

Yazılım Geliştirme Laboratuvarı - I

Ad Soyad: Anıl Ceco

Ad Soyad: Kerem Ünal

Numara 221307047

Numara: 221307052

I. PROJE KONUSU

Bu proje, on farklı sanatçının 50.000'den fazla eserinden oluşan geniş bir veri setini analiz ederek, bir sanat eserinin hangi sanatçıya ait olduğunu yüksek doğrulukla belirleyebilecek bir yapay zeka modeli geliştirmeyi amaçlamaktadır. Sanat dünyasında, eserlerin doğru şekilde sınıflandırılması ve sanatçılara özgü tarzların analiz edilmesi hem kültürel hem de akademik anlamda büyük bir öneme sahiptir.

Projede, derin öğrenme tabanlı transformer modelleri kullanılarak modern yapay zeka tekniklerinden faydalanılacaktır. Farklı transformer modelleriyle karşılaştırmalı analiz yapılacak ve her bir modelin performansı değerlendirilecektir. Veri seti, sanatçılara ait eserlerin görsel özelliklerini doğru şekilde algılayabilmesi için veri artırma ve boyutlandırma teknikleriyle optimize edilmiştir.

Proje süresince, Google Colab platformunun ücretsiz GPU olanakları kullanılarak ortak bir eğitim süreci yürütülmekte olup, model geliştirme süreci, optimize edilmiş bir eğitim stratejisiyle desteklenmektedir. Nihai hedefimiz, hem akademik hem de endüstriyel kullanım için etkili bir yapay zeka sınıflandırma çözümü geliştirmek ve sanat eserlerinin analizine yönelik inovatif bir yaklaşım sunmaktır.

I. GOOGLE COLAB

Bu projede, yapay zeka modellerimizin geliştirilmesi ve eğitimi için Google Colab platformu kullanılmıştır. Google Colab, ücretsiz GPU ve TPU kaynakları sağlayarak, yüksek hesaplama gücü gerektiren derin öğrenme projelerinde etkili bir çalışma ortamı sunar. Colab'ın sunduğu bu avantajlar, özellikle kişisel donanım kaynaklarının sınırlı olduğu durumlarda büyük bir avantaj sağlamaktadır.

II. VERİ ARTTIRMA İŞLEMLERİ

Bu proje kapsamında, derin öğrenme modellerinin performansını artırmak ve daha genelleştirilebilir bir model elde etmek için veri artırma (data augmentation) teknikleri kullanılmıştır. Veri artırma, mevcut veri setini çeşitli yöntemlerle çoğaltarak modelin farklı varyasyonları öğrenmesini sağlar. Bu projede kullanılan veri artırma teknikleri ve işlevleri şu şekildedir:

Kullanılan Veri Artırma Teknikleri

- **Multiply** ((0.6, 1.3))
Görsellerin parlaklık değerleri rastgele çarpanlarla değiştirilir. Çarpan 0.6 ile 1.3 arasında seçilerek, görsellerin parlaklıkları artırılır veya azaltılır. Bu, farklı aydınlatma koşullarına karşı modelin dayanıklılığını artırır.
- **LinearContrast** ((0.6, 1.4))
Görsellerin kontrast oranları rastgele ayarlanır.
- 0.6 değeri kontrastı azaltırken, 1.4 değeri kontrastı artırır. Bu işlem, modelin farklı

- kontrast seviyelerindeki görselleri tanıma yeteneğini geliştirir.
- **AddToHueAndSaturation** ((-15, 15))
Görsellerin renk tonu (hue) ve doygunluk (saturation) değerleri rastgele değiştirilir. Bu işlem, görsellerdeki renk varyasyonlarını artırarak, modelin farklı renk tonlarına uyum sağlamasını kolaylaştırır.
- **Flip** (0.5)
Görseller yatay ekseninde çevrilir. Bu işlem, görsellerin %50 olasılıkla simetrik versiyonlarını oluşturarak, modelin simetri durumlarını öğrenmesini sağlar.
- **PerspectiveTransform** (scale=(0.03, 0.09))
Görsellerin perspektifi rastgele değiştirilir. 0.03 ile 0.09 arasında bir ölçek değeri kullanılarak, görsellerin farklı açılardan görünümü oluşturulur. Bu, modelin perspektif değişikliklerine karşı dayanıklılığını artırır.
- **AdditiveGaussianNoise** (scale=(0, 0.03*255))
Görsellere rastgele Gauss gürültüsü eklenir. Gürültü seviyesi 0 ile 0.03*255 arasında ayarlanır. Bu teknik, görsellerdeki küçük bozulmalara karşı modelin dayanıklılığını artırmayı hedefler.

