



Tıbbi Görüntü Bölütme için MMIAU-Net Mimarisi Önerisi

Proposed Mixed Modified Inception and Self Attention U-Net for Medical Image Segmentation: MMIAU-Net

Müberra Nur AKÇAMAN

Çanakkale Onsekiz Mart Üniversitesi
Bilgisayar Mühendisliği
Çanakkale, Türkiye
mnakcaman@comu.edu.tr
ORCID: 0000-0002-6428-3826

Tolga ENSARİ

Arkansas Tech University
Dept of Engineering and Computing
Sciences
Russellville, USA
tensari@atu.edu
ORCID: 0000-0003-0896-3058

Ahmet SERTBAŞ

İstanbul Üniversitesi-Cerrahpaşa
Bilgisayar Mühendisliği
İstanbul, Türkiye
asertbas@iuc.edu.tr
ORCID: 0000-0001-8166-1211

Öz

Derin öğrenme; başta tıbbi görüntü bölütme olmak üzere birçok alanda başarılı sonuçlar elde eden bir yapay zeka yöntemidir. İnsan sağlığı için hayati önem taşıyan medikal görüntülerde, hassas analiz yapılarak kesin sonuçlara varılması önemlidir. Derin öğrenme yöntemleri yüksek hesaplama karmaşıklığı sayesinde gözden kaçabilecek en küçük hastalık detayını bile yakalayabilmektedir. U-Net derin öğrenme modeli bu alandaki yüksek başarısından dolayı en popüler mimaridir. Ancak bölütmedeki doğruluk değerleri her veri kümesinde farklı sonuçlar verdiği için başarımının iyileştirilmesine her daim ihtiyaç vardır. Kapsamlı karşılaştırma yapabilmek için bu çalışmada, herkesin erişimine açık olan PH², BOWL 2018 ve CVC-ClinicDB tıbbi veri kümeleri kullanılmıştır. U-Net, U-Net++, Attention U-Net, Residual U-Net, Residual Attention U-Net, TransUNet, Swin-Unet ve bu çalışma için modifiye edilen Self Attention U-Net ve MMIAU-Net (Mixed Modified Inception and Self Attention U-Net) ile eğitimler yapılmıştır. Analizler sonucunda önerilen modelin daha az parametre kullanarak daha yüksek başarımlara ulaştığı görülmüştür.

Anahtar sözcükler: Derin öğrenme, U-Net, Bölütme, Tıbbi görüntü işleme, Bilgisayar görü

Abstract

Deep learning is an artificial intelligence method that achieves successful results in many areas, especially medical image segmentation. It is crucial to reach definitive conclusions by performing precise analysis of medical images, which are vital for human health. Thanks to their maximum computational complexity, deep learning methods can capture even the smallest disease detail that may be overlooked. U-Net deep

learning model is the most popular architecture due to its high success in this field. However, since the accuracy values in segmentation give different results in each dataset, there is always a need to improve its performance. In order to make a comprehensive comparison, three independent medical datasets were used in this study: PH², BOWL 2018, CVC-ClinicDB, which are publicly accessible. U-Net, U-Net++, Attention U-Net, Residual U-Net, Residual Attention U-Net, TransUNet, Swin-Unet, and also with Self Attention U-Net and MMIAU-Net (Mixed Modified Inception and Self Attention U-Net), which were modified for this study. After the analysis, it was seen that the proposed model MMIAU-Net reached higher performances by using fewer parameters.

Keywords: Deep learning, U-Net, Segmentation, Medical image processing, Computer vision

1. Giriş

Hastalıkların tespiti için tıbbi cihazlardan elde edilen görüntüler doktorlar, radyologlar gibi uzmanlar tarafından incelenmekte ve onların vereceği kararlara göre tedavi süreci başlamaktadır. Ancak her ne kadar küçük bir ihtimal bile olsa insanlar tarafından verilen kararların yanlış çıkma ihtimali mümkündür. Günümüzde, medikal alanda hastalıkları teşhisleri için kullanılan; bilgisayarlı tomografi (CT), pozitron emisyon tomografisi (PET), ultrason (US), manyetik rezonans görüntüleme (MRI), röntgen, vb. gibi birçok tıbbi görüntüleme yöntemleri vardır [1]. Bir radyolog için bu tür karmaşık taramaları incelemek zaman kaybına sebep olabilmektedir. Bu sebeple derin öğrenme algoritmaları bu karmaşıklığı engeller ve teşhis yöntemlerine kolaylık sağlayarak hız kazandırır [1, 2].

Son yıllarda çokça kullanılan makine öğrenmesi yöntemleri büyük ve karmaşık veri kümeleri için yeterli olmamaktadır. Bu ihtiyaçtan dolayı makine öğrenmesinin bir alt dalı olan derin öğrenme medikal alanda kullanılmaya başlanmıştır. Derin öğrenme algoritmaları sayesinde görüntülerden bölütme

işlemi çoğu zaman başarılı bir şekilde yapılarak; mevcut hastalıkların tedavisi veyahut muhtemel hastalıkların erken teşhisi için kullanılmaktadır [3].

Tıbbi görüntülerde bölütleme çalışmaları 2015 yılında Ronneberger ve ark. tarafından önerilen U-Net mimarisinden sonra oldukça hız kazanmıştır [4]. Bu mimarinin günümüzde her türlü alanda başarılı bölütleme yapmasından dolayı birçok araştırmacı da bu tasarımın üzerinden değişiklikler yaparak yeni modeller üretmektedirler.

U-Net mimarisi bir çeşit evrişimli sinir ağıdır (Convolutional Neural Networks, CNN). Ancak geleneksel CNN'lerden farklı olarak ResNet [5] te kullanılan atlamalı bağlantılar (skip connections) kullanılmaktadır. Bu sayede bölütleme görevlerinin de üstesinden gelmektedir.

U-Net mimarileri atlamalı bağlantılar ile her ne kadar güzel sonuçlara ulaşsa da konvolüsyon işlemleri bazı lokal ve global olan özellikleri başarılı bir şekilde öğrenme yeteneklerini kısıtlamaktadır [6]. Transformatör ağları, veriler üzerindeki global bağımlılığı hesaplayabilmek için özdekkat mekanizmasını kullanmaktadır. Bu sebeple, transformer ağları [7] U-Net mimarisinde kullanılmasıyla birlikte U-Net mimarisi, lokal ve global özelliklerin çıkarılması yeteneğine de sahip olmaktadır. Bu sayede, aynı transformatör yapısı görüntü verileri üzerinde kullanılarak bilgisayarlı görü uygulamalarından olan semantik bölütleme, görüntü sınıflandırma, nesne algılama gibi işlemleri de başarıyla gerçekleştirebilen Vision Transformer (ViT) [8] geliştirildi [9]. ViT, eş zamanlı bir şekilde Multihead Self-Attention (MSA) kullanarak uzun vadeli bağımlılıkların yakalanmasına olanak sağlamaktadır [10].

Birçok görüntü işleme algoritması evrişim işlemlerinde sabit boyutlu filtreler kullanmaktadır. Modellerde kullanılabilecek doğru filtre boyutunu bulmak zaman açısından sıkıntılı olabilir. Sabit boyutlu filtrelerin kullanılması benzer boyutlu görüntüler için uygundur. Ancak analizlerde kullanılan veriler genellikle farklı boyutlardadır. Buna çözüm olarak ya hesaplama karmaşıklığına sahip olan daha derin ağlar kullanılmalıdır ya da farklı boyutlardaki evrişim filtrelerinin paralel olarak kullanılması olarak bilinen inception modülleri kullanılmalıdır [11, 12].

Çalışmanın ana katkıları aşağıdaki gibi özetlenebilir:

- Bu makale en güncel U-net tabanlı mimarilerin karşılaştırmalı analizini sunmayı amaçlamıştır.
- Tıbbi veri kümelerinde yapılan bölütleme başarımı arttırırken, parametre sayısını azaltan yeni bir ağ mimarisi MMIAU-Net önerilmektedir.
- Modelin genel bölütleme başarısını ölçmek amacıyla farklı biyomedikal alanlara ait üç veri kümesi kullanılmıştır.
- U-Net tabanlı mimariler ile karşılaştırıldığında, önerilen MMIAU-Net modelinin iyi bir başarımla sergilediği görülmüş ve görüntü bölütleme için alternatif en iyi yöntem olabileceğini gösterilmiştir.
- Önerilen model; doğruluk, IoU, dice katsayısı, duyarlılık ve özgüllük gibi çeşitli değerlendirme metrikleri ile analiz edilmiştir.

- Bu çalışmanın, tıbbi görüntülerde derin öğrenme yöntemlerinin kullanımını konu alan çalışmalara katkı sağlayacak bir kaynak olması hedeflenmiştir.

Makale dört ana bölümden oluşmaktadır. Çalışmanın ikinci bölümünde, kullanılan veri kümeleri ve modeller ayrıntılı bir şekilde tanıtılmıştır. Üçüncü bölümde elde edilen deneysel bulgular özetlenmiştir. Son olarak dördüncü bölümde ise bu çalışmadan elde edilen sonuçlar tartışılmıştır.

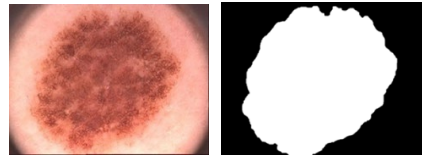
2. Materyal ve Yöntem

2.1 Veri Kümeleri

Bu çalışmada, herkesin erişimine açık olan PH², Bowl ve CVC-ClinicDB olmak üzere üç adet tıbbi veri kümesi kullanılmıştır. Veriler derin öğrenme dünyasında özellikle büyük ses getiren bölütleme mimarileri için tercih edilen veri kümeleri arasından seçilmiştir.

