

CEVAPLAR

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

Soruda verilen verisetini kullandım. Onun haricinde internetten başka veri bulamadım.

Modelde Transfer Learning metoduyla VGG16 feature extractor algoritmasını kullandım.

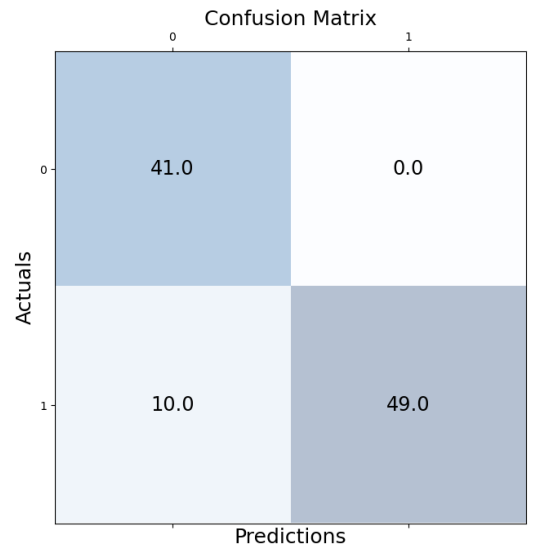
Bu modelin ilk layerlarındaki öznitelik çıkarıcıları bilindiği üzere low level olan özellikler yani temel öznitelikler bunlar çizgi tespiti gibi özellikleri çıkaran katmanlardır. Her resimde olan bu özelliklerle oynamanın önemi yoktu zaten imagenet verisiyle eğitilmiş layerlardı. Fakat son layer yani üst katman layerlar ise daha genel öznitelikleri çıkaran layerlar olduğu için bu kısımlardaki öznitelikleri yeniden eğitmek gerekecekti. Çünkü Imagenet verisetinde bizim verilerimizi içeren bir sınıf bulunmamaktadır.

Böylece modelin başarısı artırıldı.

2-İstedığınız yapay öğrenme modelini kullanarak oluşturacağınız bir sistem tasarlayın ve kaynak kodunuzu githuba yükleyin. Bu aşamada performans metriklerinizi (accuracy, f-score...) listeleyin. Görselleştirme için hazır toollardan yararlanabilirsiniz

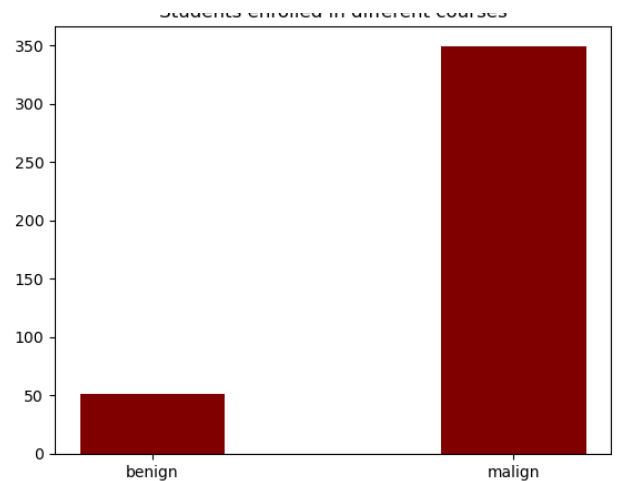
Accuracy: 0.8981881188118812
Precision: 1.0
Recall: 0.8293845504913915
F_score: 0.9058410944430779

KFOLD OLARAK EĞİTİLDİĞİNDEN GÖRSELLEŞTİRME KISMI BİRAZ EKSİK KALDI



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

Doğruluk oranımız %90 olarak oluşturuldu bu ortalama bir değerdir çünkü modelimizi kfold yöntemiyle eğittik ki tüm veriler eğitim kısmına katılsın. Burada Precision belki biraz yüksek çıkmış olabilir sanki Malign yani 1 sonucuna biraz daha meyilli olabilir ama imbalanced veriseti olması bence bunda çok etkili verisetinin dağılımında resimde görülecektir. Ben bu sorunun etkisini azaltmak için verisetinin benign sınıfından olan verilerini oversampling yoluyla artırım yoluna gittim bu da başarıyı artırmıştı zaten.



4-Bu soruda iyi bir sonuç için sizi kısıtlayan aşamalar nelerdir, farklı bir model kullanmak isteseniz hangi model/leri kullanırdınız sebebi ile açıklayınız?

Bunun dışında asıl yapılması gerek SOTA sınıflandırma algoritmalarından yararlanmak olacaktır. EfficientNet ya da Resnet gibi aslında resneti kullandım Transfer Learning metoduyla fakat high level layerları eğitmeden yüksek doğrulukta öznetelik çıkarımı yapamıyor. Vgg nin high levellerı eğitilmiş hali daha yüksek başarı sağlamıştı. Resnette high level feature extractor layerları eğitmeye kalktığımda ise laptopum dondu☹. Bu yüzden eğitemedim çünkü parametre sayısı inanılmaz yükseldi.

Tabi bazı preprocess işlemleride önemli . Veriyi modele sokmadan önce bir Denoising işleminden geçirdim. Bu da Opencv kütüphanesinden bir denoising fonksiyonu bu da başarıyı bir miktar yükseltti ancak daha iyi bir denoising işlemi yapılabilir bunun için decoder ve encoderlardan yararlanılması gerekebilir.

5-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?

Ben aslında tirads skorları üzerinden bir sınıflama gerçekleştirdim ilk amacım aslında hem görüntüden gelen bir sınıflandırma değeri ile hem de istatistiksel verilerden gelen (xml dosyası içinde bulunan) veriyle desteklenmiş bir ensemble learning algoritması yapmayı düşünmüştüm ama istatistik verisinden gelen sınıflandırma sonucuna pek güvenemedim ki orada xgboost algoritmasıyla bir sınıflandırma yapmıştım %98 başarı gelince o verilerin teşhisin kendisiyle alakalı veriler olduğunu düşündüm. Bu yüzden bunu yapmadım. Sorunun ilk kısmına gelecek olursak belki sınıflandırmaya bir değer daha eklersek ki bu da confidence olarak sonuç çıkarılacak burdan gelen veriye göre tirads skoru belirlenebilir. Yani atıyorum %98 oranında malign sonucunu veriyorsa bu 5. Seviye malign olabilir diye yorumlanabilir çünkü araştırdığım kadarıyla 2-3-4a-4b gibi sonuçlar kistin iyi kötü huylu olma olasılığına işaret ediyor.