

2023

TEKNOFEST

HAVACILIK, UZAY VE TEKNOLOJİ FESTİVALİ

SAĞLIKTA YAPAY ZEKÂ YARIŞMALARI

(Bilgisayarlı Görüyle Tarama Mamografilerinde BI-RADS

Kategorisi ve Meme Kompozisyonu Tahmini Yarışma

Kategorisi)

PROJE DETAY RAPORU

TAKIM ADI: ENGINEER DOCS

TAKIM ID: 151359

| YARIŞMA SEVİYESİ | |
|--------------------------------------|---|
| <input type="checkbox"/> ORTAÖĞRETİM | <input checked="" type="checkbox"/> YÜKSEKÖĞRETİM |

İçindekiler

| | |
|--|----------|
| 1. Proje Mevcut Durum Değerlendirmesi | 3 |
| 2. Özgünlük | 4 |
| 3. Sonuçlar ve İnceleme | 7 |
| 3.1. BI-RADS Sonuçları | 7 |
| 3.2. Kompozisyon Sonuçları | 7 |
| 4. Deney ve eğitim aşamalarında kullanılan veri setleri | 7 |
| VinDr-Mammo Veri Seti | 7 |
| RSNA Veri Seti | 8 |
| 5. Referanslar | 8 |



1. Proje Mevcut Durum Değerlendirmesi

Öncelikli olarak Proje Sunuş Raporu için yaptığımız kapsamlı literatüre taramasından projenin deney aşamasında çokça faydalandık. Raporumuzda belirttiğimiz üç ana çalışma ve o çalışmaları bulurken karşılaştığımız teknikleri detaylı bir şekilde analiz ettik. Analiz sonucu gösterdi ki belirttiğimiz metot ve algoritmalar genel hatlarıyla uygun ve gereklidir.

Sunuş Raporunda öngördüğümüz veri ön işleme bu projenin olmazsa olmazı olduğu yaptığımız deneyler sonucunda da tasdiklenmiş oldu. Belirttiğimiz gibi açık kaynaklı benzer data-setleri etkin şekilde kullandık ve diğer değinilen metotları da öngördüğümüz şekilde kullandık. Kullanım detaylarımızı Özgünlük başlığı altında bulabilirsiniz.

Yaptığımız araştırmalar sonucunda öznetelik çıkarma işleminin vazgeçilmez olduğuna karar verdik ve uygulanacak metotlar eklemiştik. Sunuş Raporuna spesifik olarak VGG-16 yapısını kullanmayı öngörmüştük. Ancak deneylerimizi VGG-16'le sınırlamaya karar verdik. Çünkü aldığımız sonuçlardan tatmin olmadık. Ayrıca saf VGG-16 yapısı kullanımı projeye uygun gelmediğinden önceden eğitilmiş (VGG-16 benzeri) yapıların kombine edilmiş hali olan kendimize özgü bir yapı oluşturduk ve buna 4HeadedBreast-Net ismini verdik. Bu yapının detaylarını Özgünlük başlığı altında bulabilirsiniz.

Aynı şekilde sınıflandırıcı da bu projenin en önemli etmenlerinden biridir. Proje Sunuş Raporunda Decision Tree tabanlı algoritmalarının Multi Layer Perceptron'lardan daha iyi sınıflandırma yaptığını belirtmiştik. LightGBM yapısının kullanılmasını planlamıştık. Ancak deney sonuçlarımız bu bilginin bizim projemizde tam olarak uygulanabilir olmadığını gösterdi. Bu nedenle modelimizin son halinde MLP yapısı sınıflandırıcı olarak tercih edildi. Benzer şekilde bu yapının detayları Özgünlük başlığı altında bulunabilir.

Sunuş Raporunda belirttiğimiz topluluk öğrenmesi metodunu uygulamadık. Çünkü geliştirdiğimiz modelin yapısı büyük oldu ve benzer yapıdaki büyük modellerden tahmin almaya çalışmak yarışma sırasında bize verilen kısıtlı süreyi aşacağını öngördük.

Son olarak yarışma içeriğine dahil olan ve metotlarda koyduğumuz konum bilgisi çıkarıcı üstüne bu rapora deney yapamadık. Bunun en önemli sebebi zamanımızın kısıtlı olmasıydı. BI-RADS ve kompozisyon deneylerimin sonucu beklentilerimin altına kaldığı için öncelikle onları iyileştirmeye odaklandık ve zamanımızı onun üzerinde harcadık.

Sunuş raporunda belirttiğimiz gibi donanım olarak Google Colab'ı tercih ettik. Genellikle Tesla T4 kullanarak modellerimizi eğittik. Ayrıca birden fazla Colab kullanarak deneyleri paralel olarak devam ettirdik. Sunuş Raporuna ek olarak ön işlemeyi Google Cloud'daki bir sanal makinede gerçekleştirdik.

