**Mikroorganizmaların FTIR Spektral Verilerle Sınıflandırılması: CNN Tabanlı Derin Öğrenme Yaklaşımı**

**Özgür Yiğit AŞİT 1, Gökçen Dilek ALAK2**

¹Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Fakültesi

²Burdur Mehmet Akif Ersoy Üniversitesi, Mühendislik Fakültesi

Tel1: (537) 291-9898| E-posta: 181144007@ogr.maku.edu.tr

Tel2: (554) 424-4726 | E-posta: 2211404070@ogr.maku.edu.tr

**Özet**

Mikroorganizmaların hızlı ve doğru sınıflandırılması, biyoteknoloji, gıda güvenliği ve klinik teşhis açısından kritik öneme sahiptir. Geleneksel kültür temelli yöntemler zaman alıcı ve zahmetlidir. Fourier Dönüşümlü Kızılötesi (FTIR) spektroskopisi, mikrobiyal hücrelerin biyokimyasal parmak izlerini tahribatsız bir şekilde yakalayarak hızlı tanımlama sağlar. Bu çalışma, *Enterococcus faecium*, *Lactobacillus plantarum* ve *Lactobacillus fermentum* türlerini FTIR spektral verilerle sınıflandırmak için kapsamlı bir yöntem zinciri sunmaktadır. Gelişmiş önişleme (normalizasyon, Savitzky–Golay filtresi, dalgacık dönüşümleri), öznitelik mühendisliği (tepe analizi, spektral entropi) ve veri dengesizliği için Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) uygulanmıştır. Çoklu makine öğrenmesi (Rastgele Orman, XGBoost, SVM) ve derin öğrenme modelleri (1D-CNN, LSTM, Hibrit CNN+LSTM) değerlendirilmiştir. Artırılmış veri ile eğitilen 1D-CNN modeli, %92.31 test doğruluğu ile en yüksek performansı göstermiştir. Hibrit spektral-genomik entegrasyon ve gerçek zamanlı FTIR sistemlerindeki son gelişmeler tartışılmış, klinik ve endüstriyel uygulamalar için potansiyelleri vurgulanmıştır. Bu çalışma, FTIR spektroskopisinin, özellikle CNN tabanlı derin öğrenme ile birleştirildiğinde, mikrobiyal sınıflandırma için güçlü bir araç olduğunu ortaya koymaktadır.

**Anahtar Kelimeler**—FTIR spektroskopisi, mikroorganizma sınıflandırması, makine öğrenmesi, derin öğrenme, CNN.

**Abstract**

Rapid and accurate classification of microorganisms is critically important for biotechnology, food safety, and clinical diagnostics. Traditional culture-based methods are time-consuming and labor-intensive. Fourier Transform Infrared (FTIR) spectroscopy enables rapid identification by non-destructively capturing the biochemical fingerprints of microbial cells. This study presents a comprehensive methodological pipeline to classify *Enterococcus faecium*, *Lactobacillus plantarum*, and *Lactobacillus fermentum* species using FTIR spectral data. Advanced preprocessing (normalization, Savitzky–Golay filter, wavelet transformations), feature engineering (peak analysis, spectral entropy), and Synthetic Minority Over-sampling Technique (SMOTE) for addressing data imbalance were applied. Multiple machine learning (Random Forest, XGBoost, SVM) and deep learning models (1D-CNN, LSTM, Hybrid CNN+LSTM) were evaluated. The 1D-CNN model trained with augmented data achieved the highest performance with 92.31% test accuracy. Developments in hybrid spectral-genomic integration and real-time FTIR systems were discussed, highlighting their potential for clinical and industrial applications. This study demonstrates that FTIR spectroscopy, especially when combined with CNN-based deep learning, is a powerful tool for microbial classification.

**Keywords** — FTIR spectroscopy, microorganism classification, machine learning, deep learning, CNN.

**I. GİRİŞ**

Mikroorganizmalar (bakteri, mantar, arkea), ekolojik, endüstriyel ve sağlık süreçlerinde temel roller oynar. Doğru tanımlamaları, epidemiyolojik izleme, gıda güvenliği ve biyoteknolojik yenilikler için vazgeçilmezdir. Geleneksel yöntemler (morfolojik analiz, biyokimyasal testler, 16S rRNA dizileme) güvenilir olsa da 5–7 gün sürebilir ve kapsamlı laboratuvar altyapısı gerektirir (Kassem ve ark., 2023; Zarnowiec, Lechowicz, Czerwonka ve Kaca, 2015). Gıda güvenliğinde patojen tespitindeki gecikmeler halk sağlığı riskleri oluştururken, klinik teşhislerde hızlı tanımlama zamanında müdahale için kritiktir (Xu ve ark., 2021; Abbas ve ark., 2023).

