Journal of Environmental Research and Engineering

JERE, 2026, 2 (1): 28-46

# Deepfake Detection and Generalization: A Comparative Study of Deep Learning Architectures

Ali Eren Tuğrul [1]* , Yakup Kutlu[1]

[1]Department of Computer Engineering, Iskenderun Technical University, Hatay, Türkiye.

**ABSTRACT**

Deepfake technology, which enables the creation of highly realistic synthetic face images using deep learning, poses significant threats to information integrity and digital trust. This study presents a comparative evaluation of deep learning architectures for image-based deepfake detection, with a particular focus on generalization across different data distributions. Five models were implemented and evaluated: MesoNet, ResNet-50, EfficientNet-B4, Xception, and an extended Xception model retrained with an additional DF40-based external dataset. Experiments were conducted on the 140k Real and Fake Faces dataset, which contains 100,000 training, 20,000 validation, and 20,000 test images, and were extended with 25,696 external training images. All models were trained using transfer learning where applicable, CrossEntropyLoss with label smoothing, AdamW optimization, cosine annealing, mixed precision, early stopping, and checkpoint selection based on validation loss. The original Xception model achieved 99.73% accuracy on the in-distribution test set, but its external test accuracy dropped to 52.74%, with a fake recall of only 6.04%. In contrast, the Xception+DF40 model reached 92.15% accuracy, 0.9828 AUC, and 84.50% fake recall on the external test set while maintaining 99.10% accuracy on the original test set. These results show that data diversity and out-of-distribution evaluation are critical for robust deepfake detection.

**ARTICLE INFO**

**Received** 2026-05-11,

**Accepted** 2026-06-29,

**Publication Date** 2026-06-30

**Keywords**:

Deepfake Detection, Transfer Learning, Xception, Generalization, Out-of-Distribution, DF40, EfficientNet, ResNet

# Deepfake Tespiti ve Genelleme: Derin Öğrenme Mimarilerinin Karşılaştırmalı Analizi


## ÖZET

Derin öğrenme kullanılarak gerçekçi sentetik yüz görüntüleri üretebilen deepfake teknolojisi, bilgi bütünlüğü, dijital güven ve mahremiyet açısından önemli tehditler oluşturmaktadır. Bu çalışmada, görüntü tabanlı deepfake tespiti için farklı derin öğrenme mimarileri karşılaştırmalı olarak incelenmiş ve özellikle farklı veri dağılımlarında genelleme başarımı değerlendirilmiştir. MesoNet, ResNet-50, EfficientNet-B4, Xception ve DF40 tabanlı harici veriyle yeniden eğitilmiş Xception+DF40 modeli uygulanmıştır. Deneyler, 100.000 eğitim, 20.000 doğrulama ve 20.000 test görselinden oluşan 140k Real and Fake Faces veri seti üzerinde yürütülmüş; genelleme başarımını artırmak için 25.696 eğitim görseli içeren DF40 tabanlı harici veri kaynağı eklenmiştir. Modeller, uygun mimarilerde transfer öğrenme, label smoothing içeren CrossEntropyLoss, AdamW, CosineAnnealingLR, karma hassasiyet, erken durdurma ve en düşük doğrulama kaybına göre checkpoint seçimi ile eğitilmiştir. Orijinal Xception modeli ana test setinde %99,73 doğruluk elde etmesine rağmen harici test setinde %52,74 doğruluğa ve yalnızca %6,04 sahte geri çağırıma düşmüştür. Buna karşılık Xception+DF40 modeli harici test setinde %92,15 doğruluk, 0,9828 AUC ve %84,50 sahte geri çağırım değerine ulaşmış; ana test setinde ise %99,10 doğruluğu korumuştur. Bulgular, deepfake tespitinde veri çeşitliliğinin ve dağılım dışı değerlendirmenin dayanıklı model geliştirme açısından kritik olduğunu göstermektedir.



## MAKALE BİLGİSİ

**Anahtar Kelimeler:**

Deepfake Tespiti, Transfer Öğrenme, Xception, Genelleme, Dağılım Dışı Test, DF40, EfficientNet, ResNet




## GİRİŞ

Günümüzde dijital medya içeriklerinin hızla artması ve yapay zeka teknolojilerinin gelişimi, özellikle yüz manipülasyonu alanında derin sahte (deepfake) içeriklerin yaygınlaşmasına yol açmıştır. Deepfake teknolojisi, gerçekçi ancak sahte görüntü ve videolar üreterek bireylerin kimliklerini taklit etme, ifadelerini değiştirme ve tamamen yeni içerikler oluşturma imkanı

sağlamaktadır. Bu durum, kimlik hırsızlığı, dolandırıcılık ve mahremiyet ihlalleri gibi ciddi sosyal ve hukuki sorunları beraberinde getirmektedir. Bu bağlamda, deepfake içeriklerin tespiti için geliştirilen modellerin performanslarının karşılaştırılması hem akademik hem de pratik açıdan büyük önem taşımaktadır (Kumar ve Verma, 2026).

Deepfake tespiti alanında kullanılan yöntemler genel olarak görüntü tabanlı ve ses tabanlı olmak üzere iki ana kategoride incelenebilir. Görüntü tabanlı yöntemlerde sahte yüzlerin gerçeklerden ayırt edilmesi için çeşitli derin öğrenme modelleri kullanılmaktadır. Kumar ve Verma tarafından yapılan karşılaştırmalı çalışmada Xception, ResNet50, EfficientNetB0, DenseNet121 ve MobileNet gibi önceden eğitilmiş modeller değerlendirilmiş; Xception modelinin %99,14 test doğruluğu ile öne çıktığı raporlanmıştır. Bu bulgu, transfer öğrenmenin deepfake tespitinde etkili olduğunu ve özellikle Xception mimarisinin karmaşık sahte yüz izlerini yakalamada güçlü bir aday olduğunu göstermektedir (Kumar ve Verma, 2026).

Bununla birlikte deepfake tespitindeki en önemli zorluklardan biri, modellerin farklı veri setleri ve daha önce görülmemiş sahte üretim teknikleri karşısında sınırlı genelleme yeteneği göstermesidir. Peng ve arkadaşları, yüz manipülasyon algoritmalarının gelişmesiyle birlikte yerel bozulma ve frekans izlerinin azalabileceğini, sıkıştırma ve iletim kaynaklı bozulmaların da tespit başarısını olumsuz etkileyebileceğini vurgulamaktadır. Bu nedenle veri artırma, çoklu kaynak öğrenimi, özellik güçlendirme, parmak izi tespiti ve zaman serisi analizi gibi stratejiler genelgeçer deepfake tespiti açısından önem kazanmaktadır (Peng et al., 2025).

Video tabanlı deepfake tespiti ise hem mekansal hem de zamansal özelliklerin birlikte analiz edilmesini gerektirir. Ma ve arkadaşlarının önerdiği 3D Dikkat Ağı modeli, video düzeyinde mekansal-zamansal özellikleri çıkaran hafif bir dikkat modülü ile FaceForensics++ veri setinde yüksek doğruluk ve Celeb-DF veri setinde güçlü çapraz veri transferi göstermiştir (Ma et al., 2023). Benzer şekilde Sun ve arkadaşlarının Sequential-Parallel Networks (SPNet) modeli, mekansal ve zamansal özellikleri sıralı olarak çıkarıp paralel biçimde birleştirerek büyük ölçekli deepfake veri setlerinde yüksek tanıma başarımı ve düşük hesaplama karmaşıklığı sunmuştur (Sun et al., 2023).