2.1.1 PH²

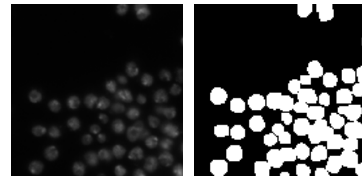
PH² veri seti [13] Portekiz Pedro Hispano Hastanesi tarafından oluşturulan, 200 adet cilt lezyon görüntüsü içeren verilerdir. Görüntüler 8 bit RGB 768x560 piksel çözünürlüğüne sahiptir. Şekil 1'de veri kümesine ait bir örnek ve bu örneğe ait orijinal maske görüntüsü verilmiştir.



Şekil- 1: PH² Orijinal görüntü ve lezyona karşılık gelen maske

2.1.2 Data Science Bowl

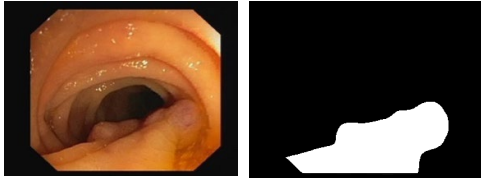
Data Science Bowl 2018 yarışmasında kullanılan hücre çekirdek bölütleme veri kümesi [14] 670 adet görüntüden oluşmaktadır. Çekirdek, tüm yaşam fonksiyonlarını kontrol eden DNA, RNA ve genleri içerir dolayısıyla çekirdekte saptanan herhangi bir bozulma kalp hastalığı, Alzheimer, kanser gibi çeşitli hastalıklara yol açmaktadır [15]. Bu veri kümesine ait bir örnek ve bu örneğe ait maske görüntüsü Şekil 2'de verilmiştir.



Şekil- 2: Bowl 2018 Orijinal görüntü ve ona karşılık gelen maske

2.1.3 CVC-ClinicDB

CVC-ClinicDB veri seti [16] 2015 yılında yapılan polip algılama yarışmasında kullanılan kolonoskopi videolarından elde edilen bir veri setidir. Bu veri kümesi 384x288 boyutlarındaki 612 adet görüntüden oluşmaktadır. Maske görüntüleri, siyah arka plan üzerinde gösterilen beyaz polipler şeklinde oluşturulmuştur. Şekil 3'te veri kümesine ait bir örnek ve bu örneğe ait orijinal maske görüntüsü verilmiştir.



Şekil- 3: CVC-ClinicDB kolonoskopi görüntüsü ve polipe karşılık gelen maske

2.2 Derin Öğrenme Modelleri

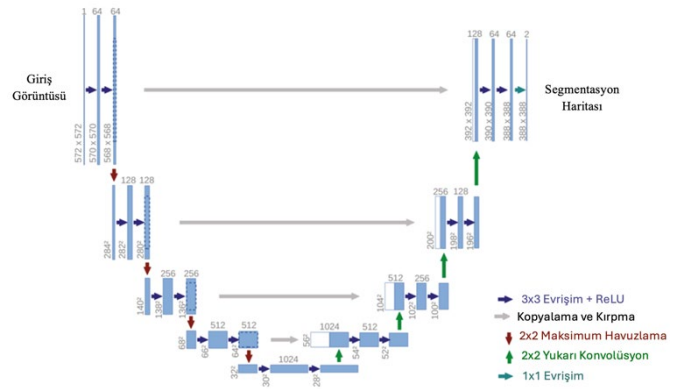
2.2.1 U-Net

U-Net [4], görüntü bölütlemesi için tasarlanmış bir evrişimli sinir ağı mimarisidir. Bu model bölütleme görevlerinde genellikle yüksek başarı ile çalıştığı için son yıllarda biyomedikal alanında sıklıkla kullanılmaktadır.

U-net ağlarında tamamen bağlı katman yoktur. Evrişim katmanları kullanılmaktadır, ayrıca her evrişim katmanında ReLU aktivasyon fonksiyonu kullanılmaktadır. U-net mimarisindeki katmanlar bir bütün olarak düşünülürse U harfini çağrıştırmaktadır. Bu sebeple ismini görüntüsünden almaktadır. Mimaride sol tarafta kodlayıcı, sağ tarafta ise kod çözücü kısmı bulunmaktadır.

Daralan yol diye de bilinen kodlayıcı, normal bir evrişimli ağ mimarisine benzemektedir. Orijinal U-Net mimarisinde ilk adım olarak verilen giriş görüntüsüne 3x3 boyutunda 64 adet filtre ile konvolüsyon yapılmasıyla başlar. Bunun yanında U-Net mimarisinde, kodlayıcı 4 adet bloktan oluşur. Bu bloklar iki tane ardışık 3x3 konvolüsyon katmanı, bir aktivasyon fonksiyonu ve bir 2x2 boyutlu maksimum havuzlama (max pooling) işleminden oluşmaktadır. Her bir adımda öznelik kanallarının sayısı iki katına çıkmaktadır. Kodlayıcının amacı bölütleme yapabilmek için görüntünün içeriğini öğrenebilmektir.

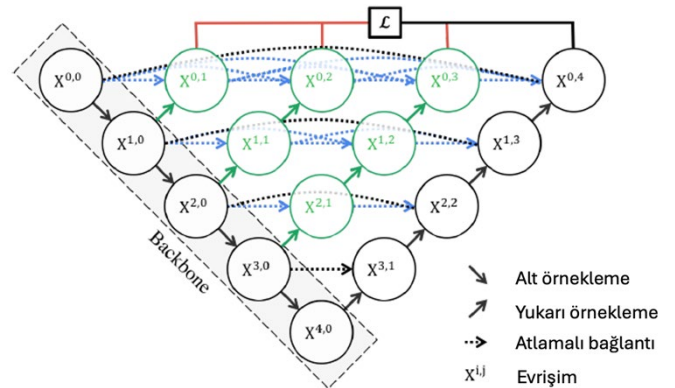
Geniş yol (expansive path) diye bilinen kod çözücü kısmı ise görüntüyü eski boyutlarına getirerek çözünürlüğünde artmasını amaçlamaktadır. Ayrıca atlama bağlantıları sayesinde kodlayıcı kısmından bağlamsal bilgileri de almaktadır. Bu kısım da 4 bloktan oluşmaktadır. Bu bloklar bir 2x2 yukarı konvolüsyon (up-conv) katmanı, öznelik haritası, iki ardışık 3x3 konvolüsyon katmanı ve aktivasyon fonksiyonundan oluşmaktadır. En son katmanda ise, 64 bileşenli özellik vektörünü istenilen sayıdaki sınıfa eşleyebilmek 1x1 konvolüsyon işlemi uygulanmaktadır. Orijinal U-Net mimarisinde toplamda 23 evrişim katmanına sahiptir. Mimari basit bir yapıya sahip olduğu için parametre sayısı azdır ve diğer algoritmalara göre eğitim süresi oldukça kısa sürmektedir. Bu mimarideki en büyük detay katmanlarının hem kendisiyle hem de karşı katmanla evrişim halinde olmasındadır. Mimarideki kullanılan atlamalı bağlantılara, bölütlenmiş görüntülerde çok küçük nesnelere için bile olsa hangi noktalara daha çok odaklanması gerektiği hakkında bilgi verilmektedir. U-Net'e ait mimari Şekil 4'te gösterilmiştir.



Şekil- 4: U-Net Mimarisi [4]

2.2.2 U-Net++

U-Net++ [17], bölütleme doğruluğunu artırmak amacıyla atlamalı bağlantı fikrini bir adım daha ileriye taşıyarak, kodlayıcı ve kod çözücü arasında semantik boşluğu dolduracak şekilde yoğun blok ve evrişim katmanları ekleyen bir derin öğrenme mimarisidir. Bu mimaride omurga kısmı olarak bilinen kodlayıcı kısmı görüntüdeki önemli özellikleri çıkaran yapıyı temsil etmektedir ve genellikle önceden eğitilmiş bir ağdan yararlanır. U-Net++ modelindeki ara katmanlar sayesinde farklı katmanlardan elde edilen özellikler daha detaylı bir şekilde aktarılmış olur ve görüntüdeki en küçük ayrıntılar bile yüksek doğruluk oranıyla elde edilir. Böylece iki katmanların arasındaki anlamsal bilgi kaybı en aza indirgenmiştir. Şekil 5'te bu mimarinin yapısı sunulmuştur.



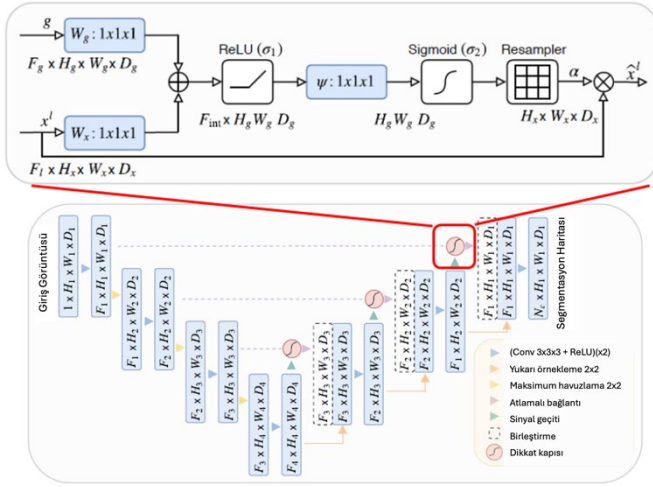
Şekil- 5: U-Net++ Mimarisi [17]

2.2.3 Dikkat U-Net

Dikkat U-Net [18], son yıllarda doğal dil işlemede kullanılan dikkat mekanizmasını atlamalı bağlantılarda kullanılmaktadır. Kod çözücüdeki her bir katman, aynı seviyedeki kodlayıcı kısmından gelen özelliklerin geçmesi gereken bir dikkat kapısına sahiptir. Her katmandan sonra bu dikkat kapılarının sürekli kullanılması, modelde aşırı hesaplama karmaşıklığına sebep olmadan, bölütleme başarımını önemli ölçüde artırır [11].

