



Derin Sinir Ağları Kullanarak Histopatolojik Görüntülerde Meme Kanseri Tespiti

Emre Dandıl^{1*}, Zafer Serin²

¹ Bilecik Şeyh Edebali Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Bilecik, Türkiye (ORCID: 0000-0001-6559-1399)

² Bilecik Şeyh Edebali Üniversitesi, Lisansüstü Eğitim Enstitüsü, Bilgisayar Mühendisliği Anabilim Dalı, Bilecik, Türkiye (ORCID: 0000-0002-5213-8517)

(Bu yayın 26-27 Haziran 2020 tarihinde HORA-2020 kongresinde sözlü olarak sunulmuştur.)

(DOI: 10.31590/ejosat.780705)

ATIF/REFERENCE: Dandıl, E. & Karaca, S. (2020). MR Spektroskopi Sinyalleri Kullanılarak LSTM Derin Sinir Ağları ile Beyinde Sahte Tümörlerin Tespiti. *Avrupa Bilim ve Teknoloji Dergisi*, (Special Issue), 451-463.

Öz

Bu çalışmada çağımızın en tehlikeli hastalıklarından birisi olan ve geç teşhis durumunda tedavisi oldukça zor bir hal alan meme kanserinin histopatolojik görüntülerde tespiti için derin sinir ağları kullanılarak bir model önerilmiştir. Çalışma ile bilgisayar destekli teşhis sistemlerinin yüksek doğruluk değerlerine ulaşarak meme kanserinin tespitinde yardımcı bir teşhis metodu kullanılması sağlanmıştır. Kullanılan veriseti 40X, 100X, 200X ve 400X yakınlaştırma oranlarına sahip ve toplamda 7909 adet histopatolojik görüntü içermektedir. Önerilen derin sinir ağı modelinde DenseNet201, Inception V3, ResNet50 ve Xception olmak üzere dört farklı ön-egitimli (pre-trained) ağı kullanılarak daha başarılı sonuçlar elde edilmiştir. Modellerin başarımlarını daha da artırmak amacıyla bırakma (dropout) ve veri artırma (data augmentation) yöntemleri kullanılmıştır. Kullanılan bu ağlar kendi arasında karşılaştırılmış ve Xception ağı ile elde edilen sonuçların diğer ağlara oranla daha başarılı olduğu görülmüştür. Xception ağı ile 200X yakınlaştırma oranında gerçekleştirilen deneysel çalışmalarda, hem diğer ağlara göre hem de diğer yakınlaştırma oranlarına göre en başarılı sonuçlara ulaşılmış ve 98.01'lik bir doğruluk (Accuracy) skoru, 97.89'luk bir hassasiyet (Precision) değeri ve 97.47'lik bir hatırlama (Recall) değeri elde edilmiştir. Xception ağının 200X yakınlaştırma oranındaki ROC eğrisi altındaki alan değeri ise 0.975 olarak hesaplanmıştır.

Anahtar Kelimeler: Meme kanseri, Histopatolojik görüntü, Sınıflandırma, Derin sinir ağları, ön-egitimli ağlar, Xception

Breast Cancer Detection on Histopathological Images Using Deep Neural Networks

Abstract

In this study, a model is proposed using deep neural network on histopathological images for the detection of breast cancer, which is one of the most dangerous diseases of our age and is very difficult to treat in case of late diagnosis. Within the study, computer-aided diagnosis systems are used as an assistant diagnostic method in the detection of breast cancer by reaching high accuracy values. The dataset has magnification ratios of 40X, 100X, 200X and 400X and contains a total of 7909 histopathological images. In the proposed deep neural network model, more successful results are obtained using four different pre-trained networks such as DenseNet201, Inception V3, ResNet50 and Xception. In order to further increase the performance of the models, dropout and data augmentation techniques are used. These networks are compared among themselves and the results obtained with the Xception network are performed to be more successful than other networks. In experimental studies carried out with the Xception network at the ratio of 200X magnification the most successful results are achieved with an accuracy score of 98.01, a precision value of 97.89 and recall value of 97.47 when comparing to other networks and other magnification ratios. The area value of under the ROC curve for the Xception network at 200X zoom ratio is 0.975.

Keywords: Breast cancer, Histopathological image, Classification, Deep neural network, pre-trained networks, Xception

¹ Sorumlu Yazar: Bilecik Şeyh Edebali Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü, Bilecik, Türkiye, ORCID: 0000-0001-6559-1399, emre.dandil@bilecik.edu.tr

1. Giriş

Kanser, nüfusu nispeten fazla bir ülke olan ABD’de kalp hastalıklarından sonra ikinci sırada ölümlerin nedeni olarak gösterilmiş ve kısa süre içerisinde de birinci sırada yer alması beklenen bir hastalıktır (Siegel, Miller, & Jemal, 2015). Dünya Sağlık Örgütü’nün (WHO) raporuna göre; dünya çapında kadınlar arasında en yaygın olarak görülen kanser %25.2’lik oran ile meme kanseridir (Stewart & Wild, 2014). Meme kanserinin hastalıklılık (Morbidity) oranı ise %14.7 olup son yıllarda kanser ölümcüllüğü (Mortality) akciğer kanserinden sonra ikinci sırada yer almaktadır (Han et al., 2017). Yine bu araştırmaya göre, yaklaşık yarım milyon meme kanseri hastası hayatını kaybetmiş ve her yıl 1.7 milyon yeni vaka ortaya çıkmaktadır. Bu vaka sayılarının giderek artması beklenmektedir. 2019 yılında yapılan bir çalışmaya göre ise toplamda 271.270 yeni meme kanseri vakası ve aynı çalışmaya göre toplamda 42.260 meme kanseri nedeniyle ölüm beklendiği öngörülmüştür (Siegel, Miller, & Jemal, 2019).

Meme kanserinin teşhisinde mamografi, ultrason ve manyetik rezonans (MR) görüntüleme gibi klasik yöntemler kullanılmaktadır. Mamografi özel bir röntgen çeşidi olarak betimlenir. Ultrason ise ses dalgaları kullanılarak tümör tespiti yapılan bir yöntemdir. MR, manyetik alanın ve radyo dalgalarının kullanılarak teşhisin yapıldığı bir yöntemdir (Dogan, Gonzalez-Angulo, Gilcrease, Dryden, & Yang, 2010). Her ne kadar meme kanserinin teşhisinde bu yardımcı görüntüleme yöntemleri kullanılsa da, kesin tanı için histopatolojik görüntülerde testler yapılması gerekmektedir. Patologlar biyopsi ile kesin hastalık tanılarını koymaktadır ancak bir hastalığın mikroskop yardımıyla detaylı bir şekilde incelenmesi çok uzun süreler alabilmektedir. Ayrıca kimi durumlarda insan kaynaklı (yorgun olma veya deneyimsizlik gibi) hatalar oluşabilmekte ve her zaman doğru sonuca ulaşmak mümkün olmamaktadır. Son yıllarda bu nedenlerden ötürü patoloji alanında makine öğrenmesi, derin öğrenme ve görüntü işleme gibi teknolojilerden faydalanılmaktadır (Wang, Khosla, Gargeya, Irshad, & Beck, 2016). Patologlar meme kanseri teşhisi için dokusal özellikleri, meme parankiminin normal yapılarındaki farklılıkları tespit ederek Hematoksilin ve Eosin (H&E) ile boyanmış dokuları mikroskop ile inceler ve buna göre bir sonuca varmaya çalışır. Uygun tedavi yönteminin seçilmesi için dokunun iyi huylu veya kötü huylu olduğunun zamanında tespit edilmesi gereklidir. Buna karşın H&E yöntemi patologlar için çok zahmetli ve zaman alıcıdır. Ayrıca uzmanlar kimi zaman kanser tespitinde ortak bir sonuca varamamaktadır. Yapay zeka ve makine öğrenmesi gibi teknolojiler kanser tespitinde daha hızlı ve tek bir insana göre daha başarılı sonuçlar ortaya koymasından ötürü bu alanlarda daha sık kullanılmaya başlamıştır. Ayrıca bu yöntem sınıflandırma doğruluğunu arttırmakta ve uzmanlar arasındaki görüş farklılıklarını azaltmaktadır (Kumar et al., 2020).

Patoloji ve onkoloji alanlarında kullanılan makine öğrenmesi modelleri histopatolojik görüntüdeki tümörün iyi huylu veya kötü huylu olduğunun sınıflandırılmasını sağlamaktadır. Tümör sadece bir kitle olarak tanımlanabilir ve bu tanıma göre her tümör insanlar için bir tehlike arz etmez ve bazı tümörler iyi huylu olabilir. Kanser dediğimiz hastalık hücrelerin kontrolsüz bir şekilde bölündüğü ve yayıldığı durumları ifade etmektedir (Öztürk & Akdemir, 2019). Derin öğrenme algoritmalarında kullanılan modeller eğitim aşamasında kendisine gösterilen onlarca örneğin özellik çıkartmasını (hücre büyüklüğü, rengi ve şekli gibi) yapar ve daha önce görmediği bir takım görüntüler üzerinde test edilerek modelin başarılı olup olmadığı gösterilir. Kullanılan modeller çok fazla piksel ve matris işlemleri yapması nedeniyle güçlü donanımlara sahip bir bilgisayara ihtiyaç duyar. Yapılan eğitimler kimi durumlarda günler sürebilmektedir. Modeller bu kadar çok gereksinime ve zamana ihtiyaç duymalarına karşın kimi durumlarda yanlış tahminlerde bulunabilmekte ve hatalar yapabilmektedir. Hatalı sonuçlar elde edilmesinin en büyük sebepleri: tümörlü hücrenin arka plandan ve diğer dokulardan ayırt edilmesinde zorluk çekilmesi, görüntüye uygulanan bazı işlemlerin özellik kaybına neden olması vb.dir. Modelleri bu eksikliklerinden kurtarmak ve hatta daha iyi sonuçlar elde etmek için önceden eğitilmiş modeller (pre-trained) ortaya çıkmıştır (Rakhlin, Shvets, Igloukov, & Kalinin, 2018).

