

# DEEP LEARNING İLE BİTKİ YAPRAKLARINDAN HASTALIK SINIFLANDIRILMASI

Elif Nur KÜÇÜK

Behlül SARIKAYA

Danışman

Prof. Dr. Aybars UĞUR

91200000203@ogrenci.ege.edu.tr 91200000200@ogrenci.ege.edu.tr aybars.ugur@ege.edu.tr

Ege Üniversitesi Fen Bilimleri Enstitüsü  
Bilgisayar Mühendisliği Anabilim Dalı

## Özet

Bitki hastalıklarının doğru tespiti, tedavi amaçlı yanlış kullanılacak kimyasallara karşı direnç oluşması, ilaçlama maliyetinin artması, hastalığın önlenememesiyle oluşabilecek salgınlara ve buna bağlı olarak ürün verimliliğinin düşmesinin önlenmesi için çok önemlidir. Hastalık çeşidinin fazla olması nedeniyle hastalıkları tespit etme ve sınıflandırma uzmanlar için bile zaman alıcı ve zordur. Bu zorluğu ortadan kaldırmak için çalışmamızda bitki yaprak hastalıklarını sınıflandırmak için derin öğrenme tabanlı bir yöntem önerilmektedir.

Çalışmamızda 3642 tane yaprak görüntüsünden oluşan [Kaggle Plant Pathology 2020-FGVC7](#) veri seti kullanılmıştır. Veri seti paslı, kabuklanmış, bir yaprak üzerinde birden fazla hastalığın bulunabileceği kombinasyon ve sağlıklı elma ağacı yaprağı görüntülerinden oluşmaktadır.

**Anahtar Kelimeler:** plant leaf diseases, Deep learning, Classification,

## Giriş

Bitki hastalıklarının hızlı bir şekilde tanımlanması, gerekli altyapı eksikliğinden dolayı dünyanın birçok yerinde zor olmaya devam etmektedir. Güncel teknolojilerle görüntüler, yüksek çözünürlüklü akıllı telefon kameraları aracılığıyla dünya çapında mevcuttur. Bu nedenle, tarlada alınan bitki görüntülerine dayanan düşük maliyetli, genel amaçlı, otomatik ve doğru bitki hastalığı sınıflandırma algoritmalarının geliştirilmesi, bitki hastalıkları ve kontrol uygulamalarını dünya çapında kullanılabilir hale getirecektir. Bu sorunu çözmek önemlidir çünkü bitki hastalıklarını erken teşhis etmek her yıl tonlarca tarımsal ürün tasarrufu sağlayabilir. Bu sadece açlığı azaltarak genel nüfusa değil, hak ettikleri hasadı almalarını sağlayarak çiftçilere de fayda sağlayacaktır.

Yaprak görüntüleri hastalıkların tespiti için bitki üzerindeki yoğunluğu ve kalıcılık süresi açısından en önemli bilgi kaynaklarından biridir. Bu nedenle bitkiler üzerinde gerçekleştirilen pek çok çalışmada hastalığın yaprak üzerinde oluşturduğu bozulma belirleyici olarak kullanılmıştır.

## Literatür İncelemesi

Araştırmalarda bir grup çalışmacı bilgisayarlı görme teknikleri üzerinden yapraklar üzerindeki hastalıkların tespit edilmesi için çalıştılar. Yapılan bilimsel araştırmalara bakıldığında, bir çok görüntü işleme methodunun ve yapay zeka tekniğinin kullanıldığını görebilmek mümkündür. Bunlardan bazıları şunlardır; [Aakanksha Rastogi](#) ve arkadaşlarının hazırladığı yapay sinir ağı ve K-means barındıran iki aşamalı yaprak hastalığı derece tespiti[1], [Pranjali B. Padol](#)'ın üzüm yapraklarındaki hastalıkların SVM sınıflandırma tekniği kullanılarak saptaması[2], farklı bir teknikle üzüm yapraklarındaki hastalıkların tespitini yapan [A. Meunkaewjinda](#) ve arkadaşları ise genetik algoritma ve destek vektör makinesi kullanmışlardır[3]. Reinald Adrian D. L. Pugoy ve diğerlerinin pirinç yaprakları hastalıklarını K-means algoritması ile sınıflandırması[4], [Gittaly Dhingra](#) ve arkadaşlarının yapmış oldukları bitki hastalığı sınıflandırma[5] Bu çalışmalarda derin öğrenme yönteminin görüntü işleme yöntemlerine göre daha yüksek doğruluk sağladığı ve daha iyi performans sergiledikleri belirtilmiştir.