III. VERİ YÜKLEME VE NORMALİZASYON

Projede kullanılan veri setleri, belirtilen bir dizin altında train (eğitim), val (doğrulama) ve test (test) olmak üzere üç ana klasör halinde organize edilmiştir. Görseller, PyTorch kütüphanesinin datasets.ImageFolder sınıfı aracılığıyla yüklenmiştir. Bu yöntem, dosya dizin yapısına dayalı olarak veri örneklerini otomatik şekilde sınıflandırarak hızlı ve etkili bir veri yükleme sağlar.

Yüklenen görsellerin modelle uyumlu hale getirilmesi için bir dizi dönüşüm uygulanmıştır. Bu dönüşümler sırasında görseller, model giriş boyutlarına uyum sağlamak amacıyla 224x224 piksel boyutuna yeniden ölçeklendirilmiş, ardından tensor formatına dönüştürülmüş ve normalize edilmiştir. Normalizasyon işlemi, verilerin belirli bir aralıkta ölçeklenerek modelin daha etkili öğrenmesini yardımcı olmuştur. Kullanılan değerler, kırmızı, yeşil ve mavi (RGB) kanalları için ortalama standart sapma ayarlanmıştır.

Veri yükleme süreci, mini partiler (batch) halinde işlem yapmayı sağlayacak şekilde tasarlanmıştır. Eğitim veri yükleyicisi, verilerin sırasını karıştırarak (shuffle) modelin her eğitim adımında farklı veri kombinasyonlarını görmesini sağlamış; doğrulama ve test yükleyicileri ise veri sırasını sabit tutarak değerlendirme süreçlerinde tutarlılığı korumuştur. Mini parti boyutu (batch size) olarak 128 seçilmiştir.

Son olarak, veri setindeki toplam örnek sayıları hesaplanmış ve eğitim sürecinde kullanılmak üzere her sınıfın isimleri belirlenmiştir. Bu işlem sonucunda veri setleri eğitim için

hazır hale getirilmiş ve aşağıdaki sınıflar tanımlanmıştır: {sınıflar}. Bu yapılandırma, modelin etkili şekilde eğitilmesi ve değerlendirilmesi için gereken temel veri işleme adımlarını kapsamaktadır.

IV. TRANSFORMATÖR MODELLERİ

A. ViT

Vision Transformer (ViT), görsel veriler üzerinde çalışan derin öğrenme modelleri için geliştirilmiş bir transformer tabanlı mimardır. Geleneksel olarak görüntü işleme problemlerinde Convolutional Neural Networks (CNN) kullanılırken, ViT tamamen transformer bloklarına dayalıdır ve dikkat mekanizmalarını kullanarak görüntüleri işler.

ViT'nin temel özellikleri şunlardır:

- **Görüntülerin Parçalara Ayrılması:** Görüntü, belirli bir boyutta (örneğin, 16x16) küçük parçalara (patch) bölünür ve her bir parça bir dizi girdisi olarak işlenir.
- **Pozisyon Kodlaması:** Görüntü parçalarının konum bilgisi korunarak modele aktarılır.
- **Dikkat Mekanizmaları:** Transformer blokları aracılığıyla görüntü üzerindeki tüm parçalar arasında ilişkiler öğrenilir.
- **Son Katman:** Tüm öğrenilen özellikler, bir sınıflandırma katmanına aktarılır ve çıktı alınır.

Bu kod, bir derin öğrenme modelinin eğitim ve doğrulama aşamalarını düzenli bir şekilde gerçekleştirmek için tasarlanmıştır. Eğitim ve doğrulama işlemleri, belirli bir sayıda dönem (epoch) boyunca verisetinin özelliklerini öğrenmek ve modelin performansını değerlendirmek amacıyla yapılır.

