Avrupa Bilim ve Teknoloji Dergisi
Özel Sayı, S. 404-410, Ekim 2019
© Telif hakkı EJOSAT'a aittir
Araştırma Makalesi

www.ejosat.com ISSN:2148-2683

European Journal of Science and Technology
Special Issue, pp. 404-410, October 2019
Copyright © 2019 EJOSAT
Research Article

Lenf Kanserine İlişkin Patoloji Görüntülerinin Makine Öğrenimi Yöntemleri ile Sınıflandırılması

Ayşe Berika Varol[1], İsmail İşeri[2*]

[1]Ondokuz Mayıs Üniversitesi, Fen bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, Samsun, Türkiye (ORCID: 0000-0003-1856-9636)
[2]Ondokuz Mayıs Üniversitesi, Fen bilimleri Enstitüsü, Bilgisayar Mühendisliği Bölümü, Samsun, Türkiye (ORCID: 0000-0002-0442-1406)

(Bu yayın International Congress on Human-Computer Interaction,Optimization and Robotic Application kongresinde sözlü olarak sunulmuştur.)

(İlk Geliş Tarihi 1 Ağustos 2019 ve Kabul Tarihi 25 Ekim 2019)

(**DOI:** 10.31590/ejosat.638372)

Özet

Son yıllarda dijital patoloji görüntülerinin sınıflandırılması konusunda yapılan görüntü işleme ve makine öğrenimi temelli çalışmalarda oldukça başarılar sonuçlar elde edildiği görülmektedir. Elde edilen yüksek doğruluk değerleri, dijital patoloji alanında makine öğrenimi temelli sistemlerin patoloji kliniklerinde patologlara yardımcı sistemler olarak kullanılabileceğini göstermektedir. Önümüzdeki 30 yıl içerisinde özellikle patoloji alanında yapay zeka ve makine öğrenimi temelli çözümlerin çok daha yüksek oranda kullanılacağı öngörülmektedir. Bu çalışmada lenf kanserinin üç farklı türüne ait dijital patoloji görüntülerin farklı makine öğrenimi teknikleri ile sınıflandırılması işlemi gerçekleştirilmiştir. Çalışma kapsamında veri seti olarak Chronic Lymphocytic Leukaemia (CLL), Follicular Lymphoma (FL) ve Mantle Cell Lymphoma (MCL) kanserlerine ilişkin dijital patolojik görüntüler kullanılarak özellik çıkarımı ve makine öğrenim algoritmalarının eğitilmesi ve karşılaştırılması yapılmıştır. Çalışmada her bir lenf kanseri türüne ait 45 adet olacak şekilde toplamda 135 adet dijital patoloji görüntüsü ön işlemlerden gerçirilerek renk yoğunluğu, piksel yoğunluğu, entropi hesabı ve morfolojik alan hesabı özellikleri elde edilmiştir. Ardından her bir görüntü için elde edilen özellik vektörleri Random Forest, K-NN, Navie Bayes, Destek Vektör Makinesi (SVM) ve K-Star algoritmalarına girdi olarak verilerek sınıflandırma işlemi gerçekleştirilmiştir. Son aşamada elde edilen değerler Özgüllük (Specificity), Hassasiyet (Precision), Geri Çağırma (Recall) ve Doğruluk (Accuracy) performans metriklerine göre hesaplanıp, algoritmaların kıyaslaması yapılmıştır. Bu yöntemler ile algoritmaların performans değerleri karşılaştırıldığında en iyi sonuç %89,72 doğruluk oranı ortalamasıyla Random Forest tarafından elde edilmiştir.

Anahtar Kelimeler - Görüntü İşleme, Makine Öğrenimi, Patoloji

Classification of Lymph Cancer Pathology Images Using Machine Learning Methods

Abstract

In recent years, it is seen that the results obtained from the image processing and machine learning based studies on the classification of digital pathology images have been achieved quite successfully. The high accuracy values show that machine learning based systems in the field of digital pathology can be used as second reader systems for pathologists in pathology clinics. It is expected that in the next 30 years, solutions based on artificial intelligence and machine learning will be used at a much higher rate, especially in the field of pathology. In this study, digital pathology images belonging to three different types of lymph cancer were classified with different machine learning techniques. In the scope of the study, feature extraction and machine learning algorithms were trained and compared by using digital pathological images of Chronic Lymphocytic Leukaemia (CLL), Follicular Lymphoma (FL) and Mantle Cell Lymphoma (MCL) cancers as data set. In this study, 45 digital pathology images of each type of lymph carcinoma totaly 135 images were passed from the preprocessing stage and then color density, pixel density and entropy calculation features were exctracted. Then, the feature vectors obtained for each image were used as input to Random Forest, K-NN, Navie Bayes, Support Vector Machine (SVM) and K-Star algorithms and the classification process was performed. In the last stage the Specificity, Precision, Recall and Accuracy

performance metrics were calculated and the algorithms were compared. The best result were obtained by Random Forest Classifier for classification of three types of lymphoma with an average accuracy of 89.72%.