Nitelik çıkarıcı olarak birçok yapı denenmiştir. Bunlar Efficient-Net, Mobile-Net-v2, Mobile-Net-v3, VGG-16, VGG-19, ViT, Res-Net, AlexNet, ResNeXt, RegNet, DenseNet, Inception-v3, Inception-ResNet-v2 ve ShuffleNet-v2'dir. Bunlardan Efficient-Net ve Mobile-NetV2 en iyi sonuçları elde ettiğimiz yapılarıdır. Bahsi geçen yapıların farklı versiyonları ve bunların 4HeadedBreast-Net içindeki diğer yapıların farklı versiyonları ile kombine edip deneyler yaptık. 4HeadedBreast-Net'in son halini ve içinde bulunan son yapının detaylarını özgünlük kısmında bulabilirsiniz.

Sınıflandırıcı için birkaç farklı algoritmanın değişik kombinasyonlarını denedik. Bu algoritmalar LightGBM, XGBoost ve MLP'dir. LightGBM ve XGBoost eğitimleri sonucunda iyi ayarlanmış MLP'den biraz daha kötü sonuçlar verdi. Bunun bir önemli sebebi elimizdeki makinalarda bu algoritmaları eğitim ve onlara MLP'de yaptığımız gibi çapraz doğrulama algoritmaları uygulayamadık. Çünkü bu algoritmalar CPU tabanlı eğitime uygun olarak

geliştirildiği için hızlandırmak için paralel işleme gereklidir ve bu bizim sahip olduğumuz makinelerde mümkün olmadı. O yüzden XGBoost ve LightGBM'i en etkili haline getiremedik. Onun yerine uygulanabilirliği daha iyi olan MLP'yi tercih ettik. Bu MLP'yi çapraz doğrulama algoritması ile parametre ayarlamasını yaptık. MLP denemelerimizde farklı katman sayıları, nöron sayıları ve dropout olasılıkları denedik Çıkış katmanı için Softmax, Logsoftmax denedik.

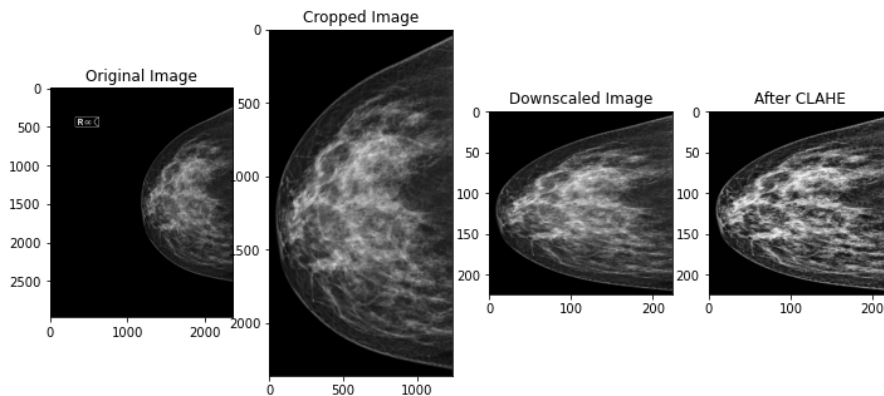
2. Özgünlük

Yapay zeka projelerinin en önemli kısımlarından bir tanesi verilerin ön işlenmesidir. Veri kalitesini arttırmak, olası hatalı verileri kaldırmak ve kullanılacak veriyi ön plana çıkarmak, yapay zeka modelinin veriyi daha hızlı ve kapsamlı öğrenmesine oldukça büyük katkı sağlar. [1] Bu yüzden, herhangi bir model uygulamaya başlamadan önce elimizdeki verileri olabildiğince sadeleştirip önemli kısımlarını ön plana çıkarmaya çalıştık.

Veri setinin boyutu en büyük problemlerden birisiydi. Yaklaşık 300 Gb yer kaplaması sürekli olarak verinin taşınımını zorlaştırıyordu. Resimlerin çok büyük bir kısmı siyah arka plandan oluşuyordu. Bu yüzden hem modelin gereksiz yerlerden fikir çıkarmaya çalışmasını engellemek hem de boyutunu küçültmek için sadece memenin kapladığı alanı kesmeye karar verdik. Burada memenin kapladığı alanı bulmak için Otsu Thresholding algoritmasını kullandık. Otsu Thresholding, bir görüntüdeki optimum eşik değerini otomatik olarak hesaplayarak, pikselleri iki gruba ayıran bir görüntü işleme algoritmasıdır. [2] Bir nevi resmi kümeleme algoritması ile iki gruba ayırıyor. Bu algoritma resimleri neredeyse orijinal boyutlarının dörtte birine düşürdü.

