

T.C.
SAKARYA ÜNİVERSİTESİ
BİLGİSAYAR VE BİLİŞİM BİLİMLERİ FAKÜLTESİ

**DERİN ÖĞRENME İLE HİSTOPATOLOJİK
RESİMLERDE METASTATİK KANSER TESPİTİ**

BSM 498 - BİTİRME ÇALIŞMASI

Sinan BAVLİ

Fakülte Anabilim Dalı : BİLGİSAYAR MÜHENDİSLİĞİ

Bu tez .. / .. / ... tarihinde aşağıdaki jüri tarafından oybirliği / oyçokluğu ile kabul edilmiştir.

.....
Jüri Başkanı

.....
Üye

.....
Üye

ÖNSÖZ

Bu bitirme çalışması, derin öğrenme tekniklerini kullanarak histopatolojik resimlerde kanser tespiti yapmayı amaçlayan bir araştırmanın ürünüdür. Bu çalışmada, modern yapay zeka yöntemlerinin sağlık alanında nasıl kullanılabileceğini ve kanser teşhisinde daha hızlı ve doğru sonuçlar elde edilmesine nasıl katkı sağlayabileceğini göstermeyi hedefledik.

Kanser, dünya genelinde en önemli sağlık sorunlarından biri olup, erken teşhis ve doğru tedavi yöntemleri hastaların yaşam süresini ve kalitesini büyük ölçüde etkileyebilmektedir. Histopatolojik resimler, kanser teşhisinde önemli bir rol oynar ve bu resimlerin doğru analizi, doğru teşhis için kritiktir. Geleneksel yöntemlerle bu analizlerin yapılması zaman alıcı ve yorucu olabilirken, derin öğrenme teknikleri bu süreci hızlandırabilir ve hata payını azaltabilir.

Bu çalışma sürecinde, literatürdeki mevcut çalışmaları inceleyerek derin öğrenme algoritmalarının histopatolojik resimlerde kanser tespiti üzerindeki etkilerini değerlendirdik. Kendi modelimizi geliştirerek, bu alandaki potansiyeli ve uygulanabilirliği test ettik. Geliştirilen modelin doğruluğunu ve etkinliğini artırmak için çeşitli optimizasyon teknikleri ve veri artırma yöntemleri kullanıldı.

Bu çalışmanın, kanser teşhisinde derin öğrenme tekniklerinin kullanımına dair literatüre katkı sağlamasını ve gelecekteki araştırmalara ışık tutmasını temenni ediyorum.

İÇİNDEKİLER

ÖNSÖZ	i
İÇİNDEKİLER	ii
SİMGELER VE KISALTMALAR LİSTESİ.....	v
ŞEKİLLER LİSTESİ	vi
TABLolar LİSTESİ	vii
ÖZET.....	vii
BÖLÜM 1. GİRİŞ.....	1
1.1. Çalışmanın Kapsamı	2
1.1.1. Veri Toplama ve Ön İşleme.....	2
1.1.2. Model Geliştirme.....	3
1.1.3. Model Eğitimi ve Değerlendirilmesi.....	3
1.1.4. Sonuçların Değerlendirilmesi.....	3
1.2. Çalışmanın Önemi.....	3
BÖLÜM 2. TEORİK ARKA PLAN:	5
2.1. Konvölüsyonel Sinir Ağları.....	7
2.2. Genelleştirme Teknikle.....	10
2.3. Modelin Değerlendirme Metrikleri.....	12
BÖLÜM 3. VERİ SETİ VE ÖN İŞLEME	16
3.1. Veri Setinin Kaynağı ve Özellikleri.....	17
3.2. Veri Artırma.....	19
3.3. Veri Setinin Bölünmesi.....	21
BÖLÜM 4. MODEL GELİŞTİRME	23
4.1. Modelin Mimarisi ve Katmanları.....	23
4.2. Modelin Eğitim Süreci.....	27
4.3. Modelin Değerlendirilmesi.....	29
4.4. Modelin Test Edilmesi.....	33
BÖLÜM 5. WEB SAYFASI ÜZERİNDEN MODELİN SUNUMU.....	35
5.1. Web Sitesi Mimarisi ve Teknolojileri.....	35
5.2. Web Sitesinin Gösterimi.....	36
5.3. Modelin Web Sitesi Üzerinden Kullanımı.....	38

BÖLÜM 6.	SONUÇ.....	40
BÖLÜM 7.	Kaynakça	41
ÖZGEÇMİŞ		42
BSM498 BİTİRME ÇALIŞMASI DEĞERLENDİRME VE SÖZLÜ SINAV TUTANAĞI		43

SİMGELER VE KISALTMALAR LİSTESİ

AUC	: Eğri altındaki alan
SGD	: stochastic gradient descent
PCam	: PatchCamelyon
ANN	: Artificial Neural Networks
LSTM	: Long Short-Term Memory
RNN	: Recurrent Neural Networks
CNN	: convolution neural network

ŞEKİLLER LİSTESİ

Şekil 1 Yapay Zeka Hiyerarşisi	7
Şekil 2 Örnek Konvolüsyonel katman yapısı.....	7
Şekil 3 Konvolüsyon işlemi	8
Şekil 4 Sigmoid grafiği	9
Şekil 5 Maxpooling işlemi	9
Şekil 6 Konfüzyon matrix grafik örneği	14
Şekil 7 Veri setinden rastgele 10 örnek	18
Şekil 8 Veri setinin hedef değişken dağılımı.....	18
Şekil 9 Veri artırma işlemi.....	20
Şekil 10 Oluşturulmuş Veri seti dosya yapısı.....	22
Şekil 11 Oluşturulmuş olan CNN ağının mimarisi	23
Şekil 12 Callbacks'in kodlanması.....	28
Şekil 13 Modelin eğitim sonuçları	29
Şekil 14 Modelin konfüzyon matrix'i	30
Şekil 15 Modelin AUC-ROC grafiği.....	31
Şekil 16 Modelin test sonuçları.....	34
Şekil 17 Web sitesinin anasayfası	36
Şekil 18 Web sitesinin hakkında kısmı.....	37
Şekil 19 Web Sitesinin iletişim kısmı.....	37
Şekil 20 Anasayfa üzerinden dosya yükleme.....	38
Şekil 21 Tahmin sonuç örneği.....	39

TABLÖLAR LİSTESİ

Tablo 1.0 Sınıflandırma Raporu	39
--------------------------------------	----

ÖZET

Anahtar Kelimeler: Yapay zeka, Derin öğrenme, Kanser tespiti, Makine öğrenmesi, Evrişimsel sinir ağı

Bu proje, derin öğrenme tekniklerini kullanarak histopatolojik resimlerde kanser tespiti yapmayı amaçlamaktadır. Kanser, erken teşhis ve doğru tedavi yöntemlerinin kritik olduğu bir hastalıktır. Histopatolojik resimler, dokuların mikroskop altında incelenmesiyle elde edilen görüntülerdir ve kanser teşhisinde önemli bir rol oynar. Geleneksel yöntemlerle yapılan analizler zaman alıcı ve yorucu olabılırken, derin öğrenme teknikleri bu süreci hızlandırabilir ve doğruluğunu artırabilir.

Projede, konvolüsyonel sinir ağları (CNN) gibi derin öğrenme algoritmalarını kullanarak bir model geliştirilmiştir. Bu model, histopatolojik resimlerde kanserli hücrelerin tespiti için eğitilmiştir. Modelin eğitimi için çeşitli veri artırma teknikleri kullanılarak modelin genelleme kabiliyeti artırılmıştır. Veri seti olarak, güvenilir ve geniş kapsamlı bir histopatolojik görüntü dataseti kullanılmıştır.

Modelin performansı, accurcay ve loss function ile değerlendirilmiştir. Elde edilen sonuçlar, modelin yüksek doğrulukla kanserli hücreleri tespit edebildiğini göstermektedir. Bu da derin öğrenme tekniklerinin histopatolojik analizlerde kullanılabilirliğini ve etkinliğini kanıtlamaktadır.

Projede oluşturulan modelin çıktılarını göstermek için bir web sitesi tasarlanmıştır. Tasarlanan web sitesi üzerinden kullanıcılar bir doku örneği girerek, girilen dokuda kanser olup olmadığını tespit edebileceklerdir.

Bu projenin sonuçları, kanser teşhisinde kullanılabilecek hızlı ve doğru bir yardımcı araç olarak derin öğrenme tekniklerinin potansiyelini ortaya koymaktadır. Gelecekte, bu tür modellerin klinik uygulamalara entegrasyonu ile doktorların teşhis sürecinde desteklenmesi ve hastaların daha hızlı bir şekilde doğru tedaviye yönlendirilmesi mümkün olacaktır.

BÖLÜM 1. GİRİŞ

Bu proje, histopatolojik resimlerde kanser tespiti için derin öğrenme tekniklerini kullanmayı amaçlayan bir araştırmadır. Histopatolojik resimler, dokuların mikroskop altında incelenmesi sonucu elde edilen görüntülerdir ve patolojistlerin kanser teşhisinde kritik bir rol oynar. Bu proje kapsamında, konvolüsyonel sinir ağları (CNN) gibi ileri seviye derin öğrenme algoritmalarını kullanarak, bu görüntülerde kanserli hücreleri otomatik olarak tespit edebilecek bir model geliştirilmiştir.

Derin öğrenme, büyük veri setlerinden öğrenme kapasitesine sahip olan yapay zeka algoritmalarının bir alt dalıdır. Bu projede kullanılan derin öğrenme modeli, çok katmanlı yapısı sayesinde karmaşık verileri analiz edebilir ve anlamlı sonuçlar üretebilir. Geliştirilen model, histopatolojik resimlerdeki kanserli hücreleri yüksek doğrulukla tespit edebilmek için eğitilmiş ve optimize edilmiştir.

Tercih edilen teknolojiler, sistemi oluşturan amaç doğrultusunda en iyi performansı verebilecekler arasından seçilmiştir. Tahminleme yapmak için kullandığımız model Python'un tensorflow kütüphaensi ile oluşturulmuştur. Oluşturduğumuz modeli entegre ettiğimiz web sitesinde ise HTML,CSS,Javascript ve pythonun flask kütüphanesi kullanılmıştır. Bu projenin başlıca amaçları şunlardır:

1. Kanser Teşhisini Hızlandırmak ve Doğruluğunu Artırmak:
 - Geleneksel yöntemlerle histopatolojik resimlerin analizi zaman alıcı ve insan hatalarına açık bir süreçtir. Bu projenin amacı, derin öğrenme tekniklerini kullanarak bu süreci otomatikleştirmek ve teşhis sürecini hızlandırmaktır. Böylece, patolojistlerin yükü hafifletilecek ve teşhis doğruluğu artırılabacaktır.
2. Derin Öğrenme Modellerinin Uygulanabilirliğini Araştırmak:
 - Derin öğrenme tekniklerinin histopatolojik resimlerde kanser tespiti üzerindeki etkinliğini ve uygulanabilirliğini araştırmak, bu projenin bir diğer amacıdır. Bu amaç doğrultusunda, farklı derin öğrenme modelleri ve veri artırma teknikleri kullanılarak en yüksek doğruluğa sahip model geliştirilmiştir.

3. Klinik Uygulamalara Katkı Sağlamak:

- Geliştirilen modelin, klinik uygulamalarda kullanılabilirliğini değerlendirmek ve gelecekte bu tür teknolojilerin sağlık alanında nasıl kullanılabileceğine dair bir yol haritası oluşturmak projenin önemli hedeflerindendir. Bu sayede, kanser teşhisinde yapay zeka destekli sistemlerin yaygınlaşması ve hasta bakımının iyileştirilmesi hedeflenmektedir.

1.1. Çalışmanın Kapsamı

Bu proje, derin öğrenme tekniklerini kullanarak histopatolojik resimlerde metastatik kanser tespiti yapmayı amaçlayan kapsamlı bir çalışmadır. Çalışmanın kapsamı, veri toplama ve ön işleme, model geliştirme, model eğitimi ve değerlendirilmesi, sonuçların analizi ve gelecekteki çalışmalar için öneriler olmak üzere beş ana bileşenden oluşmaktadır.

1.1.1. Veri Toplama ve Ön İşleme

Projenin ilk aşaması, histopatolojik görüntülerin toplanması ve bu verilerin işlenmesi sürecini içermektedir. Güvenilir ve geniş kapsamlı bir veri seti elde etmek amacıyla, çeşitli tıbbi veri tabanlarından ve literatürdeki mevcut kaynaklardan histopatolojik görüntü veri setleri incelenmiş ve en uygun veri seti seçilmiştir. Bu veriseti, kanserli ve kanserli olmayan hücreleri içeren resimlerden oluşmaktadır. Bunun yanında verisetinde ki her resmin id numarasını ve resmin kanser içerip içermediği gösteren bir binary hedef kolonuna sahip olan bir csv dosyası da mevcuttur.