Keywords: Image Processing, Machine Learning, Pathology

1. Giriş

Patoloji bilimi doku ve organlarda bulunan anormallikleri inceleyerek hücre yapılarındaki durumların analizini yaparak hastalık tanısının koyulmasına yardımcı olan bilim dalıdır. Bu sayede tanı koymamızda ve uygun tedavi yöntemini belirlememizde bize yardımcı olur (Celasun, 2018). Fakat ülkemizde patolog, teknisyen ve sekreter sayılarının materyal sayılarına göre düşük olması, var olan materyallerin bir kısmında sorun olması ve mekân donanımı ile ilgili ciddi sorunlar bulunması patoloji biliminin gelişimini kısıtlamaktadır. Ayrıca var olan arşiv verileri de belirli bir standarta uygun olarak saklanmamıştır (Yörükoğlu vd., 2009). Bu durumlara göre ülkemizde çoğu patoloji birimi dünya kalite kontrolü ve kalite güvencesi programlarına uymamaktadır (Yörükoğlu vd., 2009). Fakat son yıllarda dijital tarayıcıların gelişmesi ve çoğalması ile doku verileri dijital hale gelmiştir. Dijital patolojinin klinik iş akışını iyileştirmeye yardımcı olması, sahada cam lamellerin saklama ihtiyacını azaltması ve fiziksel lamellerin kırılma veya kaybolma riskini azaltma potansiyelinden dolayı tercih edilen bir yöntem olmuştur (Madabhushi ve Lee, 2016). Dijital patoloji verilerinin bilgisayarda işlenmesi sonucunda patoloji bilimi daha nitelikli hale gelmeye başlamıştır. Bu bilimin gelişmesi ve dijital tanı sistemlerinde kullanılmaya başlanmasıyla birlikte, hekimlerin hastalık teşhisini hızlı ve doğru bir şekilde gerçekleştirebilmesi sağlanmıştır. Accenture'ının 37.000 doktor ile yaptığı anket sonuçlarının raporuna göre doktorların %74'ü dijital platformların teşhis ve tedavi sürecinde daha doğru sonuçlar üret tiğini belirtmiştir (Digital Health). Bu sebepten günümüzde görüntü işleme temelli ve makine öğrenimi alt yapısına sahip tanı teşhis uygulamalarının önemi artmıştır.

