



# Çiğ Sütün Yağlılık Oranının Yakın Kızılötesi Spektroskopi Sinyalleri ile Tespit Edilmesi

## Determining the Fatty Rate of Raw Milk with Near Infrared Spectroscopy Signals

Tuğba Aydemir<sup>1</sup>

<sup>1</sup> Recep Tayyip Erdoğan Üniversitesi, Fizik Bölümü, [tugba\\_aydemir17@erdogan.edu.tr](mailto:tugba_aydemir17@erdogan.edu.tr)

ORCID: <https://orcid.org/0000-0003-3370-6504>

### MAKALE BİLGİLERİ

#### Makale Geçmişi:

Geliş 15 Ocak 2024  
Revizyon 22 Şubat 2024  
Kabul 22 Şubat 2024  
Online 30 Haziran 2024

#### Anahtar Kelimeler:

*Yakın kızıl ötesi spektroskopisi, çiğ süt, yağ oranı, sınıflandırma.*

### ÖZ

Süt endüstrisinde süt kalitesi hakkında bilgi veren temel parametrelerden biri sütün yağ oranıdır. Donanım ve yazılım alanındaki gelişmeler çiğ sütün içeriğinin belirlenmesinde hem pratiklik hem de yüksek doğruluk sağlayabilmektedir. Bu çalışmada geniş bir veri kümesi ile yakın kızıl ötesi spektroskopisi ölçümleri kullanılarak çiğ sütün yağlılık oranı az, orta veya çok yağlı olarak sınıflandırılmıştır. Çalışmada çiğ süt için 960 nanometre ile 1690 nm dalga boyu aralığında 2.86 nm çözünürlükle toplam 256 adet yakın kızıl ötesi spektral ölçümler kaydedilmiştir. Bu spektral ölçümlerin her biri öznitelik olarak dikkate alınmış ve Ki-Kare ve Komşu Bileşen Analizi yaklaşımları ile en etkin öznitelik kombinasyonu aranmıştır. En etkin öznitelik kombinasyonu  $k$ -en Yakın Komşuluk, Bayes, Karar Ağacı ve Destek Vektör Makineleri için en yüksek sınıflandırma doğruluğu elde edilecek şekilde belirlenmiştir. Önerilen yöntemde  $k$ -en Yakın Komşuluk sınıflandırıcısı ve Komşu Bileşen Analizi ile belirlenmiş sadece 46 spektral ölçüm değeri kullanılarak %82.32 oranında sınıflandırma doğruluğu elde edilebileceği gösterilmiştir. Sütün içeriği ürünün kalitesinin belirlenmesinin yanı sıra, ürünün fiyatının belirlenmesinde, beslenen ve süt veren hayvanların sağlığının izlenebilmesi hakkında da bilgiler verir. Bu çalışmada önerilen yöntemin pratik, yüksek doğruluklu ve üretilen sütün düzenli analizini yapabilmeye imkan tanınması ile sürünün sağlığını ve beslenme şeklinin uygunluğunu sürekli takip edilebilmesini olanaklı hale getireceği düşünülmektedir.

### ARTICLE INFO

#### Article history:

Received 15 January 2024  
Received in revised form 22 February 2024  
Accepted 22 February 2024  
Available online 30 June 2024

#### Keywords:

*Near infrared spectroscopy, raw milk, fat content, classification.*

### ABSTRACT

One of the basic parameters that gives information about milk quality in the dairy industry is the fat ratio. Developments in hardware and software can provide both practicality and accuracy in determining the content of raw milk. In this study, the fat content of raw milk was classified as low, medium or high fat using a large data set and near infrared spectroscopy measurements. In this study, a total of 256 near-infrared spectral measurements were recorded for raw milk in the wavelength range of 960 nanometers to 1690 nm with a resolution of 2.86 nm. Each of these spectral measurements was considered as a feature and the most effective feature combination was searched using Chi-Square and Neighborhood Component Analysis approaches. The most effective feature combination was determined to achieve the highest classification accuracy for  $k$ -Nearest Neighbour, Bayes, Decision Tree and Support Vector Machines. In the proposed method, it has been shown that 82.32% classification accuracy can be achieved by using only 46 spectral measurement values determined by  $k$ -Nearest Neighborhood classifier and Neighborhood Component Analysis. In addition to determining the quality of the product, the content of milk also provides information about determining the price of the product and the health of the dairy animals. It is thought that the proposed method in this study will be practical, highly accurate and allow regular analysis of the milk produced, making it possible to constantly monitor the health of the herd and the suitability of the nutrition style.

Doi: 10.24012/dumf.1420331

\* Sorumlu Yazar

## Giriş

Donanımsal ve yazılımsal gelişmelerle birlikte gıdaların üretim, dağıtım ve tüketim aşamalarında pratik, güvenilir, doğruluğu yüksek ve çevre dostu teknikler giderek artış göstermektedir. Bu teknikler hızlı sonuç vermeleri, kimyasal işlem gerektirmemeleri ve bir uzmana ihtiyaç duyulmaması özellikleri ile de tercih edilmektedir [1]. Önerilen teknikler içerisinde yakın kızılötesi spektroskopisi (YKS) esaslı yaklaşımlar bahsedilen bu özelliklerin çoğunu sağlaması bakımından gıdalardaki çeşitli nicel içerik analizlerinde yaygın olarak kullanılmaktadır. YYS 750-2000 nm dalga boyu aralığındaki elektromanyetik radyasyonun absorpsiyonu temeline dayanır.

YKS su içeriğinin tespit edilmesi [2], üzüm içeriğinin ve kalitesinin belirlenmesi [3], ekmek yapımında kullanılan unun kalitesinin ölçülmesi [4] ve elmaların sululuk oranını ve çözünür katı madde içeriğinin tespit edilmesi [5] gibi geniş yelpazede gıdaların analizlerinde başarılı sonuçlar verdiği literatürde raporlanmıştır. Bir başka uygulama alanı da çiğ sütün kalitesinin, yağ, protein, laktoz ve diğer bileşen oranlarının belirlenmesine yöneliktir. Sütün içeriği ürünün fiyatının ve kalitesinin belirlenmesinin yanı sıra, sürünün sağlığının belirlenmesi hakkında da bilgiler verir. İnek sağlığını izlemenin en iyi yolu, üretilen sütün düzenli analizini yapmak ve yağ ve protein arasındaki oranı değerlendirmektir [6].