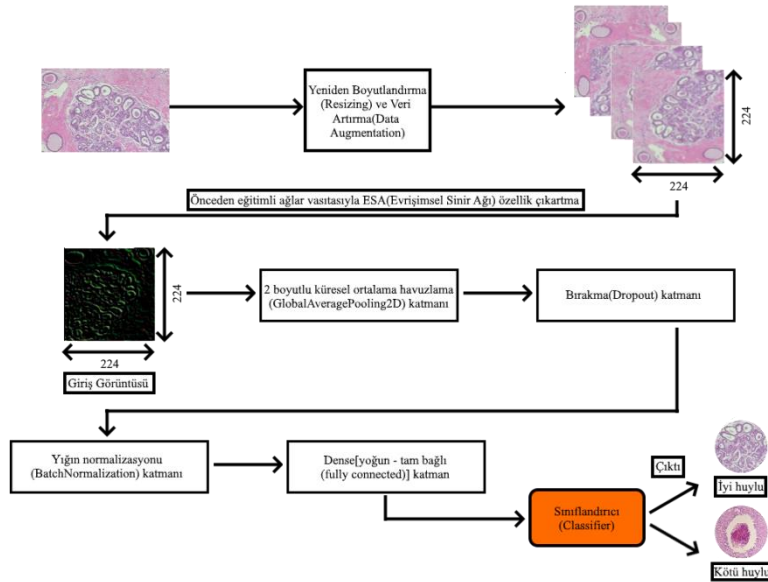
Literatürde meme kanserinin tespitinde farklı çalışmalar bulunmaktadır. 2013 yılında yapılan bir çalışmada meme kanseri sınıflandırması için PNN (Probabilistic Neural Networks), MLP (Multi Layer Perceptron) ve RBF (Radial Basis Function) modelleri ile bir çalışma yapılmıştır. Yürütülen deneysel çalışmalarda, PNN ile daha başarılı sonuçların elde edildiği görülmüştür (Azar & El-Said, 2013). Bir başka çalışmada, meme kanserinin tespiti için global veriseti üzerinde geri yayılım ağı ve Liebenberg Marquardt öğrenme fonksiyonunun kullanıldığı, ayrıca ağırlıkların DBN-NN (deep belief network path) ile başlatıldığı bir model ile yüksek bir doğruluk oranı elde edilmiştir (Abdel-Zaher & Eldeib, 2016). 2018 yılında yapılan bir çalışma da BreakHis veri seti kullanılarak meme kanseri sınıflandırması yapılmıştır. Bu çalışma da CNN (Convolutional Neural Network) ve LSTM (Long Short Term Memory)’nin bir kombinasyonu kullanılmıştır. (Nahid, Mehrabi, & Kong, 2018). Bir diğer çalışmada, 72’si tümörlü ve 184’ü tümörsüz görüntüden oluşan toplam 184 meme ultrason görüntüsü içeren verisetinde göğüslerdeki kitlelerin tespiti yapılmıştır. Çalışmada çoklu fraktal boyut özellikleri kullanılarak iyi düzeyde bir hassasiyet değerine ulaşılmıştır (Mohammed et al., 2018). 2019 yılında yapılan bir çalışmada BreakHis veriseti üzerinde tüm örnekleri etiketlemeye gerek kalmadan histopatolojik görüntülerde sınıflandırma yapabilen bir derin öğrenme modeli önerilmiştir (Sudharshan et al., 2019). Yine 2019 yılında yapılan bir diğer çalışma da tam slayt histopatolojik görüntülerde (WSI) kanserli alanları sınıflandıran bir ağ (HIC-net) önerilmiştir. Bu ağ modeli WSI’yi belirli bir düzleme bölerek penceresel olarak sınıflandırma yapmaktadır (Öztürk & Akdemir, 2019). 2020 yılında yapılan bir çalışma da meme tümörünün (CMT) köpekler için de oldukça ölümcül olduğu ve insan meme kanseri çalışmaları için oldukça güzel örnekler olduğu görülmüştür. Her iki tümör içinde hematoksilin ve eosin (H&E) ile boyanmış görüntüler kullanılan bu çalışma histopatolojik görüntüler üzerinde bir derin sinir ağı yöntemi önermiştir. Bu çalışma da VGGNet-16 ön-eğitilmiş ağı kullanılarak sınıflandırma yapılmıştır (Kumar et al., 2020).

Literatürde önerilen çalışmalar incelendiğinde, derin öğrenme ve ön-eğitilmiş ağların tıbbi alanlarda hastalık tespiti ve özellikle kanser tespitinde klasik yöntemlere oranla daha başarılı sonuçlar ortaya koydukları görülmektedir. Bu çalışmada ise, histopatolojik görüntüler kullanılarak meme kanserinin tespiti ve sınıflandırılması için ön-eğitilmiş ağlarla desteklenen derin öğrenme modeli önerilmiştir. Bu çalışmada diğer çalışmalara göre veri artırma ve bırakma gibi eklemeler yapılmış, başarıyı kanıtlanmış ön-eğitilmiş

ağlar kendi arasında karşılaştırılmış, test aşamasında veri artırma işleminin etkileri gözlemlenmiş ve bunların hepsinin sentezi bir arada sunulmuştur. Önerilen çalışmada DenseNet201, InceptionV3, ResNet50 ve Xception ön-çeritilmiş ağları kullanarak histopatolojik görüntüler üzerinde iyi huylu (benign) ve kötü huylu (malignant) meme kanser görüntüleri sınıflandırılmıştır. Böylece hekimlere karar aşamasında destek sağlayabilecek bir yardımcı uygulama geliştirilmiştir. Diğer çalışmalara göre Xception ağına oldukça başarılı sonuçlar ortaya koyduğu görülmüştür. Kullanılan veri seti, çözünürlük, kullanılan yöntemler ve gürültü için uygulanan çözümler gibi yönlerden her ağ farklı sonuçlar verebilmekte hatta her bir çalıştırma da dahi sonuçlar değişebilmektedir. Ancak yapılan testler ve uygulamalar sonucu olabildiğince eşit şartlarda karşılaştırılan bu dört ağ arasında Xception önceden eğitilmiş ağı daha başarılı sonuçlar elde edilmiştir. Çalışmada kullanılan modellerin karmaşıklık matrisleri çıkarılmış, iyi huylu ve kötü huylu olarak meme kanseri sınıflandırmasındaki başarımları daha anlaşılır bir biçimde gösterilmiştir. Ağların her bir epoktaki doğruluk değerleri grafiksel olarak çizdirilerek görsel açıdan gözlemlenmesi sağlanmıştır. Nadiren de olsa DenseNet201 ağı Xception ağından daha başarılı sonuçlar elde edilmiştir. Çalışmanın sonraki kısımları şu şekilde organize edilmiştir. İkinci bölüm materyal ve metot bölümünden oluşmakta olup, modelin genel blok diyagramı, kullanılan metotlar, veri seti gibi kavramlardan bahsedilmiştir. Üçüncü bölüm ise araştırma sonuçları ve tartışma bölümüdür. Bu bölümde elde edilen sonuçlar açık bir şekilde paylaşılmış, ağlar grafiksel ve matematiksel olarak karşılaştırılmış ve bir sonuca bağlanmıştır. Son bölümde ise çalışma ile ulaşılan sonuçlar ve gelecek için düşünülen çalışmalar sunulmuştur.

2. Materyal ve Metot

Bu çalışmada histopatolojik görüntüler üzerinde meme kanserinin tespiti için önerilen derin sinir ağlarına dayalı modelin genel blok diyagramı Şekil 1'de gösterilmiştir. Önerilen modelde görüntü alındıktan sonra ilk olarak, derin sinir ağları için oldukça önem arz eden yeniden boyutlandırma işlemi gerçekleştirilmektedir. Burada derin sinir ağlarının aldığı girdilerin aynı boyutta olması gerekmektedir ve bu nedenle girdiler CNN (Convolutional Neural Network) mimarisine girmeden önce sabit bir boyuta getirilir. Bunun yanında yapılan çalışmaya göre yüksek çözünürlüğe sahip bir görüntü bellekte daha fazla yer kaplar ve daha büyük bir sinir ağı kullanılmasına da neden olur (Hashemi, 2019). Yine aynı çalışmaya göre eğer görüntünün çözünürlüğü çok fazla küçültülürse detaylar ve görüntü de bulunan özellikler kaybolabilir. Bu gibi sebeplerden ötürü bu çalışmada yeniden boyutlandırma işlemi yatayda ve düşeyde 224x224 olarak uygulanmıştır. Bunun yanında, veri artırma ile veri seti daha fazla genişletilerek ağı daha fazla örnek sunulmuş ve ağı eğitimi yürütülmüştür. Sonraki aşamada görüntüye uygulanan özellik çıkartma ile ayırt edici özellikler belirlenmektedir. Havuzlama (Pooling) işlemi yapılarak sınıflandırmadaki görüntünün şekli herhangi bir şekilde değişse dahi, sistemin bunu tanıyabilmesi ve mevcut görüntünün özelliklerini kaybetmeden daha küçük boyutta bir matris elde edilmesi amaçlanmıştır (Tolias, Sicre, & Jégou, 2015). Ek olarak, ezberleme (Over-fitting) işleminden kaçınmak amacıyla bırakma (dropout) işlemi gerçekleştirilmiştir. Yığın normalizasyonu (batch normalization) derin sinir ağlarının hızını, performansını ve sürekliliğini arttıran bir yöntemdir (Ioffe & Szegedy, 2015) (Collis, 2017). Son olarak görüntüler sınıflandırılmak için tam bağlantılı katmandan geçer ve iyi huylu veya kötü huylu meme kanseri sınıflandırma işlemi tamamlanır.



