

TEKNOFEST

HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

SAĞLIKTA YAPAY ZEKÂ YARIŞMALARI

(Bilgisayarlı Görüyle Abdomen (Karin) Bölgesi için Hastalık

Tespiti Kategorisi)

PROJE DETAY RAPORU

TAKIM ADI: ODÜ-YAZ

TAKIM ID: 367195

İÇİNDEKİLER

	<u>Sayfa</u>
KAPAK	1
İÇİNDEKİLER	2
1. Proje Mevcut Durum Değerlendirmesi (20 puan)	3
2. Özgünlük (30 puan)	4
2.1. Grand Challenge Veriseti Uygulama Ortamı	4
2.2. VUMC Veriseti Uygulama Ortamı	6
3. Sonuçlar ve İnceleme (30 puan)	7
3.1. Grand Challenge Veriseti Uygulama Sonuçları	7
3.2. VUMC Veriseti Uygulama Ortamı	9
4. Deney ve eğitim aşamalarında kullanılan veri setleri (15 puan)	11
4.1. Grand Challenge Veriseti Temini	11
4.2. VUMC Veriseti Uygulama Ortamı	11
4.3. Veriseti Arttırma	11
5. Referanslar (5 puan)	12

ŞEKİLLER

Şekil 1. Dice benzelik katsayısının grafiksel gösterimi	7
---	---

ÇİZELGELER

	<u>Sayfa</u>
Çizelge 1. U-Net Hiper Parametreler	4
Çizelge 2. Eğitim Süreci Elde Edilen Değerler	7
Çizelge 3. Görsel Sonuçlar	8
Çizelge 4. Nesnel Sonuçlar	8
Çizelge 5. Algoritmaların Model Üzerindeki Etkisi	10
Çizelge 6. Fonksiyonların Model Üzerindeki Etkisi	10
Çizelge 7. Öğrenme Oranlarının Modeldeki Etkisi	10
Çizelge 8. Adım Sayısının Modeldeki Etkisi	10
Çizelge 9. Veri Arttırma Yöntemleri	12

1. Proje Mevcut Durum Değerlendirmesi (20 puan)

Proje Sunuş Raporu sonuçlarımızdan sonra puanımızın tam puana yakın olması sebebiyle mevcut planlamamız üzerinden yolumuza devam etmeye karar verdik.

Sağlık Bakanlığı tarafından verisetinin paylaşıldığı 31 Mayıs 2022 tarihinden önce modelimizin eğitiminde, karın bölgelerinde meydana gelebilecekleri tespit için MR görüntüleri kullanarak makine öğrenmesinden faydalانılmıştır.

Oluşturulan modelde, öncelikle görüntülerden arka plan temizlenmiş ve daha sonra üst iki kadran sağ için 1 ve sol için 2 olarak birbirinden ayrılmıştır. Sonrasında saf olarak sağ üst kadran ve sol üst kadran vektörleri çevreden bağımsız olarak üretilmiştir. Üretilen vektörler, tanı için istatistiksel hesaplamalar yapmak üzere özelliğe çıkarımında kullanılmıştır. Bu analizde alınan vektörleri temsilen; standart sapma (std), varyans (var), medyan (med), mod (mod), çarpıklık (skew), basıklık (kurt), süreklilik (cont), korelasyon (corr), entropi (ener) ve homojenlik(homo) değerleri her iki kadran için belirlenmiştir.

Görüntülerin analizinden sonra elde edilen tüm bu parametreler makine öğrenmesi için sisteme yüklenmiş ve makinenin görmediği görüntülerden, hangisinin hasta hangisinin sağlam olduğunu tahmin etmesini istenmiştir. Daha sonra istatiksel olarak anlamlı çıkan parametreleri kullanarak tekrar makine öğrenmesi yapıp yine makineden tahmin yapması istenmiştir.

Makine öğrenmesinde kullanılan veri seti; kullanacağımız modele karar vermek için daha önceden yapılan çalışmalar incelenmiş, özellikle çok sayıda ve tüm organlarla ilgili ayrı ayrı çalışma bulunması nedeniyle kullanılacak verisetlerinden yola çıkılarak kendi modelimizin eğitiminde yarışma komitesinin veri paylaşımından önce bu verisetlerinin (Bu verisetleri ile ilgili açıklamalar “**4. Deney ve eğitim aşamalarında kullanılan veri setleri**” başlığı altında detaylıca açıklanmıştır.) kullanımına karar verilmiştir.

Bu sayede organ tomografileri üzerine modelimizi daha iyi eğiterek katmanlar üzerinde gerekli iyileştirmeleri daha iyi gerçekleştirip, yarışma için verilecek frame’erde təşhis süreci hem daha kısa hem daha doğru olduğu düşünülmüştür.

Yarışma Komitesi’nin veriseti paylaşımı öncesinde, github.com [4] üzerinden paylaşılan, abdominal organ segmentasyonu için geniş ölçekli bir veri seti olan WORD veri seti kullanılarak Transfer Learning uygulaması gerçekleştirilmiştir. Bunun yanı sıra grand-challenge.org [x] üzerinden elde edilen böbrek, pankreas ve damar gibi görüntüler üzerinden çalışmalar sürdürülmüştür.

Görüntü ön-işleme adının temel amacı görüntü bölütleme ve sınıflandırma işlemlerinde ortaya çıkabilecek hatalı sonuçların önüne geçmektir.

MR görüntüleri, gürültü, kalıntı, kalite bozukluğu ve farklılıklar içerebilmektedir. Bu gürültülerin yok edilmesi ve görüntü kalitelerinin artırılması ile iyileştirilirler. MR taramalarının neden olduğu gürültü ve istenmeyen kalıntılar sınıflandırma aşamasında modelimizin verisetleri başarısını düşürebilmektedir.

Bölütleme işlemi için spatial-Fuzzy C-Means (s-FCM) metodu [5] kullanılmıştır. Geleneksel FCM metodu ile karşılaştırıldığında, s-FCM her bir pikselin diğer piksellerle olan uzaysal ilişkisini göstermektedir. Böylece, üyelik kümelerinden daha başarılı sonuçlar elde edilmekte ve kalıntılar, gürültüler daha başarılı olarak yok edilmektedir.

Bölütleme adımı tamamlandıktan sonra “sorun var/yok” işlemine karar verilmiştir. Bu karar bir sınıflandırma yöntemi tarafından verilmesine karşın, organlar üzerindeki bazı önemli özellikler karar verme sürecini kolaylaştırmaktadır.

