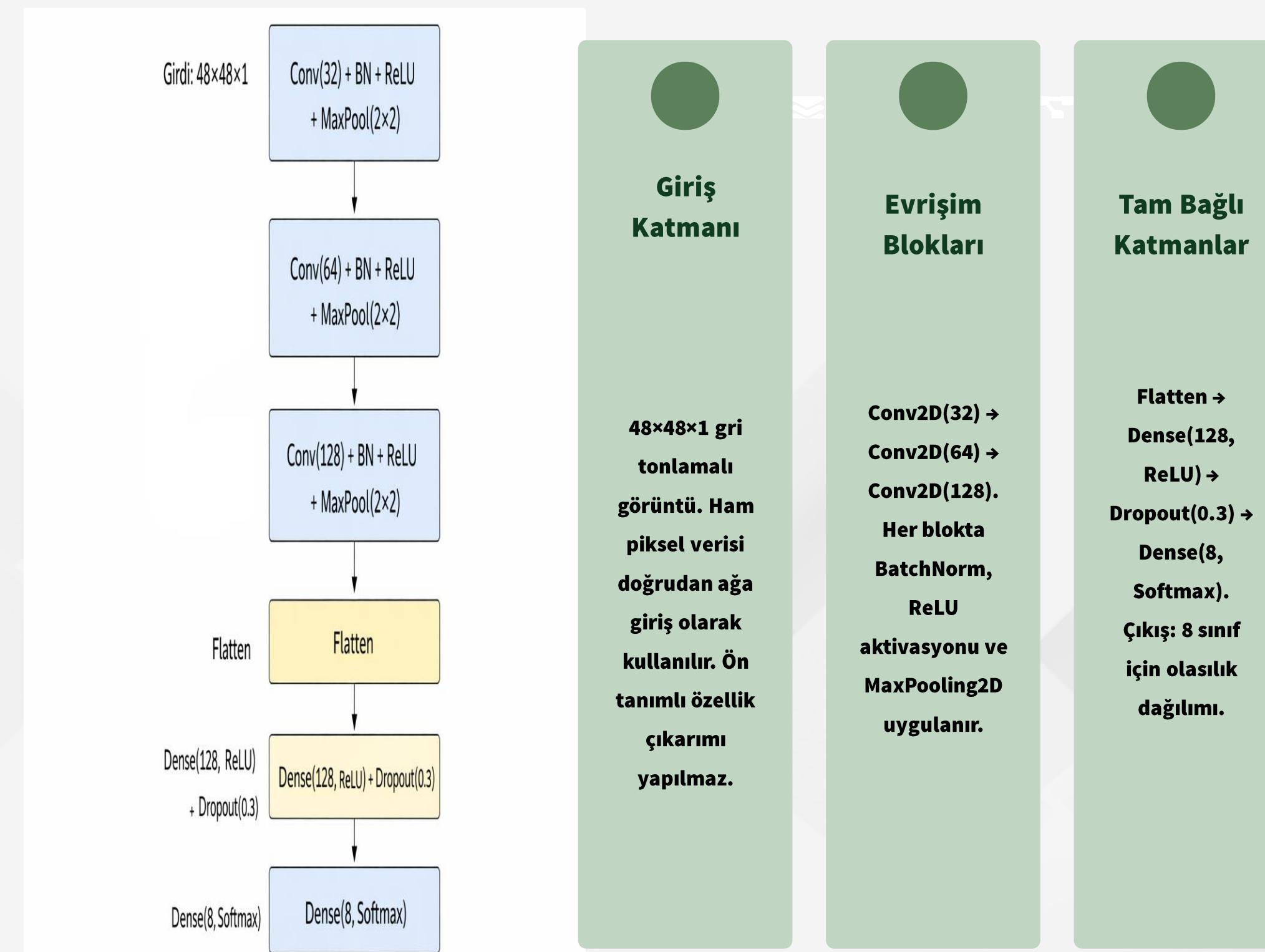
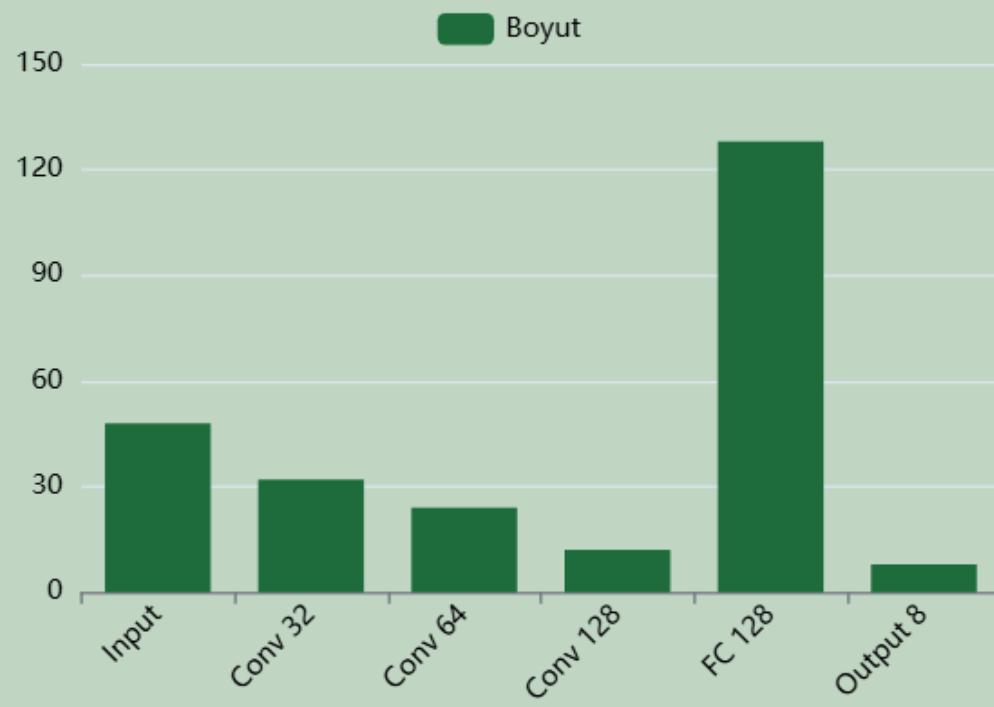
sonra