Medikal resimler kırıldıktan sonra Kontrast Sınırlı Uyarlamalı Histogram Eşitleme (CLAHE) metodu ile iyileştirilmiştir. CLAHE yöntemi, bir görüntünün kontrastını arttırmak için kullanılan bir işleme yöntemidir. Görüntüdeki piksel değerlerinin histogramını eşitler ve ardından adaptif bir şekilde sınırlar, böylece kontrast artar. [3] Bu sayede kanserli olabilecek bölgeler daha belirgin hale geliyor. Buda Convolutional Layers'dan geçtikten sonra belirgin hale gelen kısımların daha yüksek değerlere sahip olmasını sağlıyor.

Ön işlemenin son adımında çoğu öznetelik çıkarma (feature extractor) modelleri giriş katmanında resimleri 3x224x224 şeklinde alıyor. Bu nedenle, tüm resimler Lanczos algoritması ile küçültüldü. Dicom formatında olan resimleri PNG formatında kaydedildi. PNG resimlerin kayıpsız sıkıştırılmasını sağlıyor. Bu ön işleme adımları sayesinde hem veri yaklaşık 300 Gb den 481 Mb ye düşürüldü, hem de resimlerin sadece önemli kısımları alınıp belirginleştirildi. Aşağıdaki figür ön işleme adımlarını sırasıyla göstermektedir.



Figür 1: Ön İşleme Aşamaları Sırası İle

Ön işleme aşaması tamamlandıktan sonrasında örnek makalelerdeki model yaklaşımlarına baktık. Bu modellerin her birinden bir fikir çıkararak kendi modelimizi oluşturduk (4HeadedBreast-Net). Öncelikle genel olarak 4 resimli veri hem internetteki veri setlerinin hepsinde mevcut olmamasından dolayı hem de hesaplama kapasitesinden dolayı bir yada iki resim kullanıyordu. Biz tüm verileri kullanmak için 4 görüntü açısını da girdi olarak alıp her birini ayrı işleyen bir model oluşturduk. Bütün görüntü açıları öznetelik çıkarıcı modellerden geçirilerek dört adet gömülmüş öznetelik elde edildi.

Bu dört veriyi tek bir veri haline getirmek için birleştirme algoritması kullandık. Genel yaklaşımlar gömülmüş öznetelik çıktıları düzleştirilerek (flatten) birbirlerinin ucuna ekliyordu, veya bu gömülmüş öznetelik çıktılarını düzleştirip alt alta ekleyerek dikey eksende maksimumlarını veya ortalamalarını alıyordu. Fakat bu yaklaşımların ilkinde bilgi kaybı olmaması güzel bir şey olsa da üst üste bu çıktıları birleştirmek çoğu modelde $7 \cdot 10^6 \times 1$ lik bir vektör ile sonuçlanıyordu. Buda modelin parametre sayısını artırarak öğrenmeyi negatif etkiliyordu. İkinci yaklaşımda ise her parametre için diğer 3 çıktının bilgisi tamamen siliniyordu. Bizde hem parametre sayısını çok arttırmayacak hem de bilgi kaybını minimuma indirecek bir birleştirme algoritması düşündük. Aklımıza bunlardan farklı olan ve bahsettiğimiz kriterlere uyan iki metot geldi: çıktıları kanal (channel) üzerinden birleştirmek veya çıktıları dikdörtgen şekline getirmek. Matematiksel olarak ifade edecek olursak, mesela her bir görüntünün çıktısı $7 \times 512 \times 512$ olsun.

1- Kanal üzerinden birleştirme çıktısı: $28 \times 512 \times 512$

2- Dikdörtgen şeklinin çıktısı: Bu çıktıları ikili halde yan yana birleştirip sonra alt alta birleştirerek $7 \times 1024 \times 1024$ lük bir çıktı elde ettik. (Figür 3 model mimarisinde daha rahat anlaşılmaktadır.)

Bu iki yöntemi de denedik, ve en iyi sonucu ikincisi ile elde ettik. Sonrasında buradaki bilgiyi birleştirmek ve daha küçük boyutlara indirmek için birleştirilmiş çıktı Convolution Layer'dan geçirildi.