Çalışmalarımızda, otomatik çoklu organ segmentasyonunun gerçekleştiriminde Python programlama dili kullanılmıştır. 3D BT görüntülerini okumak ve üzerinde işlem yapmak için Simple ITK kütüphanesi, sayısal işlemleri için Numpy kütüphanesi ve derin öğrenme modellerini çalışırmak için Pytorch kütüphanesi kullanılmıştır.

Son olarak, modellerin sonuçlarını görselleştirmek için Matplotlib kütüphanesi kullanılmıştır. Derin öğrenme modeli ise 6 GB GDDR6 belleğe sahip NVidia GeForce RTX 3060 üzerinde çalıştırılmıştır. Değerlendirme ölçüyü olarak Dice skoru kullanılmıştır.

Tüm örnekler eğitim ve test için %70-%30 olarak ayrılmıştır.

2. Özgünlük (30 puan)

Tıbbi görüntü böülütleme, görüntü analiz çalışmalarında önemli bir yere sahip olup, tıp alanında bilgisayar destekli teşhis için büyük bir öneme sahiptir. Başta radyoloji olmak üzere, tıbbın birçok alanında elde edilen görüntülerden, ilgi bölgesinin bölütlenmesi, hastalıkların teşhisi ve tedavi planlaması için kayda değer önceliğe sahiptir. Organ, lezyon veya tümör gibi ilgi bölgesi olarak değerlendirilebilecek görüntüler üzerindeki radyolojik yapılar, görüntü böülütleme ile çıkarılabilir. Böylece, karar süreçlerinde ikincil bir araç olan destekleyici nitelikte ek materyaller elde edilmiş olmaktadır. Görüntü böülülemenin, ilgi bölgesinin yerinin belirlenmesi ve nesne bölgesinin sınırının tanımlanması olarak, iki ilişkili işlemlerden oluşan düşüncüsülebilir.

Bu çerçevede ön hazırlık aşamasında şartnamede istenenler tek tek ele alınarak model üzerinden sonuçları incelenmiştir. Karın bölgesine ait bilgisayarlı tomografi (BT) görüntülerinde çok sınıflı organ segmentasyonu, tıbbi görüntü analizi için önemlidir. Karın iç organları, şekil ve boyut varyasyonları nedeniyle büyük bireysel farklılıklarla sahip olabilir, bu da otomatik segmentasyon yöntemlerinin geliştirilmesini zorlaştırmaktadır. Bu nedenle, segmentasyon doğruluğunun iyileştirilmesi aktif bir araştırma alanı olmuştur.

Verilerin işlenmesi, evrimli sinir ağının oluşturulup eğitim aşamasının gerçekleştirilmesi ve sonuçların değerlendirilmesi için kullanılan yazılımsal ve donanımsal kaynaklar şu şekildedir;

Yazılımsal Kaynaklar:

- Python 3.9
- Anaconda3 2022.5
- Tensorflow 2
- Cuda Toolkit 11.1
- NVIDIA Driver 516

Donanımsal Kaynaklar:

- Intel i7-11800H CPU
- 64 GB 3200 Mhz DDR4L RAM
- NVIDIA GTX 3060, 6 GB GDDR6 RAM, Grafik kartı

2.1. Grand Challenge Veriseti Uygulama Ortamı

Bir ESA algoritmasının öğrenmesini sağlayan en önemli etken hiperparametrelerdir. Bu parametreler epoch, batch size ve learning rate olarak verilebilir. Bu parametreler sezgisel olarak bulunamadığından deneysel olarak karar verilmek zorundadır. Çizelge 1.'de verilen hiperparametreler tercih edilmiştir. Bu parametrelerin seçilirken elde bulunan eğitim setindeki örnek sayısı verinin türü ve litaratürde sıkılıkla kullanılan değerler baz alınmıştır. Deney sürecinde her 1 epochtaki oluşan modelin performansı göz önüne alınarak, en yüksek doğruluk oranını veren model kaydedilmiştir.

Çizelge 1. U-Net Hiper parametreler

Epoch	20
Steps Per Epoch	300
Batch Size	2
Verbose (Model Kaydetme Periyodu)	1
Loss Function	Binary crossentropy
Optimization Function	Adam
Learning Rate	1e-4

İkili formattaki imgeler için en çok tercih edilen karşılaştırma metrikleri tercih edilmiştir. Bunun için referans ve çıktı imgesi karşılaştırılarak, True Positive (TP), False Positive (FP), False Negative (FN) ve True Negative (TN) sayıları bulunmuştur. Daha sonra dengeli bir değerlendirme olması açısından elde edilen TP, FP, TN, FN değerleri ile Precision, Recall, PBC, specificity, FPR, FNR, FMeasure sonuçları bulunmuştur.

Referans imgedeki piksel $GT(x,y)$ olsun ve elde edilen çıktı imgedeki piksel $I(x,y)$ olmak üzere TP, FP, TN, FN elde etmek için aşağıdaki kurallar kullanılmıştır.

- TP: $GT(x,y)=255$ ve $I(x,y)=255$
- TN: $GT(x,y)=0$ ve $I(x,y)=0$
- FP: $GT(x,y)=0$ ve $I(x,y)=255$
- FN: $GT(x,y)=255$ ve $I(x,y)=0$

Elde edilen bu değerler ile aşağıda formülleri verilen metrikler hesaplanmıştır [13].

$$\begin{aligned}
 \text{recall} &= \frac{\text{TP}}{\text{TP} + \text{FN}} \\
 \text{specificity} &= \frac{\text{TN}}{\text{TN} + \text{FP}} \\
 \text{FPR} &= \frac{\text{FP}}{\text{TN} + \text{FP}} \\
 \text{FNR} &= \frac{\text{FN}}{\text{TN} + \text{FP}} \\
 \text{PBC} &= 100.0 * (\text{FN} + \text{FP}) / (\text{TP} + \text{FP} + \text{FN} + \text{TN}) \\
 \text{precision} &= \frac{\text{TP}}{\text{TP} + \text{FP}} \\
 \text{FMeasure} &= 2.0 * (\text{recall} * \text{precision}) / (\text{recall} + \text{precision})
 \end{aligned}$$

2.1.1. Karşılaştırılan ESA Çeşitleri

Farklı U-net ağ yapılarının başarımının ölçülmesi ve karşılaştırılması için verilen farklı ağ mimarileri aynı eğitim sistemiyle aynı hiper parametreler ile eğitime tabi tutulmuş ve aynı test veri seti ile başarımı ölçülmüştür.