1.1 Literatür Çalışmaları

Makine Öğrenimi algoritmaları sayesinde koşulların iyileştirilmesi, tedavi yöntemini modelleyerek yapılan test sonuçlarında yanlış teşhis koyma oranını azaltmak ve sonuçları belirli bir standartta koyarak hem hekimlerin hem de hastaların yaşam kaliteleri arttırılması amaçlanmaktadır (Kourou vd., 2015). Günümüzde makine öğrenmesi yöntemlerinin medikal alanda kullanımı oldukça yaygındır. Örneğin sara tanısı koymada (Erçelebi ve Subaşı,2006), meme kanseri sınıflandırmasında (Yıldız, 2012), (Abdoulaye ve Demir, 2017), (Wang vd., 2016), prostat kanseri tanı koymada (Khouzani ve Soltanian, 2003), MR görüntüleri ile beyin tümörü saptamada (Nabizadeh ve Kubat, 2015), şizofrenlik tespiti (Fan vd., 2005) ve lenf nodüllerinin kanser hücresi olup olmama durumunu tespit etme (Ishikawa vd., 2013), (Jiang vd., 2018) gibi alanlarda makine öğrenimi yöntemleri kullanılmaktadır. Abdoulaye ve diğerleri (2017), mamografi görüntülerini 3 temel aşamadan geçirmişlerdir. Görüntünün komşuluklarını inceleyerek gürültülerden arındırmışlardır ve nesnenin alan, çap, çevre gibi fiziki özelliklerini bularak örüntüyü çıkarımı yapmışlardır. Geliştirdikleri sistem ile çıkan örüntü herhangi bir yapay zekâ destekli algoritma ile eğitilerek kanser tespit sistemi geliştirmek için uygun olduğu sonucuna varmışlardır. Wang ve diğerleri (2016), meme kanseri histapoloji görüntülerinin sınıflandırılması için otomatik görüntü analizi tekniğini önermişlerdir. Nodüllerin sınıflandırılması için 4 şekil tabanlı özellik ve renk uzaylarını kullanan 138 özellik çıkarmışlardır. Ön işlem olarak arka planda kalan nesneleri aydınlatmak için alt-üst şapka dönüşümü kullanmışlardır. Ardından ROI' lerin yerini tespit edebilmek için dalgacık dönüşümü ile hücrelerin kenarlarını belirlemişlerdir. Bu işlemi büyüyen kanser hücrelerin yerini tespit etmek için yapmışlardır. Daha sonra birleşen hücreleri ayırmak için CSS köşe algılama yöntemi kullanarak çizgiler belirlenmiştir ve DSSM ile üst üste gelen hücreleri ayırmışlardır. Şekil özelliklerinde, hücrenin alan, çevre, dış merkezlilik ve yuvarlaklık özelliklerine bakarken dokusal özelliklerinde, gri seviye histogramının standart sapması, ortalaması, bağıl düzgünlük, tekdüzelik ve entropisine bakmışlardır. Sınıflandırma sırasında tüm özellikleri kullandıklarında performans sorunu ortaya çıktığı için CAGA denilen ajan genetik algoritması ile en iyi sonuç üreten özellikleri belirleyerek, seçilen özellikleri SVM algoritmasında sınıflandırmışlardır. Sonuç olarak önerdikleri yöntemin normal ve malign hücre görüntülerini sınıflandırmadaki doğruluk performans oranını %96,19 olarak bulmuşlardır. Jafari-Khouzani ve Soltanian (2003), prostat kanseri patolojik verilerinden yararlanarak gleason derecelendirme sistemini otomatik hala getirmek için çalışmalar yapmışlardır. Görüntünün enerji ve entropi özelliklerini hesaplamışlardır. Ardından özellik çıkarımı için multiwavelet (çoklu dalgacık dönüşümü), sınıflandırma için ise K-NN algoritmasını kullanmışlardır. Deney sonucunda %97 başarıya ulaşmışlardır. Ishikawa ve diğerleri (2013), patoloji görüntülerini HLAC, dalgacık dönüşümü ve Delaunay özellik çıkarıcıdan geçirerek kitlenin şekil, frekans ve konum bilgilerini elde ettikten sonra SVM algoritması ile tespit edilen hücrenin kanserli olup olmadığını saptamaya çalışmışlardır. Çalışmanın duyarlılık (sensitivity) değeri %94,6 olurken özgüllük değeri(specifity) %84,9 olarak hesaplamışlardır. Jiang ve diğerleri (2018), CLL, FL ve MCL lenf patolojileri üzerine çalışarak kendi yöntemlerini geliştirmişlerdir. Görüntüyü ön işlemden geçirerek boyutlandırdıktan sonra doku özellikleri entropi, yoğunluk ortalaması, yoğunluk standart sapması, LPQ, GLCM içeren bir özellik seti çıkarmışlardır. Çıkarılan özellikleri kullanarak patoloji görüntülerini SVM algoritması ile sınıflandırmışlardır. Bunun sonucunda ortalama ACC performans değerini %97,96 olarak elde etmişlerdir. Türkoğlu ve diğerleri (2013), veri içerisinde yer alan örüntüyü bulmak için patolojik görüntüleri logaritmik filtreden geçirdikten sonra morfolojik işlemlere tabi tutmuşlardır. Bu işlemlerin sonucunda görüntüdeki farklı olan hücreleri kırmızı kan hücrelerinden ayırt etmeyi başarmışlardır.Demir ve diğerleri (2015), akciğer nodül verilerini diğerleri gibi çeşitli filtrelerle görüntüleri işlemişlerdir. Bu sayede örüntüyü elde ederek, Dış yüzey doku özelliklerini Gray Level Cooccurrence Matrisi (GLCM) kullanarak hesaplamışlardır ve araştırmada kullanılan destek vektör makinesini, parçacık sürü algoritması ile optimize etmişlerdir. Deneyler sonucunda dış yüzey dokusal özellikler dâhil edilerek ölçüldüğünde %98,03 başarı varken, akciğer duvarına bitişik nodüllerin ve gri seviye eşik değerlerin dışında olan nodüller CAD sistemi tarafından tespit

edilememiştir. Avunduk ve diğerleri (2007), patoloji görüntüsünü yine diğerleri gibi gri seviyeye taşıyıp referans değerlerini belirleyerek resmin belirli kısımlarını almışlardır ve noktaya göre işlem yaparak yayılma algoritması uygulamışlardır. Ardından görüntüyü ikili formata çevirip beyaz kalan kısımların alanını hesaplayarak örüntüyü oluşturmuşlardır. Bu işlemlerin sonucunda alan hesaplamalarında %96,5'luk başarı elde etmişlerdir. Albayrak (2013), tarafından yapılan mitoz belirleme işlemlerinde, La*b* renk uzayında bulunan, resim bölütleme ve morfolojik işlemlerden geçirdikten sonra RasOr, RotOr ve DVM algoritmalarıyla sınıflandırmıştır. En iyi sonucu DVM yönteminde almıştır. İkinci bir yöntem olarak da piksellerde Haralick dokusal tanımlayıcı algoritması kullanarak özellik çıkarmış ve RasOr yöntemiyle sınıflandırılmasını yapmıştır. Mitozlu hücre silik ise etiketlemenin yanlış olabileceğine ve görüntünün La*b* renk uzayına dönüştürülmesinin sınıflandırma başarısına doğrudan katkı sağlamadığı sonucuna varmıştır. Demir ve diğerleri (2016), 3 tip ot tipini birbirinden ayırmak için R-G-B renk uzayını kullanarak renk ölçüm cihazı (DC) ve dijital kamera (CM) araçları ile otları ayırma işlemini gerçekleştirmiştir. Deneyin sonucunda CM olmadan da DC ile elde edilmiş görüntüyü işleyerek karşılaştırma için gerekli olan yeterli bilgileri elde edebilmeyi başarmışlardır.

