KSÜ Mühendislik Bilimleri Dergisi, 29(3), 2026
Araştırma Makalesi

KSU J Eng Sci, 29(3), 2026
Research Article

Kahramanmaras Sutcu Imam University
Journal of Engineering Sciences

Geliş Tarihi : 24.03.2026
Kabul Tarihi : 29.06.2026

Received Date : 24.03.2026
Accepted Date : 29.06.2026

İNDİRGENMİŞ ÖZNİTELİKLER VE ÇAPRAZ DOĞRULAMA STRATEJİLERİ İLE ÇOK SINIFLI OBEZİTE TAHMİNİ

MULTI-CLASS OBESITY PREDICTION USING REDUCED FEATURES AND CROSS VALIDATION STRATEGIES

Niyazi KILIÇ (ORCID: 0009-0007-4233-8762)

İstanbul Üniversitesi - Cerrahpaşa, Elektrik Elektronik Mühendisliği Bölümü, İstanbul, Türkiye

*Sorumlu Yazar / Corresponding Author: Niyazi KILIÇ, niyazi.kilic@iuc.edu.tr

ÖZET

Bu çalışmada, obezite seviyelerinin çok sınıflı tahmininde öznitelik seçimi ve çapraz doğrulama parametrelerinin etkileri incelenmiştir. Kaggle üzerinden erişilen ve Türkiye'deki bireylerden çevrim içi anket yoluyla toplanan 1610 örnekten oluşan "Obesity" veri seti kullanılmıştır. Veri seti 14 öznitelik ve dört sınıf (zayıf, normal, fazla kilolu ve obez) içermektedir. Öznitelik seçimi amacıyla Korelasyon Tabanlı Öznitelik Seçimi yöntemi uygulanmış ve hedef değişken ile en yüksek ilişkiye sahip öznitelikler belirlenmiştir. Bu sayede öznitelik sayısı 14'ten önce 9'a, ardından 7'ye düşürülmüştür. Sınıflandırma işlemi Random Forest algoritması kullanılarak gerçekleştirilmiş ve model performansı k=2 ile k=30 arasındaki farklı k-katlı çapraz doğrulama değerleri altında sistematik olarak değerlendirilmiştir. En yüksek doğruluk değeri, 9 öznitelik kullanıldığında %93,41 olarak elde edilmiştir. Sonuçlar, öznitelik sayısının azaltılmasına rağmen sınıflandırma performansının korunduğunu ve bazı durumlarda iyileştirilebildiğini göstermektedir. Ayrıca önerilen yaklaşımın, aynı veri seti üzerinde rapor edilen çalışmalarla karşılaştırıldığında rekabetçi bir performans sergilediği görülmüştür. Bulgular, uygun öznitelik seçimi ve çapraz doğrulama stratejilerinin model karmaşıklığını azaltırken yüksek sınıflandırma başarısının elde edilmesine katkı sağlayabileceğini ortaya koymaktadır.

Anahtar Kelimeler: Obezite, korelasyon tabanlı öznitelik seçimi, rastgele orman algoritması, makine öğrenmesi, çapraz doğrulama

ABSTRACT

This study investigates the effects of feature selection and cross-validation parameters on multi-class obesity level prediction. The "Obesity" dataset from Kaggle, consisting of 1,610 samples collected via an online survey of individuals in Türkiye, was used. The dataset includes 14 features and four classes (underweight, normal weight, overweight, and obese). A Correlation-Based Feature Selection approach was employed to identify the features most strongly associated with the target variable. Using different correlation thresholds, two reduced feature subsets containing 9 and 7 features were generated from the original 14-feature dataset. Classification was performed using the Random Forest algorithm, and model performance was systematically evaluated under k-fold cross-validation settings ranging from k=2 to k=30. The highest classification accuracy of 93.41% was achieved using the nine-feature subset. The results demonstrate that reducing the number of features can preserve and, in some cases, improve classification performance. Furthermore, the proposed approach exhibited competitive performance compared with previously reported studies conducted on the same dataset. The findings indicate that appropriate feature selection and cross-validation strategies can reduce model complexity while maintaining high classification accuracy. These results highlight the importance of feature selection in obesity classification and suggest that simpler models with fewer features can achieve effective and reliable predictive performance.

Keywords: Obesity, correlation-based feature selection, random forest, machine learning, cross validation.

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0).

GİRİŞ

Dünya Sağlık Örgütü'ne göre obezite, sağlığı olumsuz etkileyebilecek aşırı yağ birikimi ile karakterize edilen kronik ve karmaşık bir durum olarak tanımlanır (World Health Organization. Obesity and overweight 2025). Obezite, tip 2 diyabet ve kalp hastalıkları riskini artırır, kemik sağlığını ve üreme fonksiyonlarını etkiler ve bazı kanser türlerinin görülme olasılığını artırır (Yanıkekrem 2017). Ayrıca, uyku ve hareket kabiliyetini engelleyerek yaşam kalitesini olumsuz yönde etkileyebilir. Fazla kilo ve obeziteyi teşhis etmek için, bir kişinin ağırlığı ve boyu ölçülerek vücut kitle indeksi (VKİ) hesaplanır. VKİ, ağırlığın (kg) boyun karesine (m^2) bölünmesiyle bulunur. Yetişkinler için, 25 kg/m^2 veya daha yüksek bir VKİ fazla kilolu olarak, 30 kg/m^2 veya daha yüksek bir VKİ ise obez olarak tanımlanır (World Health Organization. Obesity and overweight 2025).

Obezite, günümüzde dünya çapında hızla yayılan ve ciddi sağlık sorunlarına yol açan önemli bir halk sağlığı sorunu haline gelmiştir. Dünya Sağlık Örgütü'nün (DSÖ) 2022 verilerine göre, dünya genelinde 2,5 milyar yetişkin fazla kilolu kategorisindedir ve bu bireylerin 890 milyondan fazlası obez olarak sınıflandırılmaktadır. Başka bir deyişle, 18 yaş ve üzeri yetişkinlerin %43'ü fazla kilolu, %16'sı ise obeziteyle yaşamlarını sürdürmektedir (World Health Organization. Obesity and overweight 2025). Türkiye İstatistik Kurumu (TÜİK) verilerine göre, boy ve kilo değerleri kullanılarak hesaplanan vücut kitle indeksi incelendiğinde, 15 yaş ve üzeri obez bireylerin oranı 2019 yılında %21,1 iken 2022 yılında %20,2'ye düşmüştür. 2022 yılında cinsiyet farklarına bakıldığında, kadınların %23,6'sının obez, %30,9'unun ise obez öncesi sınıfında yer aldığı, erkeklerin ise %16,8'inin obez, %40,4'ünün obez öncesi sınıfta olduğu görülmüştür (Türkiye İstatistik Kurumu, Türkiye Sağlık Araştırması, 2022). Tip 2 diyabet, kalp-damar hastalıkları ve bazı kanser türleri başta olmak üzere birçok kronik hastalıkla güçlü bir ilişkisi vardır. Bu nedenle obezitenin doğru şekilde tespit edilmesi ve sınıflandırılması, etkili tedavi stratejileri geliştirilmesi açısından kritik öneme sahiptir (Yanıkekrem, 2017).