Şekil 1. Histopatolojik görüntülerde meme kanserinin tespiti ve sınıflandırılması için önerilen derin sinir ağları tabanlı modelin blok diyagramı

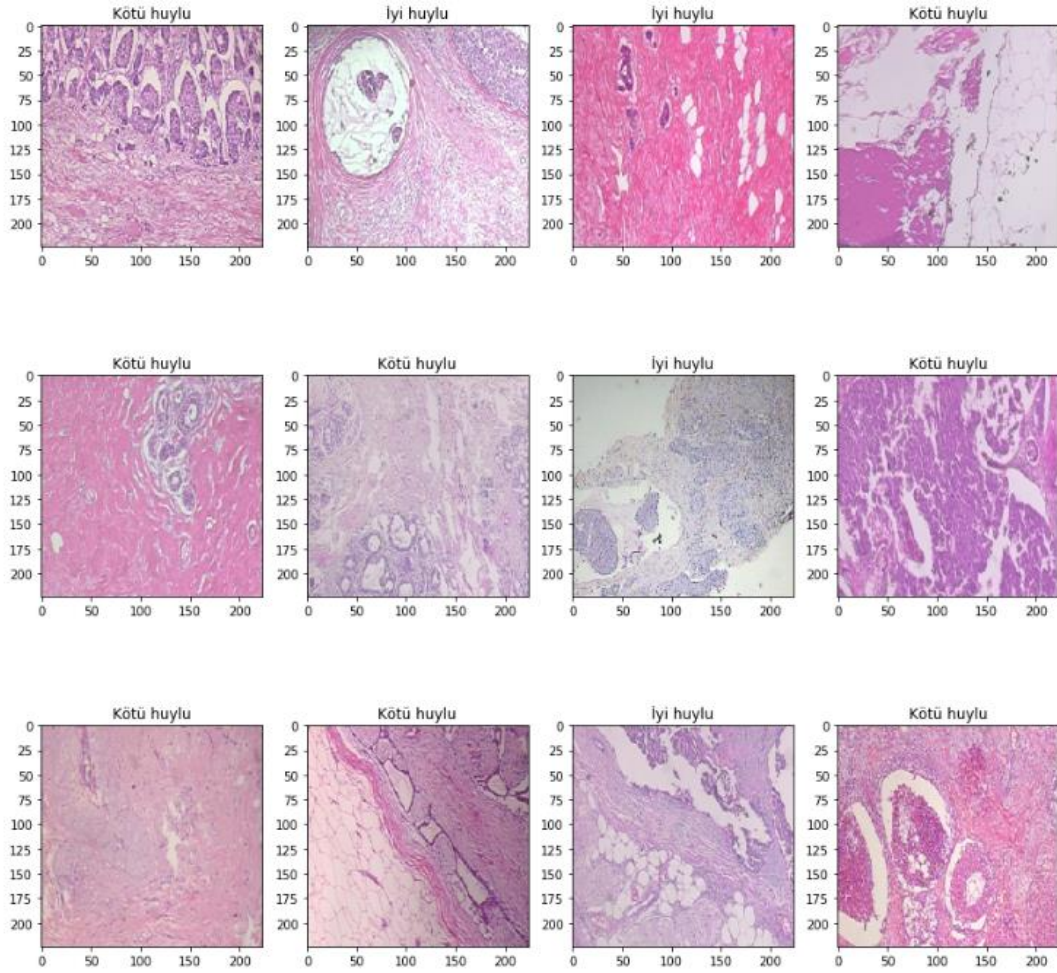
2.1. Meme Kanseri Veriseti

Bu çalışmada önerilen modelin performansını değerlendirmek amacıyla oldukça fazla ve kaliteli histopatolojik görüntü içeren BreakHis veri seti kullanılmıştır (Fabio A Spanhol, Oliveira, Petitjean, & Heutte, 2015). Bu veri setinde 82 hastadan toplanan toplamda 7909 mikroskopik görüntü bulunmaktadır. 2480 iyi huylu ve 5429 kötü huylu görüntüden oluşan bu veri seti 700x460 çözünürlüğe ve 3 kanal RGB – 8 bit derinliğe sahip PNG formatındadır. Bu veri seti P&D laboratuvarı-Patolojik Anatomi ve Sitopatoloji tarafından Parana-Brezilya'da oluşturulmuştur. Veri seti Tablo 1'de gösterildiği üzere 40X, 100X, 200X ve 400X büyütme oranlarına sahip

görüntülerden oluşmaktadır. Şekil 2’de bu veri setinde bulunan iyi huylu ve kötü huylu görüntülerin 224x224 çözünürlükteki bazı örnekleri gösterilmiştir (Fabio Alexandre Spanhol, Oliveira, Petitjean, & Heutte, 2016) (Fabio A Spanhol, Oliveira, Cavalin, Petitjean, & Heutte, 2017).

Tablo 1. BreakHis veri setindeki yakınlaştırma oranları ve görüntü sayıları

Yakınlaştırma Oranı	İyi Huylu Görüntü Sayısı	Kötü Huylu Görüntü Sayısı	Toplam
40X Oranında	652	1370	1995
100X Oranında	644	1437	2081
200X Oranında	623	1390	2013
400X Oranında	588	1232	1820
Görüntü Toplamı	2480	5429	7909

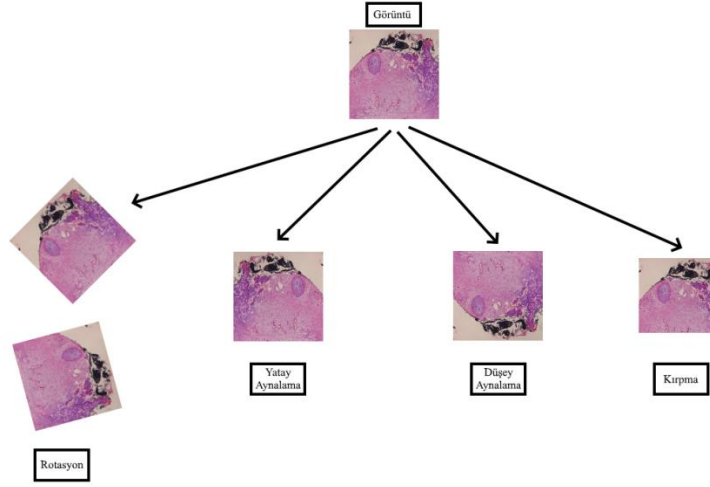


Şekil 2. BreakHis verisetinden iyi huylu ve kötü huylu örnek histopatolojik görüntüler

2.2. Veri Artırma(Data Augmentation) İşlemi

Derin öğrenme ve makine öğrenmesi algoritmalarında karşılaşılan en büyük sorun ağı eğitecek yeterince veri olmamasıdır. Yeterince veri bulunmaması bu algoritmalarda sık sık ortaya çıkan ve büyük bir sorun olan ezberleme (over-fitting) problemini meydana getirir. Bu olay kısaca ağın gördüğü eğitim verilerini (görüntü, text vb.) ezberlemesi ve eğitim verileri dışında bir girdi ile karşılaştığında başarı gösterememesidir. Bu sorundan kurtulmanın en önemli yöntemlerinden birisi veri artırımı yapılmasıdır. Bu yöntem eğitim setine uygulanır ve elde bulunan verilerin özellikleri değiştirilerek yapay (artificial) olarak daha fazla görüntü elde edilir. Bu çalışmada kullanılan verisetinde de veri artırımı yapılarak görüntüler yatay ve dikey olarak döndürülmüş, görüntülere

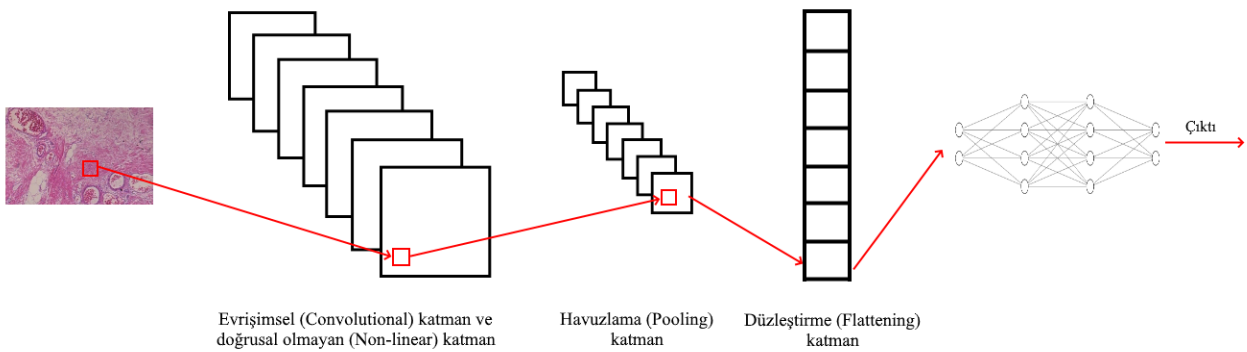
rastgele yakınlaştırma değerleri uygulanmıştır (Perez & Wang, 2017) (Taylor & Nitschke, 2017). Şekil 3'te uygulanan veri artırma işlemlerinden bazıları şekilsel olarak gösterilmiştir. Veri artırma işleminin test setine de uygulanması işleme Test-zamanı-artırması (Test time augmentation - TTA) ismi verilmiştir (Moshkov, Mathe, Kertesz-Farkas, Hollandi, & Horvath, 2020).