Literatür taramasına göre araştırmacılar patoloji görüntüsü üzerindeki kanserli kitleyi tespit edebilmek için öncelikle görüntü üzerinde örüntü çıkarımı yapmışlardır. Örüntünün özellik seçimi sırasında genellikle görüntü üzerindeki kitlenin farklı boyut, renk ya da konumda olması, görüntüde anomalilere sebep olduğundan, kitleyi tespit edebilmek için bu gibi özellikleri tercih etmişlerdir. Araştırmacılar özellik çıkarımının sonunda elde edilenleri kullanılarak, veri setinin belirli bir yüzdesi ile ağ eğitmişlerdir. Verinin geri kalan kısmı ile de ağı test ederek kanserli hücre tespit başarı oranı hesaplamışlardır. Bu kaynaklara göre (Abdoulaye ve Demir, 2017) ve (Avunduk vd., 2007) numaralı kaynaklar hem renkten hem de alan özelliklerinde yararlanırken, (Demir vd., 2016) numaralı kaynak sadece renk özelliğini kullanarak sınıflandırma yapmıştır. Ayrıca kaynak (Khouzani ve Soltanian, 2003) ve (Ishikawa vd., 2013) dalgacık dönüşümünden yararlanırken, entropi ve konum özelliklerini de sırasıyla kullanarak yüksek oranda başarı elde etmişlerdir.

Çalışmanın geri kalan kısmında 2. Bölüm' de kullanılan veri kümesinden, özellik çıkarım yönteminden ve sınıflandırma algoritmalarından bahsedilmiştir. 3. Bölüm' de kullanılan yöntemler ile elde edilen bulgular değerlendirilmiştir. 4. Bölüm' de ise elde edilen sonuç ve çıkarımlar açıklanarak makale sonlandırılmıştır.

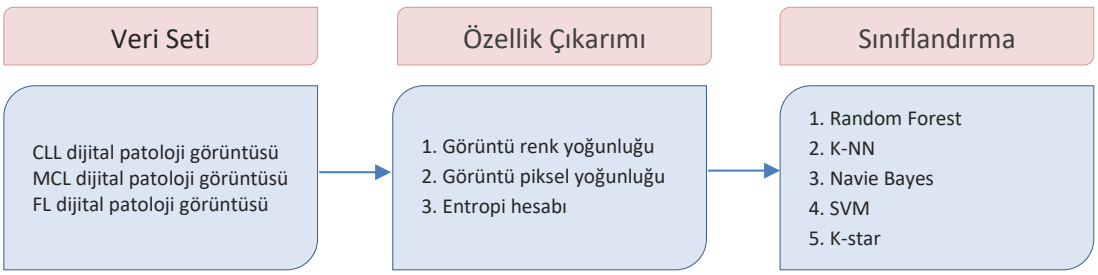
2. Materyal Ve Yöntem

Bu çalışma hazırlanırken Şekil.1 de göründüğü gibi 3 aşamalık bir yol izlenmiştir. İlk olarak 135 adet dijital patoloji verisi toplanmıştır ardından, görüntülerin piksel yoğunluğu, renk yoğunluğu, entropisi ve alan hesap özellikleri çıkarılmıştır. Özellikler çıkarıldıktan sonra bu veriler seçilen algoritmalardan geçirilerek patoloji sonuçları CLL, MCL ya da FL tümör tipleri arasında sınıflandırılmıştır. Bu lenf kanseri tipleri;

- Chronic Lymphocytic Leukemia (CLL), genellikle orta yaş ve üzerine görülen lenfositlerin kontrolsüz artmasıyla oluşan kanser türüdür. Bu kanser türü erkeklerde kadınlara oranla 2 kat daha fazla görülmektedir (Demir vd., 2012).
- Mantle Cell Lymphoma (MCL), daha çok ileri yaştaki erkeklerde görülen B lenfositinin maligne dönüşmesi sonucu olan lenf kanseri türüdür (Koçyiğit vd., 2008).
- Follicular Lymphoma (FL), yavaş ilerleyen non-Hodgkin lenfomalar arasında en sık görülen ve 3 türe sahip olan tümör tipidir (Buckstein vd., 2005).

Dijital patoloji verileri her bir tümör tipinden 45 adet alınarak işlenmiştir. Bu görüntülerin birbirinden ayrımı renk değişiklikleri ve yoğunlukları sayesinde yapılmıştır. Ayrıca görüntülerin entropileri hesaplanarak sınıfların belirsizlik aralıkları çıkarılmıştır. Bu sayede sınıflandırma işleminde entropiden de yararlanılmıştır. Özellik çıkarımı sırasında morfolojik işlemler yaparak tümör olabilecek kitlelerin alan ve çevre ölçüleri hesaplanıp veri setine eklenmesine rağmen sonuca olumlu yönde hiçbir katkısı bulunmadığı için çıkarılmıştır. Sınıflandırmanın başarı değerleri hesaplanırken özgünlük (specificity), hassasiyet (precission), geri çağırma (recall) ve doğruluk (accurancy) metriklerinden yararlanılmıştır (Nguyen ve Armitage, 2008). Bu metrikler;

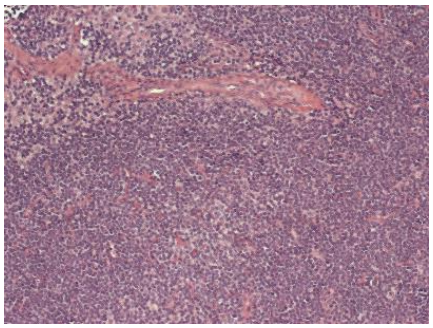
- Özgünlük (specificity), gerçekte yanlış olup, tahmin sırasında da ne kadar yanlış bulunduğunun oranını ölçer.
- Hassasiyet (precission), tahmin edilen örnekler içerisinden kaç tanesinin gerçekten doğru olduğunu ölçer.
- Geri çağırma (recall), pozitif sınıfa ait örneklerden kaç tanesinin doğru tahmin edildiğini ölçer.
- Doğruluk (accurancy), tahmin edilen doğru değerlerin oranını verir.