Obezite tespiti için çeşitli risk faktörleri tanımlamaları yapılmıştır (Telleria-Aramburu & Arroyo-Izaga, 2022). Obezite risk faktörleri, değiştirilebilir ve değiştirilemez olarak iki ana kategoriye ayrılmaktadır. Değiştirilebilir risk kategorisi; yaşam tarzı faktörlerini ve sağlıklı sosyal kuralları içermektedir. Değiştirilemez risk kategorisi ise cinsiyet ve yaş gibi faktörleri kapsamaktadır (Gutiérrez-Gallego ve ark., 2024; Chatterjee ve ark., 2020; Börnhorst ve ark., 2020). Bazı araştırmacılar, obeziteye yol açan 8 ana faktör ve bu ana faktörlere bağlı 30 alt faktör bulunduğunu belirtmişlerdir. Bu ana faktörler; fiziksel aktivite, beslenme alışkanlıkları, cinsiyet, sağlıklı beslenme farkındalığı, genetik ve metabolik hastalıklar, yaş, psikolojik etkiler ve sosyokültürel ve çevresel faktörler olarak sıralanmaktadır (Bolayır ve ark., 2023; Uğurlu ve ark., 2024). Obezite prevalansını etkileyen faktörlerin belirlenmesi ve gerekli önlemlerin alınması, olası sağlık sorunlarıyla mücadele açısından büyük önem taşımaktadır (Canoy & Buchan, 2007). Obezite sınıflandırmasında kullanılan geleneksel yöntemler genellikle çok sayıda değişken içermekte, bu durum model karmaşıklığını artırarak sınıflandırma performansını olumsuz etkileyebilmektedir. Bu nedenle, sınıflandırma doğruluğunu korurken veya artırırken daha az sayıda öznitelik kullanılması araştırmacılar için önemli bir çalışma alanı haline gelmiştir. Öznitelik seçimi, yalnızca hesaplama maliyetlerini azaltmakla kalmayıp aynı zamanda modelin yorumlanabilirliğini ve genelleme yeteneğini de artırabilmektedir (Mert ve ark., 2011; Mert ve ark., 2015). Bu kapsamda Shi ve ark. (2022), topluluk tabanlı öznitelik çıkarma teknikleri kullanarak 119 değişkeni 10 değişkene indirgemiştir. Benzer şekilde Değirmenci (2025), Korelasyon Analizi ve Bilgi Kazancı yöntemlerini kullanarak 14 özniteliği 10 özniteliğe düşürmüş ve obezite tahmin performansını değerlendirmiştir.

Son yıllarda makine öğrenmesi teknikleri sağlık alanında yaygın olarak kullanılmakta ve obezite sınıflandırmasında başarılı sonuçlar elde edilmektedir. Bununla birlikte, mevcut çalışmaların büyük bölümünde öznitelik seçiminin sınıflandırma performansı üzerindeki etkisi sınırlı düzeyde incelenmiş, farklı çapraz doğrulama stratejilerinin model başarısına katkısı ise yeterince değerlendirilmemiştir. Ayrıca model karmaşıklığı ile öznitelik sayısı arasındaki ilişkinin obezite sınıflandırması bağlamında ayrıntılı olarak ele alındığı çalışma sayısı oldukça sınırlıdır. Bu durum, daha az sayıda öznitelik kullanılarak yüksek sınıflandırma performansının elde edilip edilemeyeceği sorusunu açık bırakmaktadır.

Bu çalışmada, Köklü ve Sulak (2024) tarafından oluşturulan ve 1610 örnek ile 14 öznitelikten oluşan obezite veri seti kullanılmıştır. Veri setindeki öznitelikler Korelasyon Tabanlı Öznitelik Seçimi (CFS) korelasyon analizi yardımıyla değerlendirilmiş ve farklı büyüklüklerde indirgenmiş öznitelik kümeleri oluşturulmuştur. Daha sonra Yapay Sinir Ağları (Artificial Neural Network, ANN), Rastgele Orman Algoritması (Random Forest, RF) ve Naive Bayes (NB) sınıflandırıcıları kullanılarak performans analizleri gerçekleştirilmiştir. Ayrıca k-kat çapraz doğrulama

parametresinin sınıflandırma başarısı üzerindeki etkisi sistematik olarak incelenmiştir. Böylece öznitelik sayısındaki azalmanın model performansı ve model karmaşıklığı üzerindeki etkilerinin ortaya konulması amaçlanmıştır.

MALZEME VE YÖNTEMLER

Veri Seti

Bu çalışmada, Kaggle veri deposunda yer alan "Obesity" veri seti kullanılmıştır (Köklü & Sulak, 2024). Veri seti, Türkiye'de yaşayan bireylerden çevrim içi bir anket aracılığıyla toplanmıştır. Bireyler dört farklı sınıfa ayrılmıştır: zayıf, normal, fazla kilolu ve obez. Veri seti toplam 1610 örnek ve 14 öznitelikten oluşmaktadır. Örneklerin 73'ü zayıf, 658'i normal, 592'si fazla kilolu ve 287'si obez olarak etiketlenmiştir.

Bu veri seti; bireylerin demografik özelliklerini, beslenme alışkanlıklarını ve yaşam tarzı değişkenlerini içermesi, çok sınıflı obezite seviyelerini barındırması ve literatürde obezite sınıflandırma çalışmalarında yaygın olarak kullanılması nedeniyle tercih edilmiştir. Ayrıca yeterli örnek sayısına sahip olması, farklı makine öğrenmesi yöntemlerinin performanslarının karşılaştırılmasına olanak sağlamaktadır. Önerilen yaklaşım, benzer yapıda sağlık ve yaşam tarzı verileri içeren farklı obezite veri setlerine de uygulanabilecek niteliktedir. Bununla birlikte, yöntemin farklı popülasyonlardan elde edilen veri setleri üzerinde doğrulanması, genellenebilirliğinin değerlendirilmesi açısından önem taşımaktadır. Veri setindeki özniteliklerin tamamı, ölçüm şekilleri ve bazı temel istatistiksel özellikleri Tablo 1'de sunulmuştur.

Tablo 1. Kullanılan Veri Setindeki Öznitelik, Ölçüm ve Bazı Önemli İstatistiklerin Bilgisi

Öznitelikler	Ölçüm	Minimum	Maksimum	Ortalama	St. Sapma
Cinsiyet	Evet/Hayır	1	2	1,55	0,49
Yaş	Sayı (cm)	18	54	33,11	9,83
Boy	Evet/Hayır	150	193	167,74	7,98
Ailede obez / Aşırı kilolu var mı	Evet/Hayır	1	2	1,83	0,37
Fastfood tüketimi	Kategorik	1	2	1,77	0,44
Sebze tüketim sıklığı	Kategorik	1	3	2,06	0,74
Günlük ana öğün sayısı	Kategorik	1	3	1,87	0,63
Öğünler arası yiyecek tüketimi	Evet/Hayır	1	4	2,39	1,01
Sigara kullanımı	Kategorik	1	2	1,69	0,46
Günlük sıvı tüketimi	Evet/Hayır	1	3	2,10	0,81
Kalori tüketiminin hesaplanması	Kategorik	1	2	1,82	0,38
Fiziksel egzersiz	Kategorik	1	5	3,26	1,34
Teknoloji bağımlılığı	Kategorik	1	3	2,01	0,69
Kullanılan ulaşım türü	Kategorik	1	5	2,66	1,51

Öznitelik Seçme

Öznitelik seçimi, veri setinde yer alan değişkenler arasından hedef değişken açısından en fazla bilgi taşıyan özniteliklerin belirlenmesi ve düşük ayırt ediciliğe sahip değişkenlerin elenmesi sürecidir. Bu yaklaşım sayesinde yüksek boyutlu veri setlerinde hesaplama maliyeti azaltılmakta, aşırı öğrenme riski düşürülmekte ve modelin genelleme performansı artırılmaktadır. Obezite gibi çok faktörlü sağlık problemlerinde öznitelik seçimi, sınıflandırma üzerinde daha etkili olan risk faktörlerinin belirlenmesine olanak sağlamakta ve model performansının iyileştirilmesine katkıda bulunmaktadır. Ayrıca, daha az sayıda öznitelik kullanılması klinik karar destek sistemleri açısından sonuçların yorumlanabilirliğini artırmaktadır.

Korelasyon Tabanlı Öznitelik Seçimi

Öznitelik seçimi için Korelasyon Tabanlı Öznitelik Seçimi (Correlation-Based Feature Selection, CFS) yöntemi kullanılmıştır. Bu yöntem, her bir özniteliğin hedef değişkenle (obezite seviyesi) olan ilişkisini değerlendirerek özniteliklerin önemini ölçer. Hedef değişken nümerik ise genellikle Pearson's korelasyon katsayısı kullanılırken, kategorik bir hedef değişken söz konusu olduğunda Cramér's V veya karşılıklı bilgi (mutual information) gibi ölçütler tercih edilir (Karegowda ve ark., 2010; Wu ve ark., 2013). Bu çalışmada, hedef değişkeni 4 sınıflı kategorik bir değişken olduğundan, kategorik verilere uygun Cramér's V ölçütü kullanılmıştır.