Son yıllarda yapılan çalışmalarda bitkilerde yaprak hastalıklarının tespiti ve sınıflandırılması için derin öğrenme kapsamlı bir şekilde araştırılmıştır.[6],[7],[8],[9],[10] [Melike Sardogan](#) ve arkadaşlarının domates yapraklarındaki hastalıkları CNN ve LVQ yöntemlerine dayalı çalışmalarında 500 domates yaprağından 4 farklı hastalık sınıfını etkili bir şekilde tanıdığı belirtilmektedir.[11] [Jihen Amara](#) ve arkadaşlarının muz yaprakları üzerinde yaptıkları çalışmada hastalıkların sınıflandırılması için LeNet mimarisi Convolutional Neural Network olarak kullanılmıştır.[12] [Monu](#) ve arkadaşları dolmalık

biber bitkisi görüntü veri seti ile donatılmış bir CNN kullanarak büyük bir başarı elde etmişlerdir.[13]

Bu bilgilere bakılarak üzüm, elma, pirinç, domates ve kayısı gibi günlük yaşamımızda oldukça fazla tükettiğimiz meyvelerin hastalık tespitlerinde oldukça fazla çalışma yapıldığını söyleyebiliriz.

Ayrıca bir çok platform üzerinde gerçekleştirilen yarışmalarda ise yine domates, pirinç elma gibi meyvelerdeki hastalıkların tespiti üzerine yarışmalar düzenlenmektedir.

## Geliştirme Aşaması

Projenin geliştirme aşaması Google Colab üzerinde Python 3 ile Tesla K80 GPU kullanılarak yapılmaktadır. Görüntülerin işlenebilmesi için Python'ın OpenCv,derin öğrenme için TensorFlow ve Keras kütüphaneleri kullanılmıştır.

## Veri Hazırlama

Veri seti içerisinde train.csv,test.csv,submission.csv dosyalarının ve görüntülerin bulunduğu images klasörü ile toplamda 3645 adet veri bulunmaktadır.İlk olarak Google Drive a kayıtlı olan bu verilerimiz proje dosyasına yüklenir.Yüklenen verileri ve görüntüyü modelin gereksinimlerine göre ön işleme yapılması gerekir.Train.csv dosyasında görüntülerin isimleri ve ait oldukları hastalık sınıfı hakkında bilgi verilmektedir.

Resim boyutları 224x224 olarak ayarlanır.Train ve test olarak 1821 görüntü bulunduran toplamda iki görüntü kümesinden oluşan ve her biri 3642 tane elma yaprağı görüntüsünden oluşan veri setimizde sağlıklı, kabuklanmış,paslı ve birden fazla hastalığın bulunabileceği 4 tür görüntü bulunmaktadır.Modelin eğitilmesi için kullanılan Train veri setimizdeki bu 4 türün dağılım sayıları eşit değildir.1821 görüntüden 515 tanesi sağlıklı, 622 tanesi kabuklanmış, 592

