

Tekrarlayıcı Sinir Ağları ile Mikropların Sınıflandırılması

Talha Burak ALAKUŞ^{1*}

¹ Yazılım Mühendisliği Bölümü, Mühendislik Fakültesi, Kırklareli Üniversitesi, Kırklareli, Türkiye
^{*1} talhaburakalakis@klu.edu.tr

(Geliş/Received: 26/05/2023;

Kabul/Accepted: 07/08/2023)

Öz: Mikroplar insan vücudunun hem içinde hem de dışında yaşayan ve hemen hemen her yerde bulunan tek hücreli küçük canlılardır. Zararlı olabileceği gibi yararlı olan mikroplarda bulunmaktadır. Bu yüzden mikroplar ekolojik dengeyi sağlanmasında büyük bir rol üstlenmektedir. Ancak, bu mikropları birbirinden ayırt edebilmek çoğu zaman meşakkatli bir iş olmaktadır. Mikropların çeşitli özelliklerini anlayan ve taksonomi nüanslarını yorumlayabilen bir uzman bilgisinin gerekmesi ve laboratuvar ihtiyacının olması bu süreci hem maliyetli hem de zaman alıcı yapmaktadır. Bundan dolayı, bilgisayar destekli sistemlerin bu alanda önemi artmış ve günümüzde yapay zeka teknolojileri mikropların sınıflandırılmasında kullanılmaya başlanmıştır. Makine öğrenmesi ve derin öğrenme teknolojilerinin bu alana uygulanmasıyla yüksek seviyede doğruluk skorları elde edilmekte ve biyoçeşitlilikle ilgili değerlendirmeler otomatik bir şekilde yapılabilmektedir. Bu çalışmada da derin öğrenme modellerinden biri olan tekrarlayıcı sinir ağları kullanılarak, mikropların sınıflandırılması gerçekleştirilmiştir. Çalışma dört aşamadan meydana gelmiştir. Birinci aşamada, veriler elde edilmiştir. İkinci aşamada bu veriler normalleştirme işlemine tabi tutulmuş ve ön işlemden geçirilmiştir. Üçüncü aşamada UKSB ve TSA modelleri tasarlanmış ve sınıflandırılmıştır. Son aşamada ise sınıflandırıcıların performansları doğruluk, kesinlik, duyarlılık, F1-skor ve EAA skoru ile belirlenmiştir. Çalışmanın sonunda TSA sınıflandırıcısı ile %92,53, UKSB sınıflandırıcısıyla ise %99,85 doğruluk skoru elde edilmiştir.

Anahtar kelimeler: Mikrop tahmini, derin öğrenme, tekrarlayıcı sinir ağları.

Classification of Microbes with Recurrent Neural Networks

Abstract: Microbes are tiny single-celled creatures that live both inside and outside the human body and are found almost everywhere. There are microbes that can be harmful as well as harmless. Therefore, microbes play a major role in maintaining the ecological balance. However, distinguishing these microbes from each other is often a difficult task. The need for a specialist who understands the various characteristics of microbes and can interpret taxonomy nuances and the need for a laboratory makes this process both costly and time consuming. Therefore, the importance of computer aided systems in this field has increased and artificial intelligence technologies have begun to be used in the classification of microbes today. With the application of artificial intelligence technologies to this field, high level accuracy scores are obtained, and biodiversity assessments can be made automatically. In this study, microbes were classified by using deep learning models, which is one of the artificial intelligence methods. The study consisted of four stages. In the first stage, the data were obtained. In the second stage, these data were preprocessed and normalized. In the third stage, LSTM and RNN models are designed, and data were classified. In the last stage, the performances of the classifiers were determined by accuracy, precision, recall, F1-score, and AUC score. At the end of the study, an accuracy score of 92,53% was obtained with the RNN classifier and 99,85% with the LSTM classifier.

Key words: Microbe prediction, deep learning, recurrent neural networks.

1. Giriş

Tek hücreli bir canlı olan mikroplar hayatımızın hemen hemen her yerinde bulunmaktadır. Mikroplar zararlı olabileceği gibi yararlı da olmaktadır. Bazıları sağlıklı sindirime yardımcı olmak için insanların bağırsaklarında yaşarken [1], bazıları ciltlerde bulunmaktadır [2]. Bu yararlı mikroplar vücudun dengede kalmasına ve bağışıklık sisteminin güçlendirilmesinde yardımcı olmaktadır [3]. Bunun yanı sıra, zararlı mikroplar ise yüzeylere yapışıp bulaşıcı hastalık yayarak vücutta hastalık ve enfeksiyonlara yol açan rahatsızlıklara neden olabilmektedirler [4]. Yararlı ve zararlı olmak üzere dört farklı türde mikrop bulunmaktadır. Bunlar; bakteriler, virüsler, mantarlar ve tek hücrelilerdir.

Mikroplar havada, suda, bitkilerde, yiyeceklerde, hayvanlarda, toprakta ve suda yoğun olarak bulunabilmektedirler. Bunlara ek olarak, doğrudan ya da dolaylı temas yolu olmak üzere genellikle iki farklı yol ile yayılmaktadırlar [5]. Mikroplardan korunabilmek için yapılması gereken en temel işlem kişisel hijyene

* Sorumlu yazar: talhaburakalakis@klu.edu.tr. Yazarın ORCID Numarası: ¹ 0000-0003-3136-3341

dayanmaktadır. Kişisel hijyene önem verilmesi, kullanılmış ürünlerin tekrar kullanılmaması, eşyaların ortak kullanılmaması, vb. işlemler, mikropların bulaşmasını engelleyebilmektedir.

