

2209/A

ÜNİVERSİTE ÖĞRENCİLERİ ARAŞTIRMA PROJELERİ DESTEK
PROGRAMI

SONUÇ RAPORU

PROJE BAŞLIĞI: GRAM BOYANMIŞ LAMLARDAN , BAKTERİ
ALGILAMA VE GRAM MORFOLOJİSİNİN YAPAY ZEKAYLA
RAPORLANMASI

PROJE YÜRÜTÜCÜSÜNÜN ADI: YUSUF TARIK BALTA

DANIŞMANININ ADI: Prof. Dr. ALPER AKÇALI

GENEL BİLGİLER

PROJENİN KONUSU	Gram Boyanmış Lamlarından, Bakteri Algılama ve Gram Morfolojisinin Yapay Zekayla Raporlanması
PROJE YÜRÜTÜCÜSÜNÜN ADI	YUSUF TARIK BALTA
DANIŞMANIN ADI	Prof. Dr. ALPER AKÇALI
PROJE BAŞLANGIÇ VE BİTİŞ TARİHLERİ	29.09.2023-29.09.2024

1. Giriş

Proje, Gram boyanmış bakterilerin tanımlanması için yapay zeka teknolojilerinden faydalanmayı hedeflemektedir. Pre-trained Evrişimli Sinir Ağları (CNN) kullanılarak, bakterilerin mikroskop altındaki görüntülerinden tanımlanmasını sağlamak amaçlanmıştır. Tıbbi mikrobiyolojide, bakterilerin tanımlanması için kullanılan geleneksel yöntemler zaman ve uzman bilgisi gerektirdiğinden, bu süreçleri daha verimli hale getirecek bir model geliştirilmesi planlanmıştır.

Yapay zeka teknolojileri, özellikle derin öğrenme, tıp alanında önemli bir yer edinmiştir ve görüntü işleme gibi konularda başarılı sonuçlar vermektedir. Gram boyama yöntemi, bakterilerin mikroskop altında renk ve şekil gibi özelliklerini gözlemleyerek tanımlanmasında uzun yıllardır kullanılan bir tekniktir. Bu proje, bu yöntemin verimliliğini artırmayı ve bakterilerin gram boyanmada morfolojik yapılarının otomatik olarak tanımlanmasını sağlamayı hedeflemiştir.

Proje kapsamında, mevcut CNN modelleri kullanılarak bir yapay zeka modeli oluşturulacak ve bu model Gram boyanmış bakterilerin tanımlanmasında kullanılacaktır. Modelin doğruluğunu ve verimliliğini artırmak için en güncel kütüphaneler ve platformlar kullanılmış ayrıca laboratuvarımızdan elde edilen görsel verilerle modelin eğitilmesi sağlanmıştır. Bu model, tıp alanında önemli bir katkı sağlamakla kalmayacak, aynı zamanda ilaç, veteriner sağlık ve gıda sektörlerinde de potansiyel uygulama alanları bulacaktır. Proje, Evrişimli Sinir Ağları kullanılarak Gram boyalı lamlar üzerindeki bakterilerin tanımlanmasını sağlayarak bu alanda yenilikçi bir yaklaşım sunmayı amaçlamaktadır.

2. Rapor dönemlerinde yapılan çalışmalar

2.1. Kendi Veri Setimizin Oluşturulması ve Modelin Eğitiminde Kullanılacak Verilerin Hazırlanması (0.-3. Ay)

Başlangıçta veri setinin hazırlanması sürecinin 3 ay sürmesi ve 1000 görüntü elde edilmesi öngörülmüştü. Ancak, verilerin toplanma süresinin beklenenden uzun sürmesi ve daha fazla veriye ihtiyaç duyulması nedeniyle bu aşamanın süresi yaklaşık 2-3 ay uzadı. Ayrıca verilerin çoğaltılması ve farklı çekilmiş görüntü tiplerine sahip verilerden faydalanıp daha kapsamlı bir model eğitimi sağlanması için literatür incelenmiştir. Proje için uygun olabilecek hazır verisetleri araştırılmıştır. Bu araştırmalar sonucunda güncel ve güvenilir iki hazır veri setinden yararlanılmıştır. Bunlar M-ROSE(1) ve DIBas(2) veri setleridir. Bu veri setlerinden de örnekler alınarak toplamda 2013 görüntü elde edildi. Bu görüntülerin büyük bir kısmı (yaklaşık 1200) kendi çekimlerimizden oluşmaktadır. Kendi çekimini yaptığımız görüntülerdeki gürültüleri yok etmek amacıyla bazı ön işleme uygulamaları yapılmıştır. Ön işlemeden geçen görüntüler sınıflarına göre klasörlere yerleştirilmiştir. Fakat özellikle kendi çekimlerimizde bazı sınıfa ait görüntülerin daha sık karşılaşıldığı için o sınıfa ait bir yığılma gerçekleşmiştir. Bu nedenle modelde o sınıfların lehine bir eğilim gözlenmiştir. Bu durumu dengelemek için hazır verisetleri kullanılarak dengeleme yapılması denense de istenen denge sağlanamamıştır. Ayrıca, veri toplama sürecinde çekim yapılan cihazın bozulması, sürecin aksamasına neden olmuştur. Bu aşamadaki gecikmeler, veri toplama sürecindeki öngörülemeyen teknik sorunlar ve cihaz arızaları, proje zaman çizelgesini önemli ölçüde etkilemiştir.