Şekil 3. Örnek bir histopatolojik görüntüye veri artırma (data augmentation) işlemlerinin uygulanması

2.3. Evrişimsel Sinir Ağı (CNN – Convolutional Neural Network) ve Ön-Eğitilmiş Ağlar

Derin öğrenmede ki temel amaç yüksek-boyutlu (high-dimensional) verilerden bazı özellikler çıkartmak ve girdiler ile çıktılarını bu özellikler vasıtasıyla ilişkilendirerek bir sonuca varmaktır. Karmaşık özellikleri çıkartabilmek için derin öğrenme algoritmaları genelde çok katmanlı bir yapıya sahiptir. Evrişimsel sinir ağları derin öğrenmede kullanılan bir ağ modelidir. Şekil 4'te gösterildiği üzere görüntüdeki pikseller öncelikle evrişimsel katman vasıtasıyla özellik çıkartma (feature extraction) işlemine tabi tutulur. Özellik çıkartmada belirli bir filtre (maske) tüm görüntü üzerinde dolaştırılır ve gerçekten iyi huylu veya kötü huylu görüntünün pikseli ile oluşturulan yeni görüntü arasında belirli bir olasılıkta benzerlik kurulur. Daha sonra sisteme doğrusal olmayan (non linear – çıktıdaki değişimin girdinin değişimi ile orantılı olmadığı sistemler) gösterilir. Doğrusal olmayan katman aktivasyon katmanı olarak da anılır çünkü burada bir aktivasyon fonksiyonu kullanılır. Doğrusallığın problem olmasının nedeni basitçe tüm ağın tek bir algılayıcı (perceptron) gibi davranmasının önüne geçilmesidir. Bu çalışmada, aktivasyon fonksiyonu olarak softmax kullanılmıştır. Havuzlama katmanında ağırlık sayısı azaltılır, uygunluk kontrol edilir. Havuzlama işlemindeki amaç mevcut resmin özelliklerini kaybetmeden daha küçük boyutta bir matris (görüntü) elde etmektir. Havuzlama işleminde de özellik çıkartmadakine benzer bir filtreleme uygulanır. Düzleştirme katmanında ise klasik sinir ağı yapısı için veriler hazırlanır ve sunulur. Bu katmandaki temel amaç verileri tam bağlı katman (fully connected layer) için hazırlamaktır. Son olarak tam bağlı katmanda ise yapay sinir ağları vasıtasıyla öğrenme sağlanır (Ide & Kurita, 2017) (Paasio & Dawidziuk, 1999) (Ma et al., 2017) (Kalchbrenner, Grefenstette, & Blunsom, 2014).



Şekil 4. Çalışmada kullanılan evrişimsel sinir ağı mimarisinin genel yapısı

Ön-eğitilmiş ağlar CNN modellerinde öğrenmenin daha hızlı sağlanması, işlemlerin daha hızlı yapılması ve doğruluk başarımının artırılması için oldukça önemlidir. Bu ağlar benzer bir problem üzerinde önceden kullanılarak buna bağlı ağırlıklar alan modeller olarak bilinir. Her bir problem için tam olarak en uygun çözüm olmasalar dahi aynı işlemleri tekrar tekrar yapmanın önüne geçerler. Bu ağlar genellikle ImageNet ismi verilen ve görsel nesne tanıma üzerine kurulmuş çok büyük bir veritabanı üzerinde önceden kullanılmıştır. Bu çalışmada da ön-eğitilmiş olan ResNet50, DenseNet201, InceptionV3 ve Xception ağları kullanılmıştır.

Residual Network (ResNet) artık değerlerin (residual value) ve artık blokların (residual block) modele eklenmesiyle meydana gelir. Her iki katmanda bir sisteme farklı bir değer ekleyerek gradyen kaybolmasının (vanishing gradient) önüne geçer (Tai, Yang, & Liu, 2017).

ResNet'e benzer bir ön-eğitilmiş ağ olan DenseNet'te ResNet'ten farklı olarak 2 katmanda bir sisteme eklenen değer kendinden sonraki tüm katmanlara eklenmiştir. Bu sistem modele birçok avantaj kazandırmıştır. Bu sayede ağ daha kolay eğitilebilir hale gelmiş ve yüksek katmanlı ağlarda çoğu katmanın işlevsiz olma sorununu optimize etmiştir (G. Huang, Liu, Van Der Maaten, & Weinberger, 2017).

Inception ağı, ağ içinde ağ olarak tanımlanabilir. Evrişim katmanlarındaki filtreleme işlemlerinin ve havuzlama işleminin eş zamanlı olarak gerçekleştirilmesi temeline dayanır. Modüller halinde işlemler yapmaktadır. InceptionV3 ağı ise ek modüllerin küme normalizasyonu (batch normalization) ve full connected layer'da yardımcı sınıflayıcı olarak ağın içinde kullanır (Szegedy, Ioffe, Vanhoucke, & Alemi, 2017) (Szegedy, Vanhoucke, Ioffe, Shlens, & Wojna, 2016).

Xception ağı ise, temelde InceptionV3 ağının üzerine ekleyerek gelişen bir ağ olmuştur. Diğer ağlara göre başarıyı yakalamasındaki en önemli etken ise evrişimsel (convolutional) kısımdaki yaptığı farklılıklardır. Normal bir ağ evrişimsel bölümde bir filtreyi en, boy ve derinlik gibi çok boyutlu matrisler üzerinde gezdirerek işlemlerini meydana getirmektedir. Xception ağı bu kısımda InceptionV3'teki gelişmelere ek olarak 2 farklı yaklaşım sunmaktadır. Bunlar, akıllı derinlik evrişimi (depthwise convolution) ve akıllı nokta evrişimi (Pointwise convolution)'dir. Akıllı derinlik evrişimi bölümünde her kanalda değil de sadece tek bir kanalda işlem yaparak sonuca ulaşır. Örneğin RGB değerlerine sahip bir görüntüde 3 kanalı değil 1 kanalı kullanır. Bu birçok özellik kaybına ve oldukça başarısız sonuçlara neden olacağından akıllı nokta evrişiminde 1 kanal üzerinde işlem yapılarak elde edilen görüntü üzerinde klasik evrişim işlemi 1x1xKanal Sayısı olarak uygulanır ve sonuca ulaşılır (Chollet, 2017).

2.4. Birakma(Dropout) İşlemi

Derin öğrenme metodlarında modellerin birleştirilerek (kombinasyon) kullanılması pek çok durumda daha iyi bir sonuç verir. Bununla birlikte büyük ağların eğitimi çok fazla hesaplama gerektirir ve en iyi sonuca ulaşmak için gerekli parametreleri bulmak oldukça zorlu bir işlemdir. Ayrıca gelişmiş ağların eğitilmesi için oldukça fazla veriye ihtiyaç duyulmaktadır ve farklı ağların farklı veri setleri ile eğitilmesi için yeterli düzeyde veri bulunamayabilir. Bir şekilde bu sorunlar çözülerek modeller birleştirilerek kullanılsa dahi hepsini test anında kullanmak hızın önemli olduğu durumlarda mümkün değildir. Veri artırma işlemi modellere uygulanmasına karşın ezberleme problemi de kimi durumlarda tam olarak çözülememektedir. Bu problemler ile başa çıkmak için bırakma yöntemi önem arz etmektedir. Bırakma işleminde temel işlem eğitim sırasında nöronları (bağlantıları ile birlikte) modelden (sinir ağı) rastgele olarak düşürmektir. Bu işlem yalnızca eğitim safhasında uygulanır test aşamasında uygulanmaz. Ayrıca bırakma işlemi pek çok farklı modele uygulanabilmektedir (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014).

3. Araştırma Sonuçları ve Tartışma

Bu çalışmada ResNet50, DenseNet201, InceptionV3 ve Xception olmak üzere dört farklı ön-eğitilmiş ağ kullanılmış ve gerçekleştirilen deneysel çalışmalar ile her birisi ile elde edilen sonuçlar kendi arasında karşılaştırılmıştır. Ulaşılan sonuçlar histopatolojik görüntüler üzerinde meme kanserinin tespiti ve sınıflandırılmasında Xception ağının daha başarılı olduğunu göstermektedir. Nadiren de olsa, DenseNet201 ağı Xception ile benzer ve hatta iyi sonuçlar elde etse de bu durum bir süreklilik arz etmemektedir.

Bu çalışma da önerilen model eğitilirken kullanılan bilgisayar Intel Core i7-7700HQ 2.80GHz işlemci, 16 GB DDR4 RAM ve 4 GB NVIDIA Geforce GTX 1050 Ti grafik kartına sahiptir. Yapılan çalışma da kullanılan görüntülerin %80'i eğitim için ve kalan %20'si test için kullanılmıştır. Deneysel çalışmalar Python programlama dili ve Keras kütüphanesi ortamı kullanılarak gerçekleştirilmiştir.

Kullanılan veri setine göre görüntünün gerçek niteliğinin (GGN) kötü huylu ve model tarafından tahmin edilen niteliğinin (MTTEN) de kötü huylu olduğu durum Doğru Pozitif (DP), GGN iyi huylu ve MTTEN kötü huylu olduğu durum Yanlış Pozitif (YP), GGN kötü huylu MTTEN iyi huylu olduğu durum Yanlış Negatif (YN) ve GGN iyi huylu MTTEN'de iyi huylu olduğu durum Doğru Negatif (DN) olarak isimlendirilir. Başarım ölçümleri için kullanılan doğruluk, hassasiyet, hatırlama ve F1 skoru performans ölçütleri aşağıdaki eşitliklerde gösterilmiştir (Davis & Goadrich, 2006) (David Martin Powers, 2011). Burada dikkat edilmesi gereken durum hassasiyet ve hatırlama değerlerinin herbir sınıf için hesaplanması gerektiğidir. Bu çalışma da iyi huylu ve kötü huylu olarak 2 sınıf bulunmaktadır.