Zararlı mikroplar kişilerde enfeksiyona neden olduğu için, izlenecek tedavi yolları da farklılık göstermektedir. Bundan dolayı çeşitli ilaç tedavileri uygulanabilmektedir. Ancak, mikropların yararlı ya da zararlı olup olmadığına karar vermek uzun süren ve sıkıcı bir işlemdir [6]. Bunun yanı sıra, mikropların sınıflandırılabilmesi ve özelliklerine göre ayırt edilebilmesi için bir uzman yorumuna ihtiyaç duyulmaktadır [7]. Bu da sürecin yavaş olmasına ve laboratuvar ihtiyacı gerektirdiği için maliyetli olmasına neden olmaktadır [8]. Bu dezavantajların önüne geçebilmek için, araştırmacılar yapay zeka destekli çalışmalara yönelmişler ve hem derin öğrenme hem de makine öğrenmesi algoritmalarını kullanarak etkili sonuçlar elde etmişlerdir. [9] numaralı çalışmada araştırmacılar LR (Lojistik Regresyon – Logistic Regression), DVM (Destek Vektör Makinesi – Support Vector Machine) ve KNN (K En Yakın Komşu – K Nearest Neighbor) algoritmalarını kullanarak sınıflandırma işlemi gerçekleştirmişlerdir. Çalışma üç aşamadan meydana gelmiştir. Birinci aşamada veriler elde edilmiştir. İkinci aşamada ön işlem yapılmış ve PBA (Parçacık Bileşen Analizi – Principal Component Analysis) ile öznitelik seçimi gerçekleştirilmiştir. Son aşamada ise sınıflandırma işlemi yapılmış ve modellerin başarımları doğruluk, kesinlik, hassasiyet ve F1-skorumları ile belirlenmiştir. En yüksek başarımlar DVM ile elde edilmiş ve %98,2 oranında doğruluk gözlemlenmiştir. Başka bir çalışmada ise araştırmacılar mikro alglerin görüntülerini kullanarak sınıflandırma işlemi gerçekleştirmiştir [10]. Çalışmada görüntü verileri değerlendirilmiş ve çeşitli görüntü işleme yöntemleri kullanılarak sınıflandırma yapılmıştır. Sınıflandırma işlemi için ÇKA (Çok Katmanlı Algılayıcı – Multi-Layer Perceptron), SMO (Sıralı Minimum Optimizasyon – Sequential Minimal Optimization), Bagging ve J48 algoritmaları kullanılmıştır. Sınıflandırıcıların performansı sadece doğruluk skoru ile belirlenmiş ve en yüksek skor %97,22 ile SMO sınıflandırıcısıyla elde edilmiştir. [11] numaralı çalışmada da benzer bir yaklaşım uygulanmış ve alg görüntüleri kullanılarak sınıflandırma gerçekleştirilmiştir. Çalışmada sınıflandırıcı olarak SMO kullanılmış ve doğruluk skoru ile sınıflandırıcının performansı belirlenmiştir. Çalışmanın sonunda %98,6 oranında doğruluk elde edilmiştir. Yapılan çalışmalar incelendiğinde, elde edilen sonuçların yüksek ve etkili olduğu gözlemlenmiştir.

Bahsedilen çalışmalardan edilen başarımlar doğrultusunda, bu çalışmada da yapay zeka destekli bir sistem geliştirilmiş ve derin öğrenme algoritmaları kullanılarak mikropların sınıflandırılması yapılmıştır. Çalışma dört aşamadan meydana gelmiştir. Birinci aşamada, veriler elde edilmiştir. İkinci aşamada bu veriler normalleştirme işlemine tabi tutulmuş ve ön işleminden geçirilmiştir. Üçüncü aşamada UKSB (Uzun/Kısa Süreli Bellek – Long/Short Term Memory) ve TSA (Tekrarlayıcı Sinir Ağı – Recurrent Neural Network) modelleri tasarlanmış ve sınıflandırılmıştır. Son aşamada ise sınıflandırıcıların performansları doğruluk, kesinlik, duyarlılık, F1-skor ve EAA (Eğri Altındaki Alan – Area Under Curve) skoru ile belirlenmiştir. Çalışmanın öne çıkan noktaları şu şekilde belirlenmiştir:

- Alanyazın incelendiğinde bu alanda yapılan çalışmaların azlığı göze çarpmış ve bu alana katkının sağlanması hedeflenmiştir.
- Çalışmalar incelendiğinde derin öğrenme algoritmalarıyla bu alanda gerçekleştirilen çalışmaya rastlanmamıştır. Bu çalışma ile bu algoritmaların başarımlarının gösterilmesi ve araştırılması hedeflenmiştir.
- Alanyazında az sayıda gerçekleştirilen çalışmalarda genellikle görüntü sınıflandırılması yapılmıştır. Bu çalışmada ise görüntülerden ziyade, mikropların çeşitli öznitelikleri (sayısal veriler) kullanılmış ve sınıflandırma gerçekleştirilmiştir.