Genelleme yeteneğini artırmak amacıyla farklı öğrenme paradigmaları da önerilmiştir. Huang ve arkadaşlarının çok görevli kendi kendine karışık görüntüler yaklaşımı, sentez parametrelerine dayalı yardımcı kayıplar kullanarak kendi kendine denetimli modellerin performansını artırmıştır (Huang et al., 2023). Luo ve arkadaşlarının Forgery-aware Adaptive Vision Transformer (FA-ViT) modeli ise önceden eğitilmiş Vision Transformer parametrelerini sabit tutarken adaptif modüllerle yerel ve küresel sahtecilik ipuçlarını yakalayarak çapraz veri seti değerlendirmelerinde yüksek AUC değerleri elde etmiştir (Luo et al., 2024).

Frekans alanı bilgisi deepfake tespitinde tamamlayıcı bir rol oynamaktadır. Li ve arkadaşlarının frekans bilgisine dayalı ayrımcı özellik madenciliği yöntemi, RGB ve frekans alanı özelliklerini konum ilişkisine dayalı birleştirme modülü ile entegre ederek hem veri seti içi hem de çapraz veri seti değerlendirmelerinde başarılı sonuçlar vermiştir (Li et al., 2021). Tian ve arkadaşlarının frekans farkındalıklı dikkatli özellik füzyonu yaklaşımı ile Fang ve arkadaşlarının iki akışlı

işbirlikçi öğrenme çerçevesi de uzamsal doku farkları ile frekans ipuçlarını birlikte kullanarak genelleme başarımını artırmayı hedeflemektedir (Tian et al., 2023; Fang et al., 2025).

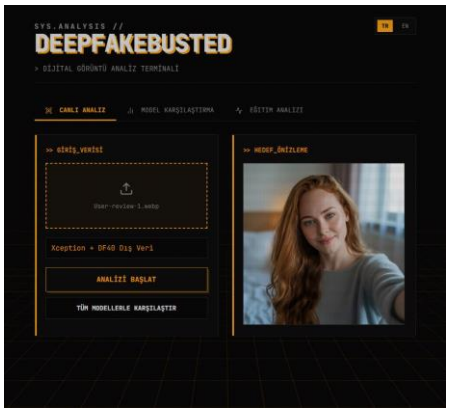
Kimlik bilgisi sızıntısı da deepfake tespitinde genelleme sorununu derinleştiren bir faktör olarak tartışılmaktadır. Dong ve arkadaşlarının tanımladığı Implicit Identity Leakage fenomeni, modellerin istemeden kimlik temsillerini öğrenerek sahtecilik izleri yerine kimliğe özgü örüntülere yönelebileceğini göstermiştir (Dong et al., 2023). Zhuang ve arkadaşlarının adversarial öğrenme tabanlı yaklaşımı ise farklı sahtecilik yöntemleri ve kimlik etkilerini azaltarak ortak ayrımcı özelliklerin öğrenilmesini hedeflemiştir (Zhuang et al., 2022).

Biyometrik yüz tanıma teknikleri de deepfake tespitinde incelenmiştir. Ramachandran ve arkadaşları, derin yüz tanıma yöntemlerinin iki sınıflı CNN tabanlı yöntemlere kıyasla daha az eğitim verisi gerektirebildiğini ve gelişmiş genelleme yeteneği sunabildiğini göstermiştir. Celeb-DF veri setinde 0,98 AUC ve 7,1 EER değerlerinin elde edilmesi, yüz tanıma temelli yaklaşımların deepfake tespitinde destekleyici bir araç olabileceğini ortaya koymaktadır (Ramachandran et al., 2021).

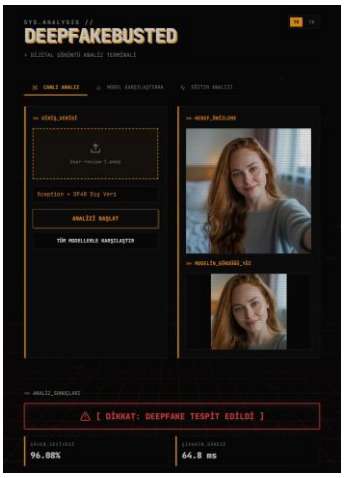
Deepfake tespiti alanındaki karşılaştırmalı çalışmalar, farklı mimarilerin güçlü ve zayıf yönlerini nesnel biçimde değerlendirmek açısından önemlidir. Deng ve arkadaşlarının benchmark çalışmasında 13 farklı yöntem, 117 model ve 882 performans ölçümü üzerinden adil koşullarda incelenmiştir (Deng et al., 2024). Kingra ve arkadaşlarının IAV-DF veri seti gibi demografik çeşitliliği yüksek kaynaklar, modellerin farklı yüz tipleri üzerindeki dayanıklılığını test etmek için yeni olanaklar sunmaktadır (Kingra et al., 2025). Gong ve Li ise 2019-2023 yılları arasındaki yaklaşımları CNN tabanlı, yarı denetimli, transformer tabanlı ve biyolojik sinyal tabanlı sınıflara ayırarak çapraz veri seti değerlendirmelerinde doğruluk oranlarının düştüğünü vurgulamıştır (Gong ve Li, 2024).

Sonuç olarak deepfake tespitinde model karşılaştırması; mimari seçimi, öğrenme stratejisi, veri seti çeşitliliği, hesaplama verimliliği ve değerlendirme metrikleri açısından çok boyutlu bir yaklaşım gerektirir. Yalnızca yüksek doğruluk değerlerine odaklanmak, özellikle dağılım dışı örneklerde sahte görüntüleri kaçırma riskini gizleyebilir. Bu nedenle doğrulukla birlikte AUC-ROC, F1-skoru, kesinlik ve özellikle sahte sınıf geri çağırımı gibi metriklerin birlikte incelenmesi gerekir.

Bu çalışmada MesoNet, ResNet-50, EfficientNet-B4, Xception ve DF40 tabanlı harici veri ile yeniden eğitilen Xception+DF40 modeli karşılaştırılmıştır. Çalışmanın temel katkısı, yalnızca kapalı veri seti başarımına değil, harici veri setlerinde genelleme başarımına da odaklanan daha dayanıklı bir Xception tabanlı model geliştirmek ve veri çeşitliliğinin deepfake tespitindeki etkisini deneysel olarak göstermektir.