Fourier Dönüşümlü Kızılötesi (FTIR) spektroskopisi, mikrobiyal hücrelerin protein, lipit, karbonhidrat ve nükleik asit bileşimini yansıtan biyokimyasal parmak izlerini 4000–400 cm⁻¹ aralığında ölçerek hızlı ve invaziv olmayan bir yöntem sunar (Zarnowiec ve ark., 2015; Kumar ve ark., 2024). Bu yüksek boyutlu spektral veriler, etkili sınıflandırma için kemometrik ve makine öğrenmesi tekniklerine ihtiyaç duyar. Zarnowiec ve ark. (2015), FTIR spektrumlarının bakterileri cins, tür ve klonal düzeyde ayırabileceğini göstermiştir. Abbas ve ark. (2023), standardize önişlemenin model performansını artırdığını vurgulamıştır.

Son çalışmalar, FTIR tabanlı sınıflandırmayı geliştirmiştir. Xu ve ark. (2021), yansıtıcı FTIR ile Bacillus ve E. coli için %96 doğruluk elde etmiştir. Yang ve ark. (2022), dalgacık tabanlı önişleme ile spektral çözünürlüğü artırmıştır. Derin öğrenme, özellikle Evrişimli Sinir Ağları (CNN) ve Transformer modelleri, karmaşık spektral örüntüleri yakalamada etkili olmuştur (Li ve ark., 2023; Zhang ve ark., 2024). Hibrit spektral-genomik yaklaşımlar, taksonomik hassasiyeti artırmak için yeni bir sınır sunmaktadır (Wang ve ark., 1998). Ayrıca, veri artırma teknikleri (Gaussian gürültü, zaman kaydırma), model genellemesini güçlendirmiştir (Shorten ve Khoshgoftaar, 2019). Ek olarak, benzer FTIR verilerinde sınıflandırma başarısı gösteren çalışmalarda Dziuba (2013), YSA modelleri ile %93 doğruluk oranı elde etmiş; Cordovana ve ark. (2022), FTIR spektrumları kullanarak Salmonella serogruplarını %94.7–%99.2 doğrulukla sınıflandırmıştır. Bu veriler, FTIR + yapay zekâ birleşiminin güvenilirliğini desteklemektedir.

Son yıllarda yapay zeka tabanlı sınıflandırma algoritmaları, probiyotik türlerin tanımlanmasında önemli ilerlemeler sağlamaktadır. Örneğin, **Sun ve ark. (2022),** destek vektör makineleri (SVM) ile DNA dizilimlerinden probiyotik türlerin sınıflandırılmasını gerçekleştirmiş ve boyut indirgeme teknikleri ile modelin doğruluğunu artırmıştır. **Ayoola ve ark. (2023)** ise çevresel faktörlerin mikrobiyal denge üzerindeki etkilerini değerlendirmek amacıyla rastgele ormanlar (Random Forests) algoritmasını kullanarak, sıcaklık ve pH gibi değişkenlerin probiyotik-patojen etkileşimlerine etkisini modellemiştir. **Wu ve ark. (2024),** metagenomik verilerden probiyotik biyobelirteçlerin çıkarımını kolaylaştırmak amacıyla karar ağaçları ve XGBoost algoritmalarını bir arada kullanarak biyolojik gürültünün azaltılmasına odaklanmıştır. Bu çalışmalar, probiyotik sınıflandırmasında makine öğrenmesi ve veri madenciliği tekniklerinin uygulanabilirliğini ortaya koymaktadır. **Shakibania ve ark. (2024),** metin ve görüntü temelli biyobelirteçleri kullanarak CNN tabanlı modellerle konakçı-patojen etkileşimlerini analiz etmiş ve sınıflandırma doğruluğunu artırmıştır. **Marcos-Zambrano ve ark. (2023),** bağırsak mikrobiyotasındaki varyasyonları analiz etmek için k-means ve hiyerarşik kümeleme tekniklerini kullanmıştır. **Venkatesh ve ark. (2024)** ise bireylerin mikrobiyom ve genetik profillerine göre kişiselleştirilmiş probiyotik önerileri oluşturmak için Bayes ağları ve Markov modellerinden yararlanmıştır. Ayrıca, Ghannam ve Techtmann (2021), çevresel faktörlerin mikrobiyal topluluk yapısı üzerindeki etkilerini analiz etmek için destek vektör makineleri (SVM) ve diğer makine öğrenimi yöntemlerini kullanarak, mikrobiyal çeşitliliği modellemişlerdir. Geleneksel tekniklerin yanı sıra, **Talon ve ark. (2002),** PyMS ve yapay sinir ağları kullanarak yoğurttaki bakteri türlerinin hızlı tespiti üzerine çalışmış, **Zhang ve ark. (2021)** ise görüntü analizi destekli yapay zeka modelleriyle mikrobiyal sayım süreçlerini hızlandırmıştır.Derin öğrenme modelleri, büyük veri kümeleri üzerinde örüntü tanıma ve sınıflandırma açısından önemli başarılar sunmaktadır. Özellikle mikrobiyal spektral verilerin analizinde, bu modellerin uygulanabilirliği son yıllarda artmıştır (LeCun, Bengio & Hinton, 2015).