Dikkat kapılarına atlamalı bağlantı ile gelen x vektörü ile kod çözücünün bir alt katmanından gelen g vektörü ilk önce 1x1 evrişim katmanından geçmektedir. Sonra ise sırasıyla ReLU aktivasyon fonksiyonu, 1x1 evrişim katmanı ve çıktısı 0 ile 1 arasında bir değere sınırlayan sigmoid katmanından

geçmektedir. Elde edilen çıktı vektörü ile kod çözücü kısımdan gelen vektör çarpılarak görüntüdeki hedeflenen bölgelere dikkat edilmesi sağlanır. Mimarinin yapısı Şekil 6'da gösterilmektedir.



Şekil- 6: Dikkat U-Net Mimarisi [18]

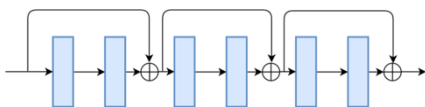
Mimaride belirtilen F, H, W, ve D terimleri özellik haritası boyutlarını belirtmek için kullanılmıştır. F_i özellik haritası kanallarının sayısını ifade etmektedir. Her bir özellik haritası ise, girdideki belirli bir özelliği temsil eden filtre ile ilişkilidir. H_i, W_i ve D_i sırasıyla U-Net mimarisinin i. katmanındaki özellik haritasının yüksekliğini, genişliğini ve derinliğini belirtmektedir. Katmanlarda bu terimlerin çarpılması ile özellik haritasının toplam boyutu hesaplanmaktadır.

Bu mimari kullandığı bu mekanizma sayesinde karmaşık içeriğe sahip görüntülerde bile hedeflenen bölgenin daha hassas bir şekilde elde edilmesini sağlamaktadır.

2.2.4 Artık U-Net

Artık bağlantılı U-Net olarak bilinen bu mimari ResNet mimarisine dayanmaktadır. Adından da anlaşılacağı üzere bu ağ önceki katmanlardan elde edilenlerin bir sonraki katmanlarda veri akışına dahil edilmesiyle oluşmaktadır [19]. Ağdaki her blokta, birinci evrişim katmanının girişi, bir atlama bağlantısı kullanılarak ikinci evrişim katmanının çıkışına eklenir. Bu bağlantılar, U-Net mimarisindeki kodlayıcı ve kod çözücü yollarında alt örnekleme veya üst örnekleme yapılmadan önce uygulanmaktadır [11].

Bir mimariye gerektiğinden fazla katman eklendikçe kaybolan gradyan sorunu (vanishing gradient problem) ortaya çıkabilir. Bu ağ yapısında ise bilgiler doğrudan katmanlara artık bağlantılar sayesinde aktarıldığından sorun çıkması mümkün oldukça önlenmektedir [20]. ResNet bloklarının dizaynı Şekil 7'de sunulmuştur.



Şekil- 7: Artık bağlantılı üç ResNet bloğu[19]

2.2.5 Dikkat Artık U-Net

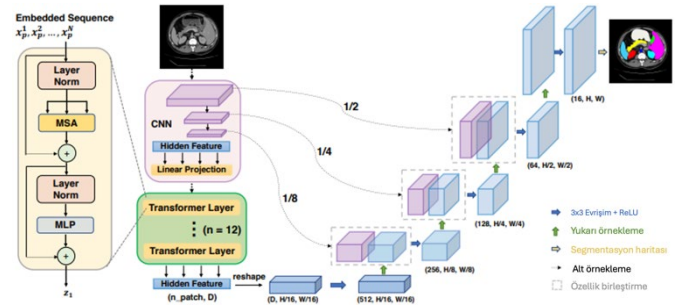
Bu model, yüksek başarımla elde edebilmek için Artık U-Net ile dikkat mekanizmasını beraber kullanmaktadır. Bilindiği üzere, U-Net ağında mekânsal bilgiler kodlayıcıdan kod çözücü kısmına atlama bağlantıları sayesinde geçmektedir. Ancak bu işlem esnasında hatalı/gereksiz bilgilerde karşıya taşınabilir. Buna çözüm olabilmesi adına dikkat mekanizması atlama bağlantılarına eklenmiştir [20]. Ayrıca bu mimariye eklenen residual (artık) bloklar, önceki katmanlardan gelen bilgileri sonraki katmanlara eksiksiz taşıyarak derin ağların bile hatasız ve verimli bir şekilde eğitilmesine olanak tanımış olur.

2.2.6 TransUnet

TransUnet [21], kodlayıcı kısmında transformator kullanan ilk Transformer-CNN melez modelidir [22]. Orijinal TransUnet mimarisi 12 adet transformator katmanından oluşmaktadır. Bir transformator katmanında; normalizasyon katmanı (Layer Norm), MSA (Multiple Sequence Alignment), normalizasyon katmanı ve MLP (Multi Layer Perceptron) bulunmaktadır.

Mimaride, ilk önce yerleşik bir dizi olarak tanımlanan görüntü parçaları belirli bir sıraya göre işlenir ve her birine pozisyon bilgisi eklenerek MSA yapısına gönderilir. MSA mekanizması, bir görüntü verisindeki piksellerin diğer tüm piksellerle olan bağımlılıklarını hesaplayarak global bağlamsal bilgi oluşturur. Bu da uzun vadeli bağımlılıkları daha iyi analiz etmesini ve başarımlarını arttırmasını sağlar [23]. MLP ise çok katmanlı yapısı sayesinde özellik vektörlerini (pikselleri) daha detaylı işlemek için kullanılır. Onlara doğrusal dönüşümler ve ReLU gibi aktivasyon işlemleri uygulayarak piksellerdeki karmaşık bilgilerin öğrenilmesini sağlar.

TransUnet, düşük seviyeli özelliklerin önemine dikkat çektiği için tıbbi görüntüleme uygulamalarında da oldukça tercih edilmekte ve başarılı sonuçlar vermektedir [24]. Ancak, çok sayıda katman ve mekanizma içerdiğinden modelin karmaşıklığı yüksektir ve görüntü bölütlerinin eğitilmesi çok zaman alabilmektedir. TransUnet mimarisi Şekil 8'da gösterildiği gibidir.



Şekil- 8: TransUnet Mimarisi [21]

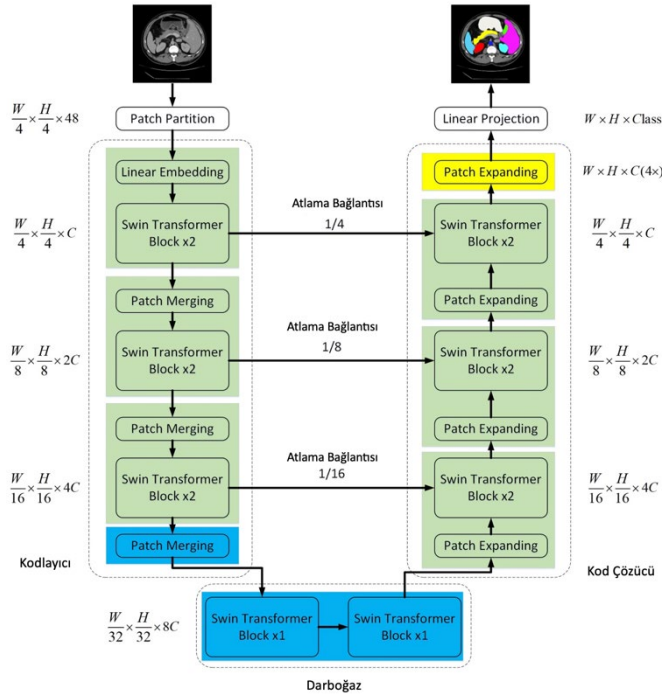
2.2.7 Swin-Unet

Swin-Unet [25] CNN kullanmayan tamamen transformator kullanan ilk modeldir. Swin transformator (Swin Transformer), normal transformatorün tek boyutlu dizisini iki boyutlu görüntü bloklarına genişletir. Mimaride öz dikkat mekanizması ile beraber pencereleri kaydırma mekanizması da kullanılmaktadır. Swin-Unet, dikkat hesaplamalarını sadece mevcut bölgenin etrafındaki küçük alanlarla sınırlayarak çalışır. Bu sayede konumsal bilgiyi daha iyi korur

ve mimarinin verimli bir şekilde çalışmasını sağlayarak başarımını artırır [23]. Model ilk aşamada parça bölme işlemi ile görüntüyü küçük parçalara ayırır ve ardından her bir parçanın özelliklerinin daha yüksek boyutlu bir uzaya yerleştirilmesini yani doğrusal yerleştirmeyi sağlar. Dönüştürülen bu parça birimleri, hiyerarşik özellik temsilleri oluşturmak için birkaç Swin Transformatör bloğundan ve parça birleştirme katmanından geçmektedir.