*Şekil 1. DeepFakeBusted web arayüzü — Canlı analiz giriş ekranı (görsel yükleme ve model seçimi)*




*Şekil 2. DeepFakeBusted web arayüzü — Analiz sonucu: deepfake tespit çıktısı (%96.08 güven, yüz kırpma önizlemesi)*

## MATERYAL VE YÖNTEM

### Veri Setleri

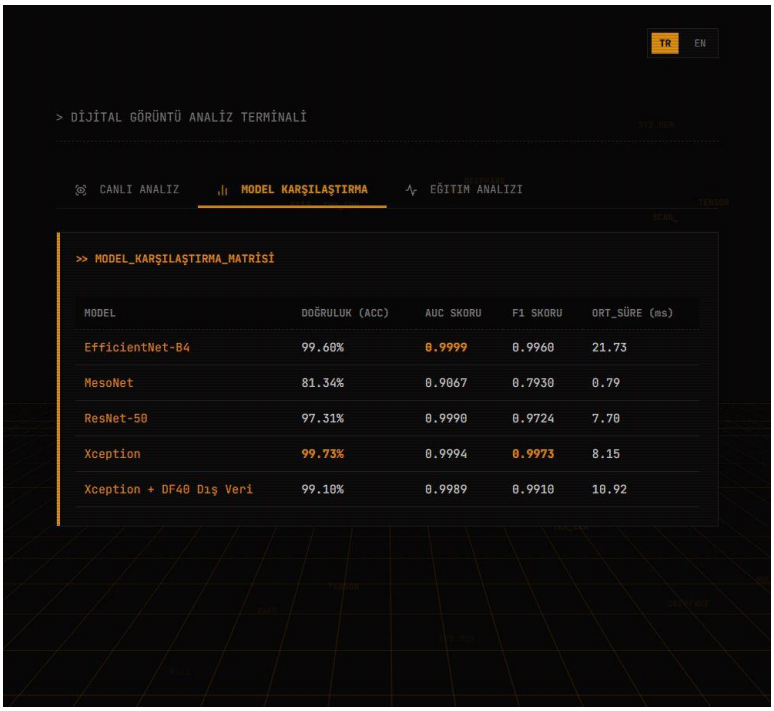
Bu çalışmada deepfake görüntü tespiti için iki farklı veri kaynağı kullanılmıştır. Ana veri seti olarak 140k Real and Fake Faces veri seti tercih edilmiştir. Veri seti eğitim için 100.000 görselden (50.000 gerçek, 50.000 sahte), doğrulama için 20.000 görselden (10.000 gerçek, 10.000 sahte) ve test için 20.000 görselden (10.000 gerçek, 10.000 sahte) oluşmaktadır. Etiketleme real=0 ve fake=1 biçiminde yapılmış, model çıktılarında özellikle sahte sınıfa ait softmax olasılığı olan fake_probability değeri kullanılmıştır.

Ana veri setindeki yüksek başarıya rağmen dış kaynaklı ve daha profesyonel deepfake örneklerinde performans kaybı gözlendiği için DF40 tabanlı ek bir harici veri kaynağı kullanılmıştır. Bu veri kaynağı eğitim için 25.696 görselden (12.848 gerçek, 12.848 sahte), doğrulama için 3.212 görselden ve test için 3.212 görselden oluşmaktadır. Final eğitimde ana veri ve harici veri birlikte kullanılarak modelin dağılım dışı örneklerdeki genelleme kapasitesi artırılmaya çalışılmıştır.

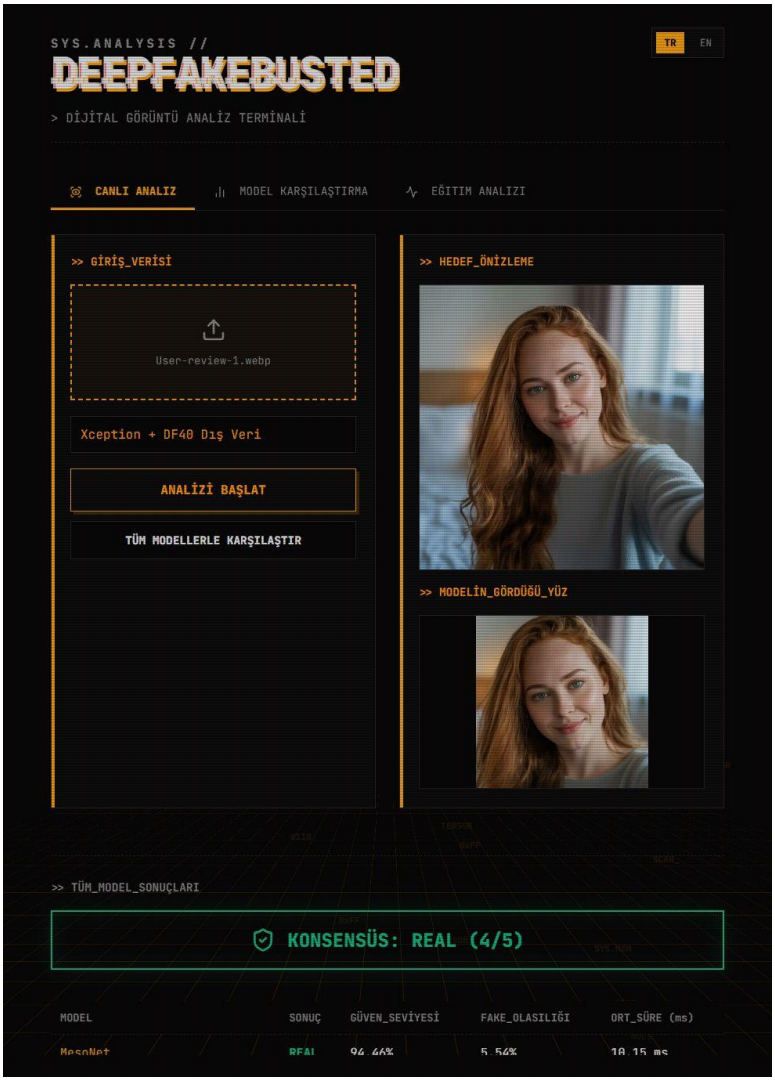
## Veri Hazırlama

Görüntüler modele verilmeden önce standart bir ön işleme hattından geçirilmiştir. Eğitim aşamasında yaklaşık 256x256 yeniden boyutlandırma, 224x224 rastgele kırpma, rastgele yatay çevirme, renk değişimi, rastgele gri tonlama, tensöre dönüştürme ve ImageNet normalizasyonu uygulanmıştır. Doğrulama ve test aşamalarında ise 224x224 yeniden boyutlandırma, tensöre dönüştürme ve ImageNet normalizasyonu kullanılmıştır. Normalizasyon değerleri mean=[0.485, 0.456, 0.406] ve std=[0.229, 0.224, 0.225] olarak belirlenmiştir; bu tercih, ResNet, EfficientNet ve Xception gibi ImageNet üzerinde ön eğitimli modellerin beklediği giriş dağılımı ile uyum sağlamak içindir.

Canlı analiz sisteminde kullanıcı tarafından yüklenen görüntüler için ek bir yüz kırpma adımı uygulanmıştır. Flask backend tarafında görüntü PIL ile RGB biçimine dönüştürülmekte, OpenCV Haar Cascade ile yüz adayları bulunmakta, göz tespitiyle aday doğrulaması yapılmakta ve en büyük yüz kutusu %35 kenar boşluğu ile genişletilerek modele verilmektedir. Böylece modelin arka plan, kıyafet veya kompozisyon yerine karar açısından en kritik bölge olan yüze odaklanması sağlanmaktadır.




