**CİLT KANSERİ TEŞHİSİ**

**ÖZCAN GÜLER**

**Bilgisayar Mühendisliği Anabilim Dalı**

**Görüntü İşlemede İleri Konular Ve Uygulamaları Dersi**

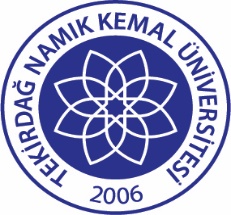
**Öğretim Görevlisi: Doç. Dr. Ahmet SAYGILI**

**2022**

**T.C.**

**TEKİRDAĞ NAMIK KEMAL ÜNİVERSİTESİ**

**FEN BİLİMLERİ ENSTİTÜSÜ**



**CİLT KANSERİ TEŞHİSİ**

**ÖZCAN GÜLER**

**1228132102**

**BİLGİSAYAR MÜHENDİSLİĞİ ANABİLİM DALI**

**GÖRÜNTÜ İŞLEMEDE İLERİ KONULAR VE UYGULAMALARI DERSİ**

**Öğretim Görevlisi: Doç. Dr. Ahmet SAYGILI**

**ARALIK-2022**

İÇİNDEKİLER

[**İÇİNDEKİLER** ii](#_Toc92185993)

[**ŞEKİLLER DİZİNİ** iii](#_Toc92185995)

[**KISALTMALAR DİZİNİ** iv](#_Toc92185997)

[**1.** **GİRİŞ** 1](#_Toc92185999)

[1.1 Literatür Özeti 2](#_Toc92186000)

[**2.** **ÖNERİLEN YÖNTEM** 6](#_Toc92186002)

[**3.** **SONUÇ VE ÖNERİLER** 15](#_Toc92186008)

[**KAYNAKLAR** 16](#_Toc92186009)

ŞEKİLLER DİZİNİ

[Şekil 1. 1. Densenet-121 eğitim ve test doğruluk oranı grafiği 3](#_Toc93496422)

[Şekil 1. 2. Densenet-121 karışıklık matrisi 3](#_Toc93496423)

[Şekil 1. 3. Densenet-121 deri kanser türlerine göre hatalı sınfılandırma oranı 4](#_Toc93496422)

[Şekil 1. 4. Densenet-121 karışıklık matrisi 4](#_Toc93496423)

[Şekil 1. 5. CNN doğruluk,test ve eğitim değerleri 5](#_Toc93496422)

[Şekil 1. 6. ResNet50 doğruluk,test ve eğitim değerleri 5](#_Toc93496423)

[Şekil 1. 7. XCeption doğruluk,test ve eğitim değerleri 5](#_Toc93496422)

[Şekil 1. 8. Precision, recall, f-measure and accuracy comparison tablosu. 5](#_Toc93496423)

[Şekil 1. 9. Mobilenet modelin karışıklık matrisi 6](#_Toc93496423)

[Şekil 1. 10. HAM1000 veri setindeki görüntü sayısına göre deri lezyonları sınıflandırması 6](#_Toc93496422)

[Şekil 1. 11. HAM1000 veri setindeki görüntü sayısına göre deri lezyonlarının vücuttaki yerine göre sınıflandırması 7](#_Toc93496423)

[Şekil 1. 12. HAM1000 veri setindeki hastaların yaş aralığı 7](#_Toc93496422)

[Şekil 1. 13. HAM1000 veri setindeki cinsiyet sayısı 8](#_Toc93496423)

[Şekil 1. 14. HAM1000 veri setindeki cilt kanseri örnekleri 9](#_Toc93496422)

[Şekil 1. 15. Önerilen CNN model 10](#_Toc93496423)

[Şekil 1. 16. Performans parametreleri 13](#_Toc93496422)

[Şekil 1. 17. Önerilen CNN modelin test ve validation doğruluk değerleri 13](#_Toc93496423)

[Şekil 1. 18. Önerilen CNN modelin test ve validation doğruluk değerleri grafiği 13](#_Toc93496423)

[Şekil 1. 19. Önerilen CNN modelin karışıklık matrisi 14](#_Toc93496422)

[Şekil 1. 20. HAM1000 veri setindeki görüntü sayısına göre deri lezyonları sınıflandırması Önerilen CNN modelin test ve validation doğruluk değerleri 14](#_Toc93496423)