U-Net mimarisindeki gibi bu tasarımda da kodlayıcı ve çözücü arasında atlama bağlantıları vardır. Bu mimarideki bağlantılar kodlayıcıdaki Swin Transformatör bloğundan kod çözücü Swin Transformatöre gitmektedir. Parça birleştirme katmanının aksine, kod çözücüdeki parça genişletme katmanı özel olarak yukarı örnekleme işlemi gerçekleştirmek için tasarlanmıştır. Son olarak, gömülü olan diziler, farklı boyutlarda ifade edilmek üzere doğrusal projeksiyon işlemi ile dönüştürülür.

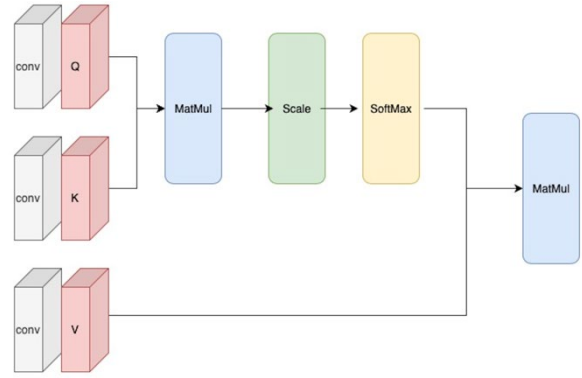
Swin-UNET modeli, tıbbi görüntülerden bölütleme yaparken SOTA (state-of-the-art, en gelişkin modeller) başarım elde etmek için swin transformatörünün güçlü yönlerinden yararlanmaktadır [26]. Swin-UNET mimarisi Şekil 9'da gösterilmiştir.



Şekil- 9: Swin-UNET Mimarisi [25]

2.2.8 Özdiğerat U-Net

Bu mimaride bölütleme için U-Net modeline dayalı bir mimari oluşturulmuş ve veri kümeleri üzerinde denenmiştir. Değıştirilmiş bu özdiğerat U-Net modeli kodlayıcı, kod çözücü ve bir adet özdiğerat modülüne sahiptir. Mimaride, özdiğerat ile özellik haritası çıkartılmaya çalışılmıştır. Özdiğeratta kullanılan sorgu, anahtar ve değeri bir önceki katmandan gelen evrişim işleminin çıktısıdır. Özdiğerat modülünden sonra elde edilen özellikler ise yukarı örnekleme blokları tarafından kullanılarak çalışmanın amacına uygun olarak görüntünün yeniden oluşturulmasını sağlamıştır. Şekil 10'da özdiğerat mimarisi verilmiştir.



Şekil- 10: Self Attention Mimarisi

2.2.9 Önerilen MMIAU-Net (Mixed Modified Inception and Self Attention U-Net) Mimarisi

2014 yılında GoogleNet tarafından duyurulan inception yapısı, ağın genişliğini ve derinliğini artırırken, işlem karmaşıklığını arttırmadığı gözlemlenmiştir [27]. Inception modüllerindeki genişleme işlemi, evrişim katmanlarındaki 1x1, 3x3, 5x5 gibi farklı boyutlardaki filtrelerin ayrıca 3x3 maksimum havuzlama işleminin paralel bir şekilde uygulanmasıyla yapılmaktadır. Burada amaç katman olarak derinleşmektense genişleme fikrine dayanmaktadır.

Bu çalışmada önerilen MMIAU-Net yapısında kullanılan inception ağı modüllerinde ilk önce 1x1'lik evrişim işlemi sonra 3x3 ya da 5x1 evrişim işlemlerini ve bunları takiben 1x3 ve 1x5 evrişim işlemleri ve maksimum havuzlama katmanları kullanılmıştır.

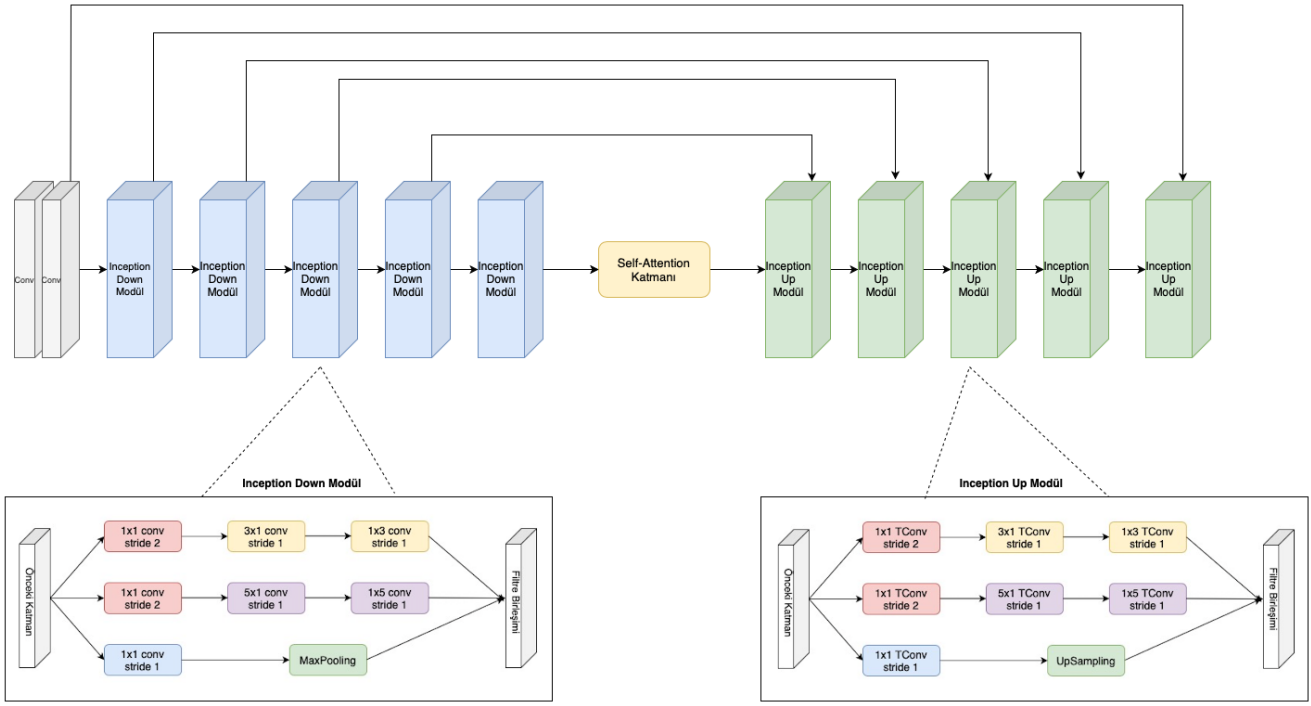
$$Q = XW^Q, K = XW^K, V = XW^V \quad (1)$$

$$\text{softmax}(z)_i = \frac{e^{z_i}}{\sum_{j=1}^N e^{z_j}} \quad (2)$$

$$Y = \text{Self Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \quad (3)$$

Önerilen modelde kodlayıcı ve kod çözücü katmanları arasında Self Attention metodu kullanılmıştır. Self Attention metodunda öncelikle X, bir önceki katmanın çıktı matrisi ve W^Q , W^K ve W^V eğitilebilir ağırlık matrisleri olmak üzere Q Sorgu, K Anahtar ve V Değer katmanları Eşitlik 1'de gösterildiği gibi hesaplanmaktadır. d_k her bir self dikkat katmanının iç boyutu olmak üzere; Q, K, V katmanları ile N elemanlı z vektörü için Eşitlik 2'deki gibi tanımlanan softmax fonksiyonu kullanılarak Eşitlik 3'de ki gibi self attention metodu uygulanmaktadır.

Bu model, diğeri U-Net tabanlı mimarilere göre hem daha iyi sonuç vermiştir hem de daha az parametre gerektirdiği için oldukça hızlı çalışmaktadır. Bu modeli geliştirmeden önce diğeri bölütleme modelleri detaylıca incelenmiş karşılaştırılmış ve self attention ile beraber inception modülünün yapılarının yüksek doğrulukla bölütleme yaptığı gözlemlendiğinden, U-Net mimarisine bu iki yapı eklenmiştir. Oluşturulan bu ağı, inception down modülleri, inception up modülleri ve self attention katmanı olmak üzere üç ana bileşenden meydana gelmektedir. Bu çalışma için oluşturulan ağı yapısı Şekil 11'de verilmiştir.



Şekil- 11: MMIAU-Net Mimarisi

2.3 Değerlendirme Metrikleri

Bu çalışmada, model başarımının değerlendirilebilmesi için doğruluk, birleşim kesişimi (Intersection over Union-IoU), dice benzerlik katsayısı (dice coefficient), duyarlılık (sensitivity) ve özgüllük (specificity) olmak üzere altı adet metrik kullanılmıştır. Bu metrikler dört parametrenin çeşitli kombinasyonlarının bir araya gelmesinden oluşmaktadır; TP, gerçek pozitiflerin sayısını, FP, yanlış pozitiflerin sayısını, FN, yanlış negatifleri ve TN gerçek negatifleri temsil etmektedir. Metriklerin formülleri aşağıya eklenmiştir:

$$\text{Dogruluk} = \frac{TN + TP}{TN + TP + FP + FN} \quad (4)$$

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (5)$$

$$\text{Dice} = \frac{2 * TP}{(2 * TP) + FP + FN} \quad (6)$$

$$\text{Duyarlılık} = \frac{TP}{TP + FN} \quad (7)$$

$$\text{Ozgulluk} = \frac{TN}{TN + FP} \quad (8)$$

Modelin eğitimi esnasında kayıp fonksiyonu olarak dice loss kullanılmıştır. Bu kayıp fonksiyonu, verilen maske ile tahmin edilen maske arasındaki örtüşmeyi hesaplamaktadır.

