



SAĞLIKTA YAPAY ZEKA YARIŞMASI

PROJE DETAY RAPORU

GÖREV: Difüzyon MR Görüntülerinden İnmenin Zamansal Sınıflandırması

Yarışma Eğitim Seviyesi: Üniversite ve Üzeri

Proje Adı: MR GÖRÜNTÜLERİNİN ZAMANSAL SINIFLANDIRILMASI

Takım Adı: MAJESTY

Takım ID: 742174

Başvuru ID: 3627621

İçindekiler Tablosu

Takım Şeması.....	3
PROBLEME EN YAKIN ÇÖZÜM SUNAN ULUSLARARASI MAKALELERİN ÖZETİ.....	3
KULLANILAN VERİ SETİ VE ALGORİTMALAR.....	4
Veri Setleri ve Veriyi Düzenleme Stratejileri: 4	
Kullanılan Algoritmalar ve Stratejiler: 5	
SONUÇLAR VE İNCELEME.....	7
Tablo:9	
Sonuçların İncelenmesi: 9	
ÖZGÜNLÜK.....	10
REFERANSLAR.....	12

1. TAKIM ŞEMASI

Takımımız 3 Yapay Zeka Mühendisliği öğrencisi, 1 Bilgisayar Mühendisliği öğrencisi ve 1 Tıp öğrencisinden oluşmaktadır. Yapay zeka öğrencisi olan takım arkadaşlarımız modellerin doğruluğu ve algoritma seçiminde görev almıştır. Tıp öğrencimiz ise veri setinin incelenmesinde ve anlamlandırılmasında görev almıştır.

2. PROBLEME EN YAKIN ÇÖZÜM SUNAN ULUSLARARASI MAKALELERİN ÖZETİ

1. Deep learning-based classification of diffusion-weighted imaging fluid-attenuated inversion recovery mismatch (Kim et al., 2025)

Bu çalışmada, DWI ve FLAIR sekansları arasındaki uyumsuzluk (DWI–FLAIR mismatch) durumu otomatik olarak tespit edilmeye çalışılmıştır. Bu uyumsuzluk, inmenin başlangıç süresini tahmin etmekte önemli bir göstergede olarak kabul edilmektedir. Bu amaçla geliştirilen derin öğrenme tabanlı konvolüsyonel sinir ağı (CNN) modeli, hem iç testlerde hem de farklı merkezlerden elde edilen harici testlerde yüksek AUC değerleriyle başarılı sonuçlar vermiştir. Bulgular, yaklaşımın radyolog yorumlarına kıyasla daha tutarlı sınıflamalar yapabildiğini göstermektedir. Ancak model yalnızca DWI ve FLAIR dizilerine dayandığından, ADC gibi potansiyel olarak faydalı ek sekanslar çalışmaya dahil edilmemiştir.

2. SDS-Net: A Synchronized Dual-Stage Network for Predicting Patients Within 4.5-h Thrombolytic Treatment Window Using MRI (Zhang et al., 2024)

Bu çalışma, zamana bağlı inme sınıflandırması için geliştirilen iki aşamalı derin öğrenme mimarisi SDS-Net'e odaklanmaktadır. Model, ilk aşamada inme lezyonlarını segmentleyip ardından bu verilerle başlangıç zamanını sınıflandırmaktadır. Böylece hem anlamlı bölgeler görüntünden doğrudan çıkarılmakta hem de sınıflama daha odaklı yapılmaktadır. DWI, ADC ve FLAIR sekanslarının birlikte kullanıldığı bu yöntem, doğruluk ve AUC açısından başarılı sonuçlar vermektedir. Segmentasyon ve sınıflandırmayı aynı anda gerçekleştirebilmesi modeli benzerlerinden ayırmaktadır. Ancak zaman sınıflaması yalnızca 4.5 saatlik eşik üzerinden yapılmış, hiperakut, subakut ve kronik ayrimı uygulanmamıştır.

3. MRI radiomic features-based machine learning approach to classify ischemic stroke onset time (Zhang et al., 2021)

Bu çalışmada, özellikle radyomik özelliklere dayalı daha klasik makine öğrenmesi yöntemleriyle sınıflandırma yapılmıştır. Araştırmacılar, DWI ve ADC sekanslarından doku yapısını yansitan binlerce özellik çıkartarak çeşitli algoritmalar eğitmiş ve özellikle yüksek duyarlılığa sahip modeller elde etmişlerdir. Bu durum, yöntemin tedaviye uygunluk belirlemekte faydalı olabileceğini göstermektedir. Yöntemin yalnızca DWI ve ADC ile çalışabilmesi önemli bir avantajdır; ancak manuel segmentasyon gerektirmesi ve sınırlı ölçükte test edilmiş olması, klinik uygulamalar için bazı kısıtlar doğurmaktadır.

