

1. Aşama Değerlendirme Soruları



Cüneyt Taha Çoban

Kavakpınar Mah. Tarkan Sok. No:5A Pendik İstanbul

+90 5319175647

cuneytcbnx@gmail.com



1. Değerlendirme Soruları Cevapları

1.1. Hangi veri setini kullandınız, ilgili veri seti için öznitelikler nelerdir, açıklayınız.

Cevap: Thyroid digital image database (TDID) veri setini kullandım. Veri setinin öznitelikleri Number, Age, Sex ve Tirads. Sağlık sektörü hakkında çok bilgim olmadığı için cinsiyet ve yaş öznitelikleri es geçtim. Number ve tirads niteliklerini kullandım.

1.2. Bulduğunuz sonucu raporlayarak, yorumlayın. Daha iyi bir model sonucu elde etmek için neler yapılabilir?

Cevap: Bulduğum sonuçta oluşturduğum model çok overfitting yaptı buda veri setinden kaynaklı olduğunu düşünüyorum. Daha çok verinin olduğu daha yüksek çözünürlük ve eşit dağılımlı veriler ile çok daha iyi sonuçlar alınabilir. Kullandığım veri setinde 58 benigns(iyi huylu) 282 maligns (kötü huylu) görüntü vardı buda modelin sadece maligns öğrenmesine sebep oldu ve genel olarak ezmiş oldu.

Eğitimde görüntü giriş boyutunu artırıp epoch ve step sayılarını artırıp daha derin bir CNN ağı oluşturup daha iyi sonuçlar elde edilebilir.

Eğitim verilerinde iyi huylu ve kötü huylu sınıflar arasındaki dengesiz veri örnekleri problemini çözmek için, `sparse_categorical_crossentropy` loss fonksiyonu yerine `weighted_binary_crossentropy` loss fonksiyonu kullanılabilir.

Sınıflandırma problemi için girdi ultrason tiroid görüntülerini derinlemesine analiz etmek için birden fazla CNN tabanlı model kullanılabilir. Her bir CNN modeli kendi mimarisine ve girdi görüntülerinin özelliklerini öğrenme özelliklerine sahip olduğundan, birden fazla CNN tabanlı modelin kullanılması, tek bir modelin kullanılmasına kıyasla daha zengin bilgilerin çıkarılmasına yardımcı olabilir.

Not: İlk sonuçları gördükten sonra veri setini data augmentation işlemleri ile büyüterek 290 benign görüntüsü elde ettim. Bunun 240'ını eğitim için 50 tanesini doğrulama için kullandım. Bunun sonucunda overfittingi biraz engellemiş oldum github paylaştığım sonuçlar veri setini büyüttükten sonra yapılan eğitimin sonuçlarıdır.

1.3. Bu soruda iyi bir sonuç için sizi kısıtlayan aşamalar nelerdir, farklı bir model kullanmak isteseydiniz hangi model/leri kullanırdınız sebebi ile açıklayınız?

Cevap: Daha derin ağları kullanabilecek donanımda bir bilgisayara sahip olmadığım için bu bana büyük bir engel teşkil ediyor. Önceki soruda belirttiğim gibi birden fazla CNN tabanlı modelin kullanılması daha iyi sonuç verebilir fakat birden fazla CNN ağını çalıştırmak için güçlü donanım ekipmanı gerektirir.

Beni kısıtlayan aşamalardan bir tanesinde veri setindeki verilerin az ve iyi görüntüye sahip olmamasıydı. Sınıflar arasındaki veri sayısı farkı çok fazlaydı. Daha büyük, eşit dağılmış ve daha düzgün bir tiroid ultrason görüntüsü olan bir veri seti bulamadım. Verileri daha düzgün hale getirmek için opencv kütüphanesi ile kenar bulma fonksiyonundan yararlanarak görüntülerdeki siyah arka planı kaldırdım ve çeşitli augmentation işlemleri ile veri setini büyüttüm.