*Şekil 3. Veri seti örnekleri — 140k Real and Fake Faces veri setinden gerçek ve sahte yüz görüntüleri*

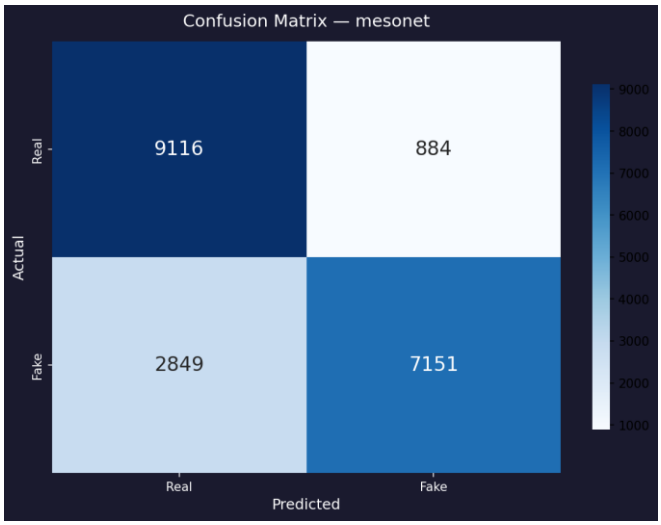



*Şekil 4. DeepFakeBusted web arayüzü — Tüm modeller karşılaştırma sonucu (konsensüs: REAL 4/5)*

Web prototipi React + Vite tabanlı bir arayüz ve Flask tabanlı bir API olarak tasarlanmıştır. Backend tarafında /api/models, /api/predict, /api/predict-all, /api/metrics, /api/logs ve /api/plots/<filename> uç noktaları ile model listesi, tek model tahmini, tüm modellerle karşılaştırma, metrikler, eğitim geçmişi ve grafikler arayüze aktarılmaktadır. Tahmin çıktıları real_probability ve fake_probability değerleriyle JSON biçiminde sunulmaktadır.

## Kullanılan Modeller

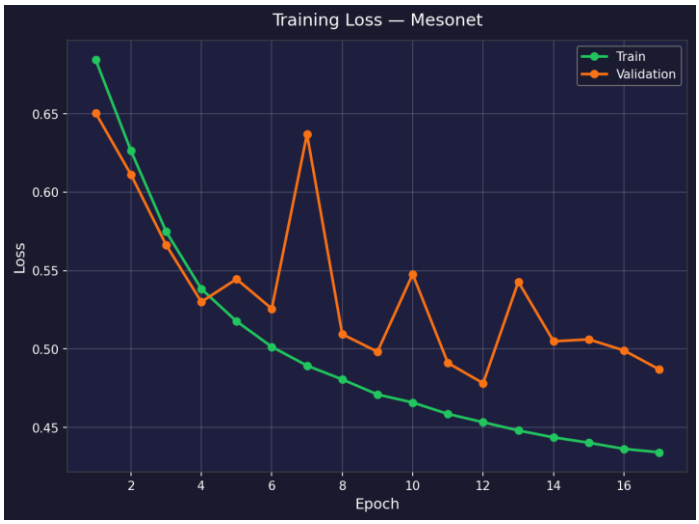
Çalışmada beş farklı derin öğrenme modeli uygulanmış ve karşılaştırılmıştır. MesoNet, deepfake tespitine özel hafif bir CNN olarak sıfırdan eğitilmiştir. ResNet-50, EfficientNet-B4 ve Xception modelleri ImageNet üzerinde ön eğitimli ağırlıklarla başlatılmış, son sınıflandırma katmanları gerçek/sahte olmak üzere iki sınıfa göre değiştirilmiştir. Xception+DF40 ise orijinal Xception modelinin DF40 tabanlı harici veriyle yeniden eğitilmiş final sürümüdür. Kod tabanında ViT desteği de bulunmasına karşın, bu çalışmanın ana karşılaştırması CNN tabanlı mimarilere odaklanmaktadır.




*Şekil 5. MesoNet modeli karışıklık matrisi — ana test seti (%81.34 doğruluk)*

## MesoNet

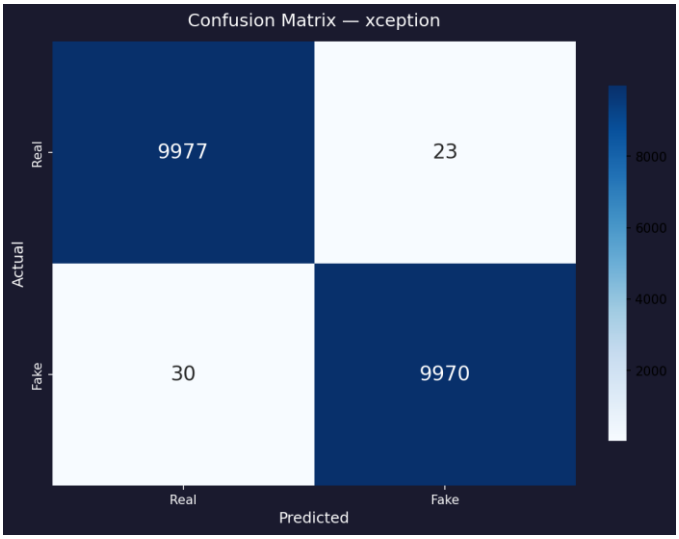
MesoNet, deepfake tespitine özel olarak tasarlanmış hafif bir konvolüsyonel sinir ağıdır. Model 3x224x224 RGB görüntüleri giriş olarak almakta; sırasıyla 3->8, 8->8, 8->16 ve 16->16 kanal geçişlerine sahip dört Conv2D, BatchNorm, ReLU ve MaxPool bloğundan oluşmaktadır. Düzleştirme işleminden sonra Dropout, 16*7*7 -> 16 Linear, LeakyReLU, Dropout ve 16 -> 2 Linear katmanları kullanılmaktadır. Batch boyutu 64, öğrenme hızı 1e-3, ağırlık azalması 1e-4 ve epoch sayısı 20 olarak belirlenmiştir. Model boyutu yaklaşık 0,09 MB olduğu için çok hızlı ve hafiftir; ancak karmaşık deepfake izlerini yakalamada daha güçlü mimarilere göre sınırlı kalmaktadır.



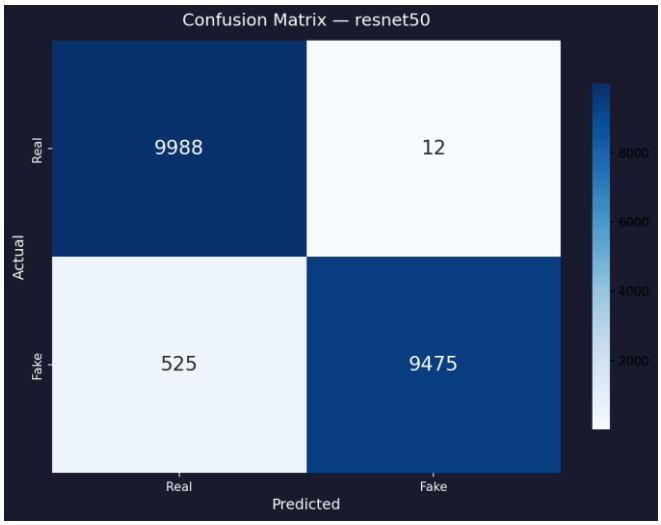

*Şekil 6. MesoNet modeli eğitim kaybı eğrisi — 20 epoch boyunca train ve validation kayıp değerleri*

ResNet-50, artık bağlantılar sayesinde derin ağlarda gradyan kaybı sorununu azaltan güçlü bir CNN mimarisidir. Bu çalışmada ImageNet üzerinde ön eğitimli ağırlıklarla kullanılmış ve son tam bağlantılı katman iki çıkışlı gerçek/sahte sınıflandırma katmanı ile değiştirilmiştir. Batch boyutu 32, öğrenme hızı 1e-4, ağırlık azalması 1e-4 ve epoch sayısı 20 olarak ayarlanmıştır. Model boyutu 89,89 MB olup ana veri setinde %97,31 doğruluk ve 0,9990 AUC elde etmiştir.

