



T.C.
HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

YÜKSEK LİSANS TEZİ

DÖNÜŞTÜRÜCÜ KULLANARAK SANAT RESİMLERİNİN
SINIFLANDIRILMASI

NERGİZ MALAL

BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Şanlıurfa
2025



**T.C.
HARRAN ÜNİVERSİTESİ**

YÜKSEK LİSANS TEZİ

**DÖNÜŞTÜRÜCÜ KULLANARAK SANAT RESİMLERİNİN
SINIFLANDIRILMASI**

NERGİZ MALAL

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI
Tez Danışmanı: Dr. Öğr. Üyesi SERDAR ÇİFTÇİ**

**Şanlıurfa
2025**

TEŐEKKÜR

Tez sürecinin her aşamasında karşıma çıkan zorluklarda, sadece akademik bilgisiyle değil, insanlığıyla da yanımda olan; sabrı, anlayışı ve yönlendirmeleriyle bana rehberlik eden kıymetli danışmanım Dr. Öğr. Üyesi Serdar Çiftçi'ye en içten teşekkürlerimi sunuyorum.

Akademik yolda ilerleme motivasyonu sağlayan Arş. Gör. Neval Karaca hocama; yüksek lisans sürecimde yanımda olan değerli arkadaşım Esmâ'ya teşekkür ederim.

Bu araştırmada yer alan tüm/kısmi nümerik hesaplamalar, TÜBİTAK ULAKBİM Yüksek Başarım ve Grid Hesaplama Merkezi'nde (TRUBA kaynaklarında) gerçekleştirilmiştir.

Tez çalışmam süresince bana sabırla destek olan eşim Yakup Malal'a; özellikle bugünlere gelmemi sağlayan, en değerlilerim olan annem Asya ve babam Selahattin'e sonsuz teşekkür ederim.

İÇİNDEKİLER

ÖZET	i
ABSTRACT	ii
ŞEKİLLER DİZİNİ	iii
ÇİZELGELER DİZİNİ	iv
KISALTMALAR	v
1. GİRİŞ	1
2. ÖNCEKİ ÇALIŞMALAR	4
2.1. Makine Öğrenmesi Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları	4
2.2. CNN Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları	5
2.3. Dönüştürücü Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları	6
3. GEREÇ VE YÖNTEM	8
3.1. Veri Kümesi	8
3.1.1. WikiArt Veri Kümesinin Tanıtımı	8
3.1.2. WikiArt Veri Kümesinin Kapsamı ve Dağılımı	8
3.2. Veri Arttırma	10
3.3. Modeller	11
3.3.1. Dönüştürücü Tabanlı Modeller	11
3.3.1.1. ViT (Vision Transformer)	11
3.3.1.2. Swin Dönüştürücü	12
3.3.1.3. DeiT	12
3.3.1.4. BEiT	12
3.3.2. CNN Tabanlı Modeller	12
3.3.2.1. ResNet	13
3.3.2.2. MobileNet	13
3.3.2.3. EfficientNet	13
3.3.2.4. ConvNeXt	14
3.4. Hiperparametrelerin Belirlenmesi	14
4. BULGULAR	16
4.1. Model Eğitim Süreci	16
4.1.1. Değerlendirme Metodolojisi	16
4.2.1. Doğruluk (Accuracy)	16
4.2. Performans Değerlendirme Metrikleri	16
4.1.2. Deneysel Ortam	16
4.2.2. Kesinlik (Precision)	17
4.2.3. Duyarlılık (Recall)	17
4.2.4. F1-Skoru (F1-Score)	18
4.3. Deney Sonuçları	18
5. TARTIŞMA	26
6. SONUÇLAR	27
7. ÖNERİLER	28
KAYNAKLAR	29
ÖZGEÇMİŞ	33

ÖZET

YÜKSEK LİSANS TEZİ

DÖNÜŞTÜRÜCÜ KULLANARAK SANAT RESİMLERİNİN SINIFLANDIRILMASI

NERGİZ MALAL

HARRAN ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ
BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI

Tez Danışman: Dr. Öğr. Üyesi SERDAR ÇİFTÇİ

Yıl: 2025, Sayfa : 33

Bu tez çalışmasında, derin öğrenme yöntemleri kullanılarak sanat resimlerinin "Tür", "Sanatçı" ve "Stil" bazında sınıflandırma performansları incelenmiştir. Sanat resimlerinin sınıflandırılmasına yönelik çalışmalarda, konvolüsyonel sinir ağı (CNN) tabanlı yöntemler, özellik çıkarma konusundaki başarıları sayesinde alanda etkili sonuçlar vermiştir. Ancak yapılarında yer alan dikkat mekanizmaları sayesinde bilgisayarlı görü alanında daha yüksek performans sergileyen dönüştürücü (transformer) tabanlı modeller, son yıllarda bu alanda daha fazla tercih edilmiştir. Bu çalışmada CNN tabanlı ResNet, MobileNet, EfficientNet ve ConvNeXt ile dönüştürücü tabanlı ViT, Swin, BEiT ve DeiT modelleri kullanılarak sınıflandırma deneyleri gerçekleştirilmiştir. Deneylerde, çevrimiçi erişime açık WikiArt veri kümesi kullanılmıştır. "Sanatçı" ve "Tür" kategorilerinde en yüksek sınıflandırma doğrulukları sırasıyla %84,90 ve %79,52 ile BEiT modeliyle elde edilirken, "Stil" kategorisinde en iyi sonuç %72,59 doğrulukla Swin modeliyle elde edilmiştir. Dönüştürücü tabanlı yöntemlerin, CNN tabanlı yöntemlere kıyasla daha başarılı sonuçlar verdiği gözlemlenmiştir. Ayrıca tez kapsamında yapılan deneylerin literatürdeki benzer çalışmalarla karşılaştırılması sonucunda, dönüştürücü tabanlı modellerin sanat resimlerinin sınıflandırılmasında mevcut çalışmalara kıyasla daha yüksek performans sunduğu görülmüştür.

ANAHTAR KELİMELELER: Transfer Öğrenme, Resim Sınıflandırma, Görü Dönüştürücü

ABSTRACT

MASTER THESIS

CLASSIFICATION OF ART PICTURES USING TRANSFORMERS

NERGİZ MALAL

HARRAN UNIVERSITY

GRADUATE SCHOOL OF NATURAL AND APPLIED SCIENCES

COMPUTER ENGINEERING DEPARTMENT

Thesis Supervisor: Assist. Prof. Dr. SERDAR ÇİFTÇİ

Year: 2025, Page : 33

This thesis explores the classification performance of artworks based on three key attributes: "Genre," "Artist," and "Style," employing deep learning techniques. Convolutional neural network (CNN)-based methods have traditionally shown strong results in art classification tasks due to their effectiveness in feature extraction. However, transformer-based models, which utilize attention mechanisms, have recently gained prominence in the computer vision domain by demonstrating superior performance. In this study, classification experiments were conducted using several CNN-based models, including ResNet, MobileNet, EfficientNet, and ConvNeXt, alongside transformer-based models such as ViT, Swin, BEiT, and DeiT. These experiments utilized the publicly available WikiArt dataset. The highest classification accuracy in the "Artist" and "Genre" categories was achieved with the BEiT model at 84.90% and 79.52%, respectively, while the best result in the "Style" category was obtained with the Swin model at 72.59% accuracy. The findings indicate that transformer-based approaches consistently outperform their CNN counterparts. Additionally, when comparing these results with previous studies in the literature, it became evident that transformer-based models offer enhanced performance in the classification of artworks.

KEYWORDS: Transfer Learning, Painting Classification, Vision Transformer

ŞEKİLLER DİZİNİ

Şekil 1.1.	Sanatçı sınıflandırmasında BEiT modelinin tahmin sonucu.	2
Şekil 3.1.	WikiArt veri kümesinden bazı örnek resimler.	8
Şekil 3.2.	WikiArt veri kümesi'nde sanatçıya göre dağılım grafiği.	9
Şekil 3.3.	WikiArt veri kümesi'nde türe göre dağılım grafiği.	10
Şekil 3.4.	WikiArt veri kümesi'nde stile göre dağılım grafiği.	10
Şekil 4.1.	BEiT modelinin tür sınıfı karmaşıklık matrisi.	21
Şekil 4.2.	WikiArt veri kümesi için eğitim ve doğrulama kayıplarının epochlara göre değişimi.	22



ÇİZELGELER DİZİNİ

Çizelge 3.1. WikiArt veri kümesi sınıf bazında özellikleri.	9
Çizelge 3.2. Model eğitiminde kullanılan hiperparametreler	14
Çizelge 4.1. Bu çalışmada elde edilen dönüştürücü tabanlı modellerin performans sonuçları.	19
Çizelge 4.2. Bu çalışmada elde edilen CNN tabanlı modellerin performans sonuçları.	20
Çizelge 4.3. Dönüştürücü ve CNN tabanlı modellerin hiperparametre karşılaştırması	22
Çizelge 4.4. WikiArt veri kümesi kullanılarak sanatçı sınıfı üzerinde çalıştırdığımız BEiT ve Swin modellerinin sonuçları ile güncel çalışmaların sonuçları.	23
Çizelge 4.5. WikiArt veri kümesi kullanılarak stil, sanatçı ve tür sınıfları üzerinde çalıştırdığımız BEiT ve Swin modellerinin sonuçları ile güncel çalışmaların sonuçları	25



KISALTMALAR

BEiT : BERT Pre-trained Image Transformer

CNN : Convolutional Neural Network

ConvNeXt : Convolutional Network

EfficientNet : Efficient Network

GAN : Generative Adversarial Networks

ResNet : Residual Network

ResNet-50 : Residual Network with 50 layers

SWIN : Shifted Window Transformer

ViT : Vision Transformer



1. GİRİŞ

Sanat resimlerinin sınıflandırılması, insan kültürünün önemli bir parçasıdır. Sanat eserlerinin doğru şekilde sınıflandırılması ise hem tarihsel hem de estetik açıdan büyük önem taşır. Günümüzde resim sanatında dijitalleşmenin artmasıyla birlikte, büyük miktarda görsel verinin analizi zorlaşmıştır. Bu nedenle, yapay zeka ve makine öğrenimi teknikleri bu süreçte önemli bir rol oynamaya başlamıştır (Özdal, 2025). Özellikle derin öğrenme algoritmalarının gücü, resim sanatına ait eserlerin türlerine, sanatçılara ve stillerine göre doğru şekilde sınıflandırılmasında büyük avantaj sağlamaktadır (Saleh ve Elgammal, 2015).