Veri ön işleme aşamasında, toplanan görüntülerin boyutlandırılması, normalizasyonu ve gerektiğinde veri artırma teknikleri kullanılarak çeşitlendirilmesi yapılmıştır. Veri artırma işlemleri, modelin genelleme yeteneğini artırmak ve aşırı öğrenmeyi önlemek amacıyla gerçekleştirilmiştir.

1.1.2. Model Gelistirme

Bu aşamada, kanserli hücrelerin tespiti için uygun bir derin öğrenme modeli geliştirilmiştir. Konvolüsyonel sinir ağları (CNN) gibi ileri düzey derin öğrenme teknikleri kullanılarak, histopatolojik resimlerin özelliklerini öğrenebilecek ve kanserli hücreleri tespit edebilecek bir model tasarlanmıştır. Modelin mimarisi, katman sayısı, filtre boyutları ve diğer hiperparametreler optimize edilerek en iyi performansın elde edilmesi sağlanmıştır.

1.1.3. Model Eğitimi ve Değerlendirilmesi

Geliştirilen model, toplanan ve işlenen veri seti üzerinde eğitilmiştir. Eğitim sürecinde, modelin doğruluğunu ve genel performansını artırmak için çeşitli optimizasyon teknikleri ve kayıp fonksiyonları kullanılmıştır. Modelin performansı, doğruluk metriği ve loss fonksiyonu kullanılarak değerlendirilmiştir.

Eğitim ve değerlendirme sürecinde, modelin aşırı öğrenmesini önlemek ve genelleme yeteneğini artırmak için çapraz doğrulama ve test veri setleri kullanılmıştır. Modelin performansı, literatürdeki diğer çalışmalarla karşılaştırılarak değerlendirilmiş ve elde edilen sonuçlar analiz edilmiştir.

1.1.4. Sonuçların Analizi

Modelin eğitim ve test aşamalarında elde edilen sonuçlar detaylı bir şekilde analiz edilmiştir. Modelin kanserli hücreleri tespit etme konusundaki doğruluğu, precision, recall ve f1 gibi önemli performans metrikleri incelenmiştir. Bu analizler, modelin güçlü ve zayıf yönlerini belirlemeye yardımcı olmuş ve gelecekte yapılacak iyileştirmeler için öneriler sunulmuştur.

1.2. Çalışmanın Önemi

Bu çalışma, derin öğrenme tekniklerini kullanarak histopatolojik resimlerde kanser tespiti yapmayı amaçlayan bir araştırmadır. Kanser, dünya genelinde en yaygın ve

ölümcül hastalıklardan biridir ve erken teşhis, hastalığın tedavisinde büyük bir rol oynamaktadır. Histopatolojik analizler, kanser teşhisinde kullanılan temel yöntemlerden biridir. Ancak, bu analizlerin geleneksel yöntemlerle yapılması zaman alıcı ve insan hatalarına açık olabilir. Bu noktada, yapay zeka ve özellikle derin öğrenme teknikleri devreye girerek önemli avantajlar sunmaktadır.

Bu çalışmanın en önemli katkılarından biri, sağlık alanında yenilikçi teknolojilerin kullanılmasını teşvik etmesidir. Derin öğrenme algoritmalarının kullanımı, histopatolojik analizlerin doğruluğunu ve hızını önemli ölçüde artırabilir. Bu, sadece patolojistlerin iş yükünü azaltmakla kalmaz, aynı zamanda hastaların daha hızlı ve doğru bir teşhis almasını sağlayarak tedavi süreçlerinin iyileşmesine katkıda bulunur. Erken teşhis, kanser tedavisinin başarısında kritik bir faktördür ve bu çalışma, bu sürecin optimize edilmesine yardımcı olur.

Bu proje, derin öğrenme modellerinin yüksek doğruluk oranlarına ulaşabileceğini ve kanserli hücrelerin tespitinde güvenilir sonuçlar verebileceğini göstermektedir. Geleneksel yöntemlerin aksine, derin öğrenme modelleri büyük veri setlerinden öğrenebilir ve karmaşık desenleri tanıyabilir. Bu, özellikle düşük kalitede veya karmaşık görüntülerin analizinde büyük bir avantaj sağlar. Modelin hassasiyet ve özgüllük oranlarının yüksek olması, yanlış pozitif ve yanlış negatif sonuçların azaltılmasına katkıda bulunur, bu da hasta güvenliği ve tedavi etkinliği açısından büyük önem taşır.

Çalışmanın bulguları, derin öğrenme tekniklerinin klinik uygulamalarda nasıl kullanılabileceğine dair önemli ipuçları sunmaktadır. Bu tür modellerin, hastanelerde ve sağlık merkezlerinde rutin olarak kullanılması, teşhis süreçlerinin daha standart ve objektif hale gelmesine yardımcı olabilir. Ayrıca, bu teknolojilerin kullanımı, sağlık hizmetlerinin erişilebilirliğini artırabilir ve kaynakların daha verimli kullanılmasını sağlayabilir.

BÖLÜM 2. TEORİK ARKA PLAN:

Derin öğrenme , makine öğrenmesinin bir alt dalıdır ve büyük miktarda veri üzerinde öğrenme yapabilen çok katmanlı yapay sinir ağlarına dayanır. Derin öğrenme modelleri, veriden otomatik olarak özellik çıkarımı yapma ve bu özellikleri kullanarak karmaşık görevleri yerine getirme yeteneğine sahiptir. Görüntü tanıma, konuşma tanıma, doğal dil işleme ve oyun oynama gibi birçok alanda çığır açan başarılar elde etmiştir.[1]

Derin öğrenme, çok katmanlı yapay sinir ağları kullanarak veriden öğrenir. Bu sinir ağları, giriş verisinden anlamlı özellikler çıkarır ve bu özellikleri kullanarak belirli görevleri yerine getirir. Derin öğrenme modelleri, tipik olarak aşağıdaki bileşenlerden oluşur:

1. **Giriş Katmanı (Input Layer):** Modelin aldığı ham veri. Örneğin, bir görüntü, ses veya metin verisi.
2. **Gizli Katmanlar (Hidden Layers):** Giriş verisinden özellikleri çıkaran ve daha soyut temsiller üreten katmanlar. Bu katmanlar, her biri birçok nörondan oluşur ve öğrenme sürecinde veri üzerinde dönüşümler gerçekleştirir.
3. **Çıkış Katmanı (Output Layer):** Modelin nihai tahminini veya sonucunu üreten katman. Örneğin, bir görüntünün sınıfını belirler.

Derin Öğrenme Türleri

Derin öğrenme çeşitli türlerde sinir ağları kullanılarak gerçekleştirilebilir. Bu türlerin her biri, farklı görevler ve veri türleri için optimize edilmiştir:

1. **Yapay Sinir Ağları (ANN):**
 - Temel yapay sinir ağları, giriş ve çıkış katmanları arasında bir veya birkaç gizli katman içerir. Her nöron, bir önceki katmandaki tüm nöronlarla bağlantılıdır.
 - Uygulamalar: Temel sınıflandırma ve regresyon görevleri.
2. **Konvolüsyonel Sinir Ağları (CNN):**

- Özellikle görüntü ve video işleme için tasarlanmış derin öğrenme modelleridir. CNN'ler, yerel özellikleri öğrenmek için konvolüsyonel katmanlar kullanır.
- Uygulamalar: Görüntü tanıma, nesne algılama, yüz tanıma.

3. Tekrarlayan Sinir Ağları (RNN):

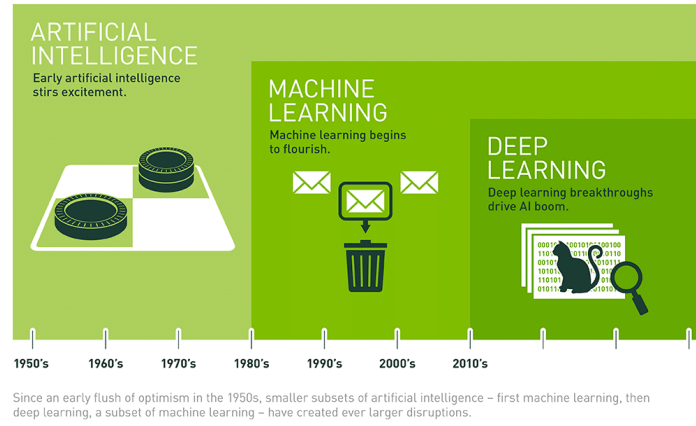
- Zaman serisi verileri veya ardışık veriler için kullanılan sinir ağlarıdır. RNN'ler, veri dizilerindeki bağımlılıkları öğrenmek için döngüsel bağlantılar içerir.
- Uygulamalar: Doğal dil işleme, konuşma tanıma, zaman serisi tahmini.

4. Uzun Kısa Süreli Bellek (LSTM):

- RNN'lerin bir türü olup, uzun vadeli bağımlılıkları öğrenmek için özel olarak tasarlanmıştır. LSTM'ler, unutma ve hatırlama kapıları ile donatılmıştır.
- Uygulamalar: Dil modeli oluşturma, metin üretimi, zaman serisi analizi.

5. Otoenkoderler (Autoencoders):

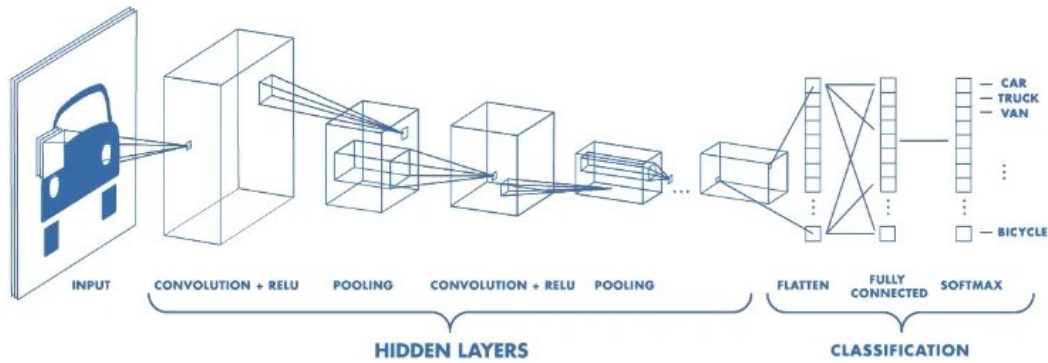
- Veriyi sıkıştırmak ve yeniden oluşturmak için kullanılan sinir ağlarıdır. Otoenkoderler, veri temsiliğini öğrenirken gürültüyü de azaltabilir.
- Uygulamalar: Anomali tespiti, veri sıkıştırma, gürültü giderme.



Şekil 1 Yapay Zeka Hiyerarşisi

2.1. Konvolüsyonel Sinir Ağları:

Konvolüsyonel Sinir Ağları, derin öğrenme algoritmaları içinde özellikle görüntü işleme alanında büyük başarı sağlamış modellerdir. CNN'ler, görüntülerdeki yerel bağımlılıkları ve mekansal hiyerarşiyi öğrenme yeteneği ile öne çıkar. Bu modeller, çeşitli katmanlardan oluşur ve her katman, giriş verisini daha yüksek seviyede bir özellik temsilcisine dönüştürür.[2]



Şekil 2 Örnek Konvolüsyonel katman yapısı

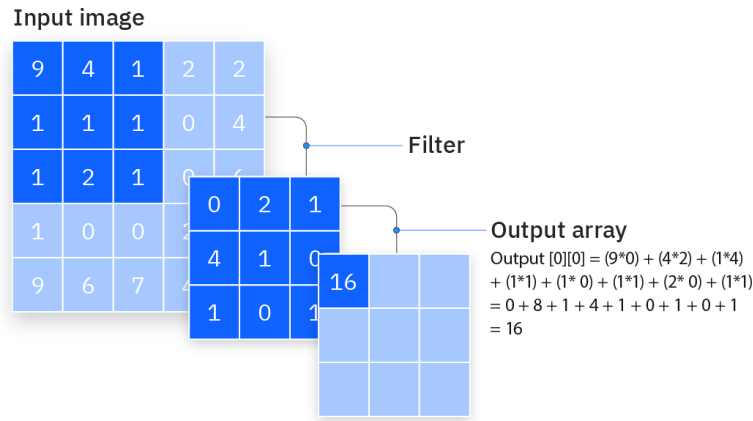
CNN'in Temel Bileşenleri ve Matematiksel Temelleri

1. Konvolüsyonel Katman (Convolutional Layer):

Evrişim katmanı, CNN'nin temel yapı taşıdır. Ağın hesaplama yükünün büyük kısmını taşır.