## Veri Bölümü

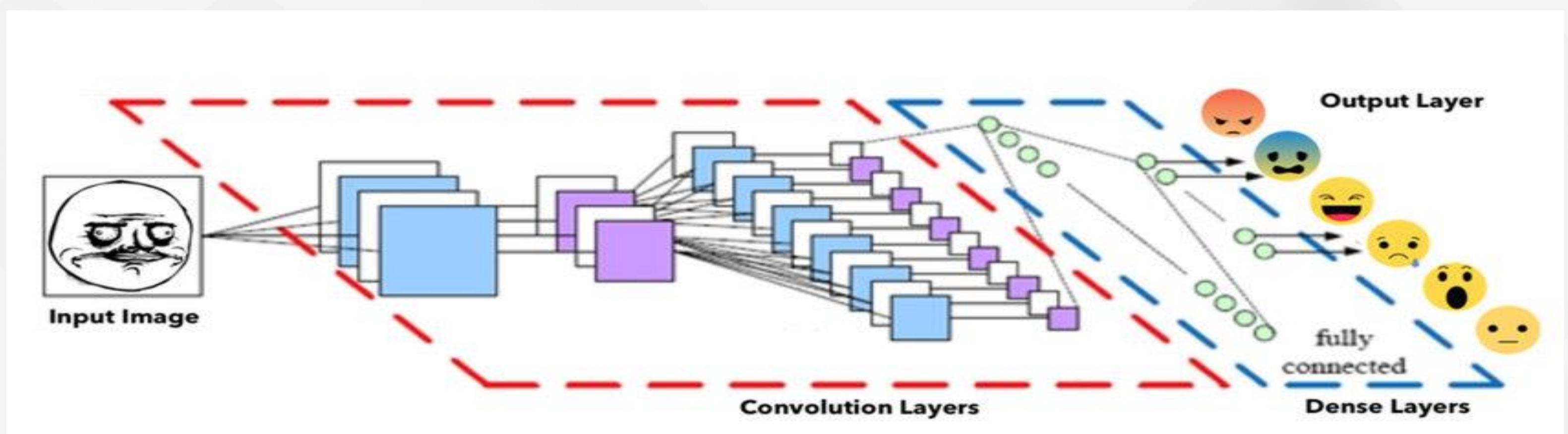
Eğitim seti %80, doğrulama seti %20 olarak ayrılmıştır. Düşük çözünürlüklü veri setlerinde bu oranların dikkatli belirlenmesi, modelin genelleme performansını önemli ölçüde etkilemektedir.

# CNN Mimarisi

Katmanlar Boyunca Özellik Haritaları



## CNN Mimarisi



# Matematiksel Temeller

## Convolution (Evrişim)

**Evrişim katmanı, I giriş görüntüsü ve K filtre/kernel olmak üzere,  $(I * K)(i,j) = \sum \sum I(i+m, j+n) \times K(m,n)$  formülüyle tanımlanır. Kenar ve kontrast gibi lokal öznitelikleri çıkarır.**

$$W(m, n) \cdot (n + m, j + X(i \sum_{m,n} = X)(i, j) * F)$$

## Softmax Function

$\sigma(z_i) = e^{z_i} / \sum_j e^{z_j}$ . Çıkış katmanında her sınıf için  $[0,1]$  aralığında olasılık üretir.

$$\frac{e^z}{\sum_{j=1}^k e^z} = (k|x = P(y$$

## Categorical Cross-Entropy

$L = -\sum_i y_i \times \log(\hat{y}_i)$ .  $y_i$  gerçek etiket,  $\hat{y}_i$  tahmin edilen olasılık.

$$\max(0, x) = \text{ReLU}(x)$$

$$(\hat{y}_k) y_k \log \sum_{k=1}^8 - = \mathcal{L}$$

# Karar Mekanizması: Softmax

Son tam bağlantılı katmanda her sınıf için bir skor üretilir:

Ardından Softmax uygulanır:

$$b + Wx = z$$

Model, en yüksek olasılığa sahip sınıfı tahmin olarak seçer.

$$\frac{e^{z_k}}{\sum_{j=1}^K e^{z_j}} = P(y=k|x)$$

# Eğitim Kurulumu

## Optimi zer

Adam  
(learning  
rate = 1e-3)

## Batch Size

64

## Epoch Sayısı

60 (veri seti  
üzerinde 60  
tam geçiş)

## Veri Artırma

Küçük  
rotasyonlar,  
yatay çevirme,  
parlaklık/kont  
rast  
değişiklikleri

## ModelCheck point

En iyi ağırlıklar  
kaydedilir

## ReduceLRO nPlateau

Doğrulama kaybı  
stabil kalırsa  
learning rate  
azaltılır

## Early Stopping

Aşırı uyum  
(overfitting)  
engellemek için  
erken durdurma  
mekanizması  
bulunur

## No Transfer Learning

Model sıfırdan eğitilmiştir (pre-trained ağ kullanılmamıştır)

# Eğitim vs Doğrulama Doğruluk Eğrisi

## Doğruluk Analizi

### Gözlemler:

Eğitim doğruluğu (train accuracy) **istikrarlı biçimde artıyor**

Doğrulama doğruluğu (validation accuracy) **artıyor fakat dalgalı**

### Akademik yorumlarım:

Model gerçekörntülerini öğrenmektedir.

Aşırı öğrenme (overfitting) **kritik seviyede değildir.**

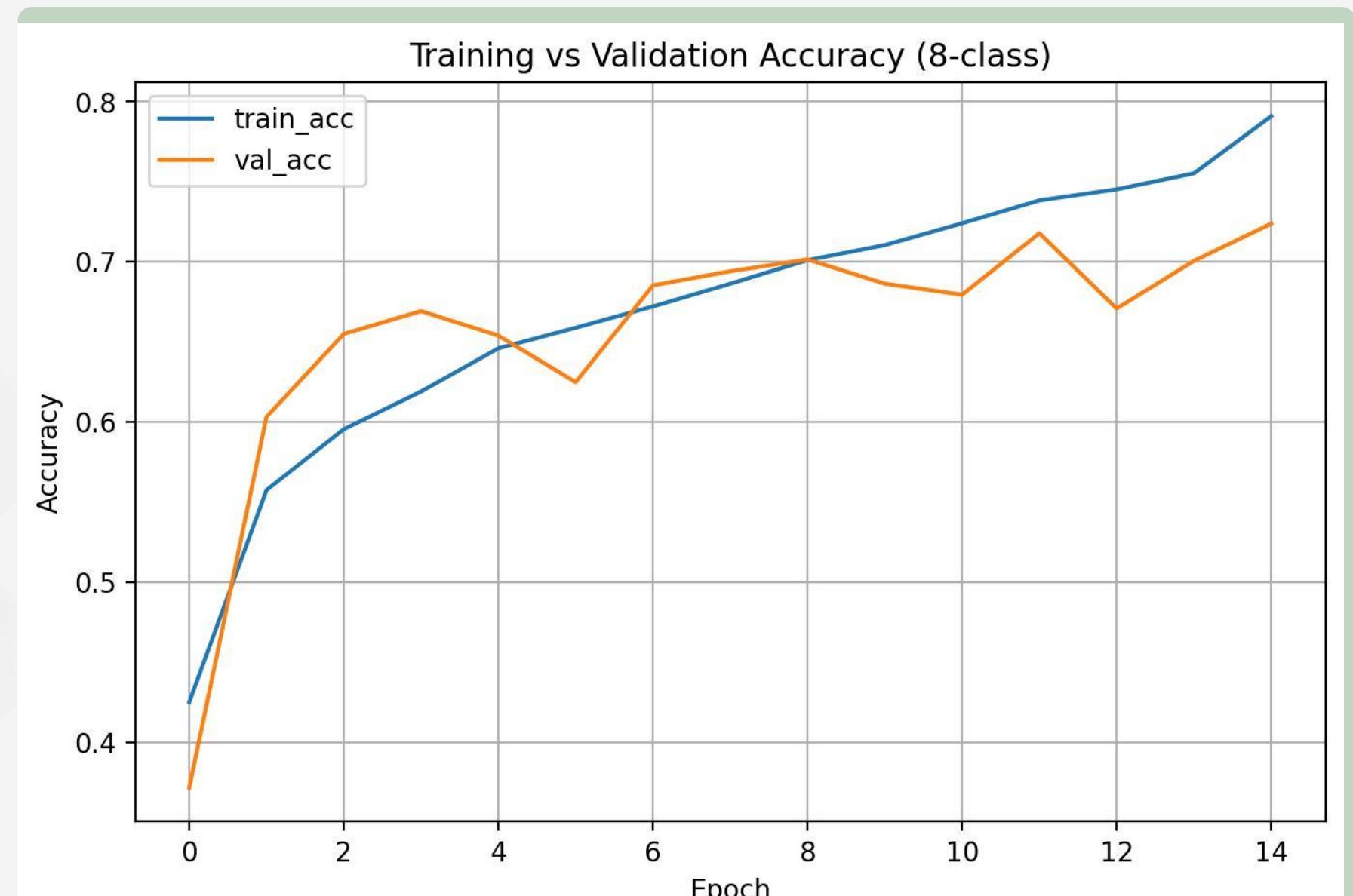
Dropout ve BatchNorm etkili çalışmaktadır.