Eğitim sürecinde, model train moduna alınır ve her veri partisi (batch) için:

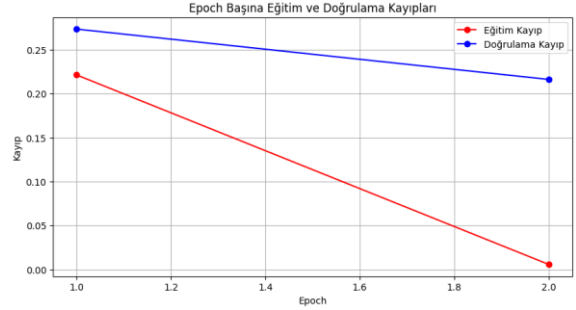
- Giriş verisi ve etiketler modele yüklenir.
- Model tahminler yapar ve bu tahminlerle gerçek etiketler arasındaki kayıp hesaplanır.
- Geri yayılım (backpropagation) yöntemi ile modelin ağırlıkları güncellenir.
- Her birkaç adımda (örneğin, 10 batch'te bir), geçici kayıp ve doğruluk bilgileri ekrana yazdırılır. Bu süreç sonunda, o döneme ait ortalama kayıp ve doğruluk değeri hesaplanır ve kayıp altına alınır.
- Doğrulama sürecinde ise model eval moduna alınır ve model, doğrulama veri kümesindeki tahminler üzerinde kayıp ve doğruluk hesaplar.
- Ağırlıklar güncellenmez, sadece performans ölçümü yapılır.
- Eğitim ve doğrulama aşamaları, belirlenen sayıda dönem boyunca tekrarlanır. Her dönem sonunda, o döneme ait kayıp ve doğruluk değerleri ekrana yazdırılır. Eğitim sonunda, toplam süre hesaplanır ve sürecin tamamlandığı belirtilir.

TEST SONUÇLARI:

Model eğitimi toplamda 42 dakika sürdü. Her 10 batch (1 batch 128 görüntü) adımda doğruluk ve kayıp verilerini görebiliyoruz. %40 civarlarında başlayan eğitim %99'lara kadar çıkarak tamamlandı.

```
Batch 370/425: Eğitim Kayıp: 0.0018, Batch Doğruluk: %100.00
Batch 380/425: Eğitim Kayıp: 0.0030, Batch Doğruluk: %100.00
Batch 390/425: Eğitim Kayıp: 0.0022, Batch Doğruluk: %100.00
Batch 400/425: Eğitim Kayıp: 0.0020, Batch Doğruluk: %100.00
Batch 410/425: Eğitim Kayıp: 0.0039, Batch Doğruluk: %100.00
Batch 420/425: Eğitim Kayıp: 0.0034, Batch Doğruluk: %100.00
Epoch Eğitim Kayıp: 0.0056, Epoch Eğitim Doğruluk: %99.92
Batch 10/12: Doğrulama Kayıp: 0.1288, Batch Doğruluk: %94.53
Epoch Doğrulama Kayıp: 0.2162, Epoch Doğrulama Doğruluk: %94.21
Toplam Eğitim Süresi: 2566.50 saniye
Eğitim tamamlandı
```

Loss grafiği:



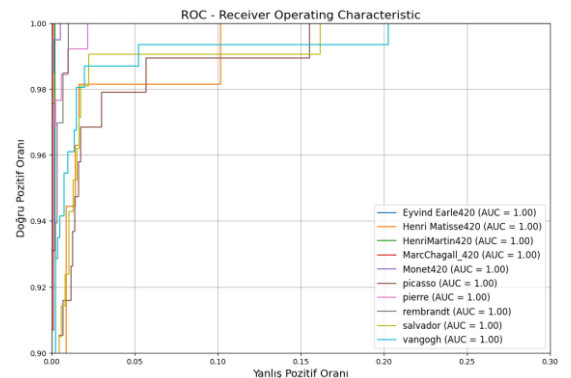
- Doğruluk (Accuracy): Modelimiz, test verisinde toplamda %95.09 doğruluk oranı elde etmiştir.
- Kesinlik (Precision): Modelin pozitif sınıfı tahmin ederken doğruluğu %95.24 olarak hesaplanmıştır.
- Hatırlama (Recall): Modelin, tüm gerçek pozitif örnekleri tespit etme oranı ise %95.09'dur.
- Özgüllük (Specificity): Modelin negatif sınıfı doğru bir şekilde tanımlama oranı %99.45 olarak hesaplanmıştır.
- F-Score (F1-Score): Precision ve recall'un harmonik ortalaması olan F1-Score değeri %95.10 olarak hesaplanmıştır.
- AUC (Area Under the Curve): Modelimizin ROC eğrisinin altındaki alan %99.85 olarak bulunmuştur.

```
Accuracy: %95.09
Precision: %95.24
Recall: %95.09
Specificity: %99.45
F-Score: %95.10
AUC: %99.85
```