Bu katman, iki matris arasında bir nokta çarpımı gerçekleştirir; bu matrislerden biri öğrenilebilir parametreler seti olarak bilinen çekirdek, diğeri ise alıcı alanın sınırlı kısmıdır. Çekirdek, bir görüntüden uzamsal olarak daha küçüktür ancak derinliği daha fazladır. Bu, eğer görüntü üç kanaldan oluşuyorsa, çekirdek yüksekliği ve genişliği uzamsal olarak küçük olacak, ancak derinlik tüm üç kanala kadar uzanacaktır.

$$(I * k)(x, y) = \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} I(x+i, y+j) \cdot k(i, j) \quad (1.1)$$



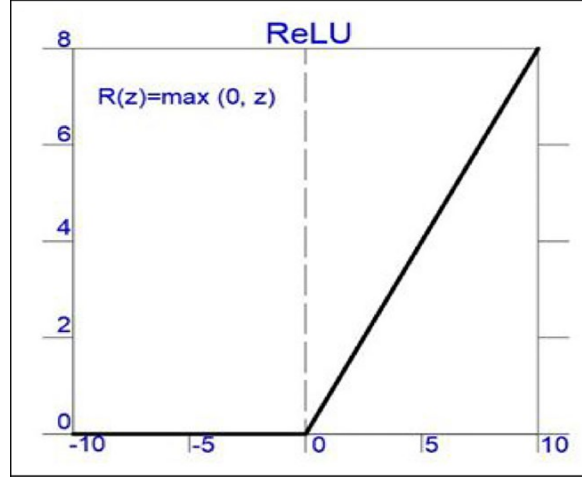
Şekil 3 Konvolüsyon işlemi

2. Aktivasyon Fonksiyonu (Activation Function):

- Aktivasyon fonksiyonları, nöronun çıktı değerini belirler. CNN'lerde yaygın olarak kullanılan aktivasyon fonksiyonu ReLU (Rectified Linear Unit) fonksiyonudur. ReLU, negatif değerleri sıfıra ayarlar ve pozitif değerleri olduğu gibi bırakır:

$$f(x) = \max(0, x) \quad (1.2)$$

- Bu, ağın doğrusal olmayan özellikleri öğrenmesini sağlar.



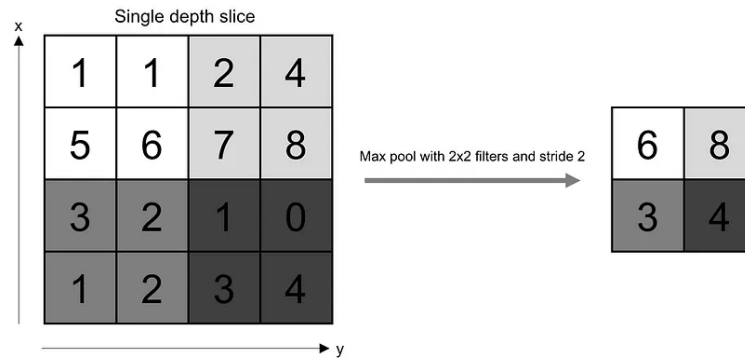
Şekil 4 Sigmoid grafiği

3. Havuzlama Katmanı (Pooling Layer):

- Havuzlama katmanları, özellik haritalarının boyutunu azaltarak hesaplama yükünü ve hafızayı azaltır. En yaygın kullanılan havuzlama yöntemleri maksimum havuzlama (max pooling) ve ortalama havuzlama (average pooling) dır.
- Maksimum havuzlama, belirli bir penceredeki en büyük değeri seçer:

$$y = \max(x_1, x_2, \dots, x_n) \quad (1.3)$$

- Havuzlama işlemi, özellik haritasının belirli bir penceresi üzerinde kaydırılarak uygulanır.



Şekil 5 Maxpooling işlemi

4. Tam Bağlantılı Katman (Fully Connected Layer):

- Tam bağlantılı katman, tüm nöronları önceki katmandaki tüm nöronlara bağlar. Bu katman, özellik haritalarını bir vektöre dönüştürür ve sınıflandırma görevini gerçekleştirir.
- Tam bağlantılı katmanın matematiksel gösterimi:

$$y = f(Wx + b) \quad (1.4)$$

- Burada W ağırlık matrisi, x giriş vektörü, b bias terimi ve f aktivasyon fonksiyonudur.

2.2. Genelleştirme Teknikleri:

Genelleştirme, bir derin öğrenme modelinin eğitim verisi dışındaki yeni ve görülmemiş veriler üzerinde ne kadar iyi performans gösterdiğini ifade eder. İyi bir genelleştirme, modelin eğitim sırasında öğrendiği bilgileri test verisi veya gerçek dünya verisi gibi yeni veriler üzerinde başarıyla uygulayabilmesi anlamına gelir. Genelleştirme, modelin sadece ezberlemesi değil, aynı zamanda verilerdeki genel desenleri ve özellikleri anlaması anlamına gelir.

Genelleştirme Problemleri

Genelleştirme problemleri genellikle iki türde ortaya çıkar:

1. Aşırı Uyum (Overfitting):

- Model, eğitim verisindeki gürültü veya rastgele desenleri öğrenir ve bu nedenle test verisinde kötü performans gösterir. Model, eğitim verisine çok fazla uyum sağladığında ortaya çıkar.

2. Eksik Uyum (Underfitting):

- Model, eğitim verisindeki temel desenleri ve ilişkileri öğrenemez. Bu durumda, model hem eğitim verisinde hem de test verisinde kötü performans gösterir.

Yaygın Kullanılan Genelleştirme Teknikleri

Genelleştirme performansını artırmak için çeşitli teknikler ve yöntemler kullanılır. İşte derin öğrenmede yaygın olarak kullanılan genelleştirme teknikleri:

1. Veri Artırma (Data Augmentation):

- Eğitim veri setinin boyutunu ve çeşitliliğini artırmak için kullanılan bir tekniktir. Veri artırma, görüntüler üzerinde döndürme, kırpma, yansıtma, parlaklık ve kontrast ayarlamaları gibi çeşitli işlemler uygulanarak gerçekleştirilir. Bu, modelin farklı varyasyonları görmesine ve daha genelleştirici olmasına yardımcı olur.

2. Düzenleştirme (Regularization):

- Düzenleştirme teknikleri, modelin aşırı uyum yapmasını önlemek için kullanılır. En yaygın düzenleştirme yöntemleri arasında L1 ve L2 düzenleştirme bulunur.
 - **L1 Düzenleştirme (Lasso):** Ağırlıkların mutlak değerlerinin toplamını ceza terimi olarak ekler.

$$L1 = \lambda \sum |w_i| \quad (1.5)$$

- **L2 Düzenleştirme (Ridge):** Ağırlıkların karelerinin toplamını ceza terimi olarak ekler.

$$L2 = \lambda \sum w_i^2 \quad (1.6)$$

3. Dropout:

- Dropout, her eğitim adımında rastgele seçilen bazı nöronları geçici olarak devre dışı bırakma tekniğidir. Bu, modelin belirli nöronlara veya özelliklere aşırı bağımlı olmasını engeller ve daha sağlam bir model oluşmasını sağlar.

$$y = f(Wx + b) \quad (1.7)$$

4. Erken Durdurma (Early Stopping):

- Eğitim sürecinde, modelin doğrulama seti üzerindeki performansı izlenir. Eğer doğrulama hatası belirli bir süre boyunca iyileşmezse, eğitim durdurulur. Bu, modelin aşırı uyum yapmasını önler.

5. Batch Normalizasyonu (Batch Normalization):

- Batch normalizasyonu, her mini-batch'in ortalama ve standart sapmasını kullanarak verileri normalleştirir. Bu, eğitim sürecini hızlandırır ve daha kararlı bir öğrenme sağlar.

6. Veri Setinin Bölünmesi (Train/Test Split):

- Eğitim ve test veri setlerinin doğru bir şekilde bölünmesi, modelin performansını gerçekçi bir şekilde değerlendirmek için önemlidir. Eğitim seti modelin öğrenmesi için kullanılırken, test seti modelin genelleştirme yeteneğini değerlendirmek için kullanılır.

7. Çapraz Doğrulama (Cross-Validation):

- Veri seti, eğitim ve doğrulama setlerine bölünerek birden fazla kez eğitilir ve test edilir. En yaygın yöntem k-katlı çapraz doğrulamadır (k-fold cross-validation), burada veri seti k eşit parçaya bölünür ve her parça bir kez doğrulama seti olarak kullanılır.

8. Veri Normalizasyonu ve Standardizasyonu:

- Verilerin ölçeklendirilmesi, modelin daha hızlı ve kararlı öğrenmesini sağlar. Normalizasyon ve standardizasyon, verilerin belirli bir aralıkta veya dağılımda olmasını sağlar.

2.3. Model Değerlendirme Metrikleri

Model değerlendirme metrikleri, bir makine öğrenmesi veya derin öğrenme modelinin performansını ölçmek için kullanılan araçlardır. Bu metrikler, modelin ne kadar iyi tahminler yaptığını, hangi alanlarda başarılı olduğunu ve hangi alanlarda iyileştirilmesi gerektiğini gösterir. Değerlendirme metrikleri, modelin doğruluğunu, hassasiyetini, hatalarını ve diğer performans göstergelerini objektif bir şekilde ölçer.

Konvolüsyonel Sinir Ağları (CNN) için en yaygın kullanılan değerlendirme metrikleri şunlardır:

1. Doğruluk (Accuracy):

- Doğruluk, doğru tahminlerin toplam tahminlere oranını ifade eder.

- Doğruluk = (Doğru Tahmin Sayısı) / (Toplam Tahmin Sayısı)
- Doğruluk, sınıflar arasında dengesiz veri dağılımı olan durumlarda yanıltıcı olabilir.

2. Hassasiyet (Precision):

- Hassasiyet, doğru pozitif tahminlerin toplam pozitif tahminlere oranıdır. Modelin doğru olarak sınıflandırdığı pozitif örneklerin oranını belirtir.
- Hassasiyet=Doğru Pozitif (TP)/Doğru Pozitif (TP)+Yanlış, Pozitif (FP)

3. Duyarlılık (Recall) :

- Duyarlılık, doğru pozitif tahminlerin toplam gerçek pozitif örneklerine oranıdır. Modelin doğru olarak tespit ettiği pozitif örneklerin oranını gösterir.
- Duyarlılık=Doğru Pozitif (TP)/Doğru Pozitif (TP)+Yanlış, Negatif (FN)

4. F1 Skoru (F1 Score):

- F1 skoru, hassasiyet ve duyarlılığın harmonik ortalamasıdır. Dengesiz veri setlerinde doğruluğa göre daha iyi bir performans ölçütü sunar.

5. Kesinlik (Specificity):

- Kesinlik, negatif sınıfların doğru olarak tahmin edilme oranını ifade eder.
- Kesinlik=Doğru Negatif (TN)/Doğru Negatif (TN)+Yanlış Pozitif (FP)

6. ROC Eğrisi (Receiver Operating Characteristic Curve) ve AUC (Area Under the Curve):

- ROC eğrisi, duyarlılık ve 1-spesifiklik (yanlış pozitif oran) değerlerini farklı eşik değerleri için gösterir. AUC, ROC eğrisinin altında kalan alanı ölçer ve modelin genel performansını değerlendirir.

- Yüksek AUC değerleri, modelin pozitif ve negatif sınıfları iyi ayırdığını gösterir.

7. Kayıp Fonksiyonu (Loss Function):

- Kayıp fonksiyonu, modelin tahminleri ile gerçek değerler arasındaki farkı ölçer. Eğitim sürecinde modelin optimize ettiği bir ölçüttür. Yaygın olarak kullanılan kayıp fonksiyonları arasında çapraz entropi (cross-entropy) ve ortalama kare hatası (mean squared error) bulunur.

$$\text{Çapraz Entropi Kaybı} = - (1/N) * \sum [y_i * \log(p_i) + (1 - y_i) * \log(1 - p_i)] \quad (1.8)$$

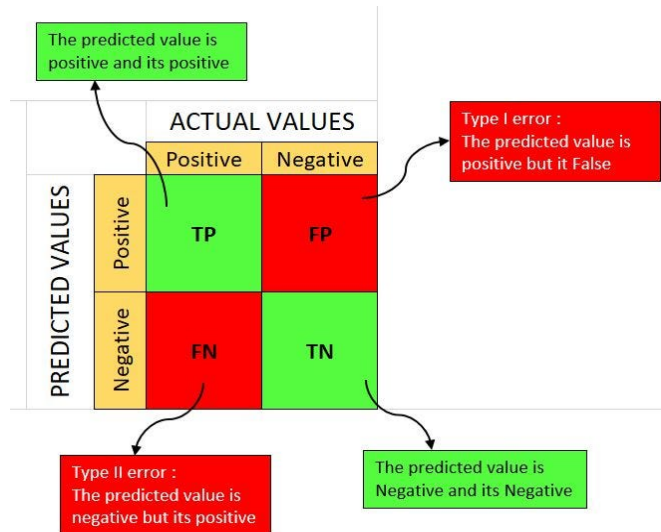
- Burada y_i gerçek etiket, p_i modelin tahmini, N toplam örnek sayısıdır.

8. Konfüzyon Matrisi (Confusion Matrix):

- Konfüzyon matrisi, modelin sınıflandırma sonuçlarını özetleyen bir tablodur. Her sınıf için doğru ve yanlış tahminlerin sayısını gösterir.