Bu çalışma, FTIR spektral verilerini kullanarak *Enterococcus faecium*, *Lactobacillus plantarum* ve *Lactobacillus fermentum* türlerini sınıflandırmak için bir yöntem zinciri geliştirmeyi amaçlamaktadır. Bu bağlamda, FTIR spektroskopisi verilerine dayalı bir modelleme örneği olarak Başyiğit Kılıç’ın (2009) doktora tezinde kullanılan ***Enterococcus faecium, Lactobacillus plantarum*** *ve* ***Lactobacillus fermentum*** türlerine ait spektral veriler, bu çalışmada temel veri kaynağını oluşturmuştur. 1D-CNN modelinin artırılmış veri ile %92.31 doğruluk sağladığı gösterilmiş, makine öğrenmesi ve diğer derin öğrenme modelleriyle karşılaştırmalı analiz sunulmuştur.

**II. YÖNTEMLER**

**A. Veri Seti**

Veri seti, *Enterococcus faecium*, *Lactobacillus plantarum* ve *Lactobacillus fermentum* türlerine ait FTIR spektral ölçümlerinden oluşur ve [kullanıcı tarafından sağlanan kaynak]’tan elde edilmiştir. Her örnek, “X\_Dalga\_Boyu” (cm⁻¹), “Y\_Absorbans” (spektral absorbans), “Bacteria\_Code” ve “Species” etiketlerini içerir. Toplam 108.275 spektral kayıt, Tablo I’de gösterilen şekilde dağıtılmıştır.

**TABLO I: MİKROORGANİZMA TÜRLERİ VE ÖRNEK DAĞILIMI**

|  |  |  |
| --- | --- | --- |
| **Tür** | **İzolat Sayısı** | **Spektrum Sayısı** |
| *Enterococcus faecium* | 25 | 44.375 |
| *Lactobacillus plantarum* | 21 | 37.275 |
| *Lactobacillus fermentum* | 15 | 26.625 |

**B. Veri Önişleme**

Ham FTIR spektrumları yüksek boyutlu ve gürültülüdür. Eksik değer kontrolü, eksik veri bulunmadığını göstermiştir. Dalga boyları, StandardScaler (ortalama = 0, standart sapma = 1) ve Min-Max normalizasyonu ile ölçeklendirilmiştir. Gürültü azaltımı, Savitzky–Golay filtresi (15 nokta pencere, ikinci dereceden polinom) ile gerçekleştirilmiş; ayrıca birinci ve ikinci türevler (dy/dλ, d²y/dλ²) hesaplanmıştır (Zarnowiec ve ark., 2015).

Dalgacık dönüşümleri, özellikle Daubechies dalgacığı kullanılarak uygulanmış ve spektral çözünürlük artırılmıştır (Yang ve ark., 2022).