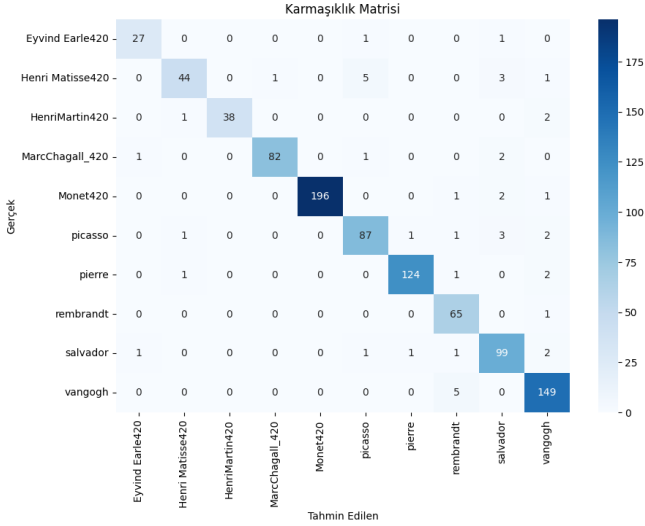
- Çıkarım zamanı: 5.36 saniye

İstenilen diğer test sonuçları da şu şekildedir:

ROC Eğrisi:



Karmaşıklık Matrisi:



B. PiT

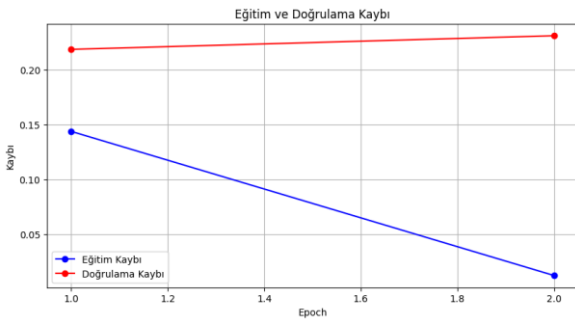
Pit base modeli, timm kütüphanesinin create_model fonksiyonu ile pit_b_224 adıyla seçilmiştir. Bu model, önceden eğitilmiş bir model olarak yüklenmiştir (pretrained=True).

Son Katman Düzenlemesi: Modelin son katmanı, 10 sınıflı sınıflandırma yapabilmesi için nn.Linear katmanı ile değiştirilmiştir. Bu işlem, modelin orijinal son katmanının çıktı özellik sayısına göre yeniden yapılandırılmasını sağlamıştır.

Her epoch boyunca, model train veri yükleyicisi kullanılarak eğitilmektedir. Eğitim sırasında her bir batch için modelin çıktıları hesaplanır ve ardından doğru sınıf ile karşılaştırılarak doğruluk (accuracy) hesaplanır. Loss fonksiyonu olarak çapraz entropi (CrossEntropyLoss) kullanılır. Bu kayıp, modelin öğrenmesini sağlamak için geri yayılım (backpropagation) algoritması ile güncellenir.

Her 10 batch'te bir, batch başına kayıp (loss) ve doğruluk (accuracy) değerleri ekrana yazdırılır. Bu sayede eğitim süreci boyunca modelin performansı izlenebilir.

Loss Grafiği:



Diğer test parametreleri de şu şekildedir:

TEST SONUÇLARI:

Accuracy (Doğruluk): %95.20, model çoğu örneği doğru sınıflandırmış.

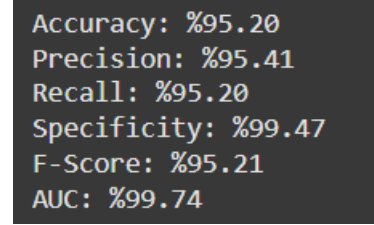
Precision (Kesinlik): %95.41, pozitif tahminlerin büyük çoğunluğu doğru.

Recall (Duyarlılık): %95.20, gerçek pozitifleri doğru şekilde tespit etmiş.