Konfüzyon matrisi, aşağıdaki dört bileşenden oluşur:

- Doğru Pozitif (TP)
- Yanlış Pozitif (FP)
- Yanlış Negatif (FN)
- Doğru Negatif (TN)



Şekil 6 Konfüzyon matrix grafik örneği

Sonuç olarak, model değerlendirme metrikleri, bir CNN modelinin performansını anlamak ve iyileştirmek için kritik öneme sahiptir. Doğruluk, hassasiyet, duyarlılık, F1 skoru ve ROC-AUC gibi metrikler, modelin farklı yönlerini değerlendirir ve genel performansını ölçer. Bu metrikler, modelin güçlü ve zayıf yönlerini belirlemeye yardımcı olur ve gelecekteki iyileştirmeler için yol gösterir.

BÖLÜM 3. VERİ SETİ VE ÖN İŞLEME

Ön işleme, ham veriyi modelin daha iyi anlaması ve daha iyi performans göstermesi için hazırlama sürecidir. CNN (Convolutional Neural Network) modelleri için ön işleme, verinin kalitesini artırarak ve modeli gereksiz karmaşıklıklardan kurtararak önemli avantajlar sağlar. İşte ön işlemenin başlıca nedenleri:

1. Veri Kalitesini Artırma:

- Ham verilerdeki gürültü, eksiklikler ve hatalar modelin performansını olumsuz etkileyebilir. Ön işleme, bu tür sorunları düzelterek veri kalitesini artırır.

2. Hesaplama Verimliliği:

- Büyük veri setleri ve yüksek çözünürlüklü görüntüler, hesaplama kaynaklarını zorlayabilir. Ön işleme, veri boyutunu ve karmaşıklığını azaltarak modelin daha hızlı ve verimli çalışmasını sağlar.

3. Genelleştirme Yeteneği:

- Modelin, eğitim verisi dışındaki yeni veriler üzerinde iyi performans göstermesi gerekir. Ön işleme teknikleri, modelin genelleştirme yeteneğini artırarak overfitting'i (aşırı uyum) önler.

4. Özellik Çıkarımı:

- Ön işleme, modelin verideki önemli özellikleri daha kolay öğrenmesini sağlar. Örneğin, görüntü normalizasyonu ve standardizasyonu, özelliklerin daha tutarlı ve anlamlı hale gelmesine yardımcı olur.

Yaygın Ön İşleme Teknikleri

CNN modelleri için yaygın olarak kullanılan ön işleme teknikleri şunlardır:

1. Görüntü Yeniden Boyutlandırma (Resizing):

- Görüntüler, modelin giriş katmanıyla uyumlu hale getirilmek üzere yeniden boyutlandırılır. Örneğin, tüm görüntülerin 224x224 piksel boyutuna getirilmesi.

2. Normalizasyon (Normalization):

- Görüntü piksellerinin değerleri genellikle 0-255 aralığında olur. Normalizasyon, bu değerleri 0-1 aralığına getirerek modelin daha hızlı ve etkili öğrenmesini sağlar.

3. Standardizasyon (Standardization):

- Verinin ortalama değeri 0 ve standart sapması 1 olacak şekilde ölçeklendirilir. Bu, özelliklerin daha dengeli ve karşılaştırılabilir olmasını sağlar.

4. Veri Artırma (Data Augmentation):

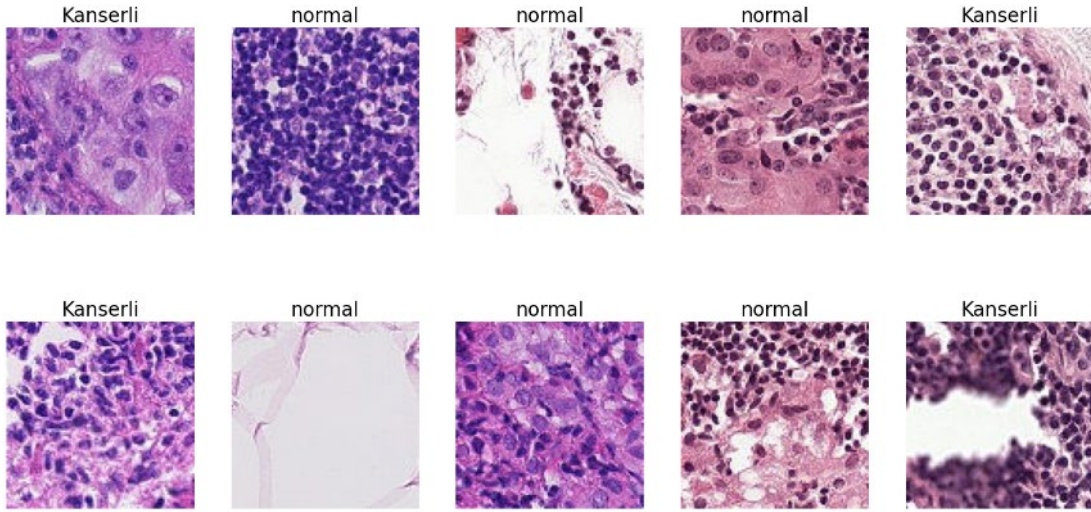
- Eğitim veri setinin boyutunu ve çeşitliliğini artırmak için görüntüler üzerinde döndürme, kırpma, yansıtma, parlaklık ayarlamaları gibi işlemler uygulanır. Bu, modelin daha genelleştirici olmasına yardımcı olur.

3.1. Veri Setinin Kaynağı Ve Özellikleri

Bu proje de PatchCamelyon (PCam) adlı veri seti kullanılmıştır. PatchCamelyon veri seti, yeni ve zorlu bir görüntü sınıflandırma veri kümesidir. Lenf düğümü kesitlerinin histopatolojik taramalarından çıkarılan 277.485 renkli görüntüden (96 x 96 piksel) oluşur. Her görüntü, metastatik dokunun varlığını belirten ikili bir etiketle bir csv dosyasında açıklanmıştır. PCam, makine öğrenimi ve derin öğrenme modelleri için yeni bir veri kaynağı imkanı sağlar: CIFAR10'dan daha büyük, ImageNet'ten daha küçük ve tek bir GPU üzerinde eğitilebilir.[3]

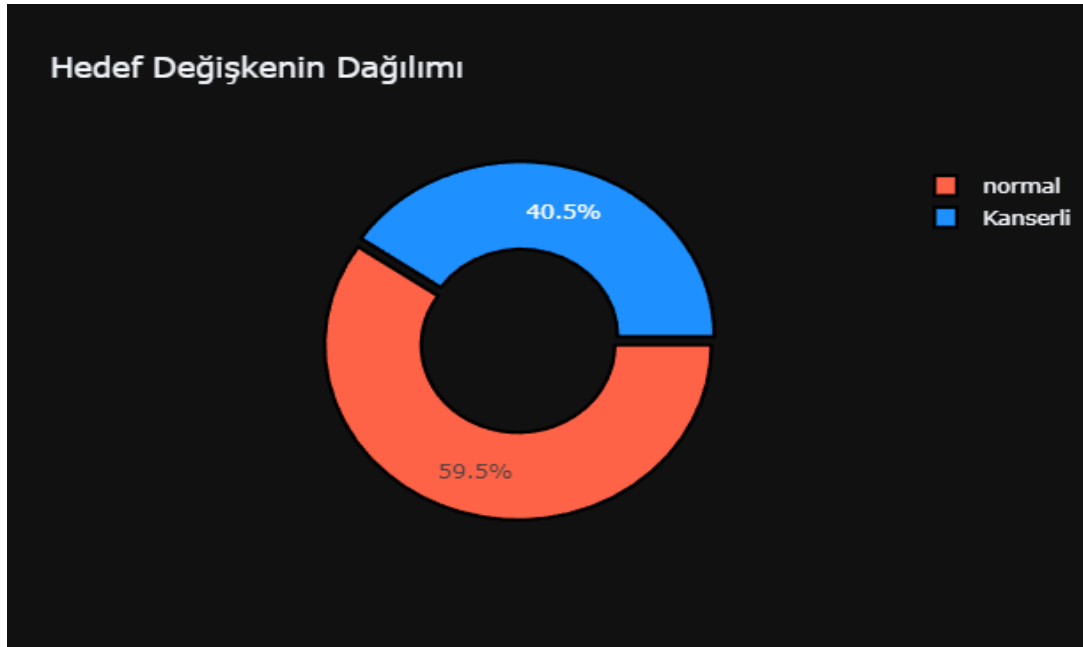
Bu veri setinde, sınıflandırmak için çok sayıda küçük patoloji görüntüsü sağlanmaktadır. Dosyalar, bir görüntü kimliğiyle adlandırılmıştır. Train_labels.csv dosyası, eğitim klasöründeki görüntüler için gerçek değerleri sağlar. Test klasöründeki görüntüler için label'ı tahmin etmek gerekmektedir. Pozitif bir label, bir yamanın merkezindeki 32x32 piksel bölgesinin en az bir tümör dokusu pikseli içerdiğini gösterir.

Train Veri Setinden Rastgele 10 Örnek Gözlem



Şekil 7 Veri setinden rastgele 10 örnek

Train veri setinde ki hedef değişkenin dağılımına baktığımda veri seti %59.5 kansersiz, %39.5 kanserli örneklerden oluşur.



Şekil 8 Veri setinin hedef değişken dağılımı

3.2. Veri Artırma

Veri artırma (Data Augmentation), makine öğrenmesi ve derin öğrenme modellerinin performansını ve genelleştirme yeteneklerini artırmak için eğitim veri setindeki çeşitliliği artıran bir tekniktir. Veri artırma, mevcut verilerin çeşitli dönüşümlerle çeşitlendirilmesiyle yapılır. Bu, modelin daha fazla veri görmesini sağlar ve aşırı uyumu (overfitting) önleyerek modelin yeni verilere daha iyi genelleme yapmasını sağlar[4].

Veri Artırmanın Önemi

1. **Aşırı Uyumun Önlenmesi:** Eğitim verilerindeki çeşitliliği artırarak modelin eğitim verisine aşırı uyum sağlamasını önler.
2. **Genelleştirme Yeteneği:** Modelin eğitim verisi dışındaki yeni ve görülmemiş verilerde daha iyi performans göstermesini sağlar.
3. **Veri Seti Boyutunu Artırma:** Küçük veri setlerinin boyutunu artırarak daha dengeli ve sağlam modeller oluşturulmasına yardımcı olur.

Yaygın Veri Artırma Teknikleri

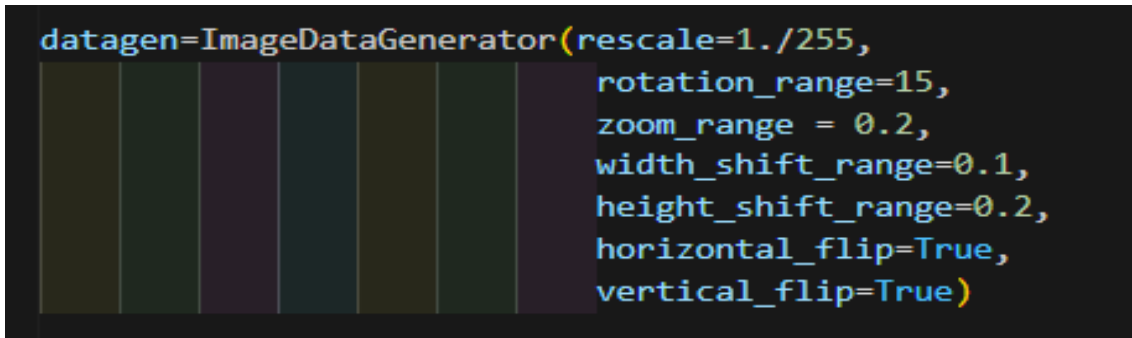
Veri artırma, çeşitli teknikler ve parametreler kullanılarak gerçekleştirilir. İşte yaygın kullanılan veri artırma teknikleri ve parametreleri:

1. **Döndürme (Rotation):**
 - Görüntülerin belirli bir açıda döndürülmesi.
 - Parametre: Döndürme açısı (örneğin, -30 ile +30 derece arasında rastgele döndürme).
2. **Yatay ve Dikey Çevirme (Horizontal and Vertical Flip):**
 - Görüntülerin yatay ve/veya dikey olarak çevrilmesi.
 - Parametre: Çevirme olasılığı (örneğin, %50 yatay çevirme).
3. **Yakınlaştırma ve Uzaklaştırma (Zoom In and Zoom Out):**
 - Görüntülerin rastgele oranlarda yakınlaştırılması veya uzaklaştırılması.
 - Parametre: Yakınlaştırma faktörü (örneğin, 0.8 ile 1.2 arasında).
4. **Kırpma (Cropping):**
 - Görüntülerin belirli bir bölümünün rastgele kesilmesi.
 - Parametre: Kırpma oranı (örneğin, %10 rastgele kırpma).

