



Kuru Üzüm Tanelerinin Sınıflandırılması için Hibrit Bir Yaklaşım

A Hybrid Approach for Raisin Grains Classification

¹Serhat KILIÇARSLAN 

¹Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bandırma/Balıkesir, Türkiye

¹skilicarslan@bandirma.edu.tr

Araştırma Makalesi/Research Article

ARTICLE INFO

Article history

Received : 8 March 2022

Accepted : 4 April 2022

Keywords:

Raisin Grains, Deep Learning, Classical Data Mining, Hybrid, Classification

ABSTRACT

Raisin grains are a very good source of energy and nutrients due to the minerals and vitamins they contain. Since raisins constitute 23% of the world in Turkey, it is important to determine the variety and quality. Traditionally, deciding on the type and quality of raisin grains is a long and costly process. In addition, experience and knowledge are important in determining the type of raisin grains. In addition, since each of the experts has different experience, knowledge and expertise, there is a difference in terms of classification of raisins. Therefore, it is important to develop a decision support system with data mining methods for the correct classification of raisin grains. In this study, we propose one hybrid model using rotation forest (RO) and deep learning algorithms of Stacked Autoencoder (SAE) for the prediction of the type of raisin grains. As a result of the experimental evaluation, the hybrid SAE-ROF method has achieved a high success rate of 91.50% in terms of performance from the classical data mining methods and deep learning methods used in the study.

© 2022 Bandırma Onyedi Eylül University, Faculty of Engineering and Natural Science. Published by Dergi Park. All rights reserved.

MAKALE BİLGİSİ

Makale Tarihleri

Gönderim : 8 Mart 2022

Kabul : 4 Nisan 2022

Anahtar Kelimeler:

Kuru Üzüm Taneleri, Derin Öğrenme, Klasik Veri Madenciliği, Hibrit, Sınıflandırma

ÖZET

Kuru üzüm içeriğinde barındırdığı mineraller ve vitaminlerden kaynaklı çok iyi bir enerji ve besin kaynağı olarak karşımıza çıkmaktadır. Dünya genelindeki kuru üzümün %23'ü Türkiye'deki topraklardan elde edilmektedir. Geleneksel olarak, kuru üzümün cinsine ve kalitesine karar vermek uzun ve maliyetli bir süreçtir. Ayrıca, kuru üzümün cinsinin belirlenmesinde tecrübe ve bilgi birikimler önem arz etmektedir. Bu nedenle, kuru üzümün doğru sınıflandırılabilmesi açısından veri madenciliği yöntemleri ile karar destek sisteminin geliştirilmesi önemlidir. Bu çalışmada, kuru üzüm tanelerinin türünün tahmini için rotasyon orman (RO) ve yığılanmış otokodlayıcı (YOK) derin öğrenme algoritmalarını kullanan bir hibrit model öneriyoruz. Deneysel değerlendirme sonucunda, hibrit YOK-RO yöntemi çalışmada kullanılan klasik veri madenciliği yöntemleri ile derin öğrenme yöntemlerinden performans açısından %91.50 ile yüksek başarı elde edilmiştir.

© 2022 Bandırma Onyedi Eylül Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi. Dergi Park tarafından yayınlanmaktadır. Tüm Hakları Saklıdır.

1. GİRİŞ

Kuru üzüm, besin değerleri açısından kalsiyum, potasyum, demir, A ve C gibi çok önemli mineralleri ve vitaminleri bünyesinde barındırmaktadır. Kuru üzüm bünyesinde bol miktarda karbonhidrat bulunduğundan dolayı iyi bir enerji kaynağı olarak karşımıza çıkmaktadır [1]. Uluslararası sert kabuk ve kuru meyveler konseyi (INC) 2019-2020 yılları arasındaki verilere göre dünyada 1.34 milyon ton kuru üzüm üretilirken, Türkiye’de bu rakam 305 bin ton kuru üzüm üretimini temsil etmektedir.

Geleneksel olarak, kuru üzümün cinsine ve kalitesine karar vermek hem uzun hem de maliyetli bir süreçtir. Ayrıca, kuru üzümün cinsinin belirlenmesinde tecrübe ve bilgi birikimler önem arz etmektedir. Ek olarak, uzmanların her birinin farklı tecrübe, bilgi birikim ve uzmanlık dallarına sahip olmalarından dolayı kuru üzümlerin sınıflandırılması açısından farklılık meydana gelmektedir. Bu nedenle, kuru üzümün doğru sınıflandırılabilmesi için veri madenciliği yöntemleri ile karar destek sisteminin geliştirilmesi önem arz etmektedir. Çalışmada, kuru üzüm tanelerinin cinsini belirlemek üzere yığınlanmış oto kodlayıcı (YOK) algoritması ve klasik veri madenciliği algoritmasından rotasyon orman (RO) ile birleştirilerek hibrit YOK-RO algoritması önerilmiştir.