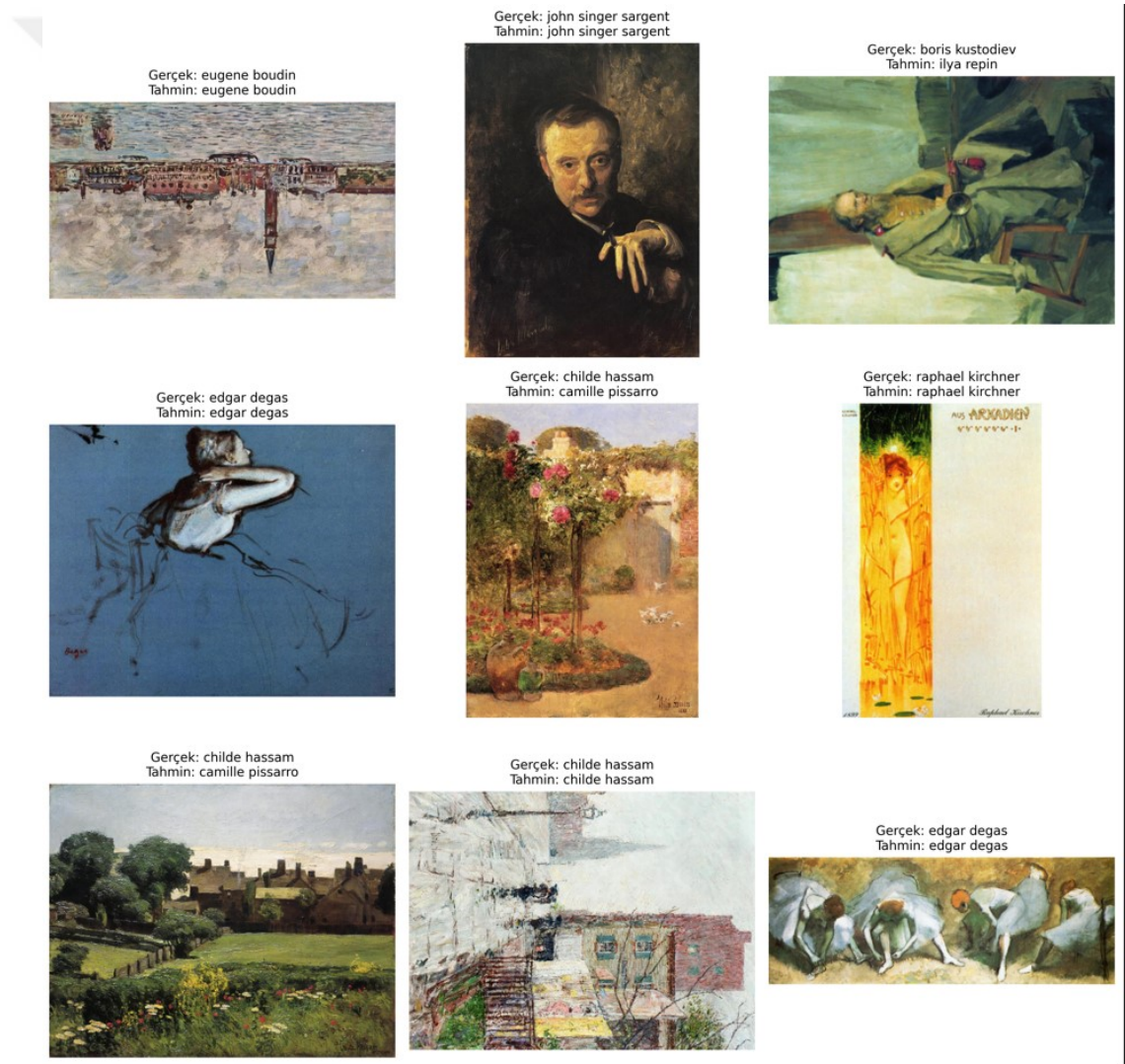
Son yıllarda sanat eserlerinin sınıflandırılması üzerine yapılan çalışmaların sayısı artmakla birlikte, bu çalışmalar genellikle sınırlı sayıdaki sanatçı veya tür üzerine odaklanmakta ve birkaç sanatçı üzerinde sınıflandırma yapılmaktadır. Ayrıca mevcut modellerin farklı sınıflandırma düzeylerindeki (sanatçı, tür, stil) dönüştürücü (Vaswani ve ark., 2017; Dosovitskiy, 2020) tabanlı derin öğrenme modelleri ile geleneksel konvolüsyonel sinir ağları CNN (LeCun ve ark., 1998) iki mimari arasındaki performans farklarını ortaya koyan güncel karşılaştırmalı çalışmaların literatürde sınırlı sayıda olduğu görülmüştür.

Bu tez çalışmasında, sanat resimlerinin sınıflandırılmasında derin öğrenme tabanlı dönüştürücü (Transformer) mimariler ile konvolüsyonel sinir ağları (CNN) kullanılarak, bu iki mimarinin başarıları karşılaştırmalı olarak incelenmiştir. Bu bağlamda, dönüştürücü tabanlı modellerin sanat resimlerinin sınıflandırılmasında CNN mimarilerine kıyasla daha yüksek doğruluk oranı sağlayıp sağlamayacağı ve bu modellerin hangi sınıflandırma türünde (sanatçı, stil, tür) daha başarılı olduğu araştırılmıştır. Her bir modelin farklı sınıflandırma görevlerindeki performansı karşılaştırılarak, mimarilerin sanat resimleri üzerindeki etkinliği değerlendirilmiştir.

Çalışmamızda BEiT (Bao ve ark., 2021), ViT (Dosovitskiy, 2020), Swin (Liu ve ark., 2021) ve DeiT (Touvron ve ark., 2021) gibi dönüştürücü tabanlı modeller ile ResNet (He ve ark., 2015), MobileNet (Howard ve ark., 2017), EfficientNet (Tan ve ark., 2019) ve ConvNeXt (Liu ve ark., 2022) gibi geleneksel CNN mimarileri kullanılmıştır. Sanat resimlerinin sınıflandırılması amacıyla seçilen bu derin öğrenme mimarileri, WikiArt veri kümesi (Tan, 2019) üzerinde eğitilmiş ve test edilmiştir.

Modellerin başarımı, doğruluk oranı (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi performans ölçütleri üzerinden karşılaştırmalı

olarak analiz edilmiştir. Şekil 1.1’de, kullanılan modeller arasında en iyi sonuçları veren BEiT'in sanat resimleri üzerinde "Sanatçı" sınıfında yaptığı tahmin sonuçlarını gösteren bir görsel yer almıştır. Görselde her sanat resmi için hem gerçek etiket hem de modelin tahmin ettiği etiketler gösterilmiştir. Bu tür görseller, modelin sınıflandırma yeteneğini ortaya koymuştur. Her bir görsel, BEiT modelinin bir sanat eseri için yaptığı tahminin doğruluğunu yansıtmıştır. Gerçek etiketler, eserin ait olduğu kategoriye; tahmin edilen etiketler ise modelin atadığı kategoriye belirtmiştir. Görsellere eklenen başlıklar, doğru ve yanlış sınıflandırılan örnekler arasında açık bir karşılaştırma yapılmıştır.



Şekil 1.1. Sanatçı sınıflandırmasında BEiT modelinin tahmin sonucu.

Tez çalışması altı bölümden oluşmaktadır. Birinci bölümde çalışmanın amacı,

kapsamı ve kullanılan yöntemler açıklanmıştır. İkinci bölümde konuyla ilgili literatür taraması yapılmış ve önceki çalışmalar değerlendirilmiştir. Üçüncü bölümde veri kümesi, kullanılan modeller ve yöntemler ayrıntılı şekilde sunulmuştur. Dördüncü bölümde deneysel sonuçlara ve bu sonuçlara dair karşılaştırmalı analize yer verilmiştir. Beşinci ve altıncı bölümlerde ise genel bir değerlendirme yapılmış ve gelecekte yapılabilecek çalışmalara yönelik öneriler sunulmuştur.



2. ÖNCEKİ ÇALIŞMALAR

Sanat resimlerinin sınıflandırılması, hem sanat tarihi hem de yapay zeka alanında önemli bir araştırma konusudur. Bu amaçla, sanatçı, tür ve stil gibi kategorilere ayrılmasını sağlayan çeşitli yöntemler geliştirilmiştir. Literatürdeki çalışmalarda, kullanılan yöntemlere göre makine öğrenmesi tabanlı, CNN tabanlı ve dönüştürücü tabanlı yaklaşımlar olarak sınıflandırılmıştır.

2.1. Makine Öğrenmesi Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları

Carneiro (2012) tarafından yapılan çalışmada, XV-XVII yüzyıllara ait 988 dini temalı sanat baskısından oluşan bir veri kümesi sunulmuş ve çeşitli yöntemlerle (Bag of Features, Label Propagation vb.) sanatsal görüntü anotasyonu ve geri getirme işlemleri değerlendirilmiştir. En iyi sonuç, geliştirilen "Inverted Label Propagation" yöntemiyle elde edilmiştir. Tan ve ark. (2019) ise koşullu generatif adversarial ağ (ArtGAN) modelini, CIFAR-10, STL-10, Oxford-102 ve CUB-200 veri kümesinde (Krizhevsky & Hinton, 2009) değerlendirmiştir. ArtGAN, CIFAR-10 üzerinde Inception skoru açısından mevcut en iyi sonuçları elde etmiştir. Choudhury (2020) sanat eserlerinin ressamalarını sınıflandırmak amacıyla WikiArt veri kümesi ile makine öğrenmesi ve derin öğrenme algoritmalarından yararlanmıştır. Çalışma, ResNet-50 transfer öğrenme yaklaşımını kullanarak %75 doğruluk elde etmiş ve sanat eserlerinden çıkarılan karmaşık özelliklerle ressam sınıflandırılmasını gerçekleştirmiştir.

Sandoval ve ark. (2021) ise insan etiketlemesi gerektirmeyen denetimsiz bir Karşıt Kümeleme Sistemi (ACS) önermiştir. ACS, denetimsiz kümeleme ve sınıflandırma modülleriyle sanat eserlerini daha yüksek güvenilirlikle etiketlemeyi başarmıştır. Abidin (2021) çalışmasında, Kaggle'dan elde edilen ve 50 sanatçının 8355 eserini içeren bir sanat eseri veri kümesi kullanılmış, 21 farklı türde eser içeren bu veri kümesinde AdaBoost algoritması en iyi sonuç veren yöntem olmuş ve %68.3 doğruluk değeri elde edilmiştir. Jeremiasz ve ark. (2022), WikiArt veri kümesini kullanarak sanat eserlerinden stil temsilleri ve renk dağılımları çıkararak büyük bir vektör tabanlı veri kümesi olan WikiArtVectors oluşturmuşlardır. Jin (2023) çalışmasında yapay zekâ (AI) ve makine öğrenmesinin (ML) tarihçesi ve gelişimini özetlemiş, özellikle derin sinir ağlarının (DNN) sanat alanındaki uygulamalarına odaklanmıştır. DNN mimarileri kullanılarak sanat eserleri üzerinde sınıflandırma ve stil tanıma işlemleri gerçekleştirilmiştir. Smith (2024), sanatçıya dayalı ince sanat sınıflandırmasının karmaşıklığını ele almış ve bu tür sınıflamaların karşılaştığı zorlukları aşmak için karmaşık yöntemler kullanmıştır. Çalışmada, yaygın olarak

kullanılan WikiArt veri kümesi kullanılmış ve sanatçılar %95 oranında filtrelenerek yalnızca en fazla esere sahip sanatçılarla sınırlı bir veri kümesi oluşturulmuştur.