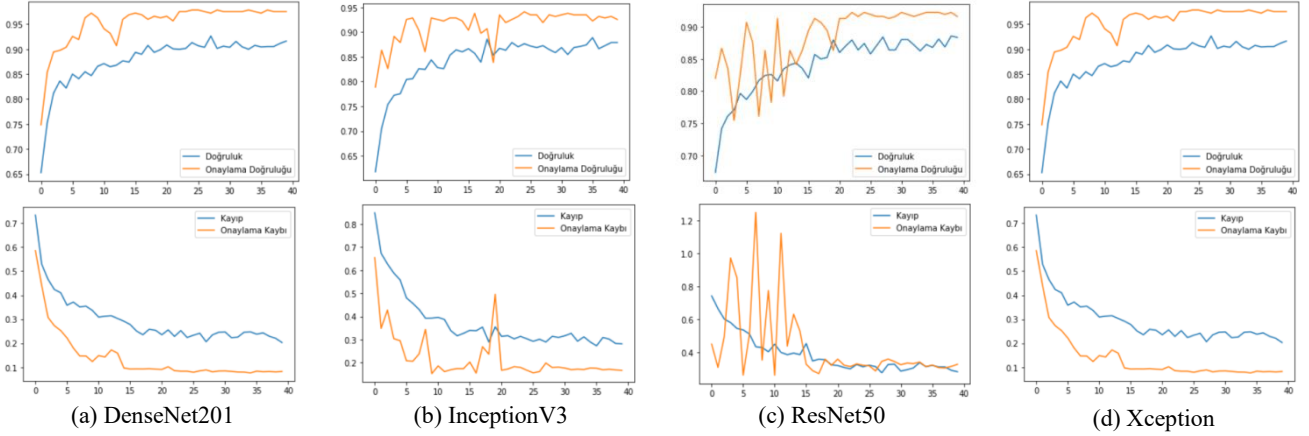
$$\text{Doğruluk(Accuracy)} = \frac{DP + DN}{(DP + DN + YP + YN)}$$

$$\text{Hassasiyet(Precision)} = \frac{DP}{(DP + YP)}$$

$$\text{Hatırlama(Recall)} = \frac{DP}{(DP + YN)}$$

$$\text{F1 Skor} = \frac{2 * (\text{Hassasiyet} * \text{Hatırlama})}{(\text{Hassasiyet} + \text{Hatırlama})}$$

Çalışmada kullanılan tüm ağlar 40 epok boyunca eğitilmiş, doğruluk, onaylama doğruluğu, kayıp ve onaylama kaybı değerlerini gösteren grafikler Şekil 5'te gösterilmiştir. Bu grafikler veri setindeki 200X yakınlaştırma oranına sahip görüntülerin sonuçlarını göstermektedir. Ağların tüm yakınlaştırma oranlarındaki detaylı analizleri ise Şekil 5'te gösterilmiştir. Elde edilen verilere göre Xception ağı ile diğer ağlara göre daha başarılı sonuçlar elde edilmiştir.



Şekil 5. DenseNet201, InceptionV3, ResNet50 ve Xception önceden eğitilmiş ağlarının 200X yakınlaştırma oranındaki başarımlarını gösteren grafikleri

Tablo 2'de sunulan sonuçlar, bu çalışma da kullanılan tüm önceden eğitilmiş ağların 40X, 100X, 200X ve 400X yakınlaştırma oranlarına sahip histopatolojik görüntüler üzerindeki eğitimlerinin sonuçlarını özetler niteliktedir. Doğruluk skoru değerleri ağların genel olarak başarımlarını göstermektedir. Bu değerlere göre ağların farklı yakınlaştırma değerlerindeki doğruluk skorları oldukça farklılık göstermektedir. Örneğin ResNet50 ağı 200X yakınlaştırma oranında 95.54'lük bir başarı elde ederken 400X yakınlaştırma oranında bu değer 90.43'e gerilemiştir. Bütün ağların doğruluk skorları genel olarak karşılaştırıldığında ise Xception ağına göstermiş olduğu başarı açık bir şekilde görülmektedir. Özellikle 200X yakınlaştırma oranında Xception 98.01 yani %98.01'lik bir başarı ortaya koymuştur. Bu tablodaki makro ortalama ve ağırlıklı ortalama değerleri her bir performans ölçütü için ayrı ayrı hesaplanır. Eşitlik (1), Eşitlik (2) ve Eşitlik (3)'te sırasıyla hassasiyet değeri, hatırlama değeri ve bu iki değer harmonik ortalaması olan F1 skor değerinin makro ortalamaları gösterilmiştir. Eşitlik (4), Eşitlik (5) ve Eşitlik (6)'da ise hassasiyet değeri, hatırlama değeri ve F1 skor değerinin ağırlıklı ortalamaları gösterilmiştir. Eşitliklerdeki n değeri veri setindeki sınıf sayısını; P_1, P_2, \dots, P_n dizisi ise her bir sınıf için hesaplanan hassasiyet değerini; R_1, R_2, \dots, R_n ise her bir sınıf için hesaplanan hatırlama değerini göstermektedir (Behera & Kumaravelan, 2019) (Vani & Rao, 2019) (Goutte & Gaussier, 2005) (David MW Powers, 2015).

$$\text{Hassasiyet Değeri (Makro Ortalama)} = \frac{\sum_{i=1}^n P_i}{n} \quad (1)$$

$$\text{Hatırlama Değeri (Makro Ortalama)} = \frac{\sum_{i=1}^n R_i}{n} \quad (2)$$

$$\text{F1 Skor Değeri (Makro Ortalama)} = \frac{2}{\frac{1}{\frac{\sum_{i=1}^n P_i}{n}} + \frac{1}{\frac{\sum_{i=1}^n R_i}{n}}} \quad (3)$$

$$\text{Hassasiyet Değeri (Ağırlık. Ort.)} = \frac{\sum_{i=1}^n |y_i| \frac{dp_i}{dp_i + yp_i}}{\sum_{i=1}^n |y_i|} \quad (4)$$

$$\text{Hatırlama Değeri (Ağırlık. Ort.)} = \frac{\sum_{i=1}^n |y_i| \frac{dp_i}{dp_i + yn_i}}{\sum_{i=1}^n |y_i|} \quad (5)$$

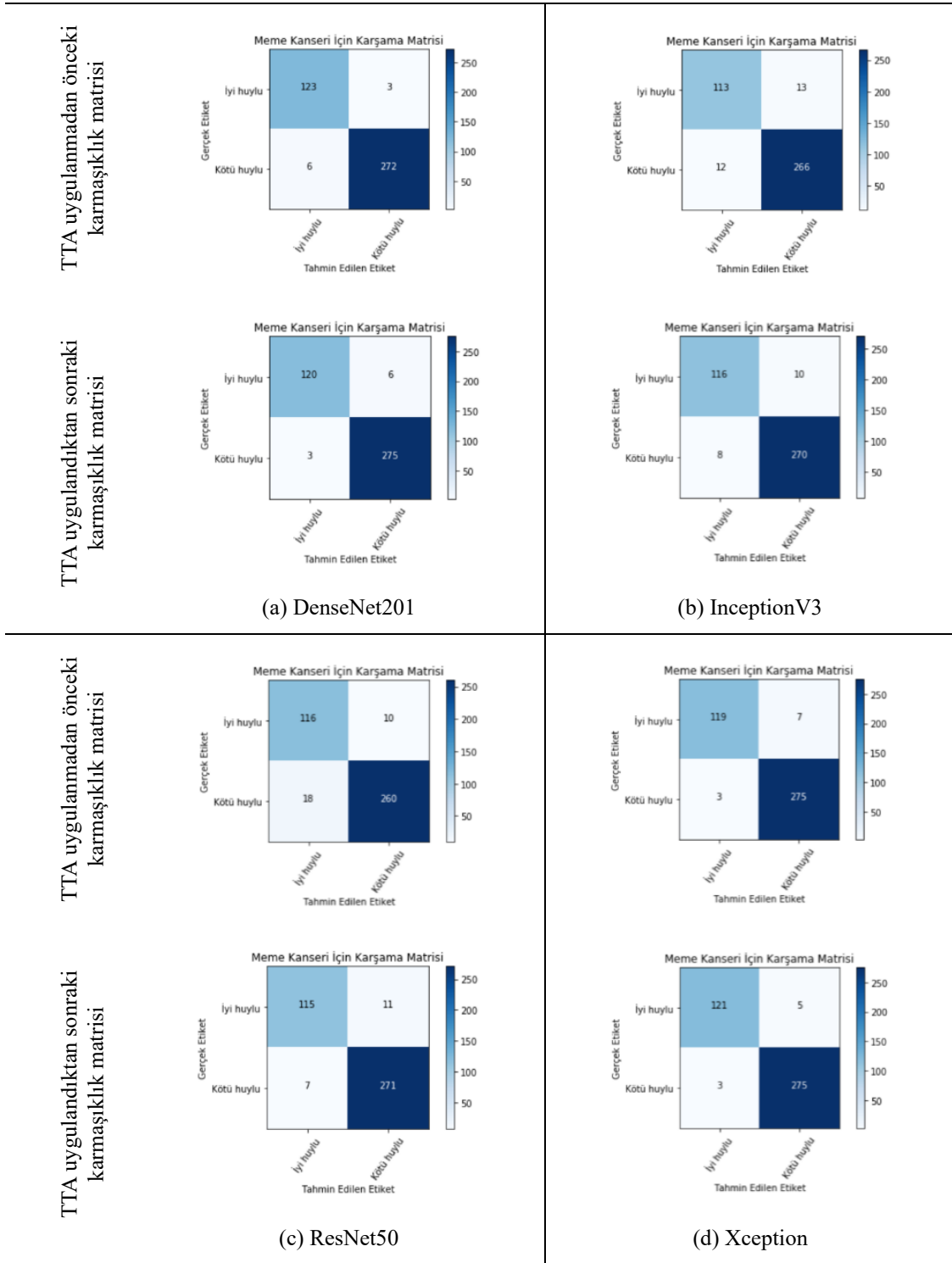
$$\text{F1 Skor Değeri (Ağırlık. Ort.)} = \frac{\sum_{i=1}^n |y_i| \frac{2dp_i}{2dp_i + yn_i + yp_i}}{\sum_{i=1}^n |y_i|} \quad (6)$$