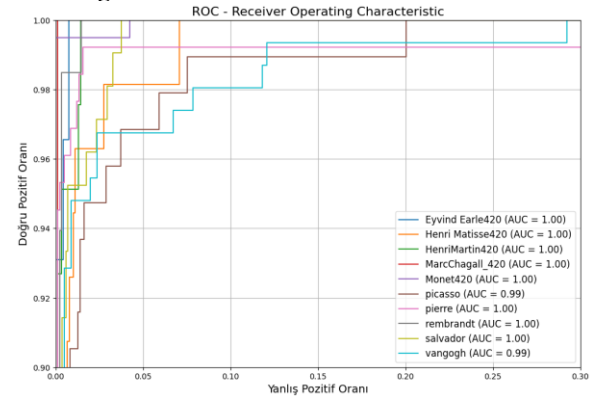
Specificity (Özgüllük): %99.47, negatif örneklerde yüksek doğruluk.

F-Score (F1 Skoru): %95.21, precision ve recall arasında iyi bir denge var.

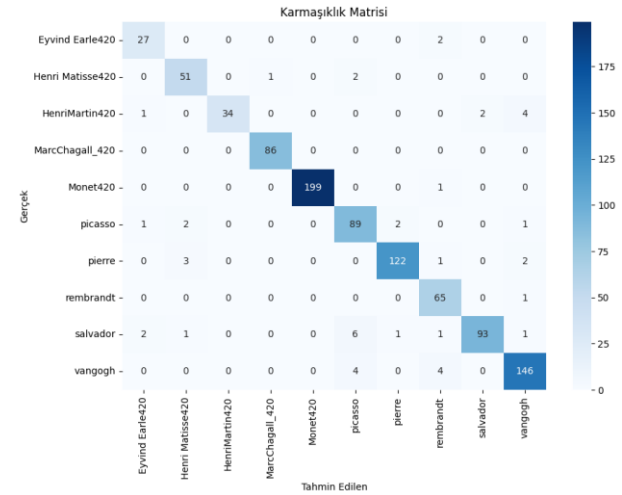
AUC: %99.74, modelin genel sınıflandırma performansı çok yüksek.



ROC Eğrisi:



Karmaşıklık Matrisi:

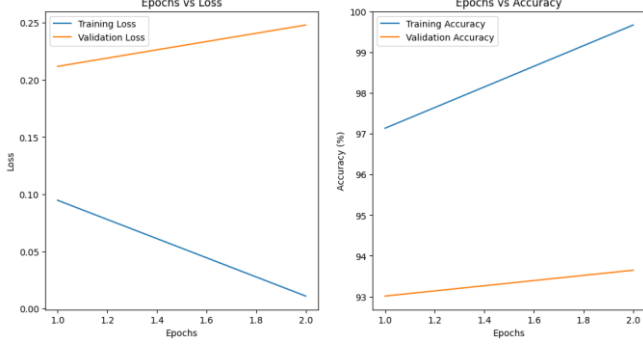


C. DeiT

DeiT (Data-efficient Image Transformers) modeli, görüntü sınıflandırma görevlerinde yüksek performans sağlayan, veri açısından verimli bir görsel transformatör modelidir. Transformer mimarisine dayanan DeiT, görüntüleri analiz etmek ve sınıflandırmak için özelleşmiş bir modeldir. Bu rapor, DeiT modelinin eğitim süreci, kullanılan veri seti ve model performansını değerlendirmektedir.

DeiT modelinin eğitim süreci, kayıp fonksiyonu olarak çapraz entropi kaybı ve optimizasyon algoritması olarak AdamW kullanılarak gerçekleştirilmiştir. Eğitim süresince modelin öğrenme süreci takip edilmiş ve her epoch sonunda kayıp ve doğruluk değerleri kaydedilmiştir.

Eğitim süreci tamamlandıktan sonra, eğitim ve doğrulama kayıpları ve doğrulukları grafikler üzerinde görselleştirilmiştir. Modelin performansı çeşitli metrikler kullanılarak değerlendirilmiş ve detaylı analizler yapılmıştır. Eğitim ve doğrulama süreçlerine ait grafikler şu şekilde çıkarılmıştır:



DeiT modelinin test veri seti üzerindeki performansını değerlendirmek amacıyla yapılan test sonuçları sunulmaktadır. Modelin doğruluğu, kesinliği, çağırısı, F1 skoru, özgüllüğü ve ROC gibi performans metrikleri hesaplanmış ve detaylı analizler yapılmıştır. Aşağıda DeiT modelinin eğitim süresi ve doğruluk oranları sonuçları görülmektedir.