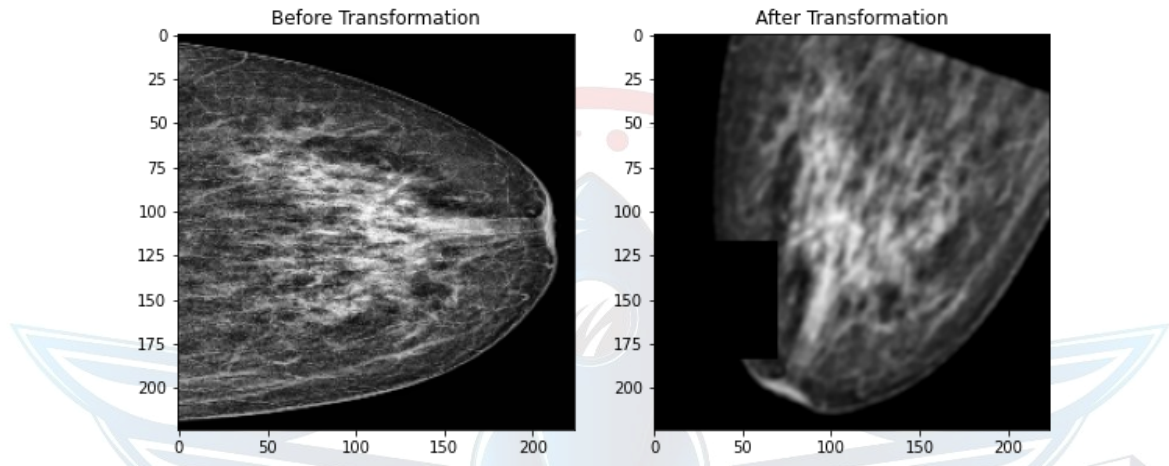
Veri setinde oldukça büyük sınıf dengesizliği hem BI-RADS türü için hem de Kompozisyon türü için var. Bu nedenle ilk olarak veri setini eğitim, doğrulama ve test kümesine ayırırken katmanlı ayırma (stratified split) kullandık. Bu metot veri setinin her sınıfını sıklığına göre bölerek, az sıklıkta olan sınıfın eğitim veya test verisinde gelme ihtimalini ortadan kaldırıyor. Modelin sıklığı az olan sınıfı daha iyi öğrenebilmesi için sıklıklarının ters orantısı ile bağlantılı penaltı terimi kullandık. Yani sıklığı az olan bir verinin doğru bilinmemesi halinde kayıp fonksiyonu daha fazla yükselerek modelin azınlık sınıfı daha iyi öğrenmesini sağlıyor.

Ayrıca modelimizin eğitim verisini ezberlemesini engelleyip aşırı öğrenmenin önüne geçmek istedik. Bu nedenle veri çoğaltma (data augmentation) algoritması uyguladık. Modelimizde işlenmek için girecek olan her örnekleme rastgele olmak koşuluyla şu işlemlerden biri ya da birkaçı uygulanmaktadır:

Yapay zeka modelinin en önemli parametrelerinden birisi de kayıp fonksiyonudur. Genellikle sınıflandırma projelerinde Cross Entropy Loss kullanılmaktadır. Fakat biz bunun yerine Focal Loss kullanmayı tercih ettik. Focal Loss, sınıf dengesizliği olan veri setlerinde nadir sınıfların daha iyi öğrenilmesini sağlamak için Cross Entropy Loss'a bir düzeltme getiren bir kayıp fonksiyonudur. Bu yöntem, daha az örneklenen sınıflara daha fazla ağırlık vererek, modelin daha dengeli bir şekilde eğitilmesini sağlar. [4] Bu sayede model hem daha hızlı öğreniyor

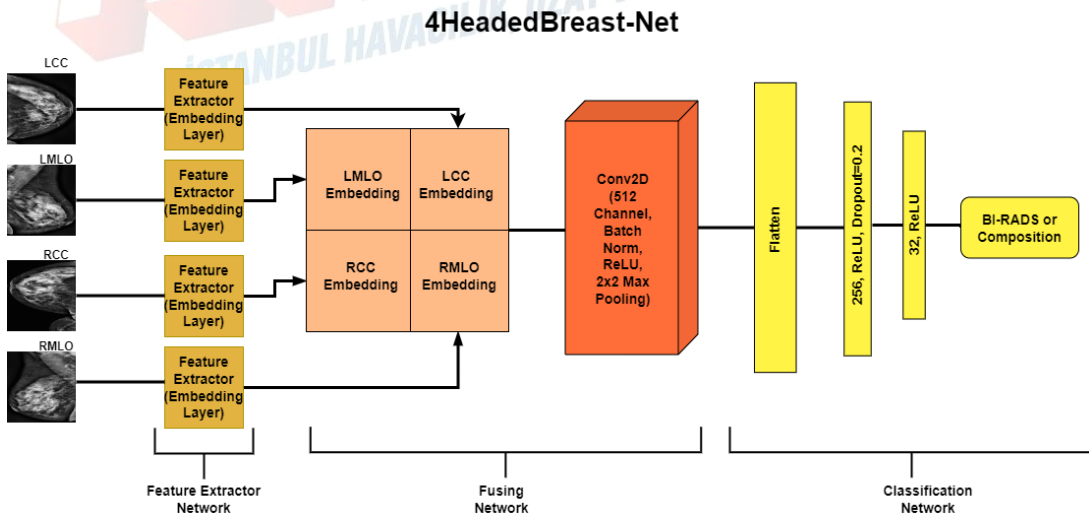
hem de modelin genelleme yeteneği artıyor.