Literatürde, kuru üzümün sınıflandırılması üzerine sınırlı sayıda çalışma olup sırasıyla, yapay sinir ağları [2], destek vektör makinesi (DVM) [3], karar ağaçları (KA) [4] gibi klasik veri madenciliği algoritmaları kullanılmıştır. Okamura ve arkadaşlarının (1993) yaptıkları çalışmada, kuru üzüm tanelerinin sınıflandırmasını bayes algoritmasını kullanarak gerçekleştirdiklerini bildirmişlerdir. Wang ve arkadaşlarının (2012) yaptıkları çalışmada, kuru üzüm tanelerini doğru şekilde sınıflandırabilmek üzere beş farklı algoritma kullanılmıştır. En iyi sınıflandırma başarısı DVM ile elde ettiklerini bildirmişlerdir [5]. Yu ve arkadaşlarının (2012) yaptıkları çalışmada, üzüm kalitelerinin sınıflandırabilmek için en küçük kareler destek vektör makinesi (EKK-DVM) sınıflandırma algoritması kullanmışlardır. Önerdikleri yöntem EKK-DVM ile kuru üzümün kalitesinin yüksek performans ile sınıflandırıldığını bildirmişlerdir [6]. Karime ve arkadaşlarının (2017) yaptıkları çalışmada, kuru üzümlerin kalitesini belirlemek ve sınıflandırmak üzerine temel bileşen analizi (TBA) kullanılarak optimum özellikler elde edilmiştir. Elde edilen yeni özniteliklere YSA ve DVM algoritmalarını kullanarak sınıflandırma başarısını incelemişlerdir. En iyi sınıflandırma başarısının DVM ile elde ettiklerini bildirmişlerdir [2]. Çınar ve arkadaşlarının (2020) yaptıkları çalışmada, kuru üzümlerin kalitesini belirlemek ve sınıflandırmak üzerine lojistik regresyon (LR), YSA ve DVM algoritmaları kullanarak sınıflandırma işlemi gerçekleştirmişlerdir. Elde ettikleri sınıflandırma sonuçlarında, DVM ile en iyi başarı oranı elde ettiklerini bildirmişlerdir [3]. Tarakcı ve Özkan (2021) yaptıkları çalışmada, UCI’den elde ettiği kuru üzüm veri kümeleri üzerinde KNN ve ağırlıklandırılmış KNN ile deneysel değerlendirme gerçekleştirmiştir. Elde ettikleri değerlendirme sonuçlarında KNN algoritmasının en iyi başarı oranı elde ettiklerini bildirmişlerdir [30]. Köklu ve arkadaşlarının (2021) yaptıkları çalışmalarda, yedi farklı türdeki kuru meyvelerin tiplerinin sınıflandırılması için makine öğrenme yöntemlerinden lojistik regresyon ve yapay sinir ağı kullanılarak sınıflandırma işlemi gerçekleştirilmiştir. Elde edilen sonuçlar ışığında en iyi başarılı sınıflandırmanın YSA ile olduğunu bildirmişlerdir [31]. Köklu ve arkadaşlarının (2021) yaptıkları çalışmada, kabak çekirdeklerini sınıflandırılması için makine öğrenmesi yöntemlerinden LR, DVM, ROF ve KNN algoritmalarını kullanmışlardır. Deneysel değerlendirmelerin sonucunda en iyi başarının DVM yöntemi ile olduğunu bildirmişlerdir [32].

Bu çalışmada, Türkiye’de elde edilmiş Besni ve Keçimen cinsi kuru üzümleri sınıflandırılması için hibrit yığınlanmış oto kodlayıcı (YOK) ve rotasyon orman (RO) algoritmasının (YOK-RO) önerilmektedir. Ayrıca kuru üzüm veri kümesi üzerine, önerilen YOK-RO’nın yanında sırasıyla YSA, DVM, LR, rastgele orman (RAO), rastgele ağaç (RA), RO ve derin sinir ağları (DSA) algoritmalarıyla analizler gerçekleştirilerek performansları karşılaştırılmıştır. Analizler sonucunda, hibrit YOK-RO yöntemi diğer veri madenciliği yöntemleri ve derin öğrenme yöntemine göre performans sonuçları açısından daha iyi olduğu gözlemlenmiştir. YOK modelinde, veri kümesindeki gizli özelliklerin derinlemesine araştırılabilmektedir. YOK, veri kümesi üzerinden özellik çıkartma konusunda iyi olmasına rağmen, sınıflandırma performansı istenilen düzeyde değildir. Bundan dolayı da YOK ile veri kümesinden derinlemesine gizli özellikler elde edildikten sonra, sınıflandırma algoritması için RO tercih edilmiştir. Böylece, kullanılan veri kümesi üzerinde derin öğrenme yöntemi YOK ile derinlemesine incelenerek yüksek başarılı sınıflandırma gerçekleştirmesi sağlanmaktadır.

Makalenin geri kalanında, ikinci bölümde çalışmada kullanılan kuru üzüm veri kümesi ve yöntemler verilmiştir. Üçüncü bölümde, deneysel değerlendirme sonuçları ve tartışma verilmiştir. Son bölümde, sonuçlar ve gelecekte yapılacak çalışma sunulmuştur.

2. MATERYAL VE METOD

2.1. Kuru Üzüm Veri Kümesi

Çalışmada, kuru üzüm veri kümesini sınıflandırılması için klasik veri madenciliği yöntemleri ve derin öğrenme yöntemleri uygulanmıştır. Deneysel olarak kullanılan veri kümesi Çınar ve arkadaşları (2020) tarafından 2020 yılında toplanmış olup 450 adet Besni ve 450 adette Keçimen cinsi üzümünden toplam 900 örnek ve 7 öznitelik bilgisine sahip kuru üzüm verisinden oluşmaktadır [3]. Tablo 1’de kuru üzümünden elde edilen veri kümesine ait özniteliklerin açıklamaları ve tanımlayıcı istatistiksel sonuçları gösterilmektedir.

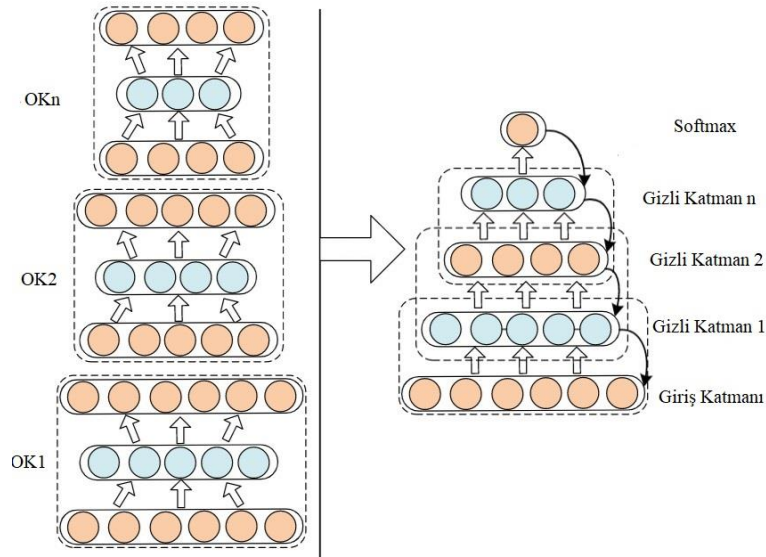
Tablo 1. Kuru üzüm veri kümesine ait özelliklerin tanımlayıcı istatistiksel sonuçları.