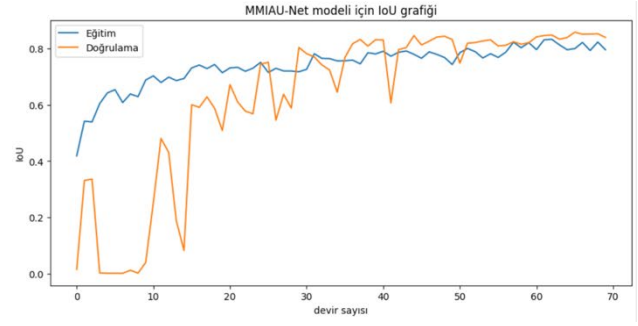
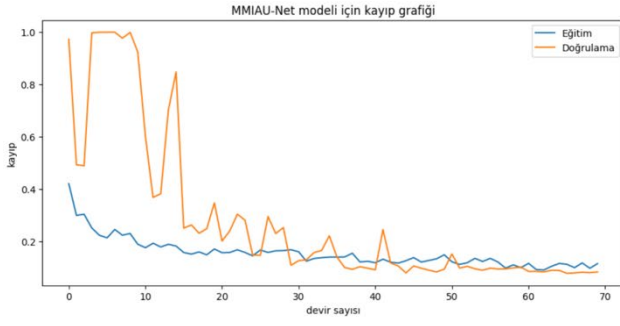
3. Deneysel Bulgular

Deneysel çalışmalar Google Colab Pro platformunda T4 GPU ile Python programlama dili üzerinde Keras kütüphanesi kullanılarak gerçekleştirilmiştir. Transformatör yapıları için Colab yeterli gelmemiş bu yüzden de TÜBİTAK ULAKBİM, Yüksek Başarımlı ve Grid Hesaplama Merkezi'nden (Truba kaynaklarından) faydalanılmıştır.

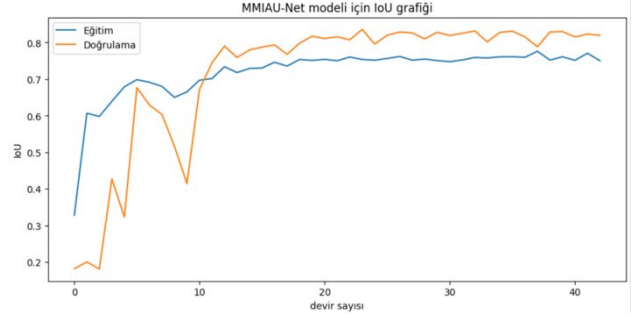
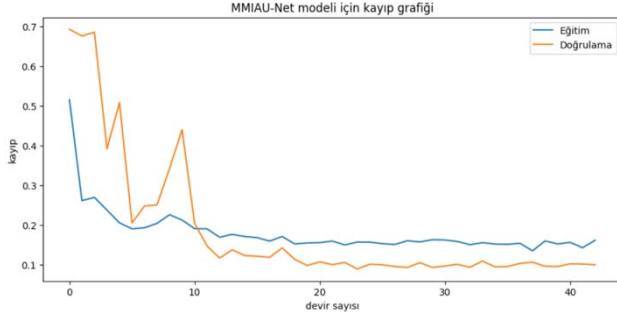
Eğitim veri kümesindeki muhtemel aşırı öğrenme problemi önlemek ve test veri setine daha iyi bir genelleme sağlamak amacıyla rastgele döndürme, çevirme gibi basit veri artırımı (data augmentation) işlemleri uygulanmıştır. Model eğitilmeden önce her bir veri kümesinin boyutları farklıdır. Doğru bir karşılaştırma yapabilmek için tüm görüntüler 256x256 piksel çözünürlüğünde yeniden boyutlandırılmıştır. Oluşturulan mimarilerde, veri kümelerine uygulanan ön işlemlerden sonra görüntüler rastgele %70 eğitim, %20 doğrulama ve %10 test olmak üzere üçe ayrılmıştır.

Hiper parametrelerin seçimi oldukça uzun bir süreç gerektirmektedir. Mevcut çalışmada üç ayrı veri seti ve dokuz farklı mimari kullanıldığından, denenen kombinasyonlar arasından en optimal değerler seçilmiştir. Modellerin eğitim esnasında aynı anda işlediği veri paket sayısını ifade eden batch boyutu 8 olarak belirlenmiştir. Eğitim esnasında tüm algoritmalarda maksimum 150 devir sayısı (epoch) kullanılmıştır. Optimizasyon algoritması olarak gradient tabanlı bir algoritma olan Adam optimizasyon fonksiyonu kullanılmıştır. Transformatör yapısı içeren modeller ile CVC-ClinicDB veri setinde eğitilen modeller için öğrenme oranı 1e-4, diğer tüm mimariler için 1e-2 tercih edilmiştir.

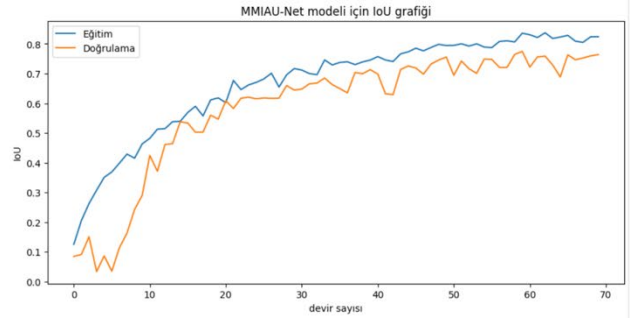
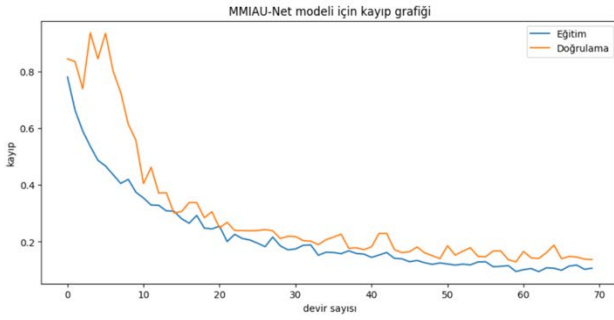
Ayrıca modellerin eğitim sürecinde Keras ait olan EarlyStopping ve ReduceLROnPlateau yöntemleri kullanılmıştır. EarlyStopping, modelin olası bir aşırı öğrenmeye yakalanmasını engelleyen bir yöntemdir. Eğitim süresince modelin başarımını izler, eğer model belirtilen epok boyunca herhangi bir ilerleme göstermiyorsa eğitim durdurulur ve modelin en başarılı sonuçları aldığı andaki değerler kaydedilir. ReduceLROnPlateau ise seçtiğimiz öğrenme oranını dinamik bir şekilde azaltarak daha iyi sonuçlar elde etmeyi amaçlar. Kullanılan modellerde bu fonksiyonda, minimum öğrenme oranı alt sınırı 1e-6 olacak şekilde ayarlanmıştır.



Şekil- 12: MMIAU-Net modelinin PH² verileri için kayıp ve IoU grafiği



Şekil- 13: MMIAU-Net modelinin BOWL 2018 verileri için kayıp ve IoU grafiği



Şekil- 14: MMIAU-Net modelinin CVC-ClinicDB verileri için kayıp ve IoU grafiği

Önerilen modelin, PH² veri kümesinde eğitilmesi sonucu elde edilen hem hata hem de IoU başarımları metrikleri incelendiğinde eğitim ve validasyon sonuçlarının 20. adımdan sonra aynı doğrultuda gittiği gözlemlenmiştir (Şekil 12). Bu da önerilen modelin ezberleme yapmadığını ve eğitim işleminin başarılı olduğunu göstermektedir. Eğitimin 60. ile 70. adımları arasında önerilen modelin hem hatasının azalmaması hem de IoU başarısının artmaması öğrenmenin bittiğini göstermektedir. Bu sebeple eğitim 70. adımda keras çerçevesinin eğitimi erken durdurma betiğince durdurulmuş olup olası bir ezberleme senaryosunun önüne geçilmesi sağlanmıştır.

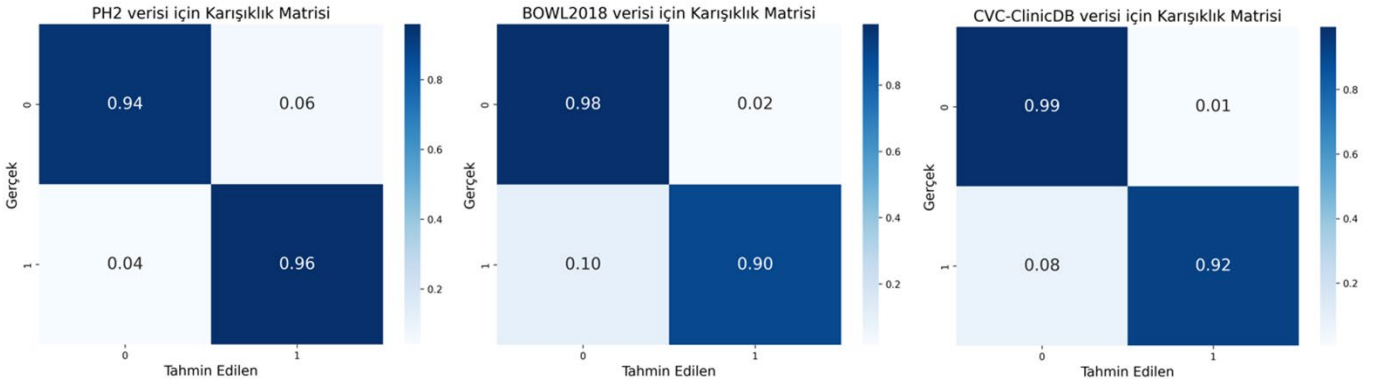
Şekil 13'te Bowl 2018 veri kümesinde eğitilen öneri modelinin kayıp ve IoU grafikleri verilmiştir. Grafiklere göre eğitimin başlangıcından itibaren devir sayısı yükseldikçe IoU oranı artarken kayıp değeri azalmaktadır. Yaklaşık 42. devir sayısında eğitimin tamamlandığı görülmektedir.