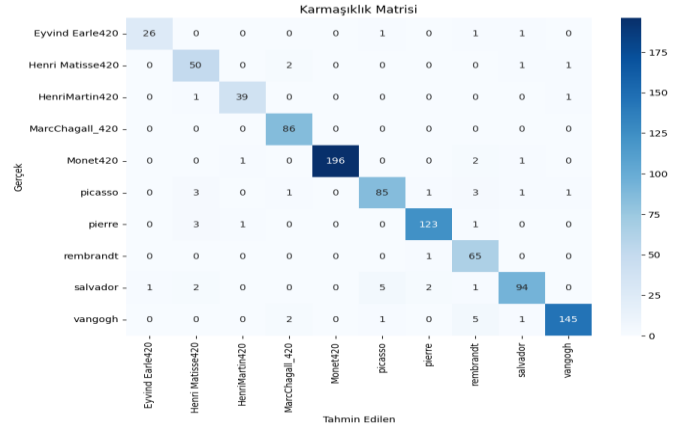
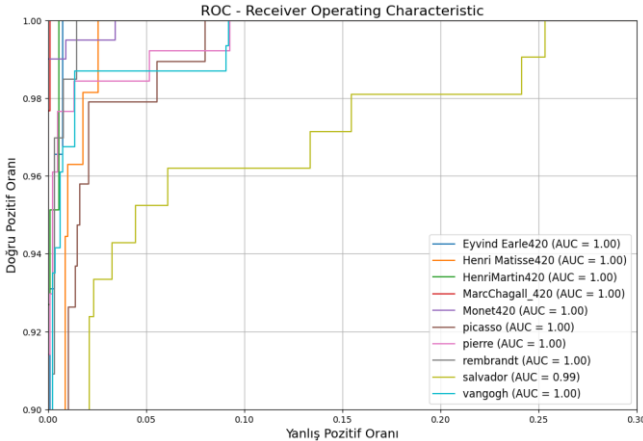
```
Batch 400/425: Train Loss: 0.0047, Batch Accuracy: 100.00%
Batch 410/425: Train Loss: 0.0023, Batch Accuracy: 100.00%
Batch 420/425: Train Loss: 0.0024, Batch Accuracy: 100.00%
Epoch Train Loss: 0.0111, Epoch Train Accuracy: 99.67%
Batch 10/12: Val Loss: 0.2669, Batch Accuracy: 93.75%
Epoch Val Loss: 0.2480, Epoch Val Accuracy: 93.65%
Total Training Time: 11161.37 seconds
Eğitim tamamlandı
```

Aşağıda deit modelinin istenilen test sonuçları verilmiştir.

```
Accuracy: %94.89
Precision: %95.12
Recall: %94.89
Specificity: %99.44
F-Score: %94.92
AUC: %99.78
```

Çıkarım Zamanı: 4.89 saniye

Modelin ROC eğrisi ve karışıklık matrisi de görüldüğü gibidir. Modelin eğitimi toplamda 11161.37 saniye yani 3 saat 10 dakika sürmüştür.

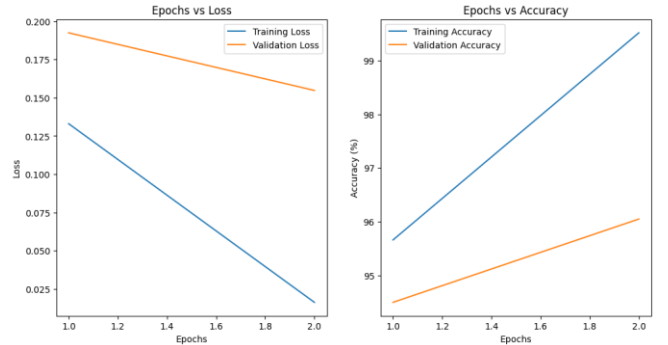


D. SwIN

Swin (Shifted Window) Transformer modeli, bilgisayarla görme alanında yüksek performanslı sonuçlar elde eden modern bir derin öğrenme modelidir. Swin Transformer, görüntüleri yerel ve küresel özellikleri etkili bir şekilde yakalamak için kaydırmalı pencere yöntemiyle işleyerek analiz eder.

