

Makine Öğrenmesi ile Bitki Hastalıklarının Tespiti

Detection of Plant Diseases by Machine Learning

Umut Barış KORKUT ve Ömer Berke GÖKTÜRK

Bilgisayar Mühendisliği Bölümü

Gazi Üniversitesi

Ankara, Türkiye

ubk.94@gmail.com, berkegokturk94@gmail.com

Oktay YILDIZ

Bilgisayar Mühendisliği Bölümü

Gazi Üniversitesi

Ankara, Türkiye

oyildiz@gazi.edu.tr

Özetçe—Bu çalışmada görüntü işleme ve makine öğrenmesi yöntemleri kullanılarak, bitki hastalıklarının otomatik tespiti gerçekleştirilmiştir. Bitki hastalıklarının zamanında ve doğru tespiti ürün kalitesi ve rekolte için önemlidir. Erken tanı ve müdahale, bitki hastalıklarından doğan maliyeti düşürmeyi ve gereksiz ilaç kullanımını azaltmayı sağlayabilir. Gerçekleştirilen bu çalışmada, farklı bitki türlerine ait yaprak görselleri toplanmış, öğrenme aktarımı yöntemi kullanılarak görsellerden özellik çıkarımı yapılmıştır. Çeşitli makine öğrenmesi yöntemleri ile önerilen model başarısı %94 elde edilmiştir.

Anahtar Kelimeler — bitkilerde hastalık tespiti; görüntü işleme; makine öğrenmesi; özellik çıkarımı; sınıflandırma.

Abstract—In this study, automatic detection of plant diseases was performed by using image processing and machine learning methods. Timely and accurate detection of plant diseases is important for crop quality and yield. Early diagnosis and intervention can reduce the cost of plant diseases and reduce unnecessary drug use. In this study, leaf images of different plant species were collected and feature extraction was done from the images using the transfer learning method. With various machine learning methods, the proposed model achieved 94 % success.

Keywords — disease detection in plants; image processing; machine learning; feature extraction; classification.

I. GİRİŞ

Tarım alanında her yıl bitki hastalıkları sebebiyle büyük zayıatlar yaşanmaktadır [4]. Hastalıkların erken tespitiyle, bu zayıatların önüne geçilebilir. Ancak bu konudaki güncel çözümler; maliyetli, sağlıksız ve başarı oranı düşüktür. Yanlış tespitlerden dolayı, her yıl ciddi miktarlarda ilaç kullanılır. Bu durum, hem ekolojik dengeyi bozar hem de maliyet oluşturur.

Gerçekleştirilen bu çalışmada, bitki hastalıklarının tespiti için; makine öğrenmesi tekniklerini kullanan, düşük maliyetli, sağlıklı ve yüksek başarı oranına sahip bir Destek Vektör Makinesi (SVM - Support Vector Machine) sunulur.

Bu çalışmada öncelikle, hastalıklı ve sağlıklı bitki görselleri elde edilir. Görseller toplanırken şekil-1’de ve şekil-2’de gösterildiği gibi arka planın beyaz olması, görselde yalnızca bir yaprak bulunması, ışıklandırmanın ve çözünürlüğün iyi olması gibi bazı kriterler göz önünde bulundurulur.

Görsellerden kriterlere uymayanları, görüntü işleme yöntemleriyle, uygun hale getirilerek veri seti oluşturulur. Oluşturulan veri setinden, makine öğrenmesi için gerekli özellikler çıkarılır. Özellik çıkarımı için, öğrenme aktarımı yöntemi kullanılır. Çıkarılan özellikler kullanılarak, makine öğrenmesi algoritmaları uygulanır. Böylece hastalıklı ve sağlıklı olmak üzere iki sınıf içeren sınıflandırma işlemi yapılır.

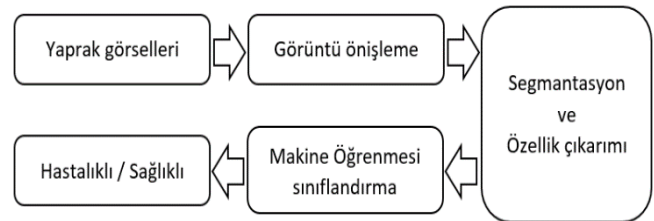
Literatürde makine öğrenmesi algoritmaları pek çok sınıflandırma problemi çözümünde kullanılmaktadır [1,2,3]. Aynı şekilde bitki hastalıklarının tespitinde de yaygın olarak kullanılmaktadır. Bashish ve arkadaşları, yapraktaki hastalıkların tespiti için K-means algoritmasını önermişlerdir. Aynı ekip, yaprak doku özelliklerinin sınıflandırılması için CCM’yi, hastalıkları altı kategoride sınıflandırmak için BPNN’yi önermişlerdir [5].