Öznitelikler	Açıklama	Mean	Std	Min	Max
Alan	Kuru üzümün sınırları içindeki piksel sayısı	79.659	35.314	25.387	235.047
Ana Eksen Uzunluğu	Üzüm üzerine çizilebilecek en uzun çizgi olan ana eksenin uzunluğu verir	405	106	226	997
Küçük Eksen Uzunluğu	Üzüm üzerine çizilebilecek en kısa çizgi olan ana eksenin uzunluğu verir	246	47	144	492
Eksantriklik	Kuru üzüm ile aynı momentlere sahip olan elipsin eksantrikliğinin bir ölçüsünü verir	0.77	0.086	0.34	0.96
Dış Bükey Alan	Kuru üzümün oluşturduğu bölgenin en küçük dışbükey kabuğunun piksel sayısını verir	82.621	36.877	26.139	278.217
Genişlik	Kuru üzümün oluşturduğu bölgenin sınırlayıcı kutudaki toplam piksel sayısına oranını verir	0.7	0.048	0.37	0.83
Çevre	Kuru üzümün oluşturduğu bölgenin sınırlayıcı kutudaki toplam piksel sayısı oranı	1104	251	619	2698
Sınıf	Besni ve Keçimen kuru üzüm sınıfı	0.33	0.47	0	1

Bu çalışmada, önerilen hibrit YOK-RO algoritması ile sırasıyla YSA, DVM, LR, RAO, RA, RO ve DSA algoritmalarıyla analizler gerçekleştirilerek performansları karşılaştırılmıştır.

2.2. Yığınlanmış Oto Kodlayıcı (YOK)

Otokodlayıcı (OK), yapay sinir ağlarının (YSA) bir çeşidi olup, giriş, gizli ve çıkış katmanlarından meydana gelmektedir. Eğitim sürecinde YSA gibi geri yayılım algoritmasını kullanan OK modeli, denetimsiz bir öğrenme gerçekleştirerek giriş verisini çıkış verisi olarak tanımlamamıza olanak tanıyan derin öğrenme yöntemidir. OK'lar çok karmaşık veri kümeleri üzerinde uygulanarak yüksek başarı performansı elde edilmesine imkan tanımaktadır. OK mimarisi, kodlayıcı ve kod çözücü olmak üzere iki birimden meydana gelmektedir. Ardı ardına OK'lar birbirine bağlanmasıyla, yani birinci OK'un gizli katmanı ikinci OK'un giriş katmanına bağlanarak yığınlanmış oto kodlayıcı (YOK) mimarisi Şekil 1'deki gibi ortaya çıkmaktadır [7-9]. Yığınlanmış oto kodlayıcıları denetimsiz yapıdan denetimli öğrenme yapısına geçişi için çıkış katmanına softmax fonksiyonu bağlanarak denetimli yapıya geçilerek sınıflandırma işlemini gerçekleştirebilmektedir [7]. Kullanılan YOK mimarisinin hiper-parametreleri Tablo 2'de gösterilmiştir. Kodlayıcı katmanından sıkıştırılmış olan veri kümesi kod çözücü katmanına ulaşırken Eşitlik 1'deki gibi hesaplanır. Kod çözücü katmanında, elde edilen gürültüden arındırılmış yeni giriş verileri elde edilebilmesi için Eşitlik 2'deki gibi hesaplama yapılarak gerçekleştirilir. Eşitlik 1 ve 2'de f aktivasyon fonksiyonunu, $\mathbf{x}_a - \mathbf{y}_b$ 'nin $\mathbf{a} - \mathbf{b}$ 'inci nöron değeri, gizli katmandaki \mathbf{y}_b 'nin \mathbf{b} 'inci nöronuna aktarılan değeri, \mathbf{n} nöron sayısı, girdi katmanındaki $\mathbf{w}_{ab} - \mathbf{w}_{ba}$ 'nın \mathbf{a} nöronundan \mathbf{b} nöronuna giden ağırlığı temsil etmektedir.

**Şekil 1.** Yığınlanmış oto-kodlayıcı mimarisi.

$$y_b = f(\sum_{a=1}^n x_a \times w_{ab}) + bias \quad (1)$$

$$x'_a = f(\sum_{b=1}^m y_b \times w_{ba}) + bias \quad (2)$$

Tablo 2. YOK hiper-parametreleri (SAE Hyper-Parameters).

Hiper-Parametreler	SAE Mimari
OK1 Gizli Nöron	8
OK2 Gizli Nöron	5
Kayıp fonksiyonu	Crossentropy
Optimizer	Adam
Aktivasyon fonksiyonu	ReLU
Öğrenme oranı	0.001
L2 Reg.	0.001
Eğitim-tekrar sayısı	100
Seyreklik oranı	0.15
Seyrek Reg.	10, 20

Tablo 2’de iki adet oto kodlayıcının art arda bağlanması sonucunda oluşan YOK mimarisinin hiper-parametreleri görülmektedir. Kurulan modelde birinci oto kodlayıcı 8 gizli nöron, ikinci oto kodlayıcı 5 gizli nörondan oluşturularak deneysel değerlendirme gerçekleştirilmiştir. Her bir oto kodlayıcı 100 tekrar ile çalıştırılmıştır. YOK’ün çalışmasında çapraz entropi maliyet fonksiyonu kullanılmıştır [8-13]. Seyreklik oranı (Sparsity Proportion) değeri 0.15 alınmıştır. Sparsity Proportion değeri, seyrek reg.’in bir parametresi olup çıktının seyrekliğini ayarlamak için kullanılmaktadır [14-17]. L2Reg., seyrek reg. değerleri eğitim aşamasında modelin ezberlemesini azaltmak ve daha iyi başarı sonucu vermek için kullanılmaktadır [18-21].