**Tablo 1.** Ana veri seti üzerinde model performans karşılaştırması




Tablo 1'de doğruluk, AUC-ROC, F1-skoru, kesinlik, geri çağırım, çıkarım süresi ve model boyutu birlikte değerlendirilmiştir. Sınıf sırası real=0 ve fake=1 olarak tutulmuş, confusion matrix değerleri bu sıraya göre yorumlanmıştır.

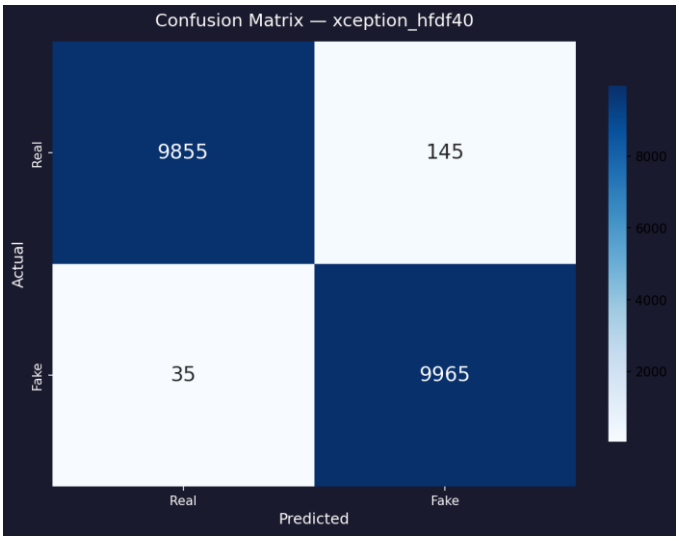



*Şekil 7. ResNet-50 modeli karışıklık matrisi — ana test seti (%97.31 doğruluk)*

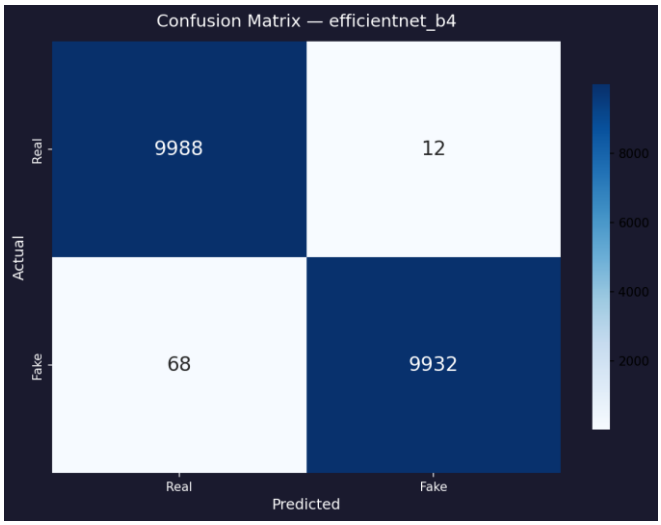
**EfficientNet-B4**

EfficientNet-B4, bileşik ölçeklendirme stratejisiyle derinlik, genişlik ve çözünürlük dengesini birlikte optimize eden verimli bir mimaridir. ImageNet üzerinde ön eğitimli ağırlıklarla başlatılmış ve son sınıflandırma katmanı iki sınıfa göre değiştirilmiştir. Batch boyutu 16, öğrenme hızı 5e-5, ağırlık azalması 1e-4 ve epoch sayısı 20 olarak belirlenmiştir. Model boyutu 67,43 MB olup ana veri setinde %99,60 doğruluk, 0,9999 AUC ve 0,995989 F1-skoru elde etmiştir. Bu sonuçlar EfficientNet-B4'ün yüksek ayrım gücü sunduğunu göstermektedir.

**Tablo 2. Harici veri seti (DF40) üzerinde model performans karşılaştırması**




Harici test sonuçları, aynı veri dağılımında elde edilen yüksek doğrulukların gerçek dünya benzeri örneklerde korunup korunmadığını görmek için ayrı olarak incelenmiştir. Bu değerlendirmede özellikle fake recall değeri kritik kabul edilmiştir; çünkü sahte bir görselin gerçek olarak sınıflandırılması uygulama açısından yüksek riskli bir hatadır.

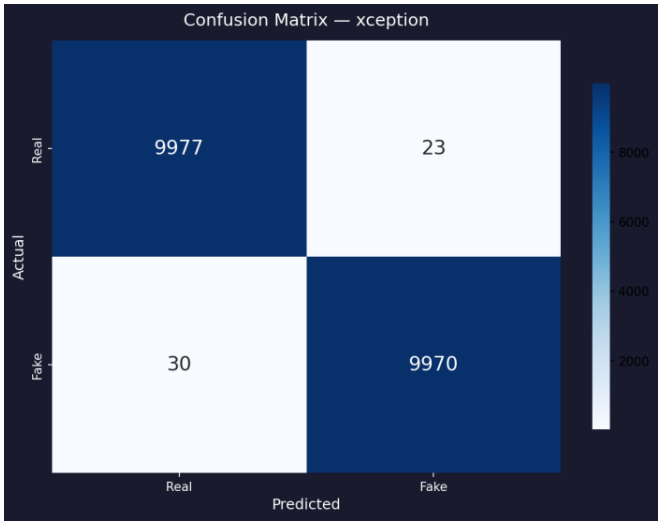



*Şekil 8. EfficientNet-B4 modeli karışıklık matrisi — ana test seti (%99.60 doğruluk)*

## Xception

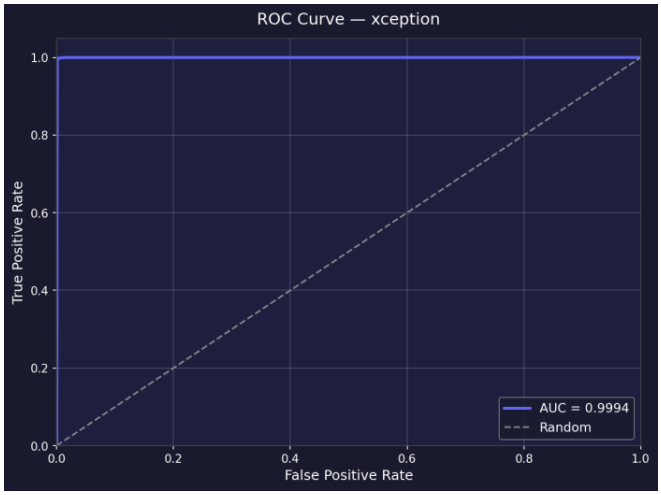
Xception, depthwise separable convolution kullanarak standart konvolüsyon işlemini ayrıştıran güçlü bir CNN mimarisidir. ImageNet ön eğitimli ağırlıklarla başlatılmış, son sınıflandırma katmanı iki sınıfa göre değiştirilmiş ve batch boyutu 16, öğrenme hızı 5e-5, ağırlık azalması 1e-4

ve epoch sayısı 20 olarak ayarlanmıştır. Model boyutu 79,60 MB olup ana veri setinde %99,73 doğruluk, 0,9994 AUC ve 0,9973 F1-skoru ile karşılaştırılan modeller arasında en yüksek kapalı veri seti doğruluğunu elde etmiştir.