2.2. CNN Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları

Agarwal ve ark. (2015), dijitalleştirilmiş resim koleksiyonlarının büyüklüğü arttıkça, bu koleksiyonlardan resimlerin düzenlenmesi ve erişilmesinde zorluk yaşandığına dikkat çekmiştir. Çalışma, resimler üzerinde özellik çıkarma ve türler ile stillerine göre sınıflandırma sorununu ele almış; altı tür ve on stil üzerinde %84.56 doğrulukla tür sınıflandırması ve %62.37 doğrulukla stil sınıflandırması elde edilmiştir. Tan ve ark. (2016) çalışmasında, büyük ölçekli sanat eserlerinin sınıflandırılması için Derin Konvolüsyonel Ağlar (Deep Convolutional Network) kullanılmıştır. Çalışmanın iki ana amacı bulunmuştur: birincisi, sanat eserlerinin sınıflandırılması için uçtan uca bir derin model eğitmek, ikincisi ise sanat eserlerinin sınıflandırılmasının nesne veya yüz tanımadan daha zorlu bir problem olduğunu savunmaktır. Çalışmada, büyük ölçekli WikiArt veri kümesi kullanılarak önerilen çözüm %68 başarı sağlamıştır. Jangtik ve ark. (2016), dijital sanat resimlerinin sanatçı sınıflandırılmasının zorluklarını ele alarak CNN-LSTM tabanlı bir model önermiş, yerel yama bilgileri kullanarak 13 sanatçı arasında çoklu sınıflandırma yapmıştır. Lecoutre ve ark. (2017), ResNet-50 kullanarak sanatsal stilin tespit edilmesi ve mevcut yöntemlerin aşılması amacını taşımıştır; WikiArt veri kümesinde (25 farklı stil için) %62 doğruluk elde edilmiştir. Cetinic ve ark. (2018), evrişimli sinir ağlarının (CNN) resim sanatları alanındaki sınıflandırma görevlerine uygulanabilirliğini inceleyerek, sanatçı (%81,94) ve tür (%77,60) sınıflandırma yapmış; stil sınıflandırmasında ise %56,43 doğruluk oranı elde etmiştir. Chu ve ark. (2018), sanat resimleri stil sınıflandırmasında derin korelasyon özelliklerini incelemiş ve WikiArt'dan toplanan 17 stil için 19.787 yağlı boya resim içeren veri kümesi kullanmıştır.

Oomen (2018), sanat resim sınıflandırması için transfer öğrenme kullanımını araştırmış ve önceden eğitilmiş CNN ağları kullanarak ağ katmanlarından çıkarılan özellik vektörlerinin birleştirilmesinin sınıflandırma doğruluğunu artırdığını göstermiştir. Yang ve ark. (2018), dijitalleştirilmiş sanat eserlerinin türünü ve dönemini sınıflandırmak için ResNet-50 modelini kullanmış ve tür sınıflandırmasında %87.8 doğruluk, dönem sınıflandırmasında ise %57.4 doğruluk elde edilmiştir. Sandoval ve ark. (2019) çalışmasında, sanat eserlerinin dijitalleştirilmesinin ardından sanat stiline doğruluğunu artırmak için iki aşamalı bir görüntü sınıflandırma yaklaşımı önermiştir. Cömert ve ark. (2021), sekiz ulusal

ressamın eserlerinden oluşan 504 resim ile ressam tanımlama işlemi gerçekleştirmiş ve MobileNet v2 modelinin %81 doğrulukla başarılı olduğunu bulmuştur. Menai ve Babahenini (2023), sanatsal tarz sınıflandırması için derin öğrenme yaklaşımlarını araştırarak, en iyi sonucu InceptionResNetV2 mimarisi kullanılmıştır. Imran ve ark. (2023), sanat eserlerini sanatsal stillerine göre sınıflandırmak için iki aşamalı bir yöntem önererek, WikiArt veri kümesi üzerinde %90,7 doğruluk elde etmiştir. Yu ve ark. (2024), sanat eserlerini sanatçıya göre sınıflandırmak için çok ölçekli piramit tabanlı bir CNN yaklaşımı önermiş ve WikiArt veri kümesinde %80 doğruluk elde edilmiştir. Zhang ve Ding (2024), medya resimlerinin stil sınıflandırması için geliştirilen ResNet50 modeline CELU aktivasyonu, bulanık havuzlama ve dikkat mekanizması entegre ederek %80,6 doğruluk elde etmiştir. Fortuna-Cervantes ve ark. (2024), ArtMD veri kümesinde üç CNN modelinin sanat medyası sınıflandırma performansını analiz etmiş ve çizim kategorisi çıkarıldığında %86 doğruluk elde etmiştir. Mezina ve Burget (2025), sanat eserlerinin dijitalleştirilmesiyle yeni gizli desenlerin keşfini ve EnsArtNet adı verilen bir CNN mimarisi geliştirilmesini amaçlamıştır; bu model, WikiArt veri kümesinde %84,93, En İyi Sanat Eserleri veri kümesinde ise %86,65 doğruluk elde etmiştir. Wang ve ark. (2025), dijital resim tarzı sınıflandırma için üç-dallı konvolüsyonel dikkat mekanizması ile iyileştirilmiş bir artıklık (residual) sinir ağı modeli geliştirmiş ve eski Yunan seramik tarzı için %100 doğruluk elde etmiştir.

2.3. Dönüştürücü Tabanlı Sanat Resimleri Sınıflandırma Çalışmaları

Krishna ve Kalluri (2019) çalışmasında, derin öğrenme ve transfer öğrenme yaklaşımlarının görüntü sınıflandırma üzerindeki etkisini ele almışlardır. Çalışma, popüler CNN mimarileri olan LeNet, AlexNet, GoogleNet, VGG16, VGG19 ve ResNet-50'nin transfer öğrenme yöntemleriyle kullanımını incelemiştir. Conde ve ark. (2021) sanat eserlerinin ince ayrıntılarını tanımda karşılaşılan zorluklara ve etiketli veri kümesi oluşturulmasının maliyetli olmasına dikkat çekmiştir. Bu çalışmada, CLIP (Contrastive Language-Image Pre-Training) kullanılarak sanat görüntüleri ve metin çiftleri üzerinde bir sinir ağı eğitilmiş ve model, önceden eğitilmeden (zero-shot) öğrenme yeteneği ile, görüntü için en uygun doğal dil açıklamasını tahmin edebilmiştir. Lu ve ark. (2021), CNN ve dönüştürücü tabanlı yöntemler (ViT) arasındaki verimliliği karşılaştırmıştır. 10.000 veri noktasından oluşan bir kelebek veri kümesi üzerinde yapılan deneyler, ViT modelinin büyük veri kümesi ile daha hassas sonuçlar verdiğini, ancak eğitim sürecinin daha maliyetli ve zaman alıcı olduğunu göstermiştir. ViT modelinin gösterdiği ve endüstri uygulamaları için iyileştirme potansiyeli sunduğu vurgulanmıştır. Zhao ve ark.

(2021) çalışmasında, sanat eserlerinin dijitalleştirilmesiyle birlikte transfer öğrenme yöntemleri kullanılarak sanat eserlerinin türleri, stilleri ve sanatçıları sınıflandırma performansları karşılaştırılmıştır. Sonuçlar, transfer öğrenme kullanımının sınıflandırma doğruluğunu artırdığını göstermiştir. Aynı araştırma ekibi (2022) yaptığı bir çalışmada, sanat resim sınıflandırmasında konvolüsyonel sinir ağlarının (CNN) etkinliğini incelemiş ve transfer öğrenme ile önceden eğitilmiş ağların daha iyi genelleme sağladığını bulmuştur. Sonuçlar, transfer öğrenme yönteminin sanat sınıflandırmasında önemli bir performans artışı sağladığını ortaya koymuştur. Liadis ve ark. (2021) çalışmasında, sanat eseri tanıma için ViT ve MLP-Mixer gibi iki derin öğrenme mimarisi kullanılmış ve her iki model sıfırdan eğitilerek WikiArt veri kümesinde 21 stil sınıfı üzerinde %39'un üzerinde doğruluk elde edilmiştir. Diem ve Mandl (2023) çalışmasında, küçük bir sanat tarihi veri kümesi üzerinde ViT ile ResNet-50'nin performansları karşılaştırılmıştır. Çalışma, 15. ile 19. yüzyıl arasındaki büyük bir portre koleksiyonu kullanarak, ViT'in %87.09 doğruluk oranı ile ResNet-50'nin %46.13 doğruluğunu geride bıraktığını göstermiştir. Schaerf ve ark. (2024) sanat eserlerinin doğrulama görevlerinde ViT'yi incelemiştir. Çalışma, Vincent van Gogh'a ait özgün resimler ve imitasyonlar içeren bir veri kümesinde Swin dönüştürücü modeli %85'in üzerinde doğrulukla EfficientNet'ten daha üstün performans gösterdiğini ortaya koymuştur. Liu ve ark. (2024) sanat resimlerinin daha doğru sınıflandırılması için konvolüsyonel transformatörler tabanlı bir tanıma sistemi geliştirmiştir. Transfer öğrenme stratejisi kullanılarak inşa edilen model, ResNet-50, VGG16, AlexNet ve ViT gibi önceden eğitilmiş ağlarla karşılaştırıldığında daha iyi performans göstermiştir ve konvolüsyonel dönüştürücülerin sanat eseri tanıma ve güvenliği alanında etkili bir yöntem olduğunu ortaya koymuştur.

3. GEREÇ VE YÖNTEM

3.1. Veri Kümesi

3.1.1. WikiArt Veri Kümesinin Tanıtımı

Veri kaynağı olarak, dijitalleştirilmiş sanat eserlerinin en büyük erişilebilir çevrimiçi arşivi olan WikiArt veri kümesi (Tan, 2019) kullanılmıştır. Koleksiyon, sanatçı bilgisi, sanatsal stil ve tür kategorizasyonu gibi çeşitli meta veri niteliklerini içerecek şekilde titizlikle hazırlanmıştır. WikiArt veri kümesinden bazı örnek resimler Şekil 3.1'de gösterilmektedir.



Artist: pierre auguste renoir
Genre: Cityscape
Style: Impressionism



Artist: claude monet
Genre: Cityscape
Style: Impressionism



Artist: camille pissarro
Genre: Cityscape
Style: Impressionism



Artist: claude monet
Genre: Cityscape
Style: Impressionism



Artist: john singer sargent
Genre: Cityscape
Style: Impressionism



Artist: claude monet
Genre: Cityscape
Style: Impressionism



Artist: john singer sargent
Genre: Cityscape
Style: Impressionism



Artist: nicholas roerich
Genre: Cityscape
Style: Art Nouveau



Artist: claude monet
Genre: Cityscape
Style: Impressionism

Şekil 3.1. WikiArt veri kümesinden bazı örnek resimler.