CVC-ClinicDB 2018 veri kümesinde eğitilen öneri modelinin kayıp ve IoU grafiği Şekil 14'te verilmiştir. Modelin kayıp grafiğinde hem eğitim hem de doğrulama eğrilerinin sıfıra yaklaşmış olması modelin hem eğitim hem de doğrulama verileri üzerinde yüksek doğrulukla çalıştığını göstermektedir.

Eğitimin 50. devir sayısından sonra model sabit bir şekilde yüzdesini koruyarak çalışmasını bitirmiştir.

Bu grafikler önerilen modelin başarımını analiz edebilmek ve eğitimin ilerlemesini takip edebilmek adına önemlidir. Üç veri kümesinde de eğitilen modelin ezberleme yapmadığı ve genellemeyi başarılı bir şekilde gerçekleştirdiği görülmektedir.

Modellerin başarımının ayrıntılı olarak değerlendirilmesinde karmaşıklık matrisleri de önemli bir rol oynamaktadır. Bu bağlamda, önerilen modelin veri kümeleri ile eğitimi sonucunda elde edilen sonuçlar kullanılarak normalize edilmiş şekilde hesaplanan karışıklık matrisleri Şekil 15'te verilmiştir. X eksenini tahmin edilen sınıfları gösterirken Y eksenini ise gerçek sınıfları göstermektedir. 0 arka plan 1 ise elde edilen bölütleme maskesini ifade etmektedir. Bu veriler doğrultusunda önerilen modelin veri kümeleri ile eğitiminde ezberleme gözükmemekte ve PH², Bowl 2018 ve CVC-ClinicDB veri kümelerinin hepsinde maske değerlerini doğru bir şekilde oluşturabildiği görülmektedir.



Şekil- 15: MMIAU-Net modeli Karışıklık matrisleri

Çizelge-1: PH² veri kümesi için elde edilen sonuçlar

Metot	Doğruluk	IoU	Dice Katsayısı	Duyarlılık	Özgüllük
U-Net	0,9608	0,8875	0,9402	0,9171	0,9849
U-Net++	0,9624	0,8893	0,9414	0,9420	0,9731
Attention U-Net	0,9584	0,8824	0,9374	0,9359	0,9710
Residual U-Net	0,9546	0,8707	0,9308	0,9657	0,9507
Attention Residual U-Net	0,9610	0,8858	0,9394	0,9467	0,9689
TransUNet	0,9325	0,8304	0,9059	0,9102	0,9665
Swin-Unet	0,9452	0,8597	0,9238	0,9247	0,9672
Self Attention U-Net	0,9442	0,8391	0,9123	0,9309	0,9459
MMIAU-Net	0,9664	0,9018	0,9483	0,9444	0,9763

Çizelge-2: BOWL 2018 veri kümesi için elde edilen sonuçlar

Metot	Doğruluk	IoU	Dice Katsayısı	Duyarlılık	Özgüllük
U-Net	0,9573	0,7901	0,8824	0,9013	0,9696
U-Net++	0,9549	0,7653	0,8668	0,8449	0,9793
Attention U-Net	0,9641	0,8190	0,9002	0,9080	0,9764
Residual U-Net	0,9647	0,8184	0,8999	0,8787	0,9835
Attention Residual U-Net	0,9573	0,7844	0,8787	0,8518	0,9808
TransUNet	0,9570	0,8006	0,8850	0,9251	0,9593
Swin-Unet	0,8431	0,6801	0,7758	0,9499	0,7946
Self Attention U-Net	0,9630	0,8165	0,8985	0,9172	0,9728
MMIAU-Net	0,9664	0,8323	0,9082	0,9283	0,9749

Çizelge-3: CVC-CLINICDB veri kümesi için elde edilen sonuçlar

Metot	Doğruluk	IoU	Dice Katsayısı	Duyarlılık	Özgüllük
U-Net	0,9861	0,8505	0,9177	0,9320	0,9912
U-Net++	0,9858	0,8475	0,9161	0,9302	0,9908
Attention U-Net	0,9750	0,7543	0,8579	0,8801	0,9842
Residual U-Net	0,9662	0,6771	0,8060	0,8250	0,9800
Attention Residual U-Net	0,9836	0,8244	0,9026	0,8873	0,9928
TransUNet	0,9637	0,6598	0,7856	0,8592	0,9736
Swin-Unet	0,9445	0,5263	0,6741	0,7312	0,9641
Self Attention U-Net	0,9849	0,8278	0,9053	0,9188	0,9907
MMIAU-Net	0,9874	0,8605	0,9241	0,9322	0,9925

Çalışmada kullanılan üç ayrı veri kümesinde eğitilen tüm modellerin deneysel sonuçları Çizelge 1, Çizelge 2 ve Çizelge 3'te sunulmuştur. Çizelgelerden anlaşılacağı üzere kullanılan

mimariler genellikle birbirlerine yakın ve başarılı bölütlemeler elde etmişlerdir. MMIAU-Net mimarisinde IoU metriği; PH² verileri için **0,9018**, Data Science Bowl 2018 verileri için

0,8323 ve CVC-ClinicDB verileri için ise **0,8605** değeri ile diğer tüm modellere göre daha iyi başarımlar elde etmiştir.

Çizelge 4’te bu çalışma için kullanılan tüm modellerin toplam parametre sayısı gösterilmektedir. Oluşturulan yeni model **8,80 milyon parametre** ile en az işlem yüküne sahiptir. Önerilen modelin daha az parametreye sahip olması eğitim ve test sürelerini kısaltarak modelin verimliliğini arttırmaktadır. Böylece büyük veri kümeleri ve kısıtlı donanım kaynağına sahip olan cihazlarla çalışırken işlem kapasitesi ve bellek kullanımını açısından önemli kazançlar sağlanmış olur. Ayrıca parametre sayısının az olması aşırı öğrenme riskini düşürerek, yeni verilere karşı daha kararlı bir başarımlar göstermesine olanak sağlamaktadır.

Çizelge-4: Metotlar ve Parametre Sayıları

Metot	Parametre Sayısı
U-Net	31,05 M
U-Net++	19,46 M
Attention U-Net	37,33 M
Residual U-Net	33,15 M
Residual Attention U-Net	39,09 M
TransUnet	98,50 M
Swin-Unet	9,36 M
Self Attention U-Net	9,96 M
MMIAU-Net	8,80 M

Şekil 16’da modellerin rastgele seçilen test verileri üzerindeki bölütleme sonuçları gösterilmektedir. Şekillerden de anlaşılacağı üzere çalışılan modeller bölütlenmek istenen bölgeleri yüksek doğruluklarla elde etmektedir.

4. Tartışma

Bu çalışmada önerilen modelin, 2020-2024 yılları arasındaki U-Net temelli en gelişkin modellerle başarı metrikleri açısından karşılaştırması Çizelge 5’te sunulmaktadır. Oluşturulan çizelge bu makale kapsamında kullanılan veri kümelerine göre sıralanmıştır.

kapısı, uzamsal dikkat ve kanal dikkati mekanizmaları ile hedeflenen cilt lezyonlarına daha fazla odaklanabilmeyi ve küçük ayrıntıları daha doğru bir şekilde yakalamayı hedeflemiştir. ISIC-2016, ISIC-2017 ve PH² veri kümelerini kullanarak eğittikleri bu model test sürecinde PH² veri kümesinde %84 Jaccard indeksi ve %91 F1 skorları elde etmiştir.

FAT-Net, Wu ve arkadaşları [29] tarafından otomatik cilt lezyonu bölütleme için geliştirilen bir transformatör ağıdır. Transformatör yapısı self attention mekanizması sayesinde global bağlam bilgilerini etkili bir şekilde analiz eder ve lezyonları daha doğru bir şekilde belirler. Model PH² veri kümesi üzerinde yaptığı testlerinde %89 IoU ve %94 dice katsayısı değerine ulaşmıştır.