➤ Epoch 15 civarında:

Train accuracy ≈ %79

Validation accuracy ≈ %72

Bu, modelin **öğrenmeye devam edebilecek potansiyele sahip olduğunu göster**



79%

Doğrulama Doğruluğu

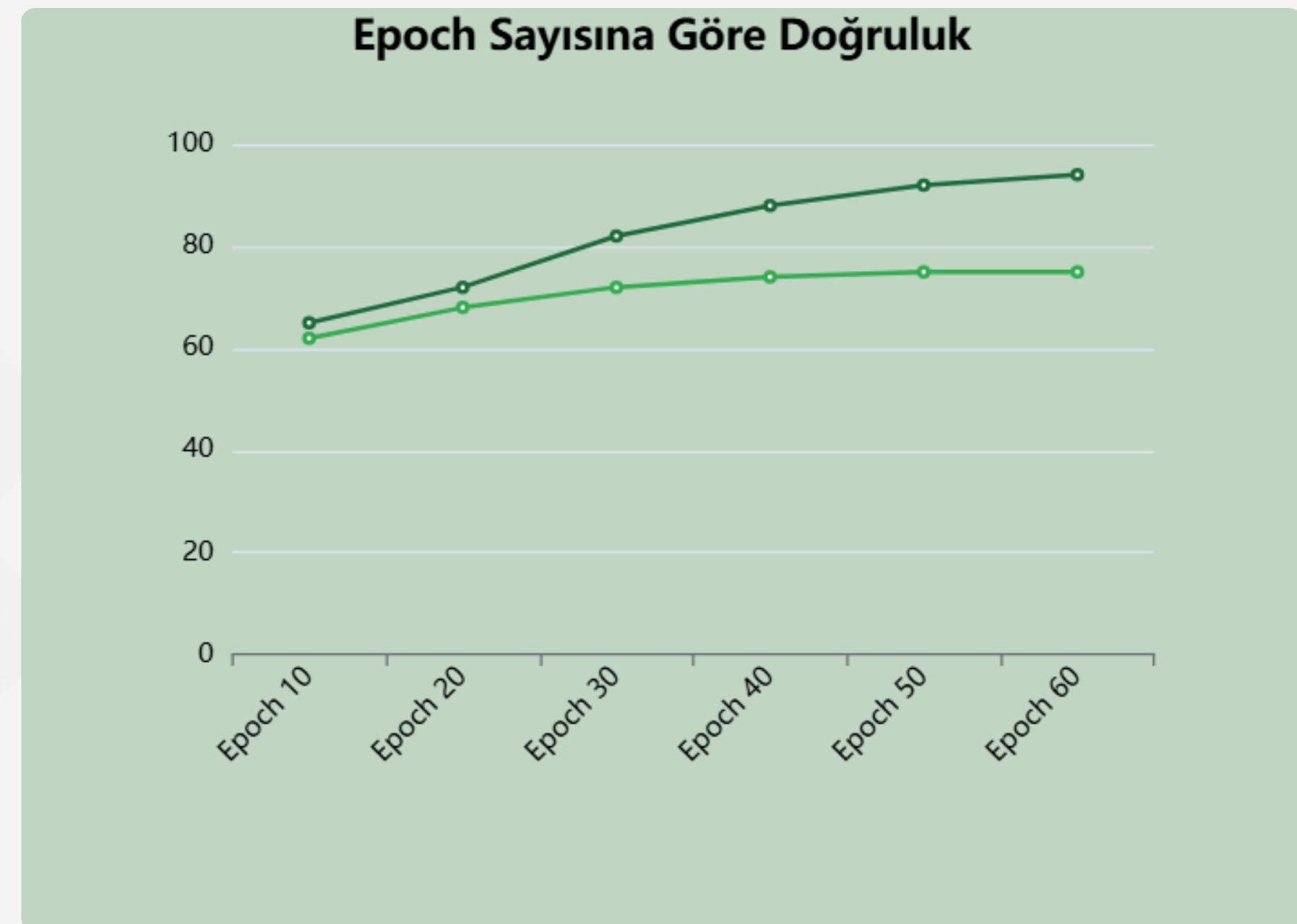
# Eğitim vs Doğrulama Doğruluk Eğrisi

## Doğruluk Analizi

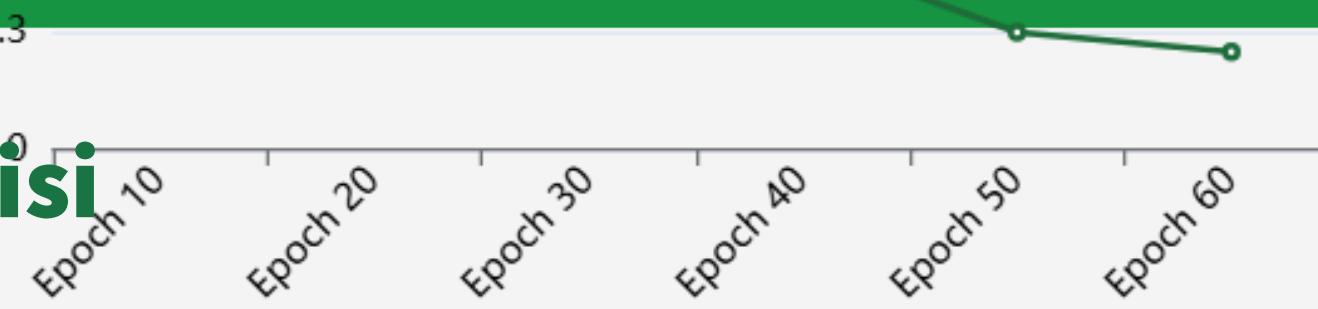
Eğitim doğruluğu yaklaşık 60. epoch sonunda %90'ı aşmıştır. Doğrulama doğruluğu %75 civarında stabilize olmuştur. Eğitim ve doğrulama eğrileri arasında ayrışma (divergence), aşırı uyumun (overfitting) varlığını gösterir.