2.1.1.1. Basit ESA

Sınırlı sayıda girdi ve çıktı katmanıyla en temel yapısıyla basit bir ESA modeli kurulmuştur. Bu modelde eğitim setindeki 512x512 boyutlu imgeler eğitim için verilirken 256x256 ya küçültülmüş ve eğitim bu boyutlar ile tamamlanmıştır. Başarım ölçülürken elde edilen 256x256 çıktı imgesi 512x512 ye tekrar ölçeklenmiş ve referans görüntüler ile karşılaştırılmıştır.

2.1.1.2. U-Net 256x256

Bu modelde temel U-net yapısı kullanılmıştır. Aktivasyon fonksiyonu RELU dur. Ayrıca eğitim setindeki 512x512 boyutlu imgeler eğitim için verilirken 256x256 ya küçültülmüş ve eğitim bu boyutlar ile tamamlanmıştır. Başarım ölçülürken elde edilen 256x256 çıktı imgesi 512x512 ye tekrar ölçeklenmiş ve referans görüntüler ile karşılaştırılmıştır.

2.1.1.3. U-Net 512x512

Bu modelde temel U-net yapısı kullanılmıştır. Aktivasyon fonksiyonu RELU dur. Ayrıca eğitim setindeki 512x512 boyutlu imgeler eğitim için doğrudan kullanılmış ve başarım ölçülürken elde edilen 512x512 çıktı imgesi referans görüntüler ile karşılaştırılmıştır. Oluşturulan bu kod ile act_func parametresi ile ağıın aktivasyon fonksiyonu dinamik olarak değiştirilebilmektedir.

2.1.1.4. U-Net 512x512 (Sigmoid)

Bu modelde temel U-net yapısı kullanılmıştır. Ancak aktivasyon fonksiyonu değiştirilerek RELU yerine Sigmoid kullanılmıştır. Eğitim setindeki 512x512 boyutlu imgeler eğitim için doğrudan kullanılmış ve başarım ölçülürken elde edilen 512x512 çıktı imgesi referans görüntüler ile karşılaştırılmıştır. 2.1.1.3. de verilen model koduna act_func parametresi sigmoid gönderilmiştir.

2.1.1.5. U-Net 512x512 (Linear)

Bu modelde temel U-net yapısı kullanılmıştır. Ancak aktivasyon fonksiyonu değiştirilerek RELU yerine Linear kullanılmıştır. Eğitim setindeki 512x512 boyutlu imgeler eğitim için doğrudan kullanılmış ve başarım ölçülürken elde edilen 512x512 çıktı imgesi referans görüntüler ile karşılaştırılmıştır. 2.1.1.3. de verilen model koduna act_func parametresi linear gönderilmiştir.

2.2. VUMC Veriseti Uygulama Ortamı

Coklu organ segmentasyonu için farklı renk uzayına sahip BT görüntülerini ile erken füzyon yöntemini içeren iki aşamadan oluşan 3D U-Net tabanlı bir model çalışması gerçekleştirilmiştir.

İlk aşamada HSV formatına dönüştürülmüş görüntüler önerilen 3D U-Net modeli kullanılarak eğitilir ve son aşamada $1 \times 1 \times 1$ evrişim işleminden geçirilerek 13 tane organ ve 1 tane organ olmadığını belirten 14 kanallı bir çıktı elde edilir. Renk uzayı seçiminde ise Ghosh ve ark. (2018) tarafından yapılan çalışma etkili olmuştur. Bu çalışmada endoskop görüntüleri üzerinde kanama bölgesini tespiti için önerilen mimari, farklı renk uzaylarında test edilmektedir ve en iyi performansın HSV renk uzayında elde edildiği görülmektedir. Bu bilgiler ışığında veri setinde kullanılan gri BT görüntüler HSV formatına dönüştürülmektedir.

İkinci aşamada ise ilk aşamada elde edilen çıktı görüntüsü, orijinal görüntü ile birleştirilir ve toplamda 15 kanallı bir görüntü elde edilir. Bu birleştirme işlemi erken füzyon olarak adlandırılmaktadır. Bu sayede segmentasyonu yapılacak organların özelliklerinin daha belirgin hale getirilmesi amaçlanmıştır. Birleştirilen bu görüntüler önerilen 3D U-Net mimarisi ile eğitilir ve tahmin segmentasyon sonuçları elde edilir.

Özelliği çıkarılacak olan görüntü kodlayıcı kısmında ağa verilir. Kodlayıcı kısmında ağa alınan görüntüler üzerinde tüm seviyelerde $3 \times 3 \times 3$ boyutunda iki tane evrişim işlemi uygulanmaktadır. Her evrişim işlemini Batch Normalization ve ReLU aktivasyon fonksiyonu takip eder.

Görüntülerdeki organların boyutlarının ve şeşinin farklı olması otomatik segmentasyon işlemini karmaşık bir görev haline getirmektedir. Özellikle küçük boyutlu organların tahmininde yüksek başarılar elde edilmesi oldukça zordur ve daha iyi sonuçlar elde edebilmek için modelin farklı bileşenleri öğrenmesi gerekmektedir. 3D U-Net mimarısında evrişim katmanlarında görüntülerdeki özelliklerin öğrenme işlemi gerçekleştirilmektedir. Fakat havuzlama katmanı sabit bir işlemidir ve bu katmanda öğrenme işlemi gerçekleştirilmemektedir. Modeldeki tüm havuzlama katmanlarının evrişim katmaları ile değiştirilmesi eğitim sırasında küçük boyutlu nesnelerin özelliklerinin kaybolmasını engelleyebilir fakat havuzlama katmanlarının kaldırılması modelin görüntülerdeki özelliklerini ezberlemesine neden olabilir.