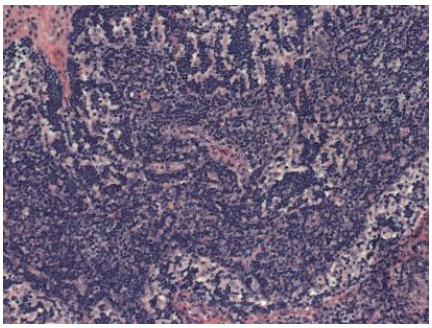
Şekil.1 *Sistemin işleyiş diyagramı*

2.1 Veri Seti

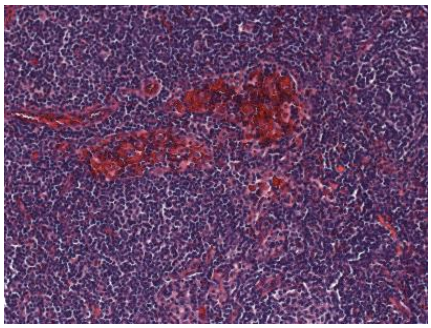
Çalışmada Lymphoma Sub-Type veri kümesi kullanılmıştır (Janowczyk, 2015). Bu veri kümesi 2015 yılında, CLL, FL ve MCL kanser türlerinden oluşturulmuş olup patoloji alanında yapılan çalışmalarda kullanılmaktadır (Janowczyk vd., 2015). Çalışma sırasında Lymphoma veri kümesinde bulunan her bir türden 45 adet alınarak toplamda 135 adet görüntü elde edilmiştir. Resim.1, Resim.2 ve Resim.3 de bu kanser türlerine örnek olarak verilmiş dijital görüntüleri mevcuttur.



***Resim.1** CLL*



***Resim.2** FL*



***Resim.3** MCL*

2.2 Özellik Çıkarımı

Veri setinin sınıflandırıcılara verildiğinde anlamlı sonuçlar çıkartabilmesi için verilerin önceden işlenerek ayırt edici özelliklerinin çıkartılması gerekir. Elimizdeki dijital patoloji verilerinde belirleyici kitle ya da ayırt edici şekiller yoktur. Yine de bu düşünceyi desteklemek için morfolojik işlemler yaparak resimlerde görünen nesnelerin alanları ve çevreleri hesaplanıp karşılaştırma yapılmıştır. Yapılan morfolojik işlemin sonucunda hiçbir ortak özellik yakalanamamıştır. Bu sebepten dolayı veriyi birbirinden ayırabilmek için renklerden yararlanmak daha doğru sonuçlar elde etmemizi sağlayacağından piksel temelli özellikler kullanılmıştır.

Bu çalışmada kullanılan 3 özellik;

I. Görüntünün Renk Yoğunluğu: RGB renk uzayının her bir eleman hesaplanıp elde edilen r-g-b değerlerinin ortalaması alınmıştır. Bu değer bize bir adet görüntünün renk yoğunluğunu verir.
II. Görüntünün Piksel Yoğunluğu: Görüntü RGB renk uzayından gri renk uzayına dönüştürülmüştür. Burada her bir pikselin sayısal bir karşılığı vardır ve bunlar matris şeklinde algılanır. Bu piksel değerlerinin aritmetik ortalaması alınarak yoğunluk hesaplanır.
III. Görüntünün Entropisi: Görüntü RGB renk uzayından Gray renk uzayına dönüştürülüp (1)'de gösterildiği gibi entropi değeri hesaplanır. Entropi görüntünün belirsizliğini ifade eder. Renk geçişlerindeki ani değişimler entropinin artmasına neden olur bu sayede verileri ayırt edebiliriz. Örneğin, duru bir resmin entropisi geçişli bir resmin entropisinden düşüktür.

$$\text{Entropy} = \sum_i P_i \log_2 P_i \qquad (1)$$

2.3 Sınıflandırma Yöntemleri

Sınıflandırma işlemi özellikleri çıkarılan verilerin çeşitli sınıflar üzerinde doğru bir şekilde dağıtılmasıdır. Veri dağıtımının başarılı olabilmesi için, sınıflandırma algoritmaları verilen eğitim kümesi tarafından eğitilir ve bu sayede verilerin dağılım şeklini öğrenirler. Eğitilen sınıflandırıcı modelleri gelen verinin özelliklerine bakarak hangi sınıfa ait olduğunu belirler ve atamasını gerçekleştirir. Bu işleyişe göre dijital patoloji verileri çeşitli sınıflandırma algoritmalarında geçirilerek, kişinin hangi tip kanser olabileceğinin olasılığı hesaplanır. Yapılan çalışmada 5 çeşit algoritma kullanılmıştır.