Çalışmanın organizasyonu şu şekilde tasarlanmıştır: İkinci kısımda veri seti ve kullanılan yöntemler hakkında bilgiler verilmiştir. Üçüncü kısımda uygulama sonuçları sergilenmiş ve her bir sınıflandırıcının sonuçları verilmiştir. Dördüncü kısımda bulgular irdelenmiş ve tartışma yapılarak çalışmanın hem avantajlarından hem de dezavantajlarından bahsedilmiştir. Son kısımda ise çalışmanın alanyazına katkılarından bahsedilmiş ve elde edilen bulguların etkileri incelenmiştir.

2. Materyal ve Yöntemler

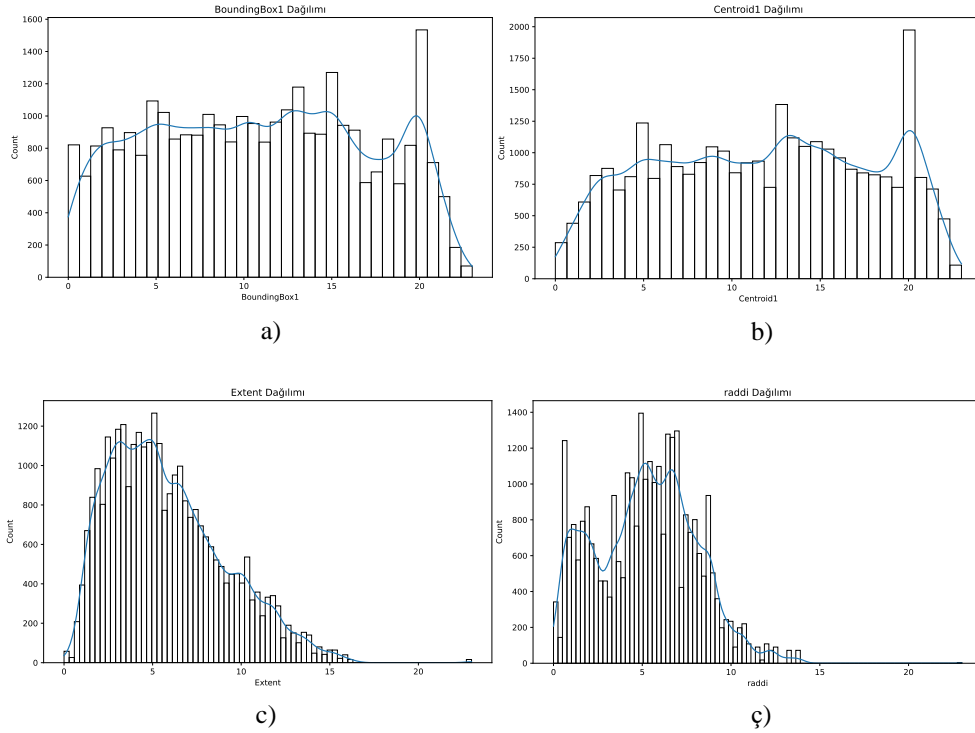
2.1. Veri seti

Çalışmada kullanılan veriler [12] numaralı çalışmadan elde edilmiştir. Çalışmada on farklı mikroorganizmaya ait veriler kullanılmıştır. Bu veriler mikroskobik görüntülerden elde edilmiş ve paylaşılmıştır. Veri setine ait çeşitli bilgiler Tablo 1’de verilmiştir.

Tablo 1. Çalışmada kullanılan veri seti hakkında genel bilgiler.

Mikroorganizma İsmi	Mikroorganizma Türü	Veri Sayısı
Aspergillus sp	Fungi	3,888
Diatom	Alg	1,818
Penicillium	Mantar	1,080
Pithophora	Alg	1,350
Protozoa	Ökaryot	3,888
Raizopus	Fungi	2,552
Spirogyra	Alg	612
Ulothrix	Alg	7,420
Volvox	Alg	4,320
Yeast	Fungi	3,600

Veri setinde toplamda 30,528 adet veri bulunmaktadır. Bu çalışmada da bu veriler kullanılmış ve çoklu sınıflandırma yapılmıştır. Sınıflandırma işlemi hem TSA hem de UKSB algoritmalarıyla gerçekleştirilmiştir. Şekil 1'de mikroplara ait birkaç özelliğin dağılımı gösterilmiştir.



Şekil 1. Verilerin çeşitli özelliklere ve etiketlere göre dağılımı: (a) boundingbox1, (b) centroid1, (c) extend, (ç) raddi.

Şekil 1'de on adet mikroorganizmaya ait yirmi dört özellikten sadece dört tanesi verilmiş ve bu özelliklerin dağılımı gösterilmiştir.

2.2. Sınıflandırma yöntemleri

Bu çalışmada sınıflandırma işlemi için TSA ve UKSB derin öğrenme algoritmaları kullanılmıştır. TSA algoritması sıralı verileri işlemek için kullanılan ve dahili bir hafızaya sahip olan güçlü bir derin sinir ağı modelidir [13]. TSA sisteminde hafıza hücreleri bulunmakta ve bir zincir oluşturmaktadır. Bu zincirli yapı sayesinde giriş verisindeki her bir eleman için aynı görev yürütülmektedir. Ancak, TSA yönteminde çeşitli problemler gözlemlenmektedir. Bu problemlerden en önemlileri kaybolan gradyan ve uçan gradyan problemleridir [14, 15]. Bu problemlerin önüne geçmek için ise başka bir tekrarlayıcı sinir ağı modeli olan UKSB önerilmiştir. UKSB