Literatürde farklı dalga boyları ile süt içeriğinin belirlenmesine yönelik raporlanan başarılı çalışmalar vardır. Süt içeriğinin belirlenmesine yönelik bir çalışmada Kawamura vd. homojenize edilmemiş sütün üç ana süt bileşenini (yağ, protein ve laktoz), somatik hücre sayısını (SCC) ve süt üre nitrojenini tahmin etmek için kalibrasyon modelleri geliştirmişlerdir [7]. Çalışmalarında 4 farklı inekten 600-1050 nm aralığında (1 nm çözünürlükle) toplam 455 örnek kaydetmişlerdir. Geliştirdikleri modelde yağ oranının testi için sadece 151 örneği kullanmışlardır. Performans ölçütü için determinasyon katsayısını kullanmışlardır ve 0.95 oranında bir başarı elde etmişlerdir. Benzer bir başka çalışmada ise Kawasaki vd. toplam 216 örneğin 72 tanesini geliştirdikleri yöntemi test etmek için kullanmışlardır ve 0.95 determinasyon katsayısı başarıları hesaplamışlardır [8]. İneklerin yemlenme farklılıklarını sütlerden alınan YYS kayıtları ile tespit edilmesine yönelik bir teknik Valenti vd. tarafından önerilmiştir [9]. Chen vd. ise Çin'de bulunan süpermarketlerden 6 farklı markadaki sütlerin makine öğrenmesi teknikleri ile sınıflandırılmasını %100 doğrulukla gerçekleştirmişlerdir. Bunun için her bir markadan 24'er olmak üzere toplam 144 örnek kullanmışlardır. Bunların 48 tanesi test işlemi için kullanılmıştır. Bir başka çalışmada ise Çelik sütün 7 özelliğini (pH, tat, koku, sıcaklık, yağlılık, renk ve bulanıklık) kullanarak sütün kalitesini yüksek, orta ve düşük olarak %99.9 sınıflandırma doğruluğu ile sınıflandırmıştır [10]. Bu çalışmalar dışında süt içeriğini YYS esaslı yaklaşımlar ile inceleyen başka çalışmalar da vardır [11-15].

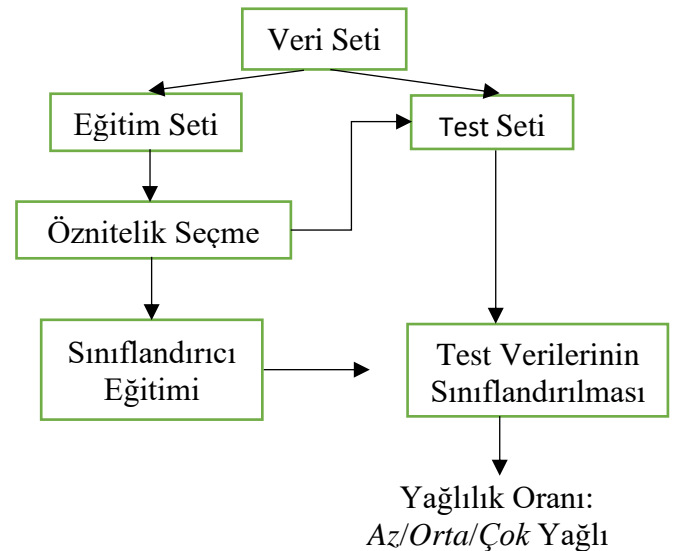
Literatürde çiğ sütün YYS sinyalleri ile farklı açılardan görece başarılı bir şekilde analiz edildiği ancak sütlerin yağ oranının az yağlı, orta yağlı ve çok yağlı olarak

sınıflandırılmamıştır. Bununla birlikte regresyon analizi ile yağlılık oranı tespit edilmeye çalışılmış ancak kararlı bir sonuca varabilmek için yeterli büyüklükte veri kümesi dikkate alınmamıştır. Bu çalışmada 2023 yılında araştırmacıların kullanımına sunulmuş ve üzerinde henüz bir bilimsel araştırma raporlanmamış 1224 örneklik çiğ süt veri kümesi ile sütlerin az yağlı, orta yağlı ve çok yağlı olarak sınıflandırılması gerçekleştirilmiştir. Ayrıca literatürde yapılan çalışmalarda ölçüm yapılan tüm dalga boyları analizde kullanılırken önerilen yöntemde yağlılık oranını tespit etmede anlamlı bilgi içeren dalga boylarının tespit edilmesi de gerçekleştirilmiştir. 4 farklı açıdan yakın kızıl ötesi (YKÖ) spektroskopisi değerleri öznitelik olarak seçilmiş ve en yüksek sınıflandırma doğruluğu %82.32 olarak Komşu Bileşen Analizi (KBA, *İngilizcesi*: Neighborhood Component Analysis) öznitelik seçme ve *k*-en yakın komşuluk (*k*-EYK, *İngilizcesi*: *k*-Nearest Neighborhood) sınıflandırıcısı ile elde edilmiştir.

Makalenin devamında öncelikle bu çalışmada kullanılan veri kümesi tanıtılmıştır. Daha sonra Özniteliklerin kullanımı ve seçimi ile sınıflandırma yöntemleri kısaca tanıtılmıştır. Çalışmanın son bölümlerinde elde edilen sonuçlar tablo ve şekillerde birlikte verilirken en son kısımda varılan bulgular, öneriler ve tartışma kısmı sunulmuştur.

## Yöntem

Bu bölümde Şekil 1'de akış diyagramı verilen yöntemin alt bileşenleri olan veri kümesi tanıtımı ile sınıflandırma için kullanılan öznitelikler ve öznitelik seçme yaklaşımları ve sınıflandırma algoritmaları anlatılmıştır.



Şekil 1. Akış şeması.

## Veri Kümesi Tanıtımı

Bu çalışmada kullanılan veri kümesi, Belçika'nın Antwerp eyaletinin deneysel süt çiftliğinde 2017 yılında sekiz haftalık bir süre boyunca çiftlikte toplanan 1224 ayrı çiğ süt örneğinin 960 nanometre (nm) ila 1690 nm dalga boyu aralığında 2.86 nm çözünürlükle YKÖ spektral ölçümlerini içermektedir [16]. Bu spektral ölçümlerde çiğ sütün yağ, protein, laktoz,

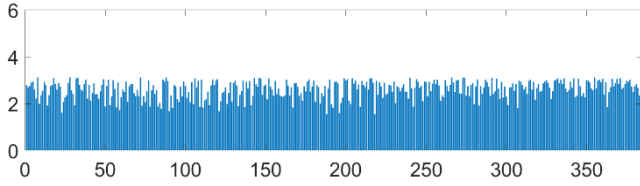
üre ve somatik hücre sayımı değerleri dikkate alınmış olsa da bu çalışmada sadece sütün yağ oranları kullanılmıştır. Bu ölçümler, 256 piksel soğutmalı InGaAs diyet dizisi YKÖ spektrometre cihazı (1.7-256 Düzlem Izgara Spektrometresi, Carl Zeiss, Jena, Almanya) kullanılarak ISO 9622 ve ISO 13366-2:2006 standartlarında kaydedilmiştir. Veri kümesi araştırmacıların genel kullanımına 18 Ağustos 2023 tarihinde sunulmuştur.

1224 çiğ süt örneği Tablo 1’de gösterilen şartlara göre az yağlı (*Sınıf 1*), orta yağlı (*Sınıf 2*) ve çok yağlı (*Sınıf 3*) olarak 3 sınıfa ayrılmıştır. Bu çalışmadaki amaç YKÖ spektral ölçümleri kullanılarak çiğ sütün hangi kategoride olduğunu tespit edebilmektir. Tablo 1’de verilen kategori şartlarına göre *Sınıf 1*’den 386, *Sınıf 2*’den 467 ve *Sınıf 3*’ten 371 örnek elde edilmiştir. Bu tabloda  $M$  tüm örnekler için (1224 çiğ süt örneği) ortalama yağ oranını ve  $S$  bunların standart sapmasını ifade etmektedir ve sırasıyla Eşitlik 1 ve Eşitlik 2’de nasıl hesaplandıkları verilmiştir. Bu eşitliklerde  $Y$  yağlılık oranını ve  $L$  toplam çiğ süt örneklem sayısını göstermektedir.