Farklı bir model kullanmak istesem Resnet50 ağını kullanabilirdim. Nedeni VGGNet, AlexNet, ResNet, DenseNet gibi ağlar yaklaşık 100-150 milyon parametre içerir. Bu büyük miktardaki parametreleri öğrenmek için güçlü bir donanım gücü gerekir (Merkezi İşleme Birim (CPU), GPU) yanı sıra büyük miktarda eğitim verisi de gerektirir. Normalde daha az eğitim verisi kullanan yüksek performanslı sistemlere ihtiyacımız olduğundan, bu sorunların tıbbi görüntü tabanlı sistemlerin performansı üzerinde güçlü olumsuz etkileri olabilir.

Resnet50 çok derin bir CNN ağı oluşturmak için değil, aynı zamanda artık ağı (ResNet) eğitmeyi de kolaylaştırmak için yeni bir yöntem önermektedir. Bağlantıyı atlama adı verilen yeni bir bağlantı türü kullanarak, bu yeni CNN mimarisi türü, bu katmanların giriş ve çıkışı tanımlama işlevine yakın olduğunda ağın bazı eğitim katmanlarını atlamasını sağlayabilir. Sonuç olarak, ağ, geleneksel CNN ağlarına kıyasla daha derin ve eğitilmesi daha kolaydır.

Not: Tıp alanında çok bilgiye sahip olmadığım için bu önerilerde bulundum. Eğer tıp alanında uzman biriyle çalışma imkanım olsa iyi huylu ve kötü huylu tiroid nodüllerinin görüntü üzerinde neye benzediğini algılayıp, klasik görüntü işleme yöntemleri ile daha iyi bir bilgi çıkarımı yapabileceğimi düşünüyorum. Örneğin iyi huylu ve kötü huylu tiroid nodülleri farklı şekil yapılarına sahip olması, renk farklılıkları gibi özellikler daha iyi bilgi çıkarımı yapmamı sağlayabilirdi.

Eğer görüntülerdeki tiroid nodüllerini kendi gözüm ile algılaya bilseydim ve bu nodülleri görüntü üzerinde etiketleyebilecek bir bilgiye sahip olsaydım veri etiketleme işlemi gerçekleştirip Mask R CNN gibi segmentasyon algoritmasıyla ultrason görüntüleri üzerindeki nodülleri tam olarak tespit edebilirdim. Bunun sayesinde ultrason sonucuna bakacak kişi veya kişilerin daha rahat bir şekilde nodülleri görmesi sağlanabilirdi.

1.4. Tirads skoru kullanarak yapacağınız bir sınıflandırma modelinde, modelinizin sonuçları ile, etiketli benign/malign teşhisi arasında uyumsuzluklar varsa, bu uyumsuzlukları nasıl karşılaştırırsınız, çözüm için öneriniz ne olur?

