Derin Öğrenme ve Uygulamaları

Final Projesi

1. **Giriş**

Derin Öğrenme ve Uygulamaları dersinin final projesinde bir derin öğrenme uygulaması kullanılmasına karar verilmiştir. Bu uygulama için kaggle platformundan bir akciğer CT görüntüsü içeren veriseti seçilmiştir. Veri seti, kontrast kullanımı ve hasta yaşı ile ilişkili CT görüntü verilerindeki eğilimleri incelemek için farklı yöntemlerin test edilmesine izin verecek şekilde tasarlanmıştır.

1. **Yöntem**
   1. **Veri Ön Hazırlık**

Veri seti içerisinde 100 tane CT görüntüsü bulunmaktadır. Bunları yarısı kanser hastası, yarısı sağlıklı olarak etiketlenmiştir. Veriler 512 x 512 formatında gelmektedir. Bunun modelimize uygun hale getirmek için 128 x 128 e yeniden şekillendirilmesi uygun görüldü.

Yeniden şekillendirme sonrası verilerin etiketleri için bir dataframe oluşturuldu. Bu etiketler dicom dosyalarının isimlerinden elde edildi. Daha sonrasında bu etikteler 0 kanser değil, 1 kanser olmak üzere tamsayı olarak dataframe’de muhafaza edildi.

A picture containing chart

Description automatically generated

**Şekil 1.** *Örnek Bir CT Verimiz*

* 1. **Eğitim ve Test Verisi Oluşturma**

Veriseti modeller üzerinde işlenmeye başlanmadan önce tahmin edilmesi hedeflenen kanser olup olmama durumunu bir veri yapısında, CT görüntüleri ise başka bir veri yapısında tutuldu. Sonrasında veriler %90 eğitim ve %10 test verisi olmak üzere bölündü.

* 1. **Modeller Seçimi**

CT görüntülerimiz için LeNet algoritması kullanılması uygun görülmüştür. Literatürde de birçok CT görüntü işleme örneğinde bu model denenmiş ve başarı yakalanmıştır. (Detection of COVID 19 from CT Image by The Novel LeNet-5 CNN Architecture) Covid 19 vakalarından elde edilen akciğer CT görüntülerinde bu model kullanıldığında yakalanan başarılar ve modelin sade oluşu seçimde etken rol oynamıştır.

A picture containing diagram

Description automatically generated

**Şekil 2.** *LeNet Modelinin Bu Proje için Uyarlanmış Şeması*

Şekil 2.’de görüldüğü gibi modelde 2 konvolüsyon katmanı, 2 max pooling katmanı bir fully connected ve 3 dense katmanı bulunmaktadır.

Text, letter

Description automatically generated

**Şekil 3.** *Modelin Google Colabda Oluşturulması*

Şekil 3.’de modelimizin katman detaylarının python dili ile işlenişi görülmektedir. Şekil 2.’de görünmeyen ayrıntılar, dense katmanlarının ilk ikisinde ReLu aktivasyon fonksiyonunun kullanılması ve son dense katmanında softmax aktivasyon katmanının kullanılmasıdır.

Table

Description automatically generated

**Şekil 4.** *Modelin Öğrenilebilir Parametreleri*

Şekil 4.’te modelimizin öğrenilebilir parametre sayısı 350592 olarak gözükmektedir.

* 1. **F1 Skoru**

**Chart

Description automatically generated**

**Sekil 5.** *Karmaşıklık Matrisi Örneği*

Bu matristen elde edilebilecek bilgiler;

**Gerçek Pozitifler (TP):** Bunlar gerçek değeri 1 ve tahmin ettiğimiz değerin de 1 olduğu örneklerdir.

**Gerçek Negatifler (TN):** Bunlar gerçek değeri 0 ve tahmin ettiğimiz değerin de 0 olduğu örneklerdir.

**Yanlış Pozitifler (FP):** Bunlar gerçek değeri 0 ancak tahmin ettiğimiz değerin 1 olduğu örneklerdir.

**Yanlış Negatifler (FN):** Bunlar gerçek değeri 1 ancak tahmin ettiğimiz değerin 0 olduğu örneklerdir.

**Doğruluk Oranı (Accuracy Rate):** Genel olarak, sınıflayıcının ne sıklıkta doğru tahmin ettiğinin bir ölçüsüdür.

(TP + TN) / TOPLAM

**Gerçek Pozitif Değerlerin Oranı (True Positive Rate):** Sınıflayıcının ne kadar gerçek pozitif değeri doğru tahmin ettiğinin bir ölçüsüdür. Hassasiyet, İsabet Oranı veya Hatırlama olarak da bilinir. (Sensitivity, Hit Rate or Recall) Mümkün olduğu kadar yüksek olmalıdır.

TP / GERÇEK POZİTİFLER

**Hassasiyet (Precision):** Tüm sınıflardan, doğru olarak ne kadar tahmin edildiğinin bir ölçüsüdür. Mümkün olduğu kadar yüksek olmalıdır. Pozitif Tahmin Edici Değer olarak da bilinir.  (Positive Predictive Value)

TP / TP + FP

**F Puanı (F Score):** Bu, gerçek pozitif değerlerin oranının (recall) ve hassasiyetin (precision) harmonik ortalamasıdır. Sınıflandırıcının ne kadar iyi performans gösterdiğinin bir ölçüsüdür ve sınıflandırıcıları karşılaştırmakta sıklıkla kullanılır.

2 \* Hassasiyet \* Gerçek Pozitif Değerlerin Oranı

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Hassasiyet + Gerçek Pozitif Değerlerin Oran

Şekil 5.’te 10 tahminden 9’unun doğru olduğu gözlemlenmiştir. Veri setimizin genişliği göz önünde bulundurulunca iyi sayılabilecek bir değerdir. Matris verilerne göre F1 skoru 0.857 olarak hesaplanmaktadır.

1. **Sonuç**

Kısıtlı veri setine göre %90 lara yaklaşan tahmin oranı ve 0.857 olan f1 skoru ile LeNet modelinin başarılı olduğu ama daha çok veri ve hiperparamete ayarları ile daha net sonuçlar elde edilebileceği gözlemlenmiştir.