2.3. Rotasyon Orman Algoritması (RO)

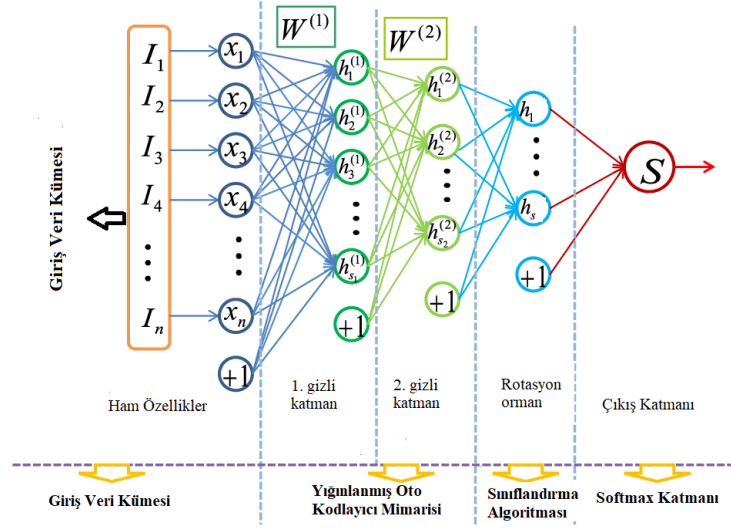
Rotasyon orman (RO) algoritması güçlü kolektif tabanlı yeni nesil algoritmadır. RO ile daha az ağaç yapısı kullanılarak daha yüksek başarı oranları elde etmek üzerine çalışmaktadır. Ayrıca, mimari olarak rastgele orman (RAO) algoritmasına benzemektedir. Algoritmada, mevcut veri kümesi içerisinde rastgele olarak veriler alınır ve alt gruplara ayrılır. Ayrılan alt grupların her birine ayrı ayrı temel bileşen analizi (TBA) uygulanarak yeni kovaryans değeri hesaplanmasıyla özellik çıkarımı gerçekleştirilir. Oluşturulan kovaryans değerlerine göre yeni rotasyon orman matrisi oluşturulur. Elde edilen yeni matrise, RO algoritması ile sınıflandırma işlemine tabi tutulmaktadır [22]. Böylelikle ortalama kombinasyon yöntemi ile her sınıfa ait veriler güven oyu alır ve sınıfı bilinmeyenlerden en yüksek güven değeri ile ilgili sınıfa ataması gerçekleştirmektedir [23-26]. TBA kullanılarak çalıştırılan RO yöntemi, özellikle seçilen tüm öznelikler gerçek değerlere sahip olduğundan dolayı geleneksel yöntemlere göre daha güçlü bir yapı kılmaktadır. Algoritma aşağıdaki adımlarla gerçekleştirilmektedir.

1. Adım: Mevcut veri kümesinden rastgele F bağımsız değişken belirlenen K sayıda alt kümeye ayrılmaktadır. Böylece F_{ij} alt kümesi oluşmaktadır.
2. Adım: Oluşturulan F_{ij} alt kümesi *bootstrap* yaklaşımı ile eğitim ve test olarak sırasıyla %75-%25 olarak ayrılır. Elde edilen veri setlerine TBA uygulanarak kovaryans matris C_{ij} hesaplanmaktadır.
3. Adım: Elde edilen C_{ij} değerleri kullanılarak rotasyon matrisi R_i oluşturulur. Bu adımlar tüm sınıflandırıcılar için tekrar edilir. Sınıflandırma işlemi Eşitlik 3’de verilmiştir. Eşitlikteki L topluluktaki sınıflandırıcının sayısını temsil etmektedir. Çalışmada $K = 3$ ve $L = 10$ olarak kabul edilmiştir.

$$\mu_j(x) = \frac{1}{L} \sum_{i=1}^L d_{i,j}(xR_i) \quad j = 1, \dots, c \quad (3)$$

2.4. Önerilen Hibrit Yığınlanmış Oto Kodlayıcılı Rotasyon Orman (YOKRO) Algoritması

Klasik veri madenciliğinde, veri kümesi üzerinde sınıflandırma işlemi gerçekleştiren modeller, yaygın olarak manuel özellik çıkarımı gerçekleştirerek sınıflandırma yapmaktadır. Elde edilen özelliklerin sınıflandırma üzerine pozitif etkisinin olmasıyla modelin dah iyi eğitilmesine yardımcı olacak Ancak, derin öğrenme algoritmaları sayesinde, özelliklerin manuel bir şekilde oluşturulması yerine, yinelemeli öğrenme özelliği ile veri kümesindeki gizli özelliklerin derinlemesine araştırılabilir. YOK, veri kümesi üzerinden özellik çıkartma konusunda iyi olmasına rağmen, sınıflandırma performansı istenilen düzeyde değildir. Bundan dolayı, YOK ile RO algoritmalarını Şekil 2’deki gibi hibrit bağlanmasıyla daha iyi sınıflandırma performansı elde edilebilmektedir. Bu işlemleri, birincisi denetimsiz öğrenme olan YOK ile özellik çıkartımı sağlanır ve ikinci adımda denetimli bir algoritma RO ile ince ayar sürecine girilerek sınıflandırma işlemi gerçekleştirilmektedir [7].

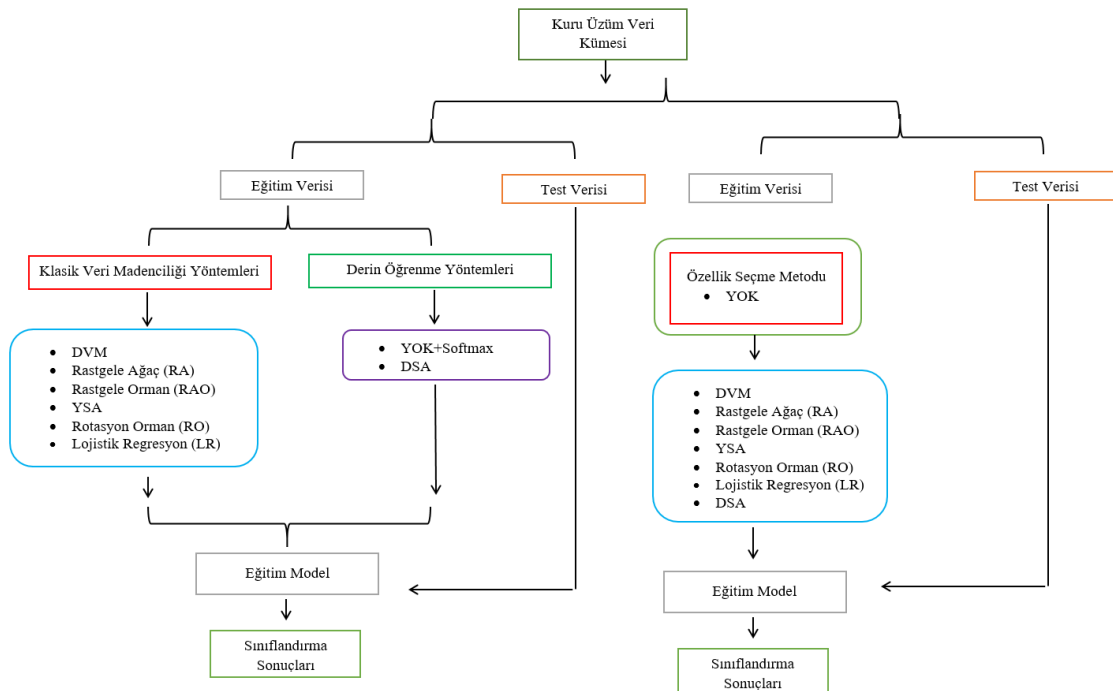


Şekil 2. Önerilen hibrit YOK-RO algoritması.