KISALTMALAR DİZİNİ

bcc Bazal Hücreli Karsinom

mel Skuamöz Hücreli Karsinom ve Melanom

akiec Aktinik Keratoz

bcc Bazal Hücreli Karsinom

bkl Seboreik keratoz

df Dermatofibroma

nv Melanositik Nevüs

vasc Vascular Skin Lesions

# GİRİŞ

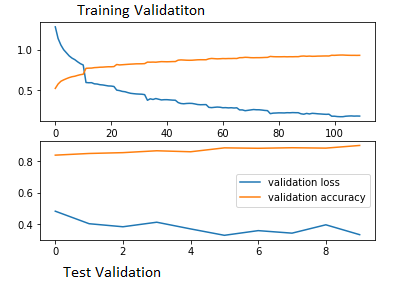
Deri kanseri, deri hücrelerinin anormal büyümesidir. Deri kanserinin nedenleri arasında Ultraviyole ışınlarına ve bronzlaşma yataklarına yoğun şekilde maruz kalmak, açık ten rengi, yüksek alkol tüketimi, ailede cilt kanseri öyküsü gösterilmektedir. Deri kanseri 3 çeşittir. Bazal Hücreli Karsinom (**bcc**), Skuamöz Hücreli Karsinom ve Melanom (**mel**). Deri kanserleri arasında en az görülen fakat en ölümcül olanı Melanom kanseridir. Melanom, melanin üreten hücrelerden kaynaklanan sıklıkla ciltte, gözlerde, meninkslerde bulunan tümördür. Melanom yılda ortalama 55.500 kişinin ölümüne neden olmaktadır bu da tüm kanser ölümlerinin %0.7’sidir. İstatistklere göre Melanom’a yakalanma riski beyaz tenlilerde %2.6, koyu tenlilerde %0.1 ve Hispaniklerde %0.6 civarındadır. Melanom, cilt kanseri ölümlerinin %90’ına neden olan en tehlikeli cilt tümörüdür ve ağırlıklı olarak Amerika Birleşik Devletleri ve Avustralya ülkelerinde görülmektedir. Birleşik Krallık’ta ise her yıl yaklaşık 46.000 cilt kanseri vakası bildirilmektedir. Deri kanseri, dünya çapında önemli ölçüde artan vaka sayısı ile ölümcül kanser türleri arasında yer almaktadır. Melanom ilk evrelerinde teşhis edilirse tedavi edilebilir ancak geç teşhis edilip tedaviye başlanılırsa ölüm riski yüksek bir hastalıktır. Hastalığın 1. evresinde hastaların 10 yıllık yaşama ihtimali %94 ile %98 arasındayken, 4. evresinde hastaların yaşama ihtimali %10 ile %15 arasındadır. Dermatologlar dermoskopi görüntülerinden cilt kanseri hastalıklarını teşhis edebilir. Tecrübesine dayanarak cilt kanserinin türünü teşhis edebilir ancak cilt kanserini tespit etmek %100 garanti değildir ve potansiyel zararlara yol açabilir. Burada olası zarar, cilt kanserine neden olmayan lezyonlar için deri biyopsisi yapılması veya cilt kanserine neden olan lezyonlar için biyopsi yapılmaması nedeniyle deri kanseri teşhisi konulamaması. Aktinik Keratoz (**akiec**), halk arasında yaşlılık lekesi olarak bilinen ve uzun süreli kontrolsüz güneşe maruz kalmaya bağlı olarak, en çok güneş gören bölgelerde görülen, deride anormal hücre gelişimini yansıtan deri değişiklikleridir. Aktinik keratozların düşük oranda deri kanserine dönüşme riskileri de vardır. Bazal Hücreli Karsinom (**bcc**) tüm cilt kanserlerinin %80’ ini oluşturur. Diğer deri kanseri çeşitlerine göre çok daha iyi seyirlidir. Açık renk ciltte ve UV ile temaslı bölgelerde daha sık görülür. Çoğu türünde ailesel yatkınlık gözlenmez. Çok nadir vücudun diğer bölgelerine yayılır. Seboreik keratoz (**bkl**), açık kahveden siyaha kadar değişebilen renkte deride oluşan kabarıklıklardır. Özellikle deri kanserleri ve siğillerle karıştırılabilir. İyi huylu tümör olarak adlandırılır. Yaş ilerledikçe daha sık olarak görülmeye başlarlar. Dermatofibroma (**df**), sıklıkla kadınlarda gözlenen şikayet vermeyen iyi huylu deri tümörleridir. Melanositik Nevüs (**nv**) sıklıkla ‘ben’ olarak da adlandırılmaktadır. Vascular Skin Lesions (**vasc**) Görüntü işleme yöntemleri, makine öğrenmesi ve derin öğrenme yöntemleri kullanılarak bilgisayar destekli tanı sistemiyle deri kanserinin olup olmadığını tahmin ederek dermatologlara yardımcı olmaktadır.