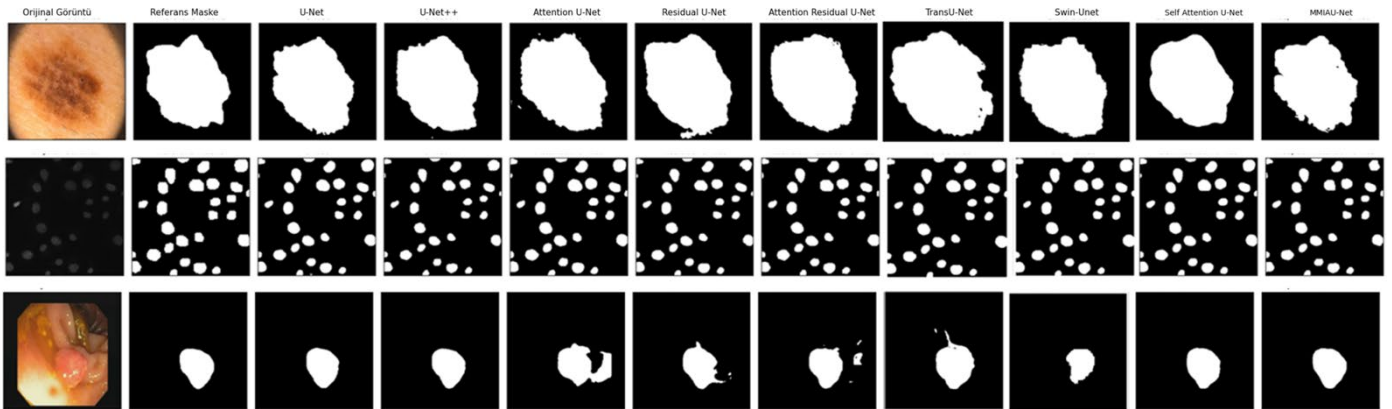
Zhang ve arkadaşları [30] oluşturdukları APT-Net modeli ile, farklı ölçeklerdeki görüntü özelliklerini birleştirmek için adaptif kodlama ve çok ölçekli özellik haritalarını aynı anda işleyebilecek çift yönlü paralel kod çözümü yapılarına sahiptir. Deneylerde model %89 IoU ve %94 dice katsayısı sonuçlarını elde etmiştir.

Hu ve arkadaşları [31] da deri lezyonu bölütleme için Attention Synergy Network (AS-Net) modelini geliştirmiştir. AS-Net hedeflediği kısımların konumsal ve ayırt edici bilgilerini doğru bir şekilde elde etmek için modelin çözümü kısmında uzamsal ve kanal isminde iki farklı dikkat yolu kullanmaktadır. Model %88 Jaccard indeksi ve %93 dice katsayısı sonuçlarına ulaşmıştır.

GSCEU-Net, Hao ve arkadaşları [32] tarafından düşük hesaplama maliyetiyle beraber yüksek bölütleme başarısı sağlayan hafifletilmiş bir model olarak sunulmuştur. PH² test veri kümesiyle %84 IoU ve %91 dice katsayısı elde etmişlerdir.

Olimov ve arkadaşları [33] biyomedikal görüntü bölütleme hızlı bir şekilde yapabilen FU-Net (Fast Unet) modelini sunmuşlardır. Bu model U-Net modelinde elde edilen başarımları arttırabilmek adına darboğaz konvolüsyon katmanları kullanmaktadır. Model eğitim için kullandığı iki veri kümesinden biri olan PH² veri kümesi için %45 IoU ve %90 dice katsayısı elde etmişlerdir.

Roy ve arkadaşları [34] mikroskopi görüntülerinde çekirdek



Şekil- 16: Tahmin edilen maske görüntüleri

Tong ve arkadaşları [28], cilt lezyonlarının bölütleme için ASCU-Net modelini geliştirmişlerdir. Modele ekledikleri dikkat

bölütleme için iki aşamalı bir model olan Nuclei-Net modelini geliştirmiştir. Bölütlemenin ilk aşamasında kolay kısımlar için

Mask Region-Based-CNN (Mask-RCNN), daha karmaşık olan çekirdeklerin üst üste binen kısımları için işaretçi kontrollü suseddi algoritmaları (marker-controlled watershed algorithms) kullanmışlardır. Böylece iki yöntemi de kullanarak bölütlemeyi yüksek doğruluk oranıyla elde etmeyi hedeflemişlerdir. Yapılan deneysel çalışmalarda Nuclei-Net Bowl 2018 test veri kümesinde %88 dice katsayısı elde etmiştir.

OAU-Net (Outlined Attention U-network) modeli, Song ve arkadaşları [35] tarafından bypass dallanma stratejisi (bypass branching strategy) kullanılarak hem yüzeysel hem de detaylı özellikleri algılayabilmek için tasarlanmıştır. Mimari de kodlayıcı olarak artık evrişim (residual convolution) ve res2convolution kullanılmıştır. Ayrıca atlamalı bağlantı kısımlarına da kontur filtresi ve attention modülü eklenmiştir. Modelin başarısı %82 Jaccard indeksi ve %90 dice katsayısı olarak ölçülmüştür.

Ibtehad ve Rahman [36] MultiResUnet modelini, U-Net'in kodlayıcı ile kod çözücü arasında özellik aktarım aşamalarında bazı uyumsuzlukların olduğunu düşündükleri için önermişlerdir. Modele özellik haritalarının daha homojen hale gelebilmesi için 'Res yolları' eklenmiştir. Ayrıca inception bloklarından esinlenerek oluşturdukları 'MultiRes' blokları çok çözünürlüklü analiz yeteneğini arttırmak için kullanılmıştır. Başarı metriği olarak Jaccard indeksi kullanarak %82 elde etmişlerdir.

Double Attention Res-U-Net, Chenarlogh ve arkadaşları [37] tarafından farklı tıbbi görüntüleme sistemlerinde bölütlemeyi kolaylaştırmak amacıyla geliştirilmiştir. Model beş özel bloklu kodlayıcı ve kod çözücü modülleri içeren iki adet ağdan oluşturulmuştur. Bu ağlar, uzamsal bilgileri korumak için modifiye edilmiş Residual yapıları içermektedir. Ayrıca farklı ölçeklerdeki bilgileri yakalayabilmek için dikkat kapıları kullanmışlardır. Modelin başarısı %75 Jaccard indeksi ve %87 dice katsayısı olarak ölçülmüştür.

Cai ve arkadaşları [38] tam konvolüsyonel ağlara dayalı olan çoğu bölütleme yönteminin, uzamsal ve bağlamsal bilgilerin kaybına sebep olduğunu belirtmiş ve bu sorunu çözmek için RAAU-Net modelini tanıtmışlardır. RAAU-Net; ResNet-18 modelini özellik çıkarıcı olarak kullanan bir kodlayıcı, çoklu algılama amacını taşıyan bir modül ve kod çözücü olmak üzere üç ana bileşenden oluşmaktadır. CVC-ClinicDB veri kümesinde %78 IoU ve %87 dice katsayısı elde etmişlerdir.

Jiang ve arkadaşları [39] tarafından geliştirilen LV-UNet modeli (a lightweight and vanilla model) özellikle mobil tıbbi cihazlarda kullanılabilir sade bir mimari sunmayı hedeflemiştir. Bu yüzden de düşük işlem gücüne sahip olması ile bilinen MobileNetV3-Large modelinden yararlanmışlardır. Modelin başarısı %85 IoU ve %91 dice katsayısı olarak ölçülmüştür.

Çizelge 5'teki sonuçlar karşılaştırıldığında; önerilen modelin, literatürdeki diğer modellere göre başarı metriklerinde daha üstün bir başarıyı sergilediği gözlemlenmiştir.

5. Sonuç

Tıbbi görüntü analizinde kullanılan en popüler yöntem derin öğrenme algoritmalarıdır. Ancak, bir derin öğrenme modelinin tüm veri kümeleri için yüksek doğruluk oranına sahip bölütleme yapma garantisi yoktur. Bu yüzden çalışılacak her veri seti için başarılı sonuçlar elde etmek isteniyorsa bu model üzerinde detaylı analizler yapılarak doğru önileme adımları uygulanmalı ve hiper parametreler dikkatlice seçilmelidir.

Son yıllarda başarısından dolayı çokça kullanılan U-Net mimarisi, hedef kısımların bölütlenmesi için kullanılan birçok modelin geliştirilmesinde öncü rolündedir. Bu bağlamda mevcut çalışmada, U-net varyantı mimarilerden bazıları, popüler veri kümelerinden olan PH², BOWL 2018 ve CVC-ClinicDB kullanılarak eğitilmiş ve sonuçlar karşılaştırılmıştır. Ayrıca yapılan analizlerin sonunda biyomedikal görüntüleme bölütlemeye kullanılabilecek yeni bir derin öğrenme yöntemi de önerilmiştir. Bulgular, bu çalışma kapsamında önerilen MMIAU-Net mimarisinin PH² veri kümesi için **%90 IoU, %94 dice benzerlik katsayısı**, Bowl2018 veri kümesi için **%83 IoU, %90 dice benzerlik katsayısı** ve CVC-ClinicDB veri kümesi için **%86 IoU, %92 dice benzerlik katsayısı** ve **8,80M** parametre sayısı ile diğer tüm U-Net tabanlı mimariye göre daha az parametre kullanarak daha iyi bir başarıyı sergilediğini göstermektedir. Önerilen özgün mimari kullanılarak elde edilen sonuçlar ile modelin güvenilirliği ve üstünlüğü ortaya koyulmuş ve gerçek tıbbi problemlere uygun bir yöntem olabileceği gösterilmiştir.