**75%**

Doğrulama Doğruluğu



# Eğitim vs Doğrulama Kayıp Eğrisi

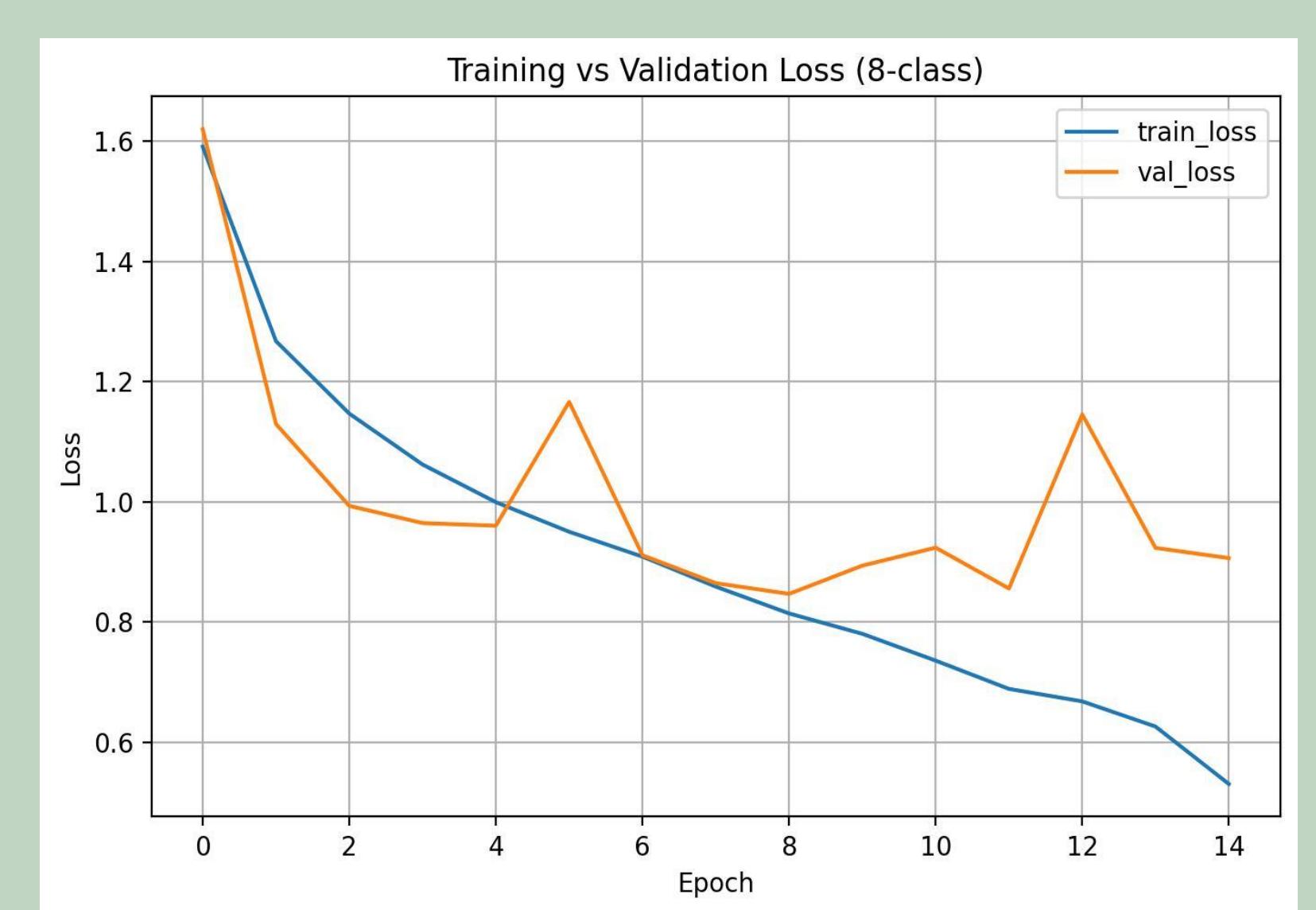


## Kayıp Yakınsaması

**Training loss:** ~1.6 → 0.16

**Validation loss:** en iyi nokta epoch 10–15

Kayıp eğrileri arasındaki ıraksamanın başlangıcı, en uygun epoch sayısını belirlemede yardımcı olmuştur.