Tablo 2’de gösterilen F1 skor ve doğruluk skor değerleri ağların başarımlarının değerlendirilmesinde kullanılan ölçütlerdir. Bu ölçütlerin hesaplanma yöntemleri birbirinden farklılık gösterebilir dahi ikisi de performans ölçütü olarak kullanılabilir ve F1 skor, hassasiyet, hatırlama ve doğruluk değerlerinin hesaplanma yöntemleri dikkate alındığında doğruluk değerinin gerçek pozitif ve gerçek negatif değerleri önem arz ettiğinde F1 skor değerinin ise yanlış negatif ve yanlış pozitif değerlerinin önem arz ettiği durumlarda kullanılmasının daha uygun olduğu söylenebilir. Yine Tablo 2’deki doğruluk ve F1 skor değerleri dikkate alındığında bu ağlar arasındaki kıyaslama Xception -> DenseNet201 -> InceptionV3 ve ResNet50 şeklinde sıralanabilir. Bu tablodan çıkarılabilecek bir diğer husus ağların yakınlaştırma oranlarına göre verdikleri sonuçların durumlara göre farklılık göstermesidir. Örneğin InceptionV3 ağı 400X yakınlaştırma oranında 92.34 gibi nispeten düşük bir doğruluk skoru değerine ulaşırken 200X değerinde 95.54 gibi oldukça yüksek bir değere ulaşmıştır. Bu da göstermektedir ki veriseti (farklı yakınlaştırma oranları), seçilen ağ, kullanılacak parametreler bunların tamamı başarımlarına etki etmekte ve buna dayalı olarak seçimler yapılmalıdır.

Tablo 2. Bütün ağlarda elde edilen başarımlarının karşılaştırılması (D.S.: Doğruluk Skoru, Has. D.: Hassasiyet Değeri, Hat.D.: Hatırlama Değeri, M.O.: Makro Ortalama, A.O.: Ağırlıklandırılmış Ortalama)

Ağ Adı	Yakınlaştırma	D.S.	Has.D. (M.O.)	Hat.D. (M.O.)	F1 Skoru (M.O.)	Has.D. (A.O.)	Hat. D. (A.O.)	F1 Skoru (A.O.)
DenseNet201	40X	97.50	96.92	97.31	97.11	97.51	97.50	97.50
	100X	94.49	94.41	92.63	93.51	94.48	94.49	94.49
	200X	97.77	97.71	97.07	97.39	97.77	97.77	97.77
	400X	95.08	94.20	94.59	94.40	95.10	95.08	95.09
InceptionV3	40X	94.75	93.55	94.45	93.99	94.83	94.75	94.79
	100X	95.21	94.96	93.78	94.37	95.19	95.21	95.20
	200X	95.54	94.98	94.59	94.79	95.53	95.54	95.53
	400X	92.34	92.15	90.13	91.13	92.32	92.34	92.33
ResNet50	40X	93.25	93.66	90.57	92.09	93.31	93.25	93.28
	100X	93.30	94.28	90.07	92.13	93.50	93.30	93.40
	200X	95.54	95.18	94.37	94.77	95.52	95.54	95.53
	400X	90.43	90.90	86.94	88.88	90.54	90.43	90.48
Xception	40X	97.25	97.11	96.49	96.80	97.24	97.25	97.24
	100X	94.97	94.39	93.82	94.10	94.95	94.97	94.96
	200X	98.01	97.89	97.47	97.68	98.01	98.01	98.01
	400X	95.35	94.16	95.46	94.80	95.49	95.35	95.42

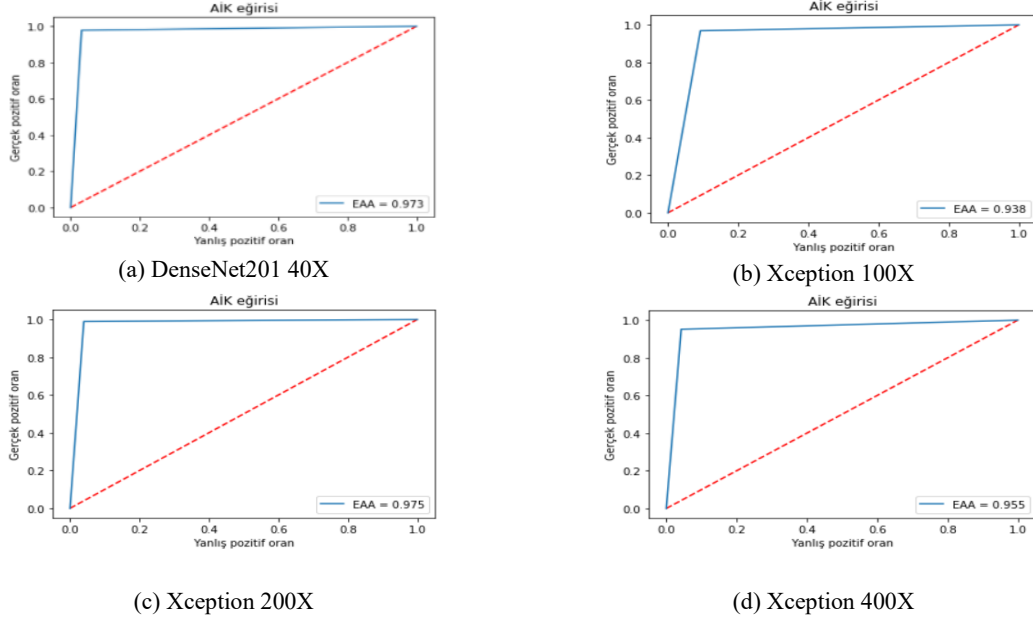
Xception ağının en başarılı olduğu 200X yakınlaştırma oranı için Xception ve diğer ağların karmaşıklık matrisi TTA işlemi uygulanmadan önce ve uygulandıktan sonra Şekil 6’da gösterilmiştir. Karmaşıklık matrislerinde ağların veri setinde yapılan ayırma sonucu görüntülere verdikleri tahminler ve görüntülerin gerçek değerleri gösterilmektedir. Xception ağının karmaşıklık matrisi incelenecek olursa (TTA işleminden sonra) toplam 404 görüntü üzerinde gerçekten iyi huylu olan 126 görüntü ve gerçekten kötü huylu olan 278 görüntü bulunmaktadır. Xception ağı iyi huyluların 121 tanesini doğru (iyi huylu tahmini yapmış) ve 5 tanesini yanlış (kötü huylu tahmini yapmış) tahmin etmiş, kötü huyluların ise 275 tanesini doğru (kötü huylu tahmini yapmış) ve 3 tanesini yanlış (iyi huylu tahmini yapmış) değerlendirmiştir. Diğer ağların karmaşıklık matrisleri de bu şekilde yorumlanabilir ve bu tahminler arasında Xception ağının çok daha başarılı olduğu görülür. Öte yandan TTA işleminin ağlara olan etkileri incelendiğinde toplam hatalı tahmin sayısı: Xception için 10’dan 8’e, ResNet50 için 20’den 18’e, InceptionV3 için 25’ten 18’e düşmüş ve DenseNet201 ağı için de olumlu veya olumsuz bir sonuç doğurmamıştır. Bu bilgiler ışığında TTA işlemi üç ağı olumlu yönde etkilemiş bir ağı ise ne olumlu ne de olumsuz olarak etkilememiştir. Buna göre TTA işleminin genel anlamda ağların başarımlarını olumlu yönde etkilediği söylenir.