Çizelge-5: Literatürdeki aynı veri kümelerinden birini kullanan en gelişkin modeller ile önerilen model karşılaştırması

Veri Kümesi	Yayın	Yıl	Metot	Başarı Değeri
PH ²	Tong ve ark. [28]	2021	ASCU-Net	%84 Jaccard indeksi %91 F1
	Wu ve ark.[29]	2022	FAT-Net	%89 IoU %94 Dice katsayısı
	Zhang ve ark. [30]	2022	APT-Net	%89 IoU %94 Dice katsayısı
	Hu ve ark. [31]	2022	AS-Net	%88 Jaccard indeksi %93 Dice katsayısı
	Hao ve ark. [32]	2023	GSCEU-Net	%84 IoU %91 Dice katsayısı
	Önerilen Model	2024	MMIAU-Net	%90 IoU %95 Dice katsayısı
Data Science Bowl	Olimov ve ark. [33]	2021	FU-Net	%45 IoU %90 Dice katsayısı
	Roy ve ark. [34]	2023	Nuclei-Net	%88 Dice katsayısı
	Song ve ark. [35]	2023	OAU-net	%82 Jaccard indeksi %90 Dice katsayısı
	Önerilen Model	2024	MMIAU-Net	%83 IoU %91 Dice katsayısı
CVC-ClinicDB	Ibtehas ve Rahman [36]	2020	MultiResUNet	%82 Jaccard indeksi
	Chenarlogh ve ark. [37]	2022	Double Attention Res-U-Net	%75 Jaccard indeksi %83 Dice katsayısı
	Cai ve ark. [38]	2024	RAAU-Net	%78 IoU %87 Dice katsayısı
	Jiang ve ark. [39]	2024	LV-UNet	%85 IoU %91 Dice katsayısı
	Önerilen Model	2024	MMIAU-Net	%86 IoU %92 Dice katsayısı

Kaynakça

- [1] Pun, N. S., Agarwal, S. *Modality specific U-Net variants for biomedical image segmentation: a survey*, Artificial Intelligence Review, 2022.
- [2] Elnakib, A., Gimel'farb, G., Suri, J. S., El-Baz, A. *Medical Image Segmentation: A Brief Survey*, 2011.
- [3] Eker, A. G., Duru, N. *Medikal Görüntü İşlemede Derin Öğrenme Uygulamaları*, Acta Infologica, vol. 5, no. 2, pp. 459-474, 2021.
- [4] Ronneberger, O., Fischer, P., Brox, T. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, Medical Image Computing and Computer-Assisted Intervention(MICCAI), 2015.
- [5] He, K., Zhang, X., Ren, S., Sun, J. *Deep Residual Learning for Image Recognition*, IEEE Conference on Computer Vision and Pattern Recognition(CVPR), 2016, pp. 770-778.
- [6] Liu, S., Zhuang, Z., Zheng, Y., Kolmanič, S. *A VAN-Based Multi-Scale Cross-Attention Mechanism for Skin Lesion Segmentation Network*, IEEE Access, · 2023.
- [7] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., Polosukhin, I. *Attention Is All You Need*. NIPS, 2017.
- [8] Dosovitskiy A., Beyer L. , Kolesnikov A. , Weissenborn D. , Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., Uszkoreit J., Houlsby N. *An image is worth 16x16 words: Transformers for image recognition at scale*, arXiv:2010.11929, 2021.
- [9] Chitty-Venkata, K. T., Emani, M., Vishwanath, V., Somani, A. K. *Neural Architecture Search for Transformers: A Survey*, IEEE Access, 2022.
- [10] Wu, H., Chen, S., Chen, G., Wang, W., Lei, B., and Wen, Z. *FAT-Net: Feature Adaptive Transformers for Automated Skin Lesion Segmentation*, Medical Image Analysis, 2022.
- [11] Siddique, N., Paheding, S., Elkin, C. P., Devabhaktuni, V. *U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications*, IEEE Access, vol. 9, pp. 82031-82057, 2021.
- [12] Szegedy C., Liu W., Jia Y., Sermanet P., Reed S., Anguelov D., Erhan D., Vanhoucke V., Rabinovich A. *Going deeper with convolutions*. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 1-9.
- [13] Mendonça, T., Ferreira, P. M., Marques, J. S., Marçal, A. R. S., Rozeira, J. *PH² - a dermoscopic image database for research and benchmarking*. 35th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), Osaka, Japan, 2013.
- [14] Caicedo, J. C., Goodman, A., Karhohs, K. W., Cimini, B. A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., Rohban, M., Singh, S., and Carpenter, A. E. *Nucleus segmentation across imaging experiments: the 2018 data science bowl*. Nature Methods, 16, 2019, pp. 1247-1253.
- [15] Yadavendra, and Chand, S. *Semantic segmentation of human cell nucleus using deep U-Net and other versions of U-Net models*. School of Computer and Systems Sciences, Jawaharlal Nehru University, New Delhi, India, 2022.

- [16] Bernal, J., Sánchez, F. J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilarifo, F. *WM-DOVA maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians*. Computerized Medical Imaging and Graphics, 2015.
- [17] Zhou, Z., Siddiquee, M. M. R., Tajbakhsh, N., Liang, J. *UNet++: Redesigning Skip Connections to Exploit Multiscale Features in Image Segmentation*, IEEE Transactions on Medical Imaging, 2019.
- [18] Oktay, O., Schlemper, J., Le Folgoc, L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N. Y., Kainz, B., Glocker, B., Rueckert, D., *Attention U-Net: Learning Where to Look for the Pancreas*, 1st Conference on Medical Imaging with Deep Learning (MIDL 2018), 2018.
- [19] Hatipoğlu, N., Bilgin, G. *Histopatolojik Görüntülerin U-Net Tabanlı Modeller Kullanılarak Bölütlenmesi*, Medical Technologies Congress (TIPTEKNO) 2021.
- [20] Sariturk, B., Seker, D. Z. *A Residual-Inception U-Net (RIU-Net) Approach and Comparisons with U-Shaped CNN and Transformer Models for Building Segmentation from High-Resolution Satellite Images*, Sensors, 2022.
- [21] Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A. L., Zhou, Y. *TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation*. 2021.
- [22] Tragakis, A., Kaul, C., Murray-Smith, R., Husmeier, D. *The Fully Convolutional Transformer for Medical Image Segmentation*. IEEE/CVF Winter Conference on Applications of Computer Vision, 2023.
- [23] Yao, W., Bai, J., Liao, W., Chen, Y., Liu, M., Xie, Y. *From CNN to Transformer: A Review of Medical Image Segmentation Models*, Journal of Imaging Informatics in Medicine, 2023.
- [24] Faria, F. T. J., Moin, M. B., Debnath, P., Fahim, A. I., Shah, F. M. *Explainable Convolutional Neural Networks for Retinal Fundus Classification and Cutting-Edge Segmentation Models for Retinal Blood Vessels from Fundus Images*, 2024.
- [25] Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M. *Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation*. 2021.
- [26] Armand, T. P. T., Bhattacharjee, S., Kim, H. C., Choi, H. K. *Transformers Effectiveness in Medical Image Segmentation: A Comparative Analysis of UNet-based Architectures*, ICAIIC 2024
- [27] Nurhopipah, A., Murdiyanto, A. W., Astuti, T. *A Pair of Inception Blocks in U-Net Architecture for Lung Segmentation*, IEEE 5th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE), 2021.
- [28] Tong X., Wei J., Sun B., Su S., Zuo Z., Wu P. *ASCU-NET: Attention Gate, Spatial and Channel Attention U-Net for Skin Lesion Segmentation*. Diagnostics, 2021, pp. 11-29.
- [29] Wu H., Chen S., Chen G., Wang W., Lei B., Wen Z. *FAT-Net: Feature adaptive transformers for automated skin lesion segmentation*. Medical Image Analysis, 2022, pp. 76.
- [30] Zhang N., Yu L., Zhang D., Wu W., Tian S., Kang X. *APT-Net: Adaptive encoding and parallel decoding transformer for medical image segmentation*. Computers in Biology and Medicine, 2022, pp. 151.
- [31] Hu K., Lu J., Lee D., Xiong D., Chen Z. *AS-Net: Attention Synergy Network for skin lesion segmentation*. Expert Systems with Applications, 2022, pp. 1238-1259.
- [32] Hao S., Wu H., Jiang Y., Ji Z., Zhao L., Liu L., Ganchev I. *GSCEU-Net: An end-to-end lightweight skin lesion segmentation model with feature fusion based on U-Net enhancements*. Information, 2023, pp. 14-23.
- [33] Olimov B., Sanjar K., Din S., Ahmad A., Paul A., Kim J. *FU-Net: fast biomedical image segmentation model based on bottleneck convolution layers*. Multimedia Systems, 2021, pp. 1-14.
- [34] Roy K., Saha S., Banik D., Bhattacharjee D. *Nuclei-Net: A multi-stage fusion model for nuclei segmentation in microscopy images*. Innovations in Systems and Software Engineering, 2023, pp. 1-8.
- [35] Song H., Wang Y., Zeng S., Guo X., Li Z. *OAU-net: Outlined Attention U-net for biomedical image segmentation*. Biomedical Signal Processing and Control, 2023, pp. 79.
- [36] Ibtihaz N., Rahman M. *MultiResUNet: Rethinking the U-Net architecture for multimodal biomedical image segmentation*. Neural networks, 2020, pp. 74-87.
- [37] Chenarlogh A.V., Shabanzadeh A., Oghli G.M., Sirjani N., Moghadam F.S., Akhavan A., Tarzamni K.M. *Clinical target segmentation using a novel deep neural network: double attention Res-U-Net*. Scientific Reports, 2022, pp. 12.
- [38] Cai S., Xiao Y., Wang Y. *Two-dimensional medical image segmentation based on U-shaped structure*. International Journal of Imaging Systems and Technology, 2024, pp. 34.
- [39] Jiang J., Wang M., Tian H., Cheng L., Liu Y. *LV-UNet: A Lightweight and Vanilla Model for Medical Image Segmentation*. arXiv preprint arXiv, 2024, pp. 2048.