tanesi paslı ve 91 tanesi birden fazla hastalığın bulunduğu sınıflara aittir. Sınıflandırma yaparken sınıfların bu şekilde eşit olarak dağılmadığı, yani her sınıf için yaklaşık aynı sayıda verinin bulunmadığı veri kümelerine dengesiz veri kümeleri denir. Dengesiz veri kümelerinde sınıflandırıcılar, genel hata oranını minimize etmeye çalışırken ağırlıklı olan sınıfları kategorize ederek, azınlık sınıfını (en az sayıda veri bulunan sınıfı) göz ardı edebilmektedir. Fakat asıl hedeflerden biri azınlık sınıfının doğru sınıflandırılması olduğu için dengesiz veri kümemizi daha dengeli hale getirmek için bazı düzenlemeler yapılmıştır. SMOTE (Synthetic Minority Oversampling Technique) yönteminde dengesiz veri dağılımının bulunduğu sınıf türü yapay olarak çoğaltılır ve dengelenme sağlanır. Yeni örnekleri sentezlemek için en yaygın kullanılan yaklaşımlardan biridir. Bu teknik [Nitesh Chawla](#) ve arkadaşlarının 2002 yılında yazdıkları Synthetic Minority Over-Sampling Technique makalesinden adını alır. SMOTE, özellik uzayında yakın olan örnekleri seçerek, özellik uzayındaki örnekler arasında bir çizgi çizer ve bu çizgi boyunca bir noktada yeni bir örnek çizerek çalışır. Azınlık sınıftan rastgele bir örnek seçilir. Daha sonra K tane komşusu bulunur. (Genellikle K=5 tane) Oluşturulan bu komşulardan biri seçilir ve özellik uzayındaki örnek ile rastgele seçilen komşu arasında herhangi bir noktada sentetik bir örnek oluşturulur. Sentetik örnekler seçilen iki örneğin dış bükey bir kombinasyonu olarak oluşturulur. Bu prosedür, azınlık sınıfı için gerektiği kadar sentetik örnek oluşturmak için kullanılabilir.

### Modelin Kurulması

Araştırmamızda görüntülerin sınıflandırılması için evrişimsel sinir ağları kullanılmıştır. CNN(Convolutional Neural Network): Bir giriş

görüntüsünü alıp, görüntüdeki çeşitli yönlere, nesnelere önem atayabilen ve birini diğerinden ayırt edebilen bir derin öğrenme algoritmasıdır. CNN görüntüyü çeşitli katmanlarda işleyerek sınıflandırılmasını sağlar.

Görüntü tanıma, sınıflandırma ve segmentasyon için yaygın olarak kullanılmaktadır. CNN modeli taslak olarak Convolutional Layer (Evrişimsel Katman), Pooling Layer (Havuzlama Katmanı) ve Fully Connected (Tam Bağlı Katman) yapılarından oluşmaktadır. Genellikle overfitting engellenmesi için çeşitli regülerizasyon yöntemleri (Dropout, Early Stopping) ile aşırı öğrenme önlenmektedir.

**Convolutional Layer:** Bu katman CNN'nin en temel katmanıdır. Görüntü üzerinde belirli bir filtre dolaştırılır ve bu filtre bir önceki katmandan gelen görüntülere evrişim işlemi uygulayarak çıktı elde edilmesini sağlar. CNN modelinin yapıtaşlarıdır. Bu katman sayesinde görüntüler üzerinde feature detection gerçekleştirilmiştir. Çıkarım yapılan özellikler üzerine aktivasyon fonksiyonu olarak relu uygulanmıştır.

**Pooling Layer :** Evrişim katmanından gelen görüntüler ve özellikler üzerinde Pooling adı verilen Havuzlama katmanı bulunur. Bu katmanda modelin karmaşıklığı, performansı düşürmeden parametre sayısını düşürerek azaltılır. Pooling yapısı olarak Max Pooling kullanılmıştır. Bu şekilde kernel boyutlarında en büyük sayı alınmıştır. Bu işlem ile sinir ağının doğru karar vermesi için yeterli bilgiyi içeren daha küçük çıktılar kullanılmış olur. Pooling gösterimin kayma boyutunu, ağ içindeki parametreleri ve hesaplama sayısını azaltmak içindir. Bu sayede ağdaki uyumsuzluk kontrol edilmiş olur.