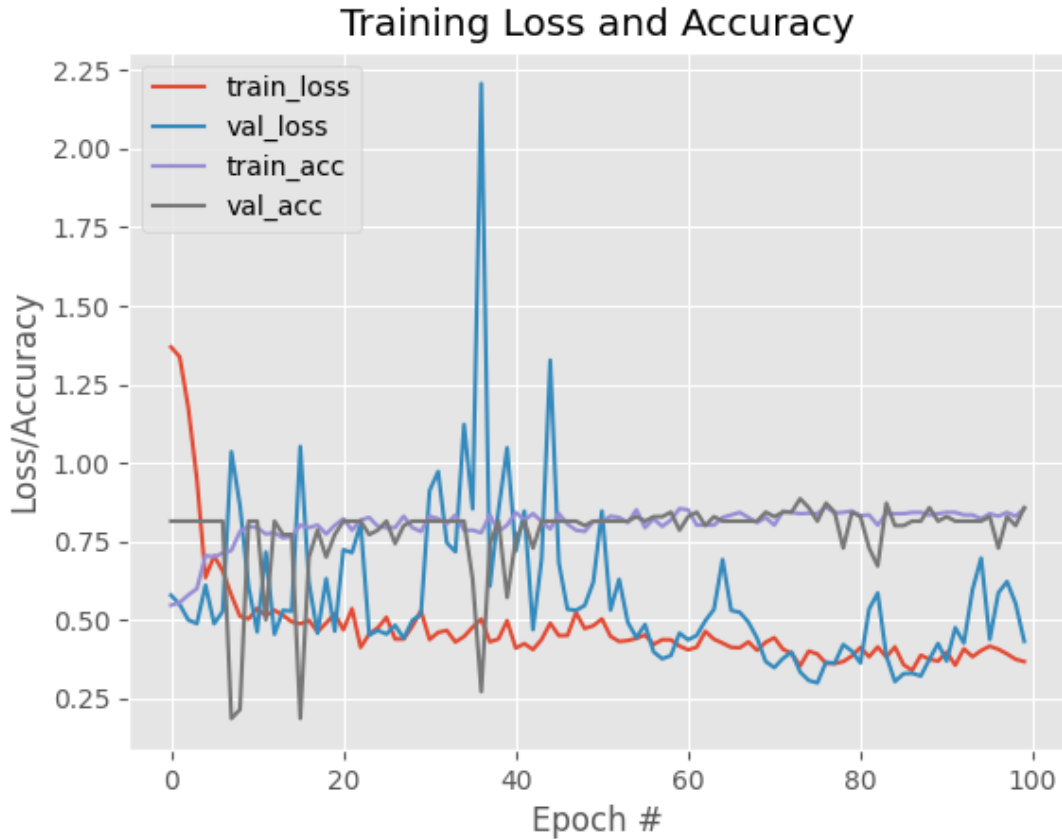
Cevap: Uyumsuzlukları karşılaştırmak için eğitimi başlatmadan önce veri setindeki görüntülerden bazılarını test için ayırmamız gerekiyor. Ben 3 benign 5 tane de malign görüntüsü olmak üzere toplamda 8 test görüntüsü ayırdım. Eğitimde kaydetmiş olduğumuz model ağırlığını yükleyerek test için ayırdığımız görüntüleri tespit ettirip, tespit sonucunda elde edilen etiketlerin doğru olup olmadığını karşılaştırabiliriz. Eğer uyumsuzluk varsa tahminler yanlış ise bunu düzeltmek için K Fold Cross Validation yapılabilir.

K Fold Cross Validation, makine öğrenmesi modelinin görmediği veriler üzerindeki performansını mümkün olduğunca objektif ve doğru bir şekilde değerlendirmek için kullanılan istatistiksel bir yeniden örnekleme(resampling) yöntemidir. K Fold Cross Validation çalışma mantığı özetle şöyledir:

- Veri seti rastgele olacak şekilde karıştırılır.
- Veri seti k gruba ayrılır.
- Seçilen grup validasyon seti olarak kullanılır.
- Diğer tüm gruplar (k-1 grup) train seti olarak kullanılır.
- Train seti kullanılarak model kurulur ve validasyon seti ile değerlendirilir.
- Modelin değerlendirme puanı bir listede saklanır.
- Son olarak değerlendirme puanlarının istatistiksel özetine bakılır. (ortalama, standart sapma, maksimum, minimum.. gibi.)