Bu çalışmada, kuru üzümde ait verilerinin sınıflandırılması için derin öğrenme algoritmalarında en iyi başarı oranına sahip YOK-RO modeli kullanılmıştır. Öncelikle, veri kümesinde anlamlı öznelikleri bulabilmek için YOK uygulanmıştır. Sınıflandırma sonuçlarının geliştirilmesi için bir sonraki aşamada da kolektif tabanlı RO algoritması kullanılmıştır. Önerilen modelin blok diyagramı Şekil 3'teki gibidir. Çalışmada kullanılan DSA mimarisi, Tablo 3'te görülmektedir. Diğer klasik veri madenciliği yöntemlerinin hiper-parametre değerleri default olarak seçilerek deneysel değerlendirmeler yapılmıştır. Deneysel değerlendirmeler python-keras kütüphanesi kullanılarak gerçekleştirilmiştir.

Tablo 3. DSA mimarisi hiper-parametreleri.

DSA Mimari				
Gizli Katman Sayısı	50 Nöron	100 Nöron	200 Nöron	300 Nöron
Öğrenme Oranı	0.001			
Aktivasyon Fonk.	ReLU			
Kayıp Fonksiyon	categorical_crossentropy			
Optimizör	SGD			
Çıkış Activastion Function	Softmax			



Şekil 3. Önerilen modelin blok diyagramı.

2.5. Değerlendirme Kriterleri

Çalışmada kullanılan veri kümesindeki veriler kuru üzüm cinsine göre Besni ve Keçimen şeklinde iki farklı sınıf ile ifade edilmektedir. Çalışmada k-fold=10 çapraz doğrulama yöntemi kullanılarak veri kümesi deneysel değerlendirmeler gerçekleştirilmiştir. 10 katlı çapraz doğrulama yönteminde veri seti 10 parçaya bölünür. Model eğitimi için 9 kısım kullanılırken, kalan 1 kısım model testi için kullanılmaktadır. Çalışmada çeşitli algoritmaların performansları karşılaştırılmıştır ve bu algoritmaların test doğrulukları verilmiştir. Çalışmada sınıflandırma algoritmalarının performansları test doğruluk, kesinlik, duyarlılık, f-skor kriterlerine göre karşılaştırılmıştır (sırasıyla Eşitlik 3, 4, 5, ve 6) [22]. Formüllerdeki TP gerçek pozitif, FP gerçek negatif, TN yanlış pozitif ve FN ise yanlış negatif ifade etmektedir [22].

$$\text{Doğruluk} = \frac{TP+TN}{TP+FP+TN+FN} \quad (3)$$

$$\text{Kesinlik} = \frac{TP}{TP+FP} \quad (4)$$

$$\text{Duyarlılık} = \frac{TP}{TP+FN} \quad (5)$$

$$F - \text{skor} = \frac{2 * \text{Kesinlik} * \text{Duyarlılık}}{\text{Kesinlik} + \text{Duyarlılık}} \quad (6)$$

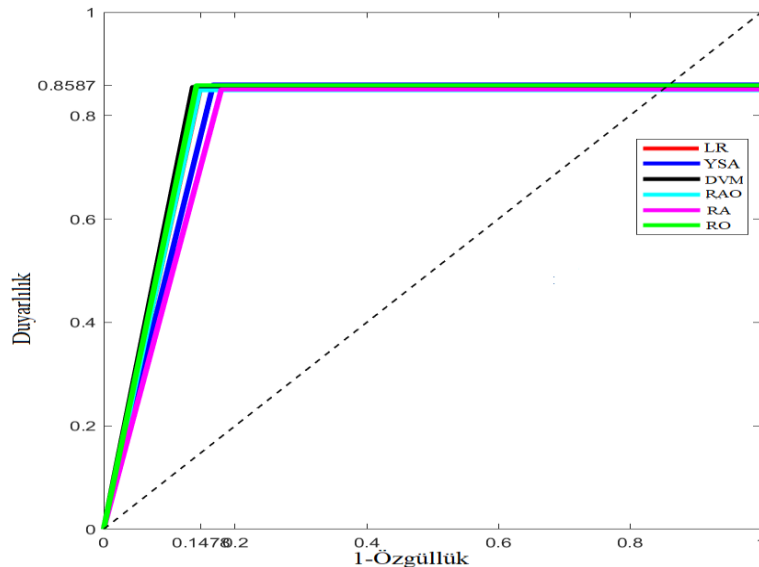
3. DENEYSEL SONUÇLAR VE TARTIŞMA

Çalışmada derin öğrenme ve klasik veri madenciliği yöntemlerinden oluşan hibrit modeller ile deneysel analizler gerçekleştirilmiştir. Bu bölümde, kuru üzüm veri kümesi üzerinden sınıflandırma işlemini gerçekleştirebilmek için hibrit yöntem YOK-RO kullanılarak elde edilen deneysel sonuçlar sunulmuştur. Çalışmada, YOK-softmax, DSA, RAO, YSA, LR ve RO algoritmalarının deneysel değerlendirmelerde Tablo 4 ve 5’de sunulmuştur. Ek olarak, YOK ile hibrit yapılmış RAO, YSA, LR, RO ve DSA algoritmalarının deneysel sonuçları Tablo 6’da sunulmuştur. Çalışmada kullanılan yöntemlere ait test modelleri ve parametreleri Tablo 4, 5 ve 6’da görülmektedir.