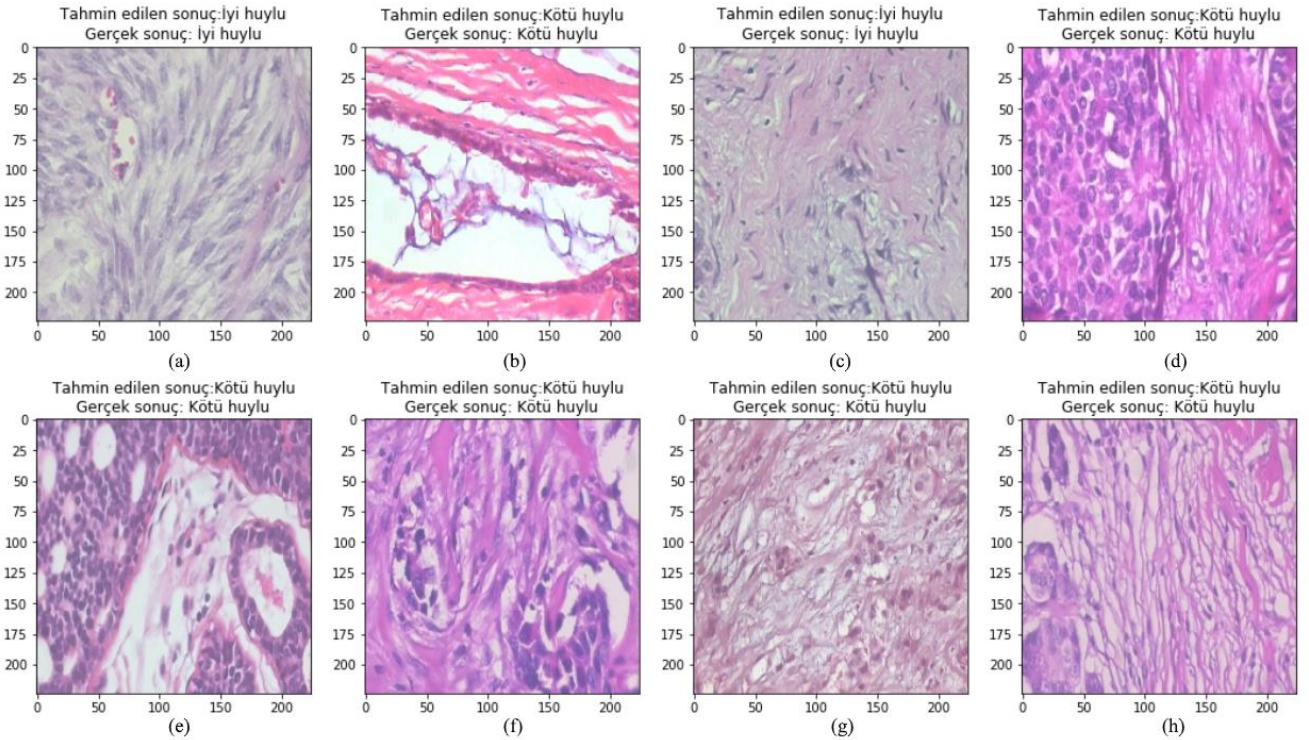
Şekil 6. 200X yakınlaştırma oranındaki görüntülerde ağların TTA'lı ve TTA'sız karmaşıklık matrisleri

Makine öğrenmesi uygulamalarında performans değerlendirmesi için en yaygın olarak kullanılan yöntemlerden birisi de AİK [Alıcı İşletim Karakteristiği – ROC (Receiver Operating Characteristic)] eğrileridir. AİK grafikleri modelin tahminlerinde ne kadar başarılı olduğunu gösterir. AİK eğrisinin altında kalan alan ise EAA [Eğri altındaki alan – AUC (Area under curve)] olarak adlandırılır. EAA değeri bir alan kapsamıyla gösterilir ve bu alan ne kadar geniş ise modelin sınıflandırma da elde ettiği başarı da o kadar yüksek olarak tanımlanır. EAA değeri için optimum değer 1 olarak ifade edilir (Bradley, 1997) (Hajian-Tilaki, 2013). Şekil 7’de DenseNet201 ağının 40X ve Xception ağının 100X, 200X ile 400X yakınlaştırma oranlarındaki AİK grafikleri ve EAA değerleri gösterilmiştir. EAA değeri, doğruluk değeri, f1 skor değeri hepsi ağların başarımlarının değerlendirilmesi için kullanılan ölçütlerdir. Kimi çalışmalara göre EAA değeri doğruluk değerine göre daha güvenilir ve ağıın başarımlarını ölçmede doğruluk değerine göre daha doğru sonuçlar vermektedir (Ling, Huang, & Zhang, 2003) (J. Huang & Ling, 2005). Şekil 7’de yer alan değerlere göre DenseNet201 ağının 40X yakınlaştırma oranındaki EAA değeri (0.973) Xception ağının EAA değerine (0.965) göre daha yüksektir. Ayrıca 40X’te F1 skor (A.O.) değerinde de DenseNet201 daha başarılı bir sonuç ortaya koymuştur. Bu da sınıflandırma da DenseNet201’in daha

başarılı olduğu anlamına gelmekte ve hatta doğruluk değerine göre de DenseNet201 ağı daha başarılıdır. Bu durumda 40X yakınlaştırma oranı özelinde DenseNet201 ağı bu dört ağ arasındaki en başarılı ağıdır. Buna karşın 40X harici bütün yakınlaştırma oranları incelendiğinde diğer tüm yakınlaştırma oranlarında Xception ağının EAA ve doğruluk değerinin daha yüksek olduğu görülmektedir. Sonuç olarak 40X yakınlaştırma oranında DenseNet201, 100X, 200X ve 400X yakınlaştırma oranında Xception ağı diğer ağlara oranla daha başarılıdır. Bu iki ağ arasında ise Xception ağının daha başarılı olduğu söylenir.

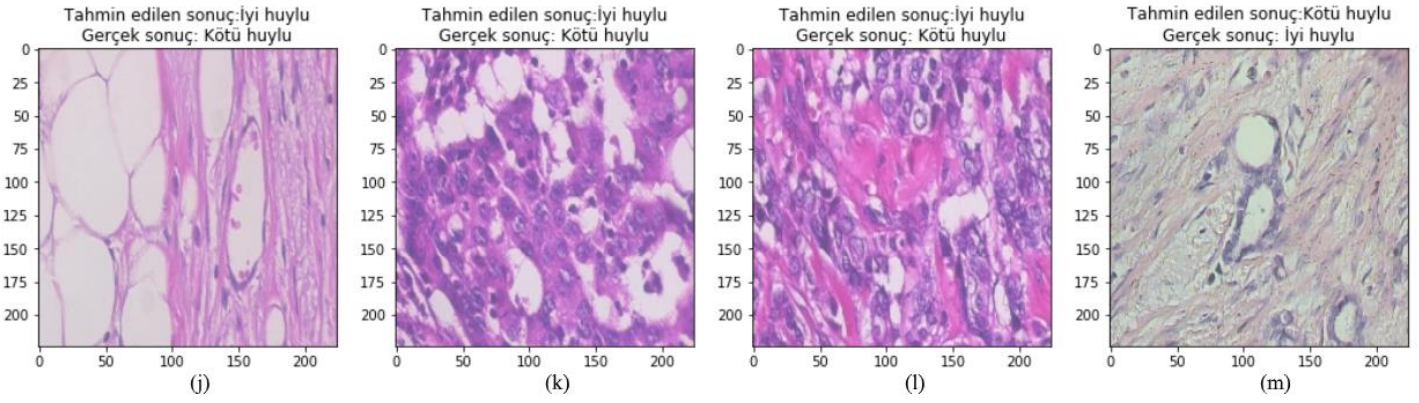


Şekil 7. Yakınlaştırma oranlarına göre ağların AIK eğrilerini ve EAA değerleri



Şekil 8. Xception ağının 200X yakınlaştırma oranındaki tahminleri ve gerçek sonuçlar.

Şekil 8'de rastgele seçilen 8 adet 200X yakınlaştırma oranındaki histopatolojik görüntü için Xception ağının yapmış olduğu tahminler ve bu görüntülerin gerçek değerleri gösterilmektedir. Bu görüntüler resizing(yeniden boyutlandırma) işlemi ile 224x224 konumuna getirilmiş ve ağın bu görüntüler için bir tahmin yapması istenmiştir. Görüldüğü üzere b, d, e, f, g ve h'de gerçek sonuçlar kötü huylu ve ağın tahmin ettiği sonuçta kötü huyludur. a ve c'de ise gerçek sonuçlar iyi huylu ve ağın tahmin ettiği sonuçta iyi huyludur. Bu 8 resim içinde Xception ağı doğru tahmin yapmıştır. Buna karşın Şekil 9'da gösterildiği üzere j, k ve l'de gerçek sonuç kötü huylu iken Xception iyi huylu tahmini yapmıştır. Son olarak m'de ise gerçek sonuç iyi huylu iken Xception kötü huylu tahmini yapmıştır. Bu görüntülerde ise Xception iyi huylu/kötü huylu meme kanseri ayırımında yanılmıştır.



Şekil 9. Xception ağıının 200X yakınlaştırma oranındaki tahminleri ve gerçek sonuçlar.

4. Sonuç

Bu çalışmada verisetinde dört farklı yakınlaştırma oranlarındaki deneysel çalışmalarda, ağlar yakınlaştırma oranından etkilenmiş ve yapı aynı olsa dahi farklı başarımlarına ulaşmışlardır. Bu da göstermektedir ki veri setlerinin farklı olması bir kenara aynı görüntüde ki farklı yakınlaştırma oranları dahi ağların başarımlarını etkilemektedir. Eğitim sırasındaki başarımların daima onaylama aşamasındakine göre daha düşük kalmıştır. Bu başlangıçta kötü bir durummuş gibi gözükse de aslında oldukça faydalı hatta neredeyse her durumda istenilen bir şeydir çünkü ağın son hali onaylama işlemi sırasında ki ağıdır ve son hali o parametrelerde bulunan ağ olacaktır. Bunun yanı sıra eğitim sırasındaki başarımlarının daha düşük olmasının sebebi ezberleme işlemine de engel olan bırakma değeridir. Bırakma değeri eğitim sırasında bazı birimleri yok sayarak parametrelerin gelişmesini sağlayarak onaylama sırasında daha sağlam ve daha doğru kararlar alabilen bir model ortaya koymaktadır bu nedenle onaylama aşamasındaki başarımların daha yüksektir. Yapılan analizler ve karşılaştırmalara göre Xception ağı ve DenseNet201 ağı her bir yakınlaştırma oranına sahip veri setinde yüksek performans metrik değerleri elde ederek diğer 2 ağa göre bu veri setinde daha başarılı olduklarını göstermişlerdir. Bu 2 başarılı ağ arasında bir karşılaştırma da ise Xception ağı çok az farkla da olsa daha başarılıdır. Ulaşılan bir diğer sonuç ise TTA işleminin sonuçlara olumlu yönde etki ettiği olmuştur.

