



OSTİM
TECHNICAL
UNIVERSITY
A N K A R A

**Klasik Makine Öğrenmesi ve Derin Öğrenme Yaklaşımları
Kullanılarak Bitki Yaprığı Hastalığı Sınıflandırması**

Ders: IYU 327 – Yapay Zeka Mühendisliği İş Yeri Uygulaması

Bölüm: Yapay Zeka Mühendisliği

Üniversite: Ostim Teknik Üniversitesi

Akademik Yıl: 2025–2026

Öğrenci: Omar A. M. Issa

Öğrenci Numarası: 220212901

Klasik Makine Öğrenmesi ve Derin Öğrenme Yaklaşımları Kullanılarak Bitki Yaprığı Hastalığı Sınıflandırması.....	1
Özet.....	3
1. Giriş ve Problem Tanımı	3
2. Veri Kümesi Tanımı	4
3. Yöntem.....	4
3.1 Veri Ön İşleme	4
3.2 Klasik Makine Öğrenmesi Taban Çizgi Modeli	5
3.3 Derin Öğrenme Modeli	5
4. Deneysel Tasarım.....	6
5. Değerlendirme Metrikleri	6
6. Sonuçlar	6
7. Tartışma ve Hata Analizi.....	7
8. Sayısal Yöntemler ve Optimizasyon Katkısı.....	7
9. Tekrar Üretilirlik ve Mühendislik Kalitesi.....	7
10. Sınırlılıklar ve Gelecek Çalışmalar.....	7
11. Sonuç.....	8

Özet

Bitki hastalıklarının erken ve doğru şekilde tespit edilmesi, tarımsal verim kayıplarının azaltılması ve ürün kalitesinin korunması açısından büyük önem taşımaktadır. Günümüzde bitki hastalıklarının teşhisi çoğunlukla uzmanlar tarafından manuel gözlem yoluyla yapılmaktadır. Bu yaklaşım zaman alıcı, öznel ve geniş tarım alanları için ölçeklenmesi zor bir yöntemdir. Bu raporda, bitki yapraklarından hastalık tespiti için uçtan uca bir görüntü tabanlı sınıflandırma çalışması sunulmaktadır.

Çalışma kapsamında, Histogram of Oriented Gradients (HOG) ve renk histogramlarına dayalı, Destek Vektör Makineleri (SVM) kullanan klasik bir makine öğrenmesi taban çizgi modeli ile ResNet-18 mimarisi kullanılarak oluşturulan transfer öğrenmeli bir derin öğrenme modeli karşılaştırılmıştır. Deneyler, PlantVillage veri kümesi üzerinde eğitim, doğrulama ve test bölümleri net bir şekilde tanımlanarak gerçekleştirilmiştir. Model performansları; doğruluk (accuracy), makro-F1 skoru, precision–recall analizleri ve karışıklık matrisleri gibi birden fazla metrik kullanılarak değerlendirilmiştir. Sonuçlar, derin öğrenme yaklaşımının sınıflar arasında daha dengeli ve yüksek performans sunduğunu, klasik yaklaşımın ise yorumlanabilirlik ve hesaplama verimliliği açısından önemli bir referans olduğunu göstermektedir. Çalışma boyunca tekrar üretilebilirlik ve deneysel tutarlılık ön planda tutulmuştur.

1. Giriş ve Problem Tanımı

Bitki hastalıkları, modern tarımda karşılaşılan en önemli problemlerden biridir ve erken aşamada tespit edilmediği takdirde ciddi ürün kayıplarına yol açabilmektedir. Birçok bitki hastalığı, yapraklar üzerinde renk değişimi, doku bozulması veya şekilsel anormallikler gibi görsel belirtilerle ortaya çıkmaktadır. Geleneksel olarak bu belirtilerin tespiti, uzman kişiler tarafından yapılan saha incelemelerine dayanmaktadır. Ancak bu süreç hem maliyetlidir hem de geniş alanlarda uygulanması zordur.

Yapay zeka bakış açısından bakıldığında, bitki yaprağı hastalık tespiti problemi denetimli çok sınıflı bir görüntü sınıflandırma problemi olarak ele alınabilir. Amaç, verilen bir yaprak görüntüsünün ilgili hastalık sınıfına veya sağlıklı sınıfına doğru şekilde atanmasıdır. Bilgisayarlı görü ve derin öğrenme alanındaki son gelişmeler, bu tür problemlerin otomatik ve güvenilir bir şekilde çözülmesini mümkün kılmıştır.

Bu projenin amacı, bitki yaprağı hastalıklarının tespiti için veri ön işleme, klasik makine öğrenmesi yöntemleri, derin öğrenme modelleri, sayısal optimizasyon ve değerlendirme adımlarını içeren uçtan uca bir sınıflandırma sistemi geliştirmek ve farklı yaklaşımları karşılaştırmalı olarak analiz etmektir.