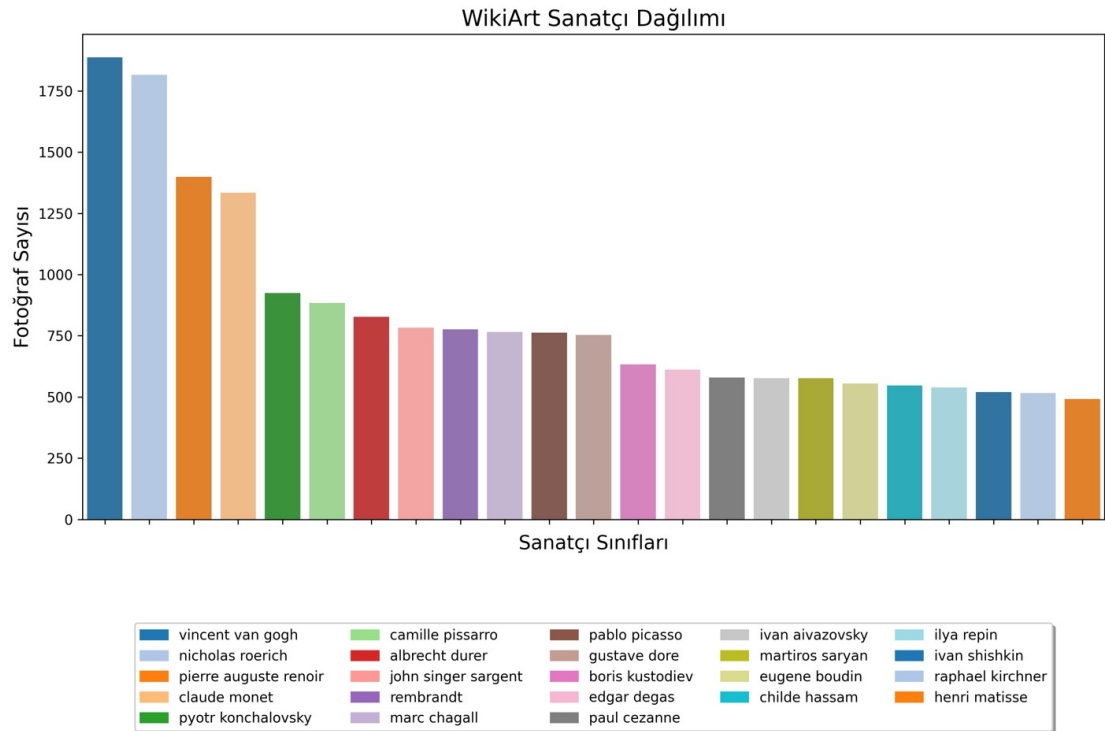
3.1.2. WikiArt Veri Kümesinin Kapsamı ve Dağılımı

Kullandığımız WikiArt veri kümesi, toplamda 80.735 farklı sanat eserini içermiştir. Sanatçı sınıflandırması için, her sanatçıya ait en az 850 resmin bulunduğu

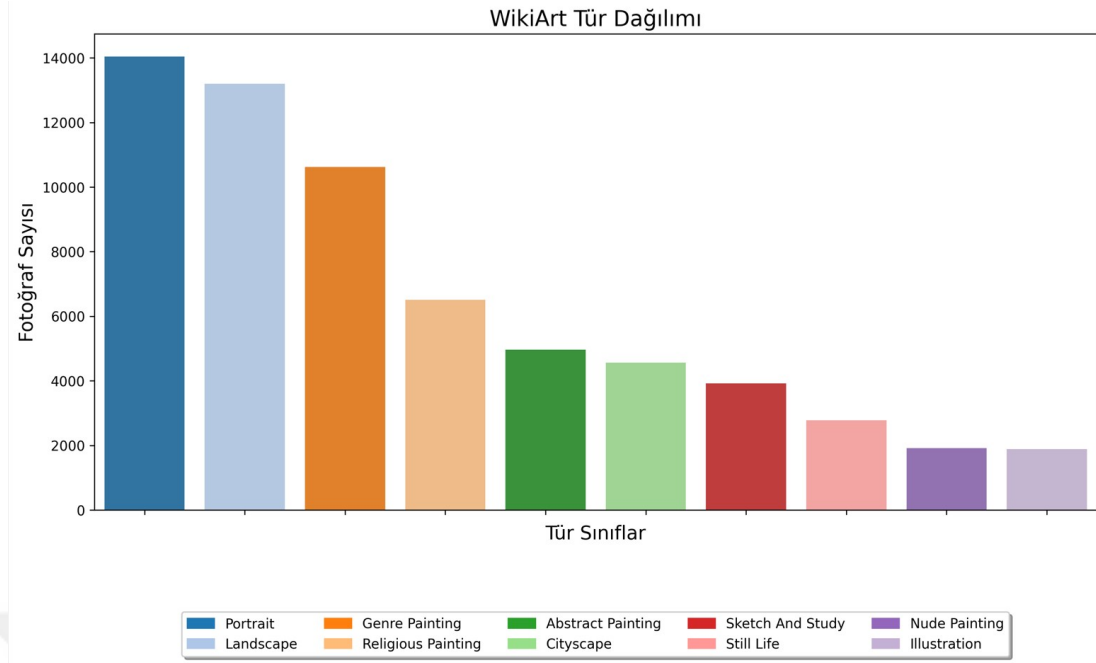
23 sanatçıdan oluşan bir alt küme kullanılmıştır. Stil sınıflandırmasına yönelik olarak, her biri 900'den fazla resim içeren 15 farklı sınıftan oluşan bir alt küme oluşturulmuştur. Tür sınıflandırması için ise her biri 1880'den fazla tablo içeren 10 sınıftan oluşan bir alt küme seçilmiştir. Çizelge 3.1'de WikiArt veri kümesindeki sınıf dağılımları ayrıntılı şekilde sunulmuştur. Sanatçı sınıfı dağılımı Şekil 3.2'de, tür sınıfı dağılımı Şekil 3.3'te ve stil sınıfı dağılımı ise Şekil 3.4'te gösterilmiştir

Çizelge 3.1. WikiArt veri kümesi sınıf bazında özellikleri.

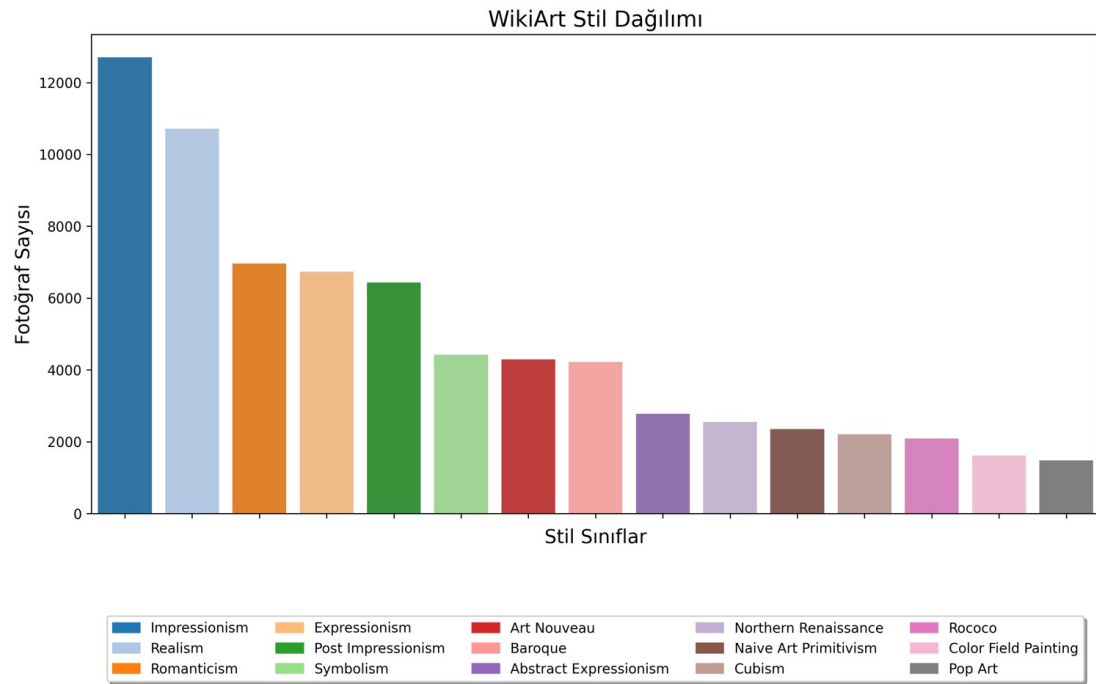
Kategori	Sınıf Sayısı	Resim Sayısı
Sanatçı	23	43.424
Tür	10	18.800
Stil	15	13.500



Şekil 3.2. WikiArt veri kümesi'nde sanatçıya göre dağılım grafiği.



Şekil 3.3. WikiArt veri kümesi'nde türe göre dağılım grafiği.



Şekil 3.4. WikiArt veri kümesi'nde stile göre dağılım grafiği.

3.2. Veri Arttırma

Veri ön işleme aşamasında, veri artırma (data augmentation) işlemleriyle sınıflar arasındaki dengesizlik giderilmeye çalışılmış ve modelin daha genel ve sağlam bir şekilde öğrenmesi sağlanmıştır. Veri artırma işlemleri kapsamında, görüntüler üzerinde rasgele dönüşümler uygulanarak çeşitlendirilmiş ve modelin farklı koşullarda öğrenebilmesi amacıyla zenginleştirilmiş bir veri kümesi

oluşturulmuştur. Uygulanan artırma işlemleri arasında rasgele 90 derece döndürme (RandomRotate90), yatay çevirme (HorizontalFlip), parlaklık ve kontrast değişimi (RandomBrightnessContrast), Gauss bulanıklığı (GaussianBlur) ve renk tonlarını değiştirme (HueSaturationValue) gibi teknikler yer almıştır. Bu işlemler, özellikle sınıf dengesizliğini azaltmak ve modelin daha çeşitli veriler üzerinde eğitilmesine olanak tanımak amacıyla gerçekleştirilmiştir.

3.3. Modeller

Bu bölümde, kullanılan farklı model türlerine ilişkin detaylı açıklamalara yer verilmiştir. Görsel sınıflandırma amacıyla başvuru alan modeller, dönüştürücü tabanlı ve CNN tabanlı olmak üzere iki ana kategori altında incelenmiştir. Model seçimi sürecinde veri kümesinin özellikleri, işlem süresi ve doğruluk oranı gibi çeşitli kriterler dikkate alınmıştır. Dönüştürücü tabanlı modeller, dikkat mekanizmaları aracılığıyla karmaşık ilişkileri öğrenebilme kapasitesi sunarken (Vaswani ve ark., 2017), CNN tabanlı modeller konvolüsyonel katmanlar yoluyla görsel özellikleri etkili bir biçimde çıkarabilmektedir (LeCun ve ark., 1998).

3.3.1. Dönüştürücü Tabanlı Modeller

Dönüştürücü tabanlı modeller, görsel veriler ve dil işleme görevlerinde büyük başarılar elde eden derin öğrenme modelleridir. Bu modellerin en önemli özelliklerinden biri, dikkat (attention) mekanizmalarını kullanmalarıdır. Dönüştürücü mimarisi, görsellerdeki pikseller arasındaki ilişkileri öğrenirken, tüm görüntüyü aynı anda işleyebilme yeteneğine sahiptir. Dönüştürücü tabanlı modeller, görüntüdeki her bir pikseli bağımsız olarak değil, tüm görüntüyü bir bütün olarak değerlendirerek, görselin genel özelliklerini daha etkili bir şekilde öğrenir. Bu yaklaşım, büyük veri kümelerinde paralel işlem yapma yeteneği sağlayarak, öğrenme sürecini daha hızlı ve verimli hale getirir (Dosovitskiy ve ark. 2020).

3.3.1.1. ViT (Vision Transformer)

Vision Transformer (ViT) (Dosovitskiy ve ark., 2020), geleneksel olarak dil işleme görevleri için geliştirilen dönüştürücü mimarisinin görsel veriler üzerinde uygulanmasıyla ortaya çıkan bir derin öğrenme modelidir. ViT, bir görüntüyü doğrudan piksel seviyesinde işlemek yerine, görüntüyü sabit boyutlu küçük parçalara (patch) böler ve her bir parçayı düzleştirerek bir vektöre dönüştürür. Bu vektörler, dönüştürücü modeline giriş olarak verilerek dikkat (attention) mekanizmaları aracılığıyla parçalar arasındaki ilişkiler öğrenilir. Böylece ViT, görüntünün genel

yapısını ve içeriğini, parçalar arasındaki bağlamı dikkate alarak anlamlandırır.