2. Model Sonuçlarının Görselleştirilmesi

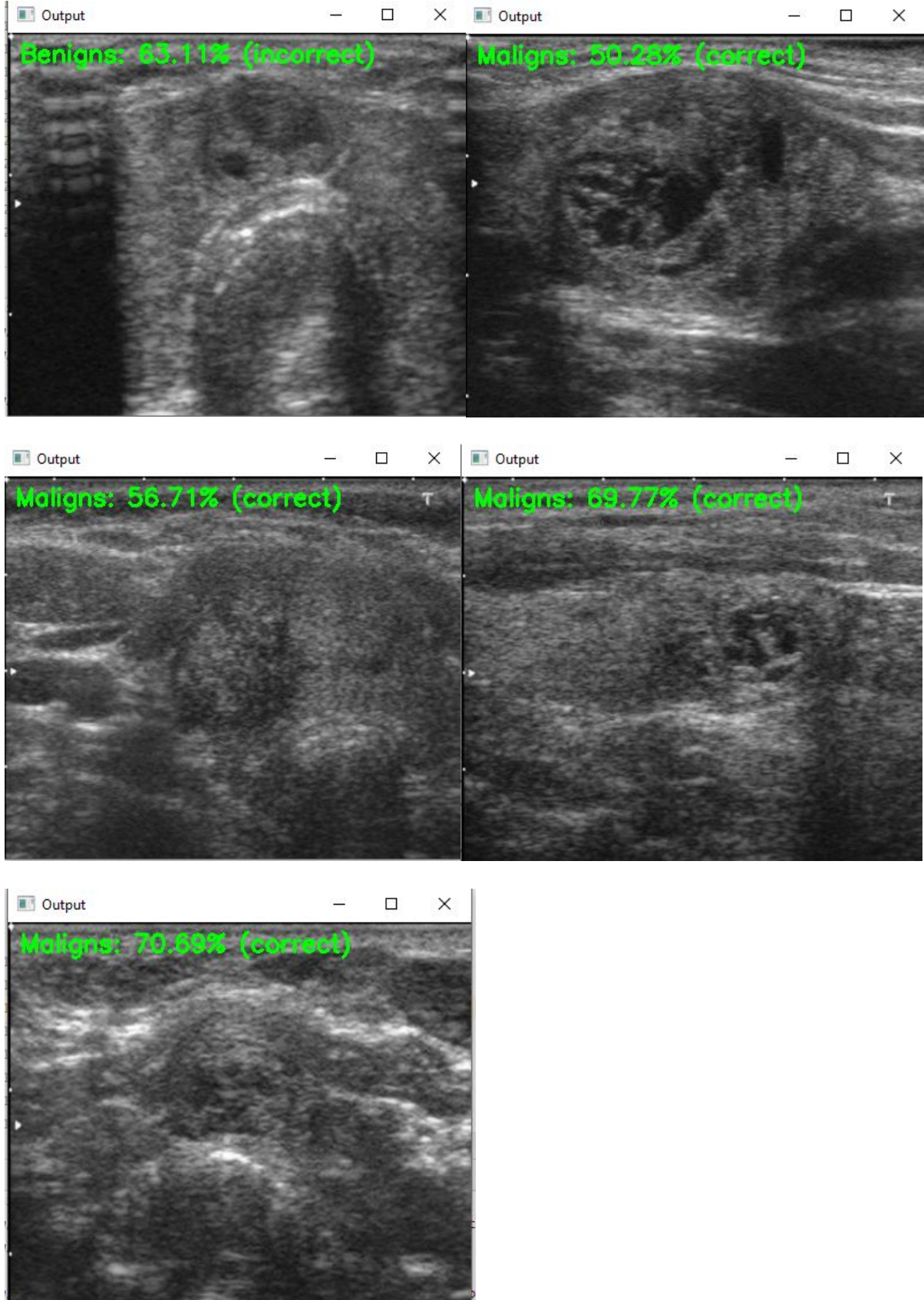
2.1 Train, Validation Loss Ve Accurarcy Değerleri



2.2 F-score Değerleri

	precision	recall	f1-score	support
Benigns (Class 0)	0.70	0.52	0.60	50
Maligns (Class 1)	0.62	0.78	0.69	50
accuracy			0.65	100
macro avg	0.66	0.65	0.64	100
weighted avg	0.66	0.65	0.64	100

2.3 Malign Test Görüntülerinin Sonuçları



2.4 Benign Test Görüntülerinin Sonuçları

