T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇEKİŞMELİ ÜRETİCİ AĞLAR İLE GERÇEKÇİ HİSTOPATOLOJİK GÖRÜNTÜ ÜRETİMİ VE EVRE TESPİTİ

Reyhan DEDE

YÜKSEK LİSANS TEZİ
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı

Danışman
Prof. Dr. Gökhan BİLGİN

Mart, 2025

T.C.
YILDIZ TEKNİK ÜNİVERSİTESİ
FEN BİLİMLERİ ENSTİTÜSÜ

ÇEKİŞMELİ ÜRETİCİ AĞLAR İLE GERÇEKÇİ HİSTOPATOLOJİK GÖRÜNTÜ ÜRETİMİ VE EVRE TESPİTİ

Reyhan DEDE tarafından hazırlanan tez çalışması 10.03.2025 tarihinde aşağıdaki jüri tarafından Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü Bilgisayar Mühendisliği Anabilim Dalı Bilgisayar Mühendisliği Programı **YÜKSEK LİSANS TEZİ** olarak kabul edilmiştir.

Prof. Dr. Gökhan BİLGİN
Yıldız Teknik Üniversitesi
Danışman

Jüri Üyeleri

Prof. Dr. Gökhan BİLGİN, Danışman
Yıldız Teknik Üniversitesi _____

Prof. Dr. Tamer ÖLMEZ, Üye
İstanbul Teknik Üniversitesi _____

Doç. Dr. Ali Can KARACA, Üye
Yıldız Teknik Üniversitesi _____

Danışmanım Prof. Dr. Gökhan BİLGİN sorumluluğunda tarafımca hazırlanan Çekişmeli Üretici Ağlar ile Gerçekçi Histopatolojik Görüntü Üretimi ve Evre Tespiti başlıklı çalışmada veri toplama ve veri kullanımında gerekli yasal izinleri aldığımı, diğer kaynaklardan aldığım bilgileri ana metin ve referanslarda eksiksiz gösterdiğimi, araştırma verilerine ve sonuçlarına ilişkin çarpıtma ve/veya sahtecilik yapmadığımı, çalışmam süresince bilimsel araştırma ve etik ilkelerine uygun davrandığımı beyan ederim. Beyanımın aksinin ispatı halinde her türlü yasal sonucu kabul ederim.

Reyhan DEDE

İmza

Aileme
Geçmişteki ve Gelecekteki Tüm Tecrübelerime

TEŞEKKÜR

Öncelikle, bilgi ve tecrübeleriyle bana rehberlik eden, sabrı ve anlayışıyla her zaman destek olan danışman hocam Prof. Dr. Gökhan BİLGİN'e sonsuz teşekkürlerimi sunarım. Hocamın yol göstericiliği bu çalışmanın her aşamasında bana ışık tutmuştur.

Araştırmalarım boyunca yanımda olan ve moral desteği sağlayan değerli aileme, sabırları ve sonsuz sevgileri için en içten şükranlarımı sunarım. Onların desteği, bu süreçte bana güç vermiştir.

Bu tez çalışmasının oluşumunda emeği geçen herkese en derin teşekkürlerimi sunarım.

Reyhan DEDE

İÇİNDEKİLER

SİMGE LİSTESİ vii

KISALTMA LİSTESİ viii

ŞEKİL LİSTESİ x

TABLO LİSTESİ xi

ÖZET xii

ABSTRACT xiv

1 GİRİŞ 1

1.1 Literatür İncelemesi . . . 2
1.2 Tezin Amacı . . . 6
1.3 Hipotez . . . 6

2 GENERATİF ÇEKİŞMELİ AĞLAR (GANs) 7

2.1 Temel Mimari . . . 7
2.2 Çekişmeli Eğitim Çerçevesi . . . 9
2.2.1 Yakınsama ve Nash Dengesi . . . 10
2.3 GAN Varyasyonları . . . 10
2.4 StyleGAN ve Varyasyonları . . . 18
2.4.1 StyleGAN3: Alias-Free Yaklaşımı ve Mimari Detayları . . . 21
2.5 GAN'ların Sınırlamaları ve Alınabilecek Önlemler . . . 23
2.5.1 Mod Çökmesi . . . 24
2.5.2 Yakınsamama Sorunları . . . 24
2.5.3 Eğitim İstikrarsızlığı . . . 25
2.6 GAN Değerlendirme Metrikleri . . . 26
2.6.1 Nitel Değerlendirme . . . 26
2.6.2 Nicel Değerlendirme . . . 27

3 DÖNÜŞTÜRÜCÜ MODELLER 32

3.1 Genel Yapı: Kodlayıcı-Çözücü Mimarisi . . . 32

3.2 Kodlayıcı Bileşenleri 32
3.2.1 Multi-Head Self-Attention Mekanizması 33
3.2.2 Point-Wise Feed-Forward Network (FFN) 33
3.2.3 Artık Bağlantılar ve Normalizasyon 34
3.3 Çözücü Bileşenleri 34
3.3.1 Masked Multi-Head Self-Attention 34
3.3.2 Encoder-Decoder Attention 35
3.3.3 Point-Wise Feed-Forward Network 35
3.4 Multi-Head Attention Mekanizması 35
3.5 Pozisyon Kodlamaları (Positional Encodings) 35
3.6 Çıkış Katmanı 37
3.7 Swin Transformer: Gelişmiş Bir Mimari 37

4 BÜYÜK DİL MODELLERİ VE ÇOK MODLU BÜYÜK DİL MODELLERİ 44
4.1 Büyük Dil Modeli (LLM) 45
4.2 Çoklu Modlu Büyük Dil Modeli (Multimodal LLM) 47

5 MATERYALLER VE YÖNTEMLER 50
5.1 Veri Kümesi 50
5.1.1 Veri Kümesine Genel Bakış 50
5.1.2 Veri Ön İşleme 50
5.1.3 Veri Artırımı ve Dengeleme 51
5.1.4 Destekleyici Veri Birleştirme 51
5.2 Görüntü Üretme Model Mimarisi 51
5.3 Model Eğitimi 53
5.3.1 Swin Transformer ile Özellik Çıkarımı 53
5.3.2 Eğitim Süreci 53
5.3.3 Sınıflandırma 55
5.3.4 Çoklu Modlu Büyük Dil Modeli ile Sınıflandırıcı 58

6 SONUÇ 63
6.1 DEĞERLENDİRME 63
6.1.1 Frechet Başlangıç Mesafesi (FID) 63
6.1.2 Swin Transformer Sınıflandırma Performansı 65
6.1.3 Çoklu Modlu Büyük Dil Modeli Sınıflandırma Performansı 66
6.2 Sınırlamalar ve Gelecek Yönelimler 66

KAYNAKÇA 68

TEZDEN ÜRETİLMİŞ YAYINLAR 73

SİMGE LİSTESİ

$D(x)$	Ayrıştırıcı ağın (Discriminator) çıktısı. x veri örneğinin gerçek olma olasılığını belirtir.
FID	Frechet Inception Distance metriği, üretilen görüntülerin gerçek görüntülere benzerliğini ölçer.
$V(D, G)$	GAN'ların minimaks optimizasyon problemini ifade eden amaç fonksiyonu.
μ_r, μ_g	Gerçek ve üretilmiş görüntülerin özellik vektörlerinin ortalamaları.
Σ_r, Σ_g	Gerçek ve üretilmiş görüntülerin özellik vektörlerinin kovaryans matrisleri.
$P_{\text{data}}(x)$	Gerçek veri dağılımını temsil eder.
$P_z(z)$	Gürültü vektörünün tanımlandığı olasılık dağılımı.
IS	Inception Score, üretilen görüntülerin kalitesini ve çeşitliliğini değerlendirir.
KL	Kullback-Leibler uzaklığı, iki olasılık dağılımı arasındaki farkı ölçer.
Tr	Matrislerin köşegen elemanlarının toplamını ifade eder.
$G(z; \theta_g)$	Üretici ağın (Generator) matematiksel ifadesi. z rastgele girdi vektörünü, θ_g ise üretici ağın öğrenilebilir parametrelerini temsil eder.

KISALTMA LİSTESİ

AdaIN	Adaptive Instance Normalization (Uyarlanabilir Örnek Normalizasyonu)
Adam	Adaptive Moment Estimation (Uyarlanabilir Moment Tahmini)
AdamW	Adam with Weight Decay (Ağırlık Azaltmalı Adam)
BCE	Binary Cross-Entropy (İkili Çapraz Entropi)
BRCA	Breast Cancer (Meme Kanseri)
CNN	Convolutional Neural Network (Evrişimli Sinir Ağı)
CPTAC	Clinical Proteomic Tumor Analysis Consortium (Klinik Proteomik Tümör Analiz Konsorsiyumu)
CV	Computer Vision (Bilgisayarlı Görü)
CycleGAN	Cycle-Consistent Adversarial Networks (Döngüsel Tutarlı Çekişmeli Ağlar)
DCGAN	Deep Convolutional GAN (Derin Konvolüsyonel GAN)
EQ-R	Rotation Equivariance (Döndürme Eşdeğerliği)
EQ-T	Translation Equivariance (Çeviri Eşdeğerliği)
FID	Frechet Inception Distance (Frechet Inception Mesafesi)
GAN	Generative Adversarial Network (Çekişmeli Üretici Ağlar)
GAN-GP	GAN with Gradient Penalty (Gradyan Ceza ile GAN)
IRFFN	Inverted Residual Feed-Forward Network (Ters Artık İleri Beslemeli Ağ)
IS	Inception Score (Inception Skoru)
LeakyReLU	Leaky Rectified Linear Unit (Sızdıran Düzeltilmiş Doğrusal Birim)
Log-CPB	Log-Spaced Continuous Position Bias (Logaritmik Aralıklı Sürekli Konum Yanlılığı)
LoRA	Low-Rank Adaptation (Düşük Dereceli Adaptasyon)

MLP	Multi-Layer Perceptron (Çok Katmanlı Algılayıcı)
miRNA	MicroRNA (Mikro RNA)
NLP	Natural Language Processing (Doğal Dil İşleme)
OOD	Out-of-Distribution (Dağılım Dışı)
PEFT	Parameter Efficient Fine-Tuning (Parametre Etkin İnce Ayar)
PGGAN	Progressive Growing GAN (Aşamalı Büyüyen GAN)
ReLU	Rectified Linear Unit (Düzeltilmiş Doğrusal Birim)
RMSProp	Root Mean Square Propagation (Karekök Ortalama Kare Yayılımı)
SGD	Stochastic Gradient Descent (Stokastik Gradyan İnişi)
SVS	Scalable Vector Graphics (Ölçeklenebilir Vektör Grafikleri)
TCIA	The Cancer Imaging Archive (Kanser Görüntüleme Arşivi)
ViT	Vision Transformer (Görüntü Transformer)
VGG	Visual Geometry Group (Görsel Geometri Grubu)
WGAN	Wasserstein Generative Adversarial Network (Wasserstein Çekişmeli Üretici Ağlar)
WXS	Whole Exome Sequencing (Tam Ekzom Dizileme)
ZeRO	Zero Redundancy Optimizer (Sıfır Fazlalık Optimizasyonu)

ŞEKİL LİSTESİ

Şekil 2.1 GAN mimarisi 8
Şekil 2.2 StyleGAN3 mimarisi 23
Şekil 3.1 Transformer mimarisi 36
Şekil 3.2 Swin Transformer versiyon 1 mimarisi 38
Şekil 3.3 İki ardışık Swin Transformer bloğu 39
Şekil 3.4 Swin Transformer versiyon 1 bloğu 40
Şekil 3.5 Swin Transformer versiyon 2 bloğu 40
Şekil 5.1 Veri kümesinden 256×256 görüntü örnekleri. 51
Şekil 5.2 Model mimarisi 52
Şekil 5.3 Gan modelinin ürettiği sentetik histopatolojik görüntüler 54
Şekil 5.4 Veri kümesinden elde edilen histopatolojik görüntü 61
Şekil 5.5 Gan modelinin ürettiği sentetik histopatolojik görüntü 61
Şekil 6.1 Güncellenen veri kümesindeki Swin transformatör sınıflandırıcısı için karışıklık matrisi 64

TABLO LİSTESİ

Tablo 2.1 GAN varyasyonlarının özellikleri ve iyileştirmeler 21
Tablo 3.1 Swin Transformer v1 ve v2 karşılaştırması 41
Tablo 3.2 Swin Transformer ve diğer mimari yaklaşımlar: Karşılaştırmalı bir analiz 41
Tablo 5.1 Farklı modellerin ve optimizasyon algoritmalarının performans metrikleri 58
Tablo 5.2 Çoklu Modlu Büyük Dil Modeli sınıflandırıcısı performansı . . . 60
Tablo 5.3 Çoklu Modlu Büyük Dil Modeli sınıflandırıcısı evre bazlı performansı 61
Tablo 5.4 Şekil 5.4'te gösterilen histopatolojik görüntü için evre bazlı toplam logits ve olasılık değerleri 61
Tablo 5.5 Şekil 5.5'te gösterilen histopatolojik görüntü için evre bazlı toplam logits ve olasılık değerleri 62
Tablo 6.1 Swin transformer sınıflandırma performansı 64

ÖZET

Çekişmeli Üretici Ağlar ile Gerçekçi Histopatolojik Görüntü Üretimi ve Evre Tespiti

Reyhan DEDE
Bilgisayar Mühendisliği Anabilim Dalı
Bilgisayar Mühendisliği Programı
Yüksek Lisans Tezi

Danışman: Prof. Dr. Gökhan BİLGİN

Bu tez çalışması, meme kanseri evreleme süreçlerini iyileştirmek ve yeni kanser evrelerini keşfetmek amacıyla, derin öğrenme tabanlı özgün bir mimari önermektedir. Önerilen model, generatif çekişmeli ağlar (GAN'lar) ve Swin Transformer modellerinin sinerjik entegrasyonuna dayanmaktadır. Bu yaklaşım, görsel verilerin yanı sıra biyolojik ve klinik bilgileri bir araya getirerek hem yüksek kaliteli hem de hastaya özgü sentetik görüntülerin üretilmesini mümkün kılmaktadır.

Model mimarisi, iki temel aşamadan oluşmaktadır. İlk aşamada, Swin Transformer kullanılarak histopatolojik görüntülerden bağlamsal olarak zengin ve çoklu ölçeklerde ayırt edici özellikler çıkarılmaktadır. Swin Transformer'ın hiyerarşik yapısı, çıkarılan bu özelliklerin hem detaylı hem de genelleştirilebilir olmasını sağlamaktadır. Elde edilen bu özellikler, GAN'ın ayırıcı bileşenine referans olarak sunulmaktadır. İkinci aşamada ise StyleGAN3, rastgele bir gürültü vektörü ile hastaya özgü bilgileri içeren birleştirilmiş bir tensörden koşullu olarak sentetik görüntüler üretmektedir. Kombine tensör, demografik bilgiler, klinik bulgular ve genetik veriler gibi çeşitli unsurları barındırmakta olup, üretilen görüntülerin hastaya özgü karakteristiklere uyum göstermesini garanti etmektedir.

Çekişmeli eğitim süreci boyunca StyleGAN3'ün üretici ve ayırıcı bileşenleri rekabetçi bir şekilde eğitilmiştir. Bu süreçte Discriminator, gerçek ve sentetik görüntüleri Swin Transformer tarafından çıkarılan özelliklerle birlikte

değerlendirerek, görsellerin görsel gerçekçiliğini ve hasta bilgileriyle tutarlılığını denetlemiştir. Birleştirilmiş tensör, her iki bileşene de koşullandırma bilgisi sağlayarak kişiselleştirilmiş görüntü üretimini desteklemiştir.

Modelin başarımı, Frechet Inception Distance (FID) metriği ile değerlendirilmiş ve 35 gibi üstün bir skor elde edilmiştir. Ayrıca, Swin Transformer tabanlı sınıflandırma sistemi, %95.03 doğruluk, %95.00 kesinlik ve %95.00 F1 skoru ile yüksek performans göstermiştir. Yeni kanser evrelerini keşfetmek amacıyla uygulanan softmax tabanlı eşik analizi, düşük güven değerlerine sahip sınıflandırma sonuçlarını yeni evrelerin adayları olarak belirlemiştir. Bu çalışmaya ek olarak CPTAC-BRCA koleksiyonundaki histopatolojik görüntüler ve bu görüntülere ait kanser evresi etiketleri kullanılarak Qwen2-VL modeli, 4-bit quantization ve LoRA tabanlı PEFT yöntemiyle ince ayar yapılarak yaklaşık %98 F1 skoru elde edilmiştir. Ayrıca, logits değerleri üzerinden belirlenen bir eşik mekanizmasıyla Out-of-Distribution (OOD) analizi gerçekleştirilmiş ve modelin tanımlanmamış yeni evreleri tespit etme kapasitesi değerlendirilmiştir.

Bu çalışma, mevcut sınıflandırma sistemlerini geliştirmek ve kanser araştırmalarına yeni bir boyut kazandırmak amacıyla, sentetik veri üretimi ve biyolojik veri entegrasyonunu bir arada kullanmıştır. Araştırma bulguları, teşhis ve tedavi planlaması gibi uygulama alanlarında önemli katkılar sunmaktadır. Gelecek çalışmalarda, multimodal büyük dil modellerinin entegrasyonu ile daha kapsamlı analizlerin yapılması ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesi hedeflenmektedir.

Anahtar Kelimeler: Meme kanseri, StyleGAN3, Swin Transformer, çoklu omik veriler, kanser evreleme.

ABSTRACT

Realistic Histopathology Image Generation and Stage Detection with Generative Adversarial Networks

Reyhan DEDE

Department of Computer Engineering
Computer Engineering Program
Master of Science Thesis

Supervisor: Prof. Dr. Gokhan BILGIN

This thesis proposes an innovative deep learning-based architecture to enhance breast cancer staging processes and discover new cancer stages. The proposed model is based on the synergistic integration of generative adversarial networks (GANs) and Swin Transformer models. This approach enables the generation of high-quality and patient-specific synthetic images by combining visual data with biological and clinical information.

The model architecture consists of two main stages. In the first stage, Swin Transformer is employed to extract contextual and multi-scale distinctive features from histopathological images. The hierarchical structure of the Swin Transformer ensures that the extracted features are both detailed and generalizable. These features are then used as a reference in the Discriminator component of the GAN. In the second stage, StyleGAN3 generates synthetic images conditionally using a random noise vector and a combined tensor containing patient-specific information. The combined tensor incorporates various elements such as demographic information, clinical findings, and genetic data, ensuring that the generated images align with the patient-specific characteristics.

During the adversarial training process, the Generator and Discriminator components of StyleGAN3 were trained competitively. In this process, the Discriminator evaluated the visual realism and consistency with patient information by analyzing real and synthetic images along with features extracted by Swin

Transformer. The combined tensor provided conditioning information to both components, supporting personalized image synthesis.

The model's performance was evaluated using the Frechet Inception Distance (FID) metric, achieving an excellent score of 35. Additionally, the Swin Transformer-based classification system demonstrated high performance, achieving 95.03% accuracy, 95.00% precision, and 95.00% F1 score. To discover new cancer stages, softmax-based threshold analysis was applied, identifying classification results with low confidence values as potential candidates for new stages.In addition to this study, the Qwen2-VL model was fine-tuned on histopathological images from the CPTAC-BRCA collection, along with their corresponding cancer stage labels, using 4-bit quantization and a LoRA-based PEFT approach, achieving an approximately 98% F1 score. Furthermore, an Out-of-Distribution (OOD) analysis was conducted using a threshold mechanism on the logits values to evaluate the model's capacity to detect previously unidentified stages.

This study utilizes synthetic data generation and biological data integration to advance existing classification systems and introduce a new dimension to cancer research. The findings of the research provide significant contributions to application areas such as diagnosis and treatment planning. Future studies aim to conduct more comprehensive analyses and develop personalized treatment strategies by integrating multimodal large language models.

Keywords: Breast cancer, StyleGAN3, Swin Transformer, multi-omics data, cancer staging.

1
GİRİŞ

Meme kanseri, dünya genelinde kadınlar arasında en sık görülen malignite türü olup kansere bağlı ölümler arasında ikinci sırada yer almaktadır. 2024 yılı itibarıyla yalnızca Amerika Birleşik Devletleri'nde yaklaşık 2.001.140 yeni vaka ve 611.720 ölümün gerçekleşeceği tahmin edilmektedir [1]. Kadınların yaşam boyu %13 oranında meme kanseri teşhisi alma riski, bu hastalığın erken teşhis ve doğru evreleme süreçlerinin ne denli önemli olduğunu göstermektedir. Erken teşhis, yalnızca sağkalım oranlarını artırmakla kalmayıp, bireylerin yaşam kalitesini iyileştirme ve tedavi süreçlerini optimize etme potansiyeline sahiptir [2].

Histopatolojik görüntüleme, meme kanseri tanısında tümör dokusunun mikroskobik yapısının incelenmesini sağlayan en temel yöntemlerden biridir. Bununla birlikte, manuel değerlendirme süreçleri zaman alıcı, maliyetli ve subjektif hatalara açık bir yapıdadır [3]. Günümüzde, bu tür zorlukların üstesinden gelmek için derin öğrenme algoritmaları giderek daha fazla tercih edilmektedir. Özellikle Generatif Adversaryal Ağlar (GAN'lar) ve Transformer tabanlı modeller, tıbbi görüntüleme ve çoklu veri analizi alanlarında çığır açıcı yenilikler sunmaktadır [4, 5].

GAN tabanlı modeller, sentetik veri üretimindeki benzersiz kapasiteleri sayesinde veri çeşitliliğini artırarak nadir görülen tümör alt tiplerinin analizini mümkün kılmaktadır [6]. StyleGAN3 gibi ileri düzey GAN modelleri, yüksek kaliteli sentetik görüntüler üreterek literatürde veri dengesizliklerini giderme açısından sıklıkla kullanılmaktadır [2, 5]. Öte yandan, Swin Transformer gibi ViT modelleri, uzun mesafeli bağımlılıkları öğrenme kapasitesiyle biyolojik ve görsel verilerin birleştirilmesine olanak tanımaktadır [7]. Bu çalışma, StyleGAN3 ile üretilen sentetik histopatolojik görüntülerin RNA, miRNA ve klinik verilerle entegre edilerek bir derin öğrenme çerçevesi oluşturulmasını önermektedir. Model, yalnızca mevcut kanser evrelerini doğru bir şekilde sınıflandırmakla kalmayıp, düşük olasılıklı sınıflandırmaları yeni evre adayları olarak işaretleyerek literatürdeki önemli bir boşluğu doldurmayı hedeflemektedir.

1.1 Literatür İncelemesi

Kanser teşhis ve evreleme süreçlerinde yapay zekâ modellerinin kullanımı giderek yaygınlaşmaktadır. Özellikle çoklu omik verilerin (genomik, transkriptomik, epigenomik, proteomik vb.) entegrasyonu, hastalığın moleküler düzeyde daha iyi anlaşılmasına olanak tanımaktadır. Çalışmalarda, kanser teşhis ve prognozunu iyileştirmek için radyomik, patomik ve laboratuvar verileri ile yapay zekâ modellerinin entegre edildiği vurgulanmıştır [8]. Son yıllarda, erken evre meme kanseri tanısında çoklu modalite veri entegrasyonu büyük önem kazanmıştır. Meme kanseri biyobelirteçlerinin keşfinde çoklu omik ve görüntüleme verilerinin birleşimi, hasta sınıflandırması ve prognostik modelleme için kritik bir yaklaşım sunmaktadır[9].

Derin öğrenme tabanlı meme kanseri teşhis sistemleri, klinik karar destek sistemleriyle entegre edilerek daha başarılı sonuçlar elde etmektedir. Örneğin, derin öğrenme tabanlı bir Bilgisayar Destekli Teşhis (CAD) sisteminin, meme kanseri tespitinde %99 başarı oranına ulaştığı belirtilmiştir[10]. Transfer öğrenme yaklaşımları, tıbbi görüntü sınıflandırmasında önemli bir rol oynamaktadır. Derin öğrenme modellerinin tıbbi görüntü analizindeki başarısını artırmak için transfer öğrenmenin kritik bir strateji olduğu gösterilmiştir [11].

Transformer tabanlı modellerin sağlık alanındaki kullanımı giderek artmaktadır. Çok modelli Transformer mimarisi, tıbbi görüntü analizine ek olarak otomatik rapor üretimi için de kullanılmaktadır[12]. GAN modellerinin performansının değerlendirilmesi, araştırmacılar için önemli bir konu olmuştur. Yaygın olarak kullanılan Frechet Inception Distance (FID) ve Inception Score (IS) gibi metriklerin güvenilirliği sorgulanmış ve yeni nesil ölçüm yöntemleri önerilmiştir[13]. Son dönem çalışmalarda, Diffusion Transformer modellerinin GAN modellerine alternatif olarak kullanılabileceği gösterilmiştir. Difüzyon modellerinde kullanılan geleneksel U-Net mimarisinin yerine Transformer tabanlı modellerin entegrasyonunun, görüntü kalitesini artırdığı ve FID skorlarını düşürdüğü belirtilmiştir[14]. Yapay zekâ tabanlı modellerin sağlık alanında kullanımı, etik açıdan önemli bazı sorunları beraberinde getirmektedir. Sağlık hizmetlerinde generatif yapay zekâ kullanımının etik boyutları ele alınmıştır [15].

Tıbbi görüntü sınıflandırması için kullanılan derin öğrenme modelleri, özellikle Generatif Adversaryal Ağlar (GAN), Otokodlayıcılar (Autoencoders) ve uzun kısa vadeli bellek (LSTM) gibi modellerle geliştirilmiştir. Tıbbi görüntü sınıflandırması için 158 çalışmayı inceleyen bir araştırmada, CNN, transfer öğrenme, GAN ve otokodlayıcıların farklı kullanım alanları değerlendirilmiştir[16]. Yeni nesil tıbbi görüntü analizinde multimodal Transformer kullanımı yaygınlaşmaktadır. Metin

ve görüntü verilerinin birlikte işlendiği multimodal Transformer mimarileri kullanılarak tıbbi raporlama süreçlerinde önemli iyileştirmeler sağlandığı belirtilmiştir [17].

Tıbbi görüntülemede derin üretken modellerin kullanımı son yıllarda büyük ilgi görmüştür. Tıbbi görüntü işleme süreçlerinde Çekişmeli Üretici Ağlar (GAN), Varyasyonel Otokodlayıcılar (VAE) ve Difüzyon Modelleri'nin (DM) karşılaştırmalı analizi yapılmıştır[18]. Transformer tabanlı üretken modellerin gelişimi ele alınmış ve özellikle tıbbi görüntü üretimindeki kullanım potansiyeli tartışılmıştır[19]. Kendi kendini denetleyen öğrenme, tıbbi görüntü sınıflandırmasında son yıllarda popüler hale gelmiştir. Farklı kendi kendini denetleyen öğrenme stratejilerinin tıbbi görüntüleme üzerindeki etkileri ele alınarak, bu modellerin etiketleme gereksinimini nasıl azalttığı gösterilmiştir[20].

Multimodal öğrenme teknikleri, yetişkin ve pediatrik beyin tümörleri için prognoz tahmini yapmada etkili olmuştur. Histopatoloji görüntülerinin gen ekspresyon profilleriyle birleştirilmesiyle geliştirilen bir derin öğrenme modelinin, tekil veri türlerine dayanan modellere göre daha iyi performans gösterdiği belirtilmiştir[8]. Günümüzde kanser teşhis ve evreleme süreçlerinde GAN ve Transformer tabanlı yapay zekâ modellerinin entegrasyonu önemli bir araştırma alanı haline gelmiştir. Çoklu omik verilerin ve tıbbi görüntülerin birleştirilmesi, daha doğru tahmin modellerinin geliştirilmesine olanak sağlamaktadır. Ancak, kullanılan metriklerin güvenilirliği ve etik kaygılar gibi faktörler, bu teknolojilerin geniş ölçekli uygulanabilirliği için dikkate alınması gereken temel unsurlardır.