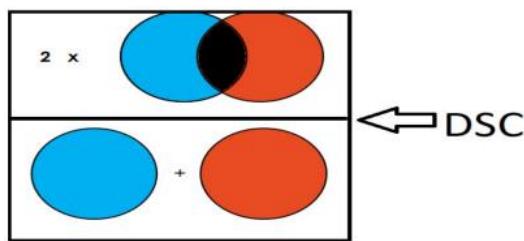
Modelimizde çoklu organ segmentasyonunda küçük organların segmentasyonunu iyileştirmek için hem havuzlama katmanın hem de evrişim katmanın avantajlarından yararlanarak her ikisinin kullanıldığı hibrit bir yapı tasarlanmıştır. Bir seviyeden diğer bir seviyeye geçiş yaparken elde edilen özellik haritasına ayrı ayrı maksimum havuzlama ve adım sayısı 2 olan $2 \times 2 \times 2$ 'lik bir evrişim işlemi uygulanmıştır. Bu işlemler sonucunda boyutu yarıya düşen iki çıktı elde edilmiştir. Maksimum havuzlama katmanın kullanılması ile modelin görüntülerdeki özelliklerini ezberlemesini ve evrişim işleminin uygulanması ile de görüntülerdeki önemli bilgilerin kaybının engellenmesi amaçlanmıştır.

Bu iki katmanın çıktıları birleştirilerek bir sonraki katmana girdi olarak verilmektedir. Kod çözme kısmında ise kodlayıcı kısmından alınan çıktı giriş görüntüsünün boyutuna ulaşılana kadar 2 adımlık $2 \times 2 \times 2$ boyutunda ters evrişim ile yukarı örneklenir. Her bir ters evrişiminden sonra ReLU aktivasyon fonksiyonu uygulanmaktadır. Her seviyede bulunan öznitelik haritası kendisine karşılık gelen kodlayıcı kısmından elde edilen öznitelik haritası ile birleştirilir. Daha sonra bu birleştirilen öz niteliklere $3 \times 3 \times 3$ 'lük iki tane evrişim uygulanır. Son katmanda $1 \times 1 \times 1$ evrişim işlemi sonucu $128 \times 128 \times 48$ boyutunda 14 kanallı bir öznitelik haritası elde edilir ve softmax aktivasyon fonksiyonundan geçirilerek üç boyutlu çoklu organ segmentasyonu gerçekleştiriliyor.

Modelimizde kayıp fonksiyonu olarak ortalama Dice kayıp fonksiyonu kullanılmıştır. Her iki aşamada elde edilen çıktıların Dice kaybı hesaplanır ve ortalaması alınır.

2.2.1. Değerlendirme Ölçütü

Modelin segmentasyon performansını değerlendirmek için Dice benzerlik katsayı (DSC) kullanılmaktadır. Bu değerlendirme metriği iki görüntü arasında eşleşen piksellerin sayısını ölçmektedir. Şekil 1'de Dice benzerlik katsayıının grafiksel gösterimi verilmiştir. Bu katsayıının değeri sıfırdan bire kadar değişir ve Dice kayıp oranı 1–DC olarak ifade edilir.



Şekil 1. Dice benzerlik katsayısunın grafiksel gösterimi

Modelimizde çok sınıflı segmentasyon gerçekleştirilmektedir. Bu yüzden her bir organ için Dice benzerlik katsayısı hesaplanmaktadır. Tüm organlar için hesaplanan Dice benzerlik katsayılarının toplamının ortalaması alınarak ortalama Dice benzerlik skoru elde edilir. Ortalama Dice benzerlik katsayısı (ADC) formülünde K simbolü organ sayısını temsil etmektedir.

$$ADC = \frac{1}{K} \sum_{j=1}^K \frac{2 \sum_{i=1}^n \sum_{j=1}^m \bar{y}_{ij,k} y_{ij,k}}{\sum_{i=1}^n \sum_{j=1}^m \bar{y}_{ij,k} + \sum_{i=1}^n \sum_{j=1}^m y_{ij,k}}$$

3. Sonuçlar ve İnceleme (30 puan)

3.1. Grand Challenge Veriseti Uygulama Sonuçları

Bölüm 2.1.'de verilen parametreler ile hazırlanan ağ modelleri, 20 epoch ve her epoch içerisinde 300 iterasyon içerek şekilde örnek veri seti ile eğitim sürecine tabi tutulmuştur. Modellerin eğitim sürecindeki Loss (kayıp) ve Accuracy (doğruluk) değerleri Çizelge 2.'de gösterildiği gibidir.

Çizelge 2. Eğitim süreci elde edilen değerler

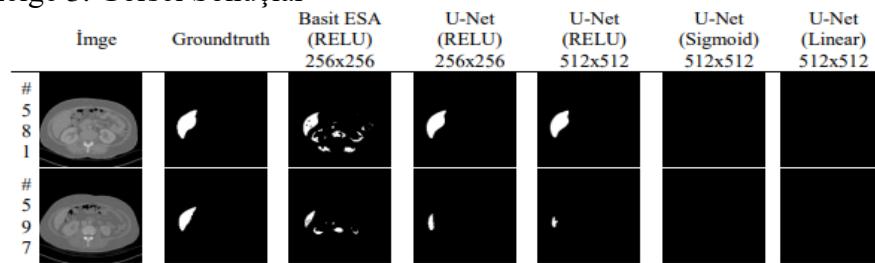
Epoch	Basit ESA		U-Net (RELU)		U-Net (RELU)		U-Net (Sigmoid)		U-Net (Linear)	
			Loss	Acc	Loss	Acc	Loss	Acc	Loss	Acc
	256x256	256x256	512x512	512x512	512x512	512x512				
1	0.2708	0.9299	0.2131	0.9285	0.2046	0.9287	0.2980	0.9303	1.1355	0.9291
2	0.2053	0.9269	0.1641	0.9269	0.1707	0.9268	0.2911	0.9268	1.1792	0.9268
3	0.1820	0.9266	0.1378	0.9266	0.1504	0.9264	0.2861	0.9266	1.1832	0.9266
4	0.1779	0.9251	0.1482	0.9249	0.1235	0.9249	0.2848	0.9249	1.2111	0.9249
5	0.1681	0.9248	0.1215	0.9311	0.1268	0.9254	0.2803	0.9254	1.2026	0.9254
6	0.1578	0.9283	0.0950	0.9641	0.1041	0.9539	0.2718	0.9281	1.1584	0.9281
7	0.1480	0.9309	0.0846	0.9725	0.0842	0.9712	0.2682	0.9286	1.1512	0.9286
8	0.1629	0.9270	0.0909	0.9699	0.0790	0.9769	0.2689	0.9272	1.1738	0.9272
9	0.1530	0.9276	0.0788	0.9773	0.0768	0.9800	0.2758	0.9231	1.2387	0.9231
10	0.1491	0.9294	0.0732	0.9813	0.0766	0.9797	0.2685	0.9260	1.1927	0.9260
11	0.1513	0.9289	0.0669	0.9834	0.0823	0.9759	0.2664	0.9265	1.1851	0.9265
12	0.1456	0.9296	0.0656	0.9835	0.0691	0.9826	0.2631	0.9275	1.1679	0.9275
13	0.1367	0.9356	0.0984	0.9638	0.0670	0.9828	0.2619	0.9277	1.1649	0.9277
14	0.1441	0.9299	0.0663	0.9817	0.0687	0.9806	0.2625	0.9272	1.1733	0.9272
15	0.1380	0.9339	0.0601	0.9871	0.0654	0.9846	0.2673	0.9250	1.2087	0.9250
16	0.1381	0.9334	0.0585	0.9863	0.0676	0.9820	0.2618	0.9272	1.1733	0.9272
17	0.1354	0.9349	0.0589	0.9860	0.0798	0.9741	0.2663	0.9252	1.2052	0.9252
18	0.1303	0.9374	0.0553	0.9884	0.0599	0.9856	0.2558	0.9295	1.1364	0.9295
19	0.1356	0.9344	0.0527	0.9887	0.0484	0.9854	0.2578	0.9286	1.1510	0.9286
20	0.1306	0.9377	0.0561	0.9882	0.0322	0.9873	0.2668	0.9249	1.2111	0.9249