2. Veri Kümesi Tanımı

Bu çalışmada kullanılan deneyler, bitki hastalıkları alanında yaygın olarak kullanılan PlantVillage veri kümesi üzerinde gerçekleştirilmiştir. PlantVillage veri kümesi, kontrollü koşullar altında çekilmiş, sağlıklı ve hastalıklı bitki yapraklarına ait RGB görüntülerden oluşmaktadır. Kontrollü çekim koşulları, modellerin arka plan gürültüsü yerine hastalığa özgü görsel özelliklere odaklanmasını sağlamaktadır.

Projede kullanılan işlenmiş veri kümesi toplamda 24 farklı sınıftan oluşmaktadır. Bu sınıflardan biri yapraksız arka plan görüntülerini temsil ederken, diğer sınıflar belirli bitki–hastalık kombinasyonları ve sağlıklı bitki sınıflarını içermektedir. Model performansının tarafsız ve güvenilir bir şekilde ölçülebilmesi için veri kümesi, sınıf dağılımları korunacak şekilde eğitim, doğrulama ve test alt kümelerine ayrılmıştır.

Veri kümesi bölünmesi şu şekildedir:

- Eğitim kümesi: 19.862 görüntü (%70)
- Doğrulama kümesi: 4.256 görüntü (%15)
- Test kümesi: 4.257 görüntü (%15)

Sınıf dağılımları incelendiğinde, bazı hastalık sınıflarında görece dengesizlikler olduğu görülmektedir. Bu nedenle, yalnızca doğruluk metriğine bağlı kalınmamış; sınıf dengesizliğini dikkate alan makro-F1 skoru gibi metrikler de değerlendirmeye dahil edilmiştir. Veri kümesine ait detaylı görsel analizler ve dağılım grafikleri Jupyter Notebook içerisinde sunulmuştur.

3. Yöntem

Geliştirilen sistem; veri ön işleme, klasik makine öğrenmesi taban çizgi modeli ve derin öğrenme tabanlı model olmak üzere üç ana bileşenden oluşmaktadır. Tüm süreç, tekrar üretilebilirlik ve adil karşılaştırma ilkeleri gözetilerek tasarlanmıştır.

3.1 Veri Ön İşleme

Tüm görüntüler, modele girdi olarak verilmeden önce sabit boyutlara yeniden ölçeklendirilmiştir. Klasik makine öğrenmesi ve derin öğrenme yaklaşımlarının hesaplama gereksinimleri farklı olduğundan, bu iki yaklaşım için farklı görüntü boyutları tercih edilmiştir. Veri kümesi, herhangi

bir model eğitimi yapılmadan önce eğitim, doğrulama ve test olarak ayrılmıştır. Bu sayede veri sızıntısının (data leakage) önüne geçilmiştir.

Ayrıca, deneylerin tekrar üretilebilirliğini sağlamak amacıyla Python, NumPy ve PyTorch kütüphaneleri için sabit rastgelelik tohumları (seed) kullanılmıştır.

3.2 Klasik Makine Öğrenmesi Taban Çizgi Modeli

Taban çizgi modeli olarak, el yapımı özniteliklere dayalı klasik bir sınıflandırma yaklaşımı kullanılmıştır. Histogram of Oriented Gradients (HOG) yöntemi, yaprak görüntülerindeki kenar ve yapısal bilgileri yakalamak amacıyla tercih edilmiştir. Buna ek olarak, renk histogramları kullanılarak hastalık belirtileriyle ilişkili renk dağılımları modele dahil edilmiştir.

Elde edilen öznitelik vektörleri standartlaştırıldıktan sonra doğrusal bir Destek Vektör Makinesi (SVM) kullanılarak sınıflandırma gerçekleştirilmiştir. SVM hiperparametreleri, eğitim verisinin sınıf dağılımı korunarak oluşturulan bir alt küme üzerinde rastgele arama (randomized search) ve çapraz doğrulama ile optimize edilmiştir. Hesaplama maliyetini düşürmek amacıyla, çıkarılan öznitelikler disk üzerinde önbelleğe alınmıştır.

3.3 Derin Öğrenme Modeli

Derin öğrenme yaklaşımı için ResNet-18 evrimsel sinir ağı mimarisi kullanılmıştır. Model, ImageNet veri kümesi üzerinde önceden eğitilmiş ağırlıklarla başlatılmış ve son sınıflandırma katmanı PlantVillage veri kümesindeki sınıf sayısına göre yeniden düzenlenmiştir.

İki farklı eğitim stratejisi değerlendirilmiştir: evrimsel omurganın dondurularak yalnızca son katmanın eğitilmesi ve üst seviyedeki bazı katmanların ince ayar (fine-tuning) ile yeniden eğitilmesi. Eğitim sürecinde çapraz entropi kayıp fonksiyonu ve Adam optimizasyon algoritması kullanılmıştır. Aşırı öğrenmenin önüne geçmek ve gereksiz hesaplamaları azaltmak için doğrulama kaybına dayalı erken durdurma (early stopping) uygulanmıştır.

4. Deneysel Tasarım

Deneysel tasarım, modeller arasında adil ve şeffaf bir karşılaştırma yapılacak şekilde oluşturulmuştur. HOG + SVM yaklaşımı taban çizgi modeli olarak belirlenirken, ResNet-18 tabanlı derin öğrenme modeli geliştirilmiş yaklaşım olarak ele alınmıştır. Tüm modeller aynı eğitim, doğrulama ve test kümeleri üzerinde değerlendirilmiştir.