Meme kanseri teşhisi ve evreleme süreçlerinde derin öğrenme tabanlı modeller son yıllarda önemli ilerlemeler sağlamıştır. Özellikle Çekişmeli Üretici Ağlar (GAN'lar), tıbbi görüntü analizi alanında devrim yaratan teknolojiler olarak ön plana çıkmaktadır. GAN'lar, sentetik veri üretimindeki üstün performansları sayesinde, eksik veri sorunlarının üstesinden gelmekte ve nadir görülen tümör alt tiplerinin analizine olanak tanımaktadır [2, 6]. GAN tabanlı modeller, meme kanseri araştırmalarında veri çeşitliliğini artırmak için sıklıkla kullanılmıştır ve StyleGAN3, bu modeller arasında yüksek kaliteli sentetik görüntüler üretme kapasitesiyle dikkat çekmektedir [5]. GAN'ların temel avantajlarından biri, yüksek kaliteli sentetik görüntüler üreterek veri dengesizliklerini azaltma kapasitesidir. Bu özellik, özellikle nadir görülen tümör alt tiplerinin analizinde araştırmacılara benzersiz bir fırsat sunmaktadır [2]. Örneğin, Mirimoghaddam ve arkadaşlarının çalışmasında, GAN tabanlı yöntemlerle HER2 pozitif meme kanseri veri setleri üzerinde başarılı sonuçlar elde edilmiş, bu da GAN'ların tıbbi görüntüleme uygulamalarında kullanım potansiyelini ortaya koymuştur [6]. Bununla birlikte, GAN'ların eğitim

süreçlerinde kararlılık sorunları ve üretilen görüntülerin doğruluğunun artırılması gibi zorluklar bulunmaktadır. Bu sorunların çözümü için gradyen cezalandırması ve spektral normalizasyon gibi yöntemler geliştirilmiştir [21].

GAN'ların yanında, Vision Transformer (ViT) modelleri de tıbbi görüntü analizi ve sınıflandırma süreçlerinde devrim yaratmıştır. ViT modelleri, uzun mesafeli bağımlılıkları öğrenme kapasitesi ve büyük ölçekli veri kümelerini işleyebilme yetenekleri sayesinde tıbbi araştırmalarda yaygın olarak kullanılmaktadır [4]. Swin Transformer, yerel pencere dikkat mekanizmasıyla hem yerel hem de küresel özellikleri öğrenme kapasitesine sahiptir ve bu özellikleriyle histopatolojik görüntülerin analizi için optimize edilmiştir [7]. ViT modellerinin avantajları arasında, biyolojik ve görsel verilerin birleştirilmesine olanak tanıyan kapasitesi bulunmaktadır. Bu, özellikle çoklu-omik verilerin histopatolojik görüntülerle entegrasyonu gibi kompleks görevlerde önem taşımaktadır [22].

Bilgisayarlı görme alanındaki hızlı gelişmeler, yaygın olarak kullanılan Evrişimli Sinir Ağlarına (Convolutional Neural Networks - CNN) alternatif olarak Transformer tabanlı mimarilerin keşfedilmesine yol açmıştır. Son çalışmalar, Vision Transformer ve Swin Transformer v2 gibi Transformer tabanlı modellerin, tıbbi görüntü analizi de dahil olmak üzere çeşitli bilgisayarlı görme görevlerinde üstün performans sergileyebileceğini göstermektedir [23, 24]. Vision Transformer modeli, Dosovitskiy ve ekibi tarafından tanıtılmış [23] ve görüntü sınıflandırma görevlerinde etkileyici sonuçlar elde etmiştir. Küresel alıcı alanı (global receptive field) ve kendine dikkat (self-attention) mekanizmalarını kullanarak, Vision Transformer uzun menzilli bağımlılıkları yakalayabilir ve karmaşık görsel desenleri etkili bir şekilde modelleyebilir. Ancak, Vision Transformer'ın performansı, girdi görüntüsünün boyutuna duyarlılığı ve kendine dikkat hesaplamalarının getirdiği yüksek hesaplama maliyeti nedeniyle sınırlı olabilir. Bu sınırlamaları aşmak için, Liu ve ekibi tarafından tanıtılan Swin Transformer v2, kaydırılmış pencereleme (shifted windowing) şeması kullanan hiyerarşik bir Transformer mimarisi önermektedir. Bu yaklaşım, Swin Transformer v2'nin çok ölçekli görsel temsilleri verimli bir şekilde modellemesini sağlarken, girdi görüntüsünün boyutuna göre doğrusal hesaplama karmaşıklığını korumasına olanak tanır.

Vision Transformer ve Swin Transformer v2'nin tıbbi görüntü analizi görevlerindeki performansları karşılaştırıldığında, çeşitli çalışmalar değerli içgörüler sunmaktadır. Transformer tabanlı modellerin tıbbi görüntü analizindeki başarısı büyük ölçüde uzun menzilli bağımlılıkları yakalama yeteneklerine ve sınırlı eğitim verileriyle başa çıkma dayanıklılıklarına bağlanmaktadır. Bununla birlikte, özellikle yüksek çözünürlüklü tıbbi görüntülerle çalışırken, Transformer

tabanlı modellerin hesaplama karmaşıklığı bir zorluk olmaya devam etmektedir. Bu sorunu çözmek için, araştırmacılar, CS-UNet modelinde olduğu gibi, Transformer mimarisine evrişim katmanlarının entegrasyonunu içeren çeşitli stratejiler önermiştir [25]. Bu bulgular, Transformer tabanlı modellerin bilgisayarlı görme, özellikle tıbbi görüntüleme alanındaki potansiyelini vurgulamakta ve performanslarını ve hesaplama verimliliklerini optimize etmek için sürekli araştırma yapılmasının önemini ortaya koymaktadır.

Multimodal veri entegrasyonu, meme kanseri teşhisinde biyobelirteçlerin keşfi ve teşhis doğruluğunun artırılması için umut vadeden bir yaklaşımdır. Lipkova ve arkadaşlarının çalışmasında, yapay zekâ yöntemleri kullanılarak multimodal veri entegrasyonunun, kanser teşhisinde doğruluğu artırdığı gösterilmiştir [22]. Ancak, mevcut literatür analiz edildiğinde, bu entegrasyonu tam anlamıyla gerçekleştiren çalışmaların sınırlı olduğu görülmektedir [26]. Özellikle RNA, miRNA ve klinik verilerin histopatolojik görüntülerle birleştirilmesine odaklanan çalışmalar, sınırlı sayıda kalmaktadır. Çoğu çalışma tek bir modaliteye odaklanmakta ve bu durum, analizlerin kapsamını ve derinliğini sınırlamaktadır [22].

Mevcut literatür, yeni kanser evrelerinin keşfi üzerine yapılan çalışmaların eksikliğine işaret etmektedir. Çoğu model, yalnızca mevcut evrelerin sınıflandırılmasına odaklanmakta ve düşük olasılıklı sınıflandırmaları yeni evrelerin adayları olarak işaretleme kapasitesinden yoksundur [27]. Bu noktada, GAN ve ViT tabanlı modellerin entegrasyonu, yeni evrelerin keşfini mümkün kılabilecek bir çözüm olarak ortaya çıkmaktadır [28]. Özellikle Swin Transformer tabanlı modellerin, düşük olasılıklı sınıflandırmaları analiz ederek yeni biyobelirteçlerin keşfi için fırsatlar sunduğu gösterilmiştir [4].

GAN ve ViT tabanlı modellerin klinik uygulamalara entegrasyonu konusunda da bazı sınırlamalar bulunmaktadır. Klinik doğrulama süreçlerinin yetersiz olması, bu modellerin sağlık hizmetlerinde daha yaygın bir şekilde kullanılmasını engellemektedir [29]. Bu eksikliklere rağmen, GAN ve ViT modellerinin entegrasyonu, literatürde önemli bir boşluğu doldurma potansiyeline sahiptir. Örneğin, Ballard ve arkadaşlarının çalışması, çoklu-omik verilerin entegre edilmesiyle biyolojik keşiflerin hızlandırılabileceğini göstermiştir [26].

Mevcut literatür, meme kanseri teşhisi ve evrelemesinde derin öğrenme modellerinin önemli ilerlemeler kaydettiğini göstermektedir. Ancak, bu alanda hala önemli eksiklikler bulunmaktadır. Özellikle, multimodal veri entegrasyonunun sınırlı olması, yeni kanser evrelerinin keşfine yönelik çalışmaların yetersiz kalması ve klinik uygulamalarda karşılaşılan doğrulama zorlukları, bu alanda daha kapsamlı

ve yenilikçi yaklaşımlara duyulan ihtiyacı ortaya koymaktadır. GAN'lar ve Transformer modelleri gibi ileri teknolojilerin potansiyeli, bu eksiklikleri gidermek için umut vaat etmektedir. Bu bağlamda, bu tez, literatürdeki bu boşlukları doldurmak ve meme kanseri evrelemesinde yeni bir perspektif sunmak amacıyla tasarlanmıştır.

1.2 Tezin Amacı

Bu çalışmanın temel amacı, literatürdeki eksiklikleri ele alarak meme kanseri evrelemesi ve yeni evrelerin keşfi için yenilikçi bir derin öğrenme çerçevesi geliştirmektir. Özellikle, StyleGAN3 modeli kullanılarak üretilen yüksek kaliteli sentetik histopatolojik görüntülerin, RNA, miRNA ve klinik veriler gibi multimodal biyolojik verilerle entegre edilmesi hedeflenmektedir. Bu entegrasyon, Swin Transformer tarafından çıkarılan görüntü özellikleriyle birleştirilerek hem görsel hem de biyolojik bilgilerin bir arada kullanıldığı bir sistem oluşturulacaktır. Bu sayede, mevcut evreleme yöntemlerinin sınırlamaları aşılacak ve daha doğru kanser evreleme sistemleri geliştirilecektir. Ayrıca, sınıflandırma sırasında düşük olasılıklı sınıflandırmaların yeni evre adayları olarak işaretlenmesiyle, literatürde önemli bir boşluk doldurulacaktır. Bu çalışma, ileri derin öğrenme yöntemleri ve multimodal veri entegrasyonuyla, meme kanseri teşhis ve tedavisinde kişiselleştirilmiş stratejilere katkı sağlamayı amaçlamaktadır.

1.3 Hipotez

Bu tezde, multimodal veri entegrasyonu ve gelişmiş derin öğrenme modelleri kullanılarak meme kanseri evreleme doğruluğunun artırılabileceği ve yeni evrelerin keşfedilebileceği öne sürülmektedir. Özellikle, RNA, miRNA ve klinik veriler gibi biyolojik bilgilerin, Swin Transformer ile histopatolojik görüntülerden çıkarılan özelliklerle birlikte StyleGAN3 modeline entegre edilmesinin, üretilen sentetik görüntülerin biyolojik gerçekliğini artırarak sınıflandırma performansını iyileştireceği düşünülmektedir. Ayrıca, sınıflandırma sırasında kullanılan softmax tabanlı eşik değer analizi ile mevcut evrelerden farklılık gösteren ve daha önce tanımlanmamış yeni kanser evrelerinin belirlenebileceği hipotez edilmektedir. Çalışma, görsel ve biyolojik bilgilerin bir arada kullanıldığı bu yaklaşımın, hem meme kanseri evrelemesinde hem de literatürde önemli katkılar sağlayacağını öngörmektedir.

2
GENERATİF ÇEKİŞMELİ AĞLAR (GANs)

Generatif Çekişmeli Ağlar (GANs), Ian Goodfellow tarafından 2014 yılında tanıtılmış ve jeneratif modelleme alanında çığır açan bir yaklaşımla, iki sinir ağı arasındaki rekabeti temel almıştır. Bu iki ağ, üretici (generator) ve ayrıştırıcı (discriminator), karşılıklı olarak birbirini geliştiren bir süreç içinde çalışır. Üretici ağ, rastgele girdi verilerini kullanarak gerçek veri örneklerine benzer sentetik veriler oluştururken, ayrıştırıcı ağ ise bu verilerin gerçek mi yoksa sentetik mi olduğunu ayırt etmeye çalışır. GAN'ların bu özgün yapısı, görüntü sentezi, doğal dil işleme, biyomedikal veri analizi ve daha birçok alanda yenilikçi uygulamaların önünü açmıştır. [30, 31]

2.1 Temel Mimari

GAN'ların mimarisi, birbirleriyle çekişmeli bir süreç içerisinde çalışan iki temel bileşenden oluşmaktadır: Üretici Ağ (Generator) ve Ayrıştırıcı Ağ (Discriminator). Bu bileşenlerin etkileşimi, GAN'ların öğrenme mekanizmasının temelini teşkil eder ve her iki ağın performansını eşzamanlı olarak iyileştiren bir öğrenme dinamiği oluşturur [30, 31]. Şekil 2.1 'de bu iki bileşenin nasıl etkileşimde bulunduğu ve temel mimari detayları gösterilmektedir.

Üretici Ağ (Generator)

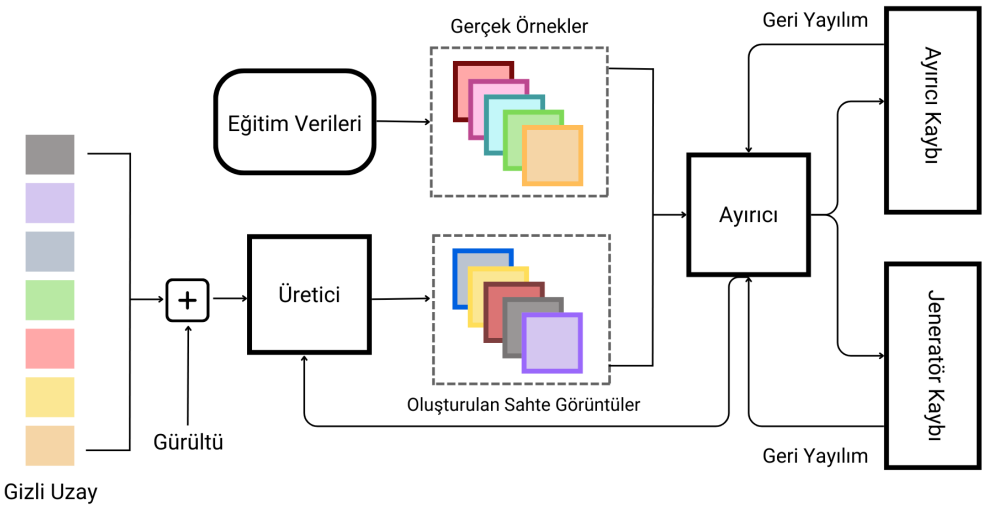
Üretici ağ, belirli bir olasılık dağılımından ($P_z(z)$) örneklenen rastgele bir girdi vektörünü (z) kullanarak, gerçek veri dağılımına benzer sentetik veri örnekleri oluşturan bir sinir ağıdır. Bu süreç matematiksel olarak şu şekilde ifade edilir:

$$G(z; \theta_g) \tag{2.1}$$

Burada 2.1:

- G, üretici ağı temsil eder.
- z, rastgele girdi vektörüdür.
- θ_g, üretici ağın öğrenilebilir parametrelerini ifade eder.

Üretici ağın nihai amacı, oluşturduğu verilerin ayrıştırıcı ağ tarafından gerçek veri olarak sınıflandırılmasını sağlamaktır. Bu doğrultuda, üretici ağ, gerçek veri dağılımını başarılı bir şekilde taklit edebilmek için parametrelerini optimize eder. Şekil 2.1'de görüldüğü üzere, üretici ağ, gizli uzaydan örneklenen gürültüyü işleyerek sahte görüntüler üretir ve bu görüntüleri ayrıştırıcıya gönderir. Bu süreç, GAN sisteminde minimaks optimizasyon problemi ile yönetilir [32, 33].



Şekil 2.1 GAN mimarisi

Ayrıştırıcı Ağ (Discriminator)

Ayrıştırıcı ağ, bir veri örneğinin gerçek ($x \sim P_{data}$) mi yoksa üretici ağ tarafından oluşturulmuş sahte ($G(z)$) bir örnek mi olduğunu ayırt etmek üzere eğitilen bir sinir ağıdır. Ayrıştırıcının verdiği çıktı şu şekilde tanımlanır:

$$D(x) \in [0, 1] \tag{2.2}$$

Burada 2.1:

- $D(x)$, x veri örneğinin gerçek olma olasılığını belirtir.

- $D(x) = 1$, verinin gerçek olduğunu, $D(x) = 0$ ise sahte olduğunu ifade eder.

Ayrıştırıcı ağ, iki temel hedef doğrultusunda optimize edilir:

1. Gerçek veri dağılımından ($x \sim P_{data}(x)$) gelen veriler için $D(x)$ değerini 1'e yaklaştırmak.
2. Üretici ağ tarafından oluşturulan sahte veriler için $D(G(z))$ değerini 0'a yaklaştırmak.

Ayrıştırıcı, verilerin gerçekliğini değerlendirme yeteneğini sürekli geliştirmekle birlikte, üretici ağın bu ayrımı yanıltma çabalarıyla karşı karşıya kalır. Bu süreç, üretici ve ayrıştırıcı arasında bir denge noktasına ulaşmayı hedefleyen dinamik bir öğrenme mekanizmasını teşkil eder. [33]

2.2 Çekişmeli Eğitim Çerçevesi

Belirtildiği gibi Generatif Çekişmeli Ağlar (GAN'lar), eğitim sürecinde minimaks bir optimizasyon problemi olarak ele alınmaktadır. Bu yapı, üretici (G) ve ayrıştırıcı (D) ağlar arasındaki dinamik bir ilişki üzerine kuruludur. Eğitim süreci aşağıdaki amaç fonksiyonu ile ifade edilir:

$$\min_G \max_D V(D,G) = E_{x\sim P_{data}(x)}[\log D(x)] + E_{z\sim P_z(z)}[\log(1 - D(G(z)))]. \quad (2.3)$$

Bu 2.2 denklemin her bir terimi şu şekilde yorumlanabilir:

- $D(x)$: Ayrıştırıcı ağın, x verisinin gerçek olduğunu tahmin etme olasılığıdır.
- $G(z)$: Üretici ağın, rastgele bir z gürültü vektörünü gerçek veriye benzer bir örneğe dönüştürmesidir.
- $P_{data}(x)$: Gerçek veri dağılımını temsil eder.
- $P_z(z)$: Gürültü vektörünün tanımlandığı olasılık dağılımıdır. [30]

Eğitim sürecinin temel prensipleri şunlardır:

1. Ayrıştırıcı ağ (D), gerçek veriler ($x \sim P_{data}$) için $\log D(x)$'i maksimize etmeye çalışır. Bu, ayrıştırıcının gerçek verileri tanımadaki başarısını artırır.

2. Ayrıştırıcı, üretici tarafından oluşturulan sentetik veriler ($x \sim G(z)$) için $\log(1 - D(G(z)))$'yi maksimize ederek bu verilerin sahte olduğunu doğru şekilde ayırt etmeye çalışır.

Bu noktada üretici ağ (G), ayrıştırıcının hata oranını artırmayı hedefler. Bu amaçla, $\log(1 - D(G(z)))$'yi minimize eder ve ayrıştırıcının ürettiği verileri gerçek olarak sınıflandırmasını sağlamaya çalışır [31] . Eğitim süreci, üretici ve ayrıştırıcı ağlar arasında bir çekişme yaratır ve bu çekişme, her iki ağın da performansını artırır.

GAN'ların bu yapısı, oyun teorisi bağlamında Nash dengesine dayanır. Nash dengesine ulaşıldığında, üretici ağın oluşturduğu sentetik veriler, ayrıştırıcı ağ için gerçek ve sahte olarak ayırt edilemez hale gelir. Bu denge, GAN modellerinin optimal performans gösterdiği durumu ifade eder [34].

2.2.1 Yakınsama ve Nash Dengesi

GAN'ların teorik olarak optimal performansı, üretici ağın gerçek veri dağılımını P_{data} tam olarak kopyaladığı ve ayrıştırıcı ağın tüm veri örnekleri için %50 olasılıkla karar verdiği durumda elde edilir. Bu durum, GAN'ların Nash dengesi olarak adlandırılır. Nash dengesinde, üretici ağ tarafından oluşturulan veriler, ayrıştırıcı ağ tarafından gerçek ve sahte olarak ayırt edilemez hale gelir. Bu eşik noktasında, GAN modeli gerçekçi ve çeşitli veri örnekleri üretme yeteneğini en üst düzeye çıkarır [34].

2.3 GAN Varyasyonları

Generatif Çekişmeli Ağlar (GAN'ler), derin öğrenme alanında veri üretimi ve sentezi konusunda devrim niteliğinde bir yenilik sunmuş ve geniş bir uygulama alanına sahip olacak şekilde geliştirilmiştir. Bu modellerin temel prensibi, bir üretici ve bir ayırt edici ağın karşılıklı rekabeti yoluyla gerçekçi veri üretmesini sağlamaktır. Klasik GAN mimarisinin sınırlamaları ve uygulama alanlarına yönelik artan gereksinimler doğrultusunda, zaman içinde farklı varyasyonlar geliştirilerek modelin kararlılığı artırılmış, eğitim süreci iyileştirilmiş ve üretilen verinin kalitesi yükseltilmiştir.

İlk olarak Conditional GAN (cGAN), modelin çıktısının belirli koşullar altında yönlendirilebilmesine olanak tanımış, ardından Deep Convolutional GAN (DCGAN), evrişimli sinir ağlarının kullanımıyla daha gerçekçi görüntü üretimi için optimize edilmiştir. Sonrasında geliştirilen InfoGAN, üretilen verinin anlamlı ve yorumlanabilir temsiller içermesini amaçlamış, Least Squares GAN (LSGAN) ve

Wasserstein GAN (WGAN) gibi yaklaşımlar ise eğitim sürecindeki istikrarsızlıkları gidermek, mod çökmesini önlemek ve üretilen verinin gerçek veri dağılımına daha yakın olmasını sağlamak üzere optimize edilmiştir.

Öte yandan, Pix2Pix ve CycleGAN, görüntüden-görüntüye dönüşüm (image-to-image translation) problemlerine yönelik önemli çözümler sunarak, farklı görüntüleme modaliteleri arasında dönüşüm yapılmasını mümkün kılmıştır. Ayrıca, Super-Resolution GAN (SRGAN) ile düşük çözünürlüklü görüntülerin yüksek çözünürlüğe ölçeklenmesi sağlanarak, detay kaybının önüne geçilmiştir.

Daha sonra geliştirilen Progressive Growing GAN (ProGAN), modelin eğitim sürecini düşük çözünürlükten yüksek çözünürlüğe kademeli olarak ilerletecek şekilde tasarlamış ve böylece yüksek kaliteli, ayrıntılı görüntüler üretebilen daha kararlı bir yapı sunmuştur. BigGAN, büyük ölçekli veri kümeleri üzerinde eğitilerek, daha yüksek çözünürlükte ve gerçekçiliği artırılmış görüntülerin üretimine olanak tanımıştır. Son olarak, StyleGAN, üretilen görüntülerin belirli stil ve içerik unsurlarına göre kontrol edilebilmesini sağlayarak, GAN mimarisi alanında önemli bir yenilik sunmuştur.

Koşullu Üretici Çekişmeli Ağ (Conditional GAN - cGAN)

Klasik GAN modeli yalnızca rastgele girdi vektörlerinden (z) veri üretirken, cGAN bu sürece ek bilgi (örneğin sınıf etiketleri veya ek özellikler) dahil ederek üretim sürecini yönlendirir [35]. cGAN'in temel motivasyonu, üretilen verinin belirli kategorilere uygun olmasını sağlamaktır. Bu modelin en büyük katkılarından biri, tıbbi görüntü sentezi ve kanser evreleme gibi alanlarda belirli hastalık kategorilerine özel veriler üretme yeteneğidir [36]. cGAN ayrıca, veri artırma (data augmentation) süreçlerinde yaygın olarak kullanılmakta olup, özellikle biyomedikal veri setlerinin genişletilmesi ve MRI görüntülerinin spesifik hastalık türlerine göre üretilmesi gibi alanlarda büyük bir başarı göstermiştir.

Derin Evrişimli Üretici Çekişmeli Ağ (Deep Convolutional GAN - DCGAN)

DCGAN, geleneksel GAN mimarisine evrişimli sinir ağları (CNN) eklenerek geliştirilmiş bir varyasyondur ve özellikle yüksek çözünürlüklü görüntü üretimi için optimize edilmiştir [37]. Standart tam bağlı katmanlar yerine evrişimli katmanlar kullanılması, modelin daha ayrıntılı ve gerçekçi görüntüler oluşturmasına olanak sağlamıştır. DCGAN'in en önemli avantajlarından biri, eğitim sürecinde kararlılığı artırmak ve mod çökmesini önlemektir. Özellikle batch normalization tekniğinin uygulanması, modelin daha iyi genelleştirme yapmasını sağlamış ve Leaky ReLU

aktivasyon fonksiyonu ile negatif değerlerin öğrenilmesi kolaylaştırılmıştır. Bu yenilikler sayesinde DCGAN, tıbbi görüntüleme, biyomedikal veri sentezi, sahte yüz üretimi ve sanat tabanlı veri üretimi gibi birçok alanda başarıyla kullanılmıştır [38].

Bilgi Maksimizasyonlu Üretici Çekişmeli Ağ (Interpretable Representation Learning GAN - InfoGAN)

Standart GAN modellerinde, üretici ağın kontrol edilebilir özellikleri öğrenmesi zor olabilir. InfoGAN (Chen et al., 2016), bilgi teorisi tabanlı bir yöntem kullanarak üretici ağın belirli özellikleri kontrol edebilmesini sağlar [36]. InfoGAN'in en büyük avantajlarından biri, üretici ağın gizli değişkenler aracılığıyla belirli özellikleri öğrenmesini sağlamasıdır. Örneğin, yüz sentezi sırasında saç rengi, yüz ifadesi veya pozisyon gibi değişkenlerin kontrol edilmesine olanak tanır. Tıbbi görüntüleme ve biyometrik güvenlik sistemleri gibi alanlarda yaygın olarak kullanılmaktadır.

En Küçük Kareler Üretici Çekişmeli Ağ (Least Squares GAN - LSGAN)

Klasik GAN modellerinde kullanılan lojistik kayıp fonksiyonları, eğitim sırasında gradyan vanishing problemlerine neden olabilmektedir. LSGAN (Mao et al., 2017), kare hata (least squares) kayıp fonksiyonunu kullanarak discriminator'ın daha güçlü ve tutarlı geri bildirimler vermesini sağlamaktadır [37]. Bu yöntem, mod çökmesini azaltarak daha dengeli bir eğitim süreci sunmakta ve üretici ağın ürettiği örneklerin gerçek verilerle daha fazla örtüşmesini sağlamaktadır. LSGAN, görüntü sentezi, sahte veri tespiti ve biyomedikal veri artırımı gibi alanlarda kullanılmaktadır.