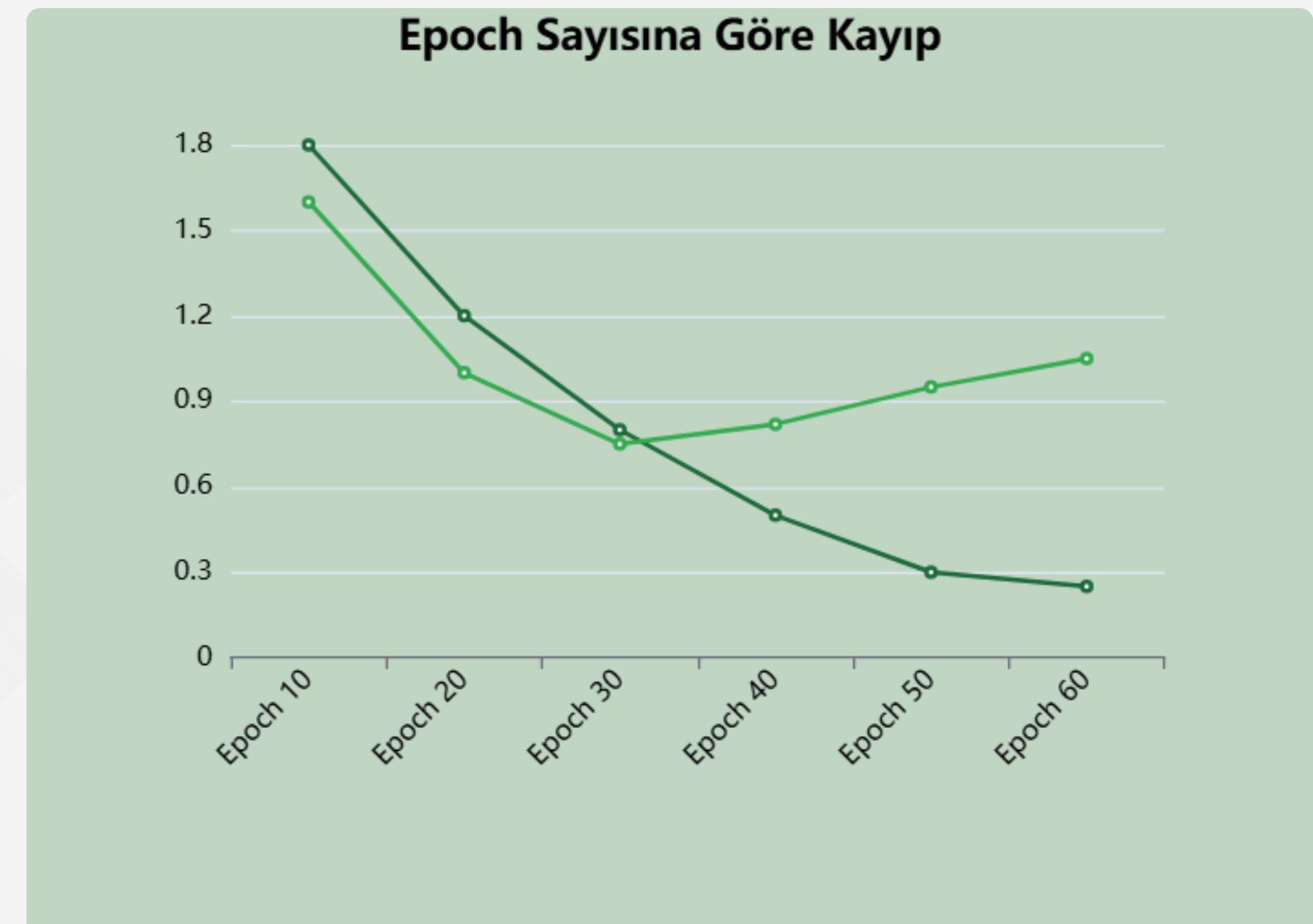
# Eğitim vs Doğrulama Kayıp Eğrisi

## Kayıp Yakınsaması

**Eğitim kaybı ilk 20 epoch boyunca hızlı bir şekilde düşmüştür. Doğrulama kaybı 30. epoch civarında en düşük değerine ulaşmış, sonrasında yükselişe geçmiştir. Bu eğri, modelin belirli bir noktada doğrulama verisi üzerinde kötüye gidişe başladığını gösterir.**

**Kayıp eğrileri arasındaki ıraksamanın başlangıcı, en uygun epoch sayısını belirlemede yardımcı olmuştur.**

$$(\hat{y}_k) y_k \log \sum_{k=1}^s - = \mathcal{L}$$



# Test Sonuçları



## Test Metrikleri

**Test Doğruluğu:** **71.6%** (60 epoch)

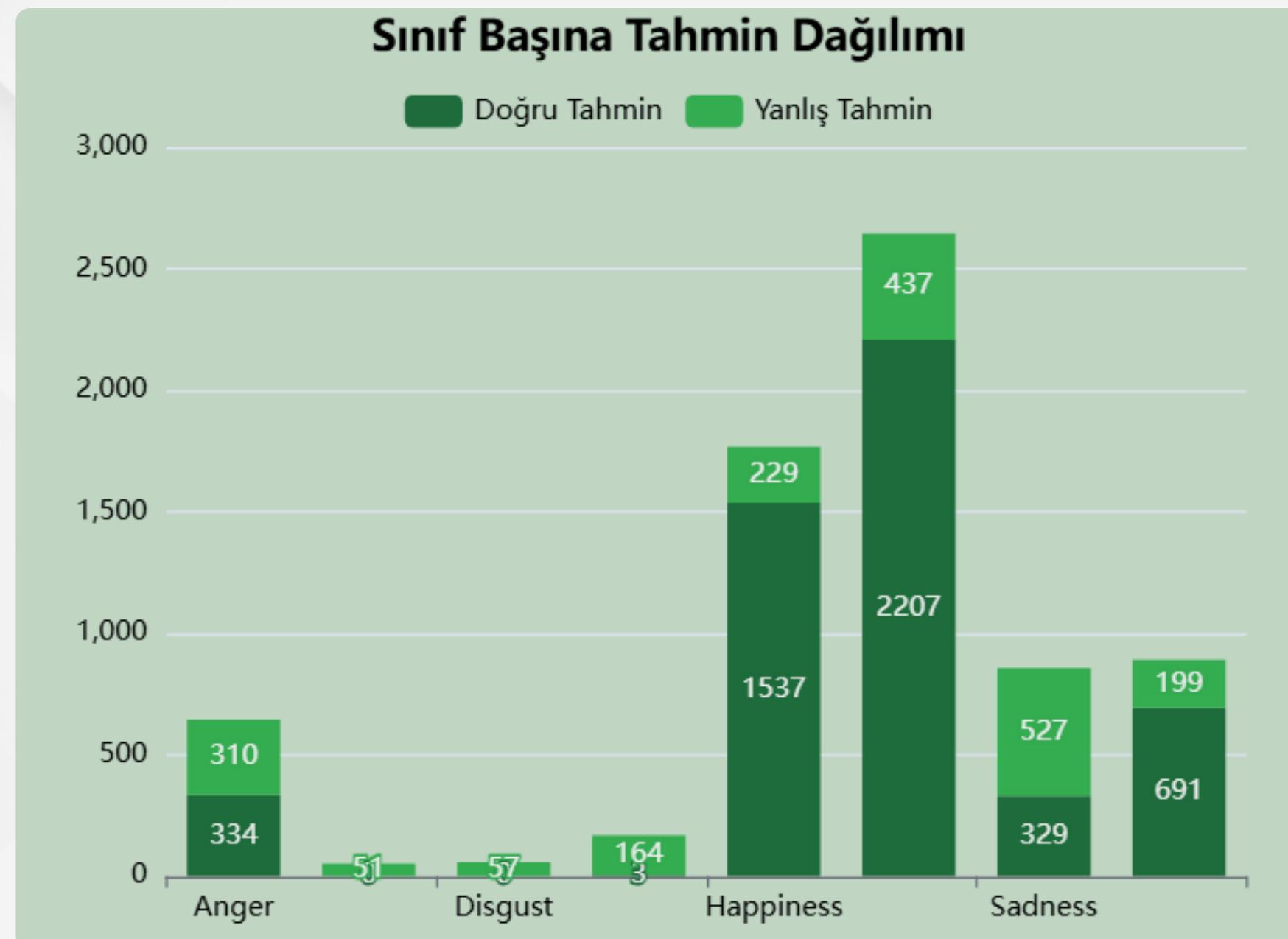
**Test Kaybı:** **1.08**



## Yorumlama

**Test verisi üzerinde %71.6 doğruluk makul bir performans göstermektedir. Ancak, doğruluk tek başına yeterli değildir**  
**- sınıf dengesizliği göz önüne alındığında, sınıf-bazında metrikler de incelenmelidir.**