4. A Machine Learning Approach for Classifying Ischemic Stroke Onset Time From Imaging (Ho et al., 2019)

Bu çalışma, dört farklı MR sekansından (DWI, ADC, FLAIR ve PWI) görüntü özellikleri çıkararak inmenin başlangıç süresini tahmin etmeye çalışmaktadır. Özellikle PWI'den elde edilen derin öğrenme temelli özellikler model performansını önemli ölçüde artırmıştır. Farklı makine öğrenmesi algoritmaları denenmiş ve karşılaştırmalı analizler yapılmıştır. Sonuçlar, modelin yalnızca görüntüye dayanarak inme başlangıcını makul doğrulukla tahmin ettiğini göstermektedir. Ancak sınıflandırma yalnızca 4.5 saat öncesi ve sonrası olarak ikili yapılmıştır.

5. Automatic detection and vascular territory classification of hyperacute staged ischemic stroke on diffusion weighted image using convolutional neural networks (Lee et al., 2023)

Bu araştırmada, transfer öğrenme temelli derin öğrenme modelleriyle DWI görüntüleri üzerinden hiperakut dönemde inme sınıflandırması yapılmıştır. Özellikle Inception-v3 mimarisi yüksek doğruluk(%86.3) ve F1 skoru(%86.2) ile dikkat çekmiştir. Model, DWI görüntüleri üzerinden inmenin hangi dolaşım bölgesinde olduğunu ve normal vakaları doğru bir şekilde ayırt edebilmektedir. Transfer öğrenme sayesinde az sayıda örnekle bile güçlü sonuçlar elde edilmiştir.

3. KULLANILAN VERİ SETİ VE ALGORİTMALAR

Veri Setleri ve Veriyi Düzenleme Stratejileri

Takım olarak modelimizi eğitmek için Teknofest tarafından paylaşılan veri setini kullandık. Buna ek olarak probleme uygun veri setleri araştırıldı ancak problemle alakalı MR CLEAN Registry, STIR (Stroke Imaging Repository) ve AISD veri setlerine erişim alınamadı.

Veri seti incelemesinden sonra ilk izlenimimiz görüntü türü (T2A, DWI ve ADC) üzerinden farklı modeller eğitmekti. Bu doğrultuda resimleri MR çeşitlerine göre sınıflandırmaya karar verdik. Veri setindeki resimlerin hangi görüntü türlerine karşılık geldiğini belirlemek ve sınıflandırmak amacıyla takımımızdaki tıp öğrencisinin bilgi birikiminden ve Google MedGemma modelinden faydalandık. Bu iş birliği sayesinde klasörlerdeki resimlerin hangi MR çeşidine karşılık geldiğini tespit ettik.

Ancak bazı hastalarda ADC görüntüleri eksiki. Bu eksikliği gidermek amacıyla, mevcut DWI görüntülerinden ADC görüntülerini üretmeyi hedefledik. Bunun için aşağıdaki formülü kullanırdık:

$$ADC = -1/b \cdot \ln(S(b)/S(0))$$

Burada:

S(b): difüzyon ağırlıklı (yüksek b değeri ile çekilmiş) görüntündür.

S(0): difüzyon ağırlıksız ($b = 0$) görüntündür.

b: kullanılan difüzyon ağırlığı değeridir.

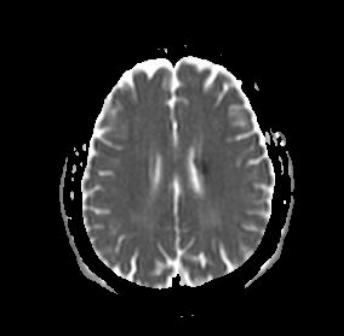
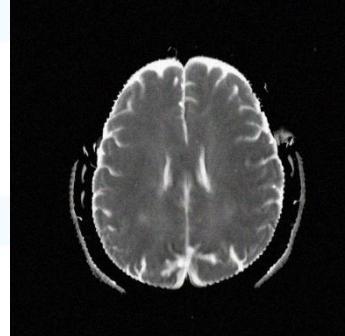
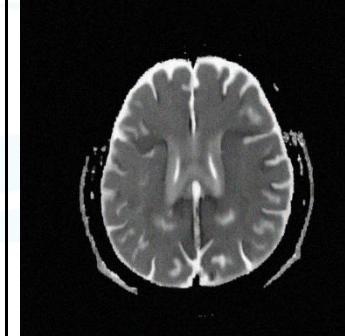
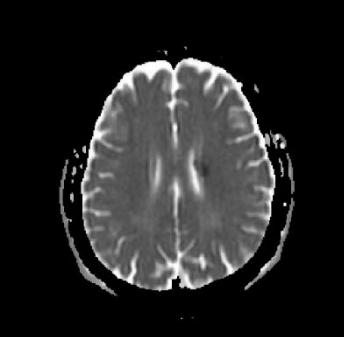
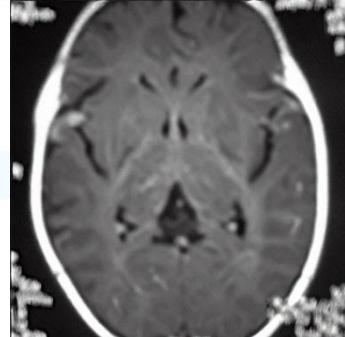
Bu süreci uygularken en temel sorunlardan biri, DWI görüntülerinde kullanılan b değerlerinin elimizde olmamasıydı. Bu nedenle bu değeri çeşitli sabit sayılar (örneğin 100, 300, 1000) olarak varsayıarak denemeler gerçekleştirdik.