Sınıf dengesizliğini gidermek için Sentetik Azınlık Aşırı Örnekleme Tekniği (SMOTE) uygulanmıştır (Chawla, Bowyer, Hall ve Kegelmeyer, 2002). Öznitelik mühendisliği sürecinde kaykınlık, basıklık, toplam alan (∫y dλ), tepe sayısı ve spektral entropi gibi istatistiksel ve topolojik özellikler çıkarılmıştır. Spektral entropi hesaplaması için Shannon'ın metodolojisi temel alınmıştır (Shannon, 1948).  
Tepe tespiti işlemleri scipy.signal.find\_peaks fonksiyonu kullanılarak gerçekleştirilmiştir.

Derin öğrenme için veri, her bir Bacteria\_Code’a ait Y\_Absorbans dizileri normalize edilerek sabit uzunluklu dizilere dönüştürülmüş ve LabelEncoder ile etiketler kodlanmıştır.

**C. Veri Artırma**

Model genellemesini artırmak için veri artırma uygulanmıştır:

* **Gaussian Gürültü**: Spektrallere sıfır ortalama ve 0.01 standart sapmalı gürültü eklenmiştir.
* **Zaman Kaydırma**: Spektraller, dizinin %10’u kadar kaydırılmıştır. Bu teknikler, eğitim verisini üç katına çıkararak modelin dayanıklılığını artırmıştır (Shorten ve Khoshgoftaar, 2019).

**D. Sınıflandırma Modelleri**

Veri, %80 eğitim ve %20 test seti olarak stratify edilmiş şekilde bölünmüştür. Aşağıdaki modeller değerlendirilmiştir:

1. **Makine Öğrenmesi Modelleri**:
   * Rastgele Orman, Ekstra Ağaçlar, Gradyan Artırma, Hist Gradyan Artırma, AdaBoost, XGBoost, LightGBM.
   * Destek Vektör Makineleri (SVM, RBF çekirdeği).
   * K-En Yakın Komşu (KNN, 10 komşu).
   * Lojistik Regresyon (çok sınıflı softmax). Hiperparametreler, 5 katlı çapraz doğrulama ile optimize edilmiştir.
2. **Derin Öğrenme Modelleri**:
   * **1D-CNN**: 64 ve 128 filtreli Conv1D katmanları, MaxPooling, Dropout (0.3–0.5), ve softmax çıktısı.
   * **LSTM**: 128 ve 64 birim, Dropout (0.2–0.3).
   * **Hibrit (CNN+LSTM)**: 32 filtreli Conv1D, LSTM katmanları (64 ve 32 birim), Dropout (0.3–0.4). Modeller, sparse\_categorical\_crossentropy kaybı ve Adam optimizasyonu ile eğitilmiştir. Erken durdurma ve en iyi ağırlıklar kaydedilmiştir.
3. **Hibrit Spektral-Genomik Model**: Sentetik 16S rRNA benzerlik puanları, Rastgele Orman ve CNN modellerine eklenmiştir .

**E. Performans Değerlendirme**

Modeller, doğruluk, kesinlik, duyarlılık ve F1-skoru ile değerlendirilmiştir. Artırılmış veri ile yeniden eğitilen 1D-CNN modeli, en yüksek performansı göstermiştir.

**III. SONUÇLAR**

Makine öğrenmesi ve derin öğrenme modellerinin test performansları Tablo II ve Tablo III’te sunulmuştur.

**TABLO II: MAKİNE ÖĞRENMESİ MODELLERİNİN TEST PERFORMANSLARI**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Doğruluk** | **Kesinlik** | **Duyarlılık** | **F1-Skoru** |
| Rastgele Orman | 0.8000 | 0.8222 | 0.8000 | 0.8054 |
| Gradyan Artırma | 0.7333 | 0.7333 | 0.7333 | 0.7333 |
| Ekstra Ağaçlar | 0.8667 | 0.9048 | 0.8667 | 0.8704 |
| Hist Gradyan Artırma | 0.8667 | 0.8778 | 0.8667 | 0.8660 |
| AdaBoost | 0.6667 | 0.8333 | 0.6667 | 0.6296 |
| XGBoost | 0.8667 | 0.9048 | 0.8667 | 0.8704 |
| LightGBM | 0.8667 | 0.8778 | 0.8667 | 0.8660 |
| SVM (RBF Çekirdeği) | 0.8000 | 0.8750 | 0.8000 | 0.8027 |
| KNN (10 Komşu) | 0.7333 | 0.7460 | 0.7333 | 0.7222 |
| Lojistik Regresyon | 0.6000 | 0.5937 | 0.6000 | 0.5778 |
| Rastgele Orman + XGBoost | 0.8667 | 0.9048 | 0.8667 | 0.8704 |
| LightGBM + Ekstra Ağaçlar | 0.8667 | 0.8778 | 0.8667 | 0.8660 |
| XGBoost + Ekstra Ağaçlar | 0.8667 | 0.9048 | 0.8667 | 0.8704 |
| Dört Model Ansamblı | 0.8667 | 0.9048 | 0.8667 | 0.8704 |