3.3.1.2. Swin Dönüştürücü

Swin dönüştürücü (Liu ve ark., 2021), görsel verileri işleme sürecinde daha verimli bir yaklaşım benimser. Bu verimlilik, görüntülerin küçük yerel parçalara ayrılması ve her bir parçanın özel dikkat katmanlarıyla işlenmesiyle sağlanır. İlk olarak, büyük görseller 16x16 piksellik küçük parçalara bölünür. Her bir parça, yerel dikkat katmanları aracılığıyla içindeki ilişkileri öğrenir. Kayan pencere tekniği kullanılarak her parça, hem kendi içindeki hem de komşularıyla olan ilişkileri anlamaya çalışır. Hiyerarşik yapı sayesinde, küçük parçalardan büyük ölçekli parçalara geçildikçe global dikkat katmanları devreye girer ve geniş alanlardaki ilişkileri öğrenir.

3.3.1.3. DeiT

DeiT (Touvron ve ark., 2021), ViT modelinin bir türevidir ve özellikle daha küçük veri kümeleri ile yüksek performans elde etmek amacıyla geliştirilmiştir. Bu model, görüntüleri küçük parçalara ayırarak işler. Her bir parça, özelliklerini taşıyan bir vektöre dönüştürülüp modelin girişine verilir. ViT'ten farkı, DeiT'in Teacher-Student yaklaşımını kullanmasıdır. Bu yaklaşımda, büyük ve önceden eğitilmiş bir model (teacher), öğrenci modeline (student) daha verimli öğrenmeyi öğretir. Öğretmen modeli, öğrenci modele doğru sınıflandırma bilgilerini aktarır ve öğrenci model, bu bilgileri taklit ederek öğrenir.

3.3.1.4. BEiT

BEiT (BERT Pretrained Image Transformer) (Bao ve ark., 2021), görsel verileri işlemek için doğal dil işleme (NLP) tekniklerinden yararlanan bir modeldir. Görüntü, genellikle 16×16 piksel boyutlarında küçük parçalara ayrılır ve her bir parça, modelin anlayabileceği özellikleri temsil eden bir vektöre dönüştürülür. Bu vektörler modele giriş olarak verilir ve dikkat (attention) mekanizmaları aracılığıyla parçalar arasındaki ilişkiler öğrenilir. BEiT, BERT modelindeki maskeli dil modellemesi yaklaşımını görsel verilere uyarlayarak, bazı görüntü parçalarını maskeleyip eksik kısımları tahmin etmeye çalışır. Böylece model, görüntüdeki küresel (global) bağlamı daha etkili bir şekilde öğrenir.

3.3.2. CNN Tabanlı Modeller

Konvolüsyonel Sinir Ağları (CNN) (LeCun ve ark., 1998) tabanlı modeller,

görsel tanıma ve sınıflandırma görevlerinde yaygın olarak kullanılan derin öğrenme modelleridir. CNN'ler, görüntü verilerini işlerken konvolüsyonel katmanlar aracılığıyla görsellerdeki önemli özellikleri çıkarır. Bu süreç, görsel verilerin giriş olarak alınmasıyla başlar; ardından, konvolüsyonel katmanlar üzerinde uygulanan filtreler sayesinde düşük seviyeli özellikler (örneğin kenarlar ve köşeler) öğrenilir. Elde edilen bilgiler, özellik haritaları (feature maps) oluşturularak temsil edilir. Aktivasyon fonksiyonları (örneğin ReLU) kullanılarak doğrusal olmayan ilişkiler öğrenilir ve havuzlama katmanları aracılığıyla boyutlar küçültülerek gereksiz detaylardan arındırılır. Son aşamada, tam bağlantılı (fully connected) katmanlar yardımıyla çıkarılan özellikler birleştirilir ve nihai sınıflandırma işlemi gerçekleştirilir. İlk katmanlar düşük seviyeli özellikleri, daha derin katmanlar ise daha karmaşık yapıları öğrenir.

3.3.2.1. ResNet

ResNet (Residual Networks) (He ve ark., 2015), derin öğrenme alanında çok derin ağların eğitimindeki zorlukları aşmayı amaçlayan bir CNN modelidir. ResNet, "residual learning" yapısını kullanır. Her bir katman, bir önceki katmandan gelen çıktıyı doğrudan kullanır. Bu doğrudan bağlantılar, kısa yol (shortcut) bağlantıları olarak bilinir. Kısa yol bağlantıları, katmanlar arasındaki öğrenme sürecini hızlandırarak, ağın daha derin katmanlarında bile etkili bir şekilde öğrenme yapılmasına yardımcı olur. ResNet'in en büyük avantajı, çok derin ağlar oluşturulmasına rağmen öğrenme sürecinde karşılaşılan kaybolan gradyan (vanishing gradient) problemini ortadan kaldırmasıdır.

3.3.2.2. MobileNet

MobileNet (Howard ve ark., 2017), mobil cihazlar ve gömülü sistemler gibi sınırlı işlem gücüne sahip donanımlarda kullanılmak üzere optimize edilmiş bir CNN tabanlı derin öğrenme modelidir. Mobil cihazların işlem gücü ve bellek kapasitesine uygun olarak geliştirilmiş olan MobileNet'in en önemli özelliği, geleneksel konvolüsyonel katmanları minimize etmek için kullanılan katman bazlı ayrık konvolüsyon (depthwise separable convolution) yaklaşımıdır. Bu yaklaşım, her bir kanal için ayrı ayrı konvolüsyon işlemi yapılmasını sağlar. Bu, geleneksel konvolüsyonel katmanlara göre daha az hesaplama gücü gerektirir, bu da modelin daha hızlı çalışmasına olanak tanır.

3.3.2.3. EfficientNet

EfficientNet (Tan ve ark., 2019), Google tarafından geliştirilen ve görüntü sınıflandırma gibi görevlerde yüksek doğruluk sağlayan verimli bir CNN mimarisidir. Derinlik, genişlik ve giriş boyutunu dengeli şekilde ölçeklendiren Bileşik Ölçekleme (Compound Scaling) yöntemiyle daha az parametreyle yüksek performans sunar. Temelinde MobileNetV2 yapısını kullanan EfficientNet, hem güçlü hem de mobil cihazlar için uygun hafif bir model ailesidir.

3.3.2.4. ConvNeXt

ConvNeXt (Liu ve ark., 2022), klasik Konvolüsyonel Sinir Ağları (CNN) mimarisini modern derin öğrenme teknikleriyle güncelleyerek geliştirilmiş bir görüntü sınıflandırma modelidir. ViT modelinden ilham alan ConvNeXt; büyük çekirdekler, GELU aktivasyonu, Layer Normalization gibi bileşenlerle CNN'in yapısını modernleştirir. Bu sayede hem hesaplama verimliliği korunur hem de yüksek doğrulukla güçlü performans elde edilir.

3.4. Hiperparametrelerin Belirlenmesi

Eğitim parametreleri, modelin performansını doğrudan etkileyen en önemli faktörlerden biridir. Bu çalışmada, modelin verimli bir şekilde eğitilebilmesi için Çizelge 3.2'de yer alan parametreler kullanılmıştır.

Çizelge 3.2. Model eğitiminde kullanılan hiperparametreler

Hiperparametre	Değer Aralığı
Batch Size	16, 32, 64
Learning Rate	1e-5, 5e-4
Epoch Sayısı	20 - 30
Warmup Steps	100 - 200

Optuna'nın Bayesian optimizasyonu, modelin hiperparametrelerini daha verimli bir şekilde aramak için kullanılmıştır. Bu optimizasyon süreci, her bir parametre denemesinin önceki sonuçlara dayanarak seçilmesini sağlar ve böylece daha az deneme ile optimal parametre kombinasyonuna ulaşılmasını hedefler. TPE (Tree-structured Parzen Estimator) algoritması, geçmiş denemelerden elde edilen verilerle hangi parametre kombinasyonlarının daha iyi sonuçlar vereceğini tahmin eder. Bu sayede, her yeni denemede daha iyi sonuçlar elde edebilmek için daha

hedeflenmiş seçimler yapılır.

Optuna, deneme sayısının (n_{trial}) 3 olarak belirlendiği bir şekilde çalıştırılmış ve yapılan üç deneme sonucunda en uygun hiperparametreler belirlenmiştir. Optimizasyon kapsamında belirlenen hiperparametreler arasında, modelin eğitim sırasında aynı anda işleyeceği örnek sayısını belirleyen toplu iş boyutu (batch size), öğrenme hızını ifade eden öğrenme oranı (learning rate), modelin tüm veri kümesi üzerinde kaç tur eğitim yapılacağını gösteren dönem sayısı (epoch) ve öğrenme oranının başlangıçta ne kadar sürede kademeli olarak artırılacağını belirleyen ısınma adımı sayısı (warmup steps) yer almaktadır. Ayrıca, optimizasyon sürecinde, modelin daha verimli bir şekilde öğrenmesi ve daha iyi sonuçlar elde etmesi için AdamW optimizatörü de kullanılmıştır.



4. BULGULAR

4.1. Model Eğitim Süreci

4.1.1. Değerlendirme Metodolojisi

Modelin performansını değerlendirmek amacıyla, sınıflandırma başarısı ve hata oranları analiz edilmiştir. Değerlendirme sürecinde her bir dönem (epoch) için eğitim kaybı (training loss) ve doğrulama kaybı (validation loss) değerleri izlenmiştir. Ayrıca, modelin genel performansı kesinlik (precision), duyarlılık (recall), doğruluk (accuracy) ve F1-skoru (F1-score) gibi sınıflandırma metrikleri ile karmaşıklık matrisi (confusion matrix) kullanılarak detaylı bir şekilde analiz edilmiştir.

4.1.2. Deneysel Ortam

Modelin eğitimi için TRUBA¹hesaplama altyapısı kullanılmıştır. Eğitim süreci sırasında, donanım olarak 50 GB sistem RAM'ine sahip, 1 adet CUDA destekli NVIDIA GPU ve 20 çekirdekli bir CPU kullanılmıştır. Yazılım konfigürasyonu olarak, RockyLinux-9.2 işletim sistemi, Python 3.x sürümü ve modelin eğitimini gerçekleştirmek için PyTorch derin öğrenme kütüphanesi ile Hugging Face Hub kullanılmıştır.

4.2. Performans Değerlendirme Metrikleri

Modelin performansını değerlendirmek amacıyla, çeşitli doğruluk ve hata ölçütleri dikkate alınarak kapsamlı bir değerlendirme yapılmıştır. Bu kapsamda kullanılan metrikler, modelin sınıflandırma başarısını, genelleme yeteneğini ve farklı sınıflar üzerindeki ayırt ediciliğini analiz etmeye olanak sağlamıştır. Elde edilen sonuçlar, modelin eğitim ve test verileri üzerindeki performansını karşılaştırmalı olarak incelemek için kullanılmıştır.