Yaptığımız bu denemelerde, sabit b değerleri kullanılarak oluşturulan ADC görüntüleri görüntü kalitesi ve parlaklık dağılımı açısından tutarsız sonuçlar verdi. Özellikle bazı görüntülerde lezyon bölgeleri hiç belirginleşmedi, yapay parazitler ve siyah-beyaz (anlam taşımayan) görüntüler oluştu. Bu nedenle bu yöntemle klinik olarak geçerli bir ADC görüntüsüne ulaşmadık.

Sonuç olarak veri setindeki eksik ADC görüntülerini DWI üzerinden üretme çabamız, b değeri bilgisine sahip olmadığımızdan başarısız olmuştur.

Veri çoğaltma stratejisi olarak kafatası sıyırmaya yönteminden de faydalandık. HD-BET aracılıyla oluşturulan kafatası sıyrılmış görüntüler, orijinal veri setine eklenecek eğitim havuzumuz genişletildi. Bu hibrit eğitim yaklaşımının, modelimizin F1 skorunu belirgin şekilde artırdığı gözlemlendi.

Kısıtlı veri erişimimiz nedeniyle, klasik veri çoğaltma yöntemlerine ek olarak difüzyon modelleriyle görüntü üretme stratejisine başvurduk. Bu doğrultuda Stable Diffusion (1.5, XL) tabanlı modeller, Flux.1 Schnell modeli ve medikal görseller için özel olarak eğitilmiş Prompt2MedImage modelini kullandık. *Tablo 1*'de HiperakutAkut veri kümesinden bir örneğin girdi olarak kullanıldığı senaryoda modellerin verdiği çıktıları görebilirsiniz.

		
Orijinal	SD 1.5 Tabanlı Model	SDXL Tabanlı Model
		<p>NOT: Prompt2MedImage modelinin metin girdisini oluşturmak için orijinal resim gelişmiş VLM modellerine açıklattırılmıştır ve metin girdisi bu açıklamaya göre belirlenmiştir.</p>
Flux.1 Schnell Modeli	Prompt2MedImage	

Tablo 1

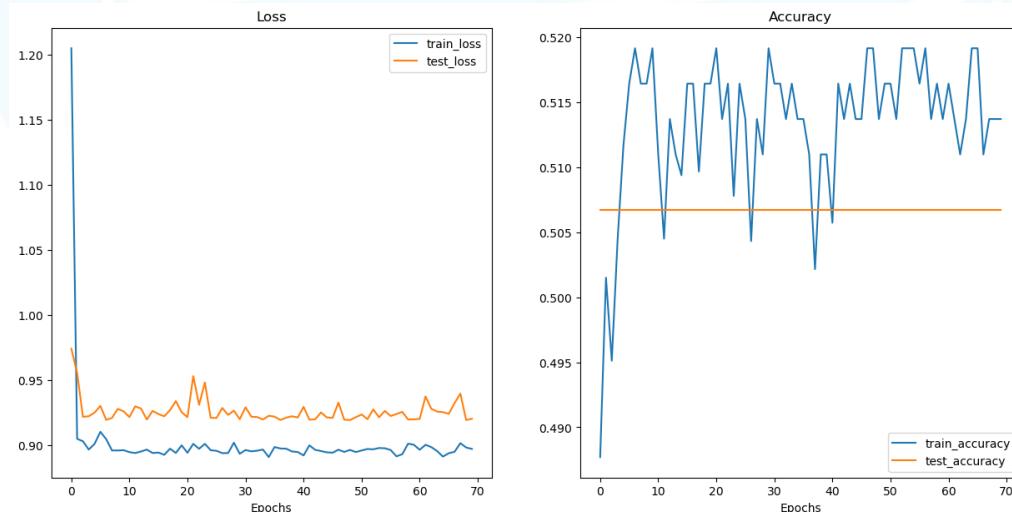
Difüzyon modelleri üzerinde yaptığım testler ışığında difüzyon modellerinin koşullandırıcı bir girdi resmi olmadan MR görüntülerini klinik tutarlılıkla oluşturamadığı; girdi resminin olduğu senaryoda ise MR görüntüsünün klinik tutarlılık sınırlarını aşmadığı, üretimlerde girdi resmine oldukça benzer görseller ürettiğini gördük. Kullanılan Algoritmalar ve Stratejiler kısmında bahsedeceğimiz hatalı veri odaklı eğitim döngüsü yönetiminin başarısını da göz önüne alduğımızda, difüzyon modelleri ile oluşturduğumuz verilerin benzerliğinin ezberleme durumuna neden olacağı endişesi ile bu stratejiyi rafa kaldırındı.