*Şekil 9. Xception modeli karışıklık matrisi — ana test seti (%99.73 doğruluk)*

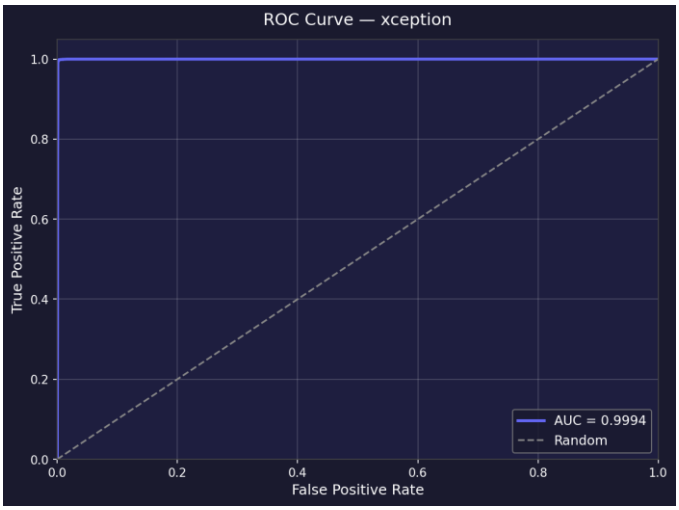
Buna karşın, aynı Xception modelinin DF40 tabanlı harici test setindeki performansı ciddi biçimde düşmüştür. Model harici testte yalnızca %52,74 doğruluk, 0,6141 AUC ve %6,04 sahte geri çağırım değeri üretmiştir. Bu sonuç, modelin ana veri seti dağılımında çok başarılı görünmesine rağmen dış veri dağılımlarında sahte örneklerin büyük bölümünü kaçırabildiğini ve veri setine özgü örüntülere aşırı bağımlı kalabildiğini göstermektedir.



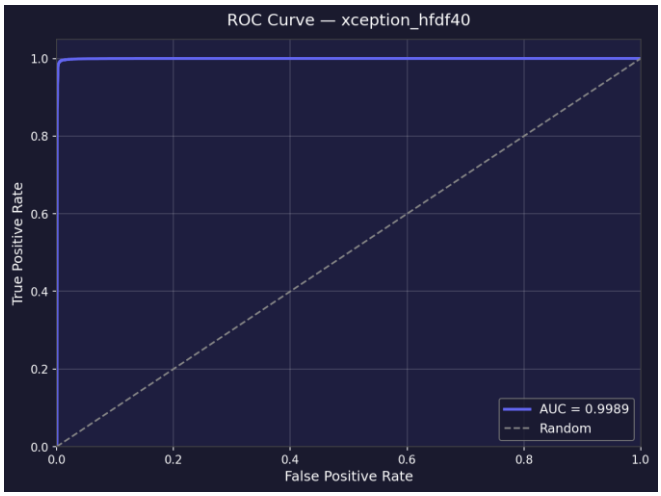

*Şekil 10. Xception modeli ROC eğrisi — ana test seti (AUC=0.9994)*

Bu bulgudan hareketle modelin genelleme başarımını artırmak amacıyla Xception mimarisi DF40 tabanlı harici veri setiyle yeniden eğitilmiş ve final model Xception+DF40 olarak adlandırılmıştır.

**Tablo 3. Xception+DF40 final modelinin harici veri setindeki performansı**



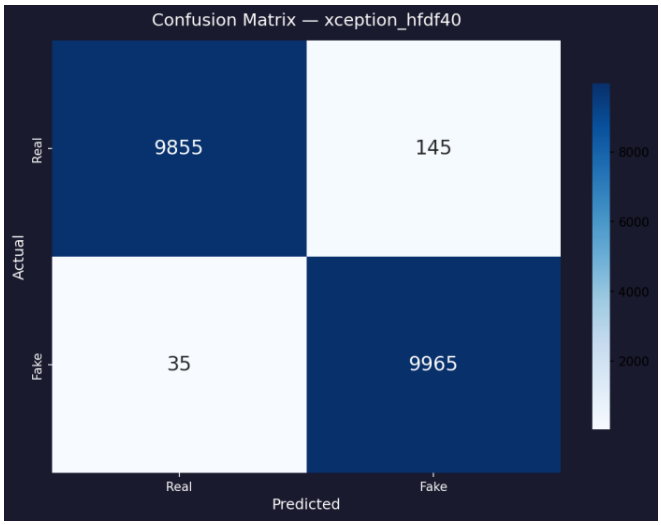

Tablo 3, orijinal Xception ile Xception+DF40 arasındaki dağılım dışı performans farkını özetlemektedir. Yeni model, ana veri setinde küçük bir doğruluk kaybı yaşamasına rağmen harici testte çok daha dengeli ve güvenilir sonuçlar üretmiştir.




*Şekil 11. Xception+DF40 modeli ROC eğrisi — harici DF40 test seti (AUC=0.9828)*

**Xception + DF40 (Final Model)**

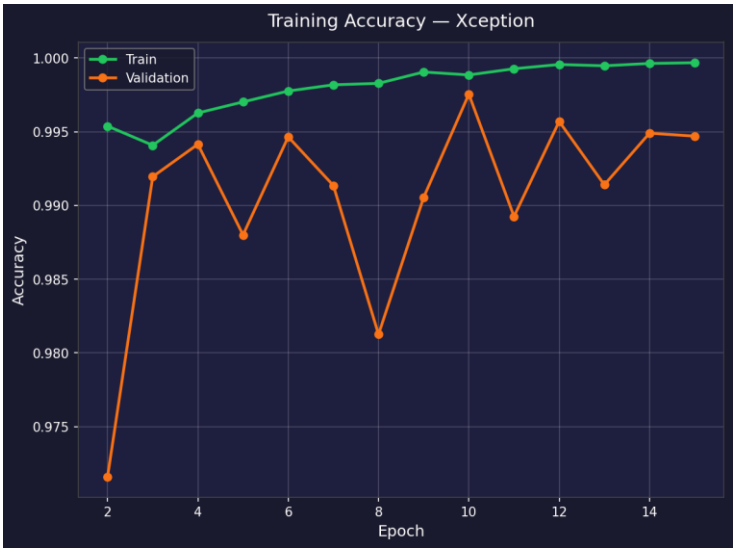
Xception+DF40, çalışmanın final modeli olup orijinal Xception modelinin harici DF40 tabanlı veriyle yeniden eğitilmesiyle elde edilmiştir. Model 8 epoch eğitilmiş, batch boyutu 16, öğrenme hızı 5e-5 ve ağırlık azalması 1e-4 olarak korunmuştur. Daha düşük epoch sayısı, güçlü bir başlangıç modelinden devam edilmesi ve erken durdurma mekanizmasıyla ilişkilidir. Model ana test setinde %99,10 doğruluk ve 0,9989 AUC elde ederken, harici test setinde %92,15 doğruluk, 0,9828 AUC ve %84,50 sahte geri çağırım değerine ulaşmıştır.