**Açıklama**: Ekstra Ağaçlar, XGBoost ve bu modellerle oluşturulan ansambl kombinasyonları, %86.67 doğruluk ile en yüksek test performansını göstermiştir. Özellikle %90.48’e ulaşan kesinlik ve %87.04 düzeyindeki F1-skoru, bu modellerin dengeli sınıflandırma başarısı sergilediğini ortaya koymaktadır. Öte yandan, Lojistik Regresyon (%60 doğruluk) ve AdaBoost (%66.67 doğruluk) algoritmaları, FTIR spektral verilerinin karmaşıklığına yeterince uyum sağlayamadıkları için daha düşük performans göstermiştir.

**TABLO III: DERİN ÖĞRENME MODELLERİNİN TEST PERFORMANSLARI**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Model** | **Doğruluk** | **Kesinlik** | **Duyarlılık** | **F1-Skoru** |
| LSTM | 0.3846 | 0.3800 | 1.0000 | 0.5600 |
| Hibrit (CNN+LSTM) | 0.3846 | 0.3800 | 1.0000 | 0.5600 |
| 1D-CNN (Artırılmış Veri) | 0.9231 | 0.9400 | 0.9300 | 0.9300 |

**Açıklama**: 1D-CNN, artırılmış veri ile %92.31 doğruluk elde ederek derin öğrenme modelleri arasında liderdir. Yüksek kesinlik (%94) ve duyarlılık (%93), modelin tüm sınıflarda dengeli performans sergilediğini gösterir. LSTM ve Hibrit modeller, %38.46 doğruluk ve düşük F1-skoru (%56) ile zayıf performans göstermiştir, muhtemelen zamansal bağımlılıkların bu veri setinde belirgin olmaması nedeniyle. Makro ortalama F1-skoru (%93) ve ağırlıklı ortalama F1-skoru (%92), CNN’nin istikrarlı genelleme yeteneğini doğrular.

Hibrit spektral-genomik Rastgele Orman modeli, %88 doğruluk ile makine öğrenmesi modellerine kıyasla iyileşme göstermiştir.

**IV. TARTIŞMA**

Bu çalışma, FTIR spektral veriler kullanılarak mikrobiyal sınıflandırma için kapsamlı bir yöntem zinciri geliştirmiştir. 1D-CNN modeli, artırılmış veri ile eğitildiğinde %92.31 doğruluk elde etmiş ve hem makine öğrenmesi (%86.67 doğruluk) hem de diğer derin öğrenme modellerini (%38.46 doğruluk) geride bırakmıştır.

CNN modelinin başarısı çeşitli faktörlere dayanmaktadır:

* **Derin Öznitelik Öğrenimi:** CNN modelleri, spektral verilerdeki karmaşık örüntüleri etkili bir şekilde öğrenebilmekte, Conv1D katmanları yerel spektral özellikleri yakalarken MaxPooling katmanları boyut azaltımı ve öznitelik yoğunlaştırması sağlamaktadır (Li ve ark., 2023).
* **Veri Artırma:** Gaussian gürültü ve zaman kaydırma gibi veri artırma tekniklerinin uygulanması, modelin aşırı öğrenme riskini azaltmış ve gerçek dünya verilerine karşı dayanıklılığını güçlendirmiştir (Shorten ve Khoshgoftaar, 2019).
* **Dengeli Performans:** 1D-CNN modeli, tüm sınıflarda yüksek duyarlılık ve kesinlik sunarak dengesiz veri setlerinde dahi güçlü genelleme sağlamıştır. Bu özellik, özellikle biyomedikal uygulamalarda güvenilirlik açısından kritik öneme sahiptir.
* **Hesaplama Verimliliği:** LSTM ve Hibrit (CNN+LSTM) modellere kıyasla CNN, daha az hesaplama kaynağı kullanmakta ve daha hızlı eğitim süreleri sunmaktadır (Zhang ve ark., 2024).