# Karışıklık Matrisi (Sayısal Değerler) - 8 Sınıf



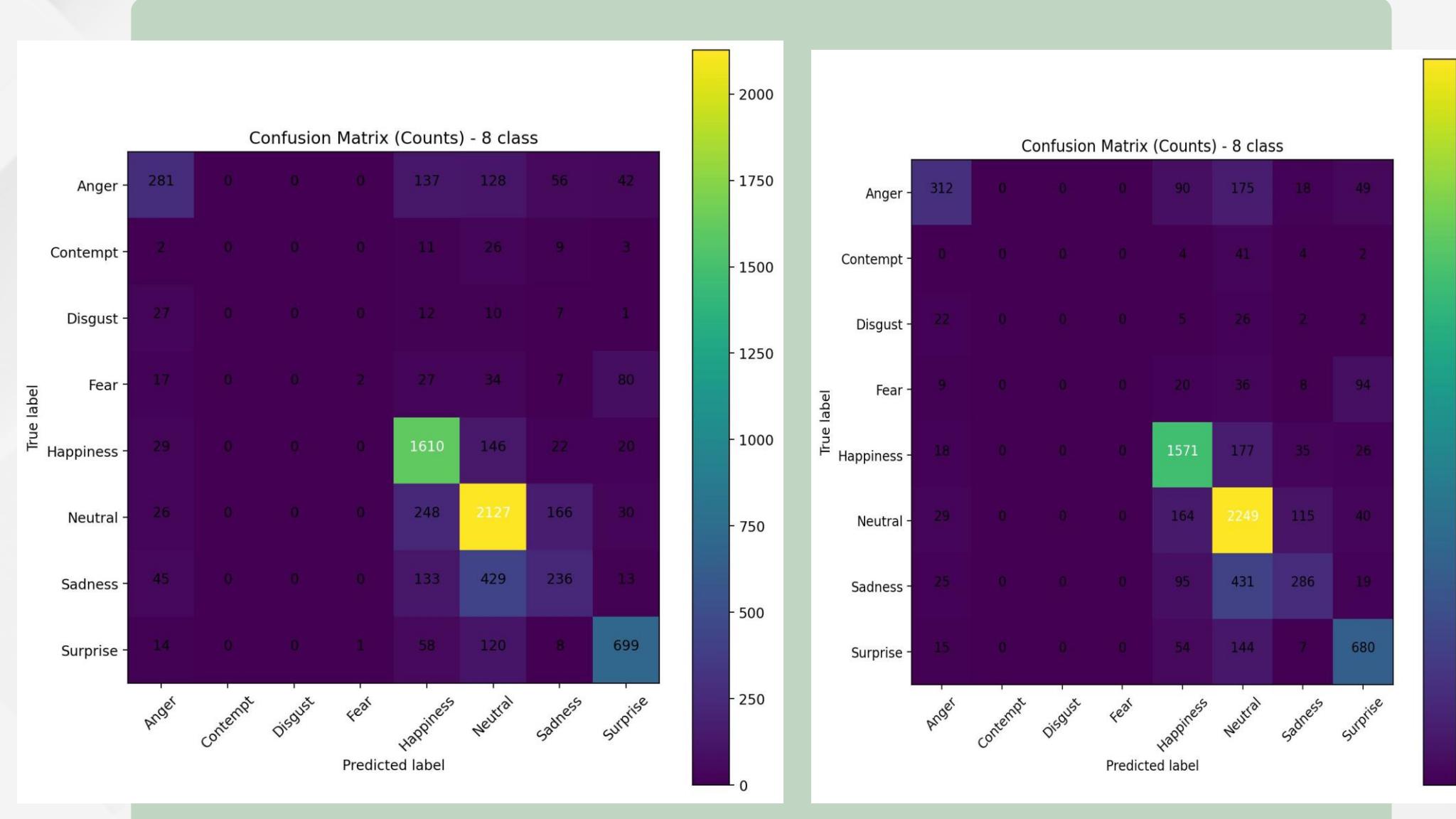
## Ana Bulguları

**Happiness ve Neutral sınıfları en yüksek doğruluk oranına sahiptir. Contempt, Disgust ve Fear sınıfları düşük performans göstermektedir. Anger sınıfı sıklıkla Neutral olarak yanlış sınıflandırılmaktadır. Fear sınıfı Surprise ile sık sık karışmaktadır.**

- **En iyi sınıflar: Happiness (%84), Neutral (%85), Surprise (%77)**
- **Zayıf sınıflar: Contempt (%0), Disgust (%0), Fear (%2)**
- **Temel karışıklık: Anger↔Neutral, Fear↔Surprise, Disgust↔Neutral**