Camargo ve Smith, bitkilerin ya da bitkinin yapraklarının RGB (Red Green Blue) renklerini; H, I3a, ve I3b’ye dönüştürmeyi tartışmışlardır. Dönüştürülen görüntü, daha sonra histogram kullanılarak analiz edilen yoğunluk dağılımlarına göre bölünür. Çıkarılan bölge; hedef bölgenin bir parçası olmayan piksel bölgesi kaldırılarak işlenir. Daha sonra her düğümün ve pikselin bütün komşuları incelenir [5].

Sannakki ve arkadaşları, hastalık tespiti ve derecelendirmesi için, K-ortalama kümeleme (KNN-k Nearest Neighborhood) algoritmasını önermişlerdir.

Al-Hiary ve ekibi, özellik çıkarımından önce yeşil piksellerin maskelenmesini önermiş ve daha doğru bir sınıflandırma elde etmişlerdir.

Önerilen model yapısı Şekil 1.1.’de gösterildiği gibidir.



Şekil 1.1. Akış diyagramı

II. ARAÇ VE YÖNTEMLER

Araştırmalar sonucu, bitki hastalıklarının genellikle yapraklarda belirti gösterdiği tespit edilmiş ve bitkinin yaprak görselinin kullanılmasına karar verilmiştir.

Görselde birden fazla yaprağın olmasının ve arka planın karışık olmasının sınıflandırma başarısını düşürdüğü gözlemlenmiştir. Temin edilecek görsellerin bir takım kriterlere uyması gerekir. Bu kriterler: görselde tek yaprak olması, arka planın beyaz ve sade olması, iyi aydınlatılmış bir ortam, yeterli çözünürlük şeklinde sıralanabilir. Bu kriterlere uyan 81 hastalıklı ve 50 sağlıklı olmak üzere 131 yaprak görseli toplanmıştır. Görseller toplanırken, belli bir yaprak türü üzerinde durulmamıştır. Sınıflandırmanın amacı, hastalık olup olmadığını bulmaktır. Ancak bu çalışmada, yeşil yapraklı bitkiler ele alındığından dolayı, görseller yeşil yapraklı bitkilere aittir. Ancak özel bir bitki türü üzerinde durulmamıştır. Hastalıklı bitki görselleri içinse, yeşil yapraklı bitkilerin, yapraklarında beliren hastalıklar ele alınmıştır. Bu konuda, özel bir hastalık türü üzerinde durulmamıştır. Hastalık belirtilerinin yapraklarda olduğu bütün görseller, sınıflandırma için elverişlidir. Şekil 2.1 ve 2.2.'de örnek yaprak görselleri görülmektedir.



Şekil 2.1. Sağlıklı bitkinin yaprak görsellerine örnek



Şekil 2.2. Hasta bitkinin yaprak görsellerine örnek

Toplanan görsellerden, %75'i eğitim ve %25'i test olmak üzere rastgele ayrılmıştır. Öğrenme aktarımı metoduyla özellik çıkarımları yapılmıştır. Öğrenme aktarımı metodunda, google tarafından geliştirilen inception v3 ön eğitim modeli kullanılmıştır. Bu model, 1000 sınıf ve 1 milyon örnek içeren bir veri setinin "image.net" üzerinde eğitilmesiyle geliştirilmiştir. Model, girdi olarak aldığı "jpg" formatındaki görsellerden, 2048 adet özniteliği vektörel öznitelik olarak çıkarmaktadır. Model, öznitelik çıkarırken, düşük seviyedeki kenar saptamadan, yüksek seviyedeki renk ve cisim saptamaya kadar birçok yöntem kullanmaktadır. Yöntemler sonucu çıkarılan her resim için 2048 özniteliğin ne olduğunun, çalışma için bir önemi yoktur. Ancak bu öznitelikler, görsel grafiklerle incelenerek, verimli özniteliklerin çıkarılma işlemi yapılabilmektedir. Kısıtlı veri setlerinde son derece yüksek başarı oranıyla çalışmaktadır.