Alanda yeni ve gelişmekte olan bu teknolojinin, ilerleyen zamanlarda daha etkin kullanılabileceğine inanıyor ve bu yöndeki araştırmalarımı süzdürüyoruz.

Kullanılan Algoritmalar ve Stratejiler

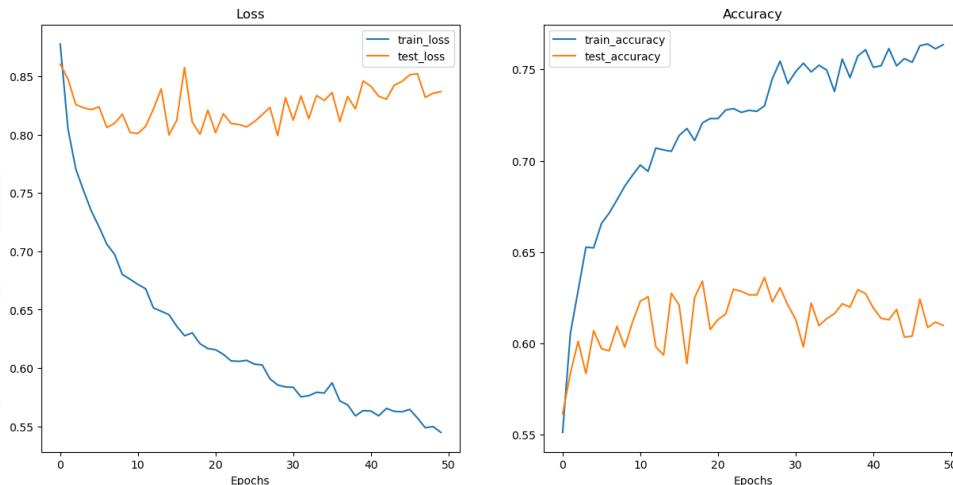
Modelimizi eğitirken veri setimizi %80-%20 train-validation şeklinde ayırdık. Bu kararın amacı, kaliteli ve iyi etiketlenmiş Teknofest verisini eğitim aşamalarında kullanıp potansiyel ek bir veri setini de test seti yapmak. Ancak herhangi bir veri setine erişim alamadığımız için bu planı gerçekleştiremedik.

Takımımızın temel stratejisi YOLO gibi temel modellerle en iyi sonucu elde edip bunu ikinci bir karar mekanizması ile desteklemekti. Temel modelleri araştırırken en yeni ve en güçlü sınıflandırma modelleri olan YOLOv11 serisi ile başladık. Bu modellerin yanında son yıllarda çok popülerleşen Transformers mimarisi kullanan modelleri de denemek istedik ve Google Vision Transformer (VIT) ile denemelerimizi yaptık. Google VIT modelini kullanırken hem ön eğitimli modeli hem de sıfırdan modeli eğitmeyi test ettik. Sıfırdan modeli eğittiğimiz testlerde model herhangi bir sonuca yakınsamadı. **Şekil 1**'de görüldüğü gibi loss, 0.89 ile 0.90 değerleri arasında sekipti.



Şekil 1

Ön eğitimli versiyonu kullandığımızda daha anlaşılır sonuçlar alabildik.