Çalışmada kullanılan görüntüler ağa girdi olarak vermeden önce aynı boyuta getirilmelidir. Bu çalışma da tek bir veri seti kullanıldığı için bu durum bir sorun olmamıştır; ancak birden fazla veri setinden görüntüler alındığı durumda da veya aynı veri setinde ama farklı boyutlarda görüntüler kullanıldığında bunların mutlaka aynı boyuta getirilmesi gerekmektedir. Model eğitiminin çok uzun süreler almaması için güçlü bir donanıma sahip bilgisayar kullanılmalıdır.

Gelecekteki çalışmalar için doğruluk değerini artırmaya yönelik daha geniş ve derin bir veri seti kullanılması amaçlanmaktadır. Yapılan diğer çalışmalardaki ve özgün veri setleri insanların kullanımına açılırsa onlarla karşılaştırmak ve yapılacak çalışmalarda kullanılması oldukça fayda sağlayacaktır. Bunun yanında donanımsal gelişmelerde makine öğrenmesi ve görüntü işleme gibi konularda oldukça önem arz ettiğinden bu alandaki gelişmeler de takip edilecektir. Yüksek bir donanım geliştirmesi ekran kartlarının ve işlemcilerin yüksek hızlara ulaşması bu alanın geleceğini çok olumlu yönde etkileyecek sürelerin kısaltılmasına vesile olacaktır. Kişisel verilerin kullanımı tabii ki çok dikkat edilmesi gereken bir konudur; fakat bazı durumlarda bunun tıbbi ve insanlık yararına olacağı durumlar için bir nebze hafifletilmesi belki de bu alan için çok daha farklı çalışmalar ortaya koyabilir.

Teşekkür

Bu çalışma da kullanılan veri setinin temin edilmiş olduğu Robotik Görme ve Görüntü Laboratuvarına (Laboratório Visão Robótica e Imagem) ayrıca veri setinin bizimle buluşmasını sağlayan Spanhol, F., Oliveira, L.S., Petitjean, C., Heutte ve L.'ya teşekkür ederiz. Ayrıca bu çalışmanın yazarları, bu çalışmayı 2019-01.BŞEÜ.25-02 proje numarası ile destekleyen Bilecik Şeyh Edebali Üniversitesi Bilimsel Araştırmalar Koordinatörlüğüne teşekkürlerini sunmaktadır.

Kaynakça

- Abdel-Zaher, A. M., & Eldeib, A. M. (2016). Breast cancer classification using deep belief networks. *Expert Systems with Applications*, 46, 139-144.
- Azar, A. T., & El-Said, S. A. (2013). Probabilistic neural network for breast cancer classification. *Neural Computing and Applications*, 23(6), 1737-1751.
- Behera, B., & Kumaravelan, G. (2019). *Performance Evaluation of Deep Learning Algorithms in Biomedical Document Classification*. Paper presented at the 2019 11th International Conference on Advanced Computing (ICoAC).
- Bradley, A. P. (1997). The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7), 1145-1159.

- Chollet, F. (2017). *Xception: Deep learning with depthwise separable convolutions*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Collis, J. (2017). Glossary of Deep Learning: Batch Normalisation: Jun.
- Davis, J., & Goadrich, M. (2006). *The relationship between Precision-Recall and ROC curves*. Paper presented at the Proceedings of the 23rd international conference on Machine learning.
- Dogan, B. E., Gonzalez-Angulo, A. M., Gilcrease, M., Dryden, M. J., & Yang, W. T. (2010). Multimodality imaging of triple receptor-negative tumors with mammography, ultrasound, and MRI. *American Journal of Roentgenology*, 194(4), 1160-1166.
- Goutte, C., & Gaussier, E. (2005). *A probabilistic interpretation of precision, recall and F-score, with implication for evaluation*. Paper presented at the European Conference on Information Retrieval.
- Hajjian-Tilaki, K. (2013). Receiver operating characteristic (ROC) curve analysis for medical diagnostic test evaluation. *Caspian journal of internal medicine*, 4(2), 627.
- Han, Z., Wei, B., Zheng, Y., Yin, Y., Li, K., & Li, S. (2017). Breast cancer multi-classification from histopathological images with structured deep learning model. *Scientific reports*, 7(1), 1-10.
- Hashemi, M. (2019). Enlarging smaller images before inputting into convolutional neural network: zero-padding vs. interpolation. *Journal of Big Data*, 6(1), 98.
- Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). *Densely connected convolutional networks*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Huang, J., & Ling, C. X. (2005). Using AUC and accuracy in evaluating learning algorithms. *IEEE Transactions on knowledge and Data Engineering*, 17(3), 299-310.
- Ide, H., & Kurita, T. (2017). *Improvement of learning for CNN with ReLU activation by sparse regularization*. Paper presented at the 2017 International Joint Conference on Neural Networks (IJCNN).
- Ioffe, S., & Szegedy, C. (2015). Batch normalization: Accelerating deep network training by reducing internal covariate shift. *arXiv preprint arXiv:1502.03167*.
- Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *arXiv preprint arXiv:1404.2188*.
- Kumar, A., Singh, S. K., Saxena, S., Lakshmanan, K., Sangaiah, A. K., Chauhan, H., . . . Singh, R. K. (2020). Deep feature learning for histopathological image classification of canine mammary tumors and human breast cancer. *Information Sciences*, 508, 405-421.
- Ling, C. X., Huang, J., & Zhang, H. (2003). *AUC: a better measure than accuracy in comparing learning algorithms*. Paper presented at the Conference of the canadian society for computational studies of intelligence.
- Ma, X., Dai, Z., He, Z., Ma, J., Wang, Y., & Wang, Y. (2017). Learning traffic as images: a deep convolutional neural network for large-scale transportation network speed prediction. *Sensors*, 17(4), 818.
- Mohammed, M. A., Al-Khateeb, B., Rashid, A. N., Ibrahim, D. A., Ghani, M. K. A., & Mostafa, S. A. (2018). Neural network and multi-fractal dimension features for breast cancer classification from ultrasound images. *Computers & Electrical Engineering*, 70, 871-882.
- Moshkov, N., Mathe, B., Kertesz-Farkas, A., Hollandi, R., & Horvath, P. (2020). Test-time augmentation for deep learning-based cell segmentation on microscopy images. *Scientific reports*, 10(1), 1-7.
- Nahid, A.-A., Mehrabi, M. A., & Kong, Y. (2018). Histopathological breast cancer image classification by deep neural network techniques guided by local clustering. *BioMed research international*, 2018.
- Öztürk, Ş., & Akdemir, B. (2019). HIC-net: A deep convolutional neural network model for classification of histopathological breast images. *Computers & Electrical Engineering*, 76, 299-310.
- Paasio, A., & Dawidziuk, A. (1999). CNN template robustness with different output nonlinearities. *International Journal of Circuit Theory and Applications*, 27(1), 87-102.
- Perez, L., & Wang, J. (2017). The effectiveness of data augmentation in image classification using deep learning. *arXiv preprint arXiv:1712.04621*.
- Powers, D. M. (2011). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation.
- Powers, D. M. (2015). What the F-measure doesn't measure: Features, Flaws, Fallacies and Fixes. *arXiv preprint arXiv:1503.06410*.
- Rakhlin, A., Shvets, A., Iglovikov, V., & Kalinin, A. A. (2018). *Deep convolutional neural networks for breast cancer histology image analysis*. Paper presented at the International Conference Image Analysis and Recognition.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2015). Cancer statistics, 2015. *CA: a cancer journal for clinicians*, 65(1), 5-29.
- Siegel, R. L., Miller, K. D., & Jemal, A. (2019). Cancer statistics, 2019. *CA: a cancer journal for clinicians*, 69(1), 7-34.
- Spanhol, F. A., Oliveira, L. S., Cavalin, P. R., Petitjean, C., & Heutte, L. (2017). *Deep features for breast cancer histopathological image classification*. Paper presented at the 2017 IEEE International Conference on Systems, Man, and Cybernetics (SMC).
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2015). A dataset for breast cancer histopathological image classification. *IEEE Transactions on Biomedical Engineering*, 63(7), 1455-1462.
- Spanhol, F. A., Oliveira, L. S., Petitjean, C., & Heutte, L. (2016). *Breast cancer histopathological image classification using convolutional neural networks*. Paper presented at the 2016 international joint conference on neural networks (IJCNN).
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1), 1929-1958.
- Stewart, B., & Wild, C. P. (2014). World cancer report 2014.
- Sudharshan, P., Petitjean, C., Spanhol, F., Oliveira, L. E., Heutte, L., & Honeine, P. (2019). Multiple instance learning for histopathological breast cancer image classification. *Expert Systems with Applications*, 117, 103-111.

- Szegedy, C., Ioffe, S., Vanhoucke, V., & Alemi, A. A. (2017). *Inception-v4, inception-resnet and the impact of residual connections on learning*. Paper presented at the Thirty-first AAAI conference on artificial intelligence.
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., & Wojna, Z. (2016). *Rethinking the inception architecture for computer vision*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Tai, Y., Yang, J., & Liu, X. (2017). *Image super-resolution via deep recursive residual network*. Paper presented at the Proceedings of the IEEE conference on computer vision and pattern recognition.
- Taylor, L., & Nitschke, G. (2017). Improving deep learning using generic data augmentation. *arXiv preprint arXiv:1708.06020*.
- Vani, S., & Rao, T. M. (2019). *An Experimental Approach towards the Performance Assessment of Various Optimizers on Convolutional Neural Network*. Paper presented at the 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI).
- Wang, D., Khosla, A., Gargeya, R., Irshad, H., & Beck, A. H. (2016). Deep learning for identifying metastatic breast cancer. *arXiv preprint arXiv:1606.05718*.