Swin modelinin eğitim süreci, kayıp fonksiyonu olarak çapraz entropi kaybı ve optimizasyon algoritması olarak AdamW kullanılarak gerçekleştirilmiştir. Eğitim süresince modelin öğrenme süreci takip edilmiş ve her epoch sonunda kayıp ve doğruluk değerleri kaydedilmiştir.

Eğitim süreci tamamlandıktan sonra, eğitim ve doğrulama kayıpları ve doğrulukları grafikler üzerinde görselleştirilmiştir. Modelin performansı çeşitli metrikler kullanılarak değerlendirilmiş ve detaylı analizler yapılmıştır. Eğitim ve doğrulama süreçlerine ait grafikler şu şekilde çıkarılmıştır:

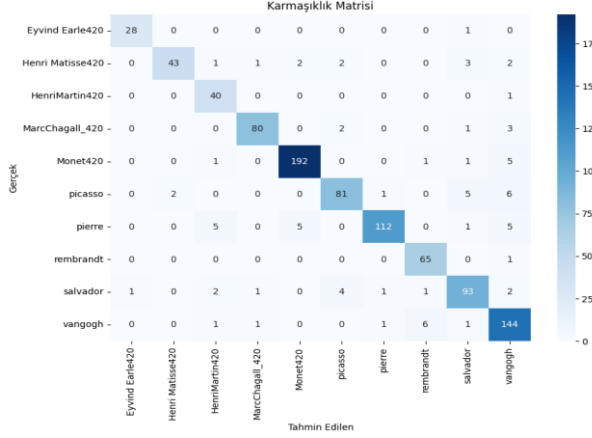
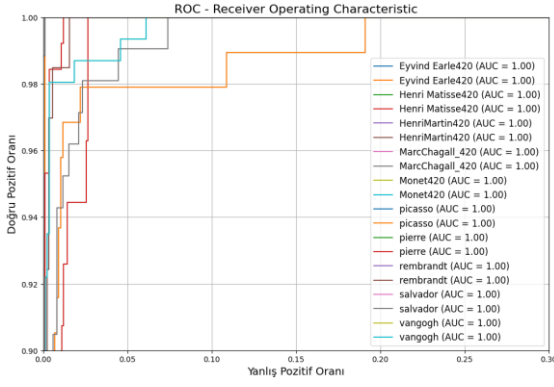


Swin modelinin test veri seti üzerindeki performans değerleri ile ilgili veriler ve grafikler aşağıdaki gibidir. Öncelikle aşağıda gördüğümüz değerler ortaya çıkmıştır.

```
Accuracy: %96.24
Precision: %96.26
Recall: %96.24
Specificity: %99.58
F-Score: %96.22
AUC: %99.88
```

Çıkarım Zamanı: 6.54 saniye

Ardından modelin ROC eğrisi ve karışıklık matrisi grafiği ortaya çıkarılmıştır.



Sonuç olarak modelin test verileri ile performansına bakıldığında 2epoch ile yüzde 96 başarı oranı sağladığı görülmektedir.

E. BEiT

Bu çalışmada, Microsoft'un BEiT (Bidirectional Encoder representation from Image Transformer) modelini kullandık. BEiT, görsel veriler üzerinde transformer tabanlı derin öğrenme yöntemleriyle etkili bir şekilde özellik çıkarımı yapar. Görsel verilerin temsili öğrenmek için bir dil modeli yaklaşımı kullanarak, görselleri anlamak için önceden eğitilmiş bir modeldir.

Model Yapılandırması:

Kullanılan model, BEiT Base'in önceden eğitilmiş versiyonudur. Model, 10 sınıf için sınıflandırma yapacak şekilde özelleştirilmiştir. BEiT, görsellerin temsili öğrenmek için AutoFeatureExtractor kullanarak görüntülerden anlamlı özellikler çıkarır. Bu modelin özellik çıkarıcı kısmı, görselleri uygun şekilde işleyerek derin öğrenme modeline uygun hale getirir.