Eğitim sürecinde görüldüğü üzere 256x256'lık modellerde Basit ESA yapısı doğruluk oranını 0.9299 dan en iyi 0.9377 oranına kadar geliştirebilirken U-Net yapısı ise 0.9285 den en iyi 0.9887 oranına kadar geliştirmiştir. Basit ESA yapısının eğitim sürecinde de öğrenme seviyesinin çok verimli olmadığı görülmektedir. U-Net (RELU) 256x256 modeli ile U-Net (RELU) 512x512 modeli eğitim süreci değerlendirildiğinde ise çok yakın bir doğruluk ile başlayıp, eğitim sonunda benzer bir doğruluk oranı değerini yakaladığı görülmektedir.

Öte yandan U-Net 512x512 modelleri arasında Sigmoid ve Linear aktivasyon fonksiyonlarına sahip olan modellerin Basit ESA modelinde olduğu gibi öğrenme sürecinde verimli olmadığı ve kabul edilebilir ölçüde öğrenmeyi gerçekleştirmediği görülmektedir.

Eğitimmiş modelin son haline test için verilen girdilere karşılık gelen groundtruth ve modelin çıktıları Çizelge 3.'te gösterilmiştir.

Çizelge 3. Görsel Sonuçlar



Çizelge 3.'teki görsel sonuçlardan anlaşılabileceği gibi önerilen U-Net (RELU) yapısı 256x256 ve 512x512 yapıları referans imgeye yakın sonuçlar üretmemiştir. Diğer yandan Basit ESA yapısı çok kötü sonuçlar üretirken, U-Net'in aktivasyon fonksiyonu değiştirildiğinde (Linear veya Sigmoid, 512x512) hiçbir sonuç üretmemiştir. Basit ESA yapısındaki filtreler küçük sayıda olduklarıdan dolayı overfitting durumu oluşturmaktadır ve kötü sonuç üretilmesine sebep olmuştur. Görüldüğü gibi Basit ESA algoritması piksellererdeki yumuşak geçişleri öğrenmemektedir. Diğer yandan U-Net (RELU) yapısı gözle dahi görülmeyen karaciğer alanlarını bülütleyebilmektedir. U-Net 512x512 için Linear ve Sigmoid fonksiyonlarının karaciğer bölgelerini öğrenemedikleri gözlenmiştir. Bunun nedeni karaciğer verisinin 0-255 arasında olduğundan RELU aktivasyon fonksiyonuyla daha uyumlu çalışmasından kaynaklanmaktadır. Ayrıca adil bir karşılaştırma olması için tüm modellerde üretilen çıktılar için eşik değeri 0.5 olarak ayarlanmıştır. RELU aktivasyon fonksiyonlu U-Net'ler 0.5 eşik değerinde en iyi sonucu üretirken aynı eşik değerinde Sigmoid ve Linear aktivasyon fonksiyonlu U-Net'ler bu eşik değerine göre başarı elde edememişlerdir.

Çizelge 4. Nesnel Sonuçlar

Metrik	Basit	U-Net	U-Net	U-Net	U-Net
	ESA	(Relu)	(Relu)	(Linear)	(Sigmoid)
	256x256	256x256	512x512	512x512	512x512
TP	565488	878526	824976	0	0
FP	754596	16218	6324	0	0
FN	340476	27438	80988	905964	905964
TN	25078128	25816506	25826400	25832724	25832724
Recall (%)	62.42	96.97	91.06	0.00	0.00
Specificity (%)	97.08	99.94	99.98	100.00	100.00
FPR (%)	2.92	0.06	0.02	0.00	0.00
FNR (%)	37.58	3.03	8.94	100.00	100.00
PBC (%)	409.55	16.33	32.65	338.82	338.82
Precision (%)	42.84	98.19	99.24	0.00	0.00
FMeasure (%)	50.81	97.58	94.97	0.00	0.00

Elde ettiğimiz sonuçlar Çizelge 4.'te, kullanılan metriklerin performansları yüzdelik olarak sergilenmiştir. Fmeasure metriğini tercih etmemizin nedeni precision ve recall değerlerinin harmonik ortalamasını allığından dolayı daha güvenilir nesnel sonuçlar vermektedir. Recall, Specificity, FPR, FNR, PBC, Precision ve F-measure metrikleri 0-1 arasında değer almaktadır. F-measure, Recall, Specificity ve Precision için en yüksek tanıma değerinde 1 değeri vermektedir. F-measure, Recall, Specificity ve Precision için en düşük tanıma değerinde 0 değeri vermektedir. Yine FPR, FNR ve PBC değerlerinde 0 yaklaşması durumunda iyi performans verdiği bilinmektedir.

Sonuçlar incelediğinde, 256x256 girdi boyutunda; U-net yapısı aşikar olarak Basit ESA yapısından daha başarılıdır. F-measure değeri referans alındığında, deneysel sonuçlara göre U-net yapısı 97.58% F-measure skoru verirken, Basit ESA yapısı sadece 50.81% F-measure skoru verebilmiştir. Precision sonuçları temel alındığında U-Net 98.19% skor verirken, Basit ESA yapısı ancak 42.84% performans skoru sergilemiştir.