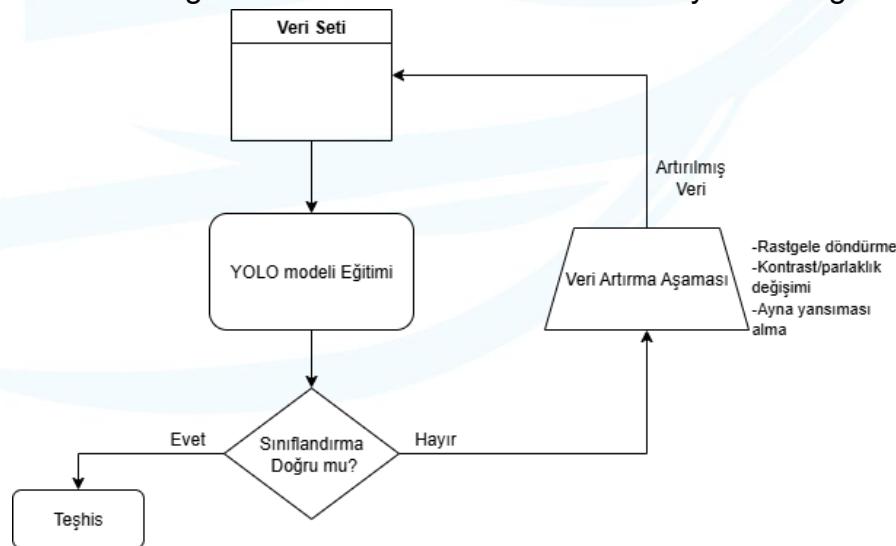
Şekil 2

Ancak Şekil 2'den de görüldüğü gibi train loss azalırken test loss arttığı için overfit olduğunu düşünüp ve YOLO modeline kıyasla kötü performans gösterdiği için çalışmalarımızı burada bıraktık.

Google VIT'e paralel yürüttüğümüz YOLOv11 ile çalışmalarda daha yüksek başarı aldık ve ana model olarak YOLOv11'i seçtik. Eğitim sürecinde kullandığımız ekstra bir adım ise hatalı veri odaklı eğitim döngüsü adımıydı.

Bu yöntem şu şekilde çalışmaktadır:

1. Mevcut model eğitim seti üzerinde tahmin yapar.
2. Hatalı sınıflandırılan örnekler belirlenir.
3. Bu örnekler üzerinde çeşitli veri artırma işlemleri uygulanır:
 - Rastgele döndürme
 - Kontrast/parlaklık değişimi
 - Ayna yansması alma
4. Üretilen veriler eğitim setine tekrar eklenerek model yeniden eğitilir.



Şekil 3

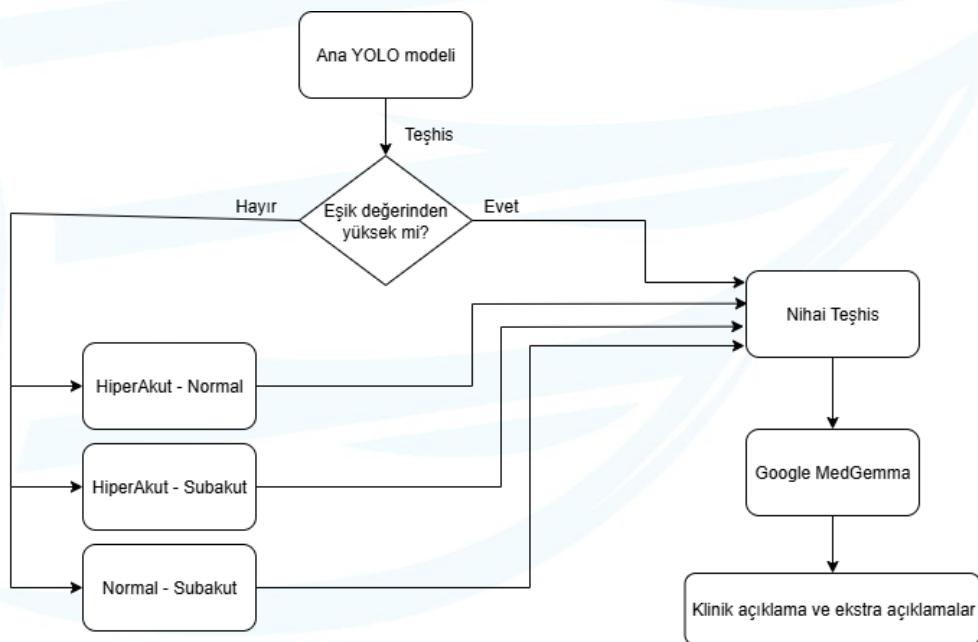
Böylece eğitim setindeki modelin öğrenmesi için yeterli olmayan veri noktalarını artırdık.

Ana modelimizi seçtikten sonra buna ek karar sistemleri araştırmaya başladık. Ek karar sistemleri şekil hem ana modelin arada kaldığı durumlarda destek olabilmek hem de sağlıkta çok dikkat edilen yapay zekanın açıklanabilirlik tarafını geliştirmek için tercih ettiğimiz yöntemdi. Bu tarafta 2 yöntem kullandık: Google MedGemma ve kendi geliştirdiğimiz iki aşamalı karar akışı.