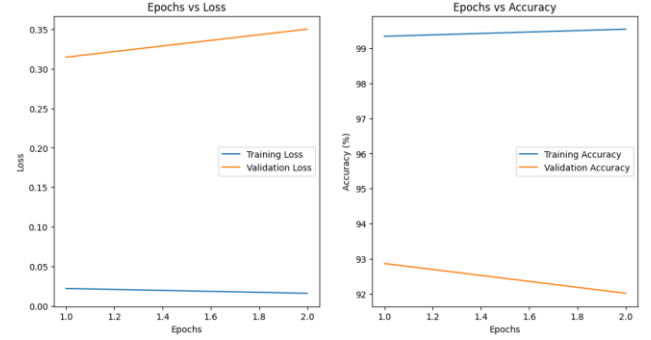
Cihaz Ayarları ve Eğitim Süreci:

Model, GPU'ya taşınarak hızlandırılmış eğitim süreci başlatılır. Modelin eğitimi için Cross-Entropy Loss kayıp fonksiyonu kullanılır. Bu fonksiyon, doğru sınıflandırma için gerekli olan hatayı minimize eder. Modelin parametrelerini optimize etmek amacıyla AdamW optimizasyon algoritması tercih edilir. Bu optimizatör, öğrenme hızını iyileştirerek modelin eğitim sürecini daha verimli hale getirir.

```
Batch 380/425: Train Loss: 0.0177, Batch Accuracy: 98.44%
Batch 390/425: Train Loss: 0.0147, Batch Accuracy: 99.22%
Batch 400/425: Train Loss: 0.0013, Batch Accuracy: 100.00%
Batch 410/425: Train Loss: 0.0088, Batch Accuracy: 100.00%
Batch 420/425: Train Loss: 0.0203, Batch Accuracy: 99.22%
Epoch Train Loss: 0.0158, Epoch Train Accuracy: 99.54%
Batch 10/12: Val Loss: 0.1273, Batch Accuracy: 96.09%
Epoch Val Loss: 0.3504, Epoch Val Accuracy: 92.03%
Total Training Time: 785.67 seconds
Eğitim tamamlandı
```

Ekstra epochtan ötürü sistemde 15 dakika olarak gözüксе de model eğitimi toplamda 50 dakika sürdü.

Loss Grafiği:

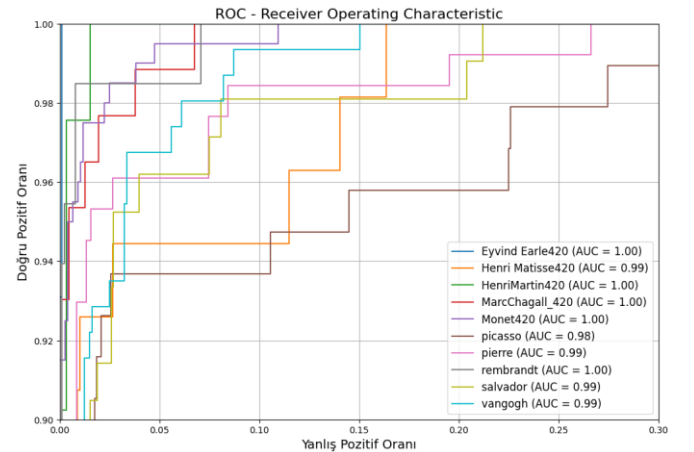


- Accuracy: %91 olarak hesaplanmıştır.
- Precision: %92 olarak hesaplanmıştır.
- Recall: %91 olarak hesaplanmıştır.
- Specificity: %99 olarak hesaplanmıştır.
- F-Score: %91 olarak hesaplanmıştır.
- AUC: %99 olarak hesaplanmıştır.

```
Accuracy: %91.65
Precision: %92.01
Recall: %91.65
Specificity: %99.05
F-Score: %91.66
AUC: %99.41
```

İstenilen diğer testler de aşağıdaki gibidir:

ROC Eğrisi:



V. SONUÇ

Sonuç olarak 5 farklı transformatör modeli ile eğitim yaparak güçlü sonuçlar elde ettik. 5 model de birbirinden farklı oranlarda başarı ile 10 adet sanatçının eserlerini %90'nın üzerinde bir doğrulukla tespit edebilmektedir. Görüntü işlemenin detaylarına yönelik güçlü bir çalışma olmuştur.

VI. KAYNAKÇA

- [1] [Vision Transformer \(ViT\)](#)
- [2] [BEiT](#)
- [3] [Swin Transformer](#)
- [4] [DeiT](#)
- [5] [Build a Deep CNN Image Classifier with ANY Images](#)
- [6] [Transformers for beginners | What are they and how do they work](#)