Basit ESA yapısının düşük performans vermesinin nedeni, sınırlı sayıda veri ile eğitildiğinde overfitting veya underfitting durumuyla karşılaşmasıdır. Yeterli sayıda filtrenin olmamasından dolayı Basit ESA yapısı doğru sonuçları üretmemektedir. Diğer yandan U-net yapısı incelendiğinde, auto-encoder paradigmınınfiltrelere taşınmasından dolayı daha iyi görsel çıktılar verebilmiştir. Eşik değerleri ile oynanarak U-Net yapısının sonuçlarında daha da iyileştirme yapılabilmektedir. Fakat adil bir karşılaştırma olması açısından eşik değeri U-Net ve Basit ESA içinde 128 olarak alınmıştır.

Diger yandan adil bir performans karşılaştırması gerçekleştirmek için karaciğer imgeleri 512x512 boyutlarında (orjinal) eğitim işlemi gerçekleştirilmişdir. Farklı aktivasyon fonksiyonlarıyla eğitim işlemi tekrarlanmıştır. Çizelge 4.'te verilen sonuçlar göz önüne alındığında Relu aktivasyon fonksiyonu iyi sonuçlar üretirken Linear ve Sigmoid fonksiyonları karaciğer bölgelerini öğrenemedikleri gözlenmiştir. Bunun nedeni karaciğer verisinin 0-255 arasında olduğundan Relu aktivasyon fonksiyonuyla daha uyumlu çalışmasından kaynaklanmaktadır.

İmgelerle 256x256 boyutlarında eğitim yapıldığında 97.58% F-measure değeri üretirken 512x512 boyutlarında eğitim yapıldığında performans 94.97% F-measure skoru vermiştir.

Performans düşüklüğünün sebebi yüksek çözünürlükte karaciğerdeki gürültülü değerlerinin öğrenmeyi kötü etkilediği gözlenmiştir. Çünkü yüksek çözünürlükte daha detaylı bölgeler de eğitim sürecine dâhil olmuştur. Sonuç olarak eğitim aşamasında imgeleri 256x256 boyutunda eğitmek ve test aşamasında çıkan 256x256'lık sonuçları 512x512 boyutlarına yükseltmenin daha başarılı sonuçlar verdiği not edilmiştir.

3.2. VUMC Veriseti Uygulama Ortamı sonuçları

Otomatik çoklu organ segmentasyonunun gerçekleştiriminde Python (3.6) programlama dili kullanılmıştır. 3D BT görüntülerini okumak ve üzerinde işlem yapmak için Simple ITK kütüphanesi, sayısal işlemler için Numpy kütüphanesi ve derin öğrenme modellerini çalıştmak için Pytorch kütüphanesi kullanılmıştır. Son olarak modellerin sonuçlarını görselleştirmek için Matplotlib kütüphanesi kullanılmıştır.

Değerlendirme ölçüyü olarak Dice skoru kullanılmıştır.

Tüm örnekler eğitim ve test için %70-%30 olarak ayrılmıştır.

Gerçekleştirilen çoklu organ segmentasyon işleminde doğru tahmin oranını artırmak için önerilen modelin eğitimi aşamasında kullanılan parametrelerin optimizasyonu gerçekleştirilmiştir.

Her bir parametre ile önerilen model 300 adım çalıştırılmıştır ve optimum sonucu veren parametreler bulunmuştur. Modelde ilk olarak, öğrenme oranı 1e-3, batch size 2 ve aktivasyon fonksiyonu ReLU olarak rastgele belirlenmiştir.

Modelde belirlenen bu parametreler ve Adagrad, Adamax ve Adam optimizasyon algoritmaları ile ayrı ayrı çalıştırılmıştır. Bu sayede önerilen modelde optimum sonuç veren optimizasyon algoritması belirlenmiştir. Adagrad, Adamax ve Adam optimizasyon algoritmasının Dice sokuruna etkisi Çizelge 5'de verilmiştir.

Çizelge 5. Algoritmaların model üzerindeki etkisi

Organ Adı	Adagrad	Adamax	Adam
Sağ Böbrek	0,876	0,908	0,918
Sol Böbrek	0,814	0,900	0,920
Safra Kesesi	0,338	0,430	0,704
Aort	0,773	0,817	0,887
Pankres	0,582	0,649	0,663
Ortalama	0,630	0,714	0,760

Çizelge 6. Fonksiyonların model üzerindeki etkisi

Organ Adı	PReLU	ReLU
Sağ Böbrek	0,928	0,918
Sol Böbrek	0,927	0,920
Safra Kesesi	0,340	0,704
Aort	0,824	0,887
Pankres	0,641	0,663
Ortalama	0,716	0,760

Çizelge 5'de verilen sonuçlar karşılaştırıldığında Adam optimizasyon algoritması ile en yüksek segmentasyon sonucu elde edilmiştir. Ayrıca tüm organların segmentasyon ortalamasında da en yüksek sonucu verdiği görülmüştür. Optimum sonucu elde edebilmek için önerilen modelin eğitiminde Adam optimizasyon algoritması kullanılmıştır. İkinci aşamada farklı aktivasyon fonksiyonlarının önerilen model üzerindeki etkisini görmek için öğrenme oranı $1e-3$, batch size 2 ve optimizasyon algoritması olarak Adam seçilmiştir. Bu parametreler ile beraber PReLU ve ReLU aktivasyon fonksiyonları ayrı ayrı çalıştırılmıştır ve bu iki aktivasyon fonksiyonunun çoklu organ segmentasyonunda Dice skoru üzerindeki etkisi Çizelge 6'da verilmiştir.