Son olarak, veri setinin artırılması ve modelin sağlamlığını (robustness) arttırmak amacı ile hem yeni veri setleri kullanıldı hem de veri artırma yöntemleri (data augmentation) kullanıldı. Veri setini genişletmek için RSNA ve VinDr Mammo veri setleri eklendi. Bu veri setlerinin miktarı ve istatistikleri dördüncü bölümde detaylı anlatılmıştır. Bunun yanı sıra, modelin daha dayanıklı olması için bütün veriler üzerinde random rotasyon (-30 ile 30 derece arasında), Gauss Bulanıklaştırma, Parça Silmesi ve Renk Titreştirilmesi uygulandı. Bu 4 veri çoğaltma yönteminin her girdi resmine random bir şekilde uygulanması sayesinde bir görüntü her seferinde farklı bir çoğaltma tekniğinden geçiriliyor. Bu sayede modelin verileri ezberlemesinin önüne geçtik, ve hassas verilere daha dayanıklı hale geldi.



Figür 2: Veri Çoğaltma Yönteminin Girdi Resim Üzerindeki Etkisi

Aşağıdaki figür modelimizin mimarisini özetlemektedir. Model 3 kısımdan oluşuyor: Feature Extractor, Fusing ve Classification Network. Feature Extractor Network, medikal görüntülerin gömülmesini sağlıyor; Fusing Network, ilk kısımdan gelen dört çıktının birleştirilip minimum bilgi kaybı ile birleştirilmesini sağlıyor; Classification Network ise eğer amaç BI-RADS türünü tahminlemekse 3 muhtemel sınıf için olasılık dönüyor. Eğer amaç kompozisyon türünü belirlemek ise 4 muhtemel sınıf için olasılık dönüyor.



Figür 3: 4HeadedBreast-Net Mimarisi

3. Sonular ve İnceleme

Sonuları deęerlendirme aısından her kategorinin ayrı ayrı (BI-RADS iin 0, 1-2, 4-5; kompozisyon iin A, B, C, D) kesinlik (precision), duyarlılık (recall) ve F1 skor deęerlerinin yanı sıra toplam model ıktılarının aęırlıklı ve aęırlıksız doęruluk deęerleri (accuracy, weighted accuracy) de ölölmüő ve kaydedilmiőtir. Ayrıca, eęitim esnasında modelin performansı doęrulama kaybı (validation loss) deęiőtini ile de takip edilmiőtir. Modellerimiz 3 veri seti üzerinde eęitilmiőtir. Bu veri setlerinin birisi Teknofest'in saęladıęı veri seti olup test kısmında sadece bu verilerin alt kümesi (subset) kullanılmıőtır. Model eęitilirken kullandıęımız doęrulama (validation) kısmında ise 3 setten hibrit olarak elde edilen bir küme üzerinde deneme yapılmıőtır.

Aőaęıda vereceęimiz sonuları elde edene kadar birok deney yaptık. Her deneyimizde metotlarımızın tahminlemelere uygunluęunu analiz ettik. Bu sayede metot geliőtirip deęiőtirdik. Ayrıca yaptıęımız her hatayı modelimizi bir basmak daha geliőtirmek iin kullandık. İlk olarak modelimizde Cross Entropy kayıp fonksiyonunu kullandık. Ancak doęrulama (validation) sonularına baktıęımızda modelimizin BI-RADS 0 kategorisini hi tahminlemedięini grdük. Bu nedenle bu metodun doęu olmadıęını düőtündük ve yeniden bir literatür taraması yaparak Focal Kayıp fonksiyonunun bizimki gibi dengesiz daęılım gsteren veri setlerinde etkili olduęunu ęrendik ve bu yzden bu fonksiyonu uyguladık. Sonularımızda gzle grölür farklar yarattı hem BI-RADS 0 kategorisini tahminlemesi düzeltildi hem de ęrenme hızı 1.5 katına ıktı.

Ayrıca preprocess kısmında birok sorunla karőtılaőtık. En önemli sorun veri setimizdeki grüntülerin farklı data boyutlarına kodlanmasıydı. Yani veri tiplerinin bazıları 8 bit bazıları 12 bazıları ise 16 bitti. Bu nedenle preprocess metotlarımızla uyumsuzluk gsteriyordu. grüntüler CLAHE iőtlemi uygulanamıyordu. Bu nedenle tüm verileri 8 bit haline getirdik. Buna ek olarak grüntülerin monochrome tipleri farklıydı bu da monochrome 2 tipinde olan grüntülerin iőtlendikten sonra bozulmasına neden oluyordu. Bu yzden de hepsini monochrome 1 tipine evirdik. Bu problemin varlıęını yaptıęımız deney sonucunda anladık. CLAHE uyguladıęımız ve uygulamadıęımız ilk deneyde ikisi arasındaki fark beklenenin ok altındaydı. Ancak tüm verileri aynı tipe getirdięimizde sonularımızda ani bir iyiliőtme oldu ki bu metodumuzun problemimiz ile uyumunu ve doęruluęunu kanıtladı.