## Literatür Özeti

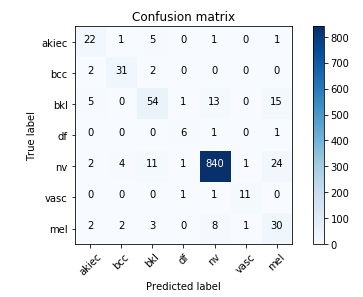
Dermoskopi, derinin alt yüzey yapılarının görsel olarak incelenmesini mümkün kılmak için gelen ışığın ve yağa batırmanın kullanımına dayanan invazif olmayan bir inceleme tekniğidir. Dermoskopi kullanılarak cilt kanseri tespiti, yardımsız gözleme dayalı tespitten daha yüksek olsa da, teşhis doğruluğu dermatoloğun eğitimine bağlıdır. Melanositik nevüslerden melanom tanısı özellikle erken evrede kolay değildir. Bu nedenle otomatik tanı aracı hekimler için olmazsa olmazdır. Uzman dermatologlar tanı için dermoskopi kullandıklarında bile melanom tanısındaki doğruluk oranı yaklaşık %75-84 olarak tahmin edilmektedir. Bilgisayar destekli tanı, tanı doğruluğunun yanı sıra hızın da artmasına yardımcı olmaktadır. Bilgisayar insandan daha zeki değildir, ancak renk değişimi, asimetri, doku özellikleri gibi insan gözünün kolayca algılayamayacağı bazı bilgileri çıkarabilir. Melanom cilt kanseri tanısını iyileştirmek için yedi noktalı kontrol listesi, ABCDE kuralı ve Menzies yöntemi gibi birçok önerilen sistem ve algoritma vardır.

XINRUI ZHUANG

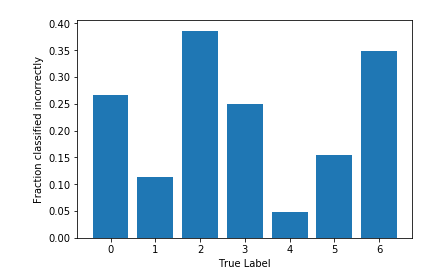
Pytorch framework’u kullanılmış. Resnet, VGG, Densenet ve Inception gibi birkaç ağdan oluşur. Hiperparametre ayarlaması olmadan, 10 epoch doğrulama setinde %90'dan fazla doğruluk elde etmek için Densenet-121 kullanılmıştır.



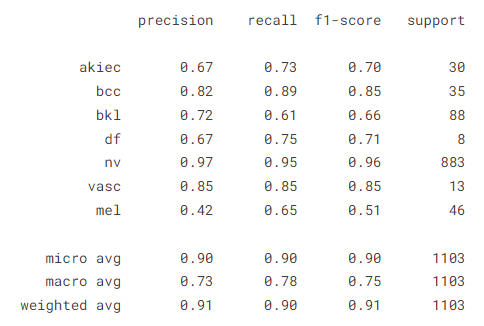
Şekil 1. 1. Densenet-121 eğitim ve test doğruluk oranı grafiği



Şekil 1. 2. Densenet-121 karışıklık matrisi



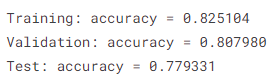
Şekil 1. 3. Densenet-121 deri kanser türlerine göre hatalı sınfılandırma oranı



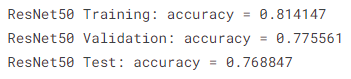
Şekil 1. 4. Densenet-121 karışıklık matrisi

JULIANA NEGRINI DE ARAUJO

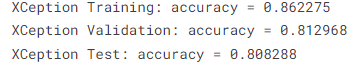
CNN, ResNet-50 ve XCeption. HAM10000 veri seti kullanılmıştır. Eğitim ve Test setleri oluşturun (80 - 20 oran) Girişi normalleştirin. En iyi uygulamaları takiben, eğitim seti verileri referans olarak kullanılarak normalleştirme yapılmalıdır. Test verileri, bilinmemesi gerektiği için normalleştirilemez. Hedef etiketleri dönüştürmek için bir Hot Encoding Eğitim setini Eğitim ve Doğrulama setlerine ayırın (90 - 10 oran). CNN, görüntüleri 3 boyutta yeniden şekillendirilmesini gerektirir (yükseklik = 75 px, genişlik = 100 px, kanal = 3).