Cramér's V, $\chi 2$ test istatistiğine dayalı bir ölçü olup, iki kategorik değişkenin bağımlılığını [0, 1] aralığında ölçer; 0 tam bağımsızlığı, 1 ise tam bağımlılığı ifade eder. Matematiksel olarak, şu şekilde hesaplanır:

$$V = \sqrt{\frac{x^2}{n.\min(k-1,r-1)}} \quad (1)$$

Burada, χ^2 Ki-kare test istatistiğini, n toplam gözlem sayısını, k sütun sayısını ve r satır sayısını göstermektedir.

Bu çalışmada CFS yönteminin tercih edilmesinin temel nedeni, sınıflandırma modelinden bağımsız bir öznitelik seçimi gerçekleştirmektir. RF eğitim sonrasında öznitelik önem derecelerini hesaplayabilmesine rağmen, bu önem değerleri kullanılan eğitim örneklerine ve model parametrelerine bağlı olarak değişebilmektedir. Buna karşılık, Cramér's V tabanlı korelasyon analizi doğrudan hedef değişken ile öznitelikler arasındaki ilişkinin değerlendirilmesine olanak sağlamaktadır. Böylece sınıflandırma işleminden önce düşük ilişkiye sahip öznitelikler elenerek veri boyutu azaltılmış, model karmaşıklığı düşürülmüş ve daha yorumlanabilir öznitelik kümeleri elde edilmiştir. Ayrıca, filtre tabanlı bir yöntem olan CFS hesaplama maliyeti açısından gömülü yöntemlere göre daha düşük bir yük oluşturmaktadır.

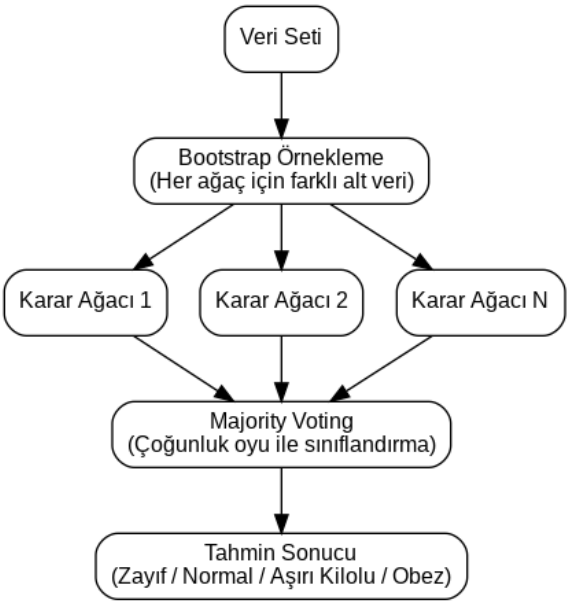
Sınıflandırma Yöntemi: Rastgele Orman Algoritması

Sınıflandırma için RF kullanılmıştır. RF, birden fazla karar ağacını birleştiren bir topluluk öğrenmesi (ensemble learning) yöntemidir. Her bir ağaç, eğitim verisinin rastgele bir alt kümesi (bootstrapping) ve özniteliklerin rastgele bir alt kümesi üzerinde inşa edilir. Sınıflandırma işlemi, tüm ağaçların tahminlerinin birleştirilmesiyle gerçekleştirilir ve nihai tahmin, en çok oy alan sınıf olarak belirlenir (Rodriguez-Galiano ve ark., 2012).

RF, aşırı öğrenmeyi azaltmak ve genelleme yeteneğini artırmak için rastgelelik prensiplerine dayanır. Ayrıca, öznitelik önem derecelerini hesaplayarak hangi özniteliklerin model için daha belirleyici olduğunu ortaya koyar. RF, öznitelik seçimi için Gini Endeksi'ni kullanır; bu endeks, bir özniteliğin sınıflara göre saflık derecesini ölçer (Pal, 2005). Verilen bir eğitim seti T için, rastgele bir durum seçilip bunun bir C_i sınıfına ait olduğu söylendiğinde, Gini Endeksi şu şekilde ifade edilebilir:

$$\sum \sum_{j \neq i} (f(Ci,T)/|T|)(f(Cj,T)/|T|) \quad (2)$$

Burada $(f(Ci,T)/|T|)$ seçilen durumun Ci sınıfına ait olma olasılığıdır. RF obezite sınıflandırması için iş akış diyagramı Şekil 1'de gösterilmiştir.



Şekil 1. RF Obezite Sınıflandırma İş Akış Diyagramı

RF'nin obezite sınıflandırmasındaki üstünlüğünü göstermek amacıyla aynı veri setine ANN (Zou ve ark., 2009) ve NB (Rish. 2001) sınıflandırıcıları da uygulanmış ve sonuçlar karşılaştırılmıştır.

K-Katkı Çapraz Doğrulama

Makine öğreniminde en yaygın kullanılan model değerlendirme tekniklerinden biridir. Bu yöntemde, tüm veri seti rastgele olarak k eşit büyüklükte alt katmana bölünür. Model daha sonra k kez eğitilir: her seferinde k-1 katman eğitim için ve kalan bir katman doğrulama/test için kullanılır. Her k yinelemeden elde edilen performans metrikleri (örneğin doğruluk, F1-skoru) ortalaması alınarak modelin genelleme performansının daha güvenilir bir tahmini elde edilir (Kohavi, 1995). K-katlı çapraz doğrulamanın basit bir eğitim-test bölünmesine göre en büyük avantajı, tüm veri noktalarını hem eğitim hem de doğrulama için kullanması, performans tahmininin varyansını azaltması ve sınırlı veri setlerinden daha iyi yararlanmasıdır (Refaeilzadeh ve ark., 2009).

Bu çalışmada k-kat çapraz doğrulama yöntemi, model performansının belirli bir eğitim-test ayrımına bağımlı kalmadan değerlendirilmesi amacıyla kullanılmıştır. Ayrıca farklı k değerlerinin sınıflandırma performansı üzerindeki etkisini incelemek ve modelin kararlılığını değerlendirmek amacıyla k değeri 2 ile 30 arasında sistematik olarak değiştirilmiştir. Bunun yanında, modelin genelleme yeteneğini artırmak amacıyla öznitelik seçimi uygulanmış ve performans değerlendirmeleri farklı veri bölünmeleri altında gerçekleştirilmiştir. Bu yaklaşım sayesinde yalnızca en yüksek performans değerinin elde edildiği k değeri değil, aynı zamanda farklı k değerleri altında model davranışındaki değişim ve performans kararlılığı da değerlendirilebilmiştir.

Her çapraz doğrulama yinelemesinde eğitim ve test alt kümeleri birbirinden bağımsız tutulmuş ve model performansı yalnızca ilgili yinelemede test amacıyla ayrılan veri alt kümesi üzerinde değerlendirilmiştir. Bu yaklaşım, performans tahminlerinin belirli bir eğitim-test ayrımına bağımlı kalmasını önleyerek daha güvenilir sonuçlar elde edilmesine katkı sağlamaktadır. Bununla birlikte, öznitelik seçimi işleminin çapraz doğrulama uygulamasından önce gerçekleştirilmiş olması nedeniyle, test verilerine ait bilgilerin öznitelik seçim sürecine dolaylı etkisi tamamen dışlanamamaktadır. Bu durumun olası etkileri çalışmanın sınırlılıkları bölümünde ayrıca tartışılmıştır.

Performans Ölçütleri

Performans değerlendirmesinde doğruluk, F1-skoru ve Cohen's Kappa ölçütleri kullanılmıştır. Doğruluk, modelin genel sınıflandırma başarısını göstermektedir. F1-skoru, hassasiyet ve duyarlılık ölçütlerini birlikte değerlendirerek sınıf dağılımlarından kaynaklanan etkileri azaltmaktadır. Cohen's Kappa istatistiği ise rastlantısal doğru sınıflandırmaları dikkate alarak modelin gerçek uyum düzeyini değerlendirmektedir. Bu nedenle söz konusu ölçütlerin birlikte kullanılması, model performansının daha kapsamlı biçimde değerlendirilmesine olanak sağlamaktadır. Bu ölçütler, sınıf dengesizliğinin sıkça görüldüğü obezite veri setlerinde modelin güvenilirliğini objektif olarak yansıtmaktadır.