modelinde, standart bir TSA’da olduğu gibi tek bir sinir ağı bulunmamaktadır. Bu modelde dört etkileşimli bir katman yer almaktadır. Genel olarak bir UKSB ağında, bellek blokları yer almaktadır. Buradaki durumlar bir sonraki bloğa aktarılmaktadır. Bu sayede veriler temelde değişiklik göstermemekte ve ileriye yönelik hareket edebilmektedir. UKSB modeli daha önceden de açıklandığı üzere kaybolan gradyan problemini çözmek için geliştirilmiş bir yöntemdir. Kaybolan gradyan problemi, büyük farklılıklar bulunan girdilerin aktivasyon sonrasında ufak değer aralıklarına indirgenmesiyle, bunların gittikçe sıfıra yaklaşması ve model tarafından algılanamaması olarak ifade edilmektedir. Bundan dolayı da öğrenme işlemi yeteri kadar iyi olamamaktadır. UKSB ise bu problemin üstesinden yapısında bulunan hücre blokları, unutmama, girdi ve çıktı kapıları aracılığıyla gelmektedir [16]. Bu nedenle çoğu durumda UKSB mimarisi daha etkili bir performans sergilemektedir.

2.3. Değerlendirme ölçütleri

Bu çalışmada derin öğrenme modellerinin başarımları doğruluk, kesinlik, hassasiyet, F1-skor, EAA ve karmaşıklık matrisi değerlendirme ölçütleriyle belirlenmiştir. Karmaşıklık matrisi bir sınıflandırma algoritmasının performansını göstermek için kullanılan bir çeşit tablodur. Performans, eğitim verisinin test edilerek belirlenmesine dayanmaktadır. Karmaşıklık matrisinde GP (Gerçek Pozitif – True Positive), GN (Gerçek Negatif – True Negative), YP (Yanlış Pozitif – False Positive) ve YN (Yanlış Negatif – False Negative) olmak üzere dört farklı parametre bulunmaktadır. GP ifadesi tahmin edilen değer pozitif ve doğru olduğunu ifade etmektedir. Bunun yanı sıra ise GN ifadesi tahmin edilen değer negatif ve doğru olduğunu göstermektedir. Bunlara ek olarak, YP ifadesi değer pozitif ancak yanlış olduğunu belirtirken, YN ifadesi değer negatif ve yanlış olduğunu ifade etmektedir. Doğruluk, F1-skor, kesinlik ve hassasiyet değerleri ise bu verilen ifadelerin çeşitli denklemlerle hesaplanmasına dayanmaktadır. Denklem 1 ve Denklem 4 arasında bu işlemlerin hesaplanması gösterilmiştir.

$$\text{Doğruluk} = \frac{GP + YN}{GP + GN + YP + YN} \quad (1)$$

$$\text{Kesinlik} = \frac{GP}{GP + YP} \quad (2)$$

$$\text{Hassasiyet} = \frac{GP}{GP + YN} \quad (3)$$

$$F1 - \text{Skor} = 2 * \left(\frac{\text{Kesinlik} * \text{Hassasiyet}}{\text{Kesinlik} + \text{Hassasiyet}} \right) \quad (4)$$

Bunun yanı sıra EAA skoru da önemli bir değerlendirme kriteridir. EAA skoru, AİK (Alıcı İşletim Karakteristiği – Receiver Operating Characteristic) eğrisinin altındaki alanı vermektedir. EAA skoru genellikle; modelin sınıfları ayırma konusunda ne kadar iyi olup olmadığını göstermektedir. Ayrıca, en iyi eşliğin belirlenmesinde de kullanılmaktadır [17]. EAA skoru yorumlanması Tablo 2’de verilen aralıklar göz önünde bulundurulmaktadır.

Tablo 2. EAA skorlarının yorumlanması [18].

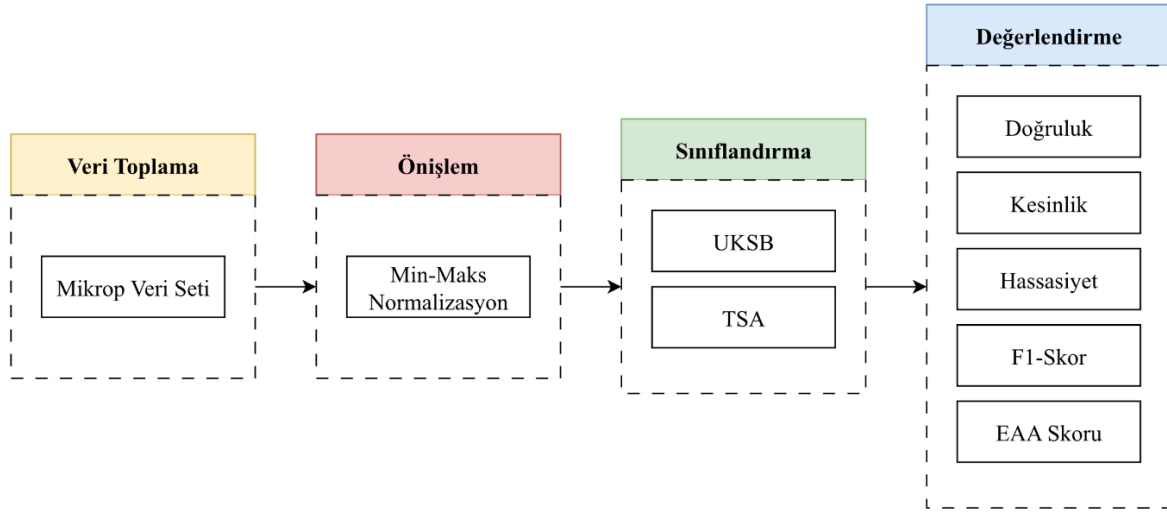
EAA Skoru	Açıklama
0,00 – 0,49	Ayırım yapılamaz
0,50 – 0,69	Zayıf sınıflandırma
0,70 – 0,79	Kabul edilebilir sınıflandırma
0,80 – 0,89	Harika sınıflandırma
0,90 – 1,00	Olağanüstü sınıflandırma