Google MedGemma: Google'ın 2025 IO etkinliğinde tanıttığı Gemma açık kaynak modellerinin sağlık alanına özel olarak fine-tune edilmiş bir versiyonudur. Bizim kullandığımız 4 milyar parametreli versiyonu yazı-resim girdilerinden yazılı çıktı çıkarabilen multimodal bir mimariye sahip. Bu modeli elimizdeki veri setini daha iyi analiz edebilmek ve klinik bir doktorun görüntüleri nasıl yorumladığını anlayabilmek için kullandık. Özellikle çıktılarında hem teşhis koyup hem de görselin neresinden bu teşhise ulaşabildiğini söylemesi açıklanabilirlik açısından büyük bir katkı sağladı.

Ancak Google'ın istediği kullanım şartlarında model'in teşhis koymak için kullanılamayacağı belirtiliyor. Dolayısıyla biz de MedGemma modelini, teşhis konulduktan sonra hastaya neden bu teşhisin konulduğunu açıklayacak yardımcı bir katman olarak kullanma kararı aldık.

İki aşamalı karar mekanizması: Ana modelimiz olan yolo modeli bazen düşük güven aralığında tahminler yapabiliyor, dolayısıyla bu düşük güvenli tahminlerin daha spesifik bir modelle yeniden değerlendirilmesinin daha iyi sonuçlar verebileceğini düşündük. Temel çalışma prensibi **Şekil 4** gösterildiği şekilde:



Şekil 4

Geliştirilen sistemde, YOLO modelinin çıktısı ilk adımda değerlendirilir. Eğer modelin tahminine ait güven skoru yeterince yüksekse, doğrudan bu tahmin son teşhis olarak kabul edilir. Ancak güven skoru belirlenen eşik değerinin altında ise, bu durumda en yüksek iki sınıfa ait tahminler alınır ve veri, bu sınıflar için özel olarak eğitilmiş 3 ikili sınıflandırma modelinden ilgili olana yönlendirilir. İlgili ikili sınıflandırma modelinin

çıktısı, bu durumda nihai teşhis olarak belirlenir. Ve açıklamalar için MedGemma'ya gönderilir.

4. SONUÇLAR VE İNCELEME

Sonuçlarımızı alırken veri setinin %20'lik validation parçasını kullandık. Eğitimlerde ilk yaklaşımımız her difüzyon türü için ayrı bir model eğitmekti. Ancak her veriyi ayrı klasörlere ayırmayı ve eksik görüntüleri elde edemediğimizden tüm veriyi karıştırımıaya karar verdik.

Yeni yaklaşımımızla ilk sonuç aldığımız yer yolo modelleriydi. Bu modelleri aynı parametreler farklı epoch sayılarıyla performans testine tuttuk. Tablo 2'de tüm varyasyonların doğruluk sonuçlarına ulaşabilirsiniz. (Tabloda özel olarak belirtildiği sürece tüm modeller 3'lü sınıflandırma modeli olup aynı veri setiyle eğitilmiştir)

Model	Micro Precision	Micro Recall	Micro Average F1	Epoch
Yolov11-n-cls	0.826	0.838	0.832	100
Yolov11-s-cls	0.799	0.799	0.799	100
Yolov11-m-cls	0.854	0.854	0.854	300
Yolov11-x-cls	0.805	0.805	0.805	300
Yolov11-m-cls HiperakutAkut-Normal (ikili sınıflandırma)	0.890	0.902	0.896	150
Yolov11-m-cls HiperakutAkut-Subakut (ikili sınıflandırma)	0.926	0.967	0.946	150
Yolov11-m-cls Normal-Subakut (ikili sınıflandırma)	0.957	0.994	0.975	150
Google VIT	-	-	0,627	50
İki Aşamalı Teşhis Modeli	0.890	0.843	0.866	-