Sayısal optimizasyon, deneysel tasarımın önemli bir parçasıdır. Klasik model için SVM hiperparametre araması yapılmış, derin öğrenme modeli için ise öğrenme oranı ve ağırlık çürütmesi (weight decay) değerleri farklı senaryolar altında denenmiştir. Erken durdurma mekanizması her iki yaklaşımda da tutarlı şekilde uygulanmıştır. Bu yapı, ders yönergesinde belirtilen en az bir taban çizgi ve bir iyileştirme/karşılaştırma şartını karşılamaktadır.

5. Değerlendirme Metrikleri

Model performansı birden fazla metrik kullanılarak değerlendirilmiştir. Genel doğruluk (accuracy) metriği raporlanmakla birlikte, tek başına yeterli görülmemiştir. Sınıf dengesizliğini dikkate alan makro-ortalama F1 skoru temel performans göstergesi olarak kullanılmıştır. Ayrıca precision-recall analizleri ve karışıklık matrisleri ile sınıf bazlı performans detaylı şekilde incelenmiştir. Tüm görsel sonuçlar Jupyter Notebook içerisinde sunulmuştur.

6. Sonuçlar

Deneysel sonuçlar, derin öğrenme tabanlı ResNet-18 modelinin klasik HOG + SVM yaklaşımına kıyasla tüm metriklerde daha yüksek performans sağladığını göstermektedir. Özellikle makro-F1 skorundaki artış, derin öğrenme modelinin sınıf dengesizliğine karşı daha dayanıklı olduğunu ortaya koymaktadır.

Karışıklık matrisi analizleri, hataların çoğunlukla görsel olarak birbirine benzeyen hastalık sınıfları arasında gerçekleştiğini göstermektedir. Klasik model, daha düşük performansa sahip olmasına rağmen, yorumlanabilirliği ve düşük hesaplama maliyeti sayesinde güçlü bir referans noktası sunmaktadır.

7. Tartışma ve Hata Analizi

Yanlış sınıflandırılan örnekler incelendiğinde, hataların çoğunlukla düşük kontrastlı görüntülerden veya hastalık belirtilerinin çok benzer olduğu sınıflardan kaynaklandığı görülmüştür. Notebook içerisinde yer alan gradyan tabanlı görselleştirme yöntemleri, derin öğrenme modelinin tahmin yaparken yaprağın hastalıklı bölgelerine odaklandığını göstermektedir.

8. Sayısal Yöntemler ve Optimizasyon Katkısı

Bu projede sayısal yöntemler önemli bir rol oynamaktadır. SVM hiperparametre optimizasyonu ve derin öğrenme modelinde yapılan öğrenme oranı taramaları, model yakınsamasını ve genelleme başarımını artırmıştır. Erken durdurma yöntemi ise aşırı öğrenmeyi azaltarak daha kararlı sonuçlar elde edilmesini sağlamıştır.

9. Tekrar Üretilbilirlik ve Mühendislik Kalitesi

Deneylerin tekrar üretilbilirliğini sağlamak için sabit rastgelelik tohumları kullanılmış, veri bölünmeleri diske kaydedilmiş ve deney ortamına ait bilgiler kayıt altına alınmıştır. Proje yapısı modüler olacak şekilde tasarlanmış ve kod okunabilirliği ile sürdürülebilirliği ön planda tutulmuştur.

10. Sınırlılıklar ve Gelecek Çalışmalar

Elde edilen başarılı sonuçlara rağmen, kullanılan veri kümesi çoğunlukla kontrollü ortamlarda çekilmiş görüntülerden oluşmaktadır. Bu durum, modelin gerçek saha koşullarındaki performansını sınırlayabilir. Gelecek çalışmalarda, farklı ışık koşullarında toplanmış verilerle eğitim yapılması, alan uyarlaması (domain adaptation) yöntemlerinin uygulanması ve gerçek zamanlı sistemlere entegrasyon hedeflenmektedir.

11. Sonuç

Bu raporda, bitki yaprağı hastalıklarının sınıflandırılması problemi klasik makine öğrenmesi ve derin öğrenme yaklaşımları kullanılarak kapsamlı şekilde ele alınmıştır. Derin öğrenme modelleri daha yüksek performans sunarken, klasik yöntemler yorumlanabilirlik ve hesaplama verimliliği açısından önemli avantajlar sağlamaktadır. Çalışma, veri işleme, modelleme, sayısal optimizasyon ve değerlendirme adımlarını içeren, tekrar üretilebilir ve mühendislik açısından sağlam bir yapay zeka çözümü ortaya koymaktadır.

Kaynakça

Hughes, D. P., & Salathé, M. (2015). An open access repository of images on plant health to enable the development of mobile disease diagnostics. *arXiv preprint arXiv:1511.08060*.

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.

Pedregosa, F., vd. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*.

Paszke, A., vd. (2019). PyTorch: An imperative style, high-performance deep learning library. *Advances in Neural Information Processing Systems*.