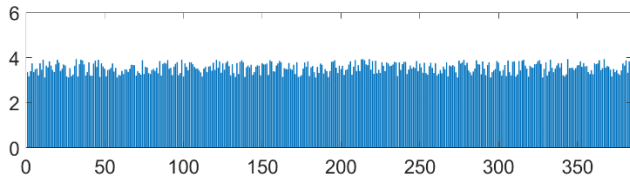
$$M = \frac{1}{L} \sum_{i=1}^L Y_i \quad (1)$$

$$S = \sqrt{\frac{\sum_{i=1}^L (Y_i - M)^2}{L - 1}} \quad (2)$$

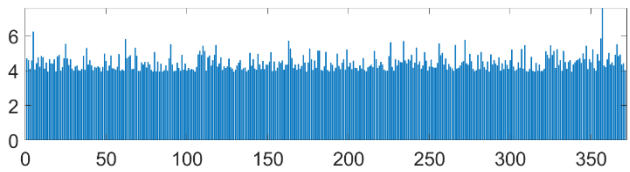
Bu örneklerin rastgele seçilen yarısı önerilen modelin eğitimi için kullanılırken diğer yarısı modelin test edilmesi için kullanılmıştır. Örneklerin yağ oranlarının dağılımı Şekil 2’de ve YKÖ spektral ölçüm değerleri Şekil 3’te gösterilmektedir.



(a)

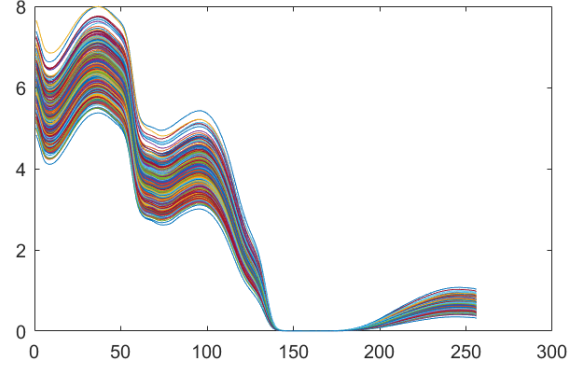


(b)

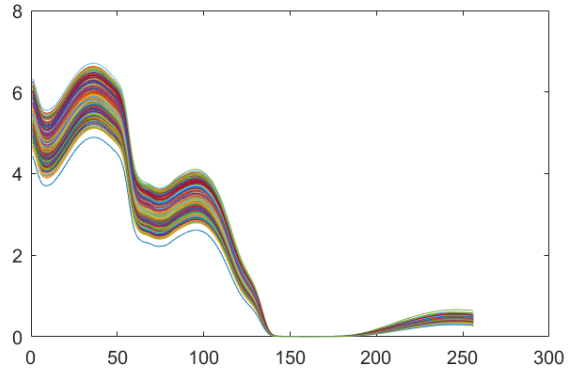


(c)

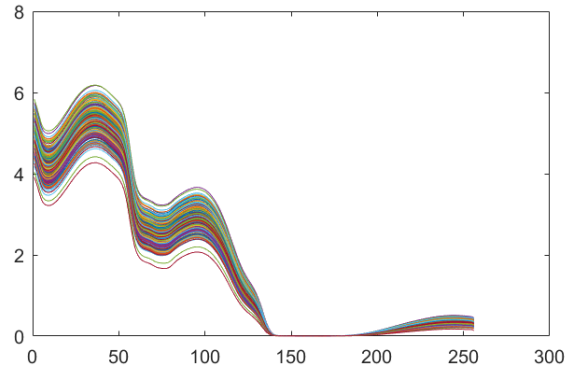
Şekil 2. Örneklerin yağ oranı dağılımları, (a) Yağ oranı 3.12’nin altında olan *Sınıf 1* örnekleri, (b) Yağ oranı 3.12 ile 3.93 arasında olan *Sınıf 2* örnekleri, (c) Yağ oranları 3.93’ün üstünde olan *Sınıf 3* örnekleri.



(a)



(b)



(c)

Şekil 3. YKÖ spektral ölçüm değerleri, (a) *Sınıf 1*’e ait spektral ölçüm değerleri, (b) *Sınıf 2*’ye ait spektral ölçüm değerleri, (c) *Sınıf 3*’e ait spektral ölçüm değerleri.

Tablo 1. Veri kümesinin kategorizasyonu.

Yağlılık Kategorisi	Kategori Şartı	Örnek Sayısı
Az Yağlı	$Sınıf 1 < \left(M - \frac{S}{2}\right) = 3.12$	386
Orta Yağlı	$M - \frac{S}{2} < Sınıf 2 < \left(M + \frac{S}{2}\right) = 3.93$	467
Çok Yağlı	$Sınıf 3 > M + \frac{S}{2}$	371

## Öznitelik Belirleme ve Öznitelik Seçme

Bu çalışmada ölçülen spektral değerler direkt olarak öznitelik değerleri olarak kullanılmıştır. Bu değerlerin dağılımı Şekil 3'te gösterilmiştir. Bu çalışmanın özgün yanlarından biri de olan sınıflandırma doğruluğunu arttıran spektral değerler dört farklı öznitelik seçme yaklaşımları ile değerlendirilmiştir [17]. Bu yaklaşımlar şunlardır: 1) En düşük spektral değerlerden başlayarak ardışıl eklemeye dayalı öznitelik seçme: Bu yaklaşımda öncelikle 960 nm'ye dayalı spektral ölçme değeri öznitelik olarak kullanılarak sınıflandırma yapılmıştır. Ardından 960 nm ile 962.6 nm spektral ölçme değeri öznitelik olarak kullanılarak sınıflandırma yapılmıştır. Bu işlem her defasında bir sonraki spektral değer eklenmesiyle en son 256 adet spektral değer birlikte kullanımı ile sınıflandırma yapıncaya kadar devam etmiştir. Yapılan bu 256 test içerisinde en yüksek sınıflandırma doğruluğunu veren spektral ölçümler en uygun öznitelikler olarak değerlendirilmiştir. 2) En yüksek spektral değerlerden başlayarak ardışıl eklemeye dayalı öznitelik seçme: Bu yaklaşımda öncelikle 1690 nm'ye dayalı spektral ölçme değeri öznitelik olarak kullanılarak sınıflandırma yapılmıştır. Ardından 1690 nm ile 1687.1 nm spektral ölçme değeri öznitelik olarak kullanılarak sınıflandırma yapılmıştır. Bu işlem her defasında bir sonraki spektral değer eklenmesiyle en son 256 adet spektral değer birlikte kullanımı ile sınıflandırma yapıncaya kadar devam etmiştir. Yapılan bu 256 test içerisinde en yüksek sınıflandırma doğruluğunu veren spektral ölçümler en uygun öznitelikler olarak değerlendirilmiştir. 3) KBA öznitelik seçme: Bu yaklaşım en anlamlı özniteliklerin azalan sırada yer aldığı, mesafeye dayalı bir özellikli ağırlıklandırma yöntemidir [18-20]. Komşuluk bileşeni analizinin temel amacı, en yakın komşu sınıflandırmasını optimize eden öznitelik alt kümesini seçmeye uygun bir ağırlıklandırma vektörü  $w$  bulmaktır. İki özellik ( $m_i$  ve  $m_j$ ) arasındaki ağırlıklı mesafe ( $N_w$ ) matematiksel olarak aşağıdaki şekilde hesaplanır:

$$N_w(m_i, m_j) = \sum_{n=1}^s w_n^2 |m_{i,n} - m_{j,n}| \quad (1)$$

Bu eşitlikte  $w_n$ ;  $n$ . öznitelikle ilişkilendirilmiş ağırlığı ifade etmektedir. Ağırlıklı mesafe  $N_w$  ile  $P_{ij}$  olasılığı arasındaki ilişki matematiksel olarak şu şekilde hesaplanabilir:

$$P_{ij} = \begin{cases} \frac{k(N_w(m_i, m_j))}{\sum_{j=1}^R k(N_w(m_i, m_j))}, & i \neq j \\ 0, & i = j \end{cases} \quad (2)$$

burada  $k(N_w(m_i, m_j))$   $m_i$  ve  $m_j$  arasındaki benzerliği tanımlayan bir çekirdek işlevidir.  $m_i$ 'nin doğru sınıflandırılma olasılığı Eşitlik 3'te verilmiştir.

$$P_i = \begin{cases} \sum_{j=1}^R p_{ij} y_{ij}, & i = j \\ 0, & i \neq j \end{cases} \quad (3)$$

burada  $y_{ij}$ ;  $y_i = y_j$  olduğunda 1 diğer durumlarda sıfırdır. Tahmin hatasını en aza indirmek için komşuluk bileşen analizinin amaç fonksiyonu Eşitlik 4'te verilmiştir.

$$A(w) = \frac{1}{R} \sum_{i=1}^R P_i + \beta \sum_{b=1}^B w_b^2 \quad (4)$$

burada  $w_b$ ,  $w$  vektöründeki  $b$ 'inci elemandır ve  $\beta$ , aşırı uyum sorununun aşılmasını sağlayan düzenleme parametresini belirtir ve bunun optimal değeri çapraz doğrulama aşamasında bulunabilir. 4) Ki-kare (*İngilizcesi*: Chi-Square) öznitelik seçme: Ki-kare, girdi değişkeni (öznitelik) ve çıktı değişkeni (sınıf etiketi) arasındaki bağımsızlık eksikliğini belirleyen, çok yaygın olarak kullanılan bir öznitelik seçme yöntemidir. İki aşamada uygulanır. Öncelikle tüm özniteliklerin ayrıklaştırılması ve bu seviyeye göre sıralanması daha yüksek bir anlamlılık düzeyi ile başlar. İkinci olarak, her bir bitişik aralık çifti için Ki-kare değeri hesaplanır. En düşük Ki-kare değerine sahip iki bitişik aralık birleştirilir. Bu işlem, hiçbir Ki-kare değeri birleştirilemeye kadar kullanılır. Bu yöntemin sonunda öznitelikler önem sırasına göre ağırlıklandırılır, yani sınıflandırıcı eğitim sürecine katkılarına göre sıralanır. Ki-kare değeri Eşitlik 5'te gösterilmiştir.

$$\text{Ki - kare}(t_s, c_i) = \frac{T(XY - WZ)^2}{(X + W)(Y + Z)(X + Z)(Y + W)} \quad (5)$$

burada  $T$  veri kümesindeki toplam örnek sayısıdır,  $X$   $c_i$  sınıfındaki  $t_s$  terimini içeren örnek sayısıdır,  $Y$  diğer sınıflardaki  $t_s$  terimini içermeyen örneklerin sayısıdır,  $W$  örnek sayısıdır  $c_i$  sınıfında  $t_s$  terimini içermeyen örneklerin sayısıdır,  $Z$  ise diğer sınıflarda  $t_s$  terimini içeren örneklerin sayısıdır.

Öznitelikler Eşitlik 6'da gösterildiği gibi standart sapma normalizasyonu uygulanarak ve uygulanmadan test edilmiştir.

$$S_N = \frac{S_i - \bar{S}}{\text{std}(S)} \quad (6)$$

Bu eşitlikte  $S_N$ ; normalizasyon uygulanmış bir örneğe ait spektral ölçümü,  $S_i$ ; Bir örneğe ait  $i$ . spektral ölçüm değerini,  $\bar{S}$ ; bir örneğe ait spektral ölçümlerin ortalama değerini ve  $\text{std}(S)$ ; bir örneğe ait spektral ölçümlerin standart sapma değerini göstermektedir.

## Sınıflandırma ve Performans Ölçme Metrikleri

Bu çalışmada çiğ sütün içerdiği yağ oranını az, orta veya çok yağlı olarak sınıflandırmak için  $k$ -EYK, Bayes, Karar Ağacı (KA, *İngilizcesi*: Decision Tree) ve Destek Vektör Makineleri (DVM, *İngilizcesi*: Support Vector Machines) sınıflandırıcıları sınanmıştır. Yukarıda da bahsedildiği gibi 1224 çiğ süt örneğinden oluşturulan veri kümesinin yarısı sınıflandırıcı eğitimi ve kalan yarısı da sınıflandırıcı testi için kullanılmıştır. Literatürde sıklıkla kullanılan bu sınıflandırıcılar aşağıda önemli hususları vurgulanarak kısaca açıklanmıştır.

Gürültüye dayanıklılığı zayıf olarak bilinen  $k$ -EYK, sınıflandırma ve regresyon problemleri için kullanılan bir makine öğrenimi algoritmasıdır. Temel fikri, bir veri

noktasını çevreleyen  $k$  en yakın komşusunu inceleyerek, bu komşuların etiketlerini kullanarak veri noktasının sınıfını veya değerini tahmin etmektir. Bu çalışmada öznitelik vektörleri arasındaki uzaklık Öklid Uzaklık ölçütü ile hesaplanmış olup, eğitim aşamasında en yüksek sınıflandırma eğitimi sağlayan  $k$  değeri 1 ile 25 arasında aranmıştır.

Genellikle metin sınıflandırma, spam filtreleme ve tıbbi teşhis gibi birçok uygulama alanında başarıyla kullanılan Bayes sınıflandırıcısı, Bayes teoreminden türemiştir ve sınıflandırma sürecinde olasılık hesaplamalarını içerir. Test örneğinin özellikleri kullanılarak, Bayes sınıflandırıcısı her sınıf için olasılıkları hesaplar. Ardından, en yüksek olasılığa sahip sınıf, test örneğine tahmin olarak atanır.

Karar ağaçları yöntemi temelde, veri kümesini bir dizi karar kuralı veya koşul ile bölerek bir ağaç yapısı oluşturur. Bu ağaç yapısı, bir veri noktasının sınıfını veya değerini belirlemek için kullanılır. Bu bölmelerin her bir noktası düğüm olarak adlandırılır ve bir karar kuralını temsil eder. Örneğin, "F özelliği 3'ten büyükse sol alt düğüme git, değilse sağ alt düğüme git" gibi bir karar kriteri olabilir.