Wasserstein Üretici Çekişmeli Ağ (Wasserstein GAN - WGAN)

Generatif Çekişmeli Ağlar (GAN'ler), veri üretiminde önemli başarılar elde etmiş olsa da, eğitim süreçlerinde istikrarsızlık ve mod çökmesi gibi temel problemlerle karşılaşmaktadır. Bu sorunların temel nedenlerinden biri, geleneksel GAN'lerde kullanılan Jensen-Shannon (JS) uzaklığının, üretilen veri dağılımı ile gerçek veri dağılımı arasındaki farkı yeterince hassas bir şekilde ölçememesidir. JS uzaklığı, iki dağılımın destek kümeleri (yani, olasılık yoğunluğu sıfırdan farklı olan bölgeleri) örtüşmediğinde maksimum değere ulaşır ve bu durumda üretici model için öğrenme süreci kesintiye uğrar.

Bu sorunu çözmek amacıyla geliştirilen Wasserstein GAN (WGAN), iki dağılım arasındaki farkı ölçmek için Wasserstein mesafesi (Earth Mover's Distance - EMD)

kullanarak daha stabil ve anlamlı bir optimizasyon süreci sunmaktadır. Wasserstein mesafesi, bir olasılık dağılımını başka bir olasılık dağılımına dönüştürmek için gereken minimum taşıma maliyetini ölçmektedir. Matematiksel olarak, bu mesafe, bir dağılımdaki kütlenin diğerine taşınması için gereken en az iş miktarını ifade eder. GAN'lerin eğitim sürecinde, Wasserstein mesafesi kullanılarak, üretici modelin her adımda ürettiği veri dağılımının gerçek veri dağılımına ne kadar yaklaştığı ölçülmekte ve böylece daha sağlıklı bir optimizasyon süreci sağlanmaktadır.

Ancak, Wasserstein mesafesinin doğru bir şekilde optimize edilebilmesi için modelin Lipschitz sürekliliği koşulunu sağlaması gerekmektedir. Lipschitz sürekliliği, bir fonksiyonun giriş değişimlerine karşı aşırı duyarlı olmamasını garanti eden bir matematiksel özelliktir. Geleneksel GAN'lerde gözlemlenen eğitim istikrarsızlığı, genellikle kayıp fonksiyonunun ani dalgalanmalar yaşamasından kaynaklanmaktadır. WGAN, bu sorunu gidermek için discriminator ağında ağırlıkların belirli bir aralıkta tutulmasını sağlayan weight clipping (ağırlık sıkıştırma) yöntemini kullanmaktadır. Ancak, bu yöntemin öğrenme sürecine bazı kısıtlamalar getirdiği gözlemlenmiş ve daha sonraki çalışmalar, gradient penalty (WGAN-GP) adı verilen bir teknik ile Lipschitz koşulunun daha etkin bir şekilde sağlanabileceğini göstermiştir.

WGAN'in sağladığı en önemli avantajlardan biri, GAN eğitim sürecinin daha stabil hale getirilmesi ve mod çökmesi probleminin büyük ölçüde azaltılmasıdır. Geleneksel GAN'lerde, üretici ağ bazen yalnızca belirli bir alt küme üzerinde yoğunlaşarak çeşitliliği kaybedebilirken, WGAN sürekli ve farklılaştırılabilir bir öğrenme sinyali sunduğundan, modelin tüm veri dağılımını daha etkili bir şekilde öğrenmesini sağlamaktadır. Ayrıca, JS uzaklığı yerine Wasserstein mesafesinin kullanılması, GAN modellerinin özellikle büyük ölçekli ve karmaşık veri kümelerinde daha güvenilir bir şekilde eğitilmesine olanak tanımaktadır [31, 34, 39].

Görüntüden Görüntüye Dönüştürme Üretici Çekişmeli Ağ (Pix2Pix)

Pix2Pix, görüntüden-görüntüye çeviri (image-to-image translation) problemlerini çözmek amacıyla geliştirilmiş bir Koşullu Generatif Çekişmeli Ağ (Conditional GAN - cGAN) modelidir. Geleneksel GAN yapılarından farklı olarak, rastgele bir girdi vektörü yerine doğrudan giriş görüntüsü kullanılarak belirli bir girişten hedeflenen bir çıktının üretilmesini amaçlar. Model, özellikle eşlenmiş veri kümeleriyle (paired dataset) eğitildiğinde yüksek doğrulukta sonuçlar üretme kapasitesine sahiptir [40].

Pix2Pix mimarisi, U-Net tabanlı bir üretici (generator) ve PatchGAN tabanlı bir ayrıştırıcı (discriminator) içerir. Üretici ağ olarak kullanılan U-Net, giriş görüntüsünü kodlayarak düşük boyutlu bir temsile dönüştürmekte ve ardından bu temsili kullanarak yüksek çözünürlüklü bir çıktı üretmektedir. Skip connection mekanizması, giriş görüntüsündeki düşük seviyeli detayların çıktı görüntüsüne daha iyi aktarılmasını sağlar. PatchGAN ayrıştırıcısı, geleneksel tam görüntü değerlendirmesi yerine görüntüyü küçük yamalar halinde analiz ederek gerçek ya da sahte ayrımını yapar. Bu sayede, modelin yerel detayları öğrenme kabiliyeti artırılmıştır [41].

Pix2Pix modelinin kayıp fonksiyonu, Generatif Çekişmeli Ağ (GAN) kaybı ile L1 kaybının birleşiminden oluşmaktadır. GAN kaybı, üretilen görüntünün gerçek veri dağılımına ne kadar yakın olduğunu değerlendirirken, L1 kaybı ise giriş ve çıktı görüntüleri arasındaki piksel bazlı farkı en aza indirerek modelin daha stabil ve hızlı eğitilmesini sağlar [40].

Pix2Pix modeli, görüntüden-görüntüye dönüşüm gerektiren birçok alanda başarıyla uygulanmıştır. Uydu görüntülerinden harita üretimi, tıbbi görüntü segmentasyonu, kara kalem çizimlerinden fotoğraf benzeri görüntü üretimi, görüntü renklendirme ve görsel stil transferi gibi çeşitli uygulama alanlarında etkili sonuçlar elde edilmiştir. Tıbbi görüntüleme alanında, MRI ve röntgen görüntülerinin işlenmesi ve analiz edilmesi gibi uygulamalarda yaygın olarak kullanılmaktadır [42].

Ancak, Pix2Pix'in en büyük sınırlamalarından biri, eğitim sürecinde eşlenmiş veri kümelerine (paired dataset) ihtiyaç duymasıdır. Modelin başarısı büyük ölçüde kullanılan veri kümesinin kalitesine bağlı olup, bu tür veri setlerinin oluşturulması çoğu durumda zaman alıcı ve maliyetlidir. Bu eksikliği gidermek amacıyla CycleGAN gibi eşlenmemiş veri setleriyle çalışabilen alternatif modeller geliştirilmiştir. Ayrıca, Pix2PixHD gibi gelişmiş versiyonlar, daha yüksek çözünürlüklü görüntü üretmek ve daha ayrıntılı sonuçlar elde etmek için mimari iyileştirmeler içermektedir [40].

Sonuç olarak, Pix2Pix modeli, koşullu GAN'lerin en önemli örneklerinden biri olarak görüntüden-görüntüye çeviri alanında devrim niteliğinde bir yaklaşım sunmuş, ancak belirli sınırlamaları nedeniyle yeni jenerasyon modellerin gelişmesine de zemin hazırlamıştır. Bu bağlamda, Pix2Pix'in sağladığı yenilikler ve sahip olduğu mimari özellikler, günümüzde birçok farklı alanda temel teşkil eden yöntemlerin ortaya çıkmasını sağlamış ve görüntü işleme alanında büyük bir etki yaratmıştır.

Eşleşmemiş Görüntüden Görüntüye Dönüştürme Üretici Çekişmeli Ağ (Cycle-GAN)

Görüntüden görüntüye dönüşüm (image-to-image translation) problemlerinde, geleneksel GAN modelleri genellikle eşleşmiş veri kümeleri gerektirmektedir. Ancak, birçok pratik uygulamada, her bir görüntünün doğrudan bir karşılığı olan etiketli bir veri kümesi bulunmamaktadır. CycleGAN, bu tür eşleşmemiş veri kümeleri ile iki farklı veri kümesi arasında dönüşüm gerçekleştirmek için geliştirilmiş bir GAN modelidir.

CycleGAN, iki farklı domain arasında dönüşüm yapabilen iki ayrı üretici ve iki ayrı ayırt edici ağdan oluşmaktadır. Modelin temel amacı, bir görüntü domaininden diğerine dönüşüm gerçekleştirirken, geriye dönüş işlemi yapıldığında orijinal görüntünün geri elde edilebilmesini sağlamaktır. Bu amaç doğrultusunda Cycle Consistency Loss (Çevrim Tutarlılığı Kaybı) adı verilen özel bir kayıp fonksiyonu kullanılmaktadır. Eğer bir görüntü önce bir domain'e dönüştürülüp, ardından ters dönüşüm yapıldığında başlangıçtaki görüntüye yakın bir sonuç üretemezse, model cezalandırılmaktadır.

CycleGAN'in optimizasyon süreci, geleneksel adversarial kayıp fonksiyonuna ek olarak çevrim tutarlılığı kaybının da dikkate alınmasını gerektirmektedir. Adversarial kayıp fonksiyonu, üretilen görüntülerin gerçek olup olmadığını ayırt edici ağlar aracılığıyla ölçerken, çevrim tutarlılığı kaybı modelin ürettiği çıktının istikrarlı ve geri dönüşümlü olmasını garanti eder [32, 43, 44]. Yüksek çözünürlüklü görüntü üretimi, geleneksel GAN'ler için önemli bir zorluk teşkil etmektedir. Standart GAN modelleri, yüksek çözünürlüklü görüntüler üretmeye çalışırken genellikle eğitim sürecinde dengesizlikler yaşamakta ve model, karmaşıklık arttıkça bozulmalara neden olabilmektedir . Progressive Growing GAN (PGGAN), bu sorunu çözmek için üretici ve ayırt edici ağların aşamalı olarak büyütüldüğü özel bir eğitim yöntemi kullanmaktadır.

Süper Çözünürlüklü Üretici Çekişmeli Ağ (Super-Resolution GAN - SRGAN)

Super-Resolution Generative Adversarial Networks (SRGAN) temsil ettiği önemli yenilikler ve geniş uygulama alanlarıyla, düşük çözünürlüklü (LR) görüntüleri yüksek çözünürlüklü (HR) görüntülere dönüştürmeyi amaçlayan bir modeldir. Geleneksel süper çözünürlük teknikleri, interpolasyon veya klasik makine öğrenimi yöntemlerine dayanırken, SRGAN, Generatif Çekişmeli Ağlar (GAN) mimarisini kullanarak daha ayrıntılı ve gerçekçi görüntüler üretmeyi hedefler.

SRGAN, temel olarak iki bileşenden oluşur: bir jeneratör (generator) ve bir

ayırt edici (discriminator). Jeneratör, düşük çözünürlüklü bir görüntüyü alarak yüksek çözünürlüklü versiyonunu üretmeye çalışırken, ayırt edici ağ ise üretilen görüntülerin gerçek olup olmadığını belirlemeye çalışır. Bu rekabetçi süreç, jeneratörün giderek daha gerçekçi görüntüler üretmesini sağlar. Standart piksel bazlı kayıpların yerine, SRGAN'ın temel yeniliklerinden biri olan perceptual loss fonksiyonunun kullanımı, görüntülerin insan gözüne daha doğal görünmesini sağlamaktadır [31].

SRGAN'ın en büyük avantajlarından biri, detayları ve dokusal özellikleri koruyarak interpolasyon bazlı yöntemlere göre çok daha gerçekçi süper çözünürlük sağlama yeteneğidir. Geleneksel yöntemler genellikle bulanıklık veya kayıp detaylar gibi problemlerle karşılaşırken, SRGAN, derin evrişimli sinir ağları (CNN) ve adversarial kayıpları birleştirerek keskin ve doğal görünümlü sonuçlar üretir. Özellikle tıbbi görüntüleme, uydu görüntü işleme, biyomedikal analiz ve uzaktan algılama gibi alanlarda yaygın olarak kullanılmaktadır [45].

Tıbbi görüntüleme bağlamında, SRGAN, düşük çözünürlüklü tıbbi taramaların (örneğin MRI ve CT taramaları) çözünürlüğünü artırarak daha doğru tanılar elde edilmesine katkı sağlar. Uydu görüntü işlemelerinde ise, düşük çözünürlüklü uydu görüntülerinin detaylarının iyileştirilmesi ve analiz edilmesi için kritik bir araç olarak değerlendirilmektedir. Bunun yanı sıra, SRGAN'ın biyometrik güvenlik sistemleri, mikroskop görüntü analizleri ve otomatik kalite kontrol sistemleri gibi endüstriyel uygulamalarda da önemli bir rol oynadığı görülmektedir [36].

Sonuç olarak, SRGAN, süper çözünürlük problemlerinde GAN tabanlı yaklaşımlar arasındaki en önemli modellerden biri olup, yüksek frekanslı detayları koruma, dokusal doğruluğu artırma ve görsel kalitede önemli iyileştirmeler sağlama yeteneği ile öne çıkmaktadır. Yapılan araştırmalar, SRGAN'ın geleneksel yöntemlere kıyasla hem nesnel hem de öznel değerlendirmelerde üstün performans gösterdiğini ortaya koymaktadır. Gelecekte, SRGAN ve türevlerinin daha derin sinir ağı yapıları, hibrit kayıp fonksiyonları ve yeni uygulama alanları ile daha da geliştirilmesi beklenmektedir [41].

Kademeli Büyüyen Üretici Çekişmeli Ağ (Progressive Growing GAN - ProGAN / PGGAN)

PGGAN, eğitim sürecine düşük çözünürlükte (örneğin, 4×4 piksel) başlayarak, modelin genel yapıyı öğrenmesini sağlar. Daha sonra, eğitim ilerledikçe yeni katmanlar eklenerek görüntü çözünürlüğü kademeli olarak artırılır. Bu aşamalı büyüme süreci, yüksek çözünürlüklü görüntüler üretirken modelin istikrarlı

bir şekilde öğrenmesini sağlar ve aşırı güncellenmelerden kaynaklanan eğitim dengesizliklerini azaltır.

PGGAN'in bir diğer yeniliği, yeni eklenen katmanların model tarafından kademeli olarak öğrenilmesini sağlayan fade-in (kademeli geçiş) yöntemidir. Yeni katmanlar doğrudan eklenmek yerine, eğitim sırasında yavaş yavaş etkin hale gelir. Bu süreç, modelin eğitim sürecine ani değişiklikler getirmesini önleyerek daha stabil bir öğrenme süreci sağlar.

Bu teknikler sayesinde, PGGAN geleneksel GAN'lere kıyasla daha yüksek çözünürlüklü ve daha gerçekçi görüntüler üretebilmektedir. Özellikle, yüz üretimi, medikal görüntüleme ve sanatsal içerik üretimi gibi alanlarda yaygın olarak kullanılmaktadır. Aşamalı büyüme stratejisi ve fade-in yöntemi sayesinde, yüksek çözünürlük gerektiren uygulamalarda istikrarlı bir şekilde eğitim yapılabilmesine olanak tanımaktadır [31, 33, 46].

Büyük Ölçekli Üretici Çekişmeli Ağ (BigGAN)

BigGAN, büyük ölçekli veri kümeleri üzerinde yüksek çözünürlüklü görüntülerin üretilmesi için geliştirilen gelişmiş bir Generatif Çekişmeli Ağ (GAN) modelidir. Geleneksel GAN modelleri, yüksek çözünürlüklü ve çeşitli görüntüler üretmekte zorluk çekerken, BigGAN genişletilmiş mimari ve yenilikçi eğitim teknikleriyle bu kısıtlamaları aşmayı hedeflemiştir. ImageNet gibi büyük ölçekli veri kümeleri üzerinde eğitilen model, sınıf koşullu GAN (class-conditional GAN, cGAN) yapısını kullanarak üretilen görüntülerin daha anlamlı ve gerçekçi olmasını sağlamaktadır [35].

BigGAN'ın en dikkat çekici özelliklerinden biri, kendiliğinden dikkat mekanizması (self-attention mechanism) ve spektral normalizasyon (spectral normalization) kullanarak geleneksel GAN'lara kıyasla daha stabil bir eğitim süreci sunmasıdır. Modelin önemli bir yeniliği olan truncation trick (kesme hilesi) yöntemi, üretilen görüntülerin çeşitliliğini kontrol ederek daha kaliteli ve daha gerçekçi görseller elde edilmesini sağlar. Ancak, bu teknik bazen görüntülerin fazla benzer olmasına ve çeşitliliğin azalmasına neden olabilir [37].

BigGAN ayrıca, eğitim sürecinde büyük batch boyutları kullanarak gradyan güncellemelerinin daha stabil olmasını sağlamış ve geleneksel GAN eğitiminde karşılaşılan mode collapse (mod çökmesi) problemini önemli ölçüde azaltmıştır. Modelin büyük ölçekli eğitim süreci, daha geniş kanal sayılarıyla çalışmasına ve böylece yüksek çözünürlüklü, detaylı ve doğal görüntüler üretmesine olanak tanımaktadır. Inception Score (IS) ve Fréchet Inception Distance (FID) gibi

değerlendirme metrikleri açısından geleneksel GAN modellerine kıyasla önemli ölçüde üstün performans göstermektedir [47].

BigGAN'ın yüksek çözünürlüklü görüntü üretme yeteneği sayesinde birçok farklı alanda uygulamaları bulunmaktadır. Sanat ve yaratıcılık alanında, stil transferi, sanatsal görüntü sentezi ve görsel üretim süreçlerinde kullanılmaktadır. Tıbbi görüntüleme alanında, MRI ve CT taramalarının çözünürlüğünü artırarak daha doğru teşhislerin yapılmasına olanak sağlamaktadır. Uydu görüntüleme ve jeo-uzamsal analiz alanında, yüksek kaliteli sentetik görüntüler oluşturarak veri analiz süreçlerini iyileştirmektedir. Otonom sürüş sistemleri için, modelin ürettiği yüksek kaliteli sentetik veriler daha kapsamlı eğitim veri kümeleri oluşturulmasına yardımcı olmaktadır [37].

Ancak, BigGAN'ın bazı önemli sınırlamaları bulunmaktadır. Yüksek hesaplama maliyetleri, modelin büyük ölçekli veri kümeleriyle eğitilmesini zorlaştırmakta ve bu süreci yüksek performanslı GPU veya TPU gibi hızlandırıcı donanımlar gerektiren bir işlem haline getirmektedir. Genelleme yeteneğinin sınırlı olması, BigGAN'ın belirli veri kümeleri üzerinde eğitildiğinde farklı veri setlerine adapte olmasını zorlaştırabilir. Kesme hilesi (truncation trick) sebebiyle modelin ürettiği görüntülerin çeşitliliği bazen düşebilir ve belirli desenlere aşırı uyum sağlayabilir [37].

Sonuç olarak, BigGAN, yüksek kaliteli ve çeşitli görüntü üretimi konusunda önemli bir dönüm noktası olmuş, ancak büyük ölçekli eğitim gereksinimleri nedeniyle uygulamada bazı sınırlamalarla karşılaşmıştır. Gelecekte, hafifletilmiş versiyonlar, daha verimli eğitim teknikleri ve transfer öğrenme stratejileri ile BigGAN'ın daha geniş kullanım alanlarına yayılması beklenmektedir [35].

Bu varyasyonlar, GAN'ların eğitim kararlılığını artırmış, küçük veri kümelerinde bile yüksek kaliteli görseller üretilmesini sağlamış ve geniş bir uygulama yelpazesi sunmuştur. GAN mimarilerindeki bu yenilikler, araştırmacılara esnek ve güçlü modellerle daha geniş bir uygulama alanı yaratma fırsatı vermiştir.

2.4 StyleGAN ve Varyasyonları

GAN'ların temel yapısı, farklı uygulama alanları ve eğitim zorluklarını çözmek için çeşitli varyasyonlarla geliştirilmiştir. GAN mimarisinin çeşitli varyasyonları, farklı kullanım alanlarına yönelik olarak geliştirilmiştir.

Üretici Çekişmeli Ağ (GAN) modelleri yıllar içinde önemli gelişmeler kaydederek

farklı problemlere yönelik çeşitli mimariler ortaya çıkarmıştır. Bu karşılaştırmada, dört GAN modeli: StyleGAN, StyleGAN2, StyleGAN-ADA ve StyleGAN3 görsel kaliteyi artırmak, eğitim sürecini daha kararlı hale getirmek ve veri etkinliğini yükseltmek için sundukları yenilikler açısından ele alınmaktadır.

StyleGAN, yılında Tero Karras ve ekibi tarafından geliştirilen ve GAN'ların görüntü üretme süreçlerinde devrim yaratan bir modeldir [48] . Geleneksel GAN yapılarından farklı olarak, giriş latent vektörünü doğrudan kullanmak yerine, bu vektörü bir ara gizli uzaya dönüştüren bir yapı sunmuştur. Bu ara uzay, özelliklerin daha ayrışabilir olmasını sağlayarak daha iyi kontrol edilebilen görseller üretmeye olanak tanımıştır. Modelin en büyük yeniliklerinden biri, Adaptive Instance Normalization (AdaIN) yöntemiyle her bir evrişim katmanına stil kodları eklenerek, görsellerin belirli özelliklerinin doğrudan manipüle edilebilmesini sağlamasıdır. Ayrıca, her bir katmana rastgele gürültü eklenerek ince detayların (örneğin, saç telleri ve çiller) daha gerçekçi hale gelmesi mümkün olmuştur. Ancak, StyleGAN modelinde belirli yapaylıklar ortaya çıkmış ve interpolasyon sırasında "su damlası" olarak adlandırılan görsel artefaktlar tespit edilmiştir [48]. Bu eksiklikler, bir sonraki versiyon olan StyleGAN2 ile giderilmeye çalışılmıştır.

StyleGAN2, yılında StyleGAN modelinin eksikliklerini gidermek için geliştirilmiş ve özellikle modelin ürettiği görüntülerdeki artefaktları azaltmaya odaklanmıştır. AdaIN'in yerine, yeni bir ağırlık modülasyon ve demodülasyon tekniği eklenerek modelin içindeki normalizasyonun neden olduğu görsel bozulmalar düzeltilmiştir. Böylece, su damlası benzeri hatalar ortadan kaldırılmış ve modelin ürettiği görseller daha keskin ve doğal hale getirilmiştir. Ayrıca, modelin gizli uzayında yapılan düzenlemelerle interpolasyon kalitesi artırılmış, böylece daha doğal geçişler sağlanmıştır. Progressive Growing yöntemi de kaldırılmış ve model, eğitim süreci boyunca çözünürlük seviyelerini değiştirmeden doğrudan tam çözünürlükte eğitilmiştir. Tüm bu yenilikler, StyleGAN2'yi önceki versiyona kıyasla daha tutarlı ve yüksek kaliteli görseller üreten bir model haline getirmiştir [49].

StyleGAN-ADA, yılında GAN eğitiminde veri eksikliğinin yarattığı aşırı öğrenme problemini çözmek amacıyla geliştirilmiştir. Klasik GAN eğitiminde, sınırlı miktarda veri kullanıldığında diskriminatör modeli, eğitim verilerini ezberleme eğiliminde olur ve bu da modelin gerçekçi yeni görseller üretmesini engeller. StyleGAN-ADA, veri artırma tekniklerini adaptif bir şekilde ayarlayarak diskriminatörün aşırı öğrenmesini engellemiştir. Modelin getirdiği en büyük yenilik, veri artırma işleminin yalnızca gerektiği kadar uygulanmasını sağlayan bir adaptasyon mekanizmasıdır. Bu sayede, StyleGAN-ADA daha az veriyle yüksek kaliteli görüntüler üretebilmiş ve veri kısıtlaması bulunan birçok alanda

kullanılabilir hale gelmiştir [50].

StyleGAN3, yılında geliştirilmiş ve GAN'ların görüntü üretme sürecindeki temel eksikliklerden biri olan dönüşüme karşı duyarsızlığı gidermeyi amaçlamıştır. StyleGAN2'de üretilen görsellerde, detayların piksel koordinatlarına yapıştığı görülmüştür. Bu sorun, modelin bazı ayrıntıları, nesnelerin yüzeyleri yerine sabit piksel konumlarına bağlamasına neden olmuş ve özellikle animasyon gibi alanlarda görsel olarak dikkat dağıtıcı sonuçlar ortaya çıkarmıştır. StyleGAN3, bu problemi çözmek için aliasing'i ortadan kaldıran yeni bir sinyal işleme yöntemi kullanmıştır. Bu sayede, modelin ürettiği tüm detaylar, doğal bir şekilde yüzeylere yapışmakta ve interpolasyon sırasında daha akıcı geçişler sağlamaktadır. Ayrıca, modelin tüm sinyallerini sürekli bir fonksiyon olarak ele alarak dönüşümlere karşı tam anlamıyla eşdeğerlik sağlamıştır. Bu sayede, StyleGAN3 özellikle video ve animasyon gibi alanlar için daha uygun bir GAN modeli haline gelmiştir [51].

Bu dört model arasındaki karşılaştırmayı özetlemek gerekirse, StyleGAN, stil tabanlı üretici mimarisi ile GAN'larda büyük bir yenilik yapmış, ancak interpolasyon sırasında yapaylıklar barındırmıştır. StyleGAN2, bu yapaylıkları gidererek modelin ürettiği görsellerin kalitesini artırmış ve daha kararlı bir eğitim süreci sunmuştur. StyleGAN-ADA, az veriyle eğitimi mümkün kılarak GAN'ların daha geniş uygulama alanlarına yayılmasını sağlamıştır. StyleGAN3 ise aliasing problemini çözerek görsellerin daha doğal ve dönüşüme duyarlı olmasını sağlayarak özellikle animasyon ve video üretimi için daha güçlü bir model haline gelmiştir.

Tablo 2.1 GAN varyasyonlarının özellikleri ve iyileştirmeler

Model	**Öne Çıkan Özellikler**	**Önceki Modellerden İyileştirmeler**
StyleGAN	Stil tabanlı bir üretici mimarisi sunar. Gürültü girişleriyle stokastik varyasyon ekleme imkanı sunar.	Adaptive Instance Normalization (AdaIN) ile daha iyi stil kontrolü sağlar. Öğrenilmiş sabit tensörler kullanılmıştır.
StyleGAN2	AdaIN yerine ağırlık demodülasyonu yöntemi eklenmiştir. Yol uzunluğu düzenlemesi uygulanmıştır.	Daha iyi Frechet Başlangıç Mesafesi (FID) skorları ve görsel kalite elde edilmiştir. Aşamalı büyüme kaldırılarak eğitim kararlılığı artırılmıştır.
StyleGAN2-ADA	Eğitim sırasında artırma gücünü dinamik olarak ayarlayan Uyarlamalı Ayırt Edici Artırma (ADA) yöntemi eklenmiştir.	Küçük veri kümeleriyle kaliteli görüntüler üretmek mümkün hale gelmiştir. Bellek kullanımı optimize edilmiştir.
StyleGAN3	Sürekli sinyallerle çalışacak şekilde tasarlanmıştır. Fourier tabanlı giriş temsili eklenmiştir.	Doku kayması sorunları giderilmiştir. Video ve animasyonlar için uzamsal eşdeğerlik geliştirilmiştir.

(Tablo 2.4)'de yaygın olarak kullanılan GAN türleri, temel özellikleri ve iyileştirmeleri tablo şeklinde de özetlenmiştir. Bu varyasyonlar, mimarilerindeki yenilikler ve teknik iyileştirmelerle GAN'ların performansını ve uygulama alanlarını genişletmiştir.