Bir dięer karőtılaőtıęımız problem ise Fuse Network'e giren 4 znitelięi gmölmüő ıktıların birleőtirilmesi idi. Öncelikle bu alanda kullanılan genel yaklaőtımları denedik. Bütün gmölmüő znitelikler düzeltirilip birbirinin üstüne eklendi. Bunu denedięimizde yaklaşık 7 milyon girdiden oluőtan bir vektör elde ettik ve bunu Classification Network'e verdięimizde sadece ilk katmanda $7 \times 10^6 \times 256$ adet parametre oluőtuyordu. Buda modelin ęrenimini zorlaőtırıyordu. Sonrasında parametre sayısını düőtürecek başka bir metot bulduk. Bu metotta ise düzeltirililen gmölmüő znitelikler birbirine alt alta eklenip dikey ekseninde maksimum veya ortalama deęerleri alınarak yeni bir vektör oluőturuluyor. Bu metotta oluőtan ıktı vektörü ilk metottakinin drtte biri uzunluęa sahip oluyordu, fakat dięer üç ıktıdan gelen bilgi tamamen siliniyordu. Aynı őekilde bu yöntem ile de istedięimiz doęruluk ve F1 skorlarına eriőtmedik. Bu yzden hem parametre sayısını daha az tutacak hem de bilgi kaybını minimuma indirecek bir birleőtirme algoritması geliőtirdik. Bunun detayları zgöllük kısmında anlatılmıőtı. Geliőtirdięimiz metot sayesinde bahsettięimiz iki probleminde özdük, ayrıca doęruluk ve F1 skorumuzu arttırmayı baőtardık.

őtü ana kadarki tecrübelerimiz bize genel geer ML bilgilerimizin ve teorik bilgilerimizin gerek hayat uygulamasından farklı gsterdi. Hibir aőamasında yaptık oldu diyemeyeceęini

ve sürekli geliştirilme payının olduğunu gördük. Ayrıca çözümlerini yaratıcı düşünmeyi zorunlu kıldığını da anladık.

Temel olarak metodlarımız sonuçlarla direkt olarak alakalı bunun sebebi ise karşılaştığımız problemleri her zaman F1 ve doğruluk skorumuzu arttıracak şekilde kendi geliştirdiğimiz yöntemler ile çözmemizdir. Yukarıda bahsedilen sorunlar verilerin ön işleme ve sınıflandırma aşamalarında olmuştur. Bu yüzden bunların hepsi ayrılmadan tek bir başlık altında anlatıldı. BI-RADS ve Kompozisyon Sonuçları kısmında ise sadece yarışmanın ilgili kısımlarının sonuçları konulmuştur.

3.1. BI-RADS Sonuçları

Verilerin BI-RADS etiketlerini tahminlemek için üçüncü figürdeki modeli kullandık. Son çıkış katmanı 3 nörondan oluşuyor, her bir nöron BI-RADS 0, 1-2 veya 4-5 olma olasılığını gösteriyor. Feature Extractor Network için değişik öznelik çıkarıcı modeller denedik. En iyi sonuçları Mobile Net V2 ile elde ettik. Teknofest'in sağladığı Sağlık Bakanlığı verisinin tamamı direkt olarak kullanıldı. Bunun haricinde dördüncü bölümde detaylı açıkladığımız RSNA ve VinDr Mammo veri setleri kullanıldı. Fakat üç veri setinin hepsinde BI-RADS1-2 etiketine sahip olan veri sayısı çok fazla olduğu için, RSNA verisinden sadece BI-RADS0 etiketine sahip olan verileri aldık. Aksi takdirde, model BI-RADS1-2 etiketini öğrenmeye daha çok yelteniyor ve diğer etiketlerin öğrenilmesi zorlaşıyordu.

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| BI-RADS0 | 0.53 | 0.55 | 0.54 | 110 |
| BI-RADS1-2 | 0.83 | 0.81 | 0.82 | 263 |
| BI-RADS4-5 | 0.79 | 0.79 | 0.79 | 224 |
| accuracy | | | 0.76 | 597 |
| macro avg | 0.71 | 0.72 | 0.72 | 597 |
| weighted avg | 0.76 | 0.76 | 0.76 | 597 |

Figür 4: BI-RADS Modelinin Sınıflandırma Skorları

| | | Confusion Matrix | | |
|------------|------------|------------------|------------|------------|
| | | BI-RADS0 | BI-RADS1-2 | BI-RADS4-5 |
| Prediction | BI-RADS0 | 60 | 27 | 26 |
| | BI-RADS1-2 | 25 | 213 | 20 |
| | BI-RADS4-5 | 25 | 23 | 178 |
| | | Ground Truth | | |

Figür 5: BI-RADS Modelinin Hata Matrisi

Figür 4'ten görebileceğimiz üzere, model BI-RADS1-2 ve BI-RADS4-5 etiketlerini çok iyi öğrenmesine rağmen BI-RADS0 etiketi için yeterli kesinlik ve duyarlılık skorlarını elde

edemiyor. Bunun sebebi elimizdeki BI-RADS0 verilerinin çok az olmasından kaynaklanıyor. Diğer etiketlere sahip olan veri sayısı neredeyse BI-RADS0'ın 4 katına denk geliyor. Bu problemin veri setinin özellikle BI-RADS0 için artmasıyla düzelebileceğine inanıyoruz. Eğer final aşamasına kalırsak, yeni gelen veriler ile bu sınıf içinde daha iyi kesinlik ve duyarlılık seviyelerine erişebiliriz. Figür 5 ise modelin hata matrisini gösteriyor. Görüldüğü üzere model BI-RADS1-2 ve 4-5'i oldukça iyi bilirken, 0 için neredeyse verilerin yarısını yanlış tanımlıyor.