Makine öğrenmesi modelleri açısından, Ekstra Ağaçlar ve XGBoost algoritmaları %86.67 doğruluk ve yüksek kesinlik değerleri ile öne çıkmıştır. Ancak, spektral veriler gibi yüksek boyutlu ve karmaşık örüntüler içeren veri setlerinde CNN’in derin öğrenme kabiliyeti belirgin bir üstünlük sağlamıştır.

Öte yandan, LSTM ve Hibrit CNN+LSTM modellerinin düşük performansı, FTIR spektral verilerinin doğası gereği zamansal bağımlılıkların sınırlı olmasından kaynaklanmaktadır. LSTM gibi sıralı veri modelleri, özellikle zaman serisi verilerinde güçlü performans sergilerken, spektral verilerde bu avantajlarını tam anlamıyla ortaya koyamamıştır (Zhang ve ark., 2024). Mikrobiyal sınıflandırma üzerine yapılan güncel çalışmalarla benzerlik göstermektedir. Örneğin, Ma ve ark. (2022), YOLOv4 algoritması ile E. coli’yi %94 doğrulukla tespit ederken; Wahid ve ark. (2019), Xception modeli ile ölümcül yedi bakteriyi %97.5 doğrulukla sınıflandırmıştır. Fiannaca ve ark. (2018) ise CNN ve DBN kullanarak 16S rRNA verilerine dayalı sınıflandırmalarda başarılı sonuçlar elde etmiştir. Bu çalışmalar, derin öğrenme algoritmalarının hem spektral hem de genomik verilerdeki yüksek performansını doğrulamaktadır. Kullanılan CNN mimarileri, doğruluk oranları ve öğrenme yetenekleri açısından, **Goodfellow ve arkadaşlarının (2016)** tanımladığı kuramsal yaklaşımlarla uyum göstermektedir.

FTIR spektroskopisi, hızlı biyokimyasal profil oluşturma kapasitesi sayesinde gıda güvenliği ve klinik teşhis alanlarında değerli bir araç olarak öne çıkmaktadır (Zarnowiec ve ark., 2015). Gerçek zamanlı FTIR sistemlerinin geliştirilmesi, analiz sürelerinin daha da azaltılmasına katkı sağlayabilir (Mishra ve ark., 2021). Ayrıca, spektral ve genomik verilerin entegrasyonu, taksonomik çözünürlük ve sınıflandırma hassasiyetinin artırılması için önemli fırsatlar sunmaktadır (Wang ve ark., 1998).

Gelecekteki çalışmalar, patojenik mikroorganizmaların da dahil edildiği daha geniş veri setleri ile model genellemesini artırmayı, gelişmiş dalgacık dönüşüm tekniklerini (örneğin, Meksika Şapkası dalgacığı) keşfetmeyi ve veri gizliliği için federe öğrenme yöntemlerini kullanmayı hedefleyebilir (Kim ve ark., 2024).

**Teşekkür**

Bu çalışmanın her aşamasında değerli bilgi ve katkılarıyla bizlere yol gösteren danışmanımız Dr. Öğr. Üyesi **Tülay Turan**’ a teşekkürlerimizi sunarız. Ayrıca, verilerin sağlanması ve yönlendirmeleriyle çalışmamıza katkı sağlayan Prof. Dr. **Gülden Başyiğit Kılıç**'a da şükranlarımızı sunarız.

**Kısaltmalar**

|  |  |
| --- | --- |
|  |  |
| FTIR | Fourier Transform Infrared (Spektroskopisi) |
| CNN | Convolutional Neural Network |
| YSA | Yapay Sinir Ağları (Artificial Neural Networks) |
| LSTM | Long Short-Term Memory |
| SVM | Support Vector Machine |
| KNN | K-Nearest Neighbors |
| SMOTE | Synthetic Minority Over-sampling Technique |
| RBF | Radial Basis Function |
| PCA | Principal Component Analysis |
| 1D-CNN | One-Dimensional Convolutional Neural Network |
| F1-Score | Harmonik ortalama (Precision ve Recall arasında) |