2.4.1 StyleGAN3: Alias-Free Yaklaşımı ve Mimari Detayları

StyleGAN3, NVIDIA tarafından geliştirilen ve StyleGAN2'nin eksikliklerini gidermek amacıyla tasarlanmış bir jeneratif adversaryal ağdır. Bu mimari, özellikle "aliasing" sorununu çözmeye ve görüntü sentezi sırasında detayların doğal bir hiyerarşik yapıda işlenmesini sağlamaya odaklanmıştır. Bu iyileştirmeler, görüntülerin daha tutarlı bir şekilde ölçeklenebilmesini ve dönüşümler sırasında geometrik uyumu korumasını mümkün kılar. Şekil 2.2 'de, StyleGAN3 mimarisinin genel yapısı ve bileşenleri detaylandırılmıştır.

Alias-Free Yaklaşım ve Mimari Yenilikler

StyleGAN3'ün temel amacı, *aliasing* sorununu çözerek daha doğal ve tutarlı görüntü üretimini sağlamaktır. Geleneksel GAN mimarilerinde, sinyal işleme tekniklerinin yetersizliği nedeniyle, görüntü detaylarının piksel koordinatlarına yapıştığı gözlemlenmiştir. Bu durum, detayların nesnelerin yüzeylerine doğru bir şekilde bağlanmasını engelleyerek, gerçekçilikten uzak sonuçlar ortaya çıkarmaktadır. StyleGAN3, bu problemi aşağıdaki yöntemlerle çözmeyi başarmıştır:

- **Sürekli Sinyal İşleme:** StyleGAN3, sinyallerin sürekli bir alanda işlendiği bir model kullanır. Nyquist-Shannon örnekleme teoremiyle uyumlu olarak, her bir sinyal bant sınırlı bir şekilde işlenir. Bu, sinyallerin detay kaybı yaşamadan katmanlar arasında aktarılmasını sağlar.
- **Kaiser Filtreleri:** Yüksek kaliteli Kaiser filtreleri, *aliasing* etkisini minimuma indirmek ve görüntü detaylarının her çözünürlük seviyesinde korunmasını sağlamak için kullanılmıştır. Bu filtreler, özellikle düşük çözünürlük seviyelerinde daha agresif bant sınırlama yaparak, sinyalin gereksiz frekans içeriklerini ortadan kaldırır.
- **Fourier Tabanlı Giriş Özellikleri:** StyleGAN2'nin sabit başlangıç özellikleri yerine, Fourier tabanlı giriş özellikleri kullanılmıştır. Bu, giriş sinyallerinin daha doğal bir hiyerarşik yapıya sahip olmasını ve dönüşümlerde tutarlılığı artırmasını sağlar [51].

Çeviri ve Rotasyon Eşdeğerliği

StyleGAN3, hem çeviri (*translation*) hem de rotasyon (*rotation*) eşdeğerliği sağlayan bir yapı ile tasarlanmıştır. Bu özellik, her bir çözünürlük seviyesindeki detayların, temel katmanlardaki büyük ölçekli özelliklere tutarlı bir şekilde bağlanmasını mümkün kılar. Çeviri eşdeğerliği (EQ-T) ve rotasyon eşdeğerliği (EQ-R) metrikleri, bu yapının performansını doğrulamış ve StyleGAN3'ün önceki jenerasyonlara göre üstünlüğünü ortaya koymuştur [51].

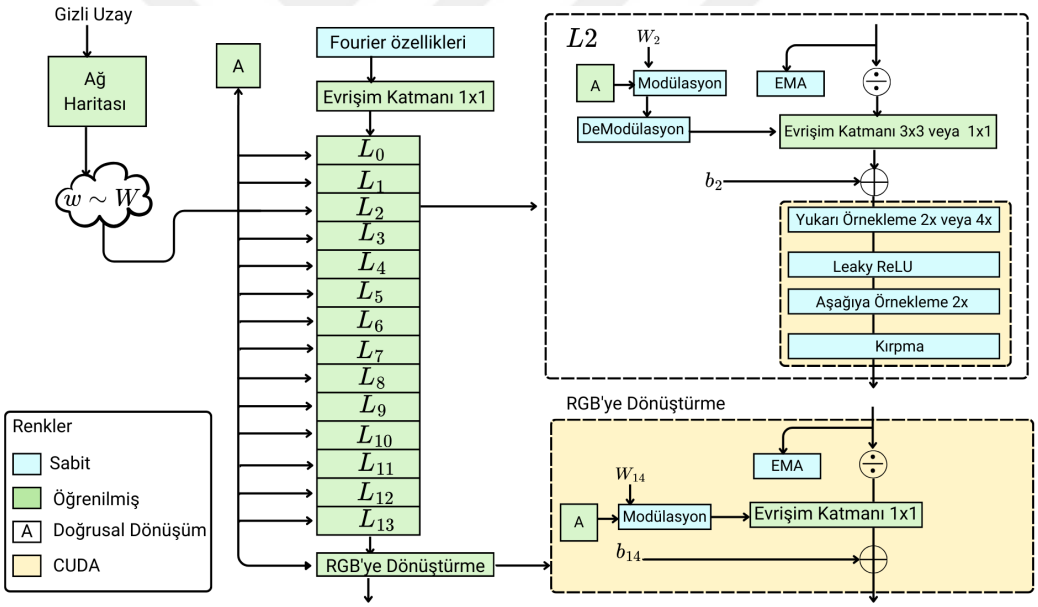
Eşdeğerliğin sağlanmasında kullanılan yöntemlerden biri, tüm aktivasyon fonksiyonlarının (ör. ReLU) yüksek frekanslı gürültü oluşturmamasını sağlamaktır. Bu amaçla, her bir aktivasyon işleminden sonra düşük geçişli bir filtreleme uygulanmıştır. Ayrıca, StyleGAN3'te kullanılan modüle edilmiş konvolüsyonlar, detayların yüzeylere doğru şekilde bağlanmasını garanti eder.

Performans Değerlendirmesi ve Uygulama Alanları

StyleGAN3'ün performansı, FFHQ, MetFaces ve AFHQ gibi çeşitli veri setlerinde değerlendirilmiştir. Bu değerlendirmeler, StyleGAN3'ün StyleGAN2'ye kıyasla hem FID hem de eşdeğerlik metriklerinde üstün sonuçlar elde ettiğini göstermiştir. Örneğin:

- FFHQ veri kümesinde FID değeri %10 iyileştirilmiş,
- Görüntü detaylarının daha doğal bir şekilde hareket ettiği ve nesne yüzeylerine bağlandığı gözlemlenmiştir.

StyleGAN3'ün sunduğu bu yenilikler, video ve animasyon üretimi gibi dinamik görüntüleme gereksinimlerine sahip alanlarda kullanım için idealdir. Bunun yanı sıra, tıbbi görüntüleme ve yapay zeka destekli tasarım gibi farklı disiplinlerde de geniş bir uygulama potansiyeline sahiptir [51].



Şekil 2.2 StyleGAN3 mimarisi

2.5 GAN'ların Sınırlamaları ve Alınabilecek Önlemler

Çekişmeli Üretici Ağlar (GANs), güçlü generatif modelleme yeteneklerine sahip olmasına rağmen, pratik uygulamalarda çeşitli sınırlamalarla karşılaşmaktadır. Bu sınırlamalar, modelin eğitim sürecinin istikrarını ve üretilen verinin çeşitliliğini etkileyerek performans kayıplarına neden olabilir. Özellikle mod çökmesi (mode collapse), yakınsamama (non-convergence) ve eğitim istikrarsızlığı (training instability) gibi sorunlar GAN tabanlı modellerin güvenilirliğini

sınırlandırmaktadır. Bu bölümde, bu sınırlamaların nedenleri incelenmekte ve literatürde önerilen çözüm yöntemleri açıklanmaktadır.

2.5.1 Mod Çökmesi

GAN'ların en yaygın sorunlarından biri olan mode collapse, üretici ağın (generator), gerçek veri dağılımının yalnızca sınırlı bir bölümünü öğrenmesi ve sürekli olarak benzer örnekler üretmesi durumudur. Mode collapse, özellikle karmaşık veri dağılımlarında modelin genelleme yeteneğini önemli ölçüde sınırlandırır ve üretilen veri çeşitliliğini azaltarak modelin gerçekçi veri üretmesini engeller [31]. Bu sorunun önüne geçmek için geliştirilen çözümlerden biri, çeşitliliği artırmaya yönelik kayıp fonksiyonlarının (diversity-promoting loss functions) kullanılmasıdır. Unrolled GAN ve Feature Matching Loss gibi teknikler, üretici ağı daha geniş bir veri dağılımını öğrenmeye teşvik eden kayıp fonksiyonları sunarak mode collapse'i azaltmayı hedefler [52]. Unrolled GAN, ayrıştırıcı ağın (discriminator) güncellenmesini geciktirerek üreticinin daha geniş bir veri uzayını keşfetmesini sağlarken, Feature Matching Loss, ayrıştırıcının yalnızca tekil örnekleri ayırt etmek yerine, özellik uzayındaki istatistikleri öğrenmesini sağlayarak çeşitliliği artırır [33]. Bir diğer etkili yöntem, minibatch ayrımcılık (minibatch discrimination) tekniğidir. Bu yöntem, ayrıştırıcının her minibatch içerisindeki örneklerin farklı olup olmadığını değerlendirmesine olanak tanır. Geleneksel ayrıştırıcılar yalnızca tek tek örnekleri değerlendirirken, minibatch ayrımcılığı, tüm minibatch içindeki çeşitliliği göz önünde bulundurarak üretici ağın aynı türden örnekleri sürekli üretmesini önler [52]. Ayrıca, Wasserstein GAN (WGAN) gibi yeni yaklaşımlar, modelin daha kararlı bir öğrenme süreci geçirmesine yardımcı olarak mode collapse'i önlemektedir. Geleneksel GAN'larda, ayrıştırıcı ağın çıktı değerleri, modelin öğrenme sürecinde dengesiz değişimlere neden olabilirken, WGAN'da Kantorovich-Rubinstein metriği kullanılarak üreticinin daha istikrarlı bir optimizasyon süreci geçirmesi sağlanır. Bu sayede üreticinin çeşitlilik içeren veri örnekleri üretme kapasitesi artırılır [31].

2.5.2 Yakınsamama Sorunları

GAN'ların eğitim süreci, üretici ve ayrıştırıcı arasındaki dinamik dengenin hassas yapısı nedeniyle çoğu zaman istikrarlı bir yakınsama noktası oluşturmakta zorlanır. Mode collapse'e ek olarak, GAN'lar genellikle optimizasyon sürecinde dengesiz gradyan güncellemeleri yaşar ve bu durum modelin yakınsamamasına veya aşırı

dalgalanmalara neden olur [33]. Üretici ağ, ayrıştırıcının kararlarını yanlış yönlendirmeye çalışırken, ayrıştırıcı ağın hızla öğrenmesi üreticinin hiçbir zaman istikrarlı bir öğrenme sürecine girmemesine yol açabilir.

Yakınsamama sorunlarını çözmek için kullanılan tekniklerden biri spektral normalizasyon (spectral normalization) yöntemidir. Bu yöntem, ayrıştırıcı ağın ağırlıklarını normalize ederek gradyan büyüklüğünün kontrol altında tutulmasını sağlar. Böylece, ayrıştırıcının aşırı güçlü hale gelerek üreticiyi tamamen başarısız hale getirmesi önlenir [31]. Özellikle GAN'ların gradyan büyüme problemini engellemek için spektral normalizasyon yaygın olarak önerilmektedir.

Bunun yanı sıra, gradyan kırpma (gradient clipping) yöntemi de modelin yakınsama sürecini istikrarlı hale getirmek için kullanılmaktadır. Gradyan büyüklüğünün belirli bir sınır içinde tutulmasını sağlayan bu yöntem, aşırı büyük değerlerin oluşmasını önleyerek modelin dengesiz güncellenmesini engeller [52].

Yakınsama sürecini iyileştiren bir diğer teknik, dinamik öğrenme oranı ayarlamalarıdır (dynamic learning rate scheduling). Sabit öğrenme oranları yerine, Adam Optimizer veya Cyclical Learning Rates gibi tekniklerin kullanılması, modelin daha esnek bir şekilde öğrenmesini sağlar. Öğrenme oranlarının eğitim sürecine bağlı olarak ayarlanması, modelin stabilitesini artırarak yakınsama sürecini optimize eder [33].

Ek olarak, Two-Time-Scale Güncelleme Kuralı (TTUR - Two-Time-Scale Update Rule), üretici ve ayrıştırıcı için farklı öğrenme oranları kullanarak iki ağın birbirine aşırı bağımlı hale gelmesini engeller ve böylece istikrarlı bir öğrenme süreci sağlar [52].

2.5.3 Eğitim İstikrarsızlığı

GAN'ların eğitimi, üretici ve ayrıştırıcı arasındaki karmaşık etkileşim nedeniyle oldukça zor bir süreçtir. Eğitim sırasında hiperparametre seçimleri, veri dağılımı ve model tasarımı gibi faktörler modelin istikrarını doğrudan etkileyebilir. Özellikle eğitim sürecinde, modelin kayıp fonksiyonunun dengesiz hareket etmesi modelin başarısız öğrenmesine neden olabilir [31].

Eğitim sürecindeki istikrarsızlıkları azaltmak için en etkili yöntemlerden biri LeakyReLU aktivasyon fonksiyonunun kullanılmasıdır. Standart ReLU yerine LeakyReLU kullanımı, gradyan akışını iyileştirerek eğitimin daha stabil hale gelmesini sağlar [52].

Bir diğer yaygın çözüm, veri çoğaltma (data augmentation) teknikleridir. Özellikle Mixup ve CutMix gibi teknikler, eğitim verilerinin çeşitliliğini artırarak modelin daha geniş bir veri dağılımını öğrenmesine yardımcı olur [33].

GAN'ların eğitim sürecinde stabiliteyi artırmanın bir diğer yolu, hiperparametre optimizasyonu (hyperparameter optimization) yapmaktır. Eğitim sürecinde doğru hiperparametrelerin belirlenmesi, modelin performansını doğrudan etkileyen kritik bir faktördür. Grid Search veya Bayesian Optimization gibi teknikler, en uygun hiperparametre kombinasyonlarını belirleyerek eğitim sürecini iyileştirebilir [31].

Ayrıca, Least Squares GAN (LSGAN) gibi özel kayıp fonksiyonlarının kullanılması, eğitim sürecinde daha kararlı gradyanlar sağlayarak modelin performansını artırabilir [52].

Son olarak, AdamW veya RMSProp gibi gelişmiş optimizasyon algoritmaları, gradyanların daha verimli kullanılmasını sağlayarak eğitim sürecinin stabilitesini artırır [33].

GAN'ların sınırlamalarını anlamak ve bu sorunlara yönelik çözüm tekniklerini uygulamak, modellerin eğitim sürecini ve üretilen çıktının kalitesini önemli ölçüde iyileştirebilir. Özellikle mode collapse ve eğitim istikrarsızlığı gibi sorunlar, uygun tekniklerin uygulanmasıyla kontrol altına alınarak GAN'ların çeşitli uygulama alanlarında daha etkili bir şekilde kullanılmasına olanak tanır.

GAN'ların sınırlamalarını anlamak ve bu sınırlamalara karşı uygun önlemleri almak, modellerin eğitim sürecini ve çıktılarının kalitesini büyük ölçüde iyileştirebilir. Özellikle mod çökmesi,yakınsamama sorunları ve eğitim istikrarsızlığı gibi sorunlar, uygun tekniklerin uygulanmasıyla kontrol altına alınabilir. Bu iyileştirmeler, GAN'ların çeşitli uygulama alanlarında daha etkili bir şekilde kullanılmasına olanak tanır.

2.6 GAN Değerlendirme Metrikleri

GAN'ların performansını değerlendirmek için hem **nitel** hem de **nicel** ölçütler kullanılmaktadır. Bu metrikler, üretilen verilerin gerçek veri dağılımına ne kadar benzediğini ve modelin çeşitlilik ile gerçeklik başarısını anlamak için kritik öneme sahiptir [31].

2.6.1 Nitel Değerlendirme

- **Mod Çökmesi (Mode Collapse) Tespiti:**

Üretilen örneklerin çeşitliliği analiz edilerek modelin gerçek veri dağılımını ne derece geniş çapta öğrenebildiği incelenir. Eğer model, farklı ve çeşitli örnekler üretmek yerine benzer veya tekrarlayan örnekler üretmeye eğilim gösteriyorsa, bu durum mode collapse olarak adlandırılır. Mode collapse, modelin yalnızca dar bir veri alt kümesine odaklandığını ve genel veri dağılımını yeterince iyi öğrenemediğini gösterir.

- **İnsan Yargısı:**

  GAN'lar tarafından üretilen görsellerin gerçekçiliği, insan gözlemi ile değerlendirilir. Her ne kadar subjektif bir yöntem olsa da, insan yargısı, modelin görsel kalite açısından başarısını anlamak için önemli bir ölçüt olarak kullanılmaktadır. İnsan değerlendiriciler genellikle kör testler ile üretilen ve gerçek görüntüler arasındaki farkları belirlemeye çalışır.

- **En Yakın Komşular (Nearest Neighbors):**

  Bu yöntem, üretilen örneklerin gerçek veri örnekleriyle olan görsel ve istatistiksel benzerliklerini analiz etmektedir. Özellikle, her bir üretilen görüntü için en yakın gerçek görüntü bulunarak modelin gerçek veri dağılımına ne kadar yakın olduğu incelenir. Eğer model, ürettiği örnekleri gerçek verilere çok yakın bir şekilde oluşturuyorsa, bu durum modelin başarılı bir şekilde öğrenme gerçekleştirdiğini gösterebilir.

2.6.2 Nicel Değerlendirme

- **Fréchet Başlangıç Mesafesi (Fréchet Inception Distance - FID):** Frechet Başlangıç Mesafesi (FID), generatif adversaryal ağların (GAN) ürettiği görüntülerin kalitesini değerlendirmek için kullanılan ve literatürde önemli bir yere sahip olan bir metrik olarak öne çıkmaktadır. Bu metrik, gerçek ve üretilmiş görüntüler arasındaki istatistiksel farklılıkları ölçer ve bu ölçümde genellikle önceden eğitilmiş bir derin öğrenme modeli olan Inception v3'ten elde edilen özellik vektörleri temel alınır. FID'nin temel amacı, üretilen görüntülerin gerçek veri kümesindeki görüntülerle uyumluluğunu nicel bir şekilde ifade etmektir. FID değeri düşük olduğunda, üretilen görüntülerin gerçek görüntülerle daha uyumlu olduğu ve dolayısıyla daha yüksek bir görsel kaliteye sahip olduğu sonucuna varılır.

  FID'nin matematiksel tanımı, iki veri kümesi arasındaki Frechet mesafesi (diğer adıyla Wasserstein-2 mesafesi) ile ifade edilir. Gerçek ve üretilmiş görüntülerin özellik vektörlerinden hesaplanan istatistiksel parametreler bu formülde kullanılır. FID'nin formülü şu şekilde tanımlanmaktadır:

$$FID = \|\mu_r - \mu_g\|^2 + \mathrm{Tr}\big(\Sigma_r + \Sigma_g - 2\sqrt{\Sigma_r \Sigma_g}\big) \tag{2.4}$$

Bu formülde 2.4 μ_r, gerçek görüntülerin özellik vektörlerinden elde edilen ortalama değerini, μ_g ise üretilmiş görüntülerin özellik vektörlerinden elde edilen ortalama değerini temsil eder. Bu ortalama değerler, iki veri kümesinin dağılımlarının merkezlerini ifade etmektedir. Σ_r ve Σ_g, sırasıyla gerçek ve üretilmiş görüntülerin özellik vektörlerinden hesaplanan kovaryans matrislerini temsil eder ve bu matrisler, veri kümelerindeki özelliklerin dağılımlarını ve aralarındaki ilişkileri yansıtır. Formüldeki $||\mu_r - \mu_g||^2$ terimi, gerçek ve üretilmiş görüntülerin ortalamaları arasındaki Öklidyen mesafeyi ifade ederek bu iki dağılımın merkezleri arasındaki farkı ölçer. $\sqrt{\Sigma_r \Sigma_g}$ terimi ise gerçek ve üretilmiş görüntülerin kovaryans matrislerinin geometrik ortalamasını ifade eder ve bu terim, iki dağılım arasındaki örtüşme derecesini yansıtır. Matrislerin köşegen elemanlarının toplamı olan Tr ifadesi ise bu hesaplamadaki diğer bir önemli unsurdur [53] .

FID'nin hesaplanması üç temel aşamada gerçekleştirilir. İlk aşamada, gerçek ve üretilmiş görüntüler, önceden eğitilmiş Inception v3 modeline giriş olarak verilir ve bu modelin son havuzlama katmanından 2,048 boyutlu aktivasyon vektörleri (özellik vektörleri) elde edilir. İkinci aşamada, bu vektörlerden her iki veri kümesi için ortalama ve kovaryans değerleri hesaplanır. Son aşamada ise yukarıda verilen formül kullanılarak FID değeri hesaplanır. Düşük bir FID skoru, üretilmiş görüntülerin gerçek görüntülerle yüksek bir uyum içinde olduğunu ve dolayısıyla yüksek kaliteli olduğunu gösterir. Buna karşın, yüksek bir FID skoru, üretilmiş görüntülerin gerçek görüntülerle olan benzerliğinin düşük olduğunu ve görsel kalitesinin zayıf olduğunu ifade eder.

FID, GAN modellerinin performansını değerlendirmek için standart bir metrik haline gelmiştir. Özellikle GAN'ların eğitim süreçlerinde model gelişiminin izlenmesi ve performans iyileştirmeleri yapılması amacıyla yaygın olarak kullanılmaktadır. FID, üretilmiş görüntülerin hem çeşitliliğini hem de gerçekçiliğini değerlendirme kapasitesine sahip olduğu için diğer metriklere kıyasla daha anlamlı sonuçlar sunmaktadır. Literatürde yapılan çalışmalar, düşük FID skorlarının, daha gerçekçi ve çeşitli görüntüler üreten modellerle ilişkili olduğunu göstermektedir [53] . Bu nedenle, FID, generatif modellerin başarı düzeyini değerlendirme noktasında etkili ve güvenilir bir araç olarak değerlendirilmektedir.

- **Başlangıç Skoru (Inception Score - IS):** Inception Score (IS), generatif adversaryal ağların (GAN) ürettiği görüntülerin değerlendirilmesinde sıklıkla

kullanılan önemli bir metrik olarak öne çıkmaktadır. Bu metrik, üretilen görüntülerin hem görsel kalitesini hem de çeşitliliğini ölçmek amacıyla geliştirilmiştir. Inception Score, bir modelin belirgin sınıflara ait net ve kaliteli görüntüler üretebilme kapasitesini ve bu sınıflar arasında çeşitlilik sağlayabilme yeteneğini aynı anda değerlendirebilmektedir. IS'nin hesaplanmasında genellikle ImageNet üzerinde eğitilmiş Inception v3 modeli kullanılmaktadır. Bu model, bir görüntünün belirli bir sınıfa ait olma olasılığını tahmin etmek üzere tasarlanmıştır ve bu tahminler, IS'nin temel bileşenlerini oluşturmaktadır.

Matematiksel olarak IS, üretilmiş görüntülerin sınıf tahminlerinden elde edilen olasılık dağılımlarının Kullback-Leibler (KL) uzaklığını kullanır. Bu metrik, bir görüntünün belirli bir sınıfa ait olma olasılığını ifade eden $p(y|x)$ ile üretilen tüm görüntülerdeki genel sınıf dağılımını temsil eden $p(y)$ arasında bir ilişki kurar. IS şu şekilde tanımlanır:

$$IS(G) = \exp\left(E_{x\sim p_g}\left[D_{KL}(p(y|x)\|p(y))\right]\right) \tag{2.5}$$

Bu formülde 2.5 $p(y|x)$, bir görüntünün belirli bir sınıfa ait olma olasılığını, $p(y)$ ise üretilen tüm görüntülerin sınıf dağılımını ifade eder. KL uzaklığı, bu iki olasılık dağılımı arasındaki farklılığı ölçer. $E_{x\sim p_g}$, üretilen görüntülerin tümü üzerinden bu uzaklıkların ortalamasını hesaplar. Sonuç olarak, bu ortalama üzerine üstel fonksiyon uygulanarak IS elde edilir.

IS, iki temel özelliği birleştirerek bir generatif modelin başarısını değerlendirir. İlk olarak, üretilen görüntülerin net bir şekilde bir sınıfa ait olması beklenir. Bu, görüntülerin belirgin nesneler içermesi ve sınıf tahminlerinin düşük entropiye sahip olması gerektiğini ifade eder. Diğer bir deyişle, Inception v3 modeli, bir görüntünün tek bir sınıfa ait olduğuna yüksek bir güvenle karar verebilmelidir. İkinci olarak, üretilen görüntülerin genel sınıf dağılımının yüksek entropiye sahip olması gerekmektedir. Bu özellik, modelin çeşitli sınıflardan görüntüler üretebilme kapasitesini ortaya koyar.

Inception Score'un hesaplanması birkaç aşamada gerçekleştirilir. İlk olarak, üretilmiş görüntüler Inception v3 modeline giriş olarak verilir ve bu süreçte her görüntü için sınıf tahminleri elde edilir. Bu tahminler $p(y|x)$ olarak kaydedilir. Daha sonra tüm üretilen görüntülerden genel sınıf dağılımı $p(y)$ hesaplanır. Bu iki dağılım arasındaki KL uzaklığı belirlenir ve tüm görüntüler üzerinden ortalaması alınır. Son olarak, bu ortalama üzerine üstel fonksiyon uygulanarak IS değeri elde edilir. Düşük bir IS, üretilen görüntülerin ya net

olmadığı ya da sınıf çeşitliliğinin düşük olduğu anlamına gelirken, yüksek bir IS, üretilen görüntülerin hem kaliteli hem de çeşitli olduğunu gösterir.

Inception Score, GAN'ların performansını değerlendirmek için yaygın olarak kullanılan bir metrik olmasına karşın, bazı sınırlamaları da bulunmaktadır. Öncelikle, yalnızca görsel kalite ve çeşitliliği değerlendirir; görüntülerin gerçeklik seviyesini doğrudan ölçmez. Ayrıca, IS'nin doğru sonuçlar verebilmesi için kullanılan modelin eğitildiği veri kümesiyle üretilen görüntülerin uyumlu olması gerekmektedir. Örneğin, Inception v3 modeli ImageNet veri kümesinde eğitildiğinden, IS'nin en iyi sonuçları ImageNet sınıflarına uygun görüntüler üreten modeller üzerinde verir [54] .

Sonuç olarak, Inception Score, generatif modellerin değerlendirilmesinde etkili ve pratik bir araçtır. Üretilen görüntülerin hem kalitesini hem de çeşitliliğini değerlendirme yeteneği, modellerin karşılaştırılmasında önemli bir avantaj sağlar. Ancak, bu metriğin diğer değerlendirme yöntemleriyle birlikte kullanılması, daha kapsamlı ve güvenilir sonuçlar elde edilmesi açısından önemlidir. Bu nedenle, IS'nin sınırlamaları dikkate alınarak, değerlendirme sürecinde tamamlayıcı metriklerle desteklenmesi önerilmektedir.