Şekil 1. 5. CNN doğruluk,test ve eğitim değerleri



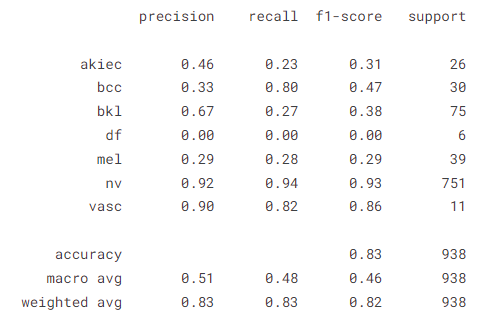
Şekil 1. 6. ResNet50 doğruluk,test ve eğitim değerleri



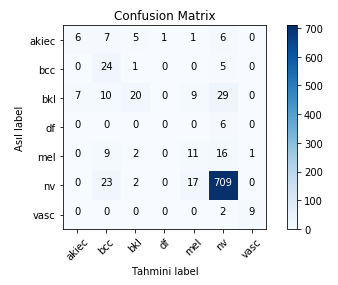
Şekil 1. 7. XCeption doğruluk,test ve eğitim değerleri

HÜZEYFE BURAK ARSLAN

Mobilenet modeli kullanılarak HAM10000 veri setinden şu sonuçlar elde edilmiştir. Batch size 50 ve epoch 15 seçilmiştir.

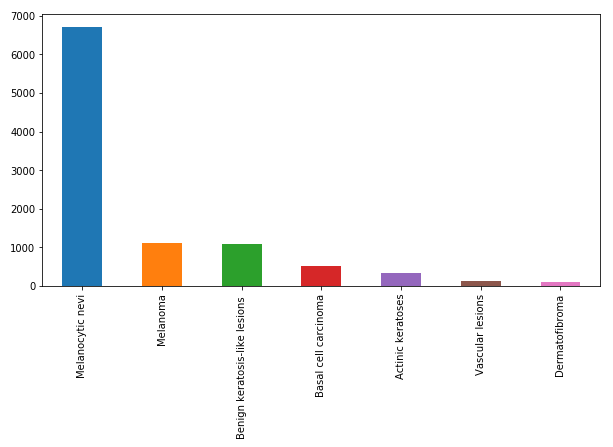


Şekil 1. 8. Precision, recall, f-measure and accuracy comparison tablosu.

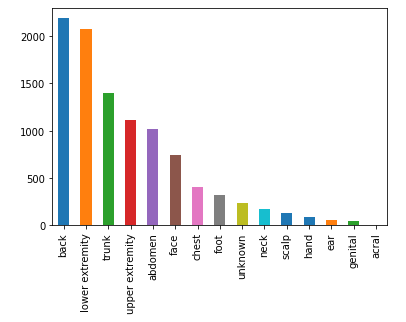


Şekil 1. 9. Mobilenet modelin karışıklık matrisi

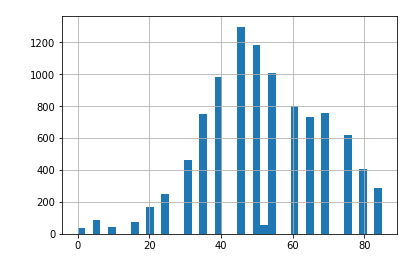
# ÖNERİLEN YÖNTEM



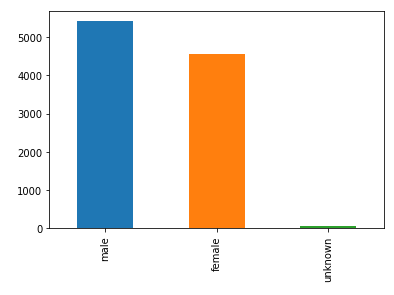
Şekil 1. 10. HAM1000 veri setindeki görüntü sayısına göre deri lezyonları sınıflandırması



Şekil 1. 11. HAM1000 veri setindeki görüntü sayısına göre deri lezyonlarının vücuttaki yerine göre sınıflandırması