EAA skorunun 0,49 ve altında olması sınıflandırma işleminin yapılmadığını ifade etmektedir. Bunun yanı sıra ise eğer EAA skoru 0,50 ve 0,69 arasındaysa bu zayıf bir sınıflandırma olarak kabul edilmektedir. EAA skorunun 0,70 ve 0,79 arasında olması kabul edilebilir sınıflandırıcıyı ifade ederken, bu değer 0,80 ve 0,89 arasında olması harika sınıflandırma işlemi yapıldığını göstermektedir.

2.4. Uygulama geliştirme ortamı

Çalışmada macOS işletim sistemine sahip bir bilgisayar üzerinde gerçekleştirilmiştir. Bilgisayar üzerinde M1 çip bulunmakta ve 8 GB RAM yer almaktadır. Apple Neural Engine’in M1 çipine dahil edilmesi, Mac cihazlarda

makine öğrenimi ve derin öğrenme görevlerinin hızını ve performansını önemli ölçüde arttırmaktadır. Son teknoloji 16 çekirdekli mimarisiyle, Apple'ın M1 çipine gömülü Neural Engine, Mac cihazlarının saniyede 11 trilyon gibi etkileyici bir işlem gerçekleştirmesini sağlamaktadır. M1 çipi, makine öğrenimi ve derin öğrenme görevlerinde olağanüstü performans sağlamak için özel olarak tasarlanmıştır. Çalışmada sınıflandırma işlemi Python programlama dili kullanılarak Anaconda, Spyder ortamında gerçekleştirilmiştir. Spyder'ın çalışmadaki versiyonu 5.4.2'dir. Anaconda, Python programlama dilinin popüler bir dağıtımıdır. Python programlama dilinin yanı sıra çeşitli veri bilimi, makine öğrenimi ve derin öğrenme araçlarını içeren zengin bir ekosisteme sahip bir platformdur. Pandas, NumPy ve SciPy gibi kütüphaneler veri işleme, matematiksel hesaplamalar ve istatistiksel analiz için kullanılmaktadır. Matplotlib ve Seaborn gibi görselleştirme kütüphaneleri ise verilerin görselleştirmesine izin vermektedir. Scikit-learn, TensorFlow ve PyTorch gibi derin öğrenme ve makine öğrenimi kütüphaneleri, çeşitli modelleme ve tahmin görevleri için sıklıkla kullanılmaktadır. Sınıflandırma işlemi için Tensorflow kütüphanesi kullanılmış ve bu kütüphanenin çeşitli komutlarından yararlanılmıştır. Ayrıca Keras kütüphanesi kullanılarak optimizasyon algoritması, kayıp fonksiyonu ve değerlendirme metrikleri hesaplanmış ve tanımlanmıştır. Çalışmanın bulguları bahsedilen değerlendirme kriterleri ile belirlenmiş ve Şekil 2'de verilen akış şemasına göre gerçekleştirilmiştir.



Şekil 2. Çalışmanın akış şeması.

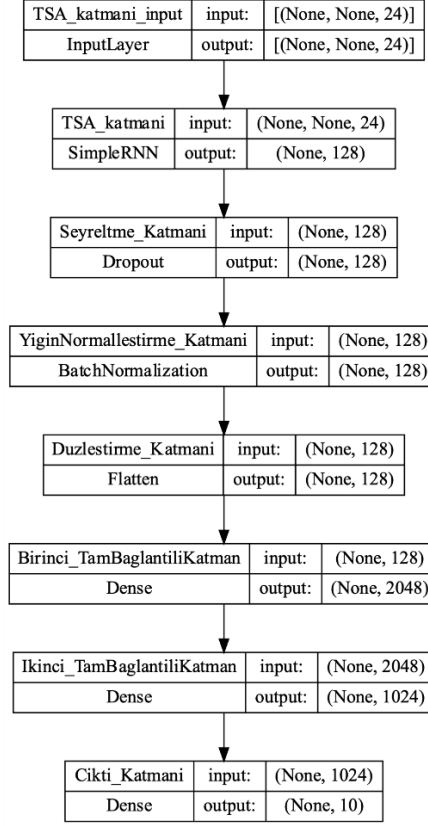
Şekil 2'de verilen akış şemasından da anlaşıldığı üzere çalışma dört aşamadan meydana gelmektedir. Birinci aşama (Verilerin Toplanması) içerisinde, kullanılan veri seti elde edilmiş ve 10 farklı mikroba ait çeşitli öznelikler kullanılmıştır. İkinci aşamada (Önilem) ise verilerin aynı değer aralığında olmasını sağlamak için normalleştirme işlemi yapılmış ve bunun için minimum-maksimum normalleştirmeye başvurulmuştur. Üçüncü aşamada (Sınıflandırma), iki farklı derin öğrenme modeli tasarlanmıştır. Bu modeller UKSB ve TSA modelleridir. Önilem aşaması sonunda elde edilen veriler, tasarlanan bu derin öğrenme modelleriyle sınıflandırılmıştır. Çalışmanın son aşamasında ise (Değerlendirme) derin öğrenme modellerinin başarımları değerlendirilmiş ve bunun için doğruluk, kesinlik, hassasiyet, F1-skor ve EAA skoru değerlendirme ölçütlerine başvurulmuştur.