- **Precision, Recall ve F1 Skoru**

  GAN'ların üretim performansını anlamada önemli bir rol oynayan bu üç metrik, modelin hem gerçekçi hem de çeşitli örnekler üretme yeteneğini değerlendirmek için kullanılır:

  - **Precision (Kesinlik):** Üretilen örneklerin ne kadar gerçekçi olduğunu ölçer. Yüksek precision değerine sahip bir model, ürettiği görüntülerin büyük çoğunluğunu gerçek verilere benzer şekilde oluşturabilmektedir.
  - **Recall (Duyarlılık):** Modelin veri çeşitliliğini ne ölçüde yansıttığını gösterir. Yüksek recall değeri, modelin gerçek veri dağılımının geniş bir yelpazesini başarılı bir şekilde taklit edebildiğini gösterir.
  - **F1 Skoru:** Precision ve Recall'un harmonik ortalaması olup, modelin genel performansını dengeli bir şekilde özetler. Eğer bir model hem gerçekçi hem de çeşitli örnekler üretiyorsa, F1 skoru yüksek olur.

  Bu üç metrik, GAN modellerinin farklı yönlerden değerlendirilmesini sağladığı için özellikle karşılaştırmalı çalışmalar ve model geliştirme süreçlerinde sıkça kullanılmaktadır [31].

- **Kernel Inspection Distance (KID):** Kernel Inspection Distance (KID), GAN'ların ürettiği ve gerçek veri arasındaki benzerliği değerlendirmek için

kullanılan bir ölçümdür. KID, *Fréchet Inception Distance (FID)* ile benzer prensiplere dayanmakla birlikte, daha az hesaplama gereksinimi ile daha hassas sonuçlar sunabilmektedir.

- KID, kernel tabanlı ölçümlerle üretilen ve gerçek veriler arasındaki benzerliği değerlendirir.
- KID, FID gibi çalışmasına rağmen, hesaplamalarda daha az bağımlılık gerektirir ve özellikle küçük veri kümelerinde daha kararlı sonuçlar verebilir.

GAN'ların karşılaştırmalı performans analizi için kullanılan bu metrik, özellikle farklı model yapılarını değerlendirirken ve üretilen görüntülerin kalite düzeyini belirlerken kritik bir rol oynar [31].

GAN değerlendirme metrikleri, modelin başarısını anlamak ve performansını optimize etmek için kritik öneme sahiptir. Hem nitel hem de nicel değerlendirmeler, modelin güçlü ve zayıf yönlerini ortaya çıkarır ve geliştiricilere daha iyi GAN mimarileri tasarlama imkanı sunar.

3
DÖNÜŞTÜRÜCÜ MODELLER

Transformer (Dönüştürücü Model), sıralı veriler üzerinde uzun menzilli bağımlılıkları etkili bir şekilde öğrenebilmek için geliştirilmiş bir yapay sinir ağı mimarisidir. İlk olarak Vaswani ve arkadaşları tarafından 2017 yılında tanıtılan bu mimari, dikkat (attention) mekanizmasını temel alarak geleneksel yöntemlerin sınırlamalarını aşmayı hedeflemiştir. Transformer modelleri, kendine dikkat (self-attention) mekanizmasını kullanarak uzun menzilli bağımlılıkları öğrenebilen bir mimariye sahiptir. Şekil 3.1 'de temel Transformer mimarisinin yapısı gösterilmektedir.

Transformer, paralel işlem yapabilme yeteneği ve sıralı verilerde bağımsız bir şekilde bağlam öğrenme kapasitesiyle hem doğal dil işleme (NLP) hem de bilgisayarla görme (CV) alanlarında devrim yaratmıştır. Mimari, encoder (kodlayıcı) ve decoder (çözücü) olmak üzere iki ana bileşenden oluşur. [30, 31]

3.1 Genel Yapı: Kodlayıcı-Çözücü Mimarisi

Transformer, sıralı verilerle çalışırken encoder(kodlayıcı) ve decoder (çözücü) olmak üzere iki ana bölümden oluşur. Encoder, giriş dizisini işler ve sabit boyutlu bir temsil (vektör) üretir. Bu temsil, girdinin anlamını kapsar ve modelin daha sonraki adımlarında kullanılır. Decoder ise bu temsili kullanarak çıktı dizisini adım adım oluşturur. Çıktıyı oluştururken her adımda önceki çıktıların bilgisini kullanır. Bu yapı, *auto-regressive* (kendine bağlı) bir süreçtir, yani model bir sonraki adımı tahmin ederken yalnızca geçmiş adımları kullanır.

3.2 Kodlayıcı Bileşenleri

Encoder, birbiri üzerine yığılmış altı katmandan oluşur (standart olarak $N = 6$). Her encoder katmanı, aşağıdaki bileşenlerden meydana gelir:

1. **Multi-Head Self-Attention Mekanizması:** Modelin her bir token'in diğer token'lerle olan ilişkisini öğrenmesini sağlar.
2. **Feed-Forward Network (FFN):** Her token'in özelliklerini daha yüksek seviyede işlemeye yardımcı olur.
3. **Artık Bağlantılar ve Normalizasyon:** Katman geçişlerinde bilgiyi dengeli hale getirerek öğrenmeyi stabil tutar.

Bu yapı, giriş dizisinin anlamını modelin daha iyi öğrenmesini sağlamak için paralel hesaplamalarla desteklenmiştir.

3.2.1 Multi-Head Self-Attention Mekanizması

Bu mekanizma, her token'in diğer tüm token'lerle olan ilişkisini öğrenmesini sağlar. Bu işlem, dikkat matrisinin (attention matrix) hesaplanmasıyla gerçekleştirilir:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \tag{3.1}$$

Bu formülde 3.1;

- **Query (Q)**, **Key (K)** ve **Value (V):** Girişten türetilen üç ayrı matristir.
- **Softmax Fonksiyonu:** Dikkat ağırlıklarını normalize ederek toplamlarının 1 olmasını sağlar.
- **Ölçeklendirme Faktörü** $\sqrt{d_k}$: Büyük çarpımların dengelenmesini sağlar.

Bu mekanizma, modelin tüm giriş token'lerini birbirleriyle ilişkilendirerek bağlamı anlamasına yardımcı olur. Ayrıca, farklı başlıklar (heads) üzerinden çalışarak modelin farklı özellikleri öğrenmesine olanak tanır.

3.2.2 Point-Wise Feed-Forward Network (FFN)

Her encoder katmanında, dikkat mekanizmasından sonra token başına doğrusal bir dönüşüm uygulanır:

$$\text{FFN}(x) = \max(0, xW_1 + b_1)W_2 + b_2 \tag{3.2}$$

Bu katman, her token'i bağımsız olarak işleyerek modelin daha yüksek seviyeli özellikleri yakalamasına yardımcı olur. FFN iki aşamalı bir yapıya sahiptir:

- İlk aşamada, bir doğrusal dönüşüm uygulanarak daha yüksek boyutlu bir uzaya çıkılır.
- İkinci aşamada, bir aktivasyon fonksiyonu (genellikle ReLU) uygulanır ve ardından tekrar giriş boyutuna geri projeksiyon yapılır.

3.2.3 Artık Bağlantılar ve Normalizasyon

- **Artık Bağlantılar (Residual Connections):** Bilginin akışını kolaylaştırarak modelin daha hızlı öğrenmesini sağlar.
- **Katman Normalizasyonu (Layer Normalization):** Modelin kararlı çalışmasını sağlamak için tüm katmanların giriş ve çıkışlarını normalize eder. Bu normalizasyon işlemi, modelin daha verimli öğrenmesini sağlar.

3.3 Çözücü Bileşenleri

Decoder (çözücü), encoder (kodlayıcı) ile benzer bir yapıya sahiptir ancak ek olarak çıktının oluşturulmasını sağlayan mekanizmalar içerir:

1. **Masked Multi-Head Self-Attention:** Gelecek token'lere erişimi engellemek için maskeleme işlemi içerir.
2. **Encoder-Decoder Attention:** Encoder çıktısını kullanarak bağlam bilgisini model içine dahil eder.
3. **Feed-Forward Network (FFN):** Encoder'daki ile aynı şekilde çalışır.
4. **Artık Bağlantılar ve Normalizasyon:** Bilginin akışını düzenler.

3.3.1 Masked Multi-Head Self-Attention

Bu mekanizma, her token'in yalnızca kendisinden önce gelen token'leri dikkate almasını sağlar. Gelecekteki token'lerin bilinmemesi için bir üçgen maskeleme uygulanır. Maskelenmiş self-attention işlemi, her adımda yalnızca bilinen bilgilere dayalı tahmin yapmaya zorlar.

3.3.2 Encoder-Decoder Attention

Decoder katmanlarının encoder tarafından üretilen bağlam bilgisini kullanmasını sağlar. Bu katmanda, query'ler decoder'dan, key ve value'ler encoder'dan gelir. Bu yapı, girdi dizisi ile çıktı dizisi arasında güçlü bağlam ilişkileri kurar ve daha anlamlı tahminler yapılmasını sağlar.

3.3.3 Point-Wise Feed-Forward Network

Encoder ile aynı şekilde çalışarak her token için bağımsız özellik işleme adımları gerçekleştirir. Bu adımda, her token'in özellikleri dönüştürülerek modelin daha yüksek seviyeli soyutlamalar yapmasına yardımcı olunur.

3.4 Multi-Head Attention Mekanizması

Transformer'ın en güçlü bileşenlerinden biri olan Multi-Head Attention, birden fazla paralel dikkat mekanizmasını içerir. Bu mekanizma, modelin farklı başlıklar altında farklı bağlamları öğrenmesini sağlar:

$$\text{MultiHead}(Q, K, V) = \text{Concat}(\text{head}_1, \ldots, \text{head}_H)W^O \tag{3.3}$$

Burada 3.3, her başlık Query, Key ve Value matrisleri üzerinden çalışarak kendi dikkat matrislerini hesaplar.

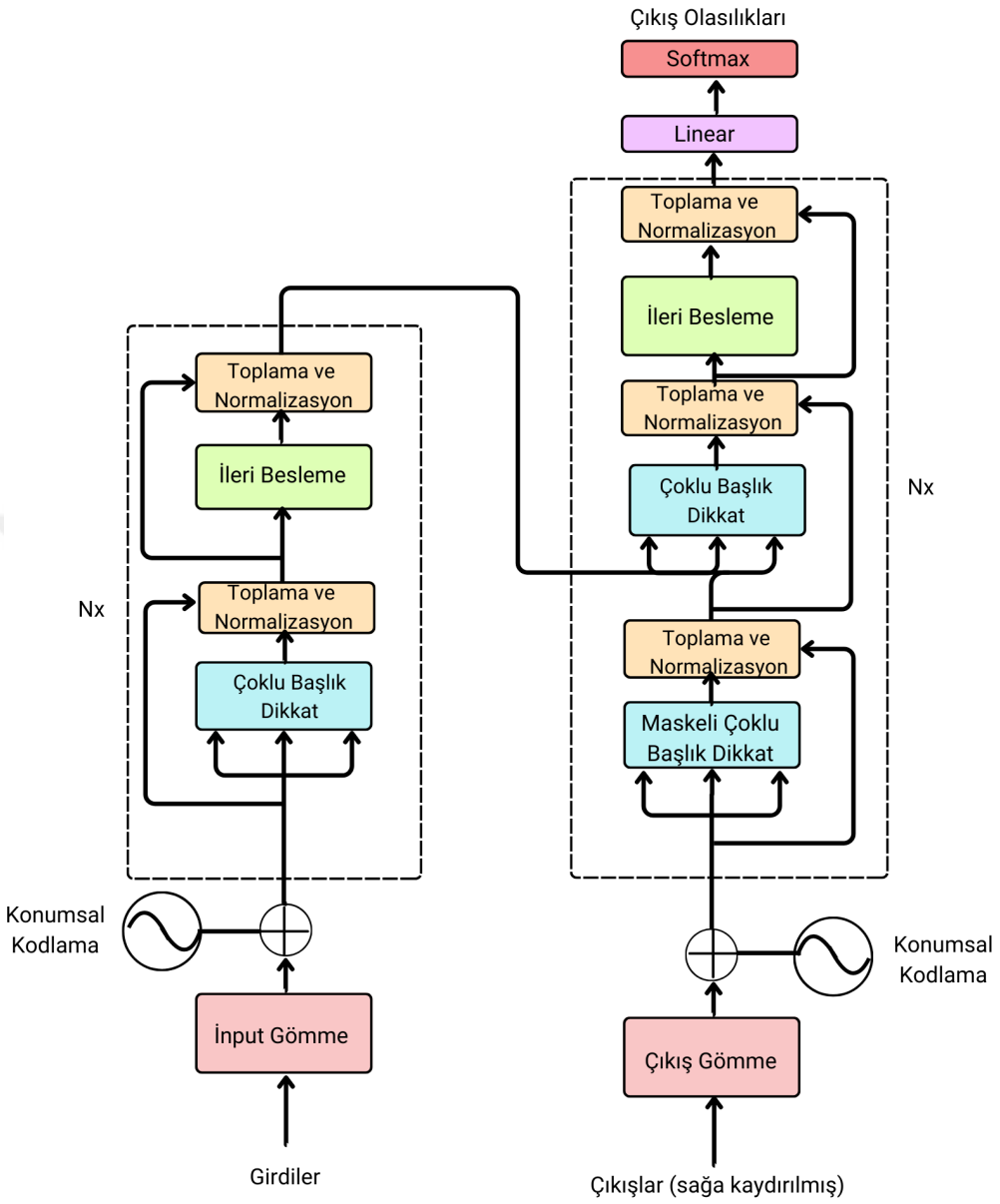
3.5 Pozisyon Kodlamaları (Positional Encodings)

Transformer, sıralı verileri işlemek için pozisyon kodlamaları (positional encodings) kullanır. Bu, sinüzoidal fonksiyonlar kullanılarak hesaplanır:

$$\text{PE}_{(pos,2i)} = \sin\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{3.4}$$

$$\text{PE}_{(pos,2i+1)} = \cos\left(\frac{pos}{10000^{2i/d_{model}}}\right) \tag{3.5}$$

Bu kodlamalar 3.4,3.5 , giriş embed'lerine eklenerek modele pozisyon bilgisini kazandırır.



Şekil 3.1 Transformer mimarisi

3.6 Çıkış Katmanı

Decoder'dan alınan çıktılar, doğrusal bir dönüşüm ve softmax fonksiyonu ile işlenerek olasılık dağılımına çevrilir. Bu süreç, dil modelleme veya çeviri gibi uygulamalarda bir sonraki token'in tahmin edilmesini sağlar.

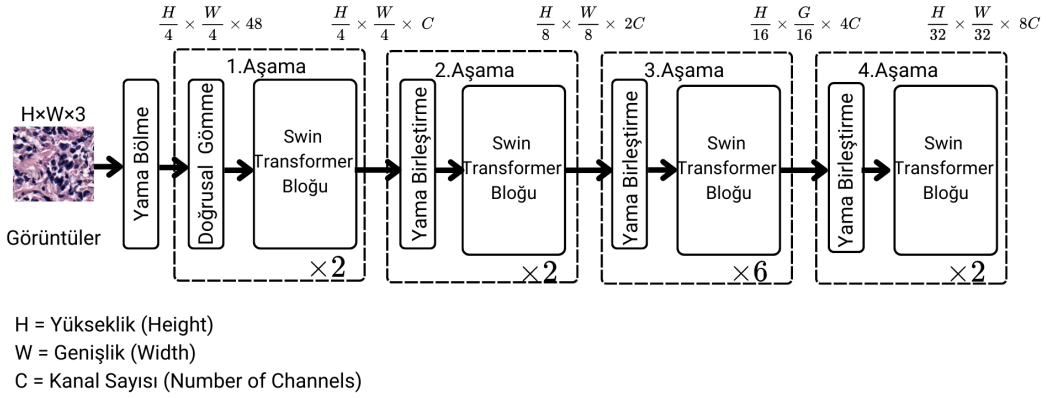
3.7 Swin Transformer: Gelişmiş Bir Mimari

Swin Transformer, Vision Transformer'ın eksikliklerini gidermek ve daha yüksek verimlilik sağlamak amacıyla geliştirilmiştir. Model, yerel özellikleri öğrenmek için kayan pencere (shifted window) mekanizmasını kullanır ve hiyerarşik bir yapı ile küresel bağlamı modelleme kapasitesini artırır.

3.7.0.1 Swin Transformer Versiyon 1

Swin Transformer, görsel veri işleme alanında saf Transformer mimarisini benimseyen ve dikkat çekici bir yenilik sunan derin bir öğrenme modelidir [55]. Şekil 3.2 'de Swin Transformer'ın ilk sürümünün mimari detayları gösterilmektedir. Hem yerel hem de küresel bağlamları dikkate alan tasarımı, görsel verilerin daha ayrıntılı ve kapsamlı işlenmesine olanak tanır. Özellikle görüntü sınıflandırma, segmentasyon ve nesne algılama gibi temel görevlerde elde ettiği yüksek başarı oranları, modelin geniş bir uygulama yelpazesinde kullanılabilirliğini göstermektedir.

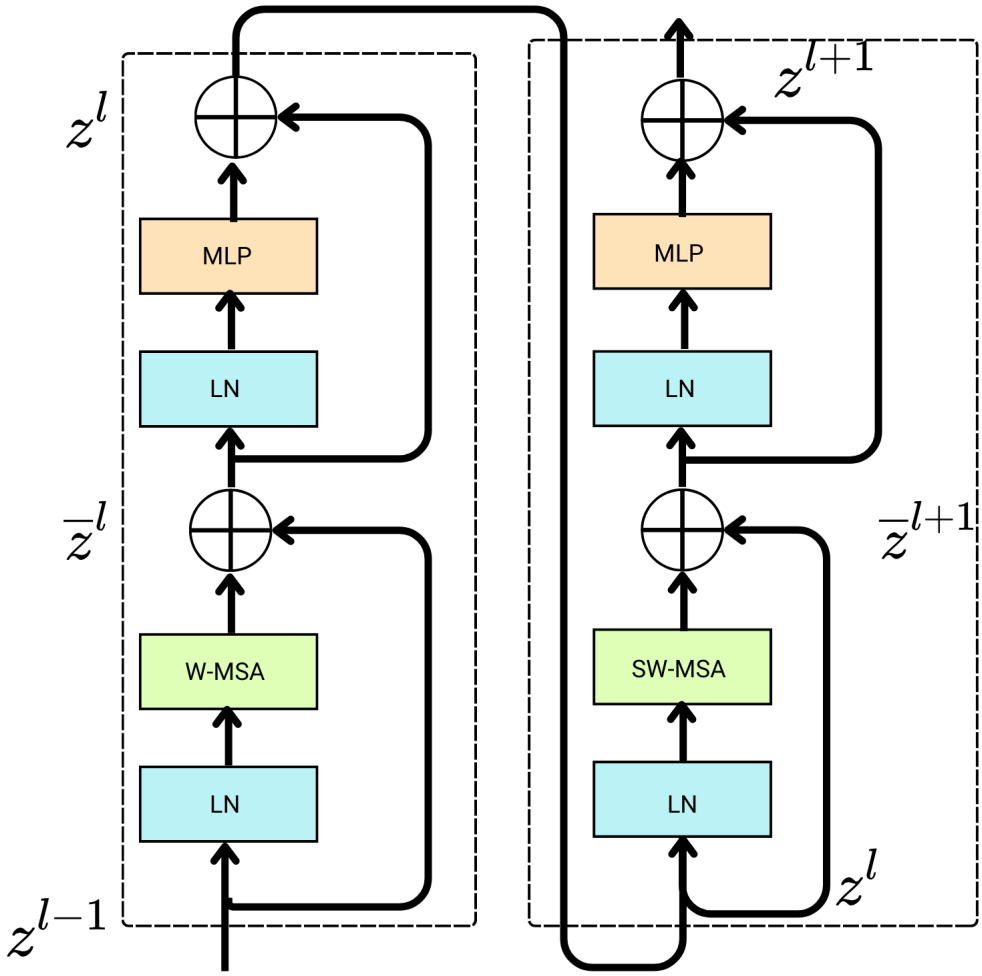
Kaydırılmış pencere dikkat mekanizması, Swin Transformer'ın temel yapı taşlarından biridir. Giriş görüntülerini sabit boyutlu küçük pencerelere bölerek, bu pencereler üzerinde dikkat işlemleri gerçekleştirir. Geleneksel pencere tabanlı yöntemlerden farklı olarak, Swin Transformer her katmanda pencereleri kaydırarak yeniden düzenler ve böylece pencereler arası bilgi akışını sağlar. Bu yenilikçi yaklaşım, modelin hem yerel detayları hem de küresel bağlamları etkili bir şekilde öğrenmesini mümkün kılar.



Şekil 3.2 Swin Transformer versiyon 1 mimarisi

Hiyerarşik yapı, Swin Transformer'ın bir diğer önemli özelliğidir. Model, her katmanda giriş görüntülerinin boyutlarını kademeli olarak küçülterek farklı çözünürlük seviyelerinde özellik temsilleri oluşturur. Bu hiyerarşik tasarım, görüntülerin hem düşük seviyeli ayrıntılarının hem de yüksek seviyeli anlamlarının etkili bir şekilde öğrenilmesini sağlar. Girdi verileri, sabit çözünürlükteki yamalara bölünerek tokenlere dönüştürülür ve ilk katmanlarda yerel özellikler öğrenilirken, sonraki katmanlarda daha karmaşık ve soyut özellikler temsil edilir.

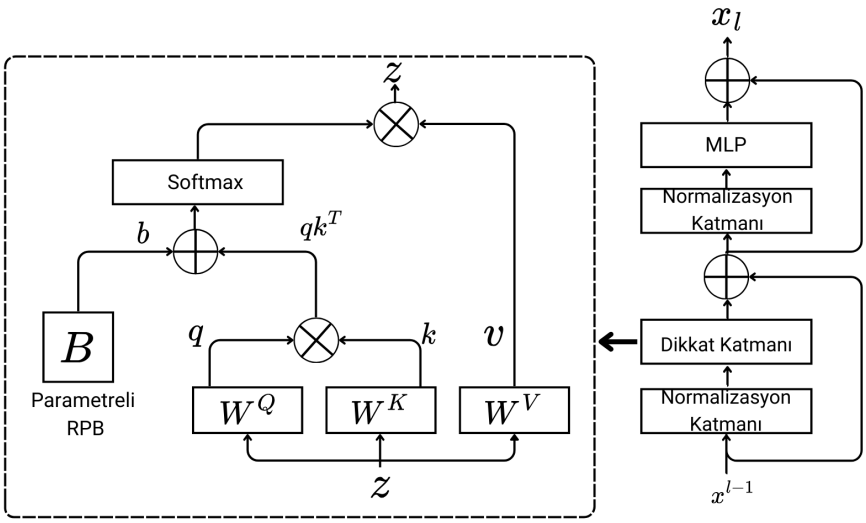
Çok kafalı dikkat mekanizması, her pencere içindeki özellikler arasındaki ilişkileri öğrenirken, kaydırılmış pencere düzenlemesi farklı pencereler arasındaki bilgi akışını sağlar. Tam bağlantılı katmanlar (MLP) ise her token'ın daha derinlemesine işlenmesine katkıda bulunur. Şekil 3.3 Burada yazarların vurgusu, kaydırılmış pencere mekanizması ve hiyerarşik yapının, yerel ve küresel bağlamların etkili bir şekilde birleştirilmesini sağlamasıdır. Bu sayede model, hem küçük ölçekli detayları hem de büyük ölçekli yapıları öğrenebilir. Ayrıca, Swin Transformer'ın esnek tasarımı, modelin görüntü işleme dışında video işleme, doğal dil işleme ve 3B nesne tanıma gibi çeşitli alanlarda da kullanılabilmesine olanak tanır. Giriş boyutuna bağlı ölçeklenebilirliği ve parametre verimliliği, modelin geniş veri setlerinde etkili bir şekilde uygulanabilmesini sağlar [55].



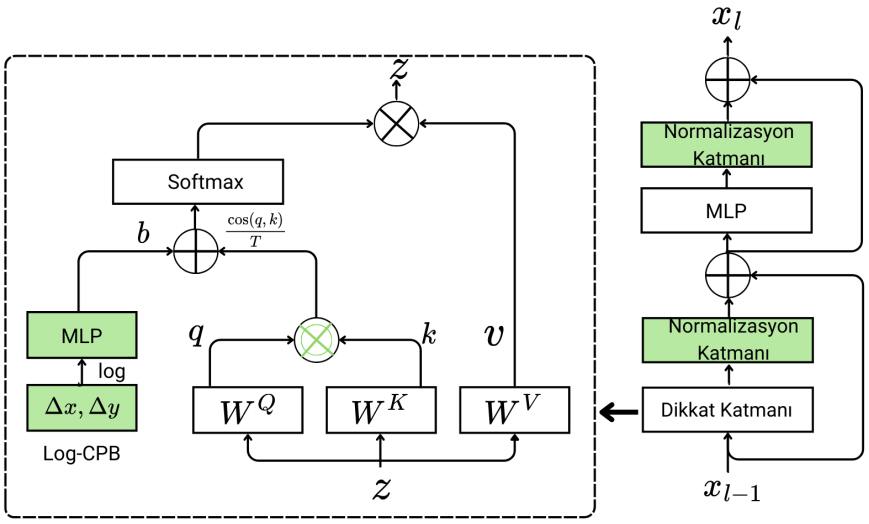
MLP: Çok Katmanlı Algılayıcı
LN: Katman Normalizasyonu
W-MSA: Pencere Tabanlı Çoklu Başlı Öz-Dikkat
SW-MSA: Kaydırılmış Pencere Tabanlı Çoklu Başlı Öz-Dikkat

Şekil 3.3 İki ardışık Swin Transformer bloğu

Sonuç olarak, Swin Transformer, kaydırılmış pencere dikkat mekanizması, hiyerarşik tasarımı ve parametre verimliliği gibi özellikleriyle görsel veri işleme alanında yenilikçi bir yaklaşım sunar ve özellikle görüntü sınıflandırma, segmentasyon ve nesne algılama gibi görevlerde üstün performans gösterir. Hem akademik çalışmalarda hem de endüstriyel uygulamalarda kullanım potansiyeli yüksek olan bu model, Transformer tabanlı mimarilerin başarısını daha da ileriye taşımaktadır.



Şekil 3.4 Swin Transformer versiyon 1 bloğu



Şekil 3.5 Swin Transformer versiyon 2 bloğu

Tablo 3.1 Swin Transformer v1 ve v2 karşılaştırması

Özellik	**Swin Transformer v1**	**Swin Transformer v2**
Normalizasyon	Pre-Norm	Res-Post-Norm
Dikkat Mekanizması	Nokta çarpımı dikkati	Ölçeklendirilmiş kosinüs dikkati
Konumsal Önyargı	Parametrelendirilmiş	Log-CPB
Ölçeklenebilirlik	Sınırlı	Geliştirilmiş (3B parametre, 1536x1536 çözünürlük)
Eğitim Kararlılığı	Büyük modellerde kararsız	Daha kararlı
Transfer Öğrenimi (Çözünürlük/Pencere Boyutu)	Bi-kübik enterpolasyon	Log-CPB ile daha etkili
Kök (Stem)	Yok	Var (evrişimli katmanlar)
Örnekleme Azaltma	Yama Birleştirme	Çift Dallanmalı Örnekleme Azaltma
İleri Beslemeli Ağ	MLP	Ters Artık İleri Beslemeli Ağ (IRFFN)

Tablo 3.2 Swin Transformer ve diğer mimari yaklaşımlar: Karşılaştırmalı bir analiz

Model	**Mimari**	**Güçlü Yönleri**	**Zayıf Yönleri**	**Swin Transformer ile Farklılıklar**
ResNet	Derin CNN, atlamalı bağlantılar	Derin ağ eğitimi, kaybolan gradyan problemi	Sınırlı küresel bağımlılık yakalama	Dikkat mekanizması ve hiyerarşik yapı kullanır.
EfficientNet	Ölçeklenebilir CNN	Verimlilik, farklı ölçeklerde performans	Göreve özgü optimizasyon eksikliği	Ölçeklendirme ve mimari yaklaşımı farklıdır.
ViT	Transformer kodlayıcısı	Büyük veri kümelerinde etkili	Yüksek hesaplama maliyeti, küçük veri kümelerinde genelleme sorunu	Hiyerarşik yapı ve kaydırılmış pencereler ile hesaplama maliyetini azaltır ve genellemeyi iyileştirir.
Inception	Çok dallı CNN	Çok ölçekli özellik yakalama	Karmaşık mimari	Dikkat mekanizması ile çok ölçekli modelleme yapar, daha basit bir yapıya sahiptir.