4.2.1. Doğruluk (Accuracy)

Modelin doğru tahmin ettiği örneklerin, toplam tahmin sayısına oranı doğruluk (accuracy) olarak adlandırılır. Makine öğrenimi ve derin öğrenme modellerinin genel performansını değerlendiren temel metriklerden biridir. Modelin

¹ <https://www.truba.gov.tr/>

doğru tahmin ettiği örneklerin (hem pozitif hem de negatif doğru tahminler) toplam tahmin sayısına oranıdır. Matematiksel formülü ise şu şekildedir:

$$\text{Dogruluk} = \frac{DP + DN}{DP + DN + YP + YN} \quad (4.1)$$

Kullanılan performans ölçütleri hesaplanırken formüllerde şu şekilde tanımlanmıştır: DP (Doğru Pozitif), modelin doğru şekilde pozitif olarak tahmin ettiği örnekleri ifade eder. DN (Doğru Negatif), modelin doğru şekilde negatif olarak tahmin ettiği örnekleri belirtir. YP (Yanlış Pozitif), modelin hatalı şekilde pozitif olarak tahmin ettiği örnekleri temsil eder. YN (Yanlış Negatif) ise modelin hatalı şekilde negatif olarak tahmin ettiği örnekleri ifade etmektedir.

4.2.2. Kesinlik (Precision)

Kesinlik, modelin pozitif tahminlerinin doğruluğunu ölçen bir metriktir. Tahmin edilen pozitif örneklerin ne kadarının gerçekten pozitif olduğunu gösterir. Formül şu şekildedir:

$$\text{Kesinlik} = \frac{DP}{DP + YP} \quad (4.2)$$

DP (Doğru Pozitif), model tarafından doğru bir şekilde pozitif olarak sınıflandırılan pozitif örneklerin sayısını ifade ederken, YP (Yanlış Pozitif) model tarafından yanlış bir şekilde pozitif olarak sınıflandırılan negatif örneklerin sayısını belirtir.

4.2.3. Duyarlılık (Recall)

Duyarlılık, modelin pozitif örnekleri ne kadar doğru tanımladığını ölçen bir metriktir. Modelin gerçek pozitif örnekleri ne kadar iyi sınıflandırdığını gösterir. Formül şu şekildedir:

$$\text{Duyarlilik} = \frac{DP}{DP + YN} \quad (4.3)$$

DP (Doğru Pozitif), model tarafından doğru bir şekilde pozitif olarak

sınıflandırılan pozitif örneklerin sayısını ifade ederken, YN (Yanlış Negatif) model tarafından yanlış bir şekilde negatif olarak sınıflandırılan pozitif örneklerin sayısını belirtir.

4.2.4. F1-Skoru (F1-Score)

F1 Skoru, Kesinlik ve Duyarlılığı dengeleyen bir metriktir. Özellikle bu değerlerden biri (Kesinlik veya Duyarlılık) düşük olduğunda kullanışlıdır. Formül şu şekildedir:

$$F1 \text{ Skoru} = 2 \times \frac{K \times D}{K + D} \quad (4.4)$$

Kesinlik (K), modelin pozitif tahminlerinin doğruluğunu ifade ederken, Duyarlılık (D) modelin pozitif örnekleri doğru şekilde tanımlama yeteneğini belirtir.

4.3. Deney Sonuçları

Bu çalışmada, farklı derin öğrenme modellerinin (Beit, Swin, ViT, DeiT, ResNet, MobileNet, EfficientNet ve ConvNext) sanatçı, tür ve stil sınıflandırma görevlerinde performansları değerlendirilmiştir. Modellerin başarısı, doğruluk, F1 skoru, kesinlik ve duyarlılık gibi yaygın performans metrikleri ile ölçülmüştür. Çizelge 4.1 ve Çizelge 4.2'de görüldüğü gibi dönüştürücü tabanlı modeller olan Beit ve Swin'in performansları, WikiArt veri kümesi (Tan 2019) üzerinde yapılan testlerde ViT ve DeiT ile karşılaştırılmıştır. ViT ve DeiT diğer modellere kıyasla geride kalmıştır. Genel olarak Beit ve Swin, daha yüksek doğruluk ve duyarlılık oranları sunmuştur. Beit, %84.90 doğruluk oranı ile “Sanatçı” sınıflandırmasında en yüksek performansı sergilerken, Swin %72.59 doğruluk oranı ile “Stil” sınıflandırmasında en yüksek performansı göstermiştir.

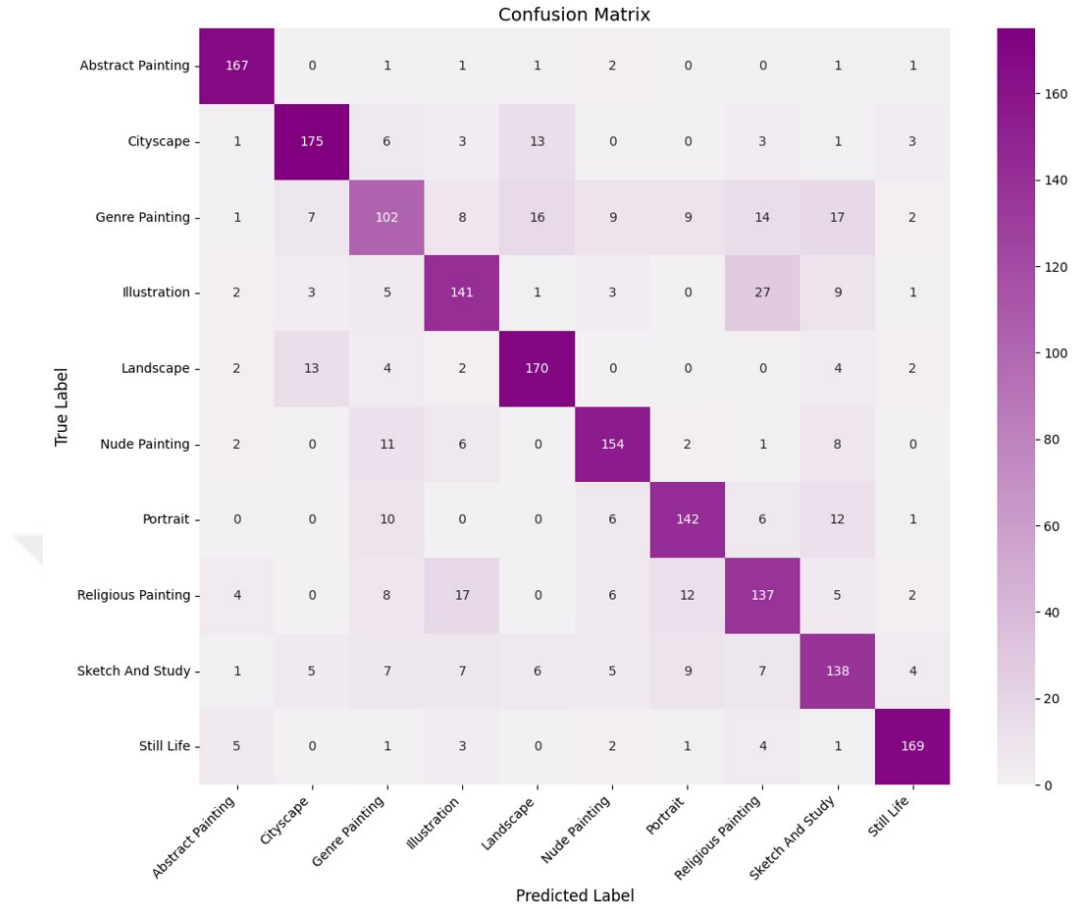
Çizelge 4.1. Bu çalışmada elde edilen dönüştürücü tabanlı modellerin performans sonuçları.

Model	Sınıf	Doğruluk (%)	F1 Skoru (%)	Kesinlik (%)	Duyarlılık (%)
BEiT	Sanatçı	84.90	84.79	84.83	84.90
	Tür	79.52	79.35	79.31	79.52
	Stil	68.07	68.27	68.91	68.07
Swin	Sanatçı	82.80	82.75	82.88	82.80
	Tür	76.44	76.20	76.12	76.44
	Stil	72.59	72.73	73.23	72.59
ViT	Sanatçı	82.02	81.98	82.02	82.02
	Tür	75.96	75.86	75.93	75.96
	Stil	68.07	68.21	68.47	68.07
DeiT	Sanatçı	79.78	79.69	79.69	79.78
	Tür	73.51	73.47	73.57	73.51
	Stil	68.52	68.42	68.82	68.52

Çizelge 4.2. Bu çalışmada elde edilen CNN tabanlı modellerin performans sonuçları.

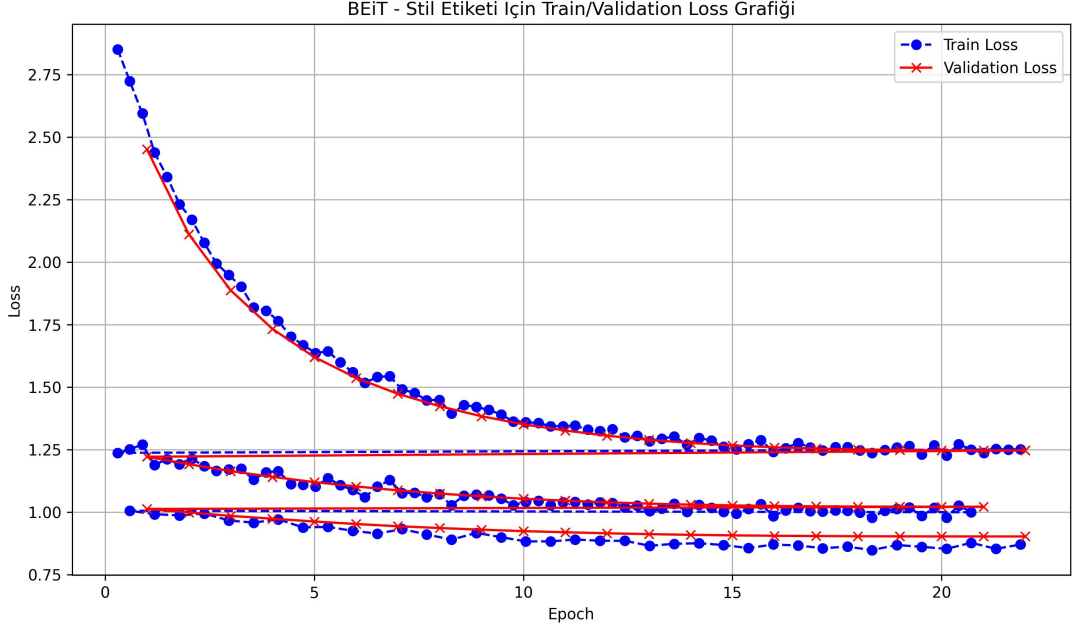
Model	Sınıf	Doğruluk (%)	F1 Skoru (%)	Kesinlik (%)	Duyarlılık (%)
ResNet	Sanatçı	76.26	76.09	76.49	76.26
	Tür	68.99	68.45	68.49	68.99
	Stil	58.07	56.81	58.12	58.07
MobileNet	Sanatçı	68.40	68.01	68.53	68.40
	Tür	69.52	69.50	69.65	69.52
	Stil	55.33	55.43	58.51	55.33
EfficientNet	Sanatçı	77.20	77.09	77.67	77.20
	Tür	71.06	70.64	70.76	71.06
	Stil	63.93	63.41	63.92	63.93
ConvNeXt	Sanatçı	79.94	79.84	79.95	79.94
	Tür	73.72	73.50	73.50	73.72
	Stil	69.26	69.24	69.66	69.26

WikiArt tür sınıfı veri kümesinin karmaşıklık matrisi Şekil 4.1'de gösterilmiştir. Tür sınıflandırmasında en iyi performans BEiT modelinden elde edilmiştir. Karmaşıklık matrisi, modelin tahminlerinin doğruluğunu görsel bir şekilde ifade etmek için kullanılmıştır. Her satır, gerçek sınıfı temsil eder ve gerçek olan kategorileri belirtirken; her sütun ise modelin yaptığı tahminleri gösterir. BEiT tür sınıfı için elde ettiğimiz karmaşıklık matrisinde, 1. satır ve 1. sütunda Abstract Painting kategorisinde 167 doğru tahmin yapılmıştır. Bu değer, başlangıçta Abstract Painting olarak kategorize edilen türün, sınıflandırma modeli tarafından doğru bir şekilde tanımlandığını göstermiştir. Modelin 2. satır ve 2. sütununda yer alan 175 değeri, modelin en fazla Cityscape türündeki sanat resimlerini doğru şekilde sınıflandırdığını ortaya koymuştur. Bu durum, modelin sanatsal türü doğru bir şekilde tanımlamada güçlü bir performans sergilediğini göstermiştir. Son olarak, 4. satır ve 8. sütunda yer alan 27 değeri, Illustration türündeki sanat resimlerini sınıflandırmada yapılan yanlışları temsil etmiştir. En fazla yanlış sınıflandırma Illustration türünde yapılmış; en fazla doğru sınıflandırma ise Cityscape türünde gerçekleştirilmiştir.