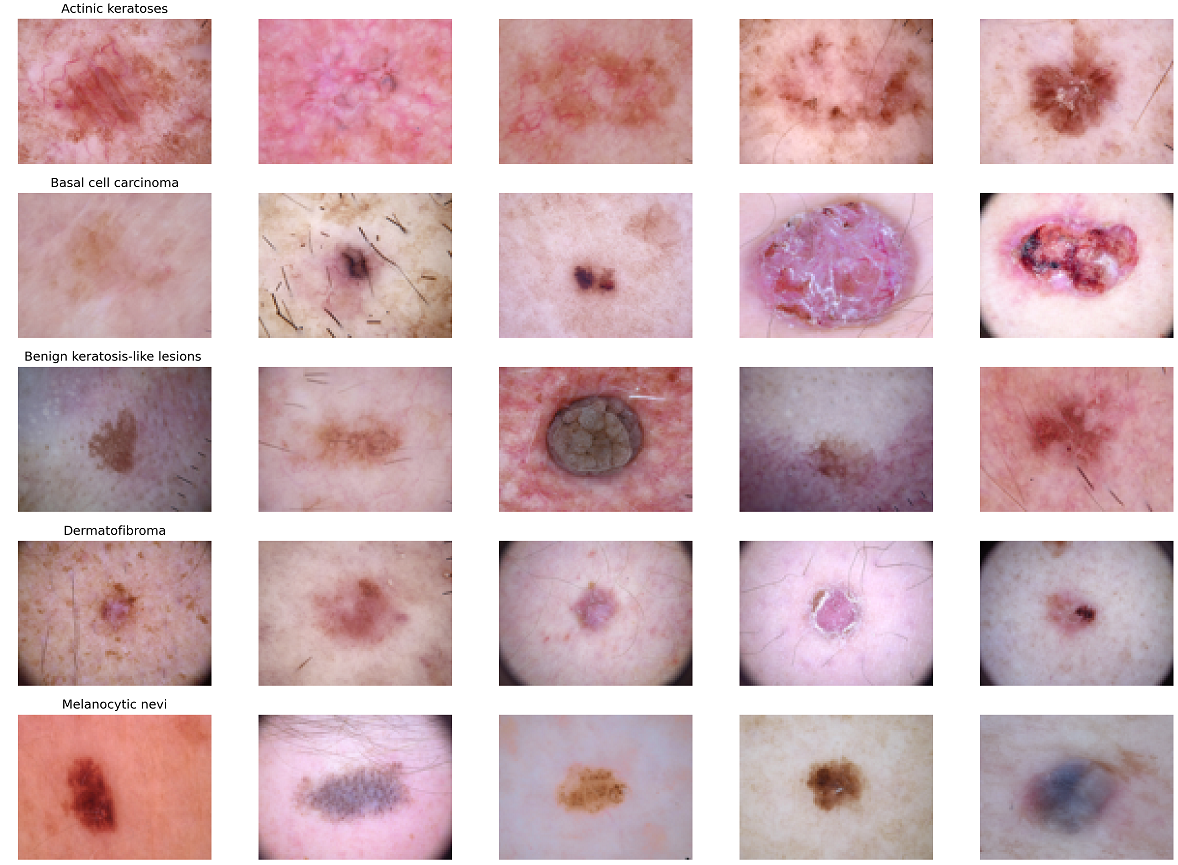


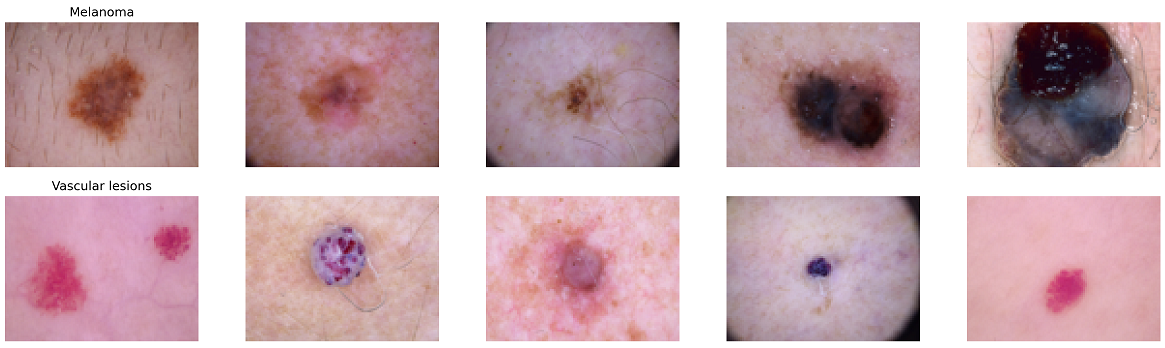
Şekil 1. 12. HAM1000 veri setindeki hastaların yaş aralığı



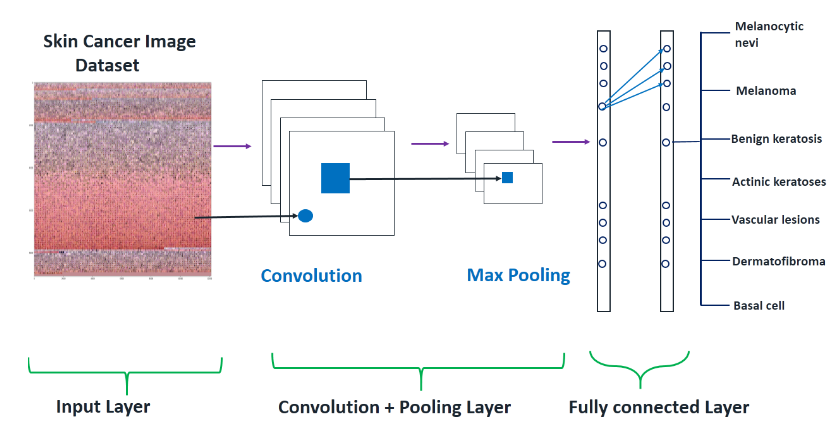
Şekil 1. 13. HAM1000 veri setindeki cinsiyet sayısı