Swin Transformer V2, orijinal Swin Transformer mimarisini temel alarak model kapasitesini artırmak ve yüksek çözünürlüklü görüntü işleme yeteneklerini geliştirmek amacıyla tasarlanmış bir modeldir. Şekil 3.4 ve Şekil 3.5 'de, Swin Transformer v1 ve v2'nin blok katman yapıları ve öznitelik çıkarma mekanizmalarındaki farklılıklar gösterilmektedir. Görsel görevlerdeki sınırları aşmayı hedefleyen bu mimari, model stabilitesini artıran yenilikçi yöntemler, yüksek çözünürlükler için adaptif mekanizmalar ve verimli bellek yönetimi optimizasyonları ile ön plana çıkmaktadır [4]. Ayrıca (Tablo 3.1) 'de Swin Transformer v1 ve v2 arasındaki temel farklar ve iyileştirmeler karşılaştırmalı olarak verilmiştir.

Mimari Yenilikler ve Teknik İyileştirmeler

Swin Transformer V2'nin mimarisinde gerçekleştirilen temel yenilikler, modelin daha geniş kapasitelere ve çözünürlük aralıklarına uyum sağlamasına olanak tanır. Bu yenilikler, modelin doğruluk oranını artırırken eğitim sürecindeki kararlılığını da sağlamaktadır:

- **Residual Post-Normalization:**
  Modelin derin katmanlarında artan aktivasyon amplitüdü nedeniyle yaşanan kararsızlık sorunlarını çözmek için rezidüel bloklardaki normalizasyon işlemleri, bloğun sonunda uygulanacak şekilde yeniden düzenlenmiştir. Bu düzenleme, özellikle büyük modellerde eğitimin stabil bir şekilde tamamlanmasını sağlamış ve aktivasyon değerlerindeki aşırı değişiklikleri minimize etmiştir [4].

- **Scaled Cosine Attention:**
  Geleneksel nokta çarpımı dikkat mekanizmasının büyük modellerde neden olduğu dengesizlikler, ölçeklenmiş kosin dikkat mekanizması ile çözülmüştür. Bu yaklaşım, dikkat haritalarındaki değerlerin daha homojen bir şekilde dağıtılmasını sağlamış ve modelin genel performansını artırmıştır.

- **Log-Spaced Continuous Position Bias (Log-CPB):**
  Orijinal Swin Transformer, pencere boyutları arasındaki geçişlerde bicubic interpolasyon yöntemi kullanırken, Swin Transformer V2'de logaritmik olarak aralıklı sürekli konum yanlılığı mekanizması geliştirilmiştir. Bu yöntem, modelin düşük çözünürlükte önceden eğitilmiş ağırlıklarının yüksek çözünürlüklere transfer edilmesinde daha başarılı sonuçlar elde edilmesini sağlamıştır.

Performans ve Değerlendirmeler

Swin Transformer V2, çeşitli görsel görevlerde önemli performans artışları sağlamıştır:

- **ImageNet-V2:** %84.0 top-1 doğruluk oranı ile önceki en iyi modeli %0.7 oranında geçmiştir.
- **COCO Nesne Algılama:** 63.1/54.4 box/mask mAP değerleri ile %1.8/%1.4'lük iyileşme kaydedilmiştir.
- **ADE20K Semantik Segmentasyon:** 59.9 mIoU ile önceki modellere göre %1.5'lik bir doğruluk artışı sağlanmıştır [4].

Bellek Kullanımı Optimizasyonları

Yüksek çözünürlüklü görüntüler ve büyük modeller, GPU bellek tüketimini önemli ölçüde artırır. Swin Transformer V2, bu sorunu çözmek için şu yöntemleri uygulamıştır:

- **Zero-Redundancy Optimizer (ZeRO):** Optimizasyon durumlarının GPU'lar arasında paylaşılmasıyla bellek tüketimi azaltılmıştır.
- **Activation Checkpointing:** Ara katman aktivasyonlarının saklanmasını optimize ederek bellek kullanımını düşürmüştür.
- **Sequential Self-Attention Computation:** Özellikle yüksek çözünürlüklerde dikkat hesaplamalarının sıralı olarak yapılmasını sağlayarak bellek yükü dengelenmiştir.

Swin Transformer V2, görsel görevlerde model kapasitesini artırarak doğruluk oranlarını önemli ölçüde iyileştiren, yenilikçi ve ölçeklenebilir bir mimaridir.Transformer tabanlı modellerin karşılaştırılması, hangi modelin hangi uygulamalar için daha uygun olduğunu anlamak açısından önemlidir. (Tablo 3.2) 'de, Swin Transformer ve diğer popüler Transformer mimarileri arasında yapılan bir karşılaştırma sunulmaktadır. Bu model, büyük ölçekli görsel modellerin potansiyelini ortaya koyarken veri verimliliğini artırmak ve bellek yönetimini optimize etmek için etkili çözümler sunmaktadır. Bu çalışmanın, görsel modellerle dil modelleri arasındaki kapasite boşluğunu kapatmaya ve bu iki alanın bütünleşik olarak modellenmesine katkı sağlaması beklenmektedir [4].

4
BÜYÜK DİL MODELLERİ VE ÇOK MODLU BÜYÜK DİL MODELLERİ

Doğal dil işleme (NLP) alanında yaşanan gelişmeler, insan dilini anlama ve üretme yeteneğine sahip büyük ölçekli yapay zekâ sistemlerinin ortaya çıkmasını sağlamıştır. Büyük Dil Modelleri (Large Language Models - LLM), milyarlarca parametre içeren sinir ağı yapılarıyla eğitilmiş, geniş veri kümeleri üzerinde öğrenme gerçekleştiren ve dilin karmaşık yapısını modelleyebilen derin öğrenme tabanlı sistemlerdir. Bu modeller, çeviri sistemleri, metin özetleme, metinden metne dönüşüm, soru-cevap sistemleri ve etkileşimli sohbet uygulamaları gibi birçok alanda etkin bir şekilde kullanılmaktadır.

LLM'lerin başarısı, 2017 yılında tanıtılan Transformer mimarisine dayanmaktadır. Transformer, geleneksel dil işleme yöntemlerinden farklı olarak kendi üzerine dikkat (self-attention) mekanizmasını kullanarak uzun mesafeli bağıntıları daha etkin bir şekilde modelleyebilmekte ve paralel hesaplama kapasitesi sayesinde büyük ölçekli veri işleme süreçlerini hızlandırmaktadır. Bu özellikler, LLM'lerin çok büyük parametre setlerine sahip olmasını ve büyük miktarda veriden öğrenerek bağlamsal anlam çıkarma yeteneklerini geliştirmesini sağlamaktadır.

Ancak insan algısı yalnızca dil tabanlı girdilerle sınırlı değildir. Günlük yaşamda görsel, işitsel ve metinsel bilgilerin birlikte işlendiği çok modlu bilişsel süreçler büyük önem taşımaktadır. Geleneksel LLM'lerin yalnızca metin tabanlı verilerle eğitilmesi, yapay zekâ tabanlı sistemlerin görseller, sesler ve diğer veri türleriyle anlamlı ilişkiler kurma yetisini kısıtlamaktadır. Bu eksiklikleri gidermek amacıyla geliştirilen Multimodal Büyük Dil Modelleri (Multimodal Large Language Models - MLLM), farklı veri türlerini entegre ederek çok modlu bilgi sentezi yapabilen yapay zekâ sistemleri olarak öne çıkmaktadır.

MLLM'ler, dil ve görsel bilgiyi birlikte işleyerek görsel ve metinsel açıklamalar oluşturma, konuşma tabanlı komutları anlamlandırma, metin-görsel ilişkilendirme

ve karmaşık çok modlu veri analizi yapma gibi görevlerde yüksek başarı göstermektedir. GPT-4V, Kosmos ve Flamingo gibi modeller, doğal dil ile görsel veriyi bütünleştirerek yapay zekâ destekli uygulamaların kapsamını genişletmektedir.

Bununla birlikte, büyük dil modellerinin eğitimi ve kullanımı, yüksek hesaplama maliyetleri ve büyük parametre setleri nedeniyle ciddi donanım gereksinimleri ortaya çıkarmaktadır. Bu modellerin daha verimli çalışmasını sağlamak için geliştirilen LoRA (Low-Rank Adaptation) ve PEFT (Parameter-Efficient Fine-Tuning) gibi teknikler, model performansından ödün vermeden eğitim sürecini optimize etmekte ve ince ayar süreçlerini hızlandırmaktadır.

Bu çalışma kapsamında, LLM ve MLLM'lerin mimari yapıları, çalışma prensipleri, model optimizasyon yöntemleri ve gelecekteki gelişim alanları detaylı bir şekilde ele alınacaktır. Özellikle Transformer mimarisinin temel bileşenleri, multimodal sistemlerin yapısal bileşenleri ve büyük ölçekli modellerin eğitimi için kullanılan optimizasyon yöntemleri incelenerek, yapay zekâ tabanlı dil işleme sistemlerinin günümüzde geldiği nokta ve gelecekteki yönelimleri akademik bir bakış açısıyla değerlendirilecektir.

4.1 Büyük Dil Modeli (LLM)

Büyük Dil Modelleri (LLM'ler), modern doğal dil işleme uygulamalarının temelini oluşturan, genellikle milyarlarca parametre içeren ve büyük ölçekli verilerle önceden eğitilmiş derin öğrenme modelleridir. LLM'lerin mimarisi, öncelikle Transformer tabanlı yapılar üzerine inşa edilmiştir. Transformer mimarisi, geleneksel RNN veya LSTM tabanlı modellere kıyasla paralel hesaplama ve uzun menzilli bağımlılıkların etkin modellenmesi açısından büyük avantajlar sağlamaktadır [56].

Temel Transformer bloğu, çok başlı (multi-head) dikkat mekanizması, katman normalizasyonu, ileri beslemeli ağ (feed-forward network) ve artık (residual) bağlantılardan oluşur. Çok başlı dikkat mekanizması, modelin giriş dizisindeki farklı alt-temsil özelliklerini aynı anda öğrenebilmesini mümkün kılar. Her dikkat başlığı, sorgu (query), anahtar (key) ve değer (value) matrisleri üzerinden hesaplamalar yapar; böylece model, farklı alt uzaysal ve anlamsal ilişkileri paralel olarak değerlendirir. Bu sayede, dilin karmaşık yapısal ve anlamsal özellikleri daha iyi yakalanır [56].

Ayrıntılı olarak ele alındığında, Transformer mimarisinde kullanılan kendine dikkat

(self-attention) mekanizması, her bir giriş elemanının diğer tüm elemanlarla olan ilişkisini değerlendirir. Bu süreçte, giriş dizisinin her öğesi hem sorgu hem de anahtar ve değer olarak işlev görür. Hesaplanan dikkat ağırlıkları, her kelimenin diğer kelimelerle olan bağlamını temsil eden yeni bir gösterim oluşturulmasına olanak tanır. Bu mekanizma, modelin özellikle uzun metinlerdeki bağlamı anlamlandırmasını sağlar ve dilin uzun menzilli bağımlılıklarını yakalar [56].

Transformer mimarisinde kullanılan konumsal kodlamalar (positional encodings), kelime sıralaması bilgilerini modele dahil eder. Standart sinüs-kosinus temelli ya da öğrenilebilir konumsal kodlamalar, modelin sıralı verideki konum bilgilerini korumasını ve böylece cümle yapısının doğru anlaşılmasını destekler. Bu, özellikle dilin sözdizimsel ve anlamsal yapısının modellenmesi açısından kritik öneme sahiptir.

LLM'ler genellikle, devasa metin koleksiyonları üzerinde kendi kendine denetimsiz (self-supervised) öğrenme ile önceden eğitilir. Bu süreçte, model "masked language modeling" veya "next token prediction" gibi görevleri kullanarak dilin yapısal ve anlamsal özelliklerini öğrenir. Ön eğitim süreci, modelin geniş bir dil bilgisini edinmesini sağlarken, sonrasında belirli görevler için ince ayar (fine-tuning) yapılarak modelin belirli uygulamalara uyarlanması mümkün hale gelir [57].

Ek olarak, günümüzde LLM mimarileri ölçeklenebilirlik prensipleri doğrultusunda geliştirilmektedir. Ölçek yasaları (scaling laws) gösteriyor ki, model boyutunun, veri miktarının ve hesaplama gücünün artması, modelin genel performansında öngörülebilir iyileşmeler sağlamaktadır. Bu durum, araştırmacıların model mimarisini daha derin ve geniş hale getirerek, daha karmaşık dil yapılarını modelleyebilmesini mümkün kılar. Ancak, bu ölçeklenebilirlik aynı zamanda yüksek hesaplama ve enerji maliyetleri gibi pratik sorunları da beraberinde getirmektedir.

LLM'lerin mimarisinde kullanılan önemli bir diğer bileşen ise, parametre verimliliğini artırmaya yönelik stratejilerdir. Özellikle "parameter-efficient fine-tuning" (PEFT) yöntemleri, tüm modelin yeniden eğitilmesine gerek kalmadan, belirli görevler için küçük miktarda parametre güncellemesi yaparak modelin performansını artırır. Bu yaklaşımlar, modelin hafızadan tasarruf etmesini ve farklı görevler arasında hızlı geçiş yapabilmesini mümkün kılar.

Modern LLM'lerde, bazı varyantlar modelin hesaplama yükünü azaltmak için seyrek (sparse) dikkat mekanizmaları, yaklaşık (approximate) dikkat hesaplamaları ya da dikkat mekanizması olmadan alternatif yaklaşımlar da kullanmaktadır. Bu tür modifikasyonlar, modelin hem eğitim hem de çıkarım (inference) aşamasında

hesaplama verimliliğini artırırken, performans kaybını minimize etmeyi hedefler.

Bunun yanı sıra, LLM mimarilerinde kullanılan ileri beslemeli ağlar, genellikle iki katmandan oluşur ve aktivasyon fonksiyonları olarak ReLU, GELU gibi fonksiyonlar tercih edilir. Bu katmanlar, dikkat mekanizması tarafından oluşturulan bağlam temsillerini işleyerek daha karmaşık ve soyut özelliklerin öğrenilmesine olanak tanır. Katman normalizasyonu ise, modelin eğitim sürecinde istikrarlı bir optimizasyon yapabilmesini sağlayarak, aşırı değerlerin etkisini azaltır.

Son olarak, LLM'ler mimarisinin başarısı, büyük veri setleri, geniş model kapasiteleri ve etkili optimizasyon algoritmalarının bir kombinasyonu ile mümkün olmaktadır. Transformer mimarisi, paralel hesaplama yetenekleri ve uzun menzilli bağımlılıkları yakalama kabiliyeti sayesinde, dilin karmaşık yapısal özelliklerini modellemede devrim yaratmıştır. Bu mimari yapı, hem metin üretimi hem de dil anlama gibi pek çok görevde üstün performans sergilemekte, dolayısıyla geniş bir uygulama yelpazesi için temel oluşturmaktadır [56, 57].

Özetle, LLM mimarileri, Transformer'ın kendine dikkat mekanizması, çok başlı dikkat, ileri beslemeli ağlar ve katman normalizasyonu gibi temel bileşenleri merkezine alarak, devasa ölçekli veriler üzerinde kendi kendine öğrenme ile eğitilmektedir. Bu yapı, dilin anlamsal ve sözdizimsel karmaşıklığını yakalamada son derece etkili olup, modern doğal dil işleme uygulamalarında devrim yaratmıştır. Araştırmalar, bu temel mimarinin üzerine yapılan yenilikler sayesinde, modelin ölçeklenebilirliği, verimliliği ve uygulama alanlarının sürekli genişlediğini göstermektedir.

4.2 Çoklu Modlu Büyük Dil Modeli (Multimodal LLM)

Multimodal LLM'ler, klasik LLM'lerin temelini oluşturan Transformer mimarisinin ötesine geçerek, yalnızca metin verisini değil, aynı zamanda görsel, ses, video ve diğer veri modalitelerini de işleyebilen ileri düzey yapay zeka sistemleridir. Bu modeller, öncelikle her modaliteye özgü encoder'lar kullanarak, ilgili veriyi modelin işleyebileceği yüksek boyutlu temsillere dönüştürür. Örneğin, görsel veriler için genellikle Vision Transformer (ViT) veya CLIP gibi modeller tercih edilirken, ses ve video gibi diğer modaliteler için de benzer temsiliyet çıkarım yöntemleri uygulanır. Elde edilen bu temsiller, daha sonra adapter modüller aracılığıyla LLM'lerin dil işleme bileşenleri ile entegre edilir. Adapter modüller, farklı modalitelerden gelen verileri aynı boyuta indirger ve ortak bir latent alanda birleşmelerini sağlar; böylece model, çapraz-dikkat (cross-attention) mekanizmaları sayesinde modaliteler arasındaki ilişkileri etkin bir biçimde öğrenir

[58, 59].

Multimodal entegrasyon stratejileri arasında erken ve geç entegrasyon yaklaşımları önemli rol oynamaktadır. Erken entegrasyonda, farklı modaliteler modelin giriş aşamasında birleştirilir ve ortak bir temsiliyet alanı oluşturulur. Bu yöntem, modaliteler arası düşük seviyeli ilişkilerin yakalanmasına olanak tanırken, geç entegrasyon stratejisinde her bir modalite önce bağımsız olarak işlenir ve daha sonra yüksek seviyeli özelliklerin birleşimi için çapraz-dikkat gibi yöntemler kullanılır. Bu yaklaşım, modalitelerin özgün özelliklerini korurken, zenginleştirilmiş bir bağlamın oluşturulmasına yardımcı olur [57, 58].

Eğitim sürecinde multimodal LLM'ler, büyük ölçekli görsel–metin veri kümeleri ve diğer multimodal veri kaynakları kullanılarak önceden eğitilir. Bu süreç, modelin hem görsel algı hem de dil üretme yeteneklerini aynı anda geliştirmesine olanak tanır. "Instruction-tuning" gibi yöntemlerle model, kullanıcı yönergelerine uygun yanıtlar verebilecek şekilde ince ayardan geçirilir; bu, modelin karmaşık görsel sorgulamalara dayalı açıklama üretme, görüntü ile ilgili detayları yorumlama ve mantıksal çıkarım yapma yeteneklerini artırır [57, 58].

Multimodal LLM'lerde ölçeklenebilirlik ve verimlilik konuları, klasik LLM'lerde olduğu gibi önemli bir araştırma alanı oluşturmaktadır. Modelin boyutunun, veri miktarının ve hesaplama kaynaklarının artması, performans artışı sağlarken aynı zamanda yüksek enerji ve hesaplama maliyetlerini de beraberinde getirir. Bu nedenle, hesaplama yükünü azaltmaya yönelik seyrek dikkat mekanizmaları, parameter-efficient fine-tuning (PEFT) ve model sıkıştırma yöntemleri gibi stratejiler uygulanmaktadır. Bu teknikler, multimodal modellerin hem eğitim hem de çıkarım aşamalarında verimliliğini artırarak, pratik uygulamalarda daha geniş kullanım olanakları sunar [57, 60].

Sonuç olarak, multimodal LLM'ler, farklı modalitelerden gelen zengin bilgileri entegre ederek, insan benzeri algı ve yorumlama yeteneklerini ortaya koymayı hedefleyen kapsamlı yapay zeka sistemleridir. Bu modeller, modaliteye özgü encoder'lar, adapter modülleri ve çapraz-dikkat mekanizmaları sayesinde, dilin ötesinde, görsel, ses ve diğer veri türlerini başarılı bir şekilde işleyebilmektedir. Bu bütünleşik mimari yapı, multimodal verilerin karmaşıklığını yönetebilme kapasitesiyle, tıp, otomotiv, robotik, eğitim ve daha birçok alanda devrim niteliğinde uygulamalara olanak tanımaktadır. Ek kaynaklardan elde edilen veriler, mimari detayların derinlemesine anlaşılmasına katkı sağlayarak, modelin ölçeklenebilirliği, verimliliği ve pratik uygulamalardaki başarısının altını çizmektedir. Böylece, multimodal LLM'ler, günümüzün karmaşık veri

ortamlarında insan benzeri performans sergileyerek, yapay genel zekaya doğru atılmış önemli bir adım olarak değerlendirilebilir.

5
MATERYALLER VE YÖNTEMLER

5.1 Veri Kümesi

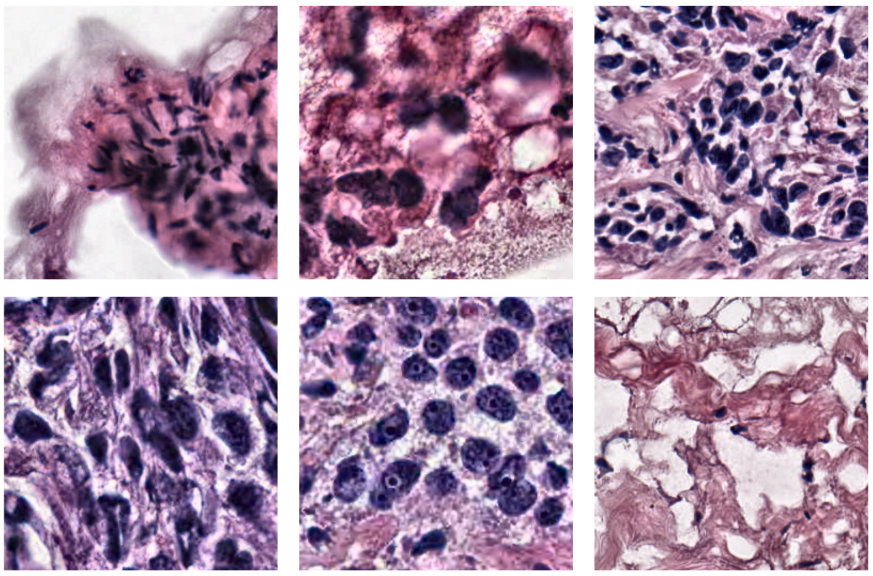
5.1.1 Veri Kümesine Genel Bakış

Bu çalışmada, Kanser Görüntüleme Arşivi (The Cancer Imaging Archive - TCIA) platformunda Klinik Proteomik Tümör Analiz Konsorsiyumu (CPTAC) tarafından sağlanan CPTAC-BRCA veri kümesi kullanılmıştır [61]. Bu veri kümesi, meme kanseri araştırmalarında yaygın olarak kullanılmakta olup, 134 hastaya ait biyolojik örnek DNA, RNA, miRNA, WXS ve klinik verilerle birlikte, SVS formatında 642 histopatolojik görüntü içermektedir.

Veri kümesinde, görüntüler 7 farklı meme kanseri evresine göre kategorize edilmiştir: Evre IA, Evre IIA, Evre IIB, Evre III, Evre IIIA, Evre IIIB ve Evre IIIC. Ancak, sınıf dağılımında önemli bir dengesizlik bulunmaktadır. Örneğin, Evre IIIC'de 4.200 örnek varken, en az temsil edilen Evre IA'da yalnızca 240 örnek bulunmaktadır. Diğer evreler de benzer şekilde değişen seviyelerde temsil edilmektedir. Veri kümesi, herhangi bir veri artırımı (augmentation) yapılmadan önce toplamda 9.065 görüntü içermektedir.

5.1.2 Veri Ön İşleme

Veri ön işleme sırasında, histopatolojik görüntüler 256×256 piksellik parçalara bölünmüştür. SVS dosyalarından düşük çözünürlüklü önizleme görüntüleri elde edilmiş, ardından ikili eşikleme yöntemi kullanılarak önemli bölgeler belirlenmiştir. Adaptif histogram eşitleme uygulanarak kontrast artırılmış ve sonuç olarak 19.515 görüntü parçası elde edilmiştir. Ayrıca, kanser evre bilgisi eksik olan 34 hasta analizden çıkarılarak, analiz kalan 100 hastanın tam verisiyle gerçekleştirilmiştir. Çalışmada kullanılan veri kümesinden bazı örnekler Şekil 5.1 'de gösterilmektedir. Bu görüntüler, 256×256 piksel boyutuna getirilerek modele giriş olarak verilmiştir.



Şekil 5.1 Veri kümesinden 256×256 görüntü örnekleri.

5.1.3 Veri Artırımı ve Dengeleme

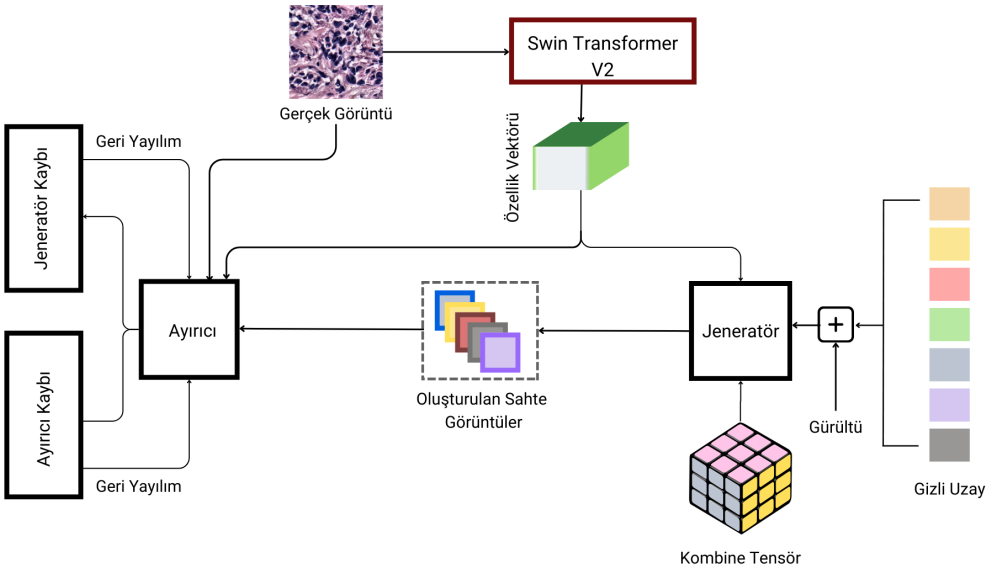
Sınıf dengesizliğini gidermek için veri artırımı (augmentation) teknikleri kullanılmıştır. Bu süreçte, tüm kanser evresi sınıflarındaki örnek sayıları, en yüksek sınıf olan Evre IIIC'nin 4.200 örneklik seviyesine getirilmiştir. Bu dengeleme sürecinde rastgele yatay ve dikey kaydırmalar, döndürmeler ve renk ayarlamaları gibi artırma teknikleri kullanılmıştır. Bu sayede, az temsil edilen sınıfların temsiliyeti artırılmış ve dengeli bir veri kümesi oluşturulmuştur. Veri artırımı sonucunda, veri kümesindeki toplam görüntü sayısı 9.065'ten 29.400'e çıkarılmış, böylece her sınıfın eğitim sürecine eşit katkı sağlaması sağlanmıştır.