3.2. Kompozisyon Sonuçları

Kompozisyon etiketleri tahmini için yine 4HeadedBreast-Net kullandık. Classification network çıkışını sınıf sayısı olan dört olacak şekilde ayarladık. Çıkan sonuçları olasılık benzeri haline getirmek için log softmax fonksiyonu kullandık. Feature Extractor Network için birçok farklı öznetelik çıkarıcılar ile deneyler yaptık. Fakat en iyi sonuçlar Effcient-Net B3 ile elde ettik. Özgünlük başlığı altında anlatılan preprocess metotlarını kullandık. Dataset olarak bir sonraki başlıkta detaylı anlatılan RSNA ve VinDr Mammo veri setleri kullandık.

| | precision | recall | f1-score | support |
|--------------|-----------|--------|----------|---------|
| A | 0.69 | 0.93 | 0.80 | 90 |
| B | 0.85 | 0.78 | 0.81 | 220 |
| C | 0.85 | 0.77 | 0.81 | 185 |
| D | 0.76 | 0.80 | 0.78 | 102 |
| accuracy | | | 0.80 | 597 |
| macro avg | 0.79 | 0.82 | 0.80 | 597 |
| weighted avg | 0.81 | 0.80 | 0.80 | 597 |

Figür 6: Kompozisyon Modelinin Sınıflandırma Skorları

| | | Confusion Matrix | | | |
|------------|---|------------------|-----|-----|----|
| | | A | B | C | D |
| Prediction | A | 84 | 16 | 14 | 7 |
| | B | 4 | 171 | 18 | 8 |
| | C | 2 | 18 | 142 | 5 |
| | D | 0 | 15 | 11 | 82 |
| | | Ground Truth | | | |

Figür 7: Kompozisyon Modelinin Hata Matrisi

Yukarıdaki figürlerde de görüldüğü gibi modelimiz test set üstünde %80 doğruluk oranına ulaştı. Kesinlikle ve duyarlılık skorlarına bakıldığında A ve D kategorisinin diğerlerinden biraz geri kaldığı görülmektedir. Bunun nedeni olarak ise veri setindeki miktarlarının az olmasının öğrenimlerini zorlaştırdığı gösterilebilir. Hata matrisine bakıldığında bilmeden A kategorisi tahminlediğini görmekteyiz. Bunun haricinde skorlar genel anlamda tatmin edici seviyelere ulaştı.

4. Deney ve eğitim aşamalarında kullanılan veri setleri

Projede verilen veri setine ek olarak RSNA ve VinDr-Mammo veri setleri kullanılmıştır.

VinDr-Mammo Veri Seti

Bu veri seti proje kapsamında sağlanan veri setiyle çokça benzerlik gösterdiği için tercih ettik. Veri seti 5,000 hastaya ait 20,000 görüntü içermektedir. Her hasta proje verisindeki gibi 4 açıdaki görüntüyü de içermektedir. Aşağıdaki istatistiksel figür 10,000 adet olarak gösterse de sadece 5,000 tanesi herkese açık olarak indirilebilmektedir. [5]

| | Breast BI-RADS | | | | | |
|----------|----------------|----------------|--------------|--------------|--------------|--------|
| | 1 | 2 | 3 | 4 | 5 | Total |
| Training | 5,362 (67.03%) | 1,871 (23.39%) | 372 (04.65%) | 305 (03.81%) | 90 (01.12%) | 8,000 |
| Test | 1,341 (67.05%) | 467 (23.35%) | 93 (04.65%) | 76 (03.80%) | 23 (01.15%) | 2,000 |
| Overall | 6,703 (67.03%) | 2,338 (23.38%) | 465 (04.65%) | 381 (03.81%) | 113 (01.13%) | 10,000 |

Figür 8: VinDr-Mammo BI-RADS Örnekleme Dağılımı [5]

| | Breast Density | | | | |
|----------|----------------|--------------|----------------|----------------|--------|
| | A | B | C | D | Total |
| Training | 40 (00.50%) | 764 (09.55%) | 6,116 (76.45%) | 1,080 (13.50%) | 8,000 |
| Test | 10 (00.50%) | 190 (09.50) | 1,530 (76.50%) | 270 (13.50%) | 2,000 |
| Overall | 50 (00.50%) | 954 (09.54%) | 7,646 (76.46%) | 1,350 (13.50%) | 10,000 |