Kullanılan yöntemde HAM10000 veri kümesi kullanılmış olup, ISIC arşivi aracılığıyla halka açık olan 10015 dermatoskopik görüntüden oluşur. Veri kümesi 7 farklı cilt kanseri görüntülerinden oluşmaktadır. Cilt kanseri türleri arasında Melanositik Nevüs, Melanom, Bazal Hücreli Karsinom, Aktinik Keratoz, Seboreik Keratoz, Dermatofibroma, Vascular Skin Lesions oluşmaktadır.





Şekil 1. 14. HAM1000 veri setindeki cilt kanseri örnekleri



Şekil 1. 15. Önerilen CNN model

Girdiden başlayarak her seferinde bir katman eklenecek olan Keras Sequential API'sini kullanılmıştır. Birincisi evrişimli (Conv2D) katmandır. Bir dizi öğrenilebilir filtre gibidir. İlk iki conv2D katmanı için 32 filtre ve son ikisi için 64 filtre ayarlanmıştır. Her filtre, çekirdek filtresini kullanarak görüntünün bir bölümünü (çekirdek boyutuyla tanımlanır) dönüştürür. Çekirdek filtre matrisi görüntünün tamamına uygulanır. Filtreler görüntünün bir dönüşümü olarak görülebilir. CNN, her yerde yararlı olan özellikleri bu dönüştürülmüş görüntülerden (özellik haritaları) ayırabilir. CNN'deki ikinci önemli katman havuzlama (MaxPool2D) katmanıdır. Bu katman basitçe bir altörnekleme filtresi görevi görür. 2 komşu piksele bakar ve maksimum değeri seçer. Bunlar, hesaplama maliyetini azaltmak için kullanılır ve bir dereceye kadar aşırı uydurmayı da azaltır. Havuzlama boyutunu (yani her seferinde havuzlanan alan boyutunu) seçmeliyiz, havuzlama boyutu ne kadar yüksekse, altörnekleme o kadar önemlidir. Evrişimli ve havuzlama katmanlarını birleştiren CNN, yerel özellikleri birleştirebilir ve görüntünün daha genel özelliklerini öğrenebilir. Dropout, her eğitim örneği için katmandaki düğümlerin bir kısmının rastgele yok sayıldığı (ağırlıklarını sıfıra ayarlayarak) bir düzenlileştirme yöntemidir. Bu, ağın bir oranını rastgele düşürür ve ağı özellikleri dağıtılmış bir şekilde öğrenmeye zorlar. Bu teknik aynı zamanda genellemeyi geliştirir ve aşırı uydurmayı azaltır. 'Relu' doğrultucudur (aktivasyon fonksiyonu max(0,x). Doğrultucu aktivasyon fonksiyonu, ağa doğrusalsızlık eklemek için kullanılır. The Flatten layer, nihai özellik haritalarını tek bir 1B vektöre dönüştürmek için kullanılır. Bu düzleştirme adımı, bazı konvolüsyon/maxpool katmanlarından sonra tamamen bağlantılı katmanlardan yararlanabilmeniz için gereklidir. Önceki evrişimli katmanların bulunan tüm yerel özelliklerini birleştirir. Sonunda, sadece yapay bir sinir ağları (YSA) sınıflandırıcısı olan iki katmandaki özellikleri kullanılmıştır. Son katmanda(Dense(10,activation="softmax")), net her sınıfın olasılık dağılımını verir. Katmanlarımız modele eklendikten sonra bir skor fonksiyonu, bir kayıp fonksiyonu ve bir optimizasyon algoritması kurmamız gerekiyor. Modelimizin bilinen etiketlere sahip görüntülerde ne kadar kötü performans gösterdiğini ölçmek için kayıp işlevini tanımlarız. Gözlenen etiketler ile tahmin edilenler arasındaki hata oranıdır. Kategorik sınıflandırmalar (>2 sınıf) için "categorical\_crossentropy" adı verilen özel bir form kullanıyoruz. En önemli işlev optimize edicidir. Bu işlev, kaybı en aza indirmek için parametreleri (çekirdek değerlerini, ağırlıkları ve nöronların yanlılığını filtreler ...) yinelemeli olarak iyileştirir. Adam iyileştirici seçilmiştir çünkü stokastik gradyan inişinin diğer iki uzantısının avantajlarını birleştiriyor. özellikle: Seyrek gradyanlı problemlerde (örn. doğal dil ve bilgisayarla görme problemleri) performansı iyileştiren parametre başına öğrenme oranını koruyan Uyarlanabilir Degrade Algoritması (AdaGrad). Ağırlık için gradyanların son büyüklüklerinin ortalamasına (örneğin, ne kadar hızlı değiştiği) dayalı olarak uyarlanan parametre başına öğrenme oranlarını da koruyan Kök Ortalama Kare Yayılımı (RMSProp). Bu, algoritmanın çevrimiçi ve durağan olmayan problemlerde (örn. gürültülü) başarılı olduğu anlamına gelir. Adam, hem AdaGrad hem de RMSProp'un faydalarının farkına varır. Adam, iyi sonuçlara hızlı bir şekilde ulaştığı için derin öğrenme alanında popüler bir algoritmadır. Metrik fonksiyon "doğruluk", modelimizin performansını değerlendirmek için kullanılır. Bu metrik işlev, model eğitilirken metrik değerlendirmeden elde edilen sonuçların kullanılmaması dışında (yalnızca değerlendirme için) kayıp işlevine benzer. Optimize ediciyi daha hızlı yakınsamak ve kayıp fonksiyonunun global minimumuna en yakın hale getirmek için, öğrenme oranının (LR) bir yöntemi kullanılmıştır. LR, optimize edicinin 'kayıp manzarasında' yürüdüğü adımdır. LR ne kadar yüksek olursa, adımlar o kadar büyük olur ve yakınsama o kadar hızlı olur. Bununla birlikte, örnekleme, yüksek bir LR ile çok zayıftır ve optimize edici muhtemelen yerel bir minimuma düşebilir. Kayıp fonksiyonunun küresel minimumuna verimli bir şekilde ulaşmak için eğitim sırasında azalan bir öğrenme oranına sahip olmak daha iyidir. Yüksek bir LR ile hızlı hesaplama süresinin avantajını korumak için, LR'yi gerekli olup olmadığına bağlı olarak (doğruluk iyileştirilmediğinde) her X adımda (epochs) dinamik olarak düşürülmüştür. Keras.callbacks'in ReduceLROnPlateau işleviyle, doğruluk 3 epoch’dan sonra iyileşmezse LR yarı yarıya azaltılacaktır.