5.1.4 Destekleyici Veri Birleştirme

Biyolojik örneklerden elde edilen DNA, RNA, miRNA, WXS ve klinik veriler, her bir hasta için sabit uzunlukta tensörlere dönüştürülmüştür. Bu tensörler daha sonra birleştirilmiş ve modelin giriş verisi olarak kullanılmıştır. Bu sayede, biyomoleküler ve klinik bilgiler modele etkili bir şekilde entegre edilmiştir.

5.2 Görüntü Üretme Model Mimarisi

Bu çalışma, derin öğrenme tabanlı bir görüntü sentezi yaklaşımıyla, yüksek kaliteli ve hastaya özgü kanser görüntüleri üretmek için tasarlanmış yeni bir mimari önermektedir. Swin Transformer ve StyleGAN3 modellerinin sinerjik entegrasyonuna dayanan bu model, hem görsel gerçekçiliği hem de hastaya özgü bilgilerin tutarlılığını koruyan sofistike bir öğrenme paradigması sunar. Önerilen



Şekil 5.2 Model mimarisi

mimari, şu şekilde işler: İlk olarak, önceden eğitilmiş bir Swin Transformer modeli, girdi kanser görüntülerinden ayırt edici özellikler çıkarır. Swin Transformer'ın hiyerarşik transformatör mimarisi, çoklu ölçeklerdeki özellikleri etkili bir şekilde yakalayarak, geleneksel yöntemlere kıyasla daha kapsamlı ve bağlamsal olarak zengin bir temsil sağlar. Bu özellikler, daha sonra Discriminator'ın eğitiminde ve değerlendirmesinde referans olarak kullanılır. İkinci olarak, StyleGAN3 mimarisi devreye girer ve rastgele bir gürültü vektörü ile hastaya özgü bilgileri içeren bir kombine tensörden koşullu olarak yeni kanser görüntüleri sentezler. Kombine tensör, hastanın demografik özellikleri, klinik bulguları veya genetik verileri gibi çeşitli bilgileri kodlayarak, üretilen görüntülerin ilgili hasta veya durumla uyumlu olmasını garantiler. StyleGAN3'ün Generator ve Discriminator bileşenleri, adversarial bir eğitim süreci boyunca rekabetçi bir şekilde eğitilir. Generator, Discriminator'ı yanıltmak için gerçekçi görüntüler üretmeye çalışırken, Discriminator ise gerçek ve sentetik görüntüleri doğru bir şekilde ayırt etmeyi öğrenir. Discriminator, üç temel girdi alır: gerçek görüntünün Swin Transformer tarafından çıkarılan özellikleri, Generator tarafından üretilen görüntü ve her iki görüntü için de ortak olan kombine tensör. Swin Transformer özellikleri, gerçekçilik değerlendirmesi için bir temel oluştururken, kombine tensör, hastaya özgü bilgilerin tutarlılığını denetler. Bu sayede, Discriminator, üretilen görüntülerin hem görsel kalitesini hem de ilgili hasta bilgilerine uygunluğunu değerlendirebilir. Kombine tensör, hem Generator'e hem de Discriminator'a koşullandırma bilgisi sağlayarak, sentezlenen görüntülerin kişiselleştirilmiş olmasını ve hedeflenen özellikleri yansıtmasını sağlar. Önerilen mimarinin, yüksek kaliteli ve hastaya özgü kanser görüntüleri üretme potansiyeli, kanser araştırmaları ve klinik

uygulamalar için önemli avantajlar sunar. Bu model, hastalıkların teşhis ve tedavisi, tedavi planlaması ve eğitim materyallerinin geliştirilmesi gibi çeşitli alanlarda kullanılabilir. Ayrıca, sentetik veri üretme yeteneği, veri kıtlığı sorununu hafifleterek, daha sağlam ve genelleştirilebilir derin öğrenme modellerinin eğitilmesine olanak tanır. Geliştirilen GAN tabanlı modelin genel yapısı Şekil 5.2 'de sunulmuştur. Bu çalışma, kanser araştırmalarında yeni ufuklar açarak, daha etkili ve kişiselleştirilmiş tedavi stratejilerinin geliştirilmesine katkı sağlayabilir.

5.3 Model Eğitimi

5.3.1 Swin Transformer ile Özellik Çıkarımı

Swin Transformer, hem yerel hem de global görsel bilgileri etkili bir şekilde yakalayabilmesi nedeniyle modelde birincil özellik çıkarıcı olarak seçilmiştir. Özellik çıkarma süreci, giriş görüntüsünün çakışmayan pencerelere bölünmesiyle başlar ve dikkat hesaplamaları her bir pencere içinde gerçekleştirilir. Hiyerarşik yapı, modelin büyük giriş boyutlarını işleyebilmesini sağlarken hesaplama verimliliğini korur. Kaydırılmış pencere mekanizması, her bloktan sonra pencerelerin kaydırılmasını sağlayarak, görüntü genelinde daha kapsamlı özellik çıkarımı yapar.

Swin Transformer mimarisi ayrıca rezidüel sonrası normalizasyon içerir; bu, her rezidüel bloğun çıktılarının ana dal ile birleştirilmeden önce normalize edilmesini sağlar. Bu yaklaşım, daha derin katmanlarda yüksek aktivasyon genliklerinin birikmesini önleyerek daha kararlı özellik çıkarımı sağlar. Ayrıca, dikkat hesaplamaları geleneksel nokta çarpımı yerine, ölçeklenmiş kosinüs benzerliği fonksiyonu kullanılarak yapılır. Bu, aşırı dikkat değerleriyle ilgili sorunları azaltır ve daha hassas özellik çıkarımını destekler.

Giriş çözünürlüğündeki değişiklikleri ele almak için, model log-aralıklı sürekli pozisyon önyargısı kullanır. Bu mekanizma, modelin farklı pencere boyutlarına aktarılması durumunda konum kodlamalarının sorunsuz şekilde interpolasyonunu sağlar. Bu özellik, Swin Transformer'ın farklı görüntü ölçeklerine etkili bir şekilde adapte olmasına olanak tanır.

5.3.2 Eğitim Süreci

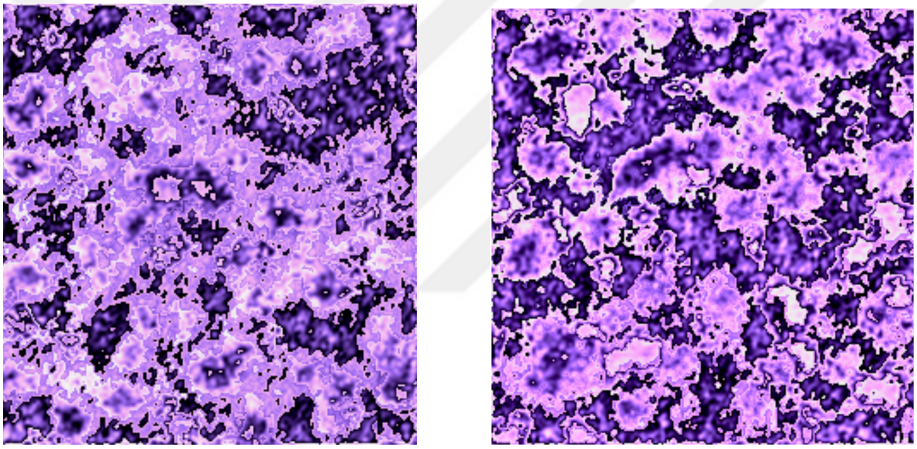
Modelin eğitim süreci, iki ana bileşen üzerine yapılandırılmıştır:

1. Swin Transformer ile özellik çıkarımı,

2. GAN mimarisi kullanılarak görüntü üretimi.

Swin Transformer, histopatolojik görüntülerden hiyerarşik özellikler çıkarırken, GAN modeli gerçek histopatolojik görüntülere oldukça benzeyen sentetik görüntüler üretmek için tasarlanmıştır.

Her eğitim iterasyonunda, bir gerçek görüntü grubu Swin Transformer ile işlenerek özellik haritaları çıkarılmış ve bu haritalar çok-omik destek verileriyle birleştirilmiştir. Bu özellikler ve destek verileri, hem GAN üreticisi hem de ayrıştırıcısı (discriminator) için giriş verisi olarak kullanılmıştır. Üretici (generator), rastgele gürültü vektörleri ve destek verilerini birleştirerek sentetik görüntüler üretirken, ayrıştırıcı, görüntülerin gerçekliğini değerlendirmiştir. Eğitim sonunda üretilen Şekil 5.3'te gösterildiği gibi 256×256 piksel sentetik histopatolojik görüntüler, modelin gerçekçi görüntüler üretme yeteneğini göstermektedir.



Şekil 5.3 Gan modelinin ürettiği sentetik histopatolojik görüntüler

Kayıp Fonksiyonları ve Optimizasyon

- Ayrıştırıcının (Discriminator) gerçek ve üretilmiş görüntüleri ayırt etme performansını optimize etmek amacıyla Hinge Loss fonksiyonu kullanılmıştır.
- Model parametreleri, optimizasyon sürecinde Adam algoritması ile güncellenmiş ve öğrenme süreci iteratif olarak gerçekleştirilmiştir.
- Üretici ağın (Generator) gradyanları, Hinge Loss'un en aza indirgenmesi esasına dayalı olarak hesaplanmıştır.
- Ayrıştırıcı ağın güncellenmesi ise, gerçek ve üretilmiş görüntüler için Hinge Loss fonksiyonunun minimize edilmesi yoluyla gerçekleştirilmiştir.

Eğitim Kararlılığı ve Düzenlemeler

- Eğitim sürecinde erken durdurma mekanizması uygulanmış olup, belirli bir epoch sonrasında Frechet Başlangıç Mesafesi (FID) metriğinde belirgin bir bozulma gözlemlendiğinde eğitim süreci sonlandırılmıştır.
- Ayrıştırıcının karar sınırlarını daha keskin belirleyebilmesini sağlamak amacıyla, gradient penalty yöntemi kullanılarak interpolasyonlu görüntüler üzerinden düzenleme gerçekleştirilmiştir.
- Modelin parametreleri, yalnızca en düşük FID metriği elde edildiğinde kaydedilmiş ve en iyi modelin korunması sağlanmıştır.

Eğitim süreci, modelin öğrenme kapasitesini en verimli şekilde değerlendirebilmek amacıyla çoklu epoch boyunca yürütülmüştür. Modelin her aşamada en iyi performansı sergileyebilmesini sağlamak adına üretici ve ayrıştırıcı ağların parametreleri, optimizasyon süreçleriyle birlikte belirli aralıklarla kaydedilmiştir. Böylece, eğitim sürecinin gerektiğinde kaldığı yerden devam ettirilebilmesi mümkün hale getirilmiştir. Eğitim döngüsü, belirlenen erken durdurma kriterlerinin karşılanması veya maksimum epoch sayısına ulaşılması durumunda sonlandırılmıştır.

5.3.3 Sınıflandırma

5.3.3.1 Veri Etiketleme ve Evre Belirleme

Ön işleme uygulanmış görüntülerin evre bilgisi, klinik destek verilerindeki hasta kimliklerinin ilgili histopatoloji görüntüleriyle eşleştirilmesi yoluyla belirlendi. Bu adım, her hastanın hastalık evresinin kendi görüntüleriyle ilişkilendirilmesi sayesinde doğru bir etiketleme yapılmasını sağladı. Evreler, sınıflandırma görevlerinde kullanılabilmesi için bir etiket kodlayıcı (label encoder) kullanılarak sayısal değerlere dönüştürüldü. Meme kanseri evrelemesinin karmaşıklığı ve doğru sınıflandırma gerekliliği göz önüne alındığında, bu eşleştirme süreci, eğitim için kullanılan veri kümesinin bütünlüğünü sağlamak açısından kritik öneme sahiptir.

Bu çalışmada, meme kanseri evrelerini sınıflandırmak ve potansiyel yeni evreleri belirlemek için çeşitli derin öğrenme modelleri kullanılmıştır. Kullanılan modeller arasında Swin Transformer, Vision Transformer (ViT), EfficientNet, ResNet ve Inception yer almaktadır. Her model, hem SGD hem de Adam optimizasyon algoritmalarıyla değerlendirilmiş ve doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi geleneksel sınıflandırma metrikleriyle test

edilmiştir. Bu metrikler, her modelin farklı kanser evreleri üzerindeki genelleme yeteneği hakkında kapsamlı bir anlayış sağlamıştır.

Deneyler sırasında Swin Transformer modeli, en iyi performansı göstermiştir. Modelin eğitimi sırasında farklı optimizasyon algoritmaları ve ağ yapıları test edilmiştir.(Tablo 5.1) 'de, bu modellerin performans ölçümleri doğruluk, kesinlik ve F1 skoru gibi metrikler üzerinden karşılaştırılmıştır. Özellikle, SGD optimizasyon algoritmasıyla eğitilen Swin Transformer modeli, %92,24 doğruluk ve %92,28 F1 skoru elde etmiştir. Swin Transformer'ın başarısı, hiyerarşik yapısı ve kaydırılmış pencere mekanizmasına dayanmaktadır. Bu mekanizma, modelin hem yerel hem de global özellikleri etkili bir şekilde öğrenmesini sağlamış, yüksek çözünürlükte optimize edilerek karmaşık görüntü detaylarını işleyebilmesine olanak tanımıştır. Bu da meme kanseri evrelerinin sınıflandırılması gibi karmaşık görsel görevlerde doğruluk oranını artırmıştır.

Ayrıca, eğitim sürecinde katman normalizasyonu kullanılarak model kararlılığı sağlanmıştır. SwinV2'deki rezidüel sonrası normalizasyon (residual post-norm), model derinleştikçe aktivasyonların kontrolsüz bir şekilde büyümesini engellemiş ve kararlı bir eğitim sürecine katkıda bulunarak tutarlı sonuçlar elde edilmesini sağlamıştır.

Ölçeklenmiş Kosinüs Dikkat (Scaled Cosine Attention), orijinal nokta çarpımı dikkat mekanizmasının yerine kullanılarak, büyük modellerde aşırı dikkat değerlerinin oluşmasını önlemiş ve dengeli bir öğrenme sürecini sağlamıştır. Bu özellik, modelin bir görüntüde doğru bölgelere odaklanmasını sağlayarak özellikle sınıflandırma görevleri için büyük önem taşımıştır. Ayrıca, log-aralıklı sürekli pozisyon önyargısı, farklı çözünürlükler arasında daha pürüzsüz geçişlere olanak tanımış ve düşük çözünürlükte önceden eğitilen modellerin yüksek çözünürlükteki görevlerde etkili bir şekilde çalışmasını sağlamıştır. Bu ayarlama, modelin farklı çözünürlükler arasında daha iyi genelleme yapmasını sağlayarak sınıflandırma performansını artırmıştır.

Swin Transformer'dan sonra en başarılı model, yine bir Transformer tabanlı mimari olan Vision Transformer (ViT) olmuştur. Vision Transformer, SGD optimizasyon algoritması ile eğitildiğinde %92,18 doğruluk ve %92,07 F1 skoru elde etmiştir. Bu çalışmada Transformer tabanlı modellerin başarısı, yerel ve global bağlamları aynı anda öğrenebilme yeteneklerinden kaynaklanmaktadır. Bu özellik, çok ölçekli özelliklerin teşhiste önemli bir rol oynadığı tıbbi görüntüleme alanında özellikle faydalıdır (Tablo 5.1).

Diğer modeller arasında EfficientNet, %91,56 doğruluk elde ederek dikkat

çekici bir performans sergilemiştir. Ancak, performansı Swin Transformer ve ViT modellerine kıyasla biraz düşük kalmıştır. EfficientNet, doğruluk ve hesaplama maliyeti arasındaki dengesiyle bilinse de, histopatoloji görüntülerinin karmaşıklığını Transformer tabanlı mimariler kadar etkili bir şekilde işleyememiştir.

ResNet ve Inception modelleri ise bu çalışmada daha düşük doğruluk ve F1 skorları ile zayıf bir performans göstermiştir. Özellikle Inception, sabit boyutlu çekirdeklerine (fixed-size kernels) bağlılığı nedeniyle çok ölçekli özellikleri yakalamakta zorlanmış ve kanser evreleri arasında genelleme yapma konusunda başarısız olmuştur (Tablo 5.1).

Swin Transformer modelinin etkinliğini daha da artırmak için eğitim-test-doğrulama ayrımı yeniden düzenlenmiş ve eğitim verisi miktarı artırılmıştır. Ek eğitim verilerinin dahil edilmesi, modelin genelleme kapasitesini artırarak daha kapsamlı bir öğrenme sürecine olanak tanımıştır. Eğitim örneklerinin çeşitliliği ve sayısının artırılmasıyla, modelin aşırı öğrenme (overfitting) riski azaltılmış ve dayanıklılığı artırılmıştır. Daha büyük bir veri kümesiyle eğitilen model, önceki test sonuçlarına kıyasla doğruluk ve F1 skorlarında önemli iyileşmeler göstermiştir.

Bu geliştirme sonucunda, Swin Transformer modeli %95,03 doğruluk, %95,00 F1 skoru, %95,03 duyarlılık (recall) ve %95,00 kesinlik (precision) değerlerine ulaşmıştır (Tablo 6.1). Bu iyileşme, sınıflandırma performansının yalnızca daha fazla veriyle eğitilerek değil, aynı zamanda veri kümesinin stratejik bir şekilde yapılandırılarak hastalığın tüm evreleri arasında dengeli bir temsil sağlanmasıyla optimize edilebileceğini göstermektedir.

Ayrıca, veri artırımı (data augmentation) ve erken durdurma (early stopping) gibi tekniklerin uygulanması, modelin performansına önemli katkılar sağlamıştır. Rastgele yatay ve dikey çevirme, döndürme ve yeniden boyutlandırma gibi veri artırımı teknikleri, eğitim setinin çeşitliliğini artırarak modelin histopatoloji görüntülerindeki varyasyonlara karşı daha dayanıklı hale gelmesini sağlamıştır. Bu teknikler, aşırı öğrenmeyi (overfitting) önlemek ve genelleme yeteneğini artırmak açısından kritik bir rol oynamıştır. Erken durdurma ise modelin aşırı eğitim yapmasını engelleyerek, görünmeyen verilere daha iyi genelleme yapmasını sağlamıştır.

Swin Transformer modelinin başarısı, aynı zamanda eğitim sürecinde StyleGAN3 tarafından üretilen sentetik görüntülerin değerlendirilmesine de olanak tanımıştır. Eğitim sürecinin ardından, StyleGAN3 tarafından üretilen sentetik histopatoloji

görüntüleri, eğitilmiş Swin Transformer modeline sınıflandırma için sunulmuştur. Bu süreç, modelin mevcut meme kanseri evrelerini doğru bir şekilde kategorize etmesini ve potansiyel yeni evre adaylarını belirlemesini sağlamıştır. Modelin güvenilir bir şekilde sınıflandıramadığı ve mevcut evrelerden farklılık gösteren örnekler, yeni evreler olarak değerlendirilmiştir. Bu, kanser evrelerinin daha detaylı bir şekilde anlaşılmasına katkıda bulunmuştur.

Sonuç olarak, Swin Transformer modeli, bu çalışmada meme kanseri evrelerini sınıflandırmada en etkili model olmuştur. Sentetik görüntülerin sınıflandırma sürecine dahil edilmesi ve eğitim verisinin genişletilmesi, model performansını daha da artırmış ve potansiyel yeni evrelerin tanımlanmasını kolaylaştırmıştır. Bu yöntem, mevcut evreleme sistemlerinin ötesine geçerek, kanser evrelerinin daha kapsamlı ve detaylı bir şekilde analiz edilmesini mümkün kılabilir.

Tablo 5.1 Farklı modellerin ve optimizasyon algoritmalarının performans metrikleri

Model	Optimizatör	Doğruluk	Kesinlik	Duyarlılık	F1 Skoru	Öğrenme Oranı
Swin	**Adam**	**0.1415**	**0.0200**	**0.1415**	**0.035**	**0.001**
Swin	**SGD**	**0.9224**	**0.9238**	**0.9224**	**0.9228**	**0.001**
ResNet	Adam	0.7986	0.7996	0.7986	0.7963	0.001
ResNet	SGD	0.5463	0.5640	0.5463	0.5371	0.001
EfficientNet	Adam	0.8633	0.8621	0.8633	0.8610	0.001
EfficientNet	SGD	0.9156	0.9149	0.9156	0.9151	0.001
ViT	Adam	0.5014	0.5192	0.5014	0.4737	0.001
ViT	SGD	0.9218	0.9208	0.9218	0.9207	0.001
Inception	Adam	0.7054	0.7347	0.7054	0.6967	0.001
Inception	SGD	0.9190	0.9181	0.9190	0.9179	0.001
Swin	Adam	0.1388	0.0193	0.1388	0.0338	0.0005
Swin	SGD	0.9163	0.9201	0.9163	0.9164	0.0005
ResNet	Adam	0.8633	0.8694	0.8633	0.8625	0.0005
ResNet	SGD	0.4946	0.4777	0.4946	0.4734	0.0005
EfficientNet	Adam	0.9184	0.9200	0.9184	0.9171	0.0005
EfficientNet	SGD	0.8844	0.8834	0.8844	0.8824	0.0005
ViT	Adam	0.6469	0.6618	0.6469	0.6394	0.0005
ViT	SGD	0.9170	0.9205	0.9170	0.9170	0.0005
Inception	Adam	0.8184	0.8199	0.8184	0.8139	0.0005
Inception	SGD	0.8837	0.8845	0.8837	0.8804	0.0005

5.3.4 Çoklu Modlu Büyük Dil Modeli ile Sınıflandırıcı

Çalışmanın bu kısmı, büyük dil modellerinin görsel-metin tabanlı öğrenme kapasitesini kullanarak histopatolojik görüntülerden kanser evresi tahmini yapmayı ve yeni evrelerin keşfini sağlamayı amaçlamaktadır. Bu doğrultuda, Qwen2-VL

modelinin ince ayar (fine-tuning) yöntemi ile histopatolojik görüntüler üzerinde eğitilmesi sağlanmış ve modelin, görsellerin içerdiği semantik bilgiyi kanser evresi etiketleri ile ilişkilendirebilmesi hedeflenmiştir. Çalışmanın metodolojisi, veri ön işleme, model mimarisinin uyarlanması, eğitim stratejileri ve Out-of-Distribution (OOD) analizi olmak üzere dört ana aşamadan oluşmaktadır.

Veri kümesi, CPTAC-BRCA koleksiyonundan elde edilen histopatolojik görüntüler ve bu görüntülere ait kanser evresi bilgisinden oluşmaktadır. Görüntülerin, belirli bir kanser evresi etiketi ile eşleştirilmesi ve modelin eğitimi için uygun bir formatta düzenlenmesi sağlanmıştır. Veri kümesi %80 eğitim, %10 doğrulama ve %10 test olacak şekilde bölünerek modelin performansını değerlendirmek için ayrılmıştır. Veri ön işleme aşamasında, görüntüler normalize edilmiş, boyutlandırılmış ve etiketleme süreci JSONL formatında gerçekleştirilmiştir. Modelin yalnızca görüntüleri anlamlandırması değil, aynı zamanda doğal dil işleme (NLP) bağlamında bir yanıt üretebilmesi hedeflenerek her veri örneği, kullanıcı-asistan diyaloğu şeklinde düzenlenmiştir.

Çalışmada kullanılan model Qwen2-VL-7B-Instruct modelidir. Bu modelin verimli şekilde eğitilebilmesi için 4-bit quantization uygulanmış, bellek kullanımı optimize edilerek daha büyük veri kümeleri ile çalışılmasına olanak sağlanmıştır. Ayrıca, LoRA (Low-Rank Adaptation) tabanlı Parameter Efficient Fine-Tuning (PEFT) yöntemi kullanılarak modelin yalnızca belirli bölümlerinin güncellenmesi sağlanmış ve gereksiz hesaplama yükü azaltılmıştır. Özellikle, vision ve dil katmanlarının birlikte optimize edilmesi sağlanarak modelin görselleri metinsel bilgilere dönüştürme yeteneği güçlendirilmiştir.

Eğitim sürecinde, modelin parametrelerini en verimli şekilde ayarlamak amacıyla modelin doğruluğu, belirlenen %10 doğrulama veri kümesi üzerinden izlenmiştir. LoRA destekli PEFT stratejisi, modelin daha düşük kaynaklarla daha yüksek doğrulukta öğrenmesini sağlamış, böylece kanser evrelerini sınıflandırma yeteneği artırılmıştır.

Eğitim sonucunda değerlendirilme metrikleri ile modelin başarısı test verisi üzerinde değerlendirildi. Model performans metrikleri ile model başarısının yüksek olduğu gözlendi. Out-of-Distribution (OOD) analizi ile modelin bilinmeyen kanser evrelerini tespit edebilme kapasitesini ölçmekdü. OOD analizi, modelin tahmin ettiği kanser evresi etiketlerinin toplam olasılıklarını değerlendirerek, modelin bilinen evrelerden birine mi ait olduğu yoksa yeni bir evre olasılığı taşıyıp taşımadığı konusunda karar vermesine yardımcı olmaktadır.

Bu süreçte, model tarafından üretilen çıktılardan logits değerleri analiz edilmiştir.

Logits hesaplaması, modelin her etiket için tahmin ettiği olasılıkları içermekte olup, bu değerlerin softmax dönüşümü ile normalize edilmesi sonucunda her kanser evresi için güven skoru elde edilmiştir. Model çıktısında tahmin edilen kanser evresinin toplam olasılığı belirlenen bir eşik değeriyle karşılaştırılmış ve eğer olasılık belirlenen eşik değerinin altında kalıyorsa, model bu görüntüyü Out-of-Distribution olarak işaretlemiştir. Bu durum, modelin bu görüntünün daha önce tanımlanan evrelerden farklı olabileceği yönünde bir çıkarım yapmasına olanak sağlamıştır. OOD analizinde her kanser evresi etiketi için toplam logits değerleri hesaplanmış, modelin en yüksek olasılık atadığı etiketin güven skoru incelenmiştir. Tüm evreler için hesaplanan olasılıkların çarpılmasıyla etiketin genel güven değeri belirlenmiştir. Eğer hiçbir kanser evresi etiketi belirlenen güven eşiğinin üzerine çıkamıyorsa, model bu görüntünün daha önce sınıflandırılmamış bir evreye ait olabileceği sonucunu çıkarmaktadır. Bu tür bir analiz, bilinmeyen veya yeni kanser evrelerinin keşfi açısından klinik araştırmalar için önemli bir adım teşkil etmektedir ve biyomedikal karar destek sistemlerinde kullanılabilecek değerli bir araç sunmaktadır.