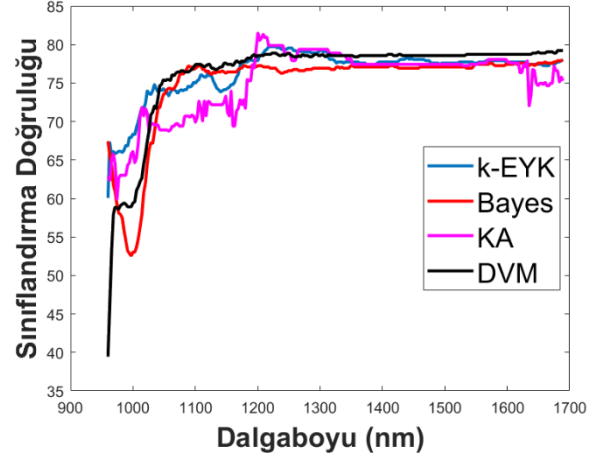
DVM'nin temel amacı, veri noktalarını sınıflar arasındaki bir hiper düzlemle en iyi şekilde ayırmaktır. Bu hiper düzlem, veri noktalarını sınıflarına göre en iyi şekilde bölme kapasitesine sahip olan düzlemdir. Destek vektörler, hiper düzleme en yakın olan veri noktalarıdır. Bu noktalar, sınıflar arasındaki marjini maksimize etmek için önemlidir. DVM'nin iki önemli hiperparametresi  $H$  ve  $\gamma$ 'dır.  $H$ , sınıflandırma hatasını düzenleyen bir parametredir. Daha yüksek  $H$  değerleri, marjinal hatayı minimize etmeye odaklanır, ancak aşırı öğrenme riskini artırabilir.  $\gamma$ , çekirdek fonksiyonunun esnekliğini kontrol eden bir parametredir. Bu çalışmada açıklanan bu dört sınıflandırıcı Matlab ortamında varsayılan parametreleri ile birlikte kullanılmıştır.

Sınıflandırıcı performansını ölçmede sınıflandırma doğruluğu ( $SD$ ) ölçütü kullanılmıştır. Sınıflandırma doğruluğu Eşitlik 7'de gösterildiği gibi doğru sınıflandırılan örnek sayısının ( $DSÖS$ ) toplam örnek sayısına ( $TÖS$ ) bölünüp 100 ile çarpılması ile yüzde cinsinden hesaplanmıştır.

$$SD = \frac{DSÖS}{TÖS} \cdot 100 \quad (7)$$

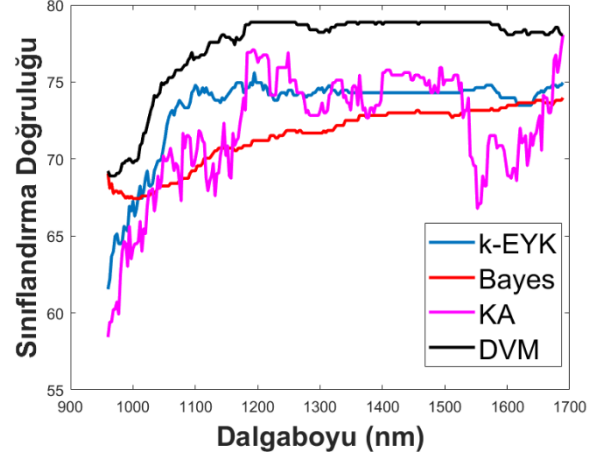
## Sonuçlar

Bu çalışmada çiğ sütün yağlılık oranı az, orta veya çok yağlı olmak üzere YKÖ spektral ölçüm verileri kullanılarak sınıflandırılmıştır. Aşağıda önışlem uygulanmış ve uygulanmamış öznitelik kullanım yaklaşımlarına ait sonuçlar verilmiştir. Şekil 4'te en düşük spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmış öznitelik seçme yaklaşımına ait sınıflandırma sonuçları sunulmuştur. Bu sonuçlara göre  $k$ -EYK, Bayes, KA ve DVM'nin en yüksek  $SD$  sonuçları sırasıyla %79.71 (960-1217.40 nm), %77.91 (960-1680.72 nm), %81.51 (960-1200.24 nm) ve %79.21 (960-1680.72 nm) olarak hesaplanmıştır.



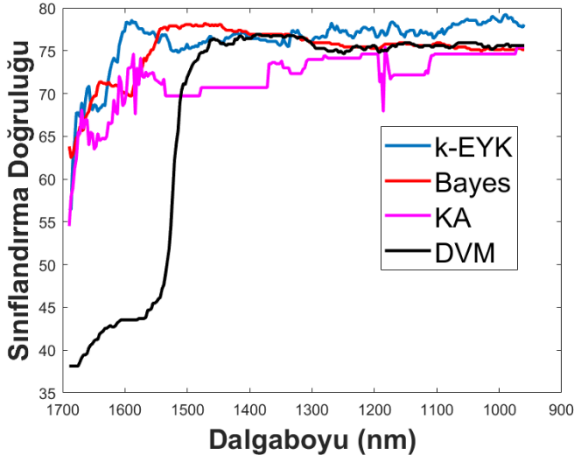
Şekil 4. En düşük spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmış öznitelik seçme yaklaşımına ait sınıflandırma sonuçları.

Şekil 5'te en düşük spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmamış öznitelik seçme yaklaşımına ait sınıflandırma algoritmalarının sonuçları sunulmuştur. Bu sonuçlara göre  $k$ -EYK, Bayes, KA ve DVM'nin en yüksek  $SD$  sonuçları sırasıyla %75.61 (960-1194.52 nm), %73.98 (960-1689.30 nm), %78.07 (960-1689.30 nm) ve %78.89 (960-1188.80 nm) olarak hesaplanmıştır.



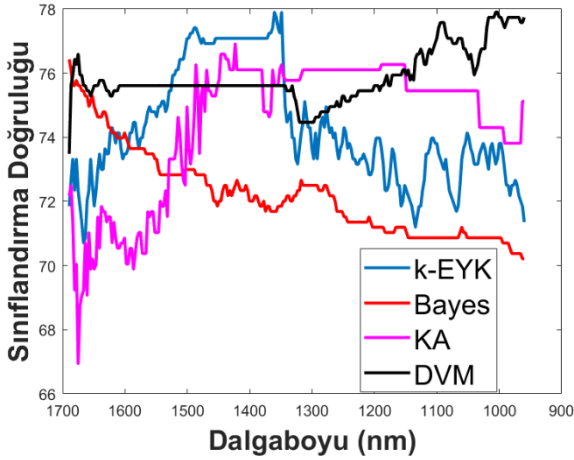
Şekil 5. En düşük spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmamış öznitelik seçme yaklaşımına ait sınıflandırma sonuçları.