2.2 Kullanılacak Modelin Oluşturulması ve Çalıştırılması (4.-9. Ay)

Bu aşamada, çeşitli modeller denenmiş ve farklı yapılandırmalarla performans iyileştirmeleri yapılmaya çalışılmıştır. Ancak, bu süreç hem beklenenden uzun sürmüş hem de istenilen performans elde edilememiştir. Denenen modellerden hiçbirisi yeterli sonuçlar verememiştir. Örneğin 5 pre-trained modellerin performans analizleri, test sonuçları ve performans sorunlarının nedenleri raporun ilerleyen kısımlarında detaylı olarak değerlendirilecektir. Bu süreç, modelin başarılı bir şekilde çalışabilmesi için sürekli denemeler ve yapılandırma değişikliklerinin gerekli olduğunu göstermiştir. Gerekli araştırmalar doğrultusunda bu değişiklikleri uygulama çalışsak da yeterli cevap alınamamıştır. Burada kullanılan pre-trained modellerin çalışma prensibinin etkili olduğu düşünülmektedir. Bu modeller, tüm görüntü üzerindeki örüntüleri takip ederek öğrenme temeline dayanmaktadır. Bizim görüntülerimizde ise noktasal tipte bakteri görüntüleri ve onun etrafında daha büyük doku parçaları bulunmaktadır. Ve modeller birazdan örneklerde görüleceği üzere bu büyük dokulara odaklanma eğilimi göstermektedir. Daha küçük alanlarda çalışması istediğinde ise yine aynı şekilde daha küçük doku parçaları, mikroskop merceğinden kaynaklı görüntüye gelen aslında anlamlı olmayan toz parçaları gibi bir çok küçük parça modelin eğitimini oldukça zorlaştırmaktadır. Modelin eğitimi sırasında aldığımız accuracy , loss score gibi metrikler sonuçları yansıtmakta başarısız kalmıştır .

2.3 Oluşturulan Modelin Performans Analizi (9.-12. Ay)

Modelin performans analizi ve test verilerinin oluşturulması beklenenden daha hızlı gerçekleştirilmiştir. Ancak, eğitilen modellerin başarısız olduğu fark edilmiştir. Performans analizinin yanı sıra, başarısızlığın nedenlerini belirlemek için kapsamlı bir sorun tespiti yapılmaya çalışılmıştır. Bu süreç, modelin başarısızlık nedenlerini anlamak ve gelecekteki iyileştirmeler için gerekli bilgileri sağlamayı amaçlamıştır. Bu

anlamda yapılan arařtırmalar sonucunda test verilerinin klasik metriklerinin yanı sıra modelin, görüntünün hangi bölümüne göre çıktı verdiđini řematize eden aktivasyon haritası yapan metotlar arařtırmıřtır. Ve bunun için Grad-CAM metodundan faydalanmıř ve sonuçları incelenmiřtir.

3. Sonuç

Birçok pre-trained model denenmiř ve hiçbirinden istenen performans alınamamıřtır. Ařađıda örnek olması açısından, dört farklı pre-trained modelin sonuçları özetlenmiřtir: InceptionV3, VGG16, VGG19 ve AlexNet. Bu modellerin eğitim çıktıları (accuracy, loss, precision, recall, F1 score vb.) ařađıda verilmiřtir: [Deđerler]. Test verileri üzerinden elde edilen sonuçlar accuracy, precision, recall ve F1 score sırasıyla řu řekildedir: [Deđerler]. Bu sonuçlar, pre-trained modellerin bu veri seti üzerinde tatmin edici performans göstermediđini ortaya koymaktadır. Bu nedenle, sınıflama yapan modellerde temel deđiřiklikler yaparak veya YOLO, FastCNN gibi segmentasyon tabanlı modeller kullanarak daha uygun ve bařarılı sonuçlar elde edilebileceđi düşünölmektedir.