Çalışmada, başlangıçta klasik veri madenciliği yöntemleri kullanılarak kuru üzüm veri kümesinin sınıflandırılmasına bakılmıştır. Sınıflandırmaya ait sonuçlar Tablo 4’de gözlemlenmektedir. Şekil 3, 4 ve 5’de görüldüğü gibi deneysel değerlendirme amacıyla kullanılan yöntemlerine göre ortalama duyarlılık ve özgülük değerleriyle ROC eğrisi çizilmiştir.

Tablo 4. Kuru üzüm veri kümesine klasik veri madenciliği yöntemi ile değerlendirme.

Model	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F-Skor(%)	MSE
Lojistik Regresyon (LR)	85.22	84.10	86.90	85.50	0.2113
Yapay Sinir Ağı (YSA)	86.33	84.60	88.90	86.70	0.1954
Destek Vektör Makinesi (DVM)	86.44	84.20	89.80	86.90	0.1356
Rastgele Orman (RAO)	85.44	83.30	88.70	85.90	0.1954
Rastgele Ağaç (RA)	82.11	82.30	81.80	82.10	0.1789
Rotasyon Orman (RO)	85.88	84.30	88.20	86.20	0.2186

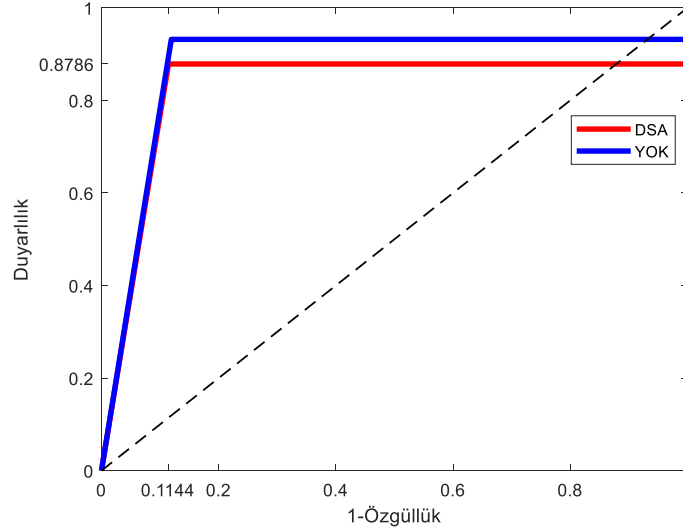


Şekil 3. Klasik veri madenciliği yöntemleriyle ilgili deneysel çalışmaların Roc eğrisi.

Tablo 4’de kuru üzüm veri kümesi üzerinde en iyi başarı oranı DVM algoritması ile doğruluk %86.44, kesinlik %84.60, duyarlılık %88.90, f-skor %86.70 ve MSE 0.1356 oranları elde edildiği görülmüştür. Şekil 3’de, ROC eğrisinde DVM’nin diğer algoritmalara göre daha başarılı bir sınıflandırma gerçekleştirdiği görülmüştür. Derin öğrenme yöntemlerinden DSA ve YOK-Softmax kullanılarak kuru üzüm veri kümesinin sınıflandırılmasına da bakılmıştır. Sınıflandırmaya ait sonuçlar Tablo 5’de sunulmuştur.

Tablo 5. Kuru üzüm veri kümesine derin öğrenme yöntemi ile değerlendirme.

Model	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F-Skor(%)	MSE
DSA	88.56	86.80	90.80	88.70	0.1854
Yığılmış Otokodlayıcı (YOK)	88.10	84.40	92.00	84.60	0.1269



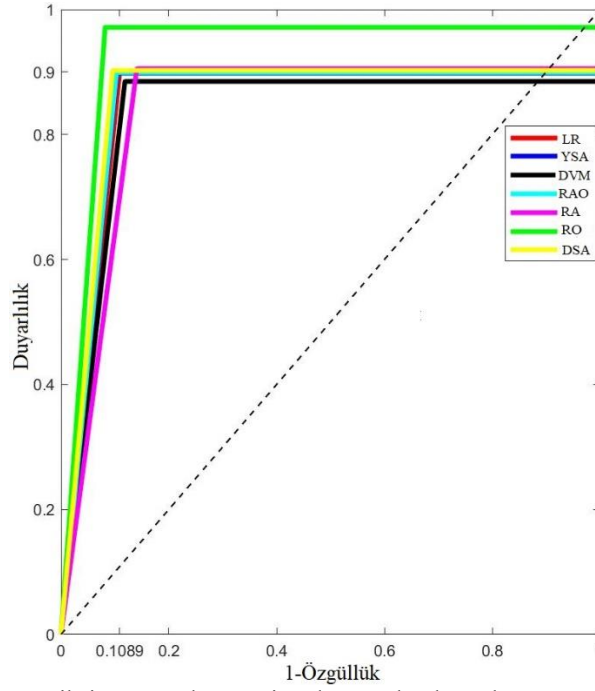
Şekil 4. Derin öğrenme yöntemleri üzerine deneysel çalışmaların Roc eğrisi.

Tablo 5’te kuru üzüm veri kümesi üzerinde k-fold=10 ile en iyi başarı oranı DSA algoritması ile doğruluk %88.56, kesinlik %86.80, duyarlılık %90.80, f-skor 0.8870 oranları elde edildiği görülmüştür. Şekil 4’te, ROC eğrisinde DSA’nın YOK’e göre daha başarılı bir sınıflandırma gerçekleştirdiği görülmüştür

Çalışmada, derin öğrenme yöntemlerinden YOK ile hibrit yapılmış RAO, YSA, LR, RO ve DSA algoritmalarının deneysel sonuçları sunulmuştur. Kuru üzüm veri kümesi üzerine uygulanan her bir algoritmanın sonuçları Tablo 6’da görülmektedir.

Tablo 6. Kuru üzüm veri kümesinin YOK ile hibrit yöntemlerin sonuçları.

Model	Doğruluk(%)	Kesinlik(%)	Duyarlılık(%)	F-Skor(%)	MSE
Lojistik Regresyon (LR)	89.11	87.60	88.40	82.80	0.1745
Yapay Sinir Ağı (YSA)	89.62	89.40	95.80	92.50	0.1482
Destek Vektör Makinesi (DVM)	88.14	88.00	95.20	91.50	0.1185
Rastgele Orman (RAO)	89.70	89.50	95.80	92.50	0.1481
Rastgele Ağaç (RA)	85.92	89.50	89.30	89.40	0.1407
Rotasyon Orman (RO)	91.50	91.80	93.70	91.23	0.1068
DSA	90.41	89.90	96.40	93.10	0.1197



Şekil 5. Hibrit yöntemler üzerine deneysel çalışmaların Roc eğrisi.