Şekil 6'da en yüksek spektral değerlerden başlayarak ardışıl eklemeye dayalı öznitelik seçme yaklaşımına ait sınıflandırma algoritmalarının normalizasyon uygulanmış sonuçları sunulmuştur. Bu sonuçlara göre  $k$ -EYK, Bayes, KA ve DVM'nin en yüksek  $SD$  sonuçları sırasıyla %79.21 (1700-991.46 nm), %78.23 (1700-1471.94 nm), %75.45 (1700-1183.08 nm) ve %76.92 (1700-1389 nm) olarak hesaplanmıştır.



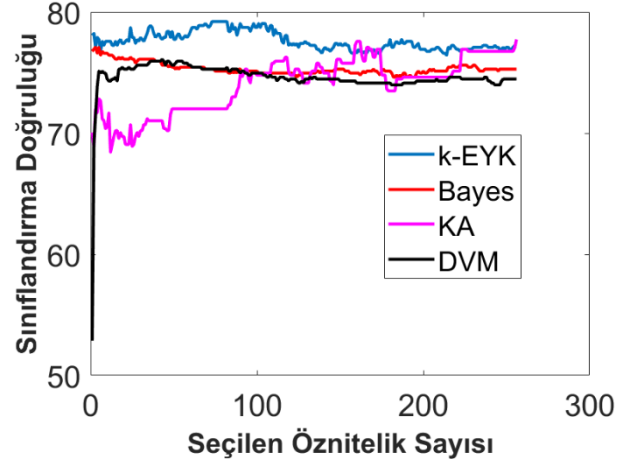
Şekil 6. En yüksek spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmış öznelik seçme yaklaşımına ait sınıflandırma sonuçları.

Şekil 7'de en yüksek spektral değerlerden başlayarak ardışıl eklemeye dayalı öznelik seçme yaklaşımına ait normalizasyon uygulanmamış sınıflandırma algoritmalarının sonuçları sunulmuştur. Bu sonuçlara göre *k*-EYK, Bayes, KA ve DVM'nin en yüksek SD sonuçları sırasıyla %77.91 (1700-1360.40 nm), %76.43 (1700-1689.30 nm), %76.92 (1700-1423.32 nm) ve %77.91 (1700-1005.76 nm) olarak hesaplanmıştır.

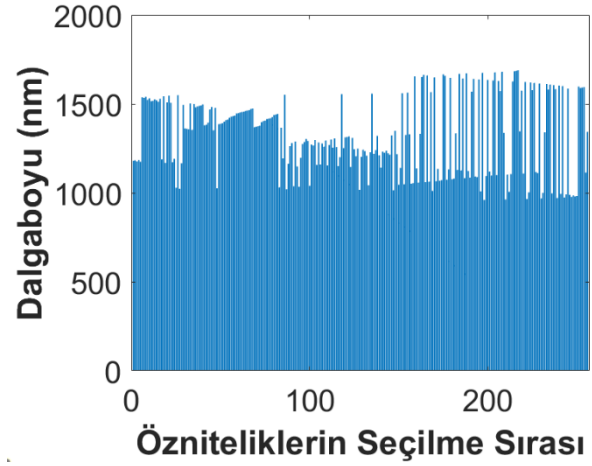


Şekil 7. En yüksek spektral değerlerden başlayarak ardışıl eklemeye dayalı normalizasyon uygulanmamış öznelik seçme yaklaşımına ait sınıflandırma sonuçları.

Şekil 8'de Ki-Kare öznelik ağırlık belirleme yaklaşımına göre elde edilen normalizasyon uygulanmış sonuçlar verilmiştir. Şekil 9'da ise bu sonuçların dikkate alınan özneliklerin sırası bar grafiği ile gösterilmiştir. Bar grafiğinden görüldüğü gibi YKÖ spektral ölçümlerin bu kriter için anlamlı bir ağırlıklandırma dağılımı görülmemektedir. *k*-EYK sınıflandırıcısı ile en yüksek SD değerini ilk 73 en ağırlıklandırılmış spektral ölçüm kullanılarak %79.21 olarak elde edilirken, Bayes ile sadece ilk 3 en ağırlıklandırılmış spektral ölçüm değerleri ile %77.09 SD hesaplanmıştır. KA sınıflandırıcısı ise en yüksek SD değerini tüm spektral ölçümleri kullanarak %77.74 olarak sağlarken, DVM ilk 43 en ağırlıklandırılmış spektral değerle %76.11 SD değeri sağlamıştır.

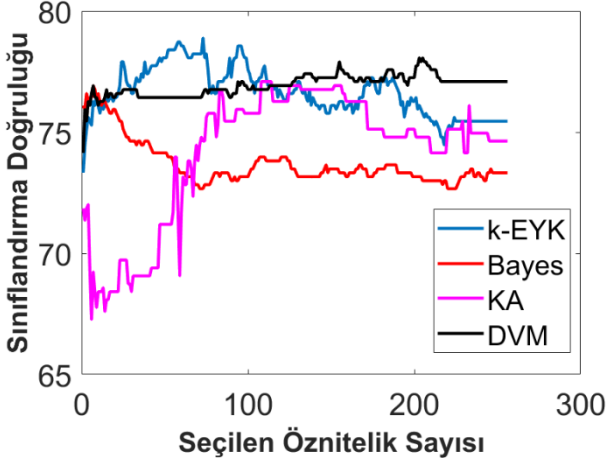


Şekil 8. Ki-Kare öznelik ağırlıklandırma ve normalizasyon uygulanmış sınıflandırma sonuçları.

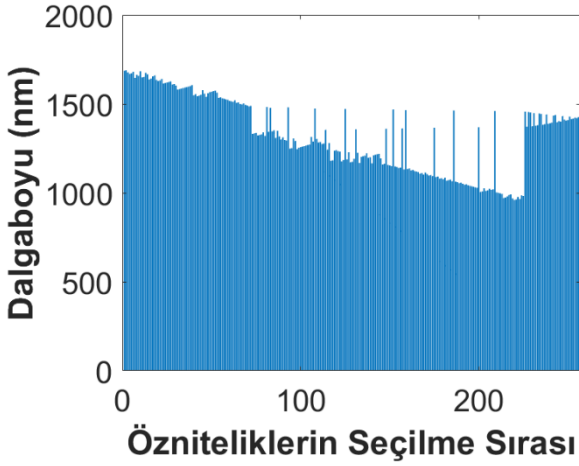


Şekil 9. Normalizasyon uygulanmış Ki-Kare öznelik ağırlıklandırma sonuçları.

Şekil 10'da Ki-Kare öznelik ağırlık belirleme yaklaşımına göre elde edilen normalizasyon uygulanmamış sonuçlar verilmiştir. Şekil 11'de ise bu sonuçların dikkate alınan özneliklerin sırası bar grafiği ile gösterilmiştir. Bar grafiğinden görüldüğü gibi YKÖ spektral ölçümlerin yüksek dalga boyları genelde öznelik olarak daha etkin belirlenmiş ve dalga boyu azaldıkça ağırlıklandırma değeri düşüş göstermiştir. *k*-EYK sınıflandırıcısı ile en yüksek SD değeri ilk 73 en ağırlıklandırılmış spektral ölçüm kullanılarak %78.89 olarak elde edilirken, Bayes ile sadece ilk 7 en ağırlıklandırılmış spektral ölçüm değerleri ile %76.92 SD hesaplanmıştır. KA sınıflandırıcısı ise en yüksek SD değerini ilk 108 en ağırlıklandırılmış spektral ölçümleri kullanarak %77.08 olarak sağlarken, DVM ilk 203 en ağırlıklandırılmış spektral değerle %78.07 SD değeri sağlamıştır.