Özellik çıkarımları yapılan veri setine, makine öğrenmesi algoritmaları; KNN, Rassal Orman, Naive Bayes, Destek Vektör Makinesi (SVM – Support Vector Machine), Sinir Ağları ve Lojistik Regresyon uygulanmıştır.

KNN algoritması, her veriyi, uzayda bir nokta olarak düşünür ve her özelliği, uzayın bir boyutu olarak ele alır. Verilerin birbirlerine uzaklığını, çeşitli uzaklık fonksiyonlarıyla hesaplayarak sınıflandırma yapar. Devamlı verilerde kullanılan üç temel uzaklık fonksiyonu vardır. Bunlardan en yaygın kullanılanları, öklid ve manhattan uzaklık fonksiyonlarıdır. Devamlı olmayan, yani kategorik verilerde ise hamming uzaklık fonksiyonu kullanılır.

$$\sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (1)$$

$$\sum_{i=1}^k |x_i - y_i| \quad (2)$$

$$D_H = \sum_{i=1}^k |x_i - y_i| \quad (3)$$

$$x = y \rightarrow D = 0$$

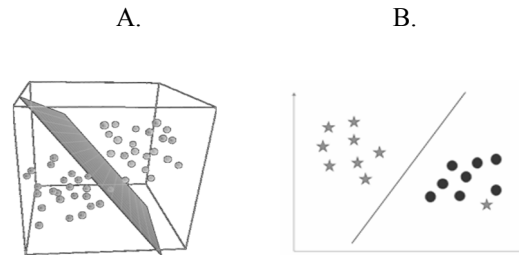
$$x \neq y \rightarrow D = 1$$

Denklem (1)- öklid - ve (2)'de – manhattan - bütün noktalar için uzaklık değerleri hesaplanır. Sınıflandırma için, en yakın nokta baz alınır. Denklem (3) serisinde, manhattan fonksiyonundan farklı olarak, işlemin sonucuna göre uzaklık 0 ya da 1 alınır. X ve y değerleri, iki noktayı temsil eder.

Rassal orman algoritması, makine öğrenimi modelinin eğitimi sırasında verilerin özelliklerine bakarak çok sayıda karar ağacı oluşturur. Karar ağaçlarının ortalamasını alarak, verinin hangi sınıfa ait olduğunu belirler.

Naive Bayes sınıflandırma yöntemi, özellikler arasındaki bağıntılara bakarak Bayes teoreminin uygulanmasına dayanır. Bayes teoremi koşullu olasılık hesaplaması yapar. Naive Bayes yönteminde girilen verinin olasılığı en yüksek olan sınıfa ait olduğu tahmin edilir. Veri setinin boyutunun çok olduğu durumlarda etkilidir.

Destek Vektör Makinesi, sınıflandırma için uzayda noktalar halinde düşünülen verileri ikiye bölerek ayırır. Uzayın her boyutu bir özellik olarak düşünülür. İki özellik için düzlem doğru ile ayrılırken, üç özellik için uzay düzlem ile ayrılır. Uzayı ikiye bölerken farklı sınıfların birbirine olan en uzak noktaları esas alınır. Şekil 2.3.A'da SVM algoritmasının, uzayı düzlemle, şekil 2.3.B'de ise düzlemi doğruyla böldüğü görülmektedir.



Şekil 2.3. A) & B) SVM'nin uzayı bölmesi

Sinir ağı, çok sayıda gizli katman içeren makine öğrenme yöntemidir. İnsan beynini taklit ederek yapılan bir yöntemdir. Yapay sinir ağları tıpkı insan beynindeki gibi birbirlerine çeşitli şekillerde bağlanmışlardır. Bu ağlar öğrenme, hafızaya alma ve veriler arasında ilişki kurma yetilerine sahiptir. Öğrenme, verilerin özelliklerinin bağlantı ağırlıklarını tekrar tekrar

ayarlayarak gerçekleřir. Derin öğrenme algoritmaları çok sayıda veri ile başarılı sonuçlar verir.

Lojistik regresyon, SVM gibi lineer bir metottur. Sınıflandırma için uzayda noktalar halinde düşünölen verileri ikiye bölerek ayırır. Uzayın her boyutu bir özellik olarak düşünöölür. Uzayı ikiye bölerken, her sınıfa ait olma olasılığını hesaplar.

$$Kesinlik = \frac{tp}{tp+fp} \quad (4)$$

$$Hassasiyet = \frac{tp}{tp+fn} \quad (5)$$

Denklem (4)'te kesinlik değeri, doğru tahmin edilen pozitif sınıfın, bütün pozitif tahminlere oranını gösterir. Böylece, pozitif olarak yapılan tahminlerin başarısı incelenmiş olur.