Figür 9: VinDr-Mammo Kompozisyon Örnekleme Dağılımı [5]

Bu veri setini kullanırken eğitim ve test için olan dataları birleştirip eğitim veri setimize ekledik. Veri setinde bazı bozuk datalar bulunmaktaydı ve biz bunları setten manuel olarak çıkardık. Ardından formatlarını projenin orijinal veri setiyle aynı hale getirip var olan işlem zincirinden (pipeline) geçirdik. VinDr veri setinde BI-RADS 0 kategorisinde veri olmamasının model eğitimini olumsuz etkileme şansı olsa da ağırlıklı kayıp metodu kullanarak bu sorunun üstesinden geldik. Ayrıca öğrenme sürecinde hiçbir veri setinde BI-RADS 3 kategorisini kullanmadığımız için VinDr seti içindeki 465 BI-RADS 3 veri örneğini setten çıkardık. BI-RADS 1 ve BI-RADS 2 verilerini tek bir etiket altında BI-RADS1-2 olarak; aynı şekilde BI-RADS 4 ve BI-RADS 5 verilerinin etiketlerini

BI-RADS4-5 olarak Teknofest'in sağladığı veri setindeki etiketleme sistemine uygun bir biçimde birleştirdik.

RSNA Veri Seti:

RSNA Screening Mammography Breast Cancer Detection isimindeki Kaggle [6] yarışmasına ait veri setidir. Bu veri setinde 11,913 hastadan alınan 54,700 görüntü bulunmaktadır. Veri setinde bazı bozuk datalar bulunmaktaydı ve biz bunları setten manuel olarak çıkardık. Bazı hastalar ya dörtten daha az mamografi resmine sahipti yada dörtten daha fazla mamografi resmine sahipti. Bu hastalardan LCC, LMLO, RCC, RMLO olmak üzere dört farklı mamografi görüntüsüne sahip olmayanları çıkardık. Dörtten fazla olanlarda ise gördüğümüz ilk dörtlüyü aldık.

BI-RADS kategorisinde BI-RADS 0, 1 ve 2 bulunmaktaydı. Biz 1 ve 2'yi birleştirip tek kategori haline getirdik. Bazı hastalarında direkt olarak BI-RADS türü etiketlenmemişti, bu hastalara ait olan mamografiler veri setinden çıkarıldı. Aynı şekilde bazı hastalara ait olan mamografilerin kompozisyon etiketi eksikti, bu hastalara ait verilerde kompozisyon modeli eğitilirken çıkarıldı.

5. Referanslar

[1] A. Famili, W.-M. Shen, R. Weber, and E. Simoudis, "Data preprocessing and Intelligent Data Analysis," *Intelligent Data Analysis*, 01-Jan-1997. [Online]. Available: <https://content.iospress.com/articles/intelligent-data-analysis/ida1-1-02>. [Accessed: 01-Mar-2023].

[2] "Otsu's method," *Wikipedia*, 24-Feb-2023. [Online]. Available: https://en.wikipedia.org/wiki/Otsu%27s_method. [Accessed: 01-Mar-2023].

[3] "Adaptive Histogram Equalization," *Wikipedia*, 12-Oct-2022. [Online]. Available: https://en.wikipedia.org/wiki/Adaptive_histogram_equalization. [Accessed: 01-Mar-2023].

[4] T.-Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollár, "Focal loss for dense object detection," *arXiv.org*, 07-Feb-2018. [Online]. Available: <https://arxiv.org/abs/1708.02002>. [Accessed: 01-Mar-2023].

[5] H. T. Nguyen, H. Q. Nguyen, H. H. Pham, K. Lam, L. T. Le, M. Dao, and V. Vu, "Vindr-Mammo: A large-scale benchmark dataset for computer-aided diagnosis in full-field digital mammography," 2022.

[6] "RSNA Screening Mammography Breast Cancer Detection," *Kaggle*. [Online]. Available: <https://www.kaggle.com/competitions/rsna-breast-cancer-detection>. [Accessed: 01-Mar-2023].

01-Mar-2023].

RAPOR TASLAKLARI İLE İLGİLİ NOT:

- Tüm raporlar akademik rapor standartlarına uygun olarak yazılmalıdır.
- Raporların içerikleri ile ilgili bilgiler yukarıda belirtilmiştir.
- Tüm raporlar “İçindekiler” ve “Kaynakça” içermelidir.
- Her rapor bir kapak sayfası içermelidir.
- Her rapor bir “İçindekiler” sayfası içermelidir.
- Raporlar sayfaları birbirini takip edecek şekilde numaralandırılmalıdır.
- Yazı tipi: Times New Roman, Punto: 12, Satır Aralıkları: 1
- “Kapak” ve “İçindekiler” kısmı hariç Rapor 10 sayfayı geçmemelidir.