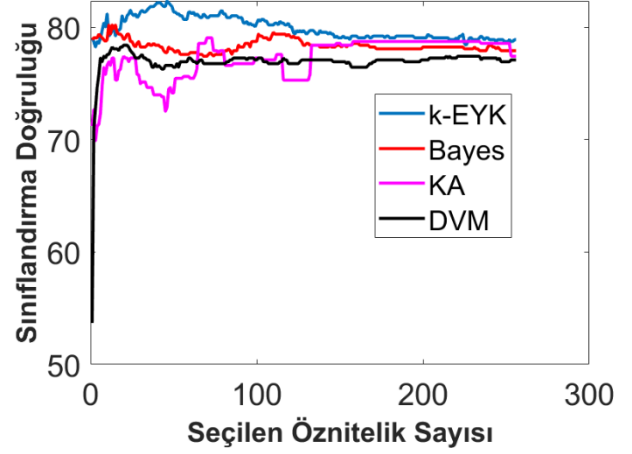


Şekil 10. Ki-Kare öznitelik ağırlıklandırma ve normalizasyon uygulanmamış sınıflandırma sonuçları.

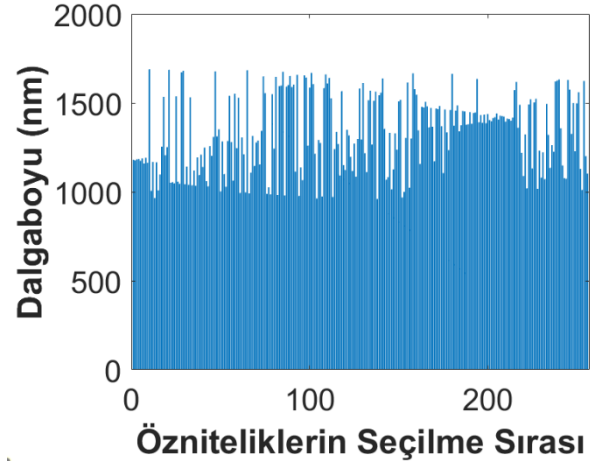


Şekil 11. Normalizasyon uygulanmamış Ki-Kare öznitelik ağırlıklandırma sonuçları.

Şekil 12'de KBA öznitelik ağırlık belirleme yaklaşımına göre elde edilen normalizasyon uygulanmış sonuçlar verilmiştir. Şekil 13'te ise bu sonuçların dikkate alınan özniteliklerin sırası bar grafiği ile gösterilmiştir. *k-EYK* sınıflandırıcısı ile en yüksek SD değerini ilk 46 en ağırlıklandırılmış spektral ölçüm kullanılarak %82.32 olarak elde edilirken, Bayes ile sadece ilk 11 en ağırlıklandırılmış spektral ölçüm değerleri ile %80.20 SD hesaplanmıştır. KA sınıflandırıcısı ise en yüksek SD değerini ilk 70 spektral ölçümleri kullanarak %79.05 olarak sağlarken, DVM ilk 20 en ağırlıklandırılmış spektral değerle %78.40 SD değeri sağlamıştır.

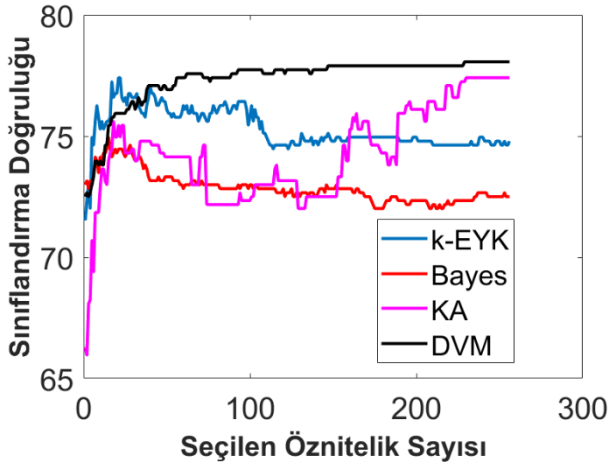


Şekil 12. KBA öznitelik ağırlıklandırma ve normalizasyon uygulanmış sınıflandırma sonuçları.

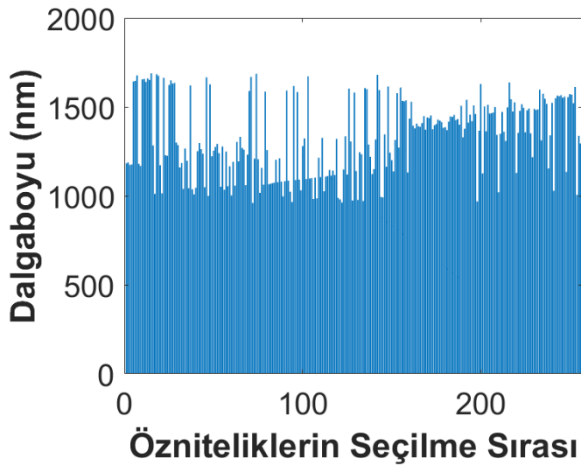


Şekil 13. Normalizasyon uygulanmış KBA öznitelik ağırlıklandırma sonuçları.

Şekil 14'te ise KBA öznitelik ağırlık belirleme yaklaşımına göre elde edilen normalizasyon uygulanmamış sonuçlar verilmiştir. Şekil 15'te ise bu sonuçların dikkate alınan özniteliklerin sırası bar grafiği ile gösterilmiştir. *k-EYK* sınıflandırıcısı ile en yüksek SD değerini ilk 21 en ağırlıklandırılmış spektral ölçüm kullanılarak %77.41 olarak elde edilirken, Bayes ile sadece ilk 15 en ağırlıklandırılmış spektral ölçüm değerleri ile %74.96 SD hesaplanmıştır. KA sınıflandırıcısı ise en yüksek SD değerini ilk 230 spektral ölçümleri kullanarak %77.41 olarak sağlarken, DVM ilk 229 en ağırlıklandırılmış spektral değerle %78.07 SD değeri sağlamıştır.



Şekil 14. KBA öznelik ağırlıklandırma ve normalizasyon uygulanmamış sınıflandırma sonuçları.



Şekil 15. Normalizasyon uygulanmamış KBA öznelik ağırlıklandırma sonuçları.