*Şekil 12. Xception+DF40 modeli karışıklık matrisi — ana test seti (%99.10 doğruluk)*

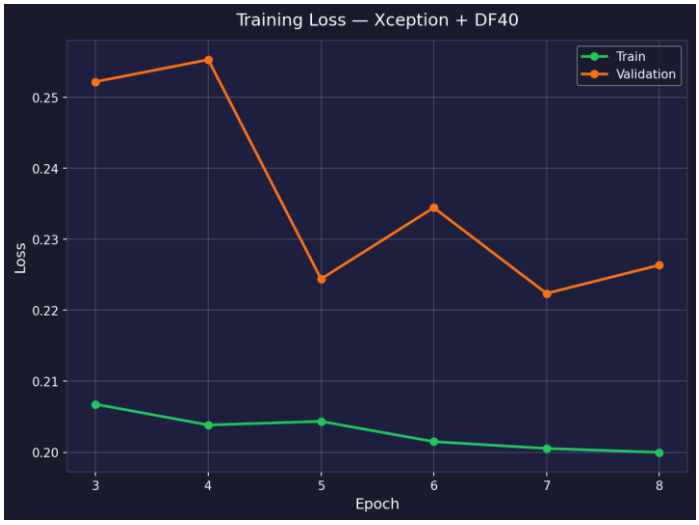
Eğitimler image_size=224, seed=42, CUDA destekli GPU, num_workers=4, pin_memory=True ve mixed precision ayarlarıyla yürütülmüştür. Kayıp fonksiyonu olarak CrossEntropyLoss(label_smoothing=0.1), optimize edici olarak AdamW ve zamanlayıcı olarak CosineAnnealingLR kullanılmıştır. Gradient clipping için max_norm=1.0 seçilmiş, doğrulama kaybı 5 epoch boyunca iyileşmediğinde erken durdurma uygulanmış ve en düşük doğrulama kaybına sahip checkpoint saklanmıştır. Büyük veri seti ve ağır mimariler nedeniyle final eğitim Google Colab Pro üzerinde Tesla T4 GPU ile gerçekleştirilmiştir.

**Tablo 4. Eğitim hiperparametreleri özeti**




Tablo 4'te model bazlı batch boyutu, öğrenme hızı, epoch sayısı ve ağırlık azalması değerleri verilmiştir. MesoNet için 64 batch ve 1e-3 öğrenme hızı; ResNet-50 için 32 batch ve 1e-4 öğrenme

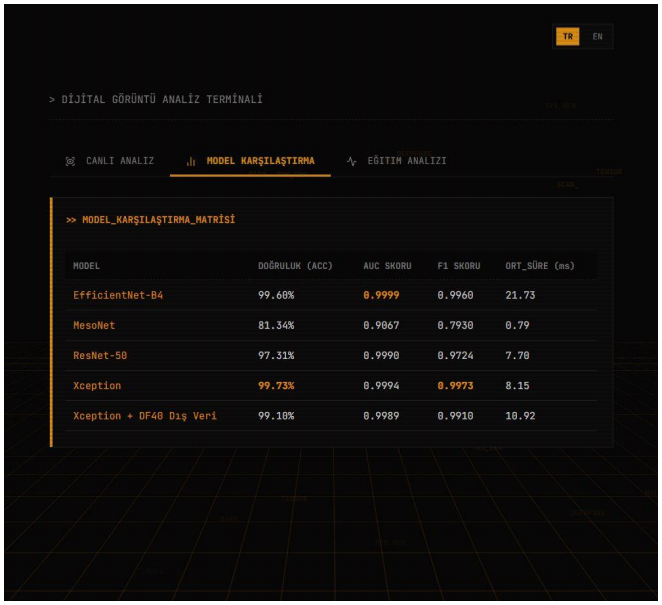
hızı; EfficientNet-B4, Xception ve Xception+DF40 için 16 batch ve 5e-5 öğrenme hızı kullanılmıştır. Tüm modellerde ağırlık azalması 1e-4 olarak sabit tutulmuştur.




*Şekil 13. Xception+DF40 modeli eğitim kaybı (Training Loss) eğrisi — train ve validation karşılaştırması*

## SONUÇLAR VE TARTIŞMA

Bu çalışma, deepfake görüntü tespiti için beş farklı derin öğrenme mimarisinin kapsamlı bir karşılaştırmasını sunmaktadır. Ana veri seti olan 140k Real and Fake Faces üzerinde en yüksek doğruluk %99,73 ile Xception modeli tarafından elde edilmiştir. EfficientNet-B4 %99,60, ResNet-50 %97,31 ve MesoNet %81,34 doğruluk sağlamıştır. Bununla birlikte bu yüksek kapalı veri seti sonuçları final model seçimi için tek başına yeterli görülmemiştir.




*Şekil 14. DeepFakeBusted web arayüzü — Model karşılaştırma sekmesi (tüm modellerin doğruluk, AUC ve F1 skorları)*

Çalışmanın en kritik bulgusu harici veri seti testinden elde edilmiştir. Orijinal Xception modeli ana test setinde %99,73 doğruluk sağlamasına rağmen DF40 tabanlı harici test setinde %52,74 doğruluğa gerilemiş, sahte geri çağırım oranı ise yalnızca %6,04 olarak ölçülmüştür. Bu oran, harici veri setindeki sahte görsellerin büyük bölümünün model tarafından gerçek olarak sınıflandırıldığını göstermektedir.

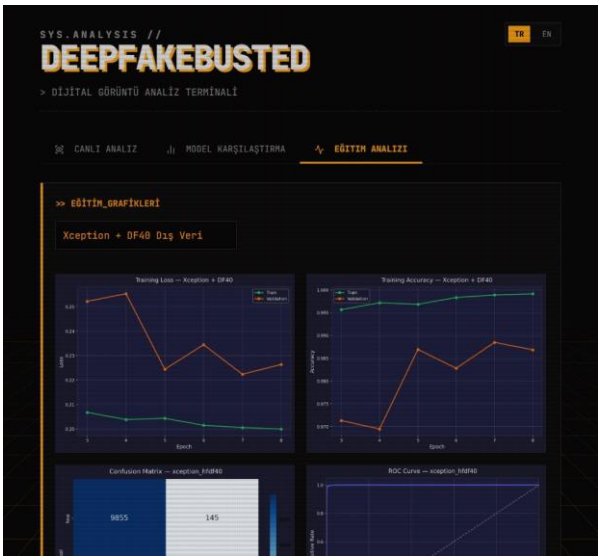
Xception+DF40 modeli harici veri setinde %92,15 doğruluk, 0,9828 AUC ve %84,50 sahte geri çağırım elde etmiştir. Ana veri setinde ise %99,10 doğruluk sağlamış, yani orijinal Xception modeline göre yalnızca %0,63 puanlık bir kapalı veri seti kaybı yaşamıştır. Bu denge, küçük bir in-domain kayıp karşılığında büyük bir out-of-domain kazanım elde edildiğini ve veri çeşitliliğinin deepfake tespitinde genelleme başarımı için belirleyici olduğunu göstermektedir.