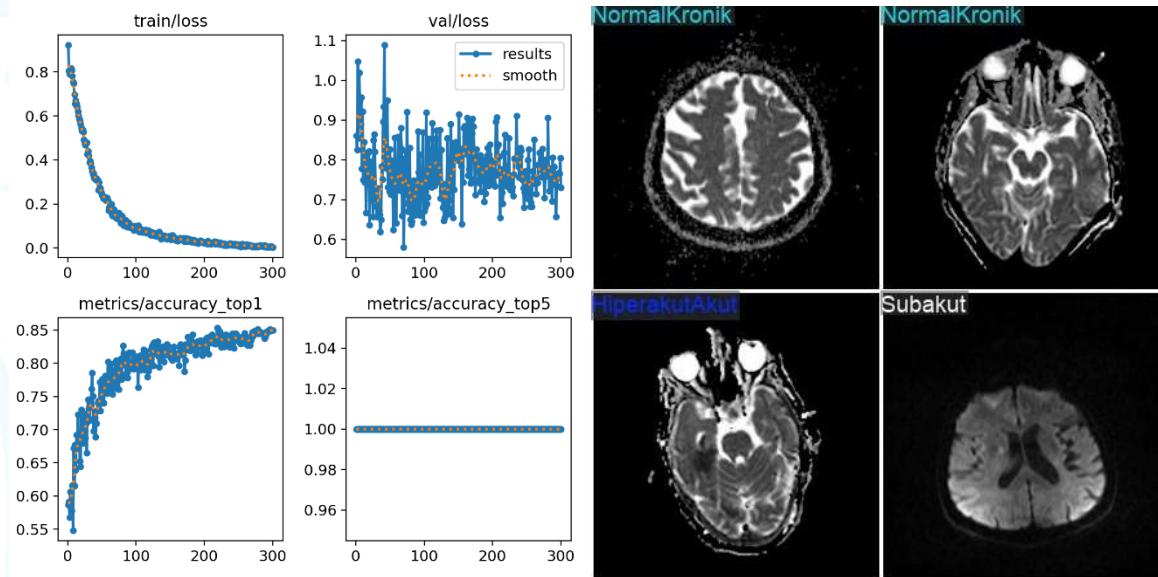
Tablo 2

Sonuçların incelenmesi

Tabloda en çok dikkat çeken yer, Yolo-n varyasyonunun az epoch sayısıyla kendinden büyük modellerden çok daha iyi sonuçlar vermesi. Bu durum, YOLO-m varyasyonunda olduğu gibi, epoch sayısı artırılarak dengelenebiliyor. Ancak az kaynak ile rekabetçi bir sonuç vermesi hızlı testler için bu model varyasyonunu oldukça kullanılabilir kılmıyor.

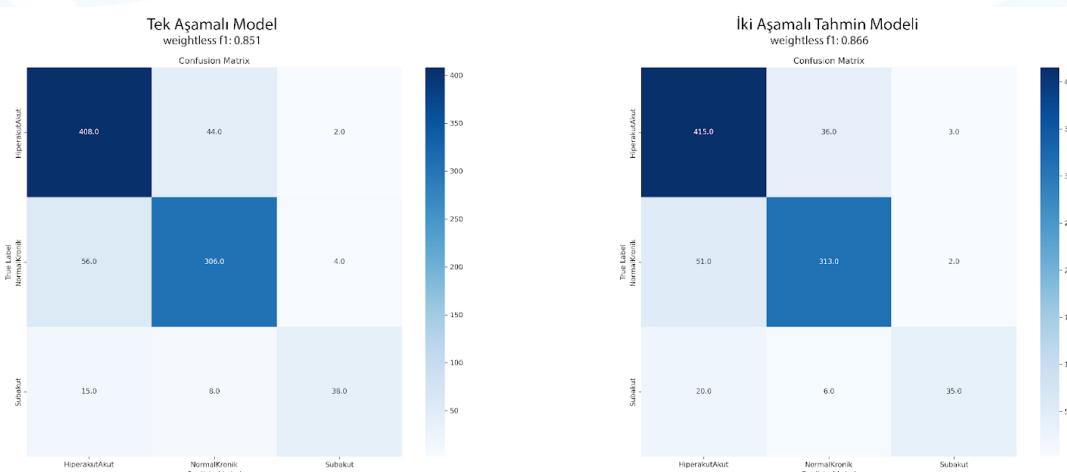
Bununla birlikte, daha yüksek doğruluk hedefleri doğrultusunda eğitim süresi daha uzun olsa da YOLO-m modeli ana model olarak tercih edilmiştir.

Kullandığımız modellerin incelemesi şu şekilde sunulmuştur: Ana tahmin modeli sonuçları ve eğitim eğrileri, iki aşamalı sınıflandırma ile confusion matrix'lerde iyileşme.



Şekil 5

İki aşamalı yöntem için eğittiğimiz ekstra 3 modelin eğitim eğrilerini maalesef yer kalmadığı için ekleyemiyoruz. Ancak confusion matrix sonuçları Şekil 6'da gösterilmekte:



Şekil 6

5. ÖZGÜNLÜK

Google ViT

Projemiz için en uygun mimariyi belirleme hedefimiz doğrultusunda, bilgisayarlı görü alanında büyük başarılar elde etmiş olan Transformer tabanlı modelleri de değerlendirdik. Bu kapsamda özellikle ön eğitimli ağırlıklara kolay erişim ve geniş bir literatür desteği sunan Google Vision Transformer (ViT) modeli üzerinde testler

gerçekleştirdik. Yaptığımız analizlerde, ViT'nin standart sınıflandırma başlıklarıyla dahi umut verici sonuçlar ürettiğini ancak en yüksek performansa ulaşması için medikal görüntülere özel tasarlanmış daha sofistik 'boyun' ve 'kafa' bileşenleri gerektirdiğini tespit ettik. Projemizin mevcut takvimi ve kaynakları dahilinde daha hızlı sonuçlar üreten ve iterasyona daha açık olan YOLOv11 mimarisine öncelik verilmiştir. Transformer tabanlı bu ileri düzey mimarilerin entegrasyonu ise projemizin final aşaması için planlanan önemli bir geliştirme hedefi olarak belirlenmiştir.