5. Parlaklık ve Kontrast Ayarları (Brightness and Contrast Adjustment):

- Görüntülerin parlaklık ve kontrastının rastgele ayarlanması.
- Parametre: Parlaklık ve kontrast faktörleri (örneğin, 0.8 ile 1.2 arasında).

Bu proje de kullanılan veri artırma parametreleri şunlardır:



Şekil 9 Veri artırma işlemi

Histopatolojik resimler, mikroskop altında alınan hücre ve doku görüntüleridir ve bu görüntülerdeki hücreler çeşitli açılarda, boyutlarda ve konumlarda olabilir. Kanser hücrelerinin tespiti için modelin bu varyasyonları öğrenebilmesi ve genelleştirme yeteneğinin artırılması gerekmektedir. Seçilen veri artırma parametreleri, bu hedefleri desteklemek amacıyla aşağıdaki nedenlerle seçilmiştir:

1. **Rescale:** Verilerin normalizasyonu, modelin daha hızlı ve stabil öğrenmesini sağlar.
2. **Rotation Range:** Hücrelerin farklı açılardaki görüntülerini öğrenerek modelin genelleştirme yeteneğini artırır.
3. **Zoom Range:** Farklı boyutlardaki hücreleri öğrenmek, modelin daha esnek ve genelleştirici olmasını sağlar.
4. **Width Shift ve Height Shift:** Hücrelerin farklı konumlarındaki varyasyonları öğrenmek, modelin farklı pozisyonlardaki hücreleri doğru bir şekilde tanımasını sağlar.

5. **Horizontal ve Vertical Flip:** Hücrelerin yönlerindeki değişiklikleri öğrenmek, modelin yatay ve dikey eksenlerdeki varyasyonlara karşı daha dayanıklı olmasını sağlar.

Bu parametrelerin kullanılması, modelin eğitim veri setindeki çeşitliliği artırarak daha sağlam ve genelleştirici bir model oluşturulmasına yardımcı olur. Bu, kanser hücrelerinin tespiti gibi kritik bir görevin daha yüksek doğrulukla yerine getirilmesini sağlar.

3.3. Veri Setinin Bölünmesi

Veri setinin bölünmesi, makine öğrenimi ve deep learning projelerinde yaygın olarak kullanılan bir yöntemdir ve genellikle veri setinin eğitim, doğrulama ve test olmak üzere üç ana bölüme ayrılmasını içerir. İşte bu bölümlerin her biri ve nedenleri:

1. Eğitim Seti (Training Set):

- **Tanım:** Modelin öğrenmesi için kullanılan veri setidir. Bu veri seti, modelin parametrelerini ayarlamak için kullanılır.
- **Neden Yapılır:** Model, veri setindeki örneklerden öğrenerek kendi ağırlıklarını günceller ve örüntüleri tanımaya çalışır.

2. Doğrulama Seti (Validation Set):

- **Tanım:** Modelin performansını değerlendirmek ve hiperparametre ayarlamaları yapmak için kullanılan veri setidir.
- **Neden Yapılır:** Modelin eğitim sürecinde aşırı öğrenme (overfitting) yapıp yapmadığını kontrol etmek için kullanılır. Doğrulama seti, modelin eğitim sırasında nasıl performans gösterdiğini anlamaya yardımcı olur ve modelin genel performansını optimize etmek için hiperparametrelerin ayarlanmasında kullanılır.

3. Test Seti (Test Set):

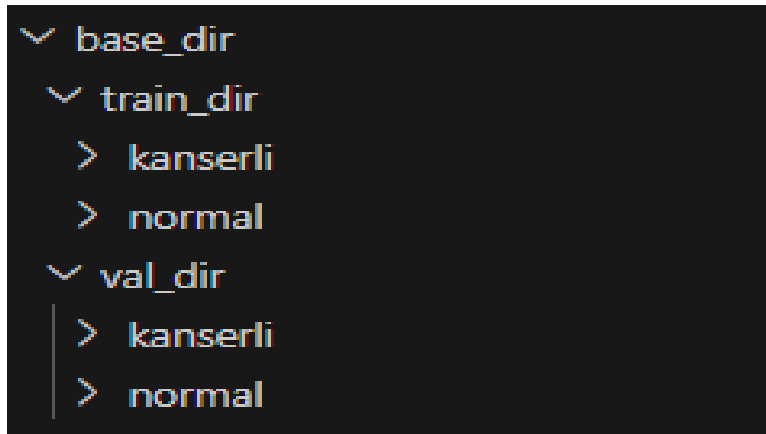
- **Tanım:** Modelin gerçek dünyada nasıl performans göstereceğini değerlendirmek için eğitim ve doğrulama sürecinden tamamen bağımsız olan veri setidir.
- **Neden Yapılır:** Modelin nihai doğruluğunu ve genelleme yeteneğini ölçmek için kullanılır. Test seti, modelin daha önce görmediği veriler üzerindeki performansını değerlendirmek için kullanılır.

Veri Setinin Bölünmesinin Nedenleri

- **Genelleme Yeteneği:** Modelin, eğitim setinde iyi performans göstermesi tek başına yeterli değildir. Modelin, daha önce görmediği yeni verilere de iyi genelleme yapabilmesi gerekir. Test seti, bu genelleme yeteneğini ölçmek için kullanılır.
- **Aşırı Öğrenmeyi Önleme:** Doğrulama seti, modelin aşırı öğrenme yapıp yapmadığını kontrol etmek için kullanılır. Aşırı öğrenme, modelin eğitim verilerine çok iyi uyum sağlaması ama yeni verilere genelleme yapamaması durumudur.
- **Model Seçimi ve Hiperparametre Ayarı:** Farklı modelleri karşılaştırmak ve en iyi model ile en iyi hiperparametreleri seçmek için doğrulama seti kullanılır.

Bu yöntem, makine öğrenimi projelerinin daha güvenilir ve genellenebilir sonuçlar üretmesini sağlar.[5]

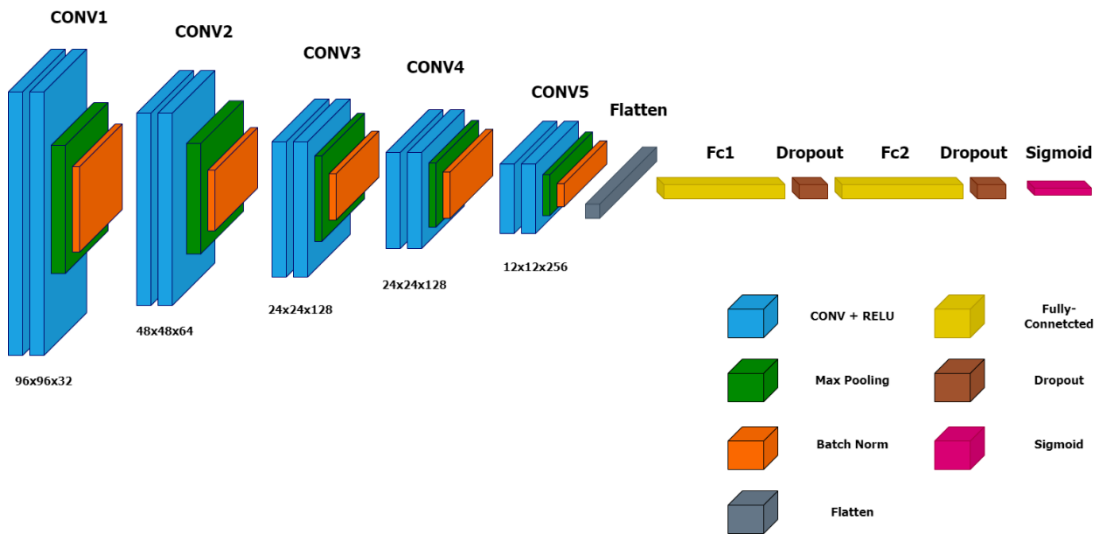
Bu proje de verilerin bölünmesi için ilk olarak sklearn train_test_split metodu kullanıldı. Train verisi %75 train,%25 validation set olarak ikiye ayrıldı. Bu ayrılma Train_labels.csv üzerinde yapıldı ardından PCAM verisetinde ki görüntüler ile 2'ye ayırdığımız csv dosyasında ki id'leri eşleştirerek base_dir dizini altında test_dir ve val_dir dosyaları onların da altında csv resimlerin csv dosyasında ki etiket bilgisine göre eşleştirilmiş resimler kanserli ve normal adlı dosyalara aktarıldı



Şekil 10 Oluşturulmuş Veri seti dosya yapısı

BÖLÜM 4. MODEL GELİŞTİRME

Kanser tespiti, erken teşhis ve tedavi planlaması açısından büyük önem taşımaktadır. Histopatolojik görüntü analizi, bu sürecin kritik bir parçasıdır ve bu alanda derin öğrenme teknikleri giderek daha yaygın bir şekilde kullanılmaktadır. Bu çalışmada, histopatolojik görüntülerden kanser tespiti yapmak amacıyla evrişimli sinir ağıları kullanılarak bir model geliştirilmiştir. CNN'ler, özellikle görüntü verisi üzerinde üstün performans gösteren derin öğrenme mimarileridir ve bu projede PatchCamelyon veri seti üzerinde eğitilmiştir.



Şekil 11 Oluşturulmuş olan CNN ağının mimarisi

4.1. Modelin Mimarisi Ve Katmanları

Model, bir dizi evrişimli katman, havuzlama katmanları, normalizasyon katmanları, yoğun bağlantılı katmanlar ve dropout katmanlarından oluşan derin bir sinir ağıdır. Aşağıda modelin her bir bileşenini ve seçilen parametrelerin faydalarını detaylı bir şekilde açıklanacaktır.

Evrişimli katmanlar, CNN'lerin temel yapı taşlarıdır ve görüntülerdeki özellikleri çıkarmak için kullanılır. Modelimizde, sırasıyla 32, 64, 128 ve 256 filtre içeren

evrişimli katmanlar kullanılmıştır. Her bir evrişim katmanında 3x3 boyutunda kernel (çekirdek) kullanılmıştır. 'relu' aktivasyon fonksiyonu, doğrusal olmayanlık ekleyerek modelin daha karmaşık örüntüleri öğrenmesini sağlar. 'Padding' parametresi 'same' olarak ayarlanmıştır, bu sayede giriş görüntüsü boyutları korunur ve kenar bilgileri kaybolmaz.

Evrişimli Katmanların Faydaları:

- **Özellik Çıkarımı:** Görüntülerdeki kenar, köşe ve dokular gibi düşük seviyeli özellikleri çıkarır.
- **Parametre Paylaşımı:** Aynı filtre, tüm görüntü üzerinde kaydırılarak hesaplama maliyetini ve modelin parametre sayısını azaltır.
- **Yerel Bağlantılar:** Görüntüdeki uzamsal ilişkileri öğrenir ve bu sayede nesnelerin konum ve biçim bağımsız özelliklerini yakalar.

Havuzlama katmanları, uzamsal boyutları küçültmek ve hesaplama maliyetini azaltmak için kullanılır. Modelimizde 2x2 boyutunda havuzlama uygulanmıştır. MaxPooling işlemi, belirli bir bölgede maksimum değeri alarak bu bölgedeki önemli özellikleri korur.

Havuzlama Katmanlarının Faydaları:

- **Boyut Azaltma:** Uzamsal boyutları küçülterek hesaplama maliyetini ve modelin parametre sayısını azaltır.
- **Özelliklerin Korunması:** Maksimum değeri alarak önemli özelliklerin korunmasını sağlar.
- **Aşırı Öğrenmeyi Önleme:** Modelin genel performansını artırarak aşırı öğrenmeyi önler.

BatchNormalization katmanları, her mini-batch'in çıktılarını normalleştirerek modelin daha hızlı ve stabil öğrenmesini sağlar. Bu katmanlar, modelin eğitimi sırasında oluşabilecek gradyan patlaması ve kaybolmasını önler.

Normalizasyon Katmanlarının Faydaları:

- **Hızlı Öğrenme:** Modelin daha hızlı ve stabil bir şekilde öğrenmesini sağlar.
- **Gradyan Patlaması ve Kaybolmasını Önleme:** Gradyanların daha dengeli bir şekilde akmasını sağlar.
- **Düzenleştirme Etkisi:** Modelin aşırı öğrenme yapmasını önleyerek genelleme yeteneğini artırır.

Modelin son karar katmanları olan yoğun bağlantılı katmanlarda, sırasıyla 256 ve 128 nöron kullanılmıştır. Bu katmanlarda 'relu' aktivasyon fonksiyonu kullanılarak doğrusal olmayanlık eklenmiştir. Ağırlıkların aşırı uyumunu (overfitting) önlemek için L2 düzenleyici (L2 regularizer) eklenmiştir.

Yoğun Bağlantılı Katmanların Faydaları:

- **Karmaşık Öğrenme:** Görüntüdeki yüksek seviyeli özellikleri öğrenir ve sınıflandırma yapar.
- **Aşırı Öğrenmeyi Önleme:** L2 düzenleyici ile aşırı öğrenme önlenir.
- **Genelleme Yeteneği:** Modelin yeni veriler üzerinde daha iyi performans göstermesini sağlar.