**Fully Connected Layer :** Mimarinin son katmanıdır.

Bu katman için öncelikle veriler tek girişli matrise dönüştürülecektir. Fully Connected Katmanındaki sinir ağları birbirine tamamen bağlanmaktadır. Her bir nöronun ayrı bir weight değeri vardır.Burada oluşturulacak ağ sadece bu tip katmanlardan oluşacak.Bunu sağlayacak işlem ise Flatten sayesinde gerçekleştirilir. Daha sonra hazırlanan bu veriler Dense ile yoğunluk filtresi verilerek bağdaştırılır. Bu katmanda girişte alınan input veriler Fully Connected Layer da kullanılmak üzere hazırlanmaktadır.

**Dropout :** Fully connected katmanlarda belli eşik değerin altındaki düğümlerin seyreltilmesinin başarımı arttırdığı gözlenmiştir. Yani zayıf bilgilerin unutulması öğrenimi arttırdığı gözlenmiştir. Bu işlemi 0 ile 1 değeri arasında belirtilen olasılık değeri ile düğümleri eksilterek uygular.

**Flattening Layer :** Bu katmanın görevi, Fully Connected Layer'a gönderilecek verileri hazırlamaktır. Yapılan işlem ise convolutional pooling katmanından gelen matrislerin tek boyutlu matrise çevrilmesidir.

Bu çalışmada kurulan model için çok sayıda katman kullanımına uygun olması nedeni ile Sequential modeli tercih edildi. Toplamda 35 katman kullanıldı. Bunlar; 8 adet Convalization katmanı, 11 adet Normalization katmanı, 4 adet Pooling katmanı ve 7 adet Regularization katmanı(dropout), çekirdek katmanı olarak ise Dense katmanı kullanıldı. Modelin gereksiz yere fazla adımlarda çalıştırılmasının önüne geçebilmek için ise erken sonlandırma işlemi gerçekleştirildi. Erken sonlandırma işlemi için ise en düşük öğrenme katsayısı olarak 0.000001 değeri belirlendi. Modelin

çalıştırılma aşamasına geçildiğinde epoch değeri olarak 100, batch size değeri olarak da 24 değeri seçildi. Her ne kadar 100 epoch değeri seçilmiş olsa da erken sonlanma özelliğinden dolayı, modelin 62. epoch tan sonra sonlandırıldığı gözlemlendi.Eğitim süresi 20 dakika olarak kaydedildi.

## Sonuç ve Tartışma

Kurulan model 62 epochta çalıştırıldığında accuracy 0.9034 olarak hesaplanmıştır.Fakat dengesiz veri kümelerinin değerlendirilmesinde accuracy bize net doğruluk oranını yansıtmayacağından daha iyi bir sonuç hesaplanması için ROC eğrisi kullanılmıştır.Herhangi bir sınıflandırma modelinin performansını kontrol etmek için en önemli değerlendirme ölçütlerinden biridir. Özellikle dengesiz veri setlerinin bulunduğu durumlarda, makine öğrenmesi algoritmalarının performansını değerlendirmek için en yaygın kullanılan ölçütlerden biridir. Ve modelin tahmininde ne kadar iyi olduğunu açıklar.ROC farklı sınıflar için bir olasılık eğrisidir. Tipik bir ROC eğrisinde X ekseninde Yanlış Pozitif Oran (FPR) ve Y ekseninde Gerçek Pozitif Orana (TPR) vardır. Model değerlendirmesinde oluşturulan iki adet grafikten birincisinde X eksen 'loss', Y eksen, 'epoch' değerini gösterip epoch sayısının artması ile loss üzerindeki değişimi göstermektedir.İkinci grafikte ise X eksen 'accuracy', Y eksen 'epoch' sayısını gösterip accuracy oranının epoch sayısının artması ile değişimini göstermektedir.Epoch sayısı arttıkça loss azalmakta ve accuracy artmaktadır.Roc eğrisi ile değerlendirme sonucunda 0.97 olarak doğru sınıflandırma oranı bulunmuştur.