## Vargılar ve Tartışma

Gıda üretim, dağıtım ve tüketim aşamalarında donanımsal ve yazılımsal gelişmelerle birlikte çeşitli tekniklerin kullanımı artmaktadır. Özellikle çiğ süt içeriğinin belirlenmesi amacıyla YKS kullanımı da önemli bir uygulama alanı bulmaktadır. Literatürde geniş spektrumda üzerinde yapılan görece başarılı çalışmalar olsa da özellikle bu çalışmaların sınırlı ölçekte veri kümeleri kullanmalarından dolayı bunların kararlılığı tartışılmaktadır. Bu çalışmada, 2023 yılında araştırmacıların kullanımına sunulmuş ve üzerinde raporlanan henüz bir çalışma olmayan 1224 örneklik çiğ süt veri kümesi kullanılmıştır. *k*-EYK sınıflandırıcısı ile KBA öznelik seçme yöntemi ile %82.32 oranında bir sınıflandırma başarımı elde edilmiştir. Bu oran sadece 46 en ağırlıklandırılmış spektral ölçüm kullanılarak elde edilmiştir. Önerilen yöntemde ayrıca standart sapma normalizasyon işleminin sınıflandırma doğruluğunu arttıracığı da ispatlanmıştır. Bu çalışma ile kullanıcıların ve araştırmacıların çok geniş spektrum kullanmalarına gerek bırakmadan daha pratik ve hızlı bir çiğ süt yağlılık oranı

belirleyebilmelerini sağlayabilen bir yöntem önerilmiştir. Bu araştırma literatürde çiğ sütün yağlılık oranını *az/orta/çok* yağlı olarak sınıflandıran ilk çalışma olduğundan sınıflandırma doğruluğu bakımından direkt bir karşılaştırma yapılamamıştır. Ancak, önerilen öznelik seçme yaklaşımı ile bu çalışmada sadece 46 spektral ölçüm kullanılması yeterli olurken [7], [8] ve [9] numaralı çalışmalarda sırasıyla 451, 451 ve 700 adet spektral ölçüm kullanılmıştır. Bu karşılaştırmadan tüm spektral ölçüm bölgesi kullanılması yerine KBA yaklaşımı ile en etkin spektral ölçümleri süt bileşenlerini belirleme için kullanılabileceği gösterilmiştir.

Gelecek çalışmalarda sınıflandırma doğruluğunun arttırabilmesi için ileri ardışıl öznelik seçme yöntemi de test edilecektir. Ayrıca, spektral ölçüm değerlerinden çarpıklık, basıklık, ortalama değer, standart sapma ve türevlerinin varyansı gibi istatistiksel öznelikler hesaplanacak ve bu öznelikler ile sınıflandırmaların yapılması düşünülmektedir.

## Kaynaklar

- [1] H. G. Yakubu, Z. Kovacs, T. Toth, & G. Bazar, "The recent advances of near-infrared spectroscopy in dairy production—A review", *Critical Reviews in Food Science and Nutrition*, 62(3), 810-831, 2022.
- [2] H. Büning-Pfaue, "Analysis of water in food by near infrared spectroscopy", *Food Chemistry*, 82(1), 107-115, 2003.
- [3] D. Cozzolino, R. G. Damberg, L. Janik, W. U. Cynkar, & M. Gishen, "Analysis of grapes and wine by near infrared spectroscopy", *Journal of Near Infrared Spectroscopy*, 14(5), 279-289, 2006.
- [4] M. L. Vigni, C. Durante, G. Foca, A. Marchetti, A. Ulrici, & M. Cocchi, "Near infrared spectroscopy and multivariate analysis methods for monitoring flour performance in an industrial bread-making process", *Analytica chimica acta*, 642(1-2), 69-76, 2009.
- [5] S. Zeng, Z. Zhang, X. Cheng, X. Cai, M. Cao, & W. Guo, "Prediction of soluble solids content using near-infrared spectra and optical properties of intact apple and pulp applying PLSR and CNN", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 304, 123402, 2024.
- [6] A. Gastélum-Barrios, G. M. Soto-Zarazúa, A. Escamilla-García, M. Toledano-Ayala, G. Macías-Bobadilla, & D. Jauregui-Vazquez, "Optical methods based on ultraviolet, visible, and near-infrared spectra to estimate fat and protein in raw milk: A review", *Sensors*, 20(12), 3356, 2020.
- [7] S. Kawamura, M. Kawasaki, H. Nakatsuji, & M. Natsuga, "Near-infrared spectroscopic sensing system for online monitoring of milk quality during milking", *Sensing and Instrumentation for Food Quality and Safety*, 1, 37-43, 2007.
- [8] M. Kawasaki, S. Kawamura, M. Tsukahara, S. Morita, M. Komiya, & M. Natsuga, "Near-infrared spectroscopic sensing system for on-line milk quality assessment in a milking robot", *Computers and electronics in agriculture*, 63(1), 22-27, 2008.



- [9] B. Valenti, B. Martin, D. Andueza, C. Leroux, C. Labonne, F. Lahalle, & A. Ferlay, "Infrared spectroscopic methods for the discrimination of cows' milk according to the feeding system, cow breed and altitude of the dairy farm", *International Dairy Journal*, 32(1), 26-32, 2013.
- [10] A. Çelik, "Using Machine Learning Algorithms to Detect Milk Quality", *Eurasian Journal of Food Science and Technology*, 6(2), 76-87, 2022.
- [11] Y. Y. Pu, C. O'Donnell, J. T. Tobin, & N. O'Shea, "Review of near-infrared spectroscopy as a process analytical technology for real-time product monitoring in dairy processing", *International Dairy Journal*, 103, 104623, 2020.
- [12] A. Khan, M. T. Munir, W. Yu, & B. R. Young, "Near-infrared spectroscopy and data analysis for predicting milk powder quality attributes", *International Journal of Dairy Technology*, 74(1), 235-245, 2021.
- [13] M. Asaduzzaman, M. Kerschbaumer, M. Bodner, N. Haman, & M. Scampicchio, "Short-wave near infrared spectroscopy for the quality control of milk", *Journal of Near Infrared Spectroscopy*, 28(1), 3-9, 2020.
- [14] C. Evangelista, L. Basiricò, & U. Bernabucci, "An overview on the use of near infrared spectroscopy (NIRS) on farms for the management of dairy cows", *Agriculture*, 11(4), 296, 2021.
- [15] H. Chen, C. Tan, Z. Lin, & T. Wu, "Classification of different liquid milk by near-infrared spectroscopy and ensemble modeling", *Spectrochimica Acta Part A: Molecular and Biomolecular Spectroscopy*, 251, 119460, 2021.
- [16] J. A. Diaz-Olivares et al., "Near-infrared spectra dataset of milk composition in transmittance mode," *Data in Brief*, vol. 51, p. 109767, Dec. 2023.
- [17] O. Aydemir, "Prediction of Six Products from the Cucurbitaceae Family Using Visible/Near-Infrared Spectroscopic Data", *Journal of Testing and Evaluation*, 51(2), 979-988, 2023.
- [18] A. K. Agrawal, & G. Chakraborty, "Neighborhood component analysis to leverage the class label information during feature selection to enhance the damage classification performance", *Structures*, 57, 105174, 2023.
- [19] J. Dhar, & N. A. Ayele, "Multi-Tier Ensemble Learning Model with Neighborhood Component Analysis to Predict Health Diseases", *IEEE Access*, 9, 138677-138715, 2021.
- [20] S. Raghu, & N. Sriraam, "Classification of focal and non-focal EEG signals using neighborhood component analysis and machine learning algorithms", *Expert Systems with Applications*, 113, 18-32, 2018.