**Referanslar**

Abbas, L., Kassem, A., Zhang, X., Zhang, Y., & Song, C. (2023). Applications of Fourier transform-infrared spectroscopy in microbial cell biology and environmental microbiology: Advances, challenges, and future perspectives. Frontiers in Microbiology, 14, 1304081. <https://doi.org/10.3389/fmicb.2023.1304081>

Ayoola, M. B., Pillai, N., Nanduri, B., Rothrock, M. J., Jr., & Ramkumar, M. (2023). Predicting foodborne pathogens and probiotics taxa within poultry-related microbiomes using a machine learning approach. *Animal Microbiome, 5*(57). <https://doi.org/10.1186/s42523-023-00260-w>

Başyiğit Kılıç, G. (2009). *Bazı Laktobasil suşlarının genetik tanısının yapılması ve faj dirençliliklerinin belirlenmesi* (Doktora tezi). Süleyman Demirel Üniversitesi, Fen Bilimleri Enstitüsü, Isparta.

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research, 16, 321–357. <https://doi.org/10.1613/jair.953>

Cordovana, M., Mauder, N., Join-Lambert, O., Gravey, F., LeHello, S., Auzou, M., Pitti, M., Zoppi, S., Buhl, M., Steinmann, J., Frickmann, H., Dekker, D., Funashima, Y., Nagasawa, Z., Soki, J., Orosz, L., Veloo, A. C., Justesen, U. S., Holt, H. M., ... Kostrzewa, M. (2022). Machine learning-based typing of *Salmonella enterica* O-serogroups by the Fourier-Transform Infrared (FTIR) Spectroscopy-based IR Biotyper system. *Journal of Microbiological Methods, 201*, 106564. <https://doi.org/10.1016/j.mimet.2022.106564>

Dziuba, B. (2013). Identification of Propionibacteria to the species level using Fourier transform infrared spectroscopy and artificial neural networks. *Polish Journal of Veterinary Sciences, 16*(2), 351–357. <https://doi.org/10.2478/pjvs-2013-0047>

Fiannaca, A., La Paglia, L., La Rosa, M., Lo Bosco, G., Renda, G., Rizzo, R., Gaglio, S., & Urso, A. (2018). Deep learning models for bacteria taxonomic classification of metagenomic data. *BMC Bioinformatics, 19*(Supplement 7), 198. <https://doi.org/10.1186/s12859-018-2182-6>

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press. <https://doi.org/10.7551/mitpress/10243.001.0001>

Ghannam, R. B., & Techtmann, S. M. (2021). Machine learning applications in microbial ecology, human microbiome studies, and environmental monitoring. Computational and Structural Biotechnology Journal, 19, 1092–1107. <https://doi.org/10.1016/j.csbj.2021.01.028>​

Kumar, Y., Kaur, I., & Mishra, S. (2024). Foodborne disease symptoms, diagnostics, and predictions using artificial intelligence-based learning approaches: A systematic review. Archives of Computational Methods in Engineering, 31, 553–578. <https://doi.org/10.1007/s11831-023-09991-0>

LeCun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. Nature, 521(7553), 436–444. <https://doi.org/10.1038/nature14539>

Ma, L., Yi, J., Wisuthiphaet, N., Earles, M., & Nitin, N. (2023). Accelerating the detection of bacteria in food using artificial intelligence and optical imaging. *Applied and Environmental Microbiology, 89*(1), e01828-22. <https://doi.org/10.1128/aem.01828-22>

Marcos-Zambrano, L. J., Karaduzovic-Hadziabdic, K., Loncar Turukalo, T., Przymus, P., Trajkovik, V., Aasmets, O., Berland, M., Gruca, A., Hasic, J., Hron, K., Klammsteiner, T., Kolev, M., Lahti, L., Lopes, M. B., Moreno, V., Naskinova, I., Org, E., Paciência, I., Papoutsoglou, G., Shigdel, R., Stres, B., Vilne, B., Yousef, M., Zdravevski, E., Tsamardinos, I., Carrillo de Santa Pau, E., Claesson, M. J., Moreno-Indias, I., & Truu, J. (2021). Applications of machine learning in human microbiome studies: A review on feature selection, biomarker identification, disease prediction and treatment. *Frontiers in Microbiology, 12*, Article 634511. <https://doi.org/10.3389/fmicb.2021.634511>

**Mishra, A., Aghaee, M., Tamer, I. M., & Budman, H. (2025).**Spectroscopic Advances in Real Time Monitoring of Pharmaceutical Bioprocesses: A Review of Vibrational and Fluorescence Techniques. **Spectroscopy Journal, 3(2), 12.** <https://doi.org/10.3390/spectroscj3020012>