# Karışıklık Matrisi (Sayısal Değerler) - 8 Sınıf

## Ana Bulguları



**Test doğruluğu: %54.5 Test Loss: 2.4276 Test**

**Accuracy: 0.6980 for 360 epoch**

**En iyi sınıflar: Happiness, Neutral, Surprise, Anger**

**Zayıf sınıflar: Contempt, Disgust**

**Test doğruluğu: %71.8 Test Loss: 0.9193 Test**

**Accuracy: 0.7181 for 10-15 epoch**

**Yorumu:**

**Happiness ve Neutral belirgin görsel örüntülere sahiptir**

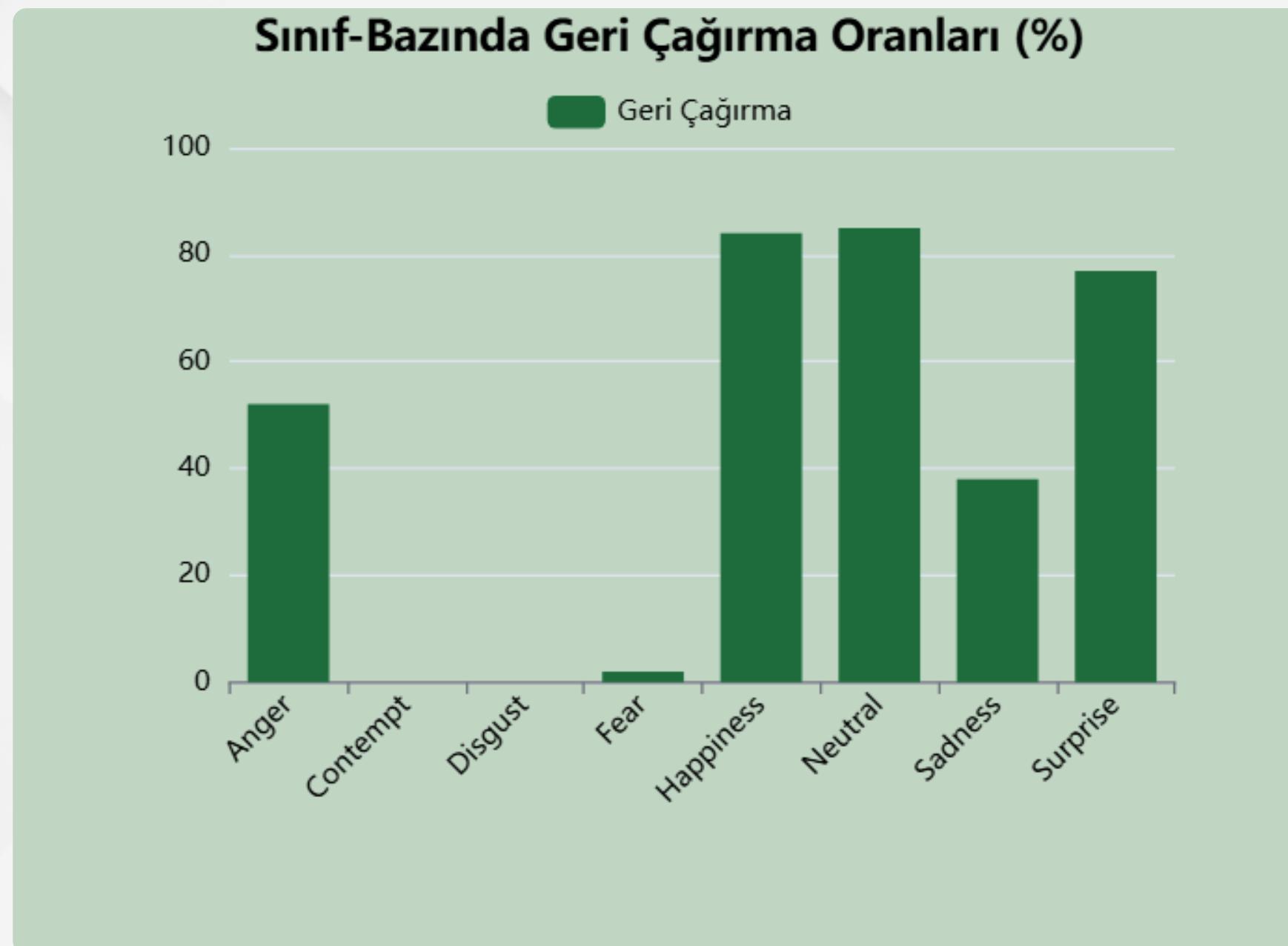
**Fear ve Disgust gibi ifadeler:**

**Görsel olarak belirsiz**

**Düşük çözünürlükte ayırt edilmesi zor**

**Bu durum model hatası değil, problem doğasının bir sonucudur.**

# Karışıklık Matrisi (Normalize Edilmiş) - 8 Sınıf

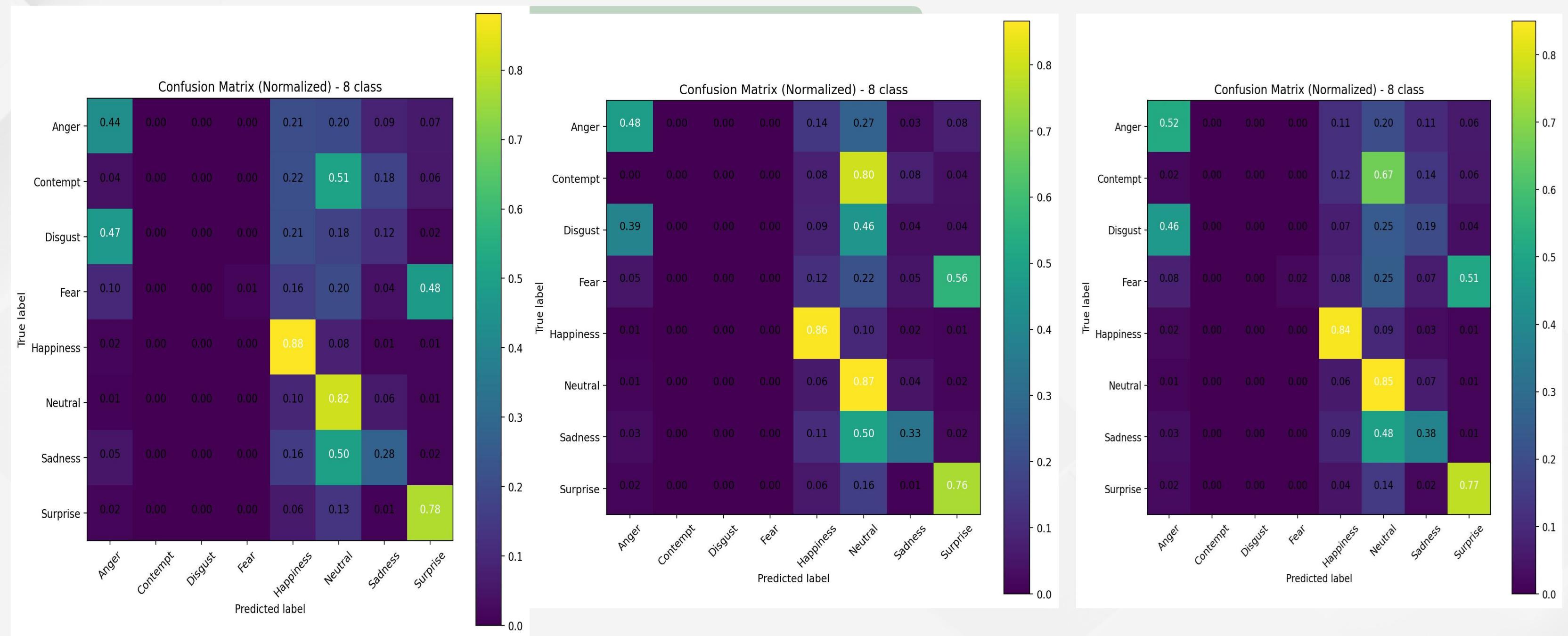


## Normalize Edilmiş Analiz

**Sınıf dengesizliğinin etkisi açıkça görülmektedir.**  
**Happiness ve Neutral gibi temsil sayısı yüksek sınıflar iyi performans gösterirken, Contempt ve Disgust gibi az temsil edilen sınıflar %0 geri çağrıma oranına sahiptir.**

- **Happiness geri çağrıma: 84% - yüksek temsil sayısı avantajı**
- **Contempt geri çağrıma: 0% - yalnızca 51 görüntü**
- **Disgust geri çağrıma: 0% - yalnızca 57 görüntü ile test edildi**

# Karışıklık Matrisi (Normalize Edilmiş) - 8 Sınıf



# Sınıflandırma Raporu

## Performans Metrikleri

**Kesinlik (Precision), Duyarlılık (Recall) ve F1-Skoru** değerleri her sınıf için hesaplanmıştır. Makro ve ağırlıklı ortalamalar genel modelin performansını temsil etmektedir.



### Kesinlik (Precision)

**Makro Ortalama: 42% | Ağırlıklı Ortalama: 72%.** Modelin pozitif tahminlerinin ne kadarı doğru olduğunu gösterir.



### Duyarlılık (Recall)

**Makro Ortalama: 38% | Ağırlıklı Ortalama: 71%.** Gerçek pozitif örneklerin ne kadarını model doğru şekilde bulduğunu gösterir.



### F1-Skoru

**Makro Ortalama: 34% | Ağırlıklı Ortalama: 70%.** Kesinlik ve duyarlılık arasında denge sağlayan harmonic ortalamasıdır.

# Örnek Tahminler

## DOĞRU VE YANLIŞ SINIFLANDIRILMIŞ YÜZ İFADLERİ



## Doğru Tahminler

Model tarafından  
doğru şekilde  
sınıflandırılan yüz  
ifadeleri. Happiness  
ve Neutral  
sınıflarında daha  
başarılı.

### CNN:

- Ağız açıklığını → Happiness
- Yüz stabilitesini → Neutral
- olarak yorumlayabilmektedir.

### ➤ CNN:

- Psikolojik bağlamı anlamaz
- Sadece piksel örüntülerine bakar

## Yanlış Tahminler

Modelin başarısız  
olduğu örnekler.  
**Anger↔Neutral,**  
**Fear↔Surprise**  
karışıklıklarında  
gözlemlenir.

## Görsel Belirsizlik

48×48 piksellik  
çözünürlük nedeniyle  
bazı ifadeler insan  
gözüne bile  
belirsizdir. Düşük  
çözünürlük, ince yüz  
detaylarının ve mikro  
ifadelerin  
kayıbmasına neden  
olarak sınıflandırma  
doğruluğunu etkiler.

# Grad-CAM Açıklanabilirlik (Explainability)



## Grad-CAM Nedir?

**Gradient-weighted Class Activation Mapping.** Model'in belirli bir tahminde hangi piksel bölgelerine odaklandığını gösterir. Kırmızı bölgeler yüksek önem, mavi bölgeler düşük önem gösterir.