Test verileriyle yapılan metriklerden alınan sonuçlar Tablo-1 'de sunulmuřtur :

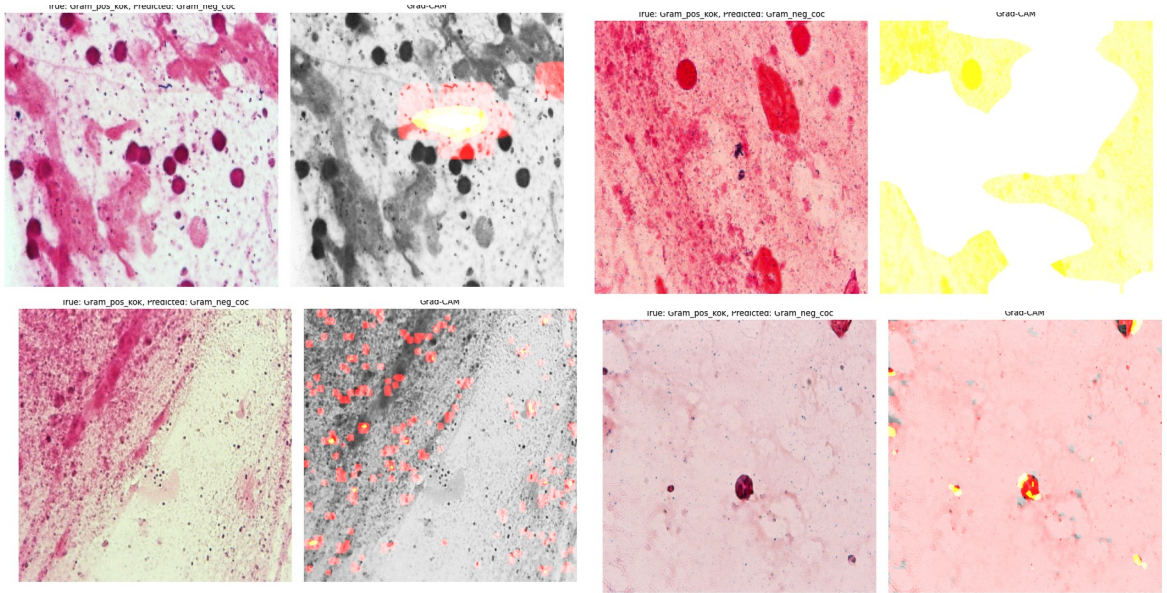
Tablo-1 Test verileri ile denenen yöntemlerin skorları

Metric	Value---VGG19	Metric	Value ---- VGG16
Accuracy	0.327	Accuracy	0.310
Precision	0.182	Precision	0.145
Recall	0.327	Recall	0.310
F1 Score	0.210	F1 Score	0.197
Metric	Value----- InceptionV3	Metric	Value----- ALEXNET
Accuracy	0.206	Accuracy	0.086
Precision	0.109	Precision	0.118
Recall	0.206	Recall	0.086
F1 Score	0.142	F1 Score	0.099

Metriklerin sonuçları, uygulanan pre-trained modellerin bu görüntüler üzerinde bařarısız olduđunu göstermektedir.

4. Çıktılar

Görsel- 1



Görsel-1: Test verileri üzerinde modellerin Grad-CAM uygulaması çıktısı. Haritalandırma göstermiştir ki model bakterilerden bağımsız bir öğrenme gerçekleşmiştir. Solda orijinal görüntü, sağ tarafta modelin hangi bölgeleri temel alarak tahminleme yaptığı aktivasyon haritası gösterilmiştir.

Kodlar ve veriler için: https://gitlab.com/yusuftarikgjhk/tubitak_2209

5. Proje ile ilgili harcama kalemleri hakkında ayrıntılı bilgi

- Microcase Dürbün Teleskop Mikroskop için Telefon Bağlantı Aparatı - Model No AL2559 - 200,00 TL

Fotoğraf çekimi esnasında mikroskop merceğinden stabil görüntüler almamız için çekim yapın cihazın sabit durması için kullanılmıştır.



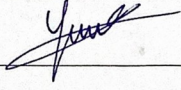
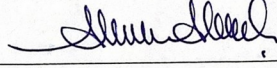
- Seagate 2tb Phenom 2 – 2.199 TL
Toplanan verilerin saklanması , eğitilmiş modellerin saklanması ve başka cihazlar üzerinde denemeler yapmak üzere taşınması için kullanılmıştır.

Harcamalar için gerekli belgeler EK-1 , EK-2 ve EK-3 de gösterilmiştir.

6. Kaynakça:

1 - Wang, X., Shi, Y., Guo, S., Qu, X., Xie, F., Duan, Z., Hu, Y., Fu, H., Shi, X., Quan, T., Wang, K., & Xie, L. (2024). A Clinical Bacterial Dataset for Deep Learning in Microbiological Rapid On-Site Evaluation. *Scientific Data*, 11(1), 1-8. <https://doi.org/10.1038/s41597-024-03370-5>

2- Zieliński, B., Plichta, A., Misztal, K., Spurek, P., Brzychczy-Włoch, M., & Ochońska, D. (2017). Deep learning approach to bacterial colony classification. *PLOS ONE*, 12(9), e0184554. <https://doi.org/10.1371/journal.pone.0184554>

PROJE YÜRÜTÜCÜSÜNÜN ADI - SOYADI - İMZA	DANIŞMANIN ADI - SOYADI - İMZA
Yusuf Tarık BALTA	Prof. Dr. Alper AKÇALI
	

Tarih : 25/09/2024