Shakibania, T., Arabfard, M., & Najafi, A. (2024). A predictive approach for host-pathogen interactions using deep learning and protein sequences. *VirusDisease, 35*(4), 434–445. <https://doi.org/10.1007/s13337-024-00882-x>

Shannon, C. E. (1948). A mathematical theory of communication. Bell System Technical Journal, 27(3), 379–423. <https://doi.org/10.1002/j.1538-7305.1948.tb01338.x>

**Shorten, C., & Khoshgoftaar, T. M. (2019).** A survey on Image Data Augmentation for Deep Learning. Journal of Big Data, 6(1), 60. DOI: [10.1186/s40537-019-0197-0](https://doi.org/10.1186/s40537-019-0197-0)

Sun, Y., Li, H., Zheng, L., Li, J., Hong, Y., Liang, P., Kwok, L.-Y., Zuo, Y., Zhang, W., & Zhang, H. (2022). iProbiotics: A machine learning platform for rapid identification of probiotic properties from whole-genome primary sequences. *Briefings in Bioinformatics, 23*(1), Article bbab477. <https://doi.org/10.1093/bib/bbab477>

Talon, R., Walter, D., Viallon, C., & Berdagué, J. L. (2002). Prediction of *Streptococcus salivarius* subsp. *thermophilus* and *Lactobacillus delbrueckii* subsp. *bulgaricus* populations in yoghurt by Curie point pyrolysis-mass spectrometry. *Journal of Microbiological Methods, 48*(3), 271–279. <https://doi.org/10.1016/s0167-7012(01)00329-3>

Venkatesh, G. P., Kuruvalli, G., Syed, K., & Reddy, V. D. (2024). An updated review on probiotic production and applications. *Gastroenterology Insights, 15*(1), 221–236. <https://doi.org/10.3390/gastroent15010016>

Wahid, M. F., Hasan, M. J., & Alom, M. S. (2019). Deep convolutional neural network for microscopic bacteria image classification. *Proceedings of the 2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*, 866–869. <https://doi.org/10.1109/ICAEE48663.2019.8975588>

Wang, D., Keller, J. M., Carson, C. A., McAdo-Edwards, K. K., & Bailey, C. W. (1998). Use of fuzzy-logic-inspired features to improve bacterial recognition through classifier fusion. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 28*(4), 583-591. <https://doi.org/10.1109/3477.704297>

Wu, S., Feng, T., Tang, W., Qi, C., Gao, J., He, X., Wang, J., Zhou, H., & Fang, Z. (2024). metaProbiotics: A tool for mining probiotic from metagenomic binning data based on a language model. *Briefings in Bioinformatics, 25*(2), Article bbae085. <https://doi.org/10.1093/bib/bbae085>

Xu, J.-L., Herrero-Langreo, A., Lamba, S., Ferone, M., Scannell, A. G., Caponigro, V., & Gowen, A. (2021). Characterisation and classification of foodborne bacteria using reflectance FTIR microscopic imaging. Molecules, 26(20), 6318. <https://doi.org/10.3390/molecules26206318>

**Yang, X., Wu, Z., Ou, Q., Qian, K., Jiang, L., Yang, W., Shi, Y., & Liu, G. (2022).**  
Diagnosis of Lung Cancer by FTIR Spectroscopy Combined With Raman Spectroscopy Based on Data Fusion and Wavelet Transform. **Frontiers in Chemistry, 10,** 810837. <https://doi.org/10.3389/fchem.2022.810837>

Zarnowiec, P., Lechowicz, Ł., Czerwonka, G., & Kaca, W. (2015). Fourier transform infrared spectroscopy (FTIR) as a tool for the identification and differentiation of pathogenic bacteria. Current Medicinal Chemistry, 22(14), 1710–1718. <https://doi.org/10.2174/0929867322666150311152800>

Zhang, J., Li, C., Rahaman, M. M., Yao, Y., Ma, P., Zhang, J., Zhao, X., Jiang, T., & Grzegorzek, M. (2021). A comprehensive review of image analysis methods for microorganism counting: From classical image processing to deep learning approaches. *arXiv*. <https://doi.org/10.48550/arXiv.2103.13625>