Şekil 4.1. BEiT modelinin tür sınıfı karmaşıklık matrisi.

Şekil 4.2’de modelin eğitim ve doğrulama kayıplarının (loss) epoch’lara göre değişimi gösterilmektedir. Epoch, modelin her eğitim örneğini bir kez işlediği tam tur sayısını ifade eder. BEiT stil sınıflandırması için yapılan analizde, eğitim verisindeki kaybın düzenli olarak azaldığı ancak doğrulama verisindeki kaybın belli bir noktadan sonra artmaya başladığı görülmüştür. Bu durum, modelin eğitim verisine fazla uyum sağladığını ve genel performansının azaldığını, yani aşırı öğrenme (overfitting) durumunun ortaya çıktığını göstermektedir.



Şekil 4.2. WikiArt veri kümesi için eğitim ve doğrulama kayıplarının epochlara göre değişimi.

Çizelge 4.3'te, modelin performansını artırmak amacıyla Optuna kütüphanesi kullanılarak gerçekleştirilen hiperparametre optimizasyonunun sonuçları yer almaktadır. Bu süreçte, kullanılan modeller için çeşitli parametre kombinasyonları değerlendirilmiş ve en iyi sonuçlar, doğrulama verisi üzerindeki başarıya göre seçilmiştir. Bu parametreler, modelin aşırı öğrenme yapmasını önlemek ve genel performansını artırmak adına eğitim aşamasında kullanılmıştır.

Çizelge 4.3. Dönüştürücü ve CNN tabanlı modellerin hiperparametre karşılaştırması

Model	Learning Rate	Batch Size	Epoch	Warmup Steps
BEiT	1.4983	16	21	189
ViT	0.00018	32	28	129
Swin	0.00026	16	26	101
DeiT	2.3096	32	23	123
ResNet-50	0.00011	16	30	103
MobileNet	0.00026	64	22	122
ConvNeXt	0.00021	32	24	105
EfficientNet	0.00045	32	20	105

Çizelge 4.4'te, literatürdeki bazı güncel yöntemlerin sınıflandırma performansları karşılaştırmalı olarak sunulmuştur. Yu ve arkadaşlarının (2024) önerdiği çok ölçekli CNN tabanlı yaklaşım, %79,11 doğruluk değeri elde etmiştir. Bu yöntem, geleneksel CNN tabanlı yapılarla belirli avantajlar sunsa da, çalışmamızda kullanılan dönüştürücü tabanlı yöntemlerle kıyaslandığında daha düşük bir performans sergilemiştir. BEiT modeli, %84,90 doğruluk oranı ile diğer yöntemlere kıyasla en yüksek başarıyı göstermiştir. Swin dönüştürücü ise %82,80 doğruluk oranı ile BEiT'ten sonra en iyi performansı gösteren ikinci model olmuştur. Bu sonuçlar, dönüştürücü tabanlı yaklaşımların sanat resimleri sınıflandırma işleminde geleneksel yöntemlere kıyasla daha yüksek başarı sağladığını ortaya koymuştur.

Çizelge 4.4. WikiArt veri kümesi kullanılarak sanatçı sınıfı üzerinde çalıştırdığımız BEiT ve Swin modellerinin sonuçları ile güncel çalışmaların sonuçları.

Yöntem	Kesinlik (%)	Duyarlılık (%)	F1 Skoru (%)	Doğruluk (%)
Multi-ölçekli CNN (Yu ve ark. 2024)	77.65	65.35	68.73	79.11
BEiT	84.83	84.90	84.79	84.90
Swin	82.88	82.80	82.75	82.80

Çizelge 4.5'de, sanat eserlerinin stil, sanatçı ve tür sınıflandırması işlemlerinde kullanılan farklı derin öğrenme tabanlı yöntemlerin doğruluk oranlarına ilişkin karşılaştırmalı sonuçlar sunulmuştur. Tablo genel olarak, klasik CNN mimarilerinden dönüştürücü tabanlı modellere geçişin, sanat resimleri sınıflandırma performansında belirgin bir iyileşme sağladığını ortaya koymuştur. Saleh ve Elgammal (2015) tarafından önerilen CNN-LMNN yöntemi, stil sınıflandırmasında %45,97, sanatçı sınıflandırmasında %63,06 ve tür sınıflandırmasında %60,28 doğruluk oranlarına ulaşmıştır. Bu yöntemi takiben, Tan ve arkadaşları (2016) tarafından geliştirilen ince ayarlı AlexNet modeli stil (%54,50), sanatçı (%74,14) ve tür (%76,02) sınıflandırmalarında daha yüksek doğruluk oranlarına erişmiştir. Cetinic ve arkadaşlarının (2018) CNN fine-tuning tabanlı yaklaşımı ise özellikle sanatçı (%81,94) ve tür (%77,60) sınıflandırmalarında öne çıkmış; stil sınıflandırmasında ise %56,43 doğruluk oranı elde etmiştir.

Yapılan çalışmamızda, dönüştürücü mimarilerine dayalı modellerin sonuçları incelendiğinde, bu yapıların klasik yöntemlere kıyasla daha yüksek doğruluk oranları sağladığı görülmüştür. BEiT modeli, sanatçı sınıflandırmasında %84,90 doğruluk oranı ile en başarılı sonuçları elde ederken; tür sınıflandırmasında %79,52 ve stil sınıflandırmasında %68,07 doğruluk oranına ulaşmıştır. İkinci en iyi dönüştürücü model olan Swin ise stil sınıflandırmasında %72,59 doğruluk oranı ile en yüksek başarıyı göstermiştir.

Bu sonuçlar, dönüştürücü tabanlı modellerin sanat resimleri gibi karmaşık ve bağlamsal öğeler içeren veri kümelerinde daha etkin temsiller öğrenebildiğini ve bu sayede sınıflandırma görevlerinde daha yüksek doğruluk sağladığını ortaya koymuştur.

Çizelge 4.5. WikiArt veri kümesi kullanılarak stil, sanatçı ve tür sınıfları üzerinde çalıştırdığımız BEiT ve Swin modellerinin sonuçları ile güncel çalışmaların sonuçları

Method	Sınıf	Doğruluk (%)
CNN- LMNN (Saleh & Elgammal 2015)	Stil	45.97
	Sanatçı	63.06
	Tür	60.28
Fine-tuned AlexNet(Tan ve ark.2016)	Stil	54.50
	Sanatçı	74.14
	Tür	76.02
CNN fine-tuning (Cetinic ve ark. 2018)	Stil	56.43
	Sanatçı	81.94
	Tür	77.60
BEiT	Stil	68.07
	Sanatçı	84.90
	Tür	79.52
Swin	Stil	72.59
	Sanatçı	82.80
	Tür	76.44

5. TARTIŞMA

Yaptığımız çalışmada, sanat resimlerinin sanatçı, stil ve tür sınıflandırması için geliştirilen dönüştürücü tabanlı modellerin, geleneksel CNN tabanlı yöntemlere kıyasla daha üstün performans gösterdiği gözlemlenmiştir. Çizelge 4.1 ve Çizelge 4.2'de yaptığımız çalışmanın sonuçlarına göre, BEiT tabanlı model sanatçı sınıflandırmada %84.90 doğruluk oranı ile en yüksek başarıyı göstermiştir. Bu durum, modelin sanatçıların karakteristik çizim özelliklerini ve stilistik farklılıklarını yüksek doğrulukla yakalayabildiğini ortaya koymuştur. Kesinlik, duyarlılık ve F1 skoru gibi dengeleyici metriklerde de BEiT modelinin tutarlı bir performans sergilemesi, modelin sadece belirli sınıflarda değil, genel dağılımda da etkili olduğunu göstermiştir. Swin dönüştürücü modeli ise %82.80 doğruluk oranı ile BEiT'e yakın bir performans sergilemiş, özellikle duyarlılık ve F1 skoru metriklerinde güçlü sonuçlar elde etmiştir. Bununla birlikte, Çizelge 4.4'te gösterildiği gibi Multi-ölçekli CNN yönteminin (Yu, 2024) dönüştürücü tabanlı modellere kıyasla düşük performans göstermesi, geleneksel CNN mimarilerinin özellikle karmaşık stilistik öğeleri ve bağlamsal ilişkileri yakalama konusunda yetersiz kaldığını göstermiştir.

Çizelge 4.4 ve 4.5'de yer alan literatür karşılaştırmaları incelendiğinde, geçmiş çalışmalarda CNN-LMNN (Saleh & Elgammal, 2015) gibi yöntemlerin stil sınıflandırmasında %45.97 gibi sınırlı doğruluk oranları elde ettiği, daha yeni CNN tabanlı yaklaşımlarda (Tan ve ark., 2016; Cetinic ve ark., 2018) ise kısmi iyileşmeler sağlandığı görülmüştür. Bu yöntemler, yaptığımız çalışmada dönüştürücü tabanlı modellerin ulaştığı doğruluk seviyelerine kıyasla daha düşük kalmıştır. Özellikle stil sınıflandırması gibi soyut kategorilerde bile SWIN modelinin %72.59 doğruluk oranına ulaşması, dönüştürücü mimarilerinin sanat resimlerinin karmaşık yapısal özelliklerini anlamada büyük bir avantaj sunduğunu göstermiştir. Bu bağlamda, dönüştürücü tabanlı modellerin, sanat resim sınıflandırma gibi görsel ve bağlamsal bilgiyi birlikte işleyen görevlerde daha geniş bir potansiyele sahip olduğu sonucuna varılmıştır.