I. Random-Forest: Karar ağacı algoritmasını n defa çalıştırılarak daha iyi tahmin yapabilmeyi sağlar. Karar ağaçları veri setinden rastgele seçilmiş alt kümelerdir. Bu karar ağaçları bir araya gelerek karar ormanlarını oluştururlar. Kayıp verili veya dengesiz veri setlerinde iyi sonuçlar üretir(Ayyadevara & Ayyadevara, 2018).
II. K-NN: Örnek tabanlı bir algoritmadır. Sınıflandırmak istenen veriye en yakın n tane komşunun sınıfına bakarak yeni verinin sınıflandırılmasını yapar (Peterson, 2009).
III. Navie Bayes: Gürültüye duyarsız ve eksik verileri ihmal eden istatiksel bir sınıflandırıcıdır. Veri setindeki özelliklerin birbirinden bağımsız olmasını ister. Sisteme belli bir oranda öğretilmiş veri verilerek ağı eğitilir ve daha sonra verilen test verilerini doğru tahmin etmesini bekler (Nagar, Dabas, & Gupta, 2017).
IV. SVM: Değişkenler arasındaki örüntülerin bilinmediği veri setlerinde çalıştırılmak üzere önerilen istatiksel sınıflandırıcıdır. Sınıflandırma işlemini bir düzlemde bulunan verileri sınır çizerek iki gruba ayırır. Bu sınır verilerin birbirinden en uzak olduğu yer olarak belirlenir(Wang & Lin, 2014).
V. K-Star: Entropi tabanlı uzaklık fonksiyonunu kullanan örnek tabanlı sınıflandırıcıdır (Cleary vd., 1995).

3. Bulgular Ve Tartışma

Uygulamanın materyal ve yöntemleri belirlendikten sonra, patoloji veri setinden uygun (dijital patoloji görüntülerini homojen bir şekilde ayırabilmeyi sağlayan özellikler) özellikler çıkarılıp çeşitli sınıflandırma algoritmalarından geçirilmiştir. En uygun özellikler renk yoğunluğu, piksel yoğunluğu ve görüntünün entropisi olarak tespit edilmiştir. Alan hesaplamalarının özellik çıkarımında hiçbir ayırt edici özelliği olmadığından dolayı sınıflandırma aşamasında veri setine konulmamıştır. Özellikleri çıkarılan veri seti çeşitli algoritmalardan geçirildikten sonra Tablo.1-5 arasında göründüğü gibi 5 adet karmaşıklık matrisi elde edilmiştir. Algoritmadaki başarı seviyelerini tespit edebilmek için tablolarda yer alan değerleri kullanarak özgünlük (specificity), hassasiyet (precission), geri çağırma

(recall) ve doğruluk (accurancy) metrikleri hesaplanmıştır. En iyi sonuç Tablo.9' da göründüğü gibi %89,72 doğruluk oranıyla Random Forest algoritmasıdır. Tüm sınıflandırma algoritmaları için k-fold değeri 10 olarak belirlenmiştir ve modeli eğitmek için eğitim veri setine %60, test veri setine %40 veri ayırmak üzere toplamda 135 veri kullanılmıştır.

Tablo 1. Random Forest karışıklık matrisi

		TAHMİN SINIF		
		CLL	FL	MCL
GERÇEK SINIF	CLL	15	-	2
	FL	-	17	1
	MCL	2	3	14

Tablo 2. K-NN karışıklık matrisi

		TAHMİN SINIF		
		CLL	FL	MCL
GERÇEK SINIF	CLL	15	1	1
	FL	-	15	
	MCL	5	4	10

Tablo 3. Navie Bayes karışıklık matrisi

		TAHMİN SINIF		
		CLL	FL	MCL
GERÇEK SINIF	CLL	8	4	5
	FL	3	15	-
	MCL	7	9	3

Tablo 4. SVM karışıklık matrisi

		TAHMİN SINIF		
		CLL	FL	MCL
GERÇEK SINIF	CLL	15	2	-
	FL	-	15	3
	MCL	-	9	2

Tablo 5. K-Star karışıklık matrisi

		TAHMİN SINIF		
		CLL	FL	MCL
GERÇEK SINIF	CLL	11	3	3
	FL	-	15	3
	MCL	2	2	15

Tablo 6. Algoritmaların özgüllük değer karşılaştırması

Özgüllük (Specificity)	**CLL**	**FL**	**MCL**	**Özgüllük Ortalaması**
Random Forest	%93,93	%90,62	%91,42	%92,99
K-Nn	%88,23	%83,33	%96,77	%91,07
Navıe Bayes	%64,28	%45,83	%82,14	%64,08
Svm	%100	%60,71	%90,90	%83,87
K Star	%93,75	%83,87	%81,25	%86,29

Tablo.6 da algoritmaların özgünlük değeri hesaplanarak, gerçekte yanlış olan değerlerin algoritma tarafından yüzde kaç oranında yanlış tahmin edildiğini ölçmüştür. Elde edilen bu sonuçlara göre, Random Forest algoritması %92,99 özgüllük ortalaması ile tahminlerinde en iyi başarıyı elde etmiştir.