Veri Büyütme

İsteğe bağlı adımdır. Overfitting(uydurma) sorununu önlemek için HAM 10000 veri setimizi yapay olarak genişletmemiz gerekiyor. Mevcut veri kümenizi daha da büyütebiliriz. Fikir, varyasyonları yeniden oluşturmak için eğitim verilerini küçük dönüşümlerle değiştirmektir. Etiketi aynı tutarken dizi temsilini değiştiren şekillerde eğitim verilerini değiştiren yaklaşımlar, veri artırma teknikleri olarak bilinir. İnsanların kullandığı bazı popüler büyütmeler, gri tonlamalar, yatay döndürmeler, dikey döndürmeler, rastgele kırpmalar, renk titremeleri, ötelemeler, döndürmeler ve çok daha fazlasıdır. Bu dönüşümlerden sadece birkaçını eğitim verilerimize uygulayarak, eğitim örneklerinin sayısını kolayca ikiye veya üçe katlayabilir ve çok sağlam bir model oluşturabiliriz. Veri artırma için şunlar seçilmiştir: Bazı eğitim görüntülerini rastgele 10 derece döndürme, %10 yakınlaştırma, görüntüleri genişliğin ve yüksekliğin %10'u kadar rastgele kaydırma.

Deneysel kurulum

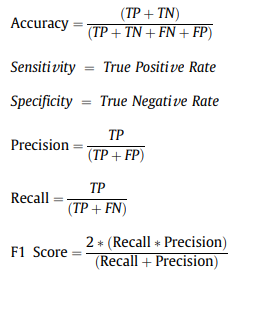
AMD Ryzen 5 1600, GTX1060 6GB, 8 GB RAM, Windows 10 (64 bit) makine ile test edilmiştir. Spyder IDE (Anaconda3) python 3.9 programı kullanılmıştır.

Modeli çalıştırma

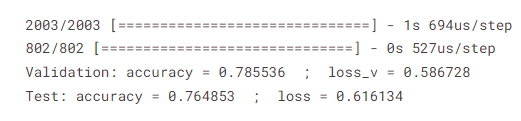
Batch size 10 ve 50 epoch seçilmiştir, batch size ne kadar küçük olursa modelimiz o kadar verimli bir şekilde eğitilecektir.