6. SONUÇLAR

Bu çalışmada, WikiArt veri kümesindeki sanat resimleri; stil, sanatçı ve tür kategorilerine göre sınıflandırılmış, CNN tabanlı modeller ile dönüştürücü tabanlı mimariler karşılaştırılmıştır. Araştırmanın temel amacı, farklı model yaklaşımlarının sanat eseri sınıflandırmasındaki başarısını sayısal verilerle ortaya koymak ve bu alanda hangi model grubunun daha etkin performans sergilediğini belirlemek olmuştur. Yapılan deneysel çalışmalar, BEiT ve Swin gibi dönüştürücü tabanlı mimarilerin, doğruluk, F1 skoru, kesinlik ve duyarlılık gibi ölçütlerde CNN tabanlı modellere kıyasla genel olarak daha üstün performans sergilediğini ortaya koymuştur. Sanatçı sınıflandırması görevinde en yüksek doğruluk oranı %84.90 ile BEiT modelinden elde edilmiş ve bu sonuç, literatürde yer alan önceki çalışmalarda sunulan sonuçlardan daha yüksek çıkmıştır. Stil sınıflandırması ise tüm sınıflar arasında en zorlayıcı olanı olarak öne çıkmıştır; bu işlemde en başarılı model, %72.59 doğruluk oranı ile Swin modeli olmuştur. Öte yandan, tür sınıflandırması açısından da BEiT modeli, %79.52 doğruluk oranıyla ilk sırada yer almış ve burada dönüştürücü tabanlı modellerin daha tutarlı çıktılar üretebildiği gözlemlenmiştir. Bu kapsamda yapılan çoklu görev sınıflandırması deneyleri, sanat resimlerinin sınıflandırılmasında dönüştürücü tabanlı yaklaşımların güçlü birer alternatif olduğunu ortaya koymuştur. Dönüştürücü mimarilerinde yer alan kendi öğelerine odaklanma (self-attention) mekanizmasının, görsel öğeler arasındaki bağlamsal ilişkileri daha etkin şekilde yakalayabilmesi, özellikle sanatçı tahmini görevinde yüksek doğruluk oranlarının elde edilmesini sağlamıştır. Elde edilen bu sonuçlar, yalnızca akademik anlamda katkı sunmakla kalmayıp, özellikle BEiT ve Swin gibi modellerin tercih edilmesinin müze arşivleme sistemleri, çevrimiçi sanat katalogları ve akıllı sanat asistanları gibi uygulamalarda sınıflandırma doğruluğunu artıracığı öngörülmüştür. Sonuç olarak, bu çalışma, hem farklı model mimarilerinin detaylı metriklerle kıyaslandığı bir ortam hem de sanat resimleri sınıflandırması alanında kullanılabilir somut model önerileri sunmuştur.

7. ÖNERİLER

WikiArt veri kümesinde bazı sanatçı, stil ve tür kategorilerinin oldukça az sayıda örnek içermesi, sınıflar arasında ciddi bir dengesizliğe yol açarak makine öğrenimi modellerinin eğitiminde az temsil edilen sınıflarda düşük doğruluk ve genel başarımlarına neden olmuştur. Bu sorunu hafifletmek amacıyla çeşitli veri artırma teknikleri uygulanmış ve veri kümesinin büyük boyutu, eğitim sürecinde bellek ve işlem gücü açısından bazı kısıtlamalar olmuştur. Gelecek çalışmalarda, daha yüksek donanım kapasitesine sahip sistemlerin tercih edilmesi ya da veri örneklerinin dengeli alt kümeler hâlinde işlenmesi faydalı olacaktır. Bu nedenle yalnızca örnek sayısı belirli bir eşik üzerinde olan sanatçılarla sınırlı bir eğitim veri kümesi oluşturulmuş, ancak bu durum sanatçı sınıflandırma işleminde dengesizliği artırarak modellerin genelleme kabiliyetini olumsuz etkilemiştir. Bu sınırlamaları aşmak için dengeli alt kümeler oluşturulabilir, ağırlıklı kayıp fonksiyonları kullanılabilir, dış bellek sistemleriyle veri yönetimi optimize edilebilir ve daha güçlü donanımlar ya da bulut tabanlı GPU/TPU altyapıları tercih edilebilir. Ayrıca, WikiArt dışındaki farklı, daha dengeli sanat resimleri veri kümeleri entegre edilerek veri çeşitliliği artırılabilir.

KAYNAKLAR

- Agarwal, S., Karnick, H., Pant, N., & Patel, U. (2015). Genre and style based painting classification. In *2015 IEEE Winter Conference on Applications of Computer Vision* (pp. 588–594). IEEE.
- Carneiro, G., et al. (2012). Artistic image classification: An analysis on the printart database. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision* (pp. 198–211). Springer Berlin Heidelberg.
- Cétnic, E., Lipic, T., & Grgic, S. (2018). Fine-tuning convolutional neural networks for fine art classification. *Expert Systems with Applications*, *114*, 107–118.
- Choudhury, M. D. (2020). Automated identification of painters over WikiArt image data using machine learning algorithms (Doctoral dissertation). National College of Ireland.
- Chu, W. T., & Wu, Y. L. (2018). Image style classification based on learnt deep correlation features. *IEEE Transactions on Multimedia*, *20*(9), 2491–2502.
- Conde, M. V., & Turgutlu, K. (2021). Clip-art: Contrastive pre-training for fine-grained art classification. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- Cömert, C., Özbayoğlu, M., & Kasnakoğlu, C. (2021). Painter prediction from artworks with transfer learning. In *2021 7th International Conference on Mechatronics and Robotics Engineering (ICMRE)*. IEEE.
- Diem, S., & Mandl, T. (2023). Automatic classification of portraits: Application of transformer and CNN-based models for an art historic dataset. In *LWDA* (pp. 192–206).
- Fortuna-Cervantes, J. M., et al. (2024). Evaluation of CNN models with transfer learning in art media classification in terms of accuracy and class relationship. *Computación y Sistemas*, *28*(1), 233–244.
- Gonthier, N., Gousseau, Y., & Ladjal, S. (2021). An analysis of the transfer learning of convolutional neural networks for artistic images. In *Pattern Recognition. ICPR International Workshops and Challenges* (pp. 320–329). Springer International Publishing.
- Iliadis, L. A., Nikolaidis, S., Sarigiannidis, P., Wan, S., & Goudos, S. K. (2021).

Artwork style recognition using vision transformers and MLP-Mixer. *Technologies*, 10(1), 2.

- Imran, S., et al. (2023). Artistic style recognition: Combining deep and shallow
- Jangtik, K. A., Ho, T. T., Yeh, M. C., & Hua, K. L. (2017). A CNN-LSTM framework for authorship classification of paintings. In *2017 IEEE International Conference on Image Processing (ICIP)* (pp. 2866–2870).
- Jeremiasz, J., et al. (2022). WikiArtVectors: Style and color representations of artworks for cultural analysis via information theoretic measures. *Entropy*, 24(9), 1175.
- Jin, S. (2023). Exploring the causes of artwork misclassification through machine learning (Doctoral dissertation). State University of New York at Stony Brook.
- Khalil, M., Khalil, A., & Ngom, A. (2023). A comprehensive study of vision transformers in image classification tasks. *arXiv preprint arXiv:2312.01232*.
- Krishna, S. T., & Kalluri, H. K. (2019). Deep learning and transfer learning approaches for image classification. *International Journal of Recent Technology and Engineering (IJRTE)*, 7(5S4), 427–432.
- Kumar, A., Mu, J., & Vasconcelos, N. (2024). IntroStyle: Training-free introspective style attribution using diffusion features. *arXiv preprint arXiv:2412.14432*.
- Liu, Y., Bai, H., & Wang, J. (2024). Fine-art recognition using convolutional transformers. *PeerJ Computer Science*, 10, e2409.
- Lu, K., Xu, Y., & Yang, Y. (2021). Comparison of the potential between transformer and CNN in image classification. *ICMLCA 2021; 2nd International Conference on Machine Learning and Computer Application*, VDE.
- Menai, B. L., & Babahenini, M. C. (2023). The effect of optimizers on CNN architectures for art style classification. *International Journal of Computing and Digital Systems*, 13(1), 353–360.
- Obeso, A. M., et al. (2019). Saliency-based selection of visual content for deep convolutional neural networks: Application to architectural style classification. *Multimedia Tools and Applications*, 78(8), 9553–9576.

- Oh, C. (2018). Automatically classifying art images using computer vision. University of Washington Libraries.
- Oomen, E. (2018). Classification of painting style with transfer learning (Master's thesis). Tilburg University, School of Humanities and Digital Science, Tilburg, The Netherlands. Supervisor: Prof. Dr. Eric Postma; Second reader: Dr. E.A. Keuleers. ANR: 1247723.
- Rodriguez, C. S., Lech, M., & Pirogova, E. (2018). Classification of style in fine-art paintings using transfer learning and weighted image patches. In *2018 12th International Conference on Signal Processing and Communication Systems (ICSPCS)*. IEEE.
- Rui, W., Liu, Y., Liu, J., Li, Z., & Chen, S. (2020). A novel deep learning-based approach for art image classification and style recognition. *PLOS ONE*, *15*(5), e0248414.
- Sandoval, C., Pirogova, E., & Lech, M. (2019). Two-stage deep learning approach to the classification of fine-art paintings. *IEEE Access*, *7*, 41770–41781.
- Sandoval, C. E., Pirogova, M., & Lech, M. (2021). Adversarial learning approach to unsupervised labeling of fine art paintings. *IEEE Access*, *9*, 81969–81985.
- Schaerf, L., Postma, E., & Popovici, C. (2024). Art authentication with vision transformers. *Neural Computing and Applications*, *36*, 11849–11858.
- Shaha, M., & Pawar, M. (2018). Transfer learning for image classification. In *2018 Second International Conference on Electronics, Communication and Aerospace Technology (ICECA)*. IEEE.
- Smirnov, S., & Eguizabal, A. (2018). Deep learning for object detection in fine-art paintings. *2018 Metrology for Archaeology and Cultural Heritage (MetroArchaeo)*. IEEE.
- Smith, N. C. (2024). Exploration of few-shot learning for artist-based fine art classification (PhD dissertation). Tilburg University.
- Tan, W. R., Chan, C. S., Aguirre, H. E., & Tanaka, K. (2016). Ceci n'est pas une pipe: A deep convolutional network for fine-art paintings classification. In *2016 IEEE International Conference on Image Processing (ICIP)* (pp. 3703–3707). Phoenix, AZ, USA.

- Tang, W., et al. (2024). Application of painting style recognition and imitation technology based on deep learning in art education. SSRN 5031075.
- Yang, S., Oh, B. M., Merchant, D., Howe, B., & West, J. (2018). Classifying digitized art type and time period. In *Proceedings of the 1st Workshop on Data Science for Digital Art History-Tacking Big Data*.
- Yemelienenko, T., Tkachenko, I., Masclef, T., Scuturici, M., & Miguet, S. (2025). Artwork recommendations guided by foundation models: Survey and novel approach. *Multimedia Tools and Applications*, 1-22.
- Yu, Q., & Shi, C. (2024). An image classification approach for painting using improved convolutional neural algorithm. *Soft Computing*, 28, 847–873.
- Zhang, X., & Ding, T. (2024). Style classification of media painting images by integrating ResNet and attention mechanism. *Heliyon*, 10(6), e27178.
- Zhao, Q., & Zhang, R. (2025). Classification of painting styles based on the difference component. *Expert Systems with Applications*, 259, 125287.
- Zhao, W., Jiang, W., & Qiu, X. (2022). Big transfer learning for fine art classification. *Computational Intelligence and Neuroscience*, 1764606.
- Zhao, W., Zhou, D., Qiu, X., & Jiang, W. (2021). Compare the performance of the models in art classification. *PLOS ONE*, 24(8), e0248414.