Değerlendirmede doğrulukla birlikte AUC-ROC, F1-skoru, kesinlik ve geri çağırım metrikleri kullanılmıştır. Deepfake tespitinde özellikle sahte örneklerin gerçek sanılması kritik bir hata olduğundan fake recall ayrıca incelenmiştir. Bu yaklaşım, tek metrikli değerlendirmelerin yanıltıcı olabileceğini ve model seçiminin gerçek kullanım riskleriyle birlikte yapılması gerektiğini göstermektedir.

Model boyutu açısından MesoNet 0,09 MB ile çok hafif bir seçenek sunarken performansı güçlü CNN mimarilerinin gerisinde kalmıştır. EfficientNet-B4 67,43 MB ve Xception 79,60 MB boyutlarıyla yüksek başarı ve makul model büyüklüğü arasında iyi bir denge sağlamıştır. ResNet-50 ise 89,89 MB ile en büyük model olmasına karşın EfficientNet-B4 ve Xception kadar yüksek doğruluk üretememiştir.

Eğitim sürecinde label smoothing, AdamW, cosine annealing, gradient clipping ve erken durdurma gibi düzenlileştirme ve optimizasyon stratejileri kullanılmıştır. Bu yapı, modelin aşırı emin tahminler üretmesini azaltmayı ve doğrulama performansına göre en uygun checkpoint'i seçmeyi amaçlamaktadır. Dış veri eklendikten sonra doğrulama başarımının yüksek seviyede kalması, modelin yalnızca ana veri setini ezberlemek yerine daha çeşitli örüntülerden öğrenme yaptığını desteklemektedir.

Çalışmanın uygulama tarafında React + Vite arayüzü ve Flask backend ile çalışan bir prototip geliştirilmiştir. Kullanıcı tek bir görsel yükleyebilmekte, seçilen modelle tahmin alabilmekte, tüm modelleri aynı görsel üzerinde karşılaştırabilmekte ve eğitim analizi ekranında loss, accuracy, ROC eğrisi ve confusion matrix grafiklerini inceleyebilmektedir. Yüz kırpma adımı, gerçek kullanımda arka plan ve kompozisyon kaynaklı yanıltıcı etkileri azaltmak için sisteme eklenmiştir.




*Şekil 15. DeepFakeBusted web arayüzü — Eğitim analizi sekmesi (Xception+DF40 kayıp ve doğruluk eğrileri)*

Çalışmanın sınırlılıkları da dikkate alınmalıdır. Sistem video yerine görüntü tabanlı çalışmaktadır; Haar Cascade tabanlı yüz tespiti yan profil, düşük ışık veya kapalı yüz durumlarında hata yapabilir; deepfake üretim teknikleri sürekli değiştiği için modelin yeni ve çeşitli veri setleriyle düzenli olarak güncellenmesi gerekir. Ayrıca model açıklanabilirliği sınırlıdır ve kararın hangi görsel ipuçlarına dayandığını göstermek için Grad-CAM gibi yöntemlerle desteklenebilir. Genel olarak bu çalışma, deepfake görüntü tespitinde yalnızca kapalı veri seti başarımının değil, harici veri setlerinde genelleme kapasitesinin de temel değerlendirme ölçütü olması gerektiğini göstermiştir.

## KAYNAKÇA

Deng, J., Lin, C., Hu, P., Shen, C., Wang, Q., Li, Q., & Li, Q. (2024). Towards Benchmarking and Evaluating Deepfake Detection. IEEE Transactions on Dependable and Secure Computing. https://doi.org/10.1109/tdsc.2024.3369711

Dong, S., Wang, J., Ji, R., Liang, J., Fan, H., & Ge, Z. (2023). Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization. 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr52729.2023.00389

Fang, S., Zhang, Z., & Song, B. (2025). Deepfake Detection Model Combining Texture Differences and Frequency Domain Information. ACM Transactions on Privacy and Security.

Gong, L. Y., & Li, X. J. (2024). A Contemporary Survey on Deepfake Detection: Datasets, Algorithms, and Challenges. Electronics. https://doi.org/10.3390/electronics13030585

Huang, P., Han, Y., Chu, E., Chen, J., & Hua, K. (2023). Multi-Task Self-Blended Images for Face Forgery Detection. ACM Multimedia Asia 2023. https://doi.org/10.1145/3595916.3626426

Kingra, S., Aggarwal, N., & Kaur, N. (2025). Assessing deepfake detection methods: a comparative evaluation on novel large-scale Asian deepfake dataset. International Journal of Data Science and Analytics.

Kumar, M., & Verma, B. (2026). Performance Evaluation of Face Forgery Detection Models: A Comparative Study. Lecture Notes in Networks and Systems.

Li, J., Xie, H., Yu, L., Gao, X., & Zhang, Y. (2021). Discriminative Feature Mining Based on Frequency Information and Metric Learning for Face Forgery Detection. IEEE Transactions on Knowledge and Data Engineering. https://doi.org/10.1109/tkde.2021.3117003

Luo, A., Cai, R., Kong, C., Ju, Y., Kang, X., Huang, J. (2024). Forgery-aware Adaptive Learning with Vision Transformer for Generalized Face Forgery Detection. IEEE Transactions on Circuits and Systems for Video Technology. https://doi.org/10.1109/tcsvt.2024.3522091

Ma, Z., Mei, X., & Shen, J. (2023). 3D Attention Network for Face Forgery Detection. 2023 4th Information Communication Technologies Conference (ICTC). https://doi.org/10.1109/ictc57116.2023.10154671

Peng, C., Chen, T., Liu, D., Guo, H., Wang, N., & Gao, X. (2025). Revisiting face forgery detection towards generalization. Neural Networks.

Ramachandran, S., Nadimpalli, A. V., & Rattani, A. (2021). An Experimental Evaluation on Deepfake Detection using Deep Face Recognition. 2021 International Carnahan Conference on Security Technology (ICCST). https://doi.org/10.1109/iccst49569.2021.9717407

Sun, R., Zhao, Z., Shen, L., Zeng, Z., Li, Y., Veeravalli, B. (2023). An Efficient Deep Video Model For Deepfake Detection. 2023 IEEE International Conference on Image Processing (ICIP). https://doi.org/10.1109/icip49359.2023.10222682

Tian, C., Luo, Z., Shi, G., & Li, S. (2023). Frequency-Aware Attentional Feature Fusion for Deepfake Detection. ICASSP 2023. https://doi.org/10.1109/icassp49357.2023.10094654

Zhuang, W., Chu, Q., Yuan, H., Miao, C., Liu, B., & Yu, N. (2022). Towards Intrinsic Common Discriminative Features Learning for Face Forgery Detection Using Adversarial Learning. 2022 IEEE International Conference on Multimedia and Expo (ICME). https://doi.org/10.1109/icme52920.2022.9859586