Denklem (5)'te hassasiyet değeri, doğru tahmin edilen pozitif sınıfın, bütün pozitif sınıflara oranını gösterir. Pozitif sınıf için sınıflandırma başarısını inceleme fırsatı verir.

III. SONUÇLAR VE TARTIřMA

KNN algoritması en yüksek sınıflandırma başarısına komřu sayısı 9 olduėunda ulaşmıştır. %78'lik bir başarı söz konusudur. řekil 3.1'de ve Tablo 1'de, KNN algoritmasının performansı gösterilmiştir.

		Tahmin		Σ
		0	1	
Gerçek	0	62	19	81
	1	12	38	50
Σ		74	57	131

řekil 3.1. KNN algoritması karmařıklık matrisi

řekil 3.1'de görölen 0, hastalıklı bitkileri, 1 ise saėlıklı bitkileri temsil eder.

TABLO 1. KNN kesinlik – hassasiyet değeri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.58	0.78	0.67	9
1	0.90	0.79	0.84	24
Ortalama/Toplam	0.82	0.79	0.80	33

Rassal orman algoritması en yüksek sınıflandırma başarısına estimator sayısı 11 olduėunda ulaşmıştır. %82'lik başarı söz konusudur. řekil 3.2'de ve Tablo 2'de Rassal Orman algoritmasının performansı gösterilmiştir.

		Tahmin		Σ
		0	1	
Gerçek	0	76	5	81
	1	24	26	50
Σ		100	31	131

řekil 3.2. Rassal orman algoritması karmařıklık matrisi

TABLO 2. Rassal orman kesinlik – hassasiyet değeri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.67	0.67	0.67	9
1	0.88	0.88	0.88	24
Ortalama/Toplam	0.82	0.82	0.82	33

SVM algoritması en yüksek sınıflandırma başarısına c parametresi 10, gamma parametresi 0.001 olduėunda ulaşmıştır. %94'lük bir başarı söz konusudur. řekil 3.3'de ve Tablo 3'de SVM algoritmasının performansı gösterilmiştir.

		Tahmin		Σ
		0	1	
Gerçek	0	70	11	81
	1	10	40	50
Σ		80	51	131

řekil 3.3. SVM algoritması karmařıklık matrisi

TABLO 3. SVM algoritması kesinlik – hassasiyet değeri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.89	0.89	0.89	9
1	0.96	0.96	0.96	24
Ortalama/Toplam	0.94	0.94	0.94	33

Lojistik regresyon algoritması en yüksek sınıflandırma başarısına c parametresi 1 olduėunda ulaşmıştır. %94'lük bir başarı söz konusudur. řekil 3.4'de ve Tablo 4'de Lojistik Regresyon algoritmasının performansı gösterilmiştir.

		Tahmin		Σ
		0	1	
Gerçek	0	74	7	81
	1	9	41	50
Σ		83	48	131

řekil 3.4. Lojistik regresyon algoritması karmařıklık matrisi

TABLO 4. Lojistik regresyon kesinlik – hassasiyet değeri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.89	0.89	0.89	9
1	0.96	0.96	0.96	24
Ortalama/Toplam	0.94	0.94	0.94	33

Naive bayes algoritmasının sınıflandırma başarısı %88 olmuřtur. řekil 3.5'de ve Tablo 5'de Naive Bayes algoritmasının performansı gösterilmiştir.

		Tahmin		
		0	1	Σ
Gerçek	0	60	21	81
	1	4	46	50
Σ		64	67	131

Şekil 3.5. Naive bayes algoritması karmaşıklık matrisi

TABLO 5. Naive bayes kesinlik – hassasiyet değerleri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.73	0.89	0.80	9
1	0.95	0.88	0.91	24
Ortalama/Toplam	0.89	0.88	0.88	33

Sinir ağı algoritması en yüksek sınıflandırma başarısına, maksimum 400 iterasyon ve 40 gizli katman olduğunda ulaşılmıştır. Sinir ağı ile %91’lik bir başarı söz konusudur. Tablo 6’da Sinir Ağları algoritmasının performansı gösterilmiştir.