## Referanslar

- [1] Aakanksha Rastogi; Ritika Arora; Shanu Sharma; "Leaf disease detection and grading using computer vision technology & fuzzy logic." 2015 2nd International Conference on Signal Processing and Integrated Networks (SPIN)
- [2] Pranjali B. Padol; Anjali A. Yadav; "SVM classifier based grape leaf disease detection." 2016 Conference on Advances in Signal Processing (CASP).
- [3] A. Meunkaewjinda; P. Kumsawat; K. Attakitmongkol; A. Srikaew; "Grape leaf disease detection from color imagery using hybrid intelligent system." 2008 International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology, ECTI-CON.
- [4] Reinald Adrian D.L. Pugoy; Vladimir Y. Mariano; "Automated rice leaf disease detection using color image analysis." Proc. SPIE 8009, Third International Conference on Digital Image Processing (ICDIP 2011).
- [5] Gittaly Dhingra; Vinay Kumar; Hem Dutt Joshi; "Study of digital image processing techniques for leaf disease detection and classification." Multimedia Tools and Applications 77, 19951-20000 (2018)
- [6] Mehmet Emin Özgüven; Kemal Adem; "Automatic detection and classification of leaf spot disease in sugar beet using deep learning algorithms." Physica A: Statistical Mechanics and its Applications Volume 535, 1 December 2019, 122537
- [7] Malusi Sibiya; Mbuyu Sumnwanyambe; "A Computational Procedure for the Recognition and Classification of Maize Leaf Diseases Out of Healthy Leaves Using Convolutional Neural Networks." AgriEngineering 2019, 1(1), 119-131
- [8] Geetharamani G; Arun Pandian j; "Identification of plant leaf diseases using a nine-layer deep convolutional neural network." Computers & Electrical Engineering, Volume 78, September 2019, Pages 536.
- [9] Ramar Ahila Priyadharshini; Selvaraj Aravazhagan; Madakannu Arun; Annamalai Mirnalini; "Maize leaf disease classification using deep convolutional neural network." Neural Computing and Applications 31, 8887-8895 (2019)
- [10] V. Suma; R. Among Shetty; Rishab F. Tated; Sunku Rohan; Triveni S. Pujar; "CNN based Leaf Disease Identification and Remedy Recommendation System." 2019 International conference of Electronics, Communication and Aerospace Technology (ICECA)
- [11] Melika Sardoğan; Adem Tuncer; Yunus Özkan; "Plant Leaf Disease Detection and Classification Based on CNN with LVQ Algorithm." 2018 International Conference on Computer Science and Engineering (UBMK)
- [12] Amara Jihen; Bouaziz Bassem; Algergawy Alyased; "A deep Learning-based Approach for Banana Leaf Diseases Classification" Conference: Datenbanksysteme für Business, Technologie und Web (BTW 2017), 17. Fachtagung des GI-Fachbereichs „Datenbanken und Informationssysteme" (DBIS), 6.-10. März 2017, Stuttgart, Germany, Workshopband
- [13] Monu Bhagat; Dilip Kumar; Rehan Mahmood; Bidhudhendra Pati; "Bell Pepper Leaf Disease Classification Using CNN" 2nd International Conference on Data, Engineering and Applications (IDEA)-2020

SMOTE: Nitesh V. Chawla; Kevin W. Bowyer; Lawrence O. Hall; W. Philip Kegelmeyer; "SMOTE: Synthetic Minority Over-sampling Technique" Journal of Artificial Intelligence Research 16 (2002) 321-357