Model Değerlendirmesi

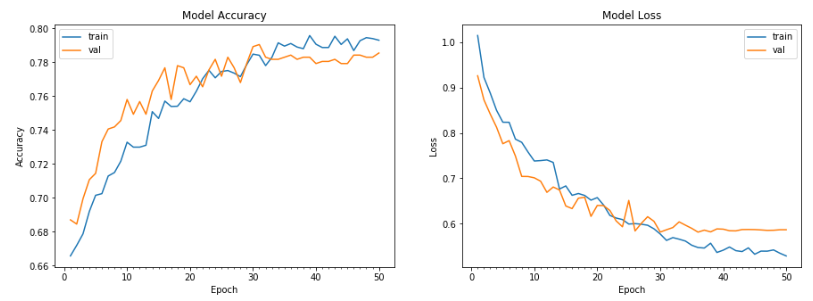
Bu adımda, modelimizin test doğruluğunu ve doğrulama doğruluğunu kontrol edeceğiz, karışıklık matrisini çizeceğiz ve ayrıca her türün yanlış sınıflandırılmış görüntü sayısını kontrol edeceğiz. Model %77 doğruluğa sahip insan gözüyle algılamaya kıyasla yine de verimli.



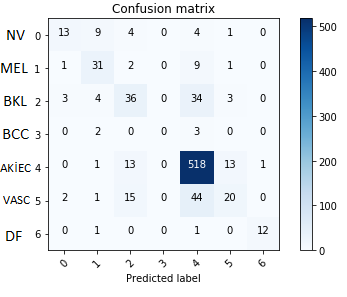
Şekil 1. 16. Performans parametreleri



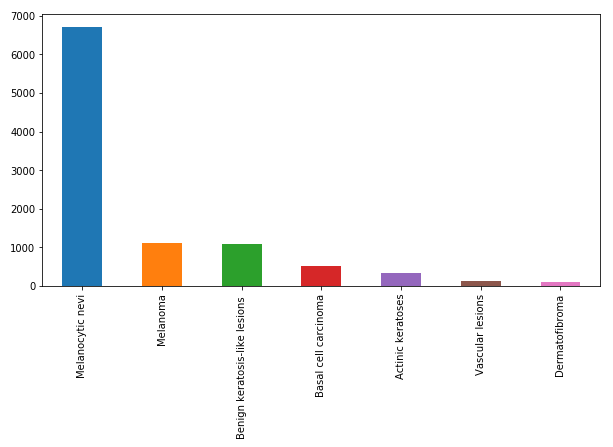
Şekil 1. 17. Önerilen CNN modelin test ve validation doğruluk değerleri



Şekil 1. 18. Önerilen CNN modelin test ve validation doğruluk değerleri grafiği



Şekil 1. 19. Önerilen CNN modelin karışıklık matrisi



Şekil 1. 20. HAM1000 veri setindeki görüntü sayısına göre deri lezyonları sınıflandırması

# SONUÇ VE ÖNERİLER

Uzman dermatologlar tanı için dermoskopi kullandıklarında bile melanom tanısındaki doğruluk oranı yaklaşık %75-84 olarak tahmin edilmektedir Diğer modellerle karşılaştırıldığında modelimizin doğruluk oranı düşük olsa da %77 doğruluk oranıyla dermatologlara hastalığın teşhis edilmesinde yardımcı olabilir. Modelimize çeşitli ayarlamalar yaparak doğruluk oranı %80’in üzerine çıkması beklenmektedir. İlerleyen zamanda mobil uygulama geliştirilerek sistemin pratik olması amaçlanmaktadır.

KAYNAKLAR

Ravi Manne, Snigdha Kantheti, Sneha Kantheti (2020). Classification of Skin cancer using deep learning, ConvolutionalNeural Networks - Opportunities and vulnerabilities- A systematic Review. Erişim adresi: <https://doi.org/10.46501/IJMTST061118>

XINRUI ZHUANG Erişim adresi: <https://www.kaggle.com/code/xinruizhuang/skin-lesion-classification-acc-90-pytorch>

JULIANA NEGRINI DE ARAUJO. Erişim adresi: <https://www.kaggle.com/code/jnegrini/ham10000-analysis-and-model-comparison>

HÜZEYFE BURAK ARSLAN. Erişim adresi: <https://www.kaggle.com/code/burakarslan38/cilt-kanserini-alg-layan-yapay-zeka-algoritmas/notebook>

MANU SIDDHARTHA. Erişim adresi: <https://www.kaggle.com/code/sid321axn/step-wise-approach-cnn-model-77-0344-accuracy/notebook>

Philipp Tschandl, Cliff Rosendahl, Harald Kittler (2018). The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *SCIENTIFIC DATA | 5:180161 | DOI: 10.1038/sdata.2018.161*