Tablo 7. Algoritmaların hassasiyet değer karşılaştırması

Hassasiyet (Precision)	**CLL**	**FL**	**MCL**	**Hassasiyet Ortalaması**
Random Forest	%88,23	%85	%82,35	%85,19
K-Nn	%75	%75	%71,42	%73,80
Navıe Bayes	%44,44	%53,57	%37,5	%45,17
Svm	%100	%57,69	%40	%65,89
K Star	%84,61	%75	%71,42	%77,01

Tablo.7 de algoritmaların hassasiyet değeri hesaplanarak, tüm sınıflardan ne kadar doğru tahmin değeri elde edildiği ölçülmüştür. Bu hesaplamaya göre, Random Forest algoritması %85,19 hassasiyet ortalaması ile en iyi başarıyı elde etmiştir.

Tablo 8. Algoritmaların geri çağırma değer karşılaştırması

Geri Çağırma (Recall)	**CLL**	**FL**	**MCL**	**Geri Çağırma Ortalaması**
Random Forest	%88,23	%94,44	%73,68	%85,45
K-Nn	%88,23	%100	%52,63	%80,28

Navıe Bayes	%47,05	%83,33	%15,78	%48,72
Svm	%88,23	%83,33	%18,18	%63,24
K Star	%64,70	%83,33	%78,94	%75,65

Tablo.8 de ise algoritmaların geri çağırma değeri hesaplanarak, tüm pozitif sınıflardan ne kadar doğru tahmin edildiğini ölçmüştür. Elde edilen sonuçlara göre, Random Forest algoritması %85,45 geri çağırma ortalaması ile tahminlerinde en iyi başarıyı elde etmiştir.

Tablo 9. Algoritmaların doğruluk değer karşılaştırması ve doğruluk ortalaması

Doğruluk Oranı (Accurancy)	**CLL**	**FL**	**MCL**	**Doğruluk Ortalaması**
Random Forest	%92,02	%92,01	%85,01	%89,72
K-Nn	%85,10	%88,88	%80	%84,66
Navıe Bayes	%57,02	%61,90	%55,31	%58,10
Svm	%94,11	%69,56	%72,72	%78,79
K Star	%83,67	%83,67	%80,39	%82,57

Son olarak Tablo.9 da algoritmaların doğruluk değeri hesaplanarak, tahmin edilen doğru değerlerin oranı ölçülmüştür. Elde edilen sonuçlara göre, Random Forest algoritması %89,72 doğruluk ortalaması ile tahminlerinde en iyi başarıyı elde etmiştir.

4. Sonuç

Çalışma da görüntü işleme yöntemlerinin dijital patoloji verilerinde kullanılması ile tümör tipinin belirlenerek doktorlar için hastalığı teşhis etme aşaması kolaylaştırılmaya çalışılmıştır. Çalışma sırasında bazı MCL görüntülerinin diğer tümör tiplerine benzemesinden dolayı genel doğruluk oranını düştüğü gözlenmiştir. Burada en iyi sonucu Random Forest algoritması vermiştir. Bu çalışma da hiçbir alan hesabı veya dönüştürücüler (Khouzani ve Soltanian, 2003), (Demir vd., 2015) olmadan sadece renklerden ve entropiden yararlanarak sonucu gidilmeye çalışılmıştır. Sonuç olarak Tablo.9 da gözüktüğü gibi %89,72'lik doğruluk oranı (accurancy rate) elde edilmiştir. Gelecekte Random Forest algoritması üzerinde daha fazla durularak doğruluk performans değerinin arttırılması hedeflenmektedir.

Kaynaklar

Abdoulaye, I. B. C., & Demir, Ö. (2017). Mamografi Görüntülerinden Kitle Tespiti Amacıyla Öznitelik Çıkarımı.

Albayrak, A. (2013). Histopatolojik görüntülerde mitoz belirleme.

Avunduk, M. C. & Sezgin, E. (2007). Patolojik Görüntülerin Bilgisayarlı Analiz Programı İle Değerlendirilmesi, Selçuk Üniversitesi Dijital Arşiv

Ayyadevara, V. K., & Ayyadevara, V. K. (2018). Random Forest. In Pro Machine Learning Algorithms (pp. 105–116). https://doi.org/10.1007/978-1-4842-3564-5_5

Buckstein, R., Pennell, N., & Berinstein, N. L. (2005). What is Remission in Follicular Lymphoma And What is its Relevance?. Best Practice & Research Clinical Haematology, 18(1), 27-56

Celasun, B., (2018). Patoloji Nedir?, http://www.turkpath.org.tr/content.php?id=35 , (Erişim: 10.03.2019)

Yörükoğlu, K., Usubütün, A., Doğan, Ö., Önal, B., & Aydın, Ö. (2009). Türkiye'de Patoloji Laboratuvarlarının Genel Profili. *Türk Patoloji Dergisi*, *25*, 19-28.

Cleary, J. G., & Trigg, L. E. (1995). K*: An Instance-Based Learner Using An Entropic Distance Measure, In Machine Learning Proceedings, pp. 108-114

Demir, B., Çetin, N., & Kuş, Z. A. (2016). Görüntü İşleme Tekniği ile Yabancı Ot Renk Özelliklerinin Belirlenmesi. Alınteri Zirai Bilimler Dergisi, 31(2), 59-64.