Çalışmada, Tablo 6'da YOK ile hibrit yapılmış RAO, YSA, LR, RO ve DSA algoritmalarının deneysel sonuçları görülmektedir. Eğitim sonucunda en iyi başarı oranı hibrit YOK-RO modelinden alınmıştır. YOK modelinde, veri kümesindeki gizli özelliklerin derinlemesine araştırılabilmektedir. YOK, veri kümesi üzerinden özellik çıkartma konusunda iyi olmasına rağmen, sınıflandırma performansı istenilen düzeyde değildir. Bundan dolayı da YOK ile veri kümesinden derinlemesine gizli özellikler elde edildikten sonra, sınıflandırma algoritması için RO tercih edilmiştir. Böylece, kullanılan veri kümesi üzerinde derin öğrenme yöntemi YOK ile derinlemesine incelenerek yüksek başarılı sınıflandırma gerçekleştirilmesi sağlanmıştır. Hibrit YOK-RO algoritması ile doğruluk %91.50, kesinlik %91.80, duyarlılık %90.70, f-skor %91.23 ve MSE 0.1068 oranları elde edildiği görülmüştür. Şekil 5'de, ROC eğrisinde YOK-RO yönteminin diğer yöntemlere göre daha başarılı bir sınıflandırma gerçekleştirdiği görülmüştür.

Yapılan tüm deneylerde en iyi başarı oranı YOK-RO ile elde edildiği görülmektedir. Önerilen hibrit modeller arasında ikinci en iyi başarı oranı YOK-DSA algoritmasından doğruluk %90.41, kesinlik %89.90, duyarlılık %96.40, f-skor %93.10 ve MSE 0.1197 oranları elde edildiği görülmüştür. YOK ile hibrit yapılan algoritmaların, normal hallerine göre daha iyi başarı oranları elde edildiği gözlemlenmektedir. Bunun en önemli sebebi, YOK algoritmasında, veri kümesindeki gizli özelliklerin derinlemesine araştırılması özelliği yatmaktadır. Böylece, YOK ile veri kümesindeki gizli özellikler gün yüzüne çıkartılarak sınıflandırma aşamasında da sınıflandırma algoritmalarının kullanılmasıyla başarının artırılacağı görülmüştür. Deneysel değerlendirmeler sonucunda, YOK-RO yüksek performanslı doğruluk oranı vermiş olması şartıdır. Önerilen YOK-RO yöntemi diğer yöntemlere göre gizli kalmış özellikleri gün yüzüne çıkartarak sınıflama problemlerinde iyi performans verdiği görülmektedir [23, 27-29].

4. SONUÇ

Geleneksel olarak, kuru üzümün cinsine ve kalitesine karar vermek uzun ve maliyetli bir süreçtir. Bu nedenle, kuru üzümün doğru sınıflandırılabilmesi açısından veri madenciliği yöntemleri ile karar destek sisteminin geliştirilmesi önem arz etmektedir. Çalışmada, kuru üzüm tanelerinin cinsini belirlemek üzere yığılmış oto kodlayıcı (YOK) algoritması rotasyon orman algoritması hibrit bir yöntem olarak önerilmiştir. Deneysel değerlendirmeler sonucunda, hibrit YOK-RO yöntemi tüm klasik veri madenciliği yöntemlerine kıyasla %91.50 başarı sonucuyla daha yüksek performans gösterdiği gözlemlenmektedir. Elde edilen sonuçlar değerlendirildiğinde YOK-RO, doğrusal olarak ayrılmayan sınıflama problemlerinde iyi performans verdiği görülmektedir. Ek olarak, benzer yöntemler arasında da global optimum değere en çok yakınsayan yöntem olduğu görülmüştür. Gelecekteki çalışmada, kuru meyve görüntüleri üzerinden, derin öğrenme yöntemleri kullanılarak hızlı karar verebilecek karar destek sistemi geliştirilmesi amaçlanmaktadır.

Yazar Katkıları

Yazar çalışmaya eşit oranlı katkı sunmuştur.

Çıkar Çatışması

Makale yazarları, aralarında herhangi bir çıkar çatışması olmadığını beyan ederler.