Aşırı öğrenmeyi önlemek amacıyla kullanılan dropout katmanları, bazı nöronların rastgele olarak devre dışı bırakıldığı katmanlardır. Bu çalışmada %30 dropout uygulanmıştır.

Dropout Katmanlarının Faydaları:

- **Aşırı Öğrenmeyi Önleme:** Rastgele nöronları devre dışı bırakarak modelin genelleme yeteneğini artırır.
- **Düzenleştirme:** Modelin daha robust ve stabil öğrenmesini sağlar.

Modelin nihai kararını veren çıkış katmanında tek bir nöron ve sigmoid aktivasyon fonksiyonu kullanılmıştır. Bu sayede ikili sınıflandırma problemi çözülür.

Modelin öğrenme sürecini optimize etmek için Adam (Adaptive Moment Estimation) optimizasyon algoritması kullanılmıştır. Adam, öğrenme hızını (learning rate) ve momentum terimlerini otomatik olarak ayarlayarak daha hızlı ve stabil bir öğrenme süreci sağlar. Modelin optimizasyon ve derleme aşamaları aşağıdaki gibidir:

Adam Optimizatörünün Faydaları ve Seçilen Parametreler

- **Learning Rate (Öğrenme Hızı):** 0.0001 olarak belirlenmiştir. Küçük bir öğrenme hızı, modelin daha küçük adımlarla ve daha dikkatli öğrenmesini sağlar, bu da modelin daha doğru sonuçlar üretmesine yardımcı olur.

- **Beta 1 ve Beta 2:** 0.9 ve 0.999 olarak ayarlanmıştır. Bu parametreler, hareketli ortalama ve ikinci moment tahmini için kullanılır ve modelin daha dengeli öğrenmesini sağlar.
- **Epsilon:** 1e-07 olarak belirlenmiştir. Bu küçük sabit, sıfıra bölme hatalarını önlemek için kullanılır.
- **Amsgrad:** False olarak ayarlanmıştır. Bu parametre, öğrenme hızının adaptif olarak ayarlanmasında kullanılan bir yöntemdir.

Adam Optimizatörünün Faydaları:

- **Hızlı ve Kararlı Öğrenme:** Öğrenme hızını ve momentum terimlerini otomatik olarak ayarlayarak daha hızlı ve stabil bir öğrenme süreci sağlar.
- **Adaptif Öğrenme:** Her parametre için ayrı öğrenme hızı belirleyerek modelin daha verimli öğrenmesini sağlar.
- **Genelleme Yeteneği:** Modelin daha geniş veri setleri üzerinde daha iyi performans göstermesini sağlar.

Derleme Parametreleri

- **Loss Function (Kayıp Fonksiyonu):** 'binary_crossentropy' olarak seçilmiştir. Bu fonksiyon, ikili sınıflandırma problemleri için uygun bir kayıp fonksiyonudur ve modelin doğru sınıflandırma yapmasını sağlar.
- **Metrics (Değerlendirme Metrikleri):** 'accuracy' olarak belirlenmiştir. Modelin doğruluğunu ölçmek için kullanılır.

Evrişimli Sinir Ağları, özellikle görüntü verisi üzerinde üstün performans gösteren derin öğrenme mimarileridir. Bu mimarilerin histopatolojik görüntülerden kanser tespiti için uygun olmasının başlıca nedenleri şunlardır:

- **Özellik Öğrenme:** CNN'ler, manuel özellik mühendisliğine gerek kalmadan görüntülerdeki önemli özellikleri otomatik olarak öğrenir.
- **Uzamsal Bilgi:** CNN'ler, görüntülerdeki uzamsal ilişkileri koruyarak nesnelerin konum ve biçim bağımsız özelliklerini yakalar.
- **Derinlik ve Karmaşıklık:** Birden fazla evrişim ve havuzlama katmanı ile düşük seviyeli özelliklerden yüksek seviyeli özelliklere kadar karmaşık örüntüleri öğrenir.
- **Genelleme Yeteneği:** Normalizasyon ve dropout gibi teknikler kullanılarak aşırı öğrenme önlenir ve modelin yeni veriler üzerindeki performansı artırılır.

Geliştirilen bu CNN modeli, histopatolojik görüntülerden kanser tespiti yapma konusunda etkili bir araçtır. Modelin derin yapısı ve seçilen hiperparametreler, veri setindeki karmaşık örüntüleri öğrenmesini ve genelleme yeteneğini artırmasını sağlar. Bu çalışma, derin öğrenme tekniklerinin tıbbi görüntü analizi gibi kritik alanlarda nasıl kullanılabileceğini göstermektedir ve kanser tespiti için önemli bir adım oluşturmaktadır.

4.2. Modelin Eğitim Süreci

Modelin eğitim süreci, eğitim verilerinin belirli bir sayıda döngü (epoch) boyunca ağ üzerinde işlenmesi ve bu süreçte modelin parametrelerinin optimize edilmesini içerir. Eğitim sürecinde kullanılan temel bileşenler ve yöntemler aşağıda detaylandırılmıştır.

Modelin eğitimi için Adam optimizasyon algoritması kullanılmıştır. Adam, stochastic gradient descent (SGD) yöntemine dayanan ve öğrenme oranını adaptif olarak ayarlayabilen bir optimizasyon yöntemidir. Kullanılan parametreler şu şekildedir:

- **learning_rate**: 0.0001 - Öğrenme oranı, modelin ağırlıklarının ne kadar hızlı güncelleneceğini belirler. Düşük bir öğrenme oranı daha yavaş ve potansiyel olarak daha kararlı bir öğrenme süreci sağlar.
- **beta_1**: 0.9 - İlk moment tahminine verilen ağırlık.
- **beta_2**: 0.999 - İkinci moment tahminine verilen ağırlık.
- **epsilon**: 1e-07 - Sayısal kararlılığı artırmak için kullanılan çok küçük bir değerdir.
- **amsgrad**: False - AMSGrad varyantının kullanılıp kullanılmayacağını belirler.

Model bu optimizasyon yöntemi kullanılarak, **binary_crossentropy** kayıp fonksiyonu ile derlenmiştir. Kayıp fonksiyonu, modelin tahminleri ile gerçek değerler arasındaki farkı ölçer ve bu farkın minimize edilmesi amaçlanır. Performans metriği olarak **accuracy** (doğruluk) kullanılmıştır.

Eğitim sürecinde kullanılan callback'ler, modelin performansını optimize etmek ve eğitim sürecini daha verimli hale getirmek için kullanılır. Bu projede üç farklı callback kullanılmıştır: EarlyStopping, ModelCheckpoint ve ReduceLROnPlateau.

```
# callbacks

# EarlyStopping: Stop training when a monitored quantity has stopped improving.
early_stop=EarlyStopping(monitor='val_loss', patience=3, verbose=1)

# ModelCheckpoint: Save the model after every epoch.
save_path='model_history.keras'
checkpoint=ModelCheckpoint(save_path, monitor='val_accuracy', verbose=1, save_best_only=True, mode='max')

# reduce_lr: Reduce learning rate when a metric has stopped improving.
reduce_lr=ReduceLROnPlateau(monitor='val_loss', factor=0.5, patience=1, verbose=1, min_lr=0.000001)
```

Şekil 12 Callbacks'in kodlanması

1. EarlyStopping

Bu callback, izlenen metrik belirli bir süre boyunca iyileşme göstermediğinde eğitim sürecini durdurur. Kullanılan parametreler:

- **monitor:** 'val_loss' - İzlenen metrik doğrulama kaybı.
- **patience:** 3 - Metrikte iyileşme olmadan beklenen epoch sayısı.
- **verbose:** 1 - Detaylı bilgi çıktısı.

Yorum: Bu parametreler, modelin gereksiz yere fazla epoch boyunca eğitilmesini engeller ve aşırı öğrenmenin (overfitting) önüne geçilmesine yardımcı olur.

2. ModelCheckpoint

Bu callback, her epoch sonunda modeli kaydeder. Sadece en iyi performansa sahip model kaydedilir. Kullanılan parametreler:

- **monitor:** 'val_accuracy' - İzlenen metrik doğrulama doğruluğu.
- **verbose:** 1 - Detaylı bilgi çıktısı.
- **save_best_only:** True - Sadece en iyi modelin kaydedilmesini sağlar.
- **mode:** 'max' - En yüksek doğruluğa sahip modelin kaydedileceğini belirtir.

Yorum: Bu callback, en iyi performansı sağlayan modelin elde tutulmasını ve eğitim sürecinde herhangi bir noktada en iyi modele geri dönülebilmesini sağlar.

3. ReduceLROnPlateau

Bu callback, izlenen metrik iyileşme göstermediğinde öğrenme oranını azaltır. Kullanılan parametreler:

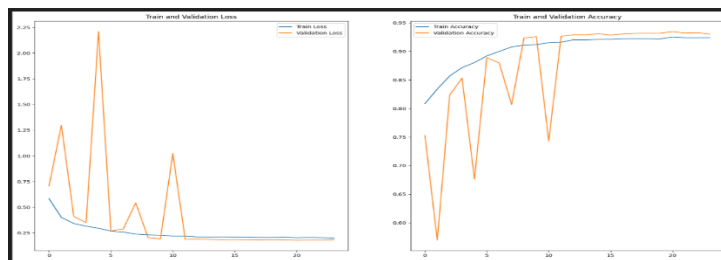
- **monitor:** 'val_loss' - İzlenen metrik doğrulama kaybı.
- **factor:** 0.5 - Öğrenme oranının ne kadar azaltılacağını belirler.
- **patience:** 1 - Metrikte iyileşme olmadan beklenen epoch sayısı.
- **verbose:** 1 - Detaylı bilgi çıktısı.
- **min_lr:** 0.000001 - Öğrenme oranının düşebileceği en düşük değer.

Yorum: Bu callback, öğrenme sürecinde duraklamaları önlemek ve modelin daha iyi bir minimuma ulaşmasını sağlamak için öğrenme oranını dinamik olarak ayarlar.

Model, 25 epoch boyunca eğitilmiştir. EarlyStopping callback'i sayesinde, doğrulama kaybı üç epoch boyunca iyileşme göstermediğinde eğitim süreci otomatik olarak durdurulmuştur. Bu, gereksiz epoch'lar boyunca eğitim yapılmasını önlemiş ve zaman kazandırmıştır. ModelCheckpoint callback'i ile en iyi doğrulama doğruluğuna sahip model kaydedilmiş ve ReduceLROnPlateau callback'i ile modelin öğrenme oranı gerektiğinde azaltılmıştır.

Eğitim süreci boyunca kullanılan bu callback'ler, modelin eğitim performansını optimize etmiş ve en iyi sonuçların elde edilmesini sağlamıştır. Model, eğitim verisi üzerinde iyi performans gösterirken, doğrulama verisi üzerinde de yeterli genel performans sağlamıştır. Bu sayede, overfitting ve underfitting gibi problemler minimize edilmiştir.

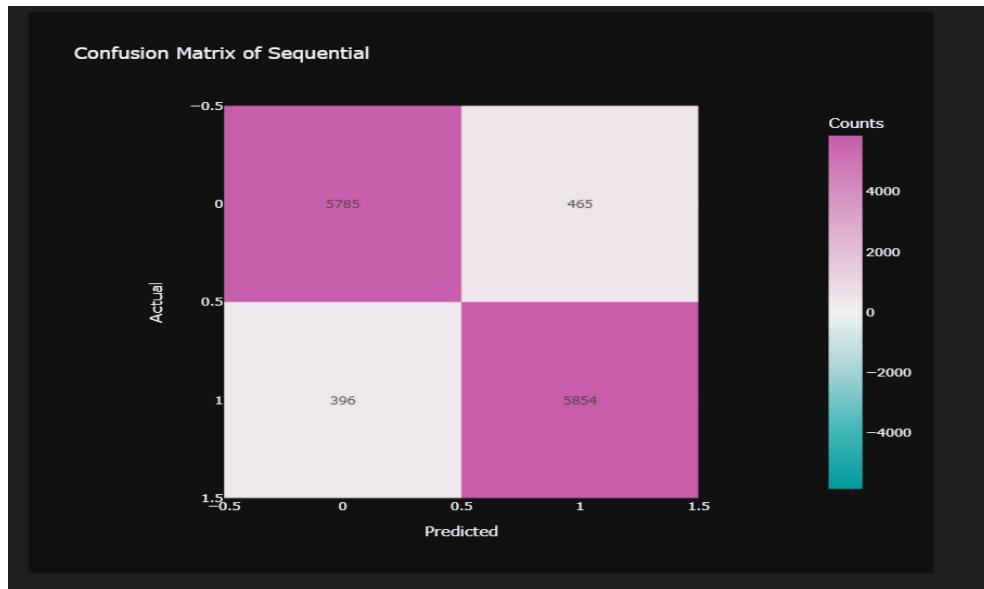
4.3. Modelin Değerlendirilmesi



Şekil 13 Modelin eğitim sonuçları

Grafikte, eğitim ve doğrulama kayıplarının epoch'lar boyunca değişimi gösterilmiştir. Eğitim kaybı düzenli olarak azalmaktadır, bu da modelin eğitim verisine daha iyi uyum sağladığını göstermektedir. Ancak, doğrulama kaybında belirgin dalgalanmalar gözlemlenmiştir. Bu dalgalanmalar, modelin bazı epoch'larda overfitting'e eğilimli olabileceğini işaret eder.