YOLOv11

Sistemimizin ana motoru olarak en güncel nesil tespit ve sınıflandırma mimarilerinden olan YOLOv11 serisi tercih edilmiştir. Literatürdeki birçok çalışmanın daha eski mimarilere odaklanmasıının aksine YOLOv11'in sunduğu modern yapı, daha az kaynakla daha hızlı eğitim ve yüksek doğruluk sağlamıştır.

İki Aşamalı Teşhis

Projemizin en temel özgünlüğü, standart 'tek seferde teşhis' yaklaşımının ötesine geçen hiyerarşik ve güven odaklı bir karar mekanizması geliştirmemizdir. Bu sistemde, ana sınıflandırma modelimiz (YOLOv11) bir tahmini düşük bir güven skoruyla ürettiğinde bu belirsiz vaka bir kenara atılmaz. Bunun yerine en olası iki sınıf için özel olarak eğitilmiş, daha hassas ikili sınıflandırma modellerinden oluşan bir uzmanlar kuruluna yönlendirilir. Bu yaklaşım, özellikle birbirine çok benzeyen subakut ve kronik gibi evrelerin ayrimında teşhis doğruluğunu ve sistemin genel güvenilirliğini belirgin şekilde artırmaktadır.

Google MedGemma ile Açıklanabilir Teşhis

Sağlıkta yapay zekanın klinik kabulündeki en büyük engellerden biri olan 'kara kutu' problemini aşmak amacıyla, sistemimize bir açıklanabilirlik katmanı entegre ettik. Bu katman için, Google'in Mayıs 2025'te kullanımına sunduğu en yeni medikal çok modlu modeli olan MedGemma kullanılmıştır. Sistemimiz nihai teşhisini koymuktan sonra ilgili MR görüntüsü ve teşhis sonucu MedGemma'ya sunulur. MedGemma, bu teşhisin neden konulduğunu, görüntüdeki hangi anormalliliklere dayandığını klinik bir dil ile açıklayan bir rapor üretir. Bu özgün kullanım, modelimizin sadece 'ne' söylediğini değil, 'neden' söylediğini de ortaya koymak hekim-yapay zeka iş birliğine ve hasta iletişimine doğrudan katkı sağlar.

Difüzyon Modelleri ve Sanal Veri Oluşturma

Her ne kadar eğitimde kullanamamış olsak da MR görüntülerinden difüzyon modelleri ile sanal veri oluşturulabileceğini test ettilik. Bunun kullanım alanları ve eğitime entegrasyonu hakkında çalışmalarımıza devam etmekteyiz.

6. REFERANSLAR

Ho, K. C., Speier, W., Zhang, H., Scalzo, F., El-Saden, S., & Arnold, C. W. (2019). A Machine Learning Approach for Classifying Ischemic Stroke Onset Time From Imaging. *IEEE transactions on medical imaging*, 38(7), 1666–1676. <https://doi.org/10.1109/TMI.2019.2901445>

Kim, P.J., Kim, D., Lee, J. et al. Deep learning-based classification of diffusion-weighted imaging-fluid-attenuated inversion recovery mismatch. *Sci Rep* 15, 5924 (2025). <https://doi.org/10.1038/s41598-025-90214-w>

Lee, K. Y., Liu, C. C., Chen, D. Y., Weng, C. L., Chiu, H. W., & Chiang, C. H. (2023). Automatic detection and vascular territory classification of hyperacute staged ischemic stroke on diffusion weighted image using convolutional neural networks. *Scientific reports*, 13(1), 404. <https://doi.org/10.1038/s41598-023-27621-4>

Zhang, X., Luan, Y., Cui, Y., Zhang, Y., Lu, C., Zhou, Y., Zhang, Y., Li, H., Ju, S., & Tang, T. (2025). SDS-Net: A Synchronized Dual-Stage Network for Predicting Patients Within 4.5-h Thrombolytic Treatment Window Using MRI. *Journal of imaging informatics in medicine*, 38(3), 1681–1689. <https://doi.org/10.1007/s10278-024-01308-2>

Zhang, Y. Q., Liu, A. F., Man, F. Y., Zhang, Y. Y., Li, C., Liu, Y. E., Zhou, J., Zhang, A. P., Zhang, Y. D., Lv, J., & Jiang, W. J. (2022). MRI radiomic features-based machine learning approach to classify ischemic stroke onset time. *Journal of neurology*, 269(1), 350–360. <https://doi.org/10.1007/s00415-021-10638-y>