## Neden Önem Taşır?

**Açıklanabilirlik Gerekliği.** Black-box CNN modellerinin kararlarının anlaşılması kritiktir. Grad-CAM analiz ile modelin rasgele karar vermediği kanıtlanabilir.



## Bulgular

**Model Odak Bölgeleri.** Model gözler, ağız ve kaşları odaklanmaktadır - duygusal ifadede anlamı olan bölgeler. Bu, modelin semantik olarak doğru özellikler öğrendiğini gösterir.

# Grad-CAM Nedir? (Teorik Açıklama)

Grad-CAM, modelin **hangi bölgelerden etkileneerek karar verdiği** rib neretsög .ridimetnöy kilrilibanalıkça

**Amaç:**

Model gerçekten yüzün anlamlı bölgelerine mi bakıyor!

**Grad-CAM Matematiksel Tanımı**

Son konvolüsyon katmanının özellik haritaları:

$$^{u \times v} \mathbb{R} \ni {}^k A$$

Hedef sınıf skoru:

$${}^c y$$

Ağırlıklar:

$$\frac{{}^c y \partial}{{}^k A \partial} \sum_j \sum_i \frac{1}{Z} = {}^c_k \alpha$$

**Grad-CAM haritası:**

ReLU, yalnızca **pozitif katkıları** tutar.

$$\text{ReLU} \left( \sum_k \alpha_k^c A^k \right) = {}_{\text{Grad-CAM}}^c L$$

## Örnek Tahminler gradcam uguladiktan sonra// Grad-CAM Görsel Yorumu

Kırmızı bölgeler → karar için kritik  
Mavi bölgeler → önemsiz



Örnek:

- Happiness → ağız , Anger → kaşlar ve gözler , Surprise → göz ve ağız
- Bu, modelin **anlamsız alanlara değil**, yüze odaklandığını kanıtlar.

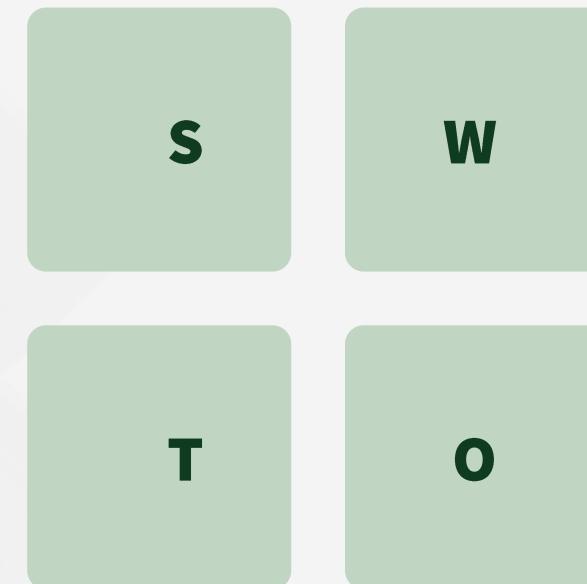
# Bazı Sınıflar Neden Başarısız?

## Sınıf Dengesizliği

**FER2013 veri seti dengesizdir.** Happiness ve Neutral sınıfları binlerce örneğe sahipken, Contempt ve Disgust yalnızca 50-60 örneğe sahiptir. Model, nadir sınıfları yeterince öğrenememektedir.

## Etiket Gürültüsü

**FER2013 veri seti kısmen el ile etiketlenmiştir** ve hata içerebilir. Bu, modeli yanıltıcı sinyallerle eğitir.



## Düşük Çözünürlük

48×48 piksellik gri tonlamalı görüntüler, insan gözüne bile bazı ifadeleri belirsiz göstermektedir. Yüksek çözünürlük RGB görüntüler daha iyi sonuç verecektir.

## Görsel Benzerlik

Anger ve Neutral, Fear ve Surprise gibi sınıflar görsel olarak son derece benzerdir. Mikro ifadeleri ve ince farklılıklarını ayırt etmek zordur.

**Not:** Bu, kodlama hatası değil, problem sınırlamasıdır.

# Akademik Yorumlama

## Hiyerarşik Öznitelik Öğrenmesi

CNN modeli, ilk katmanlarda kenarları ve kontrast gibi düşük-seviye öznitelikleri, sonraki katmanlarda göz, ağız ve kaş gibi yüksek-seviye semantik öznitelikleri, son katmanlarda ise duygusal ifade örüntülerini öğrenir. Bu hiyerarşik yapı, sinir ağlarının gücünün kaynağıdır.

## Psikoloji Olmayan Öğrenme

CNN, görüntü piksellerini matematiksel işlemlerle işleyerek olasılık dağılımları üretir. Model, duygusal psikolojisini öğrenmez, sadece piksel-duygu korelasyonlarını bulur. Bu, yapay zeka'nın ve insan algısının temel farklılığıdır.

## Literatür Tutarlılığı

Elde edilen sonuçlar (%71.6 test doğruluğu) FER2013 üzerine yapılmış önceki araştırmalarla tutarlıdır. ResNet veya EfficientNet gibi daha karmaşık modeller daha yüksek doğruluk elde etmiştir, ancak basit CNN'nin sonuçları makuldür.

# Sonuçlar



## Ham Piksel CNN Başarısı

Elle yazılan özniteliklere ihtiyaç olmadan, ham pikseller doğrudan CNN'ye verilebilir. Model, %71.6 makul test doğruluğu elde etmiştir. End-to-end öğrenme pratik ve etkilidir.



## Basit Mimar, Yüksek Reproducibility

Kullanılan CNN mimarisi hafif ve basittir (3 konvolüsyon bloğu). Kod tam olarak açık ve kolayca tekrarlanabilir. GPU olmadan CPU'da bile çalıştırılabilir.



## Açıklanabilirlik Doğrulama

Grad-CAM analiz ile modelin anlamlı yüz bölgelerine (göz, ağız, kaş) odaklandığı doğrulanmıştır. Model rasgele karar almamakta, denetimli öğrenme başarı göstermektedir.

# İlerideki Çalışmalar



## Sınıf Dengeleme

- **Class weighting, Focal Loss gibi tekniklerle dengesiz sınıfları ağırlıklandıırın**
- **Alternatif olarak, az temsil edilen sınıflar için veri artırma**



## Aşırı Uyum Azaltma

- **Daha yüksek Dropout oranları, L1/L2 regularization**
- **EarlyStopping callback'ı daha agresif tuning**



## Transfer Öğrenme

- **ResNet, EfficientNet, MobileNet gibi pre-trained modellerden başlayın**
- **Fine-tuning ile daha yüksek doğruluk**



## Yüksek Çözünürlük

- **48×48 yerine 128×128 veya 256×256 piksellik RGB görüntülerini kullanın**
- **Görsel bilgi artışı performansı önemli ölçüde iyileştirir**



## Ek Veri Setleri

- **FER2013 yanında RAF-DB, AffectNet gibi diğer veri setleriyle eğitin**
- **Cross-dataset doğrulama model genelleştirmesini gösterir**

# Teşekkür Ederim

Sorularınız ve görüşleriniz için teşekkür ederim. Tartışma ve ek açıklamalar için hazırlım.