Çizelge 6'da verilen sonuçlar karşılaştırıldığında her iki aktivasyon fonksiyonunda sonuçların birbirine yakın olduğu gözlenmektedir. PReLU aktivasyon fonksiyonunda sağ böbrek, sol böbrek, karaciğer, mide, inferior vena kava ve böbrek üstü bezlerinde çok az bir farkla ReLU aktivasyon fonksiyonuna göre daha iyi sonuç elde edilmiştir. Diğer organlarda ve tüm organların segmentasyon ortalamasında ReLU aktivasyon fonksiyonunun daha yüksek sonuç verdiği görülmektedir. Ortalama segmentasyonda daha yüksek başarı elde etmesi ve PReLU aktivasyon fonksiyonuna göre hesaplama maliyetinin daha düşük olması sebebiyle modelin eğitiminde kullanılması için aktivasyon fonksiyonu olarak ReLU seçilmiştir. Öğrenme oranının önerilen model üzerindeki etkisini tespit edebilmek için batch size 2, Adam optimizasyon algoritması ve ReLU aktivasyon fonksiyonu kullanılmıştır. Modelde öğrenme oranları $1e-3$, $1e-4$ ve $1e-5$ olarak sırayla çalıştırılmıştır ve Çizelge 7'de Dice skor sonuçları verilmiştir.

Çizelge 7. Öğrenme oranlarının modeldeki etkisi

Organ Adı	$1e-3$	$1e-4$	$1e-5$
Sağ Böbrek	0,918	0,910	0,758
Sol Böbrek	0,920	0,890	0,737
Safra Kesesi	0,704	0,413	0,222
Aort	0,887	0,813	0,710
Pankres	0,663	0,601	0,353
Ortalama	0,760	0,480	0,476

Çizelge 8. Adım sayısının modeldeki etkisi

Organ Adı	100	200	300	400	500
Sağ Böbrek	0,911	0,919	0,918	0,921	0,913
Sol Böbrek	0,906	0,908	0,920	0,906	0,907
Safra Kesesi	0,652	0,621	0,704	0,715	0,652
Aort	0,824	0,856	0,887	0,884	0,874
Pankres	0,644	0,665	0,663	0,641	0,646
Ortalama	0,731	0,750	0,760	0,759	0,757

Çizelge 7'de verilen sonuçlar karşılaştırıldığında $1e-4$ ve $1e-5$ öğrenme oranlarının hızının düşük olması nedeniyle önerilen modelin 300 adım sırasında çalıştırılması eğitim için yeterli gelmemiştir ve modelin performansı düşmüştür. Bu sebepten dolayı $1e-3$ öğrenme oranı seçiminde önerilen modelin sol böbrek üstü bezi haricinde diğer tüm organlarda ve tüm organların segmentasyon ortalamasında en yüksek sonucu verdiği gözlemlenmiştir.

Modelde, parametre optimizasyonu sonucunda öğrenme oranı $1e-3$, aktivasyon fonksiyonu ReLU, optimizasyon algoritması Adam ve batch size 2 olarak belirlenmiştir. Belirlenen optimum parametreler ile önerilen model farklı adım sayılarında çalıştırılmıştır ve sonuçları Çizelge 8'de verilmiştir. Verilen sonuçlar incelemişinde 100, 200 ve 300 adım sırasında segmentasyon doğruluk oranının arttığı görülmektedir. 300 adım sayısından sonra doğruluk oranı düşüşe geçmektedir.

4. Deney ve eğitim aşamalarında kullanılan veri setleri (15 puan)

4.1. Grand Challenge Veriseti Temini

<https://chaos.grand-challenge.org/> adresinde sunulan Abdominal BT ve MRI (T1 ve T2 ağırlıklı) ile KiTS19 verisetleri kullanılmıştır [12,14,15]. Veri kullanımı için ilgili sağlayıcıdan kullanıcı kaydı oluşturularak verilere erişilmiştir. kullanılan BT görüntüleri 17 farklı (anonimleştirilmiş) hastadan alınmıştır. Bu hastalar; tümör, lezyon veya herhangi bir hastalığa sahip olmayan sağlıklı hastalardır. Verilerin alındığı kaynaka da belirtildiği üzere: BT görüntüleri üst abdomen bölgesinden portal venöz fazda kontrast madde enjekte edilerek alınmıştır.

Eğitim seti toplamda 2579 adet abdominal CT görüntüsünü içermektedir. Her bir görüntüye karşılık gelen groundtruth görüntüsü; mevcut görüntüdeki karaciğerin bulunduğu bölge için 1 (bir), bulunmadığı bölge için 0 (sıfır) piksel değerlerini içeren görüntülerden oluşmaktadır. Bu görüntülerden 102 tanesi test amaçlı olarak eğitim yapılmadan önce eğitim setinden çıkarılmıştır.

Bunların yanı sıra; github.com [4] üzerinden paylaşılan, abdominal organ segmentasyonu için geniş ölçekli bir veri seti olan WORD veri seti kullanılarak Transfer Learning uygulaması gerçekleştirilmiştir.

Bu verisetinde, algoritma araştırmaları ve klinik uygulama geliştirme için büyük ölçekli bir Tüm Karın Organı Veri Kümesi (WORD) oluşturulmuştur. Bu veriseti, 150 abdominal BT verileri (30495 dilim) içeren 16 organ görüntüsünü barındırmaktadır.

4.2. VUMC Veriseti Temini

Çalışmalarda kullanılan diğer veriseti; Vanderbilt Üniversitesi Tıp Merkezinin (VUMC) çoklu organ segmentasyon yarışması (Multi-Atlas Labeling Beyond the Cranial Vault – Workshop and Challenge) için sağladığı karın BT görüntülerini içeren veri seti kullanılmıştır[16]. Bu veri setinde 30 karın BT görüntüsü bulunmaktadır. BT görüntülerinin hacim boyutları 512x512x85 ile 512x512x198 çözünürlük 0.54x0.54 mm² ile 0.98x0.98 mm² ve dilim kalınlığı 2.5 mm - 5.0 mm. arasında değişmektedir. Her bir BT görüntüsü dalak, sağ böbrek, sol böbrek, safra kesesi, yemek, borusu, karaciğer, mide, aort, inferior vena kava, portal ven ve splenik ven, pankreas, sağ böbrek üstü bezi, sol böbrek üstü bezi olmak üzere toplam 13 organ eğitimli kişiler tarafından manuel olarak etiketlenmiştir.