TABLO 6. Sinir ağları kesinlik – hassasiyet değerleri

	Kesinlik	Hassasiyet	F1 Sonucu	Destek
0	0.80	0.89	0.84	9
1	0.96	0.92	0.94	24
Ortalama/Toplam	0.91	0.91	0.91	33

IV. SONUÇ VE GELECEK KAPSAM

Bu çalışmada, gözetimli makine öğrenmesi teknikleri yaprak görsellerini hastalıklı ve sağlıklı olmak üzere iki sınıfa ayırmak için kullanılmıştır. Toplanan yaprak görsellerden, yapay sinir ağları algoritması ile önceden eğitilmiş inceptionv3 modeli kullanılarak öğrenim aktarması yöntemiyle özellik çıkarımları yapılmıştır. Sınıflandırma sonuçlarının gösterdiği üzere SVM ve lojistik regresyon algoritmaları iyi sonuçlar vermiştir. Bu çalışmada kullanılan yöntemler hastalık varlığının yanı sıra hastalık türünü de tespit etmek için kullanılabilir. Sınıflandırma başarı oranını yükseltmek için veri setindeki görsel sayısı artırılabilir. Yapay sinir ağları yöntemi genelde veri seti büyüdüğünde diğer algoritmalarla oranla daha iyi sonuçlar verir.

KAYNAKLAR

- [1] Berikol, G., Kula, S., Yıldız, O., "Machine Learning Techniques in Diagnosis of Pulmonary Embolism", *Journal of Clinical and Analytical Medicine*, 2015
- [2] Arslan, A., Yıldız, O., "Automated Auscultative Diagnosis System for Evaluation of Phonocardiogram Signals Associated with Heart Murmur Diseases", *Gazi University Journal of Science*, 31(1): 112-124, 2018
- [3] Karakış, Rukiye, İnan Güler, and Ali Hakan Işık. "Feature selection in pulmonary function test data with machine learning methods." *Signal Processing and Communications Applications Conference (SIU)*, 2013 21st. IEEE, 2013.
- [4] James W. C., "Assessment of Plant Diseases and Losses", *Annual Review of Phytopathology*, 27-48, 1974.
- [5] Revathi P., Hemalatha M., "Classification of Cotton Leaf Spot Diseases Using Image Processing Edge Detection Techniques", *International Conference on Emerging Trends in Science Engineering and Technology (INCOSSET)*, doi:10.1109/incoset.2012.6513900
- [6] C. C. Tucker and S. Chakraborty. "Quantitative assessment of lesion characteristics and disease severity using digital image processing", *J. Phytopathology*, 1997, vol.145, pp.273–278.
- [7] S. Ahmad, J. F. Reid, M. R. Paulsen, and J. B. Sinclair, "Color classifier for symptomatic soybean seeds using image processing PlantDisease", 1999, vol. 83, pp.320–327.
- [8] J. C. Lai, S. K. Li, B. Ming, N. Wang, K. R. Wang, R. Z. Xie, and S. J. Gao., "Advances in research on computer-vision diagnosis of cropdiseases. Scientia Agricultura Sinica", 2009, vol. 42, pp.1215–1221.
- [9] C. H. Bock, A. Z. Cook, P. E. Parker, and T. R. Gottwald., "Automated image analysis of the severity of foliar citrus canker symptoms. PlantDisease", 2009, vol.93, pp. 660–665.
- [10] C. H. Bock, G. H. Poole, P. E. Parker, and T. R. Gottwald., "Plant disease severity estimated visually, by digital photography and imageanalysis, and by hyperspectral imaging", *Critical Reviews in Plant Sciences*, March 2010, vol. 29, pp. 59–107.
- [11] H. Al-Hiary, S. Bani-Ahmad, M. Reyalat, M. Braik, and Z. ALRahamneh, "Fast and accurate detection and classification of plantdiseases", *International Journal of Computer Applications*, 2001, vol. 17, pp.31–38.
- [12] D. W. Zhang and J. Wang, "Design on image features recognition system of cucumber downy mildew based on BP algorithm", *Journal of Shenyang Jianzhu University (Natural Science)*, May 2009, vol. 25, pp. 574–578.
- [13] D. T. Zhao, Y. H. Chai, and C. L. Zhang, "Inspection of soybean frog-eye spot based on image procession", *Journal of Northeast Agricultural University*, vol. 41, pp. 119–124, April 2010.
- [14] Y. L. Cui, P. F. Cheng, X. Z. Dong, Z. H. Liu, and S. X. Wang, "Image processing and extracting color features of greenhouse diseased leaf", *Transactions of the CSAE*. vol. 21(Supp.), pp. 32–35, December 2005.