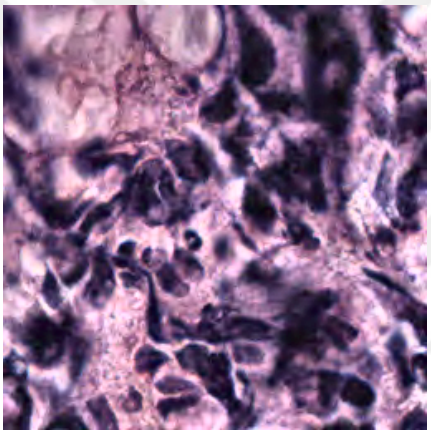
Bu metodoloji, görsel ve metinsel bilgileri birlikte işleyerek kanser evrelerini anlamlandırmaya ve yeni evreleri keşfetmeye yönelik yenilikçi bir yaklaşım sunmaktadır. Modelin yalnızca mevcut evreleri doğru bir şekilde sınıflandırmasını değil, aynı zamanda bilinmeyen evreleri de tespit edebilmesini sağlamak çalışmanın temel katkılarından biri olmuştur. Kullanılan LoRA tabanlı fine-tuning yöntemi, yüksek verimli eğitim süreçleri ile modelin performansını artırmış ve OOD analizi sayesinde bilinmeyen evrelerin keşfine yönelik önemli bir aşama kaydedilmiştir. Bu yöntem, derin öğrenme modellerinin tıbbi görüntü analizi alanındaki potansiyelini ortaya koyarak, gelecekte biyomedikal veri analizi ve klinik karar destek sistemleri için önemli bir temel oluşturacaktır.

Tablo 5.2 Çoklu Modlu Büyük Dil Modeli sınıflandırıcısı performansı

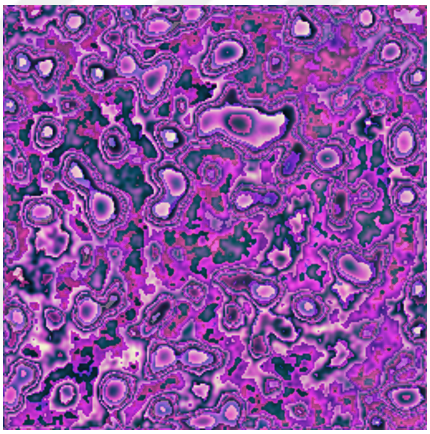
Metrik	**Değer**
Accuracy (Doğruluk)	0.9795
Kesinlik (Precision)	0.9632
Duyarlılık (Recall)	0.9652
F1 Skoru	0.9622

Tablo 5.3 Çoklu Modlu Büyük Dil Modeli sınıflandırıcısı evre bazlı performansı

Evre	Kesinlik	Duyarlılık	F1 Skoru	Örnek Sayısı
Stage IA	0.8778	0.9195	0.8982	19
Stage IIA	1.0000	1.0000	1.0000	428
Stage IIB	0.9652	0.9652	0.9652	152
Stage III	0.9096	0.9446	0.9268	30
Stage IIIA	0.9063	0.9063	0.9063	215
Stage IIIB	0.9621	0.9621	0.9621	49
Stage IIIC	0.9467	0.9467	0.9467	34



Şekil 5.4 Veri kümesinden elde edilen histopatolojik görüntü



Şekil 5.5 Gan modelinin ürettiği sentetik histopatolojik görüntü

Tablo 5.4 Şekil 5.4'te gösterilen histopatolojik görüntü için evre bazlı toplam logits ve olasılık değerleri

Evre	Logits	Güven Skoru
Stage IA	65.0000	0.0000
Stage IIA	109.0000	0.0599
Stage IIB	111.7500	0.9369
Stage III	68.8750	0.0022
Stage IIIA	94.6250	0.0000
Stage IIIB	102.5000	0.0001
Stage IIIC	93.2500	0.0000

Tablo 5.5 Şekil 5.5'te gösterilen histopatolojik görüntü için evre bazlı toplam logits ve olasılık değerleri

Evre	**Logits**	**Güven Skoru**
Stage IA	65.0000	0.0003
Stage IIA	108.7500	0.4530
Stage IIB	107.5000	0.1298
Stage III	72.0000	0.3537
Stage IIIA	95.5000	0.0000
Stage IIIB	105.2500	0.0137
Stage IIIC	96.8750	0.0000

6
SONUÇ

6.1 DEĞERLENDİRME

Bu çalışmada, meme kanseri evrelerini keşfetmek için StyleGAN3 ve Swin Transformer'ı entegre eden önerilen koşullu GAN modelimizin değerlendirilmesi, nitel ve nicel metriklerin bir kombinasyonu kullanılarak gerçekleştirilmiştir. Bu metrikler arasında, sentetik görüntülerin kalitesini ölçmek için Frechet Başlangıç Mesafesi (FID) ve Swin Transformer modelinin performansını değerlendirmek için doğruluk, kesinlik, duyarlılık ve F1 skoru gibi standart sınıflandırma metrikleri yer almıştır. Ayrıca, çıkarım sırasında eşik tabanlı softmax olasılık analizi kullanılarak yeni kanser evrelerinin keşfedilme potansiyeli araştırılmıştır.

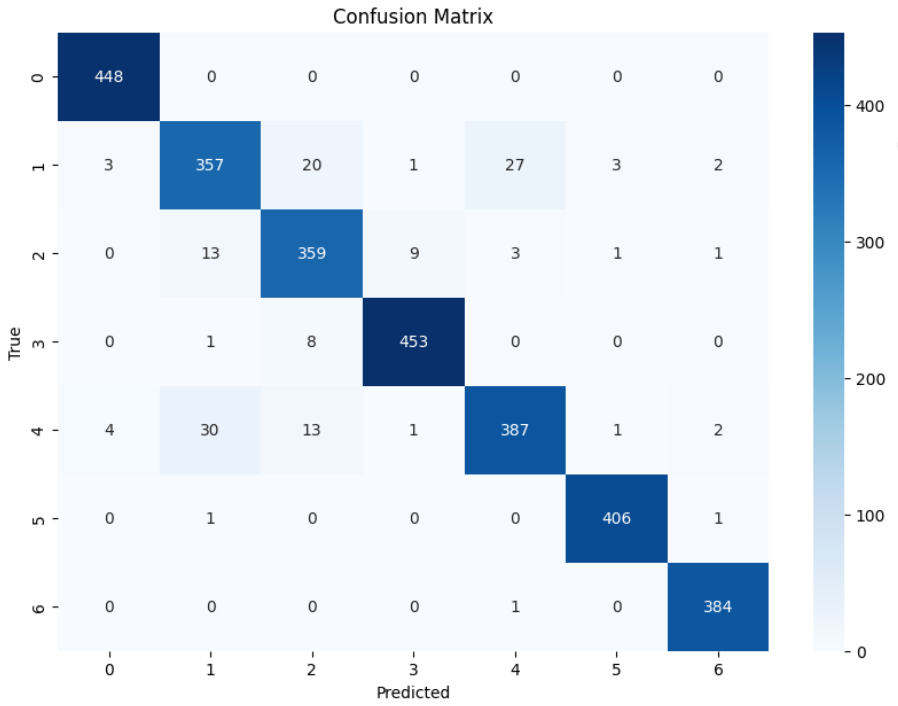
6.1.1 Frechet Başlangıç Mesafesi (FID)

FID metriği, StyleGAN3 tarafından üretilen sentetik histopatoloji görüntülerinin görsel kalitesini ve gerçek görüntülerle olan benzerliğini değerlendirmek için kullanılmıştır. Daha düşük bir FID skoru, üretilen görüntüler ile gerçek görüntüler arasındaki daha yüksek bir benzerliği ifade eder. Üretilen görüntüler için FID skoru 35 olarak hesaplanmıştır, bu da sentetik veriler ile gerçek veriler arasında makul bir görsel benzerlik olduğunu göstermektedir. Ancak, modelin daha ince detayları yakalamada iyileştirme potansiyeli bulunduğu da bu sonuçtan anlaşılmaktadır.

Swin Transformer'ın aynı veri kümesiyle ince ayar yapılmış olması göz önüne alındığında, bu FID skoru, sentetik görüntülerin etkili bir sınıflandırma için yeterince gerçekçi olduğunu ortaya koymaktadır. FID değeri 35, modelin biyolojik açıdan anlamlı görüntüler üretebilme kapasitesini desteklemekte ve bu görüntülerin Swin Transformer tarafından daha ileri analizler için kullanılabileceğini göstermektedir.

Tablo 6.1 Swin transformer sınıflandırma performansı

Metrik	**Değer**
Doğruluk (Accuracy)	0.9503
Kesinlik (Precision)	0.9500
Duyarlılık (Recall)	0.9503
F1 Skoru	0.9500



Şekil 6.1 Güncellenen veri kümesindeki Swin transformatör sınıflandırıcısı için karışıklık matrisi

6.1.2 Swin Transformer Sınıflandırma Performansı

Gerçek histopatolojik görüntüler üzerinde ince ayar yapılan Swin Transformer modeli, hem gerçek hem de sentetik veriler üzerinde test edilmiştir.

Tablo 6.1' de yer alan performans metrikleri değerleri, modelin görüntülerden karmaşık özellikleri yakalayabilme ve farklı meme kanseri evrelerini doğru bir şekilde sınıflandırabilme yeteneğini ortaya koymaktadır. Tablo 6.1"de gösterildiği gibi, Swin Transformer sınıflandırıcısı %95 doğruluk ve %95 F1 skoru ile başarılı bir performans sergilemiştir. Bu metriklerin hem gerçek hem de sentetik görüntülerde tutarlı olması, GAN tarafından üretilen verilerin etkinliğini ve modelin genel başarısını doğrulamaktadır.

6.1.2.1 Karmaşıklık Matrisi Analizi

Model, çoğu evrede güçlü bir sınıflandırma performansı sergilemiş, ancak benzer görsel özelliklere sahip evreler arasında (örneğin, Evre IIA ve Evre IIB ya da Evre IIIB ve Evre IIIC) küçük yanlış sınıflandırmalar yapmıştır. Bu yanlış sınıflandırmalar, büyük olasılıkla bitişik evreler arasındaki biyolojik benzerliklerden kaynaklanmaktadır. Ek omik verilerin (örneğin, RNA, miRNA) modele dahil edilmesi, bu zorlu durumlar arasında daha etkili bir ayrım yapılmasına yardımcı olabilir. Modelin sınıflandırma başarısını ölçmek amacıyla karışıklık matrisi oluşturulmuştur. Şekil 6.1'de Swin Transformer sınıflandırıcısının ürettiği tahminler ile gerçek etiketlerin karşılaştırılması gösterilmektedir.

6.1.2.2 Yeni Evre Keşfi için Softmax Olasılık Analizi

Yeni kanser evrelerinin keşfedilme olasılığını araştırmak için çıkarım sırasında eşik tabanlı bir softmax olasılık analizi uygulanmıştır. Swin Transformer modelinin softmax çıktısı, bilinen sınıflar (kanser evreleri) arasında bir olasılık dağılımı sağlar. Deneysel gözlemlere dayanarak, başlangıçta %30 bir eşik değeri belirlenmiştir: modelin herhangi bir görüntü için güveni (sınıf olasılığı) bu eşik değerinin altında olduğunda, bu görüntü potansiyel olarak daha önce tanımlanmamış yeni bir evre olarak işaretlenmiştir.

Bu eşik, bir başlangıç noktası olarak seçilmiş olup, ek doğrulama ve çapraz doğrulama deneyleriyle daha da optimize edilebilir. Düşük güven değerleri, bazı görüntülerin daha önce görülmemiş veya nadir kanser evrelerini temsil edebileceğini gösterebilir. Bu yaklaşım, yeni evrelerin keşfine katkıda bulunma potansiyeline sahiptir.

6.1.3 Çoklu Modlu Büyük Dil Modeli Sınıflandırma Performansı

(Tablo 5.2)'de yer alan sonuçlar, geliştirilen çok modlu büyük dil modelinin (LLM) sınıflandırma başarımını ayrıntılı bir şekilde ortaya koymaktadır. Bu tabloda doğruluk (accuracy), kesinlik (precision), duyarlılık (recall) ve F1 skoru gibi temel metrikler üzerinden modelin performansı değerlendirilmektedir. Elde edilen yüksek doğruluk oranı, çok modlu yaklaşımın hem görüntü hem de metinsel/biyolojik verileri birlikte kullanarak daha güçlü bir temsil kapasitesine ulaştığını göstermektedir.

Ayrıca, (Tablo 5.4) incelendiğinde, Şekil 5.4'de gösterilen veri kümesinden örneklenen histopatolojik görüntünün gerçek evre bilgisi IIB iken, geliştirilen çok modlu büyük dil modelinin yüksek güven skoru ile IIB evresi için doğru tahminde bulunduğu görülmektedir. (Tablo 5.5) incelendiğinde ise, eğitilen GAN modelinden elde edilen ve Şekil 5.5'de gösterilen histopatolojik görüntünün evre bilgisi IIA olarak tahmin edilmiştir. Bu sonuca göre, geliştirilen çok modlu büyük dil modelinin yeni evre keşfi için belirlenen eşik değerine bağlı olarak kullanılabileceği öngörülmektedir.

(Tablo 5.3)'de yer alan sonuçlar, özellikle nadir görülen örneklerin doğru sınıflandırılmasının, tek modelli yaklaşımlarda sıklıkla karşılaşılan veri dengesizliği problemini bu modelle belirgin biçimde azalttığını göstermektedir.

(Tablo 5.3)'deki sonuçlardan görüldüğü üzere, özellikle erken evre vakalarda modelin duyarlılık ve kesinlik değerleri diğer modellere kıyasla daha yüksek başarı sergilemektedir. Bu durum, erken teşhisin kritik önem taşıdığı onkolojik vakalarda büyük çok modlu modelin güvenilir bir ön tarama veya destekleyici tanı mekanizması olarak kullanılabileceğini göstermektedir. Buna ek olarak, literatürdeki benzer çalışmalarla karşılaştırıldığında, modelin hata payının düşük olması, çoklu veri kaynağını aynı anda işleyebilen büyük dil modellerinin klinik uygulamalarda kullanılabilirliğini artırmaktadır.

6.2 Sınırlamalar ve Gelecek Yönelimler

Model güçlü bir performans sergilemekle birlikte, birkaç sınırlama bulunmaktadır. Mevcut değerlendirme, tümör bölgelerinin segmentasyonu veya mekansal analizini içermemektedir; bu tür analizler, tümör heterojenitesine dair daha derin içgörüler sağlayabilir. Ayrıca, yeni evreleri belirlemek için kullanılan eşik tabanlı softmax analizi hâlâ başlangıç aşamasında olup, klinik açıdan anlamlılık sağlamak için daha fazla doğrulama gerektirmektedir.

Bu çalışmanın bulguları, histopatolojik görüntülerin ve destek verilerin analizinde GAN tabanlı yöntemlerin Swin Transformer ile birlikte kullanılmasının, yeni meme kanseri evrelerini etkili bir şekilde tanımlama potansiyeline sahip olduğunu göstermektedir. Çeşitli geleneksel,transformer modelleri ile yapılan sıflandırma metodu ile yeni evre keşfi tespit etme işlemi desteklenmiştir. Ayrıca multimodal büyük dil modeli ile yapılan çalışma ile yapılan çalışma multimodal büyük dil modelleri ile yapılabilecek ileri seviye analizlere destek olmuştur. Gelecekteki çalışmalarda özellikle, multimodal LLM'ler, histopatolojik görüntüler, destek veriler ve diğer omik verileri analiz ederek yeni evrelerin daha ayrıntılı bir şekilde tanımlanmasına katkı sağlayabilir.

KAYNAKÇA

[1] R. L. Siegel, A. N. Giaquinto A. Jemal, "Cancer statistics, 2024," *CA: a cancer journal for clinicians*, c. 74, no. 1, ss. 12–49, 2024.

[2] R. S. Fitzpatrick, "Using generative models to create synthetic medical imaging data to boost model performance on sparse demographic groups," yük. lis. tezi, University of Hawai'i at Manoa, 2024.

[3] O. Tanimola, O. Shobayo, O. Popoola O. Okoyeigbo, "Breast Cancer Classification Using Fine-Tuned SWIN Transformer Model on Mammographic Images," *Analytics*, c. 3, no. 4, ss. 461–475, 2024.

[4] Z. Liu ve diğ., "Swin transformer v2: Scaling up capacity and resolution," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, ss. 12 009–12 019.

[5] M. Y. Ansari, M. Qaraqe, R. Righetti, E. Serpedin K. Qaraqe, "Unveiling the future of breast cancer assessment: a critical review on generative adversarial networks in elastography ultrasound," *Frontiers in Oncology*, c. 13, s. 1 282 536, 2023.

[6] M. M. Mirimoghaddam ve diğ., "HER2GAN: Overcome the scarcity of HER2 breast cancer dataset based on transfer learning and GAN model," *Clinical Breast Cancer*, c. 24, no. 1, ss. 53–64, 2024.

[7] J. H. L. Goh ve diğ., "Comparative Analysis of Vision Transformers and Conventional Convolutional Neural Networks in Detecting Referable Diabetic Retinopathy," *Ophthalmology Science*, s. 100 552, 2024.

[8] S. Steyaert, Y. L. Qiu, Y. Zheng, P. Mukherjee, H. Vogel O. Gevaert, "Multimodal deep learning to predict prognosis in adult and pediatric brain tumors," *Communications Medicine*, c. 3, no. 1, s. 44, 2023.

[9] J. Ahmad ve diğ., "Deep learning empowered breast cancer diagnosis: Advancements in detection and classification," *Plos one*, c. 19, no. 7, e0304757, 2024.

[10] H. E. Kim, A. Cosa-Linan, N. Santhanam, M. Jannesari, M. E. Maros T. Ganslandt, "Transfer learning for medical image classification: a literature review," *BMC medical imaging*, c. 22, no. 1, s. 69, 2022.

[11] S. Raminedi, S. Shridevi D. Won, "Multi-modal transformer architecture for medical image analysis and automated report generation," *Scientific Reports*, c. 14, no. 1, s. 19 281, 2024.

[12] Y. Ozaki, P. Broughton, H. Abdollahi, H. Valafar A. V. Blenda, "Integrating omics data and AI for cancer diagnosis and prognosis," *Cancers*, c. 16, no. 13, s. 2448, 2024.

[13] R. Kumar, P. Kumbharkar, S. Vanam S. Sharma, "Medical images classification using deep learning: a survey," *Multimedia Tools and Applications*, c. 83, no. 7, ss. 19 683–19 728, 2024.

[14] W. Peebles S. Xie, "Scalable diffusion models with transformers," *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, ss. 4195–4205.

[15] A. Llinas-Bertran, M. Butjosa-Espın, V. Barberi J. A. Seoane, "Multimodal data integration in early-stage breast cancer," *The Breast*, s. 103 892, 2025.

[16] E. Betzalel, C. Penso E. Fetaya, "Evaluation metrics for generative models: An empirical study," *Machine Learning and Knowledge Extraction*, c. 6, no. 3, ss. 1531–1544, 2024.

[17] Y. Ning ve diğ., "Generative artificial intelligence and ethical considerations in health care: a scoping review and ethics checklist," *The Lancet Digital Health*, c. 6, no. 11, e848–e856, 2024.

[18] B. Şener, "Tıbbi görüntülemede derin üretken modeller: bir literatür taraması," *Politeknik Dergisi*, ss. 1–1,

[19] E. Y. Zhang, A. D. Cheok, Z. Pan, J. Cai Y. Yan, "From turing to transformers: A comprehensive review and tutorial on the evolution and applications of generative transformer models," *Sci*, c. 5, no. 4, s. 46, 2023.

[20] S.-C. Huang, A. Pareek, M. Jensen, M. P. Lungren, S. Yeung A. S. Chaudhari, "Self-supervised learning for medical image classification: a systematic review and implementation guidelines," *NPJ Digital Medicine*, c. 6, no. 1, s. 74, 2023.

[21] J. Wang, O. Alamayreh, B. Tondi M. Barni, "Open set classification of gan-based image manipulations via a vit-based hybrid architecture," *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, ss. 953–962.

[22] J. Lipkova ve diğ., "Artificial intelligence for multimodal data integration in oncology," *Cancer cell*, c. 40, no. 10, ss. 1095–1110, 2022.

[23] M. Cen ve diğ., "A Novel and Efficient Digital Pathology Classifier for Predicting Cancer Biomarkers Using Sequencer Architecture," *The American Journal of Pathology*, c. 193, no. 12, ss. 2122–2132, 2023.

[24] C. Matsoukas ve diğ., "Is it time to replace CNNs with transformers for medical images?" *arXiv preprint arXiv:2108.09038*, 2021. erişim adresi: https://arxiv.org/abs/2108.09038.

[25] Q. Liu ve diğ., "Optimizing vision transformers for medical image segmentation," *ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, IEEE, 2023, ss. 1–5.

[26] J. L. Ballard, Z. Wang, W. Li, L. Shen Q. Long, "Deep learning-based approaches for multi-omics data integration and analysis," *BioData Mining*, c. 17, no. 1, s. 38, 2024.

[27] M. L. Abimouloud, K. Bensid, M. Elleuch, M. B. Ammar M. Kherallah, "Advancing breast cancer diagnosis: token vision transformers for faster and accurate classification of histopathology images," *Visual Computing for Industry, Biomedicine, and Art*, c. 8, no. 1, s. 1, 2025.

[28] K. Rautela, D. Kumar V. Kumar, "Improved GAN for image resolution enhancement using ViT for breast cancer detection," *International Journal of Imaging Systems and Technology*, c. 34, no. 2, e22998, 2024.

[29] A. Saad, N. Ghatwary, S. M. Gasser M. S. ElMahallawy, "Automatic image generation and stage prediction of breast cancer immunobiological through a proposed IHC-GAN model," *BMC Medical Imaging*, c. 25, no. 1, s. 6, 2025.

[30] I. Goodfellow ve diğ., "Generative adversarial nets," *Advances in neural information processing systems*, c. 27, 2014.

[31] P. Sharma, M. Kumar, H. K. Sharma S. M. Biju, "Generative adversarial networks (GANs): Introduction, Taxonomy, Variants, Limitations, and Applications," *Multimedia Tools and Applications*, ss. 1–48, 2024.

[32] J. Veiner, "Generative Adversarial Networks Based on a General Parameterized Family of Generator Loss Functions," yük. lis. tezi, 2023.

[33] J. Gui, Z. Sun, Y. Wen, D. Tao J. Ye, "A review on generative adversarial networks: Algorithms, theory, and applications," *IEEE transactions on knowledge and data engineering*, c. 35, no. 4, ss. 3313–3332, 2021.

[34] M. Arjovsky, S. Chintala L. Bottou, "Wasserstein generative adversarial networks," *International conference on machine learning*, PMLR, 2017, ss. 214–223.

[35] T. Chakraborty, U. R. KS, S. M. Naik, M. Panja B. Manvitha, "Ten years of generative adversarial nets (GANs): a survey of the state-of-the-art," *Machine Learning: Science and Technology*, c. 5, no. 1, s. 011 001, 2024.

[36] M. Ali, M. Ali, M. Hussain D. Koundal, "Generative adversarial networks (gans) for medical image processing: Recent advancements," *Archives of Computational Methods in Engineering*, ss. 1–14, 2024.

[37] G. Iglesias, E. Talavera A. Dıaz-Álvarez, "A survey on GANs for computer vision: Recent research, analysis and taxonomy," *Computer Science Review*, c. 48, s. 100 553, 2023.

[38] Y. Skandarani, P.-M. Jodoin A. Lalande, "Gans for medical image synthesis: An empirical study," *Journal of Imaging*, c. 9, no. 3, s. 69, 2023.

[39] S. Gurusubramani B. Latha, "Generative Adversarial Network-a Comparative Analysis," *2022 1st International Conference on Computational Science and Technology (ICCST)*, IEEE, 2022, ss. 523–527.

[40] A. Dash, J. Ye G. Wang, "A review of generative adversarial networks (GANs) and its applications in a wide variety of disciplines: from medical to remote sensing," *IEEE Access*, c. 12, ss. 18 330–18 357, 2023.

[41] V. L. T. De Souza, B. A. D. Marques, H. C. Batagelo J. P. Gois, "A review on generative adversarial networks for image generation," *Computers & Graphics*, c. 114, ss. 13–25, 2023.

[42] A. N. Wu, R. Stouffs F. Biljecki, "Generative Adversarial Networks in the built environment: A comprehensive review of the application of GANs across data types and scales," *Building and Environment*, c. 223, s. 109 477, 2022.

[43] Ü. M. Çağlar, "Improving classification performance of endoscopic images with generative data augmentation," yük. lis. tezi, Middle East Technical University, 2022.

[44] J.-Y. Zhu, T. Park, P. Isola A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," *Proceedings of the IEEE international conference on computer vision*, 2017, ss. 2223–2232.

[45] A. A. Nayak, P. Venugopala B. Ashwini, "A Systematic Review on Generative Adversarial Network (GAN): Challenges and Future Directions," *Archives of Computational Methods in Engineering*, ss. 1–34, 2024.

[46] T. Karras, T. Aila, S. Laine J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation. arXiv 2017," *arXiv preprint arXiv:1710.10196*, ss. 1–26, 2018.

[47] A. Figueira B. Vaz, "Survey on synthetic data generation, evaluation methods and GANs," *Mathematics*, c. 10, no. 15, s. 2733, 2022.

[48] T. Karras, "A Style-Based Generator Architecture for Generative Adversarial Networks," *arXiv preprint arXiv:1812.04948*, 2019.

[49] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen T. Aila, "Analyzing and improving the image quality of stylegan," *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, ss. 8110–8119.

[50] T. Karras, M. Aittala, J. Hellsten, S. Laine, J. Lehtinen T. Aila, "Training generative adversarial networks with limited data," *Advances in neural information processing systems*, c. 33, ss. 12 104–12 114, 2020.

[51] T. Karras ve diğ., "Alias-free generative adversarial networks," *Advances in neural information processing systems*, c. 34, ss. 852–863, 2021.

[52] Z. Wang, Q. She T. E. Ward, "Generative adversarial networks in computer vision: A survey and taxonomy," *ACM Computing Surveys (CSUR)*, c. 54, no. 2, ss. 1–38, 2021.

[53] Y. Yu, W. Zhang Y. Deng, "Frechet inception distance (fid) for evaluating gans," *China University of Mining Technology Beijing Graduate School*, c. 3, 2021.

[54] S. Barratt R. Sharma, "A note on the inception score," *arXiv preprint arXiv:1801.01973*, 2018.

[55] Z. Liu ve diğ., "Swin transformer: Hierarchical vision transformer using shifted windows," *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, ss. 10 012–10 022.

[56] M. A. K. Raiaan ve diğ., "A review on large language models: Architectures, applications, taxonomies, open issues and challenges," *IEEE access*, c. 12, ss. 26 839–26 874, 2024.

[57] D. Huang, C. Yan, Q. Li X. Peng, "From large language models to large multimodal models: A literature review," *Applied Sciences*, c. 14, no. 12, s. 5068, 2024.

[58] J. Wu, W. Gan, Z. Chen, S. Wan S. Y. Philip, "Multimodal large language models: A survey," *2023 IEEE International Conference on Big Data (BigData)*, IEEE, 2023, ss. 2247–2256.

[59] S. N. Wadekar, A. Chaurasia, A. Chadha E. Culurciello, "The evolution of multimodal model architectures," *arXiv preprint arXiv:2405.17927*, 2024.

[60] M. Xu ve diğ., "A survey of resource-efficient llm and multimodal foundation models," *arXiv preprint arXiv:2401.08092*, 2024.

[61] D. P. Clark, M. Gulley, M. Marko, J. Samayoa, R. T. Kelly S. Sinha, *CPTAC Breast Cancer (CPTAC-BRCA) Collection [Data set]*, Online; available: `https://doi.org/10.7937/TCIA.2021.YI7CP4XS`, 2021.

TEZDEN ÜRETİLMİŞ YAYINLAR

Konferans Bildirisi

1. R. Dede ve G. Bilgin, "Deep Learning Based Discovery of New Breast Cancer Stages: StyleGAN3 and Swin Transformer Approach," 2024 Medical Technologies Congress, TIPTEKNO 2024 , Muğla, Türkiye, 2024