Veri setinde bulunan her BT taraması 512x512x85 ve 512x512x195 boyutunda değişmektedir. Her bir BT taramasının boyutunu eşletemek için x ve y düzleminde görüntümüzün boyutu $\frac{1}{4}$ oranında küçültülmüştür. Z ekseni ise dilim boyutlarını ayarlamak için etiketli görüntülerde organların bulunduğu başlangıç ve bitiş dilimleri tespit edilmiştir ve tahmin edilmesi istenen organların bulunmadığı kısımlar çıkarılmıştır. Bu işlemler sonucunda her bir BT görüntüsünün z ekseni farklı boyutlar elde edilmektedir. Eğitim aşamasında her bir görüntünün boyutunu eşletemek için rastgele bir başlangıç dilimi belirlenir ve bu başlangıç diliminden itibaren art arda gelen 48 tane dilim ile eğitim gerçekleştirilir. Sonuç olarak, ağa alınacak görüntünün boyutu 128x128x48x1 olarak ayarlanmaktadır. Son kısmındaki 1 sayısı görüntünün tek kanallı olduğunu göstermektedir. Önerilen model iki aşamalıdır ve ilk aşamasında Simple ITK kütüphanesinde bulunan Colormap fonksiyonu ile veri setinde bulunan gri görüntüler HSV renk uzayına sahip görüntülere dönüştürülmüştür. Bu fonksiyon ile tek kanallı görüntüler 0 ile 1 arasında normalize edilir ve 1 değerine sahip piksellere renk haritasındaki ilk renk atanarak RGB renk uzayına sahip görüntüler elde edilir. RGB renk uzayına dönüştürülen görüntüler istenilen renk uzayına (HSV) dönüştürülür.

Kullanılan veri setinin az sayıda görüntü içermesi sebebiyle rastgele seçilen görüntülere -5 ve 5 derece arasında rastgele döndürme işlemi uygulanmıştır. Aynı zamanda rastgele seçilen görüntüler üzerinde %25 ve %50 aralığında rastgele belirlenen bir oran ile görüntü yakınlaştırma işlemi gerçekleştirilir. Ek olarak her bir BT görüntüsünün içeriği tüm dilimler x ve y ekseniinde iki parçaya ayrılmıştır ve görüntülerdeki dilim sayıları iki katına çıkmıştır. Bu parçalara ayrılmış görüntülerden rastgele 48 tane dilim seçilerek görüntü, giriş görüntüsü ile aynı boyutta 128x128x48 olarak ayarlanmıştır.

4.3. Veri Arttırma:

Veri artırma, ESA tabanlı derin öğrenme tabanlı çalışmalarında birçok avantaj sağlamaktadır.

Bunlardan en önemlisi overfitting denilen filtrelerin öğrenmesini olumsuz etkileyen etkenin ortadan kaldırılmasıdır. Kısıtlı veri setlerinde modelin ezberlenmiş değerleri öğrenmesi yerine verinin çeşitli varyasyonları oluşturularak bu sorunun üstesinden gelmeye çalışılır. Çizelge 9.'da verilen veri artırma teknikleri tercih edilmiştir.

Cizelge 9. Veri Arttırma Yöntemleri

rotation_range(rotasyon aralığı)	0.2
width_shift_range (yatay kaydırma aralığı)	0.05
height_shift_range (dikey kaydırma aralığı)	0.05
shear_range (shear kaydırma aralığı)	0.05
zoom_range (zoom aralığı)	0.05
horizontal_flip (yatay aynalama)	True
fill_mode (döndürme sonrası boşlukları kapatma yöntemi)	'nearest'

5. Referanslar (5 puan)

1. Yang, S., Fang, B., Tang, W., Wu, X., Qian, J. ve Yang, W., 2017. Faster R-CNN based microscopic cell Detection. In 2017 International Conference on Security, Pattern Analysis, and Cybernetics (SPAC), IEEE, 345-350.
2. Ren, S., Girshick, R. ve Sun, J., 2017. Faster R-CNN: Towards real-time object detection with region proposal networks. IEEE Transactions on Pattern Analysis & Machine Intelligence, 39(6), 1137-1149.
3. Bengio, Y., 2009. Learning deep architectures for AI. Foundations and trends® in Machine Learning, 2(1), 1-127.
4. <https://github.com/HilLab-git/WORD> “WORD: BT görüntüsünden abdominal organ segmentasyonu için geniş ölçekli bir veri seti, kıyaslama ve klinik olarak uygulanabilir çalışma” Erişim Tarihi: 02.12.2021.
5. GAMBINO, O., DAIDONE, E., SCIORTINO, M., PIRRONE, R., ARDIZZONE, E., Automatic skull stripping in MRI based on morphological filters and fuzzy c-means segmentation. Conf Proc IEEE Eng Med Biol Soc, 2011: 5040-3, 2011.
6. Purushotham, S., Meng, C., Che, Z., Liu, Y., (2018), Benchmarking deep learning models on large healthcare datasets, Journal of Biomedical Informatics, 83, 112-134
7. Hu, Z., Tang, J., Wang, Z., Zhang, K., Zhang, L., Sun, Q., (2018), Deep learning for image-based cancer detection and diagnosis –A survey, Pattern Recognition, 83, 134-149.
8. ARAKERI, M. P., REDDY, G. R. M., Computer-aided diagnosis system for tissue characterization of brain tumor on magnetic resonance images.: 1-17, 2013.
9. Janney, S. S., & Chakravarty, S. (2019). Deep learning in medical and surgical instruments. Bioelectronics and Medical Devices, 833–855.
10. Liu, S., Wang, Y., Yang, X., Lei, B., Liu, L., Xiang Li, S., Wang, T. (2019). Deep Learning in Medical Ultrasound Analysis: A Review., Engineering. 261-275
11. Selvikvåg Lundervold, A., & Lundervold, A. (2018). An overview of deep learning in medical imaging focusing on MRI. Zeitschrift Für Medizinische Physik. 102-127
12. <https://grand-challenge.org/> “BT-MR Sağlıklı Karın Organ Segmentasyonu” Erişim Tarihi: 11.03.2022
13. Ghoneim, S, 2019, Accuracy, Recall, Precision, F-Score, Specificity, which to optimize on?, <https://towardsdatascience.com/accuracy-recall-precision-f-score-specificitywhich-to-optimize-on-867d3f11124>, Erişim Tarihi : 19.03.2022
14. The Cancer Imaging Archive, wiki.cancerimagingarchive.net, Erişim Tarihi: 01.04.2022
15. The KiTS19 Grand Challenge, kits19.grand-challenge.org, Erişim Tarihi: 01.04.2022
16. Xu Z. 2015. Multi-atlas labeling beyond the cranial vault - workshop and challenge www.synapse.org/#/Synapse:syn3193805/wiki/2177, Erişim Tarihi: 01.04.2022