3. Uygulama Sonuçları

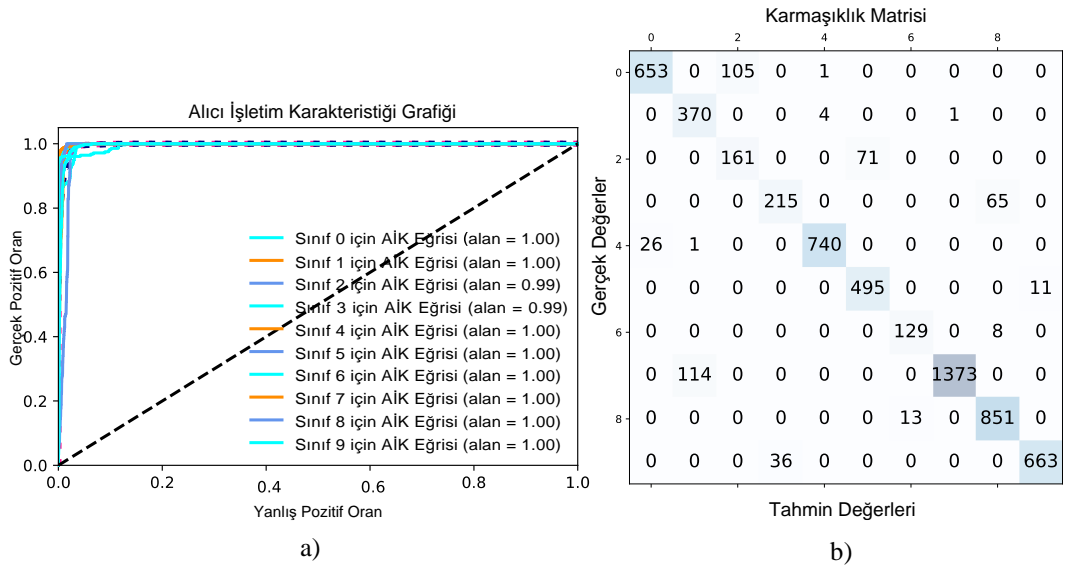
3.1. TSA modelinin uygulama sonuçları

Çalışma kapsamında geliştirilen TSA modelinin parametreleri deneme-yanılma yaklaşımı ile belirlenmiş ve en iyi sonuçlara göre değerlendirilmiştir. Modelin doğrulanması ise eğitim-test verisinin ayrıştırılmasına göre yapılmış ve verilerin %80'i eğitim, geri kalan %20'si test amacıyla kullanılmıştır. Eğitim işlemi 250 iterasyon ile gerçekleştirilmiştir. Modelin kaybı için kategorisel çapraz entropi kullanılmış ve en iyileme işlemi için ise Adam yöntemine başvurulmuştur. Şekil 3'te geliştirilen modelin parametreleri grafik üzerinde gösterilmiştir. Tablo 3'te TSA sınıflandırıcısının uygulama sonuçları verilmiştir. Tablo 3'te verilen bulgular doğrultusunda, TSA modeli ile %92,53 oranında bir doğruluk değeri elde edilmiştir. Bu model ile %88,66 oranında kesinlik değeri hesaplanırken,

bu modelin hassasiyet değeri ise %90,50 olmuştur. Bunlara ek olarak, F1-skor değeri %89,57 olarak hesaplanmıştır. Modelin EAA skoru ise 0,99 olmuş ve olağanüstü bir sınıflandırma gerçekleştirilmiştir. Şekil 4'te bu modelin AİK grafiği ve karmaşıklık matrisi verilmiştir.



Şekil 3. Geliştirilen TSA sınıflandırıcısının parametreleri.



Şekil 4. TSA modelinin değerlendirme grafikleri: (a) alıcı işletim karakteristiği, (b) karmaşıklık matrisi (Sınıf 0 *Aspergillus sp.*, Sınıf 1, *Diatom*, Sınıf 2 *Penicillium*, Sınıf 3 *Pithophora*, Sınıf 4 *Protozoa*, Sınıf 5 *Raizopus*, Sınıf 6 *Spirogyra*, Sınıf 7 *Ulohris*, Sınıf 8 *Volvox* ve Sınıf 9 *Yeast* etiketlerini ifade etmektedir.)

Tablo 3. TSA modelinin sınıflandırma sonuçları.