KAYNAKÇA

- [1] T. Uzun, B. Hallaç, A. Altın, and G. Kaya, "Midyat/Mardin ve Beşiri/Batman İlçelerinde Satışa Sunulan Bazı Kuru Üzüm Çeşitlerinin Fizikokimyasal Özelliklerinin Karşılaştırılması", Turkish Journal of Agricultural Engineering Research, vol.1, no. 2, pp. 404-414, 2020.
- [2] N. Karimi, R. R. Kondrood, and T. Alizadeh, "An intelligent system for quality measurement of Golden Bleached raisins using two comparative machine learning algorithms", Measurement, vol. 107, pp. 68-76, 2017.
- [3] İ. Çınar, M. Koklu, and P. D. Ş. Taşdemir, "Classification of Raisin Grains Using Machine Vision and Artificial Intelligence Methods", Gazi Mühendislik Bilimleri Dergisi (GMBD), vol. 6, no. 3, pp. 200-209, 2020.
- [4] K. Mollazade, M. Omid, and A. Arefi, "Comparing data mining classifiers for grading raisins based on visual features", Computers and electronics in agriculture, vol. 84, pp. 124-131, 2012.
- [5] S. Wang, K. Liu, X. Yu, D. Wu, and Y. He, "Application of hybrid image features for fast and non-invasive classification of raisin", Journal of food engineering, vol. 109, no. 3, pp. 531-537, 2012.
- [6] X. Yu, K. Liu, D. Wu, and Y. He, "Raisin quality classification using least squares support vector machine (LSSVM) based on combined color and texture features", Food and Bioprocess Technology, vol. 5, no. 5, pp. 1552-1563, 2012.
- [7] K. Adem, S. Kılıçarslan, and O. Cömert, "Classification and diagnosis of cervical cancer with stacked autoencoder and softmax classification", Expert Systems with Applications, vol. 115, pp. 557-564, 2019.
- [8] S. Kılıçarslan, K. Adem, and M. Celik, "Diagnosis and classification of cancer using hybrid model based on ReliefF and convolutional neural network", Medical hypotheses, vol. 137, no. 109577, 2020.
- [9] S. Kılıçarslan, M. Celik, and Ş. SAHİN, "Hybrid models based on genetic algorithm and deep learning algorithms for nutritional Anemia disease classification", Biomedical Signal Processing and Control, vol. 63, no. 102231, 2021.
- [10] M. Cui, Y. Wang, X. Lin, and M. Zhong, "Fault diagnosis of rolling bearings based on an improved stack autoencoder and support vector machine", IEEE Sensors Journal, vol. 21, no. 4, pp. 4927-4937, 2020.
- [11] O. Kaynar, Z. Aydın, and Y. Görmez, "Sentiment Analizinde Öznitelik Düşürme Yöntemlerinin Oto Kodlayıcı Derin Öğrenme Makinaları ile Karşılaştırılması", Bilişim Teknolojileri Dergisi, vol. 10, no. 3, pp. 319-326, 2017.
- [12] Y. Bengio, "Learning deep architectures for AI", Foundations and Trends in Machine Learning, vol. 2, no. 1, pp. 1-127, 2009.
- [13] W. Chen, S. Gou, X. Wang, X. Li, and L. Jiao, "Classification of PolSAR Images Using Multilayer Autoencoders and a Self-Paced Learning Approach", Remote Sensing, vol. 10, no. 1, pp. 1-17, 2018.
- [14] U. Erkan, "A precise and stable machine learning algorithm: eigenvalue classification (EigenClass)", Neural Computing and Applications, vol. 33, no. 10, pp. 5381-5392, 2021.
- [15] I. Pacal, and D. Karaboga, "A Robust Real-Time Deep Learning Based Automatic Polyp Detection System", Computers in Biology and Medicine, no. 104519, 2021.
- [16] I. Pacal, D. Karaboga, A. Basturk, B. Akay, & U. Nalbantoglu, "A comprehensive review of deep learning in colon cancer", Computers in Biology and Medicine, no. 104003, 2020.
- [17] F. O. Ozkok, and M. Celik, "A hybrid CNN-LSTM model for high resolution melting curve classification", Biomedical Signal Processing and Control, vol. 71, no. 103168, 2022.
- [18] K. Adem, and S. Kılıçarslan, "COVID-19 Diagnosis Prediction in Emergency Care Patients using Convolutional Neural Network", Afyon Kocatepe Üniversitesi Fen Ve Mühendislik Bilimleri Dergisi, vol 2. no. 2, pp. 300-309, 2021.
- [19] K. Adem, "Diagnosis of breast cancer with Stacked autoencoder and Subspace kNN", Physica A: Statistical Mechanics and its Applications, vol. 551, no. 124591, 2020.
- [20] M. A. Bülbül, and C. Öztürk, "Optimization, Modeling and Implementation of Plant Water Consumption Control Using Genetic Algorithm and Artificial Neural Network in a Hybrid Structure", Arabian Journal for Science and Engineering, vol. 47, no. 2, pp. 1-15, 2021.
- [21] S. Memiş, S. Enginoğlu, and U. Erkan, "Numerical data classification via distance-based similarity measures of fuzzy parameterized fuzzy soft matrices", IEEE Access, vol. 9, pp. 88583-88601, 2021.
- [22] S. Kılıçarslan, and M. Çelik, "Rotasyon orman sınıflandırma algoritması kullanarak kronik böbrek rahatsızlığının tahmini", Journal of Science and Technology of Dumlupınar University, no. 43, pp. 21-34, 2019.
- [23] J. J. Rodríguez, L. I. Kuncheva, and C. J. Alonso, "Rotation forest: A New classifier ensemble method", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 2810, pp. 1619-1630, 2006.
- [24] E. Akçetin, and U. Çelik, "İstenmeyen Elektronik Posta (Spam) Tespitinde Karar Ağacı Algoritmalarının Performans Kıyaslaması", Journal of Internet Applications & Management/İnternet

- Uygulamaları ve Yönetimi Dergisi, vol. 5, no. 2, pp. 1-20, 2014.
- [25] Ö. H. Namlı, and T. Özcan, “Makine Öğrenmesi Algoritmaları Kullanarak Gişe Hasılatının Tahmini”, *Yönetim Bilişim Sistemleri Dergisi*, vol. 3, no. 2, pp. 130-143, 2017.
- [26] D. Çıtak, and D. Sabancı, “ Response surface methodology and hydrophobic deep eutectic solvent based liquid phase microextraction combination for determination of cadmium in food and water samples”, *Journal of Food Measurement and Characterization*, vol. 15, no. 2, pp. 1843-1850, 2021.
- [27] S. Kiliçarslan, and M. Celik, “RSigELU: A nonlinear activation function for deep neural networks”, *Expert Systems with Applications*, vol. 174, no. 114805, 2021.
- [28] N. K. Okamura, M. J. Delwiche, and J. F. Thompson, “Raisin grading by machine vision”, *Transactions of the ASAE (USA)*, vol. 2, no. 36, pp. 485-492, 1993.
- [29] S. Abuzir, and Y. Abuzir, "Data Mining For CO2 Emissions Prediction In Italy." *Mühendislik Bilimleri ve Araştırmaları Dergisi* , vol. 3, no. 1, pp. 59-68, 2020.
- [30] F. Tarakci, and I. A. Ozkan, "Comparison of classification performance of kNN and WKNN algorithms." *Selcuk University Journal of Engineering Sciences*, vol. 20, no. 2, pp. 32-37, 2021.
- [31] M. Koklu, R. Kursun, Y. S. Taspınar, and I. Cinar, “Classification of Date Fruits into Genetic Varieties Using Image Analysis”, *Mathematical Problems in Engineering*, vol. 2021, pp. 1-13, 2021.
- [32] M. Koklu, S. Sarigil, and O. Ozbek, “The use of machine learning methods in classification of pumpkin seeds (*Cucurbita pepo* L.)”, *Genetic Resources and Crop Evolution*, vol. 68, no.7, pp. 2713-2726, 2021.