Demir, Ö., & Yılmaz Çamurcu, A. (2015). Computer-Aided Detection of Lung Nodules Using Outer Surface Features. *Bio-Medical Materials And* Engineering, 26(S1), S1213-S1222.

Demir, V., Kahraman, S., Katgı, A., Pişkin, Ö., Özsan, G. H., Demirkan, F., Ündar. B. & Özcan, M. A. (2012). Kronik Lenfositik Lösemi Hastalarının Genel Klinik Değerlendirilmesi. Dokuz Eylül Üniversitesi Tıp Fakültesi Dergisi, 26(1), 9-19

Digital Health, Accenture, Erişim adresi https://www.accenture.com/us-en/health-industry-index

Erçelebi, E., & Subasi, A. (2006). Classification Of EEG For Epilepsy Diagnosis in Wavelet Domain Using Artifical Neural Network And Multi Linear Regression. *In 2006 IEEE 14th Signal Processing and Communications Applications*.

Fan, Y., Shen, D., & Davatzikos, C. (2005, October). Classification of Structural İmages Via High-Dimensional Image Warping, Robust Feature Extraction, and SVM. *In International Conference on Medical Image Computing and Computer-Assisted Intervention* (pp. 1-8). Springer, Berlin, Heidelberg.

Ishikawa, T., Takahashi, J., Takemura, H., Mizoguchi, H., & Kuwata, T. (2013, October). Gastric Lymph Node Cancer Detection of Multiple Features Classifier for Pathology Diagnosis Support System. In *2013 IEEE International Conference on Systems, Man, and Cybernetics*, pp. 2611-2616

Jafari-Khouzani, K., & Soltanian-Zadeh, H. (2003). Multiwavelet Grading of Pathological Images of Prostate. *IEEE Transactions on Biomedical Engineering*, *50*(6), 697-704.

Janowczyk, A, (2015). Digital Histology, Deep Learnıng, Erişim adresi http://www.andrewjanowczyk.com/use-case-7-lymphoma-sub-type-classification/

Janowczyk, A., & Madabhushi, A. (2016). Deep Learning for Digital Pathology Image Analysis: A Comprehensive Tutorial with Selected Use Cases. Journal of Pathology Informatics, 7

Jiang, H., Li, Z., Li, S., & Zhou, F. (2018, October). An Effective Multi-Classification Method for NHL Pathological Images. In *2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC)* (pp. 763-768). IEEE.

Koçyiğit, İ., Kaynar, L. & Çetin, M. (2008). Hematopoetik Kök Hücre Biyolojisi. Türkiye Klinikleri Hematology-Special Topics, 1(2), 16-22

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine Learning Applications in Cancer Prognosis and Prediction. *Computational And Structural Biotechnology Journal*, *13*, 8-17.

Madabhushi, A., & Lee, G. (2016). Image Analysis and Machine Learning in Digital Pathology: Challenges and Opportunities. pp. 170-175

Nabizadeh, N., & Kubat, M. (2015). Brain Tumors Detection and Segmentation in MR Images: Gabor Wavelet vs. Statistical Features. *Computers & Electrical Engineering*, *45*, 286-301.

Nagar, H., Dabas, C., & Gupta, J. P. (2017). Navie Bayes and K-Means Hybrid Analysis for Extracting Extremist Tweets. International Journal of Control Theory and Application, 10(4), 209–221.

Nguyen, T. T., & Armitage, G. J. (2008). A Survey of Techniques for Internet Traffic Classification Using Machine Learning. IEEE Communications Surveys and Tutorials, 10(1-4), 56-76.

Peterson, L. (2009). K-nearest neighbor. Scholarpedia, 4(2), 1883. https://doi.org/10.4249/scholarpedia.1883

Sezgin, E. (2007). Patolojik Görüntülerin Bilgisayarlı Analiz Programı ile Değerlendirilmesi (Doctoral Dissertation, Selçuk Üniversitesi Fen Bilimleri Enstitüsü).

Türkoğlu, İ., Şengür, A., & Toraman, S. (2013) Tıbbi Görüntülerden İstenen Bir Örüntünün Ayrıştırılması. Yıldız Teknik Üniversitesi DSpace Kurumsal Arşiv

Yıldız, O., Bilge, H. Ş., Akcayol, M. A., & Güler, İ. (2012). Meme Kanseri Sınıflandırması İçin Veri Füzyonu Ve Genetik Algoritma Tabanlı Gen Seçimi. *Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi*, *27*(3).

Yörükoğlu, K., Usubütün, A., Doğan, Ö., Önal, B., & Aydın, Ö. (2009). Türkiye'de Patoloji Laboratuvarlarında Kalite Kontrol. *Türk Patoloji Dergisi*, *25*, 29-37.

Wang, P., Hu, X., Li, Y., Liu, Q., & Zhu, X. (2016). Automatic Cell Nuclei Segmentation And Classification Of Breast Cancer Histopathology Images. *Signal Processing*, *122*, 1-13.

Wang, P. W., & Lin, C. J. (2014). Support vector machines. In Data Classification: Algorithms and Applications (pp. 187–204). https://doi.org/10.1201/b17320