Doğruluk	Kesinlik	Hassasiyet	F1-Skor	EAA Skoru
%92,53	%88,66	%90,50	%89,57	0,99

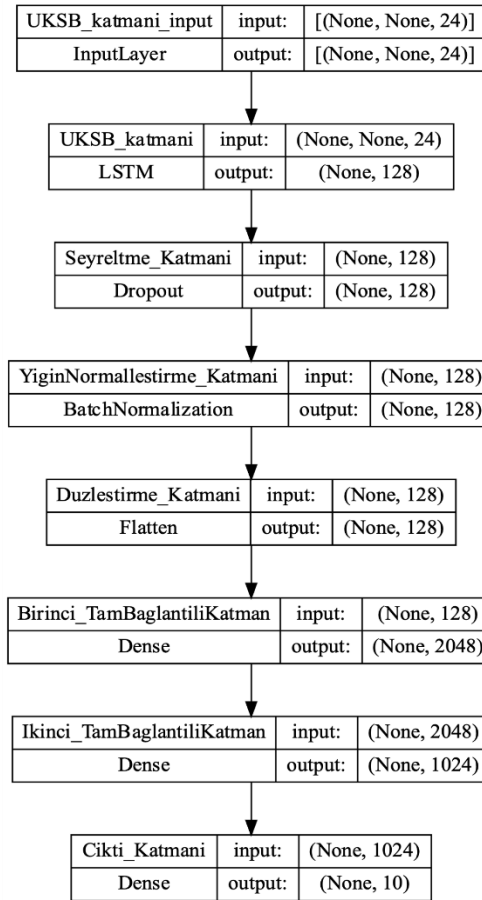
3.2. UKSB modelinin uygulama sonuçları

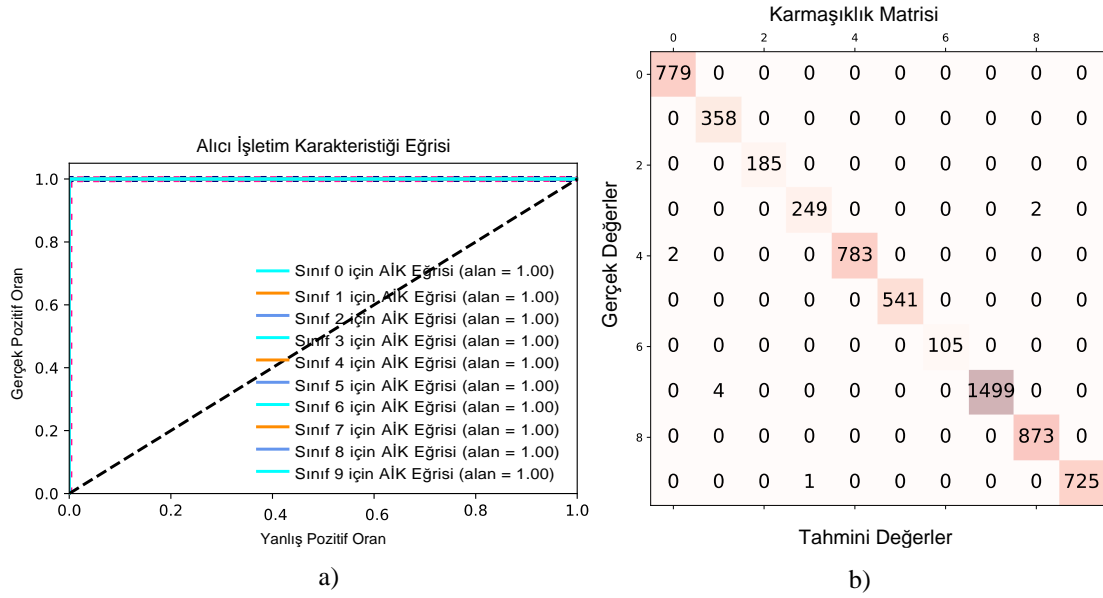
Çalışma kapsamında geliştirilen UKSB modelinin parametreleri deneme-yanılma yaklaşımı ile belirlenmiş ve en iyi sonuçlara göre değerlendirilmiştir. Modelin doğrulanması ise eğitim-test verisinin ayrıştırılmasına göre yapılmış ve verilerin %80'i eğitim, geri kalan %20'si test amacıyla kullanılmıştır. Eğitim işlemi 250 iterasyon ile gerçekleştirilmiştir. Modelin kaybı için kategorisel çapraz entropi kullanılmış ve en iyileme işlemi için ise RMSProp yöntemine başvurulmuştur. Şekil 5'te geliştirilen modelin parametreleri grafik üzerinde gösterilmiştir. Tablo 4'te UKSB sınıflandırıcısının uygulama sonuçları verilmiştir.

Tablo 4. UKSB modelinin sınıflandırma sonuçları.

Doğruluk	Kesinlik	Hassasiyet	F1-Skor	EAA Skoru
%99,85	%99,80	%99,85	%99,82	1,00

Tablo 4'te verilen bulgular doğrultusunda, UKSB modeli ile %99,85 oranında doğruluk değeri elde edilmiştir. Bu model ile %99,80 oranında kesinlik değeri elde edilirken, bu modelin hassasiyet değeri ise %99,85 olmuştur. Bunlara ek olarak, F1-skor değeri %99,82 olarak hesaplanmıştır. Modelin EAA skoru ise 1,00 olmuş ve olağanüstü bir sınıflandırma gerçekleştirilmiştir. Şekil 6'da bu modelin AİK grafiği ve karmaşıklık matrisi verilmiştir.

**Şekil 5.** Geliştirilen UKSB sınıflandırıcısının parametreleri.



- Farklı derin öğrenme yöntemlerinin kullanılması ve irdelenmesi gerekmektedir. Bu sayede bu çalışmada elde edilen bulgular daha sağlıklı bir şekilde yorumlanabilir.