Eğitim doğruluğu düzenli olarak artmakta ve doğrulama doğruluğu da genellikle yükselmektedir. Ancak doğrulama doğruluğunda da bazı dalgalanmalar mevcuttur. Bu dalgalanmalar, modelin zaman zaman doğrulama verisine iyi genelleme yapamadığını gösterebilir.



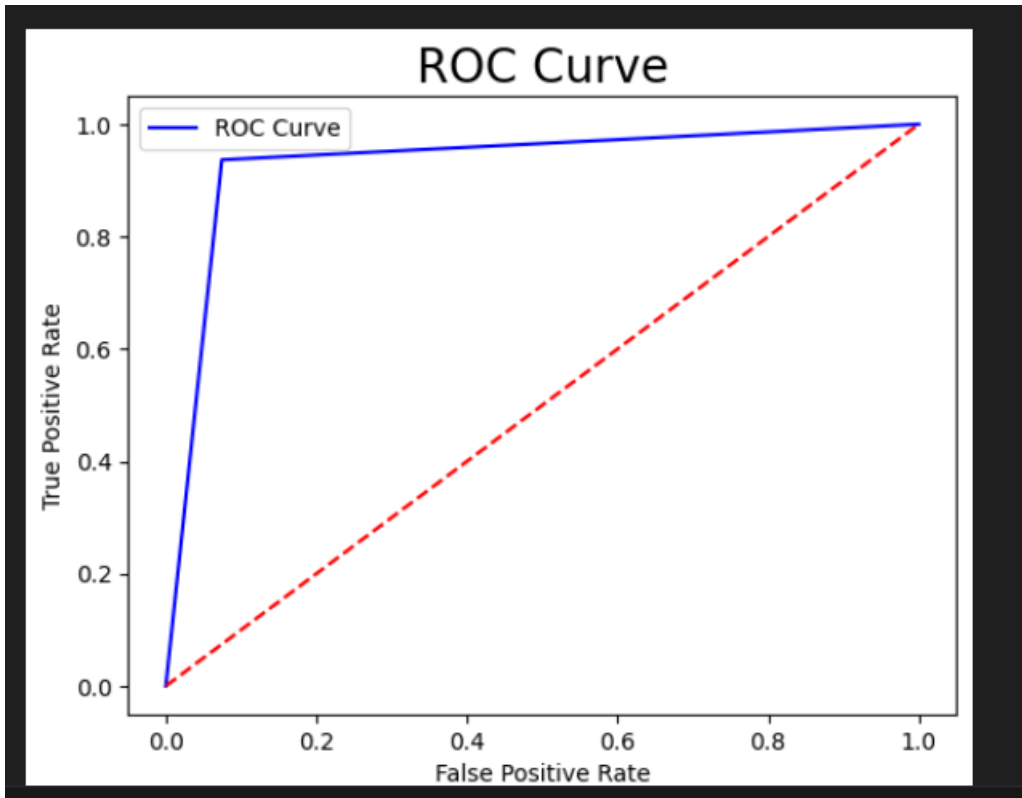
Şekil 14 Modelin konfüzyon matrix'i

Karışıklık matrisi, modelin sınıflandırma performansını değerlendirmek için kullanılır. Matristeki değerler, modelin gerçek pozitifler, gerçek negatifler, yanlış pozitifler ve yanlış negatifler ile ilgili doğru ve yanlış tahminlerini gösterir:

- **Gerçek Pozitifler (TP):** 5785
- **Yanlış Pozitifler (FP):** 465
- **Yanlış Negatifler (FN):** 396
- **Gerçek Negatifler (TN):** 5854

Bu deęerler, modelin doęruluęunu ve hata oranlarını hesaplamak için kullanılır.

ROC eęrisi, modelin sınıflandırma performansını gösterir. Eęri altındaki alan (AUC), modelin ayırma yeteneęini ölçer. Bu eęrinin grafięinde, modelin neredeyse mükemmel bir ayırım kapasitesine sahip olduęu görölmektedir (AUC deęeri 1'e çok yakın).



Şekil 15 Modelin AUC-ROC grafięi

Sınıflandırma raporunda precision, recall ve F1-score deęerleri řu řekildedir:

- **Precision (Kesinlik):** Modelin doęru pozitif tahminlerinin tüm pozitif tahminlere oranı.
 - Sınıf 0: 0.94
 - Sınıf 1: 0.93
- **Recall (Duyarlılık):** Modelin doęru pozitif tahminlerinin tüm gerçek pozitiflere oranı.
 - Sınıf 0: 0.93

- Sınıf 1: 0.94
- **F1-score:** Precision ve recall'un harmonik ortalaması.
 - Sınıf 0: 0.93
 - Sınıf 1: 0.93
- **Accuracy (Doğruluk):** Tüm doğru tahminlerin toplam tahminlere oranı: 0.93

Tablo 1.0.Sınıflandırma raporu

Kategori	Precision	Recall	F1-score	Support
0	94%	93%	93%	6250
1	93%	93%	93%	6250
Değerlendirme	Accuracy	Macro avg (kategoriler arasındaki ortalama performans)	Weighted avg (Farklı kategori boyutları dikkate alınarak dengeli ortalama)	Toplam Örnekler
	93%	93%	93%	12500

Bu değerler, modelin genel olarak yüksek bir doğruluk ve güvenilirlik seviyesine sahip olduğunu göstermektedir.

Eğitim ve doğrulama kayıpları grafiğine bakıldığında, eğitim kaybının sürekli bir azalma gösterdiği, ancak doğrulama kaybının dalgalanmalar sergilediği görülmektedir. Bu durum, modelin bazı epoch'larda doğrulama verisine iyi genelleme yapamadığını ve overfitting'e yatkın olduğunu işaret eder. Eğitim doğruluğu ve doğrulama doğruluğu grafiği de bu durumu desteklemektedir; eğitim doğruluğu sürekli artarken, doğrulama doğruluğunda dalgalanmalar gözlenmiştir.

Karışıklık matrisi, modelin genel performansını değerlendirmek için önemlidir. Gerçek pozitif ve gerçek negatiflerin yüksek olması, modelin doğru tahminler yapma konusunda başarılı olduğunu göstermektedir. Yanlış pozitif ve yanlış negatif

sayılarının nispeten düşük olması, modelin güvenilirlik düzeyinin yüksek olduğunu işaret eder.

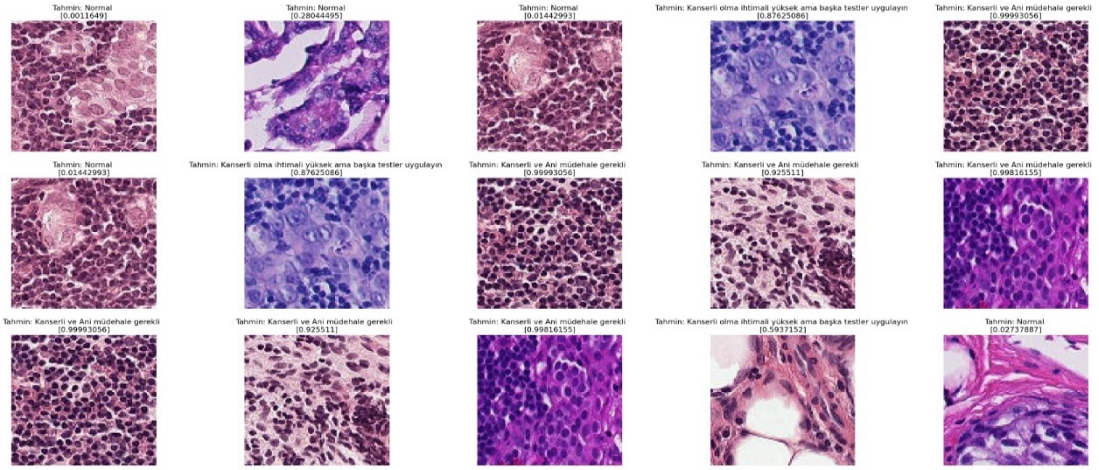
ROC eğrisi ve AUC değeri, modelin sınıflandırma performansını öne çıkarmaktadır. AUC değerinin 1'e çok yakın olması, modelin neredeyse mükemmel bir ayırım yeteneğine sahip olduğunu göstermektedir. Bu, modelin farklı sınıflar arasında yüksek bir doğrulukla ayırım yapabildiğini kanıtlar.

Sınıflandırma raporu, precision, recall ve F1-score değerleri ile modelin sınıflandırma yeteneğini ayrıntılı bir şekilde ortaya koymaktadır. Her iki sınıf için de yüksek precision ve recall değerleri, modelin hem pozitif hem de negatif örnekleri başarılı bir şekilde tanımlayabildiğini göstermektedir. F1-score değerlerinin yüksek olması, modelin hem precision hem de recall açısından dengeli bir performansa sahip olduğunu gösterir.

Modelin genel performansı ve elde edilen sonuçlar, projenin amaçlarına ulaşmada başarılı olduğunu göstermektedir. Gelecek çalışmalarda, daha büyük veri setleri ve farklı model mimarileri ile performansın daha da artırılması mümkündür. Ayrıca, modelin klinik ortamda uygulanabilirliğini değerlendirmek için daha fazla test yapılması önerilir.

4.4. Modelin Test Edilmesi

Modelin testi için test veri setinden 20 örnek seçildi. Tahmini yapılan resmin prediction sonucu %90'nın üstünde ise "Kanserli ve ani müdahale gerekli " sonucu gösterilir. Prediction sonucu %50'den büyük ve %90'dan küçük ise "Kanserli olma ihtimali yüksek ama başka testler deneyin" sonucu gösterilir. Eğer prediction sonucu %50'den küçük ama %30'dan büyükse "Kanserli olma ihtimali düşük ama başka testler uygulayın" sonucu gösterilir. Diğer durumlarda ise "Normal" sonucu gösterilir. Bu model sonrası yapılan testtir. Web arayüzünde farklı sonuç cümleleri kullanılacaktır.



Şekil 16 Modelin test sonuçları

BÖLÜM 5. WEB SAYFASI ÜZERİNDEN MODELİN SUNUMU

Histopatolojik resimlerden kanser tespiti, modern tıbbın en zorlu ve en kritik alanlarından biridir. Bu süreç, genellikle patologlar tarafından mikroskop altında hücrelerin incelenmesini gerektirir ve bu da zaman alıcı ve insan hatasına açık bir işlemdir. Bu projenin amacı, konvolüsyonel sinir ağları (CNN) kullanarak histopatolojik resimlerden otomatik olarak kanser tespiti yapabilen bir model geliştirmektir. Böylece, kanser teşhisinde doğruluk ve hız artırılabilir, patolojik analizlerin etkinliği artırılabilir ve hastaların erken teşhis ile tedavi şansı yükseltilebilir.

Modelin web sayfası üzerinden sunulması, kullanıcıların kolayca erişebileceği ve kullanabileceği bir platform sağlayarak, projenin daha geniş bir kitleye ulaşmasını ve pratik bir uygulama alanı bulmasını sağlar. Bu bölümde, modelin web sayfası üzerinden sunumunun teknik detayları ve bu yaklaşımın sağladığı avantajlar ele alınacaktır.

5.1. Web Sitesi Mimarisi Ve Teknolojileri

Projenin web sitesi, kullanıcıların histopatolojik resimleri yükleyip modelin tahmin sonuçlarını alabilecekleri bir platform olarak tasarlanmıştır. Bu mimari, frontend ve backend bileşenlerinden oluşmaktadır.

Frontend kısmı, kullanıcı arayüzünün oluşturulması için HTML, CSS ve JavaScript kullanılarak geliştirilmiştir. Bu teknolojiler, kullanıcıların etkileşimli ve kullanıcı dostu bir deneyim yaşamasını sağlar.

HTML (HyperText Markup Language): Web sayfasının yapısını oluşturmak için kullanılmıştır. Formlar, düğmeler ve görüntü alanları gibi temel bileşenler HTML ile tanımlanmıştır.

CSS (Cascading Style Sheets): Web sayfasının görünümünü ve düzenini belirlemek için kullanılmıştır. CSS, kullanıcı arayüzünü estetik ve kullanıcı dostu hale getirir.

JavaScript: Kullanıcı etkileşimlerini yönetmek ve dinamik içerik sağlamak için kullanılmıştır. JavaScript, kullanıcıdan gelen görüntülerin backend'e gönderilmesi ve tahmin sonuçlarının alınarak gösterilmesi için önemli bir rol oynar.