Doğruluk (Accuracy)

Modelin toplam doğru tahminlerinin tüm örnek sayısına oranıdır. Genel performansı gösterir.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \tag{3}$$

Burada; *TP*: True Positive (Doğru Pozitif), *TN*: True Negative (Doğru Negatif), *FP*: False Positive (Yanlış Pozitif), *FN*: False Negative (Yanlış Negatif).

F1-Skoru (F1-Score)

Precision (Hassasiyet) ve Recall (Duyarlılık) ölçütlerinin harmonik ortalamasıdır. Dengesiz veri setlerinde ve çok sınıflı problemlerde dengeli bir performans göstergesi sağlar.

$$F1 = 2x\frac{Precision\ x\ Recall}{Precision+Recall} \tag{4}$$

$$Precision = \frac{TP}{TP+FP} \tag{5}$$

$$Recall = \frac{TP}{TP+FN} \tag{6}$$

Cohen's Kappa İstatistiği

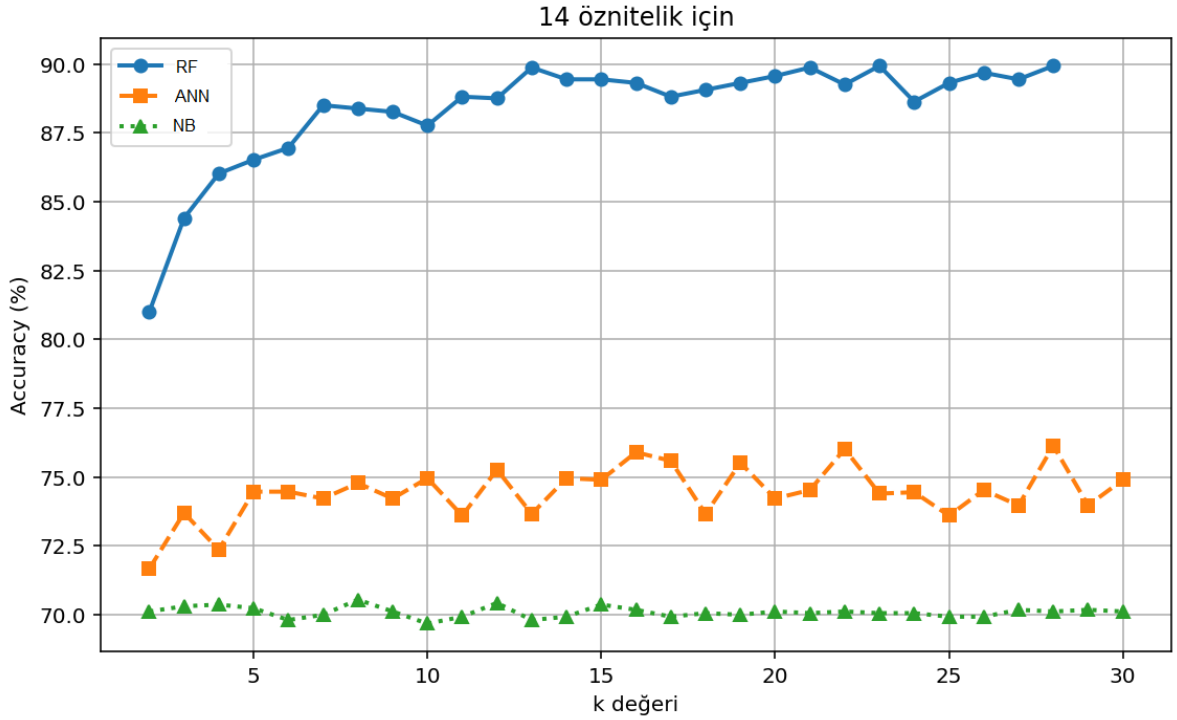
Rastgele tahminden bağımsız olarak gözlenen ve beklenen uyum arasındaki farkı ölçer. 0 ile 1 arasında değer alır; 1 mükemmel uyumu, 0 rastgele tahminleri gösterir.

$$\kappa = \frac{Po-Pe}{1-Pe} \quad (7)$$

Burada; *Po:* Gözlenen uyum oranı (accuracy), *Pe*: Rastgele tahminler yapıldığında beklenen uyum oranıdır (Sınıf olasılıklarının karelerinin toplamı).

BULGULAR VE TARTIŞMA

Bu bölümde, obezite seviyelerinin dört sınıflı (zayıf, normal, fazla kilolu, obez) sınıflandırılması amacıyla gerçekleştirilen deneysel sonuçlar sunulmakta ve yorumlanmaktadır. Çalışmada, Kaggle platformundan elde edilen ve 1610 örnek ile on dört öznitelikten oluşan orijinal veri seti temel alınmıştır. Öncelikle veri setindeki tüm öznitelikler (14 öznitelik) üzerinde üç farklı makine öğrenmesi algoritması uygulanarak temel performans değerleri belirlenmiştir. Model değerlendirmelerinin güvenilirliğini artırmak ve k-katlı çapraz doğrulamanın parametre duyarlılığını incelemek amacıyla, k değeri 2'den 30'a kadar sistematik olarak değiştirilerek her senaryo için doğruluk, F1-skoru ve Cohen's Kappa istatistikleri hesaplanmıştır. Bu yaklaşım, düşük k değerlerinde gözlenen yüksek varyans ile yüksek k değerlerinde elde edilen daha stabil performans tahminleri arasındaki ilişkiyi ortaya koymayı amaçlamaktadır. Şekil 2, on dört öznitelik için k-çapraz doğrulamada k=2 ila 30 aralığında üç farklı sınıflandırıcının doğruluk oranlarını göstermektedir. RF yöntemi tüm k değerlerinde diğer yöntemlere kıyasla belirgin biçimde daha yüksek doğruluk sağlamıştır. ANN orta düzeyde dalgalanma gösterirken, NB yöntemi k değerlerine karşı oldukça kararlı ancak düşük doğrulukta seyretmiştir.



Şekil 2. Bütün (14) Öznitelikler Giriş Alındığında K Değerine Karşılık Gelen Obezite Doğruluk Sonuçları

Tablo 2'de; RF, ANN ve NB sınıflandırıcılarına ait doğruluk, F1-skoru ve Cohen's Kappa değerlerinin minimum, maksimum ve ortalama istatistikleri ile birlikte, bu değerlere karşılık gelen k değerleri de ayrıca belirtilmiştir. Bu özet tablo, farklı k değerlerinin model tahminlerinin güvenilirliği ve kararlılığı üzerindeki etkisini karşılaştırmalı olarak görmeyi kolaylaştırmakta; özellikle RF'nin hemen hemen tüm k aralığında üstün ve nispeten daha az dalgalı performans sergilediği gözlemlenmektedir.

Modelin hesaplama karmaşıklığını azaltmak, yorumlanabilirliğini artırmak ve genelleme yeteneğini iyileştirmek amacıyla CFS yöntemi kullanılarak öznitelik seçimi gerçekleştirilmiştir. Bu kapsamda, hedef sınıfla (obezite seviyesi) en yüksek korelasyon katsayısına sahip öznitelikler belirlenmiştir. Korelasyon analizi sonucunda, hedef değişken ile korelasyon katsayısı 0,15'in üzerinde olan 9 öznitelik belirlenmiştir. Ardından, daha yüksek ayırt edicilik sağlamak amacıyla korelasyon katsayısı 0,20'nin üzerinde olan öznitelikler seçilmiş ve öznitelik sayısı 14'ten 7'ye indirgenmiştir. Bu işlem sonucunda veri boyutu %50 oranında azaltılmıştır. Özniteliklerin obezite seviyesi ile olan korelasyon katsayıları Şekil 3'te gösterilmiştir. Şekil 3 incelendiğinde yaş, sebze tüketim sıklığı ve kullanılan ulaşım

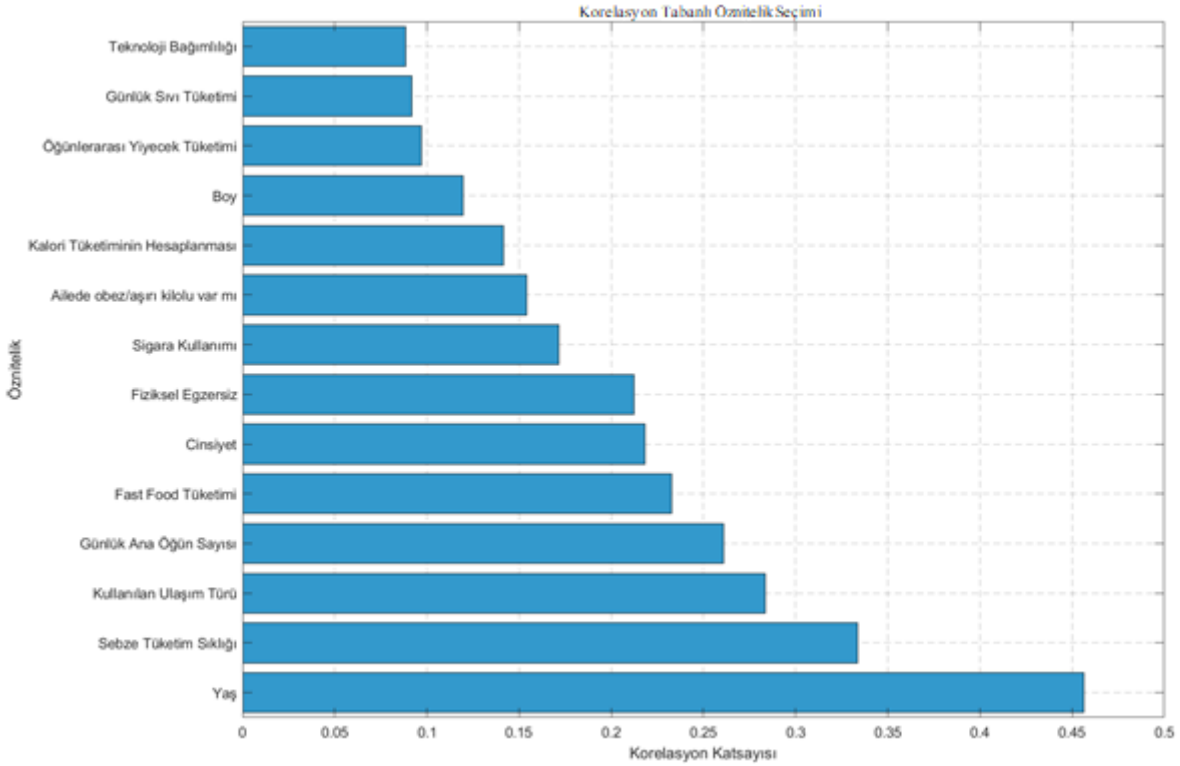
türü değişkenlerinin obezite seviyesi ile görece daha yüksek ilişki gösterdiği, teknoloji bağımlılığı ve günlük sıvı tüketimi gibi değişkenlerin ise daha düşük korelasyon değerlerine sahip olduğu görülmektedir.

Tablo 2. Tüm Öznitelikler (14) Kullanıldığında Elde Edilen Performans Sonuçları

Performans Ölçütü	İstatistik Değerleri	RF	ANN	NB
Doğruluk (%)	Minimum	80,99(k=2)	71,67(k=2)	69,68(k=10)
	Maksimum	89,93(k=23)	74,65(k=28)	70,55(k=8)
	Ortalama	89,45(k=15)	74,43(k=23)	70,10(k=9)
F-Skoru	Minimum	0,808(k=2)	0,713(k=2)	0,693(k=10)
	Maksimum	0,899(k=23)	0,760(k=28)	0,702(k=8)
	Ortalama	0,884(k=15)	0,744(k=23)	0,7(k=9)
Kappa İstatistiği	Minimum	0,708(k=2)	0,563(k=2)	0,537(k=10)
	Maksimum	0,846(k=23)	0,638(k=28)	0,55(k=8)
	Ortalama	0,823(k=15)	0,612(k=23)	0,543(k=9)

Korelasyon eşik değerleri belirlenirken özniteliklerin korelasyon dağılımı incelenmiş ve hem yeterli sayıda özniteliğin korunmasına hem de anlamlı bir boyut indirgeme sağlanmasına imkân veren değerler tercih edilmiştir. Bu amaçla 0,15 ve 0,20 eşik değerleri kullanılarak iki farklı öznitelik kümesi oluşturulmuş ve öznitelik sayısındaki azalmanın sınıflandırma performansı üzerindeki etkileri karşılaştırmalı olarak değerlendirilmiştir.

Seçilen özniteliklerin sınıf ayrıştırıcılığı üzerindeki etkisini değerlendirmek amacıyla aynı üç sınıflandırıcı (ANN, RF ve NB) indirgenmiş veri setleri üzerinde yeniden eğitilmiş ve test edilmiştir. Elde edilen sonuçlar Tablo 3 ve Tablo 4'te sunulmuştur. Ayrıca indirgenmiş öznitelikler kullanılarak gerçekleştirilen obezite sınıflandırmalarına ait doğruluk değerlerinin k parametresine göre değişimi Şekil 4 ve Şekil 5'te gösterilmiştir.



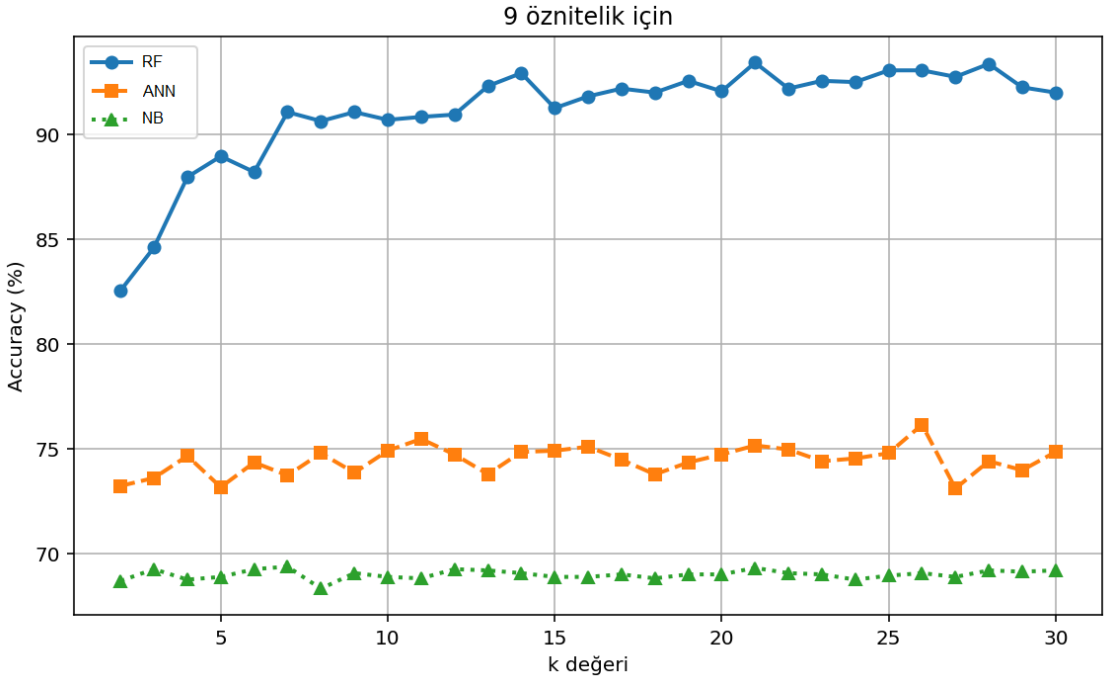
Şekil 3. Kullanılan Özniteliklerin Obezite Seviyesi İle Korelasyon Katsayısı

İndirgenmiş öznitelikler kullanılarak gerçekleştirilen obezite sınıflandırma sonuçları incelendiğinde, RF yönteminin tüm k değerlerinde ANN ve NB yöntemlerine kıyasla daha yüksek performans sergilediği belirlenmiştir. Dokuz öznitelik kullanıldığında en yüksek sınıflandırma doğruluğu k=21 için %93,41 olarak elde edilirken, yedi öznitelik kullanıldığında en yüksek doğruluk değeri k=28 için %91,98 olarak hesaplanmıştır. Bununla birlikte, dikkat çekici bulgulardan biri öznitelik sayısının azaltılmasına rağmen sınıflandırma performansının düşmemesi, aksine artış göstermesidir. Bu durum, veri setinde yer alan tüm özniteliklerin sınıflandırma açısından eşit öneme sahip olmadığını

ve düşük ayırt ediciliğe sahip değişkenlerin model performansını olumsuz etkileyebileceğini göstermektedir. Sonuçlar, uygun öznitelik seçiminin yalnızca veri boyutunu azaltan bir ön işlem adımı olmadığını, aynı zamanda sınıflandırma başarısını doğrudan etkileyen önemli bir faktör olduğunu ortaya koymaktadır.

Tablo 3. Seçilmiş Dokuz (9) Öznitelik Kullanıldığında Elde Edilen Performans Sonuçları

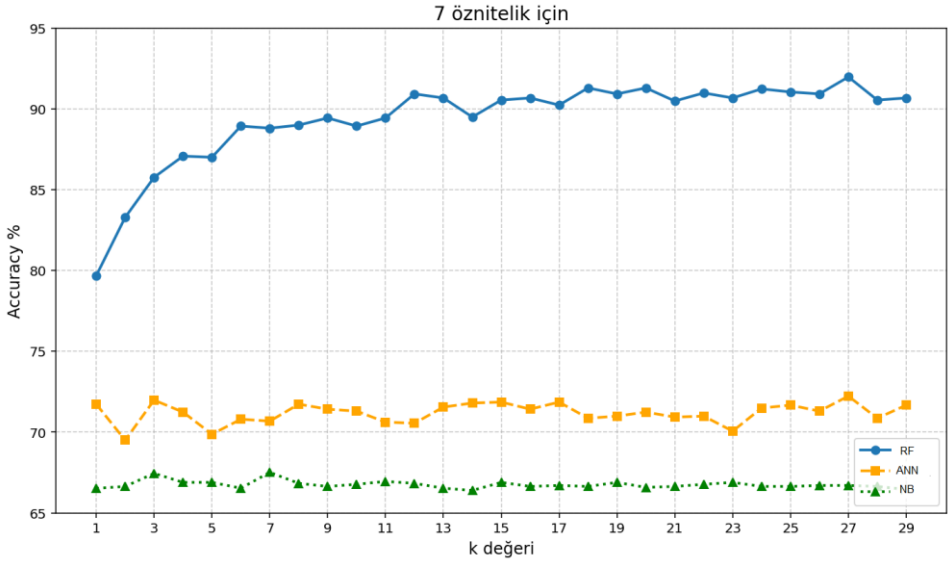
Performans Ölçütü	İstatistik Değerleri	RF	ANN	NB
Doğruluk (%)	Minimum	82,54(k=2)	73,10(k=27)	68,32(k=8)
	Maksimum	93,41(k=21)	76,15(k=26)	69,37(k=7)
	Ortalama	91,06(k=9)	74,45(k=17)	69(k=20)
F-Skoru	Minimum	0,826(k=2)	0,728(k=27)	0,677(k=8)
	Maksimum	0,934(k=21)	0,759(k=26)	0,688(k=7)
	Ortalama	0,91(k=9)	0,744(k=17)	0,682(k=20)
Kappa İstatistiği	Minimum	0,735(k=2)	0,59(k=27)	0,512(k=8)
	Maksimum	0,9(k=21)	0,636(k=26)	0,53(k=7)
	Ortalama	0,864(k=9)	0,61(k=17)	0,522(k=20)



Şekil 4. Seçilmiş 9 (Dokuz) Öznitelik Giriş Alındığında K Değerine Karşılık Gelen Obezite Doğruluk Sonuçları

Tablo 4. Seçilmiş Yedi (7) Öznitelik Kullanıldığında Elde Edilen Performans Sonuçları

Performans Ölçütü	İstatistik Değerleri	RF	ANN	NB
Doğruluk (%)	Minimum	79,68(k=2)	69,50(k=3)	66,39(k=15)
	Maksimum	91,98(k=28)	72,23(k=28)	67,51(k=8)
	Ortalama	89,41(k=10)	71,2(k=5)	66,76(k=11)
F-Skoru	Minimum	0,796(k=2)	0,682(k=3)	0,645(k=15)
	Maksimum	0,878(k=28)	0,713(k=28)	0,66(k=8)
	Ortalama	0,894(k=10)	0,71(k=5)	0,65(k=5)
Kappa İstatistiği	Minimum	0,69(k=2)	0,523(k=3)	0,48(k=15)
	Maksimum	0,92(k=28)	0,568(k=28)	0,50(k=8)
	Ortalama	0,84(k=10)	0,558(k=5)	0,485(k=5)

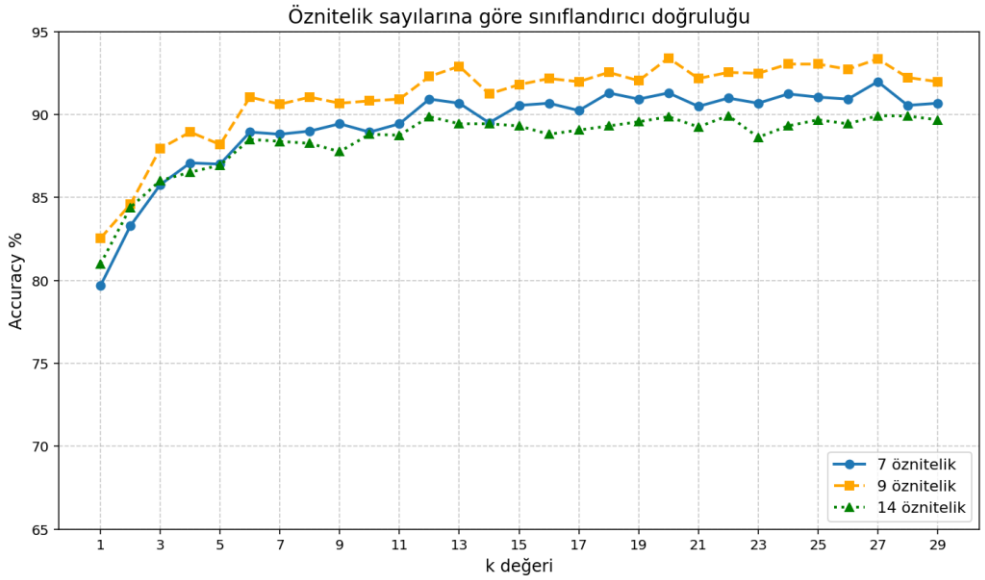


Şekil 5. Seçilmiş 7 (Yedi) Öznitelik Giriş Alındığında K Değerine Karşılık Gelen Doğruluk Sonuçları

Elde edilen performans sonuçları, çok sınıflı obezite sınıflandırmasında RF yönteminin diğer sınıflandırıcılara kıyasla daha yüksek başarı sağladığını göstermektedir. RF sınıflandırıcısına ait performans ölçütleri, farklı öznitelik sayıları dikkate alınarak Tablo 5'te sunulmuştur. Ayrıca, RF yönteminin sınıflandırma doğruluğunun üç farklı öznitelik kümesine göre değişimi Şekil 6'da gösterilmiştir.

Tablo 5. RF Sınıflandırıcısının Seçilmiş Özniteliklere Göre Performans Sonuçları

Performans Ölçütü	İstatistik Değerleri	RF		
		7 Öznitelik	9 Öznitelik	14 öznitelik
Doğruluk (%)	Minimum	79,68(k=2)	82,54(k=2)	80,99(k=2)
	Maksimum	91,98(k=28)	93,41(k=21)	89,93(k=23)
	Ortalama	89,41(k=10)	91,06(k=9)	89,45(k=15)
F-Skoru	Minimum	0,796(k=2)	0,826(k=2)	0,808(k=2)
	Maksimum	0,878(k=28)	0,934(k=21)	0,899(k=23)
	Ortalama	0,894(k=10)	0,91(k=9)	0,884(k=15)
Kappa İstatistiği	Minimum	0,69(k=2)	0,735(k=2)	0,708(k=2)
	Maksimum	0,92(k=28)	0,9(k=21)	0,846(k=23)
	Ortalama	0,84(k=10)	0,864(k=9)	0,823(k=15)



Şekil 6. RF Yönteminin Farklı Öznitelik Sayıları Ve K Değerine Göre Obezite Doğruluklarının Değişimi

Literatürde, aynı veri seti üzerinde yapılan çalışmalarda en yüksek doğruluk değerleri genellikle topluluk (ensemble) yöntemlerle rapor edilmiştir. Örneğin, (Arslan & Yapici, 2025), Ekstra Ağaçlar Sınıflandırması (Extra Trees Classifier) ile SMOTE dengesizlik giderme tekniğini birleştirerek %91,93 doğruluk elde etmişlerdir; ancak bu sonuç tam 14 öznitelik kullanılarak hesaplanmıştır. Benzer şekilde, Değirmenci (2025), Karşılıklı Bilgi (Mutual Information) tabanlı özellik seçimiyle RF modelini optimize ederek 10 öznitelik üzerinde %92,36 doğruluk oranı elde etmişlerdir. Köklü ve Sulak (2024) ise özellik seçimi veya dengesizlik giderme uygulamadan RF yöntemini kullanarak 14 öznitelik ile yalnızca %87,82 doğruluk düzeyine ulaşmışlardır. Buna karşılık, bu çalışmada performans artışı, veri dengeleme yöntemi kullanılmadan CFS ve çapraz doğrulama stratejilerinin birlikte değerlendirilmesiyle elde edilmiştir. Bu durum, veri setine uygun özniteliklerin belirlenmesinin sınıflandırma performansı üzerinde önemli bir etkiye sahip olabileceğini göstermektedir. Literatürde aynı veri seti ile yapılmış çalışmaların karşılaştırılması Tablo 6'da verilmiştir.

Tablo 6. Aynı Veri Seti Üzerinde Gerçekleştirilen Çalışmaların Yöntem Ve Performans Açısından Karşılaştırılması

Çalışma / Yazar(lar)	Yöntem / Model	Öznitelik Sayısı	En Yüksek Doğruluk (Accuracy %)
Uzun Arslan & Şenyer Yapıcı (2025)	Ekstra Ağaçlar Sınıflandırması + SMOTE	14	91,93
Ali Değirmenci (2025)	RF + Karşılıklı Bilgi	10 (seçilmiş)	92,36
Köklü & Sulak (2024)	RF	14	87,82
Bu çalışma	RF + CFS	9 (seçilmiş)	**93,41**
Bu çalışma	RF + CFS	7 (seçilmiş)	91,98

Tablo 6'da sunulan sonuçlar, önerilen yaklaşımın aynı veri seti üzerinde rapor edilen çalışmalarla karşılaştırıldığında yüksek sınıflandırma performansı sergilediğini göstermektedir. Bununla birlikte, çalışmalar arasında veri bölme stratejileri, ön işleme yöntemleri ve değerlendirme süreçleri bakımından farklılıklar bulunabileceğinden sonuçların doğrudan bir üstünlük göstergesi olarak değerlendirilmemesi gerekmektedir. Buna rağmen, herhangi bir veri dengeleme veya örnek çoğaltma yöntemi kullanılmadan ve öznitelik sayısı 14'ten 9'a düşürülerek yüksek performans değerlerinin elde edilmiş olması, öznitelik seçiminin sınıflandırma başarısındaki önemine işaret etmektedir. Elde edilen bulgular, tüm özniteliklerin sınıflandırma performansına aynı düzeyde katkı sağlamadığını ve uygun öznitelik seçiminin model başarısını artırabileceğini göstermektedir. Ayrıca öznitelik sayısının azaltılması, veri toplama sürecinin sadeleştirilmesine ve modelin yorumlanabilirliğinin artırılmasına katkı sağlamaktadır. Bu yönüyle önerilen yaklaşım, yalnızca yüksek doğruluk elde etmeyi değil, aynı zamanda daha yalın ve etkin sınıflandırma modellerinin geliştirilmesini de desteklemektedir.

ÇALIŞMANIN SINIRLILIKLARI VE GELECEK ÇALIŞMALAR

Bu çalışmanın bazı sınırlılıkları bulunmaktadır. İlk olarak, öznitelik seçimi işlemi çapraz doğrulama uygulamasından önce gerçekleştirilmiştir. Bu yaklaşım, özniteliklerin veri seti üzerindeki genel ayırt ediciliğini değerlendirmek amacıyla tercih edilmiş olmakla birlikte, performans tahminleri üzerinde sınırlı düzeyde iyimser bir etkiye neden olabilmektedir. Gelecekte gerçekleştirilecek çalışmalarda, bu etkinin tamamen ortadan kaldırılması amacıyla öznitelik seçiminin her çapraz doğrulama katmanı içerisinde yeniden uygulanması planlanmaktadır.

İkinci olarak, çalışma yalnızca Türkiye'de yaşayan bireylerden elde edilen tek bir veri seti üzerinde gerçekleştirilmiştir. Ayrıca çalışmada veri dengeleme amacıyla SMOTE veya benzeri sentetik örnek üretme yöntemleri kullanılmamış olup, sınıflandırma performansları özgün veri dağılımı üzerinde değerlendirilmiştir. Bu nedenle elde edilen sonuçların farklı popülasyonlara genellenebilirliği sınırlı olabilir. Ayrıca çalışmada yalnızca RF, ANN ve NB sınıflandırıcıları değerlendirilmiş olup, diğer güncel makine öğrenmesi ve derin öğrenme yöntemleri kapsam dışında bırakılmıştır.

Gelecekte gerçekleştirilecek çalışmalarda, önerilen yaklaşımın farklı obezite veri setleri üzerinde doğrulanması, öznitelik seçiminin her çapraz doğrulama katmanında yeniden uygulanması ve XGBoost, LightGBM, CatBoost ile derin öğrenme tabanlı yöntemler gibi güncel yaklaşımlarla karşılaştırılması planlanmaktadır. Ayrıca veri dengesizliğini gidermeye yönelik yöntemlerin ve farklı öznitelik seçimi stratejilerinin sınıflandırma performansı üzerindeki etkileri de araştırılabilir.

SONUÇ

Bu çalışma, çok sınıflı obezite tahmininde öznitelik indirgemesi ve çapraz doğrulama stratejilerinin model başarısı üzerindeki önemli etkisini göstermiştir. Elde edilen bulgular, tüm değişkenleri modele dahil etmek yerine hedef değişkenle güçlü ilişki gösteren özniteliklere odaklanmanın doğruluğu %87,82'den %93,41'e çıkarabildiğini ortaya koymaktadır. 14 öznitelikten 9'a indirgeme sayesinde hem aşırı öğrenme riski azalmış hem de model daha sağlam hale gelmiştir. Rastgele Orman algoritmasının farklı k değerleri altında sergilediği istikrarlı performans, önerilen yaklaşımın yüksek genelleme kabiliyetine sahip olduğunu doğrulamaktadır. Bu çalışmanın temel katkısı, yüksek doğruluk değerlerinin yalnızca karmaşık modeller veya ek veri işleme teknikleri ile değil, uygun öznitelik seçimi ve doğrulama stratejileri kullanılarak da elde edilebileceğini göstermesidir. Ayrıca farklı k değerlerinin sistematik olarak incelenmesi sayesinde model performansının kararlılığı değerlendirilmiş ve öznitelik sayısının azaltılmasının sınıflandırma başarısını olumsuz etkilemediği, aksine iyileştirebildiği ortaya konulmuştur. Bu yönüyle çalışma, obezite tahmini problemlerinde daha sade, yorumlanabilir ve düşük maliyetli modellerin geliştirilmesine katkı sağlamaktadır.

Sonuç olarak, Türkiye örneklemiyle gerçekleştirilen bu analiz, klinik karar destek sistemlerinde daha az veri girişiyle güvenilir teşhisler yapılabileceğini ortaya koymaktadır. Gelecek çalışmalarda bu modelin derin öğrenme yöntemleriyle hibrit hale getirilmesi veya farklı popülasyonlardan elde edilmiş daha geniş veri setleriyle test edilmesi, dijital sağlık araçlarının obeziteyle mücadeledeki etkinliğini daha da artıracaktır.

Yapay Zeka Katkı Beyanı

Bu makale tamamen herhangi bir yapay zeka aracının yardımı olmadan yazılmış, düzenlenmiş, analiz edilmiş ve hazırlanmıştır. Metin, veri analizi ve şekiller dahil tüm içeriğin yalnızca yazarlar tarafından oluşturulduğunu beyan ederim.

KAYNAKLAR

Acharjee, A., Ament, Z., West, J. A., Stanley, E., & Griffin, J. L. (2016). Integration of metabolomics, lipidomics and clinical data using a machine learning method. *BMC Bioinformatics*, 17(Suppl. 15), Article 440. https://doi.org/10.1186/s12859-016-1292-2

Arslan, R. U., & Yapici, I. S. (2025). A novel hybrid appRFch to enhancing obesity prediction. *The European Physical Journal Special Topics*, 234(15), 4843–4861. https://doi.org/10.1140/epjs/s11734-025-01620-0

Bolayır, B., Ölçer, D., Tekbıyık, B., Kara, G., & Çatuk, H. (2023). Obezite hastalığına sebep olan faktörlerin analitik hiyerarşi prosesi (AHP) yöntemiyle değerlendirilmesi. *Gümüşhane Üniversitesi Sağlık Bilimleri Dergisi*, 12(4), 1571–1589. https://doi.org/10.37989/gumussagbil.1366631

Börnhorst, C., et al. (2020). The role of lifestyle and non-modifiable risk factors in the development of metabolic disturbances from childhood to adolescence. *International Journal of Obesity*, 44, 2236–2245. https://doi.org/10.1038/s41366-020-00671-8

Canoy, D., & Buchan, I. (2007). Challenges in obesity epidemiology. *Obesity Reviews*, 8(Suppl. 1), 1–11. https://doi.org/10.1111/j.1467-789X.2007.00310.x

Chatterjee, A., Gerdes, M. W., & Martinez, S. G. (2020). Identification of risk factors associated with obesity and overweight—A machine learning overview. *Sensors*, 20, 2734. https://doi.org/10.3390/s20092734

Dugan, T. M., Mukhopadhyay, S., Carroll, A., & Downs, S. (2015). Machine learning techniques for prediction of early childhood obesity. *Applied Clinical Informatics*, 6, 506–520. https://doi.org/10.4338/ACI-2015-03-RA-0036

Ellis, K., Kerr, J., Godbole, S., Staudenmayer, J., & Lanckriet, G. (2016). Hip and wrist accelerometer algorithms for free-living behavior classification. *Medicine and Science in Sports and Exercise*, 48(5), 933–940. https://doi.org/10.1249/MSS.0000000000000840

Gutiérrez-Gallego, A., et al. (2024). Combination of machine learning techniques to predict overweight/obesity in adults. *Journal of Personalized Medicine*, 14, 816. https://doi.org/10.3390/jpm14080816

Karegowda, A. G., Manjunath, A. S., & Jayaram, M. A. (2010). Comparative study of attribute selection using gain ratio and correlation based feature selection. *International Journal of Information Technology and Knowledge Management*, 2(2), 271-277.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (pp. 1137–1145).

Köklü, N., & Sulak, S. A. (2024). Using artificial intelligence techniques for the analysis of obesity status according to the individuals' social and physical activities. *Sinop Üniversitesi Fen Bilimleri Dergisi*, 9(1), 217–239. https://doi.org/10.33484/sinopfbd.1445215

Mert, A., Kilic, N., & Akan, A. (2011, September). Breast cancer classification by using support vector machines with reduced dimension. In Proceedings of the 53rd International Symposium ELMAR-2011 (pp. 37–40). IEEE. https://doi.org/10.1109/ELMAR.2011.6059974

Mert, A., Kılıç, N., Bilgili, E., & Akan, A. (2015). Breast cancer detection with reduced feature set. *Computational and Mathematical Methods in Medicine*, 2015, Article 265138. https://doi.org/10.1155/2015/265138

Pal, M. (2005). Random forest classifier for remote sensing classification. *International Journal of Remote Sensing*, 26(1), 217–222. https://doi.org/10.1080/01431160412331269698

Refaeilzadeh, P., Tang, L., & Liu, H. (2009). Cross-validation. In L. Liu, & M. T. Özsu (Eds.), Encyclopedia of Database Systems (pp. 532–538). Springer. https://doi.org/10.1007/978-0-387-39940-9_565

Rish, I. (2001). An empirical study of the naive Bayes classifier. In Proceedings of the IJCAI 2001 Workshop on Empirical Methods in Artificial Intelligence (pp. 41–46).

Rodriguez-Galiano, V. F., Ghimire, B., Rogan, J., Chica-Olmo, M., & Rigol-Sanchez, J. P. (2012). An assessment of the effectiveness of a random forest classifier for land-cover classification. *ISPRS Journal of Photogrammetry and Remote Sensing*, 67, 93–104. https://doi.org/10.1016/j.isprsjprs.2011.11.002

Shi, X., et al. (2021). An ensemble-based feature selection framework to select risk factors of childhood obesity for policy decision making. *BMC Medical Informatics and Decision Making*, 21(1), 222. https://doi.org/10.1186/s12911-021-01580-0

Telleria-Aramburu, N., & Arroyo-Izaga, M. (2022). Risk factors of overweight/obesity-related lifestyles in university students: Results from the EHU12/24 study. *British Journal of Nutrition*, 127(6), 914–926. https://doi.org/10.1017/S0007114521001483

Triantafyllidis, A., et al. (2020). Computerized decision support and machine learning applications for the prevention and treatment of childhood obesity: A systematic review of the literature. *Artificial Intelligence in Medicine*, 104, Article 101844. https://doi.org/10.1016/j.artmed.2020.101844

Türkiye İstatistik Kurumu. (2022). Türkiye sağlık araştırması. https://data.tuik.gov.tr/ Erişim 16.02.2026.

Uğurlu, D., Yapıcı, H., Ünver, R., & Gülü, M. (2024). Comparison of obesity and physical activity levels of adult individuals by examining dietary habits with different parameters. *Journal of Health Sciences and Medicine*, 7(3), 301–307. https://doi.org/10.37944/jhsm.1450444

World Health Organization. (2025). Obesity and overweight. https://www.who.int/news-room/fact-sheets/detail/obesity-and-overweight/ Erişim 16.02.2026.

Wu, B., Zhang, L., & Zhao, Y. (2013). Feature selection via Cramer's V-test discretization for remote-sensing image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 52(5), 2593–2606. https://doi.org/10.1109/TGRS.2013.2263510

Yanıkkerem, E. (2017). Obezitenin Kadın Sağlığına Etkileri. *Kocaeli Üniversitesi Sağlık Bilimleri Dergisi*, 3(1), 37-43. https://doi.org/10.30934/kusbed.359281

Zou, J., Han, Y., & So, S. S. (2009). Overview of artificial neural networks. In D. J. Livingstone (Ed.), *Artificial Neural Networks: Methods and Applications* (pp. 14–22). Humana Press. https://doi.org/10.1007/978-1-60327-101-1_2

Zou, Q., Qu, K., Luo, Y., Yin, D., Ju, Y., & Tang, H. (2018). Predicting diabetes mellitus with machine learning techniques. *Frontiers in Genetics*, 9, Article 515. https://doi.org/10.3389/fgene.2018.00515