5. Sonuç

Bu çalışmada çeşitli mikropların özellikleri kullanılarak sınıflandırılması yapılmıştır. Çalışma kapsamında mikroskopik görüntülerden elde edilen verilere ait veri seti kullanılmış ve sınıflandırma işlemi dört aşamada gerçekleştirilmiştir. Birinci aşamada veri setinden veriler elde edilmiştir. İkinci aşamada veriler ön işleme tabi tutulmuş ve minimum-maksimum normalleştirme işlemiyle verilerin değeri [0-1] aralığına indirgenmiştir. Üçüncü aşamada derin öğrenme modelleri tanımlanmış ve hem TSA hem de UKSB sınıflandırıcıları kullanılmıştır. Son aşamada ise bu sınıflandırıcıların performansları doğruluk, kesinlik, hassasiyet, F1-skor ve EAA skorları ile ölçülmüştür. TSA sınıflandırıcısıyla %92,53 oranında doğruluk, 0,99 oranında EAA skoru elde edilmiştir. UKSB ile yapılan sınıflandırma işleminde ise %99,85 oranında doğruluk elde edilirken, EAA skoru ise 1,00 olarak ölçülmüştür. EAA skorlarına göre değerlendirme yapıldığında iki sınıflandırıcının da olağanüstü bir sınıflandırma yaptığı gözlemlenmiştir. Bu çalışma ile derin öğrenme modellerinin mikropları sınıflandırmadaki başarımı gösterilmiş ve alanyazına katkı sağlanmıştır. Bunlara ek olarak, herhangi bir öznelik çıkarımı yapılmamış ve öznelik seçimine başvurulmamıştır. Bu dezavantajlara rağmen başarılı bir sınıflandırma işlemi gerçekleştirilmiştir. Sonraki çalışmalarda, mikropların türleri sınıflandırılacak ve türler ile mikroplar arasındaki biyoçeşitlilik araştırılacaktır.

Kaynaklar

- [1] Oka A, Sartor RB. Microbial-based and microbial-targeted therapies for inflammatory bowel diseases. *Dig Dis Sci* 2020; 65: 757-788.
- [2] Boxberger M, Cenizo V, Cassir N, Scola BL. Challenges in exploring and manipulating the human skin microbiome. *Microbiome* 2021; 9.
- [3] Yu K, Pieterse CMJ, Bakker PAHM, Berendsen RL. Beneficial microbes going underground of root immunity. *Plant Cell Environ* 2019; 42: 2860-2870.
- [4] Schirmer M, Garner A, Vlamakis H, Xavier RJ. Microbial genes and pathways in inflammatory bowel diseases. *Nat Rev Microbiol* 2019; 17: 497-511.
- [5] Björk JR, Diez-Vives C, Astudillo-Garcia C, Arcihe EA, Montaya JM. Vertical transmission of sponge microbiota is inconsistent and unfaithful. *Nat Ecol Evol* 2019; 3: 1172-1183.
- [6] Buszewski B, Rogowska A, Pomastowski P, Zloch M, Railean-Plugaru V. Identification of microorganisms by modern analytical techniques. *J AOAC Int* 2017; 100(6): 1607-1623.
- [7] Ganegoda S, Chinthaka SDM, Manage PM. Geosmin contamination status of raw and treated waters in Sri Lanka. *J Natl Sci Found Sri Lanka* 2019; 47(2): 245-259.
- [8] Ferone M, Gowen A, Fanning S, Scannell AGM. Microbial detection and identification methods: Bench top assays to omics approaches. *Compr Rev Food Sci Food Saf* 2020; 19(6): 3106-3129.
- [9] Dhindsa A, Bhatia S, Agrawal S, Sohi BS. An improvised machine learning model based on mutual information feature selection approach for microbes classification. *Entropy* 2021; 23(2).
- [10] Promdaen S, Wattuya P, Sanevas N. Automated microalgae image classification. *Procedia Comput Sci* 2014; 29: 1981-1992.
- [11] Coltelli P, Barsanti L, Evangelista V, Frassanito AM, Gualtieri P. Water monitoring: automated and real time identification and classification of algae using digital microscopy. *Environ Sci Process Impacts* 2014; 16(11): 2656 – 2665.
- [12] Dhindsa A, Bhatia S, Agrawal S, Sohi BS. Dataset for efficient microbes classification system. *Mendeley Data* 2021; 4.
- [13] Cui Z, Ke R, Pu Z, Wang Y. Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction. *arXiv* 2019; 1-11.
- [14] Ribeiro AH, Tiels K, Aguirre LA, Schön TB. Beyond exploding and vanishing gradients: analysing RNN training using attractors and smoothness. In: 23rd International Conference on Artificial Intelligence and Statistics; 26-28 August 2020; Online: pp. 2370-2380.
- [15] Takeuchi D, Yatabe K, Koizumi Y, Oikawa Y, Harada N. Real-time speech enhancement using equilibrated RNN. In: International Conference on Acoustics, Speech, and Signal Processing; 04-08 May 2020; Barcelona, Spain.
- [16] Yu Y, Si X, Hu C, Zhang J. A review of recurrent neural networks: LSTM cells and network architectures. *Neural Comput* 2019; 31(7): 1235-1270.
- [17] Rosset S. Model selection via the AUC. In: 21st International Conference on Machine Learning; 04-08 July 2004; New York.
- [18] Hosmer DW, Lemeshow S, Sturdivant RX. Applied Logistic Regression. 3rd ed. John Wiley and Sons, Inc.
- [19] Mahmud M, Kaiser MS, McGinnity TM, Hussain A. Deep learning in mining biological data. *Cognit Comput* 2021; 13: 1-33.