Backend kısmı, Flask framework kullanılarak geliştirilmiştir. Flask, Python tabanlı hafif bir web framework'tür ve hızlı ve kolay bir şekilde web uygulamaları geliştirmeyi sağlar.

Flask Framework: Flask, HTTP isteklerini işlemek, modelin tahmin işlemlerini gerçekleştirmek ve sonuçları geri döndürmek için kullanılmıştır.

Model Entegrasyonu: Flask backend, kullanıcıdan gelen görüntüyü alır, gerekli ön işlemleri gerçekleştirir ve modeli kullanarak tahmin yapar. Tahmin sonuçları JSON formatında frontend'e geri döner.

5.2.Web Sitesinin Gösterimi

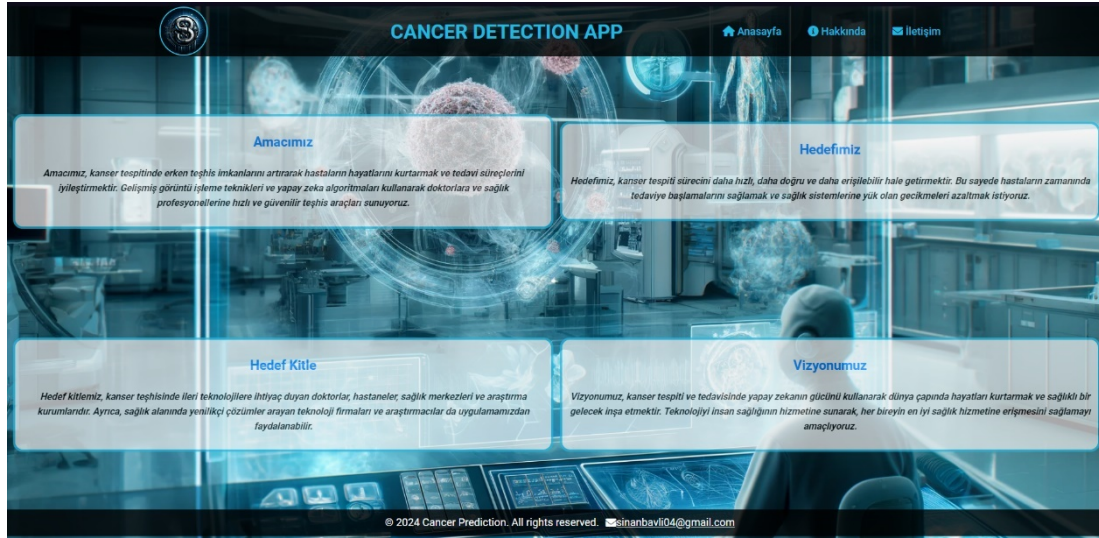
Oluşturulan Web Arayüzü 4 kısımdan oluşmaktadır. Bunlar Anasayfa, Hakkında, İletişim ve Sonuç ekranı. Sonuç ekranına yalnız tahminleme yapıldıktan sonra erişilebilir.

Home kısmında kısımda kullanıcı tarafından histopatolojik resim yüklenir ve tahmin yap butonuna basılarak entegre edilmiş model tarafından tahmin yapılır.



Şekil 17 Web sitesinin anasayfası

Hakkında kısmında amacımız,hedefimiz, hedef kitle ve vizyonumuz başlıklı yazılar vardır.



Şekil 18 Web sitesinin hakkında kısmı

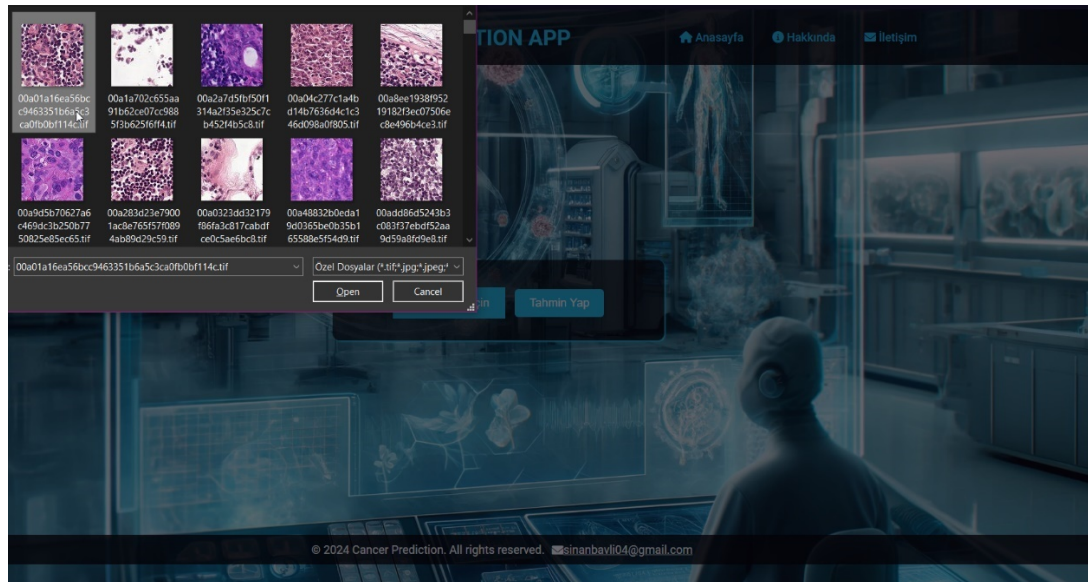
İletişim kısmında address, telefon,mail ve çeşitli sosyal medya bilgileri mevcuttur.



Şekil 19 Web Sitesinin iletişim kısmı

5.3. Modelin Web Sitesi Üzerinden Kullanımı

Modelin kullanımında ilk olarak kullanıcı tarafından bir histopatolojik resim yüklenir web sitesine anasayfa üzerinden ardından tahmin et butonuna basılır daha sonra web sitesine entegre edilmiş model tahminleme yapar ve sonuçlar Sonuç ekranında gösterilir. Tahminleme sonucunda 4 farklı sonuç oluşabilir. Eğer tahminleme sonucu başarısı %50'ın üstünde ise “Yüklenen doku örneğinde %(başarı oranı) ihtimalle metastatik kanser olduğu tespit edildi. Lütfen ani müdahale edin” uyarısı gösterilir. Eğer modelin başarısını oranı %50'dan küçükçe “Yüklenen doku örneğinde metastatik kanser olmadığı tespit edildi.” uyarısı gösterilir .



Şekil 20 Anasayfa üzerinden dosya yükleme



Şekil 21 Tahmin sonuç örneği

Yukardaki şekil 21’de görüldüğü üzere girilen resimde %99.02 ihtimalle kanser tespit edilmiştir.

BÖLÜM 6. SONUÇ

Bu bitirme projesi kapsamında geliştirilen model, histopatolojik resimlerden kanser tespiti yapabilmekte ve bu alanda önemli bir başarı sağlamaktadır. Konvolüsyonel sinir ağları (CNN) kullanılarak oluşturulan model, hem doğruluk hem de güvenilirlik açısından yüksek performans sergilemiştir. Modelin web sayfası üzerinden sunulması, kullanıcıların kolayca erişim sağlayabilmesi ve sonuçlara hızlı bir şekilde ulaşabilmesi açısından büyük avantajlar sunmaktadır.

Web tabanlı uygulama, kullanıcılara histopatolojik görüntüleri yükleyerek anında kanser tespiti yapabilme imkanı tanımakta, bu da klinik ortamlarda hızlı ve etkili karar verme süreçlerini desteklemektedir. Ayrıca, web uygulamasının kolay güncellenebilir ve bakım yapılabilir olması, modelin sürekli olarak iyileştirilmesini ve en güncel sürümünün kullanıcıların hizmetine sunulmasını sağlamaktadır.

Bu proje, kanser teşhisinde yapay zeka ve derin öğrenme yöntemlerinin etkin kullanımını göstererek, sağlık hizmetlerinin kalitesini artırmaya yönelik önemli bir adım atmıştır. Modelin yüksek doğruluk ve güvenilirlik seviyeleri, gerçek dünyada uygulanabilirliğini kanıtlamakta ve klinik ortamda kullanılabilirliğini desteklemektedir.

Projenin başarısı, gelecekte daha büyük veri setleri ve farklı model mimarileri ile yapılacak çalışmalar için sağlam bir temel oluşturmaktadır. Ayrıca, modelin klinik ortamda gerçek zamanlı uygulamalar için optimize edilmesi ve doğrulama süreçlerinin devam ettirilmesi, bu tür yapay zeka tabanlı çözümlerin sağlık sektöründe daha geniş bir şekilde benimsenmesini sağlayabilir.

Sonuç olarak, bu bitirme projesi, histopatolojik resimlerden kanser tespiti konusunda yenilikçi ve etkili bir çözüm sunmakta ve yapay zeka teknolojilerinin sağlık alanındaki potansiyelini gözler önüne sermektedir. Projenin sunduğu bu ileri görüşlü yaklaşım, kanser teşhisinde doğruluk, hız ve etkinliği artırarak, hastaların yaşam kalitesine olumlu yönde katkıda bulunacaktır.

BÖLÜM 7. Kaynakça

- [1] C. Slebos Elsa R. Flores Christine H. Chung, *Deep Learning of Histopathology Images at the Single Cell Level*.
- [2] Deep learning in product design," December 14, 2022, Mickael Brossard, Jacomo Corbo, Marie Klaeyle, and Bill Wiseman
- [3] B. S. Veeling, J. Linmans, J. Winkens, T. Cohen, M. Welling. "Rotation Equivariant CNNs for Digital Pathology.
- [4] Luis Perez ve Jason Wang, The Effectiveness of Data Augmentation in Image Classification using Deep Learning
- [5] Khalid M. Kahloot ve Peter Ekler, Algorithmic Splitting: A Method for Dataset Preparation

ÖZGEÇMİŞ

Sinan Bavli, 06.08.2000 tarihinde Ağrı’da doğdu. İlk, orta ve lise eğitimini İstanbul’da tamamladı. 2018 yılında Sakarya Üniversitesi Bilgisayar Mühendisliği Bölümü’nü kazandı. 2023 yılında T-RUPT şirketinde veri analisti olarak staj yaptı. Ardından Smartera adlı şirkette 3 ay boyunca veri analisti olarak çalıştı.

BSM 498 BİTİRME ÇALIŞMASI DEĞERLENDİRME VE SÖZLÜ SINAV TUTANAĞI

KONU : DERİN ÖĞRENME İLE HİSTOPATOLOJİK RESİMLERDE METASTATİK
KANSER TESPİTİ
ÖĞRENCİLER :SİNAN BAVLİ

Değerlendirme Konusu	İstenenler	Not Aralığı	Not
Yazılı Çalışma			
Çalışma klavuza uygun olarak hazırlanmış mı?	x	0-5	
Teknik Yönden			
Problemin tanımı yapılmış mı?	x	0-5	
Geliştirilecek yazılımın/donanımın mimarisini içeren blok şeması (yazılımlar için veri akış şeması (dfd) da olabilir) çizilerek açıklanmış mı?			
Blok şemadaki birimler arasındaki bilgi akışına ait model/gösterim var mı?			
Yazılımın gereksinim listesi oluşturulmuş mu?			
Kullanılan/kullanılması düşünülen araçlar/teknolojiler anlatılmış mı?			
Donanımların programlanması/konfigürasyonu için yazılım gereksinimleri belirtilmiş mi?			
UML ile modelleme yapılmış mı?			
Veritabanları kullanılmış ise kavramsal model çıkarılmış mı? (Varlık ilişki modeli, noSQL kavramsal modelleri v.b.)			
Projeye yönelik iş-zaman çizelgesi çıkarılarak maliyet analizi yapılmış mı?			
Donanım bileşenlerinin maliyet analizi (prototip-adetli seri üretim vb.) çıkarılmış mı?			
Donanım için gerekli enerji analizi (minimum-uyku-aktif-maksimum) yapılmış mı?			
Grup çalışmalarında grup üyelerinin görev tanımları verilmiş mi (iş-zaman çizelgesinde belirtilebilir)?			
Sürüm denetim sistemi (Version Control System; Git, Subversion v.s.) kullanılmış mı?			
Sistemin genel testi için uygulanan metotlar ve iyileştirme süreçlerinin dökümü verilmiş mi?			
Yazılımın sızma testi yapılmış mı?			
Performans testi yapılmış mı?			
Tasarımın uygulamasında ortaya çıkan uyumsuzluklar ve aksaklıklar belirtilerek çözüm yöntemleri tartışılmış mı?			
Yapılan işlerin zorluk derecesi?	x	0-25	
Sözlü Sınav			
Yapılan sunum başarılı mı?	x	0-5	
Soruları yanıtlama yetkinliği?	x	0-20	
Devam Durumu			
Öğrenci dönem içerisindeki raporlarını düzenli olarak hazırladı mı?	x	0-5	
Diğer Maddeler			
Toplam			

DANIŞMAN (JÜRİ ADINA):

DANIŞMAN İMZASI:

