



# Düzce Üniversitesi Bilim ve Teknoloji Dergisi

*Araştırma Makalesi*

## Karar Destek Makineleri ve Rastgele Orman Ağaçları Yöntemleri ile Vücut Yağ Yüzdesinin Tahmini

 Burhan BARAKLI<sup>a</sup>,  Ahmet KÜÇÜKER<sup>a,\*</sup>

<sup>a</sup> *Elektrik Elektronik Mühendisliği Bölümü, Mühendislik Fakültesi, Sakarya Üniversitesi, Sakarya, TÜRKİYE*

\* *Sorumlu yazarın e-posta adresi: kucuker@sakarya.edu.tr*

DOI: 10.29130/dubited.815454

### ÖZET

Obezite, önemli bir sağlık problemidir ve yaygınlığı giderek artmaktadır. Bu hastalığın, diğer birçok hastalığın tetikleyicisi ve habercisi olduğu bilinmektedir. Obezite hastalığının tedavi sürecinden önce, vücut yağ yüzdesinin doğru bir şekilde tespit edilmesi önemlidir. Yüksek maliyetli yöntemler ile vücut yağ yüzdesi kesin olarak ölçülmektedir. Bu çalışmada, kişilerden alınan antropometrik veri seti ile vücut yağı yüzdesi tespitinin doğru ve maliyetsiz bir şekilde tahmin edilebilmesi için destek vektör regresyonu ile rastgele orman ağaçları regresyon yöntemleri uygulanmıştır. Regresyon yöntemlerinde, model parametre değerleri, veri sayısı, özellik sayısı ve özellik seçimi tahmin başarımında önemlidir. 13 özellikli vücut yağ yüzdesi veri seti baz alınarak çalışmada, literatürde sıklıkla kullanılan 25 istatistiksel yöntem (çarpıklık, merkezi moment, basıklık vb.) ile yeni bir veri seti oluşturulmuş ve elde edilen yeni veri setinin başarımlarının literatürde yer alan diğer çalışmalardan yüksek olduğu görülmüştür. Regresyon parametrelerinin ızgara tarama yöntemleriyle belirlenmesi sayesinde tahmin doğrulukları artırılmıştır. Ayrıca özellik azaltma yöntemleri ile vücut yağ yüzdesi ile yüksek ilintili özellikler belirlenmiştir. Seçilen özellikler ile gerçekleştirilen regresyon yöntemlerinin tahmin başarımlarının da benzer diğer çalışmalardan yüksek olduğu gözlenmiştir. En iyi ortalama karesel hata değerleri olarak, Rastgele Orman Ağaçları Yöntemi ve istatistiksel yöntemle oluşturulan yeni veri seti ile gerçekleştirilen deneyde 2,2519 değeri elde edilirken, Karar Destek Makinaları ve en iyi 6 F-skor değerine sahip özellikler ile yapılan regresyon deneyinde 3,174 değerine ulaşılmıştır.

**Anahtar Kelimeler:** *Vücut yağ yüzdesi, Destek vektör regresyonu, Rastgele orman ağaçları regresyonu*

## Estimation of Body Fat Percentage Using Support Vector Machine and Random Forest Methods

### ABSTRACT

Obesity is a significant health problem, and its prevalence is increasing. It is known that this disease is the trigger and precursor of many other diseases. Before the treatment of obesity disease, it is important to determine the body fat percentage correctly. Body fat percentage can be measured precisely with high-cost methods. However, these methods are not common to use. In this study, support vector regression and random forest tree regression methods were applied to accurately and cost-effectively estimate body fat percentage with anthropometric data set taken from individuals. In regression methods, model parameter values, number of data, number of features and feature selection are important in prediction performance. In the study based on a 13-featured body fat percentage data set, a new data set was created with 25 statistical methods (skewness, central moment, kurtosis, etc.) frequently used in the literature, and it was observed that the performance of the new data set was higher than other studies in the literature. Estimation accuracy has been increased by determining the regression parameters with grid scanning methods. In addition, features that are highly correlated with body fat percentage were determined with

feature reduction methods. It was observed that the prediction success performance of the regression methods performed with the selected features was higher than other similar studies. As the best mean square error values, the value of 2.2519 was obtained in the experiment performed with the Random Forest Trees Method and the new data set created by the statistical method, while the value of 3.174 was reached in the regression experiment with Decision Support Machines and the properties with the best 6 f-score values.

**Keywords:** *Body fat percentage, Support vector regression, Random forest*

---

Geliş: 23/10/2020, Düzeltme: 16/03/2021, Kabul: 25/03/2021

## I. GİRİŞ

Çağımızın önemli hastalıkları arasında olan obezite, insan hayatını olumsuz etkileyen ve birçok ölümcül hastalığın oluşmasına neden olan ciddi bir sorundur [1]. Obezite, vücutta bulunan yağ kütlelerinin, yağsız kütleyle oranla aşırı miktarda artması olarak tanımlanmaktadır. Bu sağlık problemi hem çocuklarda hem de yetişkinlerde giderek artan bir seyir izlemektedir. Dolayısıyla bireylere ve devletlere ekonomik bir yük getirmektedir [2].

Vücudun temel enerji kaynaklarından olan vücut yağlarının, vitaminlerin emilimi ve taşınması, vücut ısısının ayarlanması, soğuğa karşı koruma ve hayati organları koruma gibi görevleri vardır. Vücut yağ miktarının yetersiz olması durumunda ise üşüme, vitaminsizlik, halsizlik, çabuk hastalanma ve hormonal dengesizlik gibi sorunlar ortaya çıkmaktadır. Vücut yağ miktarının fazla olması durumunda da obezite problemi, Tip 2 diyabet şeker hastalığı, hipertansiyon, kalp hastalıkları, kısırlık, eklem hastalığı olan osteoartrit (kireçlenme), safra kesesi taşı, karaciğer yağlanması, uyku apnesi, venöz (toplardamar yetmezliği), cilt hastalıkları, solunum zorlukları, depresyon ve bazı tümörler dahil olmak üzere bir dizi sağlık problemine sebep olmaktadır [3-4].

Hasta olan kişilerin bir şekilde izlenmesi veya hastalığın teşhisi için, obezitenin kesin ölçüm sonuçlarına ihtiyaç bulunmaktadır. Obezite, aşırı vücut yağı durumu olarak tanımlanır ve doğrudan herhangi bir niceliksel ölçüt oluşturmaz [5]. Kilo, vücut yağının basit bir niceliksel örneğidir ve ölçülmesi kolaydır, ancak kişinin genel yapısı hakkında bilgi olmadan neredeyse anlamsız bir niceliktir. Ancak vücuttan elde edilen farklı niceliksel bilgiler ile anlamlandırılabilir tanımlamalar türetilmiştir. Örneğin Dünya Sağlık Örgütü tarafından kabul gören Vücut Kitle Endeksi (Body Mass Index (BMI)), vücut boy ve kilo bilgilerinin kullanılması ile formüle edilmiştir [6].

Obeziteyi ölçmek için yaygın olarak kullanılan BMI, ucuz ve uygun bir yoldur. Ancak BMI, yetersiz sayıda parametrelerden hesaplandığından sağlık sonuçlarını tahmin etmekte genellikle yetersiz performans göstermektedir [7]. Bel çevresi veya bel-kalça oranı gibi ilave parametreler ile BMI'nin yetersizlikleri giderilmeye çalışılmıştır [8]. Ayrıca BMI, yağ ve yağsız kütle arasında ayırım yapamayan beslenme durumunu göstermektedir [9]. Vücut yağı oranı ise, obeziteyi ölçmenin daha iyi bir yolu olan gerçek vücut kompozisyonu durumunu göstermektedir [10].

Vücut yağ kütlelerinin kendisinin veya vücut yağ yüzdesinin (Body Fat Percentage (BFP)) doğrudan ölçümü daha iyi bir yaklaşım olmasına rağmen, ölçümü zordur ve geniş bir kullanım için uygun değildir. Doğru ölçüm yapan yöntemler arasında su altı tartımı [11,12], X-ışını absorpsiyometrisi [13], biyoelektrik empedans analizi [14], manyetik rezonans görüntüleme [15], havada yer değiştirme pletismografisi [16] ve yakın kızılötesi etkileşim [17] gibi yöntemler bulunmaktadır. Ancak, bu yöntemler özel ekipman gerektirdiğinden, vücut yağının doğrudan ölçümü için zor, uygunsuz ve maliyetlidir. Bu nedenle, BFP'nin temel sosyodemografik veriler (yaş, cinsiyet), temel antropometrik veriler (ağırlık, boy, bel çevresi) ve rutin kan alımından elde edilen temel laboratuvar parametreleri gibi kolayca ölçülebilir parametrelerden tahmin edilebilmesi önemlidir [18-20].

BFP'nin tahminini bir regresyon problemi olarak ele alan birkaç çalışma bulunmaktadır. Bir tahmin modeli oluşturabilmek için kişinin yaşı, kilosu, boyu ve vücut çevresi gibi kolayca ölçülebilir özellikleri dikkate alan bir çalışma gerçekleştirilmiştir [21]. Destek vektör makinesi (Support Vector Machine (SVM)) kullanarak vücut yağını tahmin etmede etkili sonuçlar elde edilmiştir [22,23]. Yapay sinir ağlarını (YSA) [10,20] kullanarak yapılan çalışmalarda, BMI hesaplarını baz alan diğer çalışmalara [24-25] göre yüksek başarı oranı gözlenmiştir. Daha az parametre ile BFP tahmini için iyi sonuçlar elde edebilen yeni akıllı hibrit yaklaşımlar [26,27] tarafından önerilmiştir. Tıbbi cihazlardan, elektronik-tıbbi sağlık kayıtlarından, akıllı telefon uygulamalarından elde edilen verilerden BFP tahmini gerçekleştirilen bir çalışmada, lojistik regresyon, YSA ve karar ağacı dahil olmak üzere farklı makine öğrenme algoritmaları kullanılmıştır. Çalışmada regresyon tahmin modellerinin, BFP'nin tahmini için iyi performansla sahip olduğu belirtilmiştir [28]. Fotopletismografi ile elde edilen verilerden makine öğrenmesi yöntemleri kullanılarak BFP tahmininde etkili bir yöntem [29]'da sunulmuştur.

Literatürden elde edilen izlenim, regresyon yöntemlerinin uygun, maliyetsiz ve başarılı çıktılar ürettiğidir. Bu nedenle, bu çalışmada destek vektör regresyonu (SVR) [30] ve rastgele karar ağaçları (RF) [31] BFP'nin tespiti için kullanılmıştır. Gerçek bir veri tabanından elde edilen temel sosyodemografik veri seti (yaş, cinsiyet) ve temel antropometrik veri seti (ağırlık, boy, bel çevresi) ile BFP'in bu parametrelerden ne kadar iyi tahmin edilebileceği araştırılmıştır. Ayrıca orijinal veri setinden yeni özellikler çıkartılmış ve veri seti genişletilmiştir. BFP değerinin tahmini hem orijinal veri setinden hem de genişletilmiş veri setinden faydalanılarak tespit edilmeye çalışılmıştır. Bunların yanı sıra özellik seçimi ve özellik azaltımının regresyon yöntemlerinin başarısına olan katkısı incelenmiştir. Özellik seçimi için tek değişkenli doğrusal regresyon (TDDR) ve özellik azaltımı için ise temel bileşenler analizi (PCA) yöntemi uygulanmıştır.

## **II. GEREÇ VE YÖNTEMLER**

BFP tahmini genel blok diyagramı Şekil 1 ile verilmiştir. Orijinal veri setine, veri seti artırımı ile oluşturulan veri setlerine iki farklı regresyon yöntemi uygulanmıştır. Regresyon başarımını arttırmak için en iyi parametre tahmini algoritmaları kullanılarak regresyon parametrelerin doğru seçilmesi hedeflenmiştir. Eğitim verisi ve test verisinin başarımını doğru elde edebilmek amacıyla k katmanlı çapraz doğrulama algoritması kullanılmıştır.



*Şekil 1. Çalışmanın organizasyonu*

Bu çalışmada 252 kişiden elde edilen elde edilen 13 özellikli bir veri seti (vs1) üzerinde çalışılmıştır [32]. Ölçüm sonuçlarından elde edilen antropometrik verilerin (vs1=Tablo 1 [X1-X13]) açıklamaları Tablo 1'de verilmiştir. Veri setindeki BFP'nin belirlenmesi için [33]'ten yararlanılmıştır.

vs1'deki özellikler ve istatistik yöntemler uygulanarak 25 özellikli yeni bir veri seti (vs2=Tablo 1 [X14-X38]) oluşturulmuştur. Üçüncü bir veri seti olan vs3 ise vs1 ve vs2 birleştirilerek elde edilmiştir.

Bir bağımsız ile bir bağımlı değişken arasındaki ilişkiyi belirlemeye yarayan TDDR yöntemi (F-skor başarımı) kullanılarak, BFP ile ilgili özelliğin ilintisi Tablo 1'de üçüncü sütunda verilmiştir [34]. BFP ile ilintisi yüksek olan özellikler, vs1 veri setinde X6 adlı özellik (abdomen çevresi=karın çevresi) ve vs2 veri setinde ise dağılımın simetrik olup olmama derecesini belirten (çarpıklık) X15 adlı özellik olmuştur.

Orijinal veri seti vs1'in ikili özellikleri arasındaki ilişkinin dağılımı Şekil 2'de gösterilmiştir. Şekilden de görülebileceği gibi farklı iki özellik birbiri ile ilişkilidir. Bazı ikili özellikler doğrusal olmayan bir

şekilde ilişkili iken (örneğin, Bilek ve Yaş), çoğu özellik doğrusal olarak birbiri ile ilişkilidir (örneğin, Boyun ve Ağırlık, Göğüs ve Abdomen). Ek olarak, ağırlık özelliğinin diğer özelliklerle güçlü bir ilişkisi varken boy özelliği diğer özelliklerle daha zayıf ilişki göstermektedir.

**Tablo 1.** Veri setlerine ait özellikler

Ad	Özellik	F-skor	Ad	Özellik	F-skor
X1	Yaşı (Yıl)	5.47	X21	Hijort Hareketlilik Katsayısı	18.28
X2	Boy (cm)	0.34	X22	Hijort Karmaşıklık Katsayısı	12.26
X3	Ağırlık (kg)	25.06	X23	Maksimum değeri	0.34
X4	Boyun çevresi (cm)	15.31	X24	Medyan değeri	6.59
X5	Göğüs çevresi (cm)	36.89	X25	Mutlak Sapma	24.72
X6	Abdomen çevresi(cm)	56.68	X26	Minimum değeri	7.47
X7	Kalça çevresi (cm)	25.99	X27	Merkezi Moment	13.23
X8	Uyluk çevresi (cm)	19.21	X28	Ortalama değeri	3.31
X9	Diz çevresi (cm)	15.69	X29	Ortalama Eğri Uzunluğu	7.63
X10	Ayak Bileği çev. (cm)	4.18	X30	Ortalama Enerji	27.79
X11	Açık Kol Biceps çevresi (cm)	14.66	X31	Ortalama Karakök RMS değeri	28.72
X12	Ön kol çevresi (cm)	7.63	X32	Standart hata	10.20
X13	Bilek çevresi (cm)	7.47	X33	Standart Sapma	10.20
Y	Vücut Yağ Yüzdesi		X34	Şekil Faktörü	17.73
X14	Kurtosis - Basıklık	53.34	X35	Tekil Değer Ayrışımı	28.72
X15	Skewness - Çarpıklık	54.67	X36	"%25" için kesilmiş ortalama değeri	41.59
X16	Çeyrekler arası genişlik	45.49	X37	"%50" için kesilmiş ortalama değeri	41.02
X17	Değişim Katsayısı	25.18	X38	5 Ortalama Teager Enerjisi	3.31
X18	Geometrik ortalama	34.32			
X19	Harmonik ortalama	28.01			
X20	Hijort Aktivite Katsayısı	10.06			

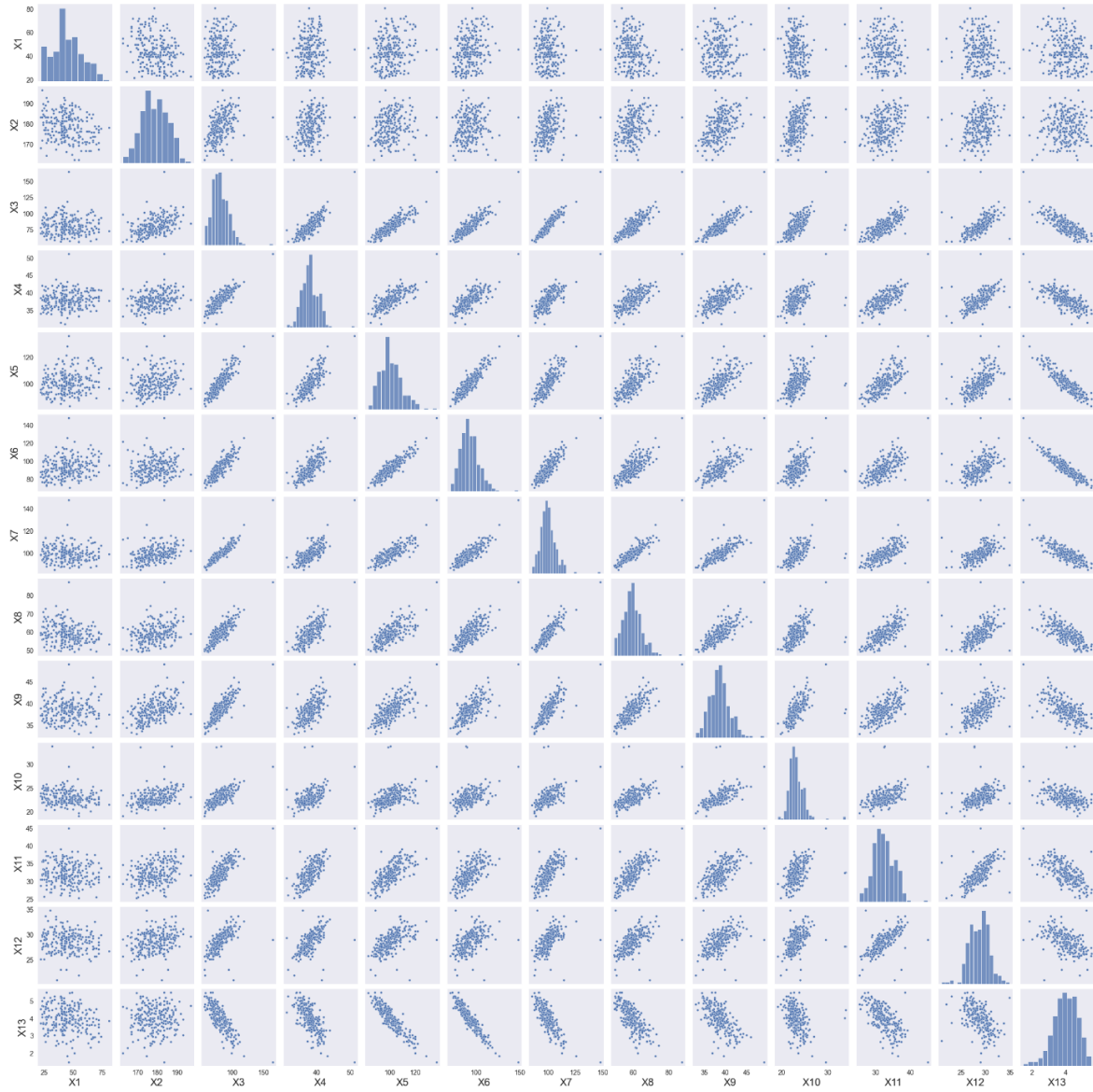
## A. TEK DEĞİŞKENLİ DOĞRUSAL REGRESYON ANALİZİ

Regresyon analizi yöntemi, değişkenler veya özellikler arasındaki ilintinin gücü ve yönü hakkında bilgi veren bir ölçüm yöntemidir. TDDG analizi [34] ise, bir bağımlı değişken ve bir bağımsız değişken arasındaki ilişkiyi sayısal olarak belirleyen bir analiz yöntemidir. Bu analizle bağımlı ve bağımsız değişkenler arasındaki ilişki bir doğru denklemi ile formüle edilmektedir. Vektör olarak verilen  $X1, X2, \dots, XM$  bağımsız değişkenler ve  $Y1, Y2, \dots, YM$  bağımlı değişkenlerin her biri normal dağılıma sahip iki popülasyondan aynı şekilde dağıtılmış örnekler olmak üzere bir lineer regresyon hipotez fonksiyonu Eş. 1. ile verilmektedir.

$$h_{\theta}(x) = \theta_1 x + \theta_0 \quad (1)$$

Eş. 2'de  $\theta$  parametrelerini optimum olarak hesaplayabilmek için bir maliyet fonksiyonu verilmiştir. Maliyet fonksiyonun minimum değerini hesaplamak amacıyla Bayır İniş Algoritması yöntemi [35] kullanılmış ve Eş. 3'te verilmiştir.  $\alpha$  öğrenme katsayısı ve  $m$  örnek sayısı olmak üzere, amaç fonksiyonu  $\theta_0^*, \theta_1^* = \underset{\theta_0, \theta_1}{\operatorname{argmin}} J(\theta_0, \theta_1)$ 'dir.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=0}^m (h(x^{(i)}) - y^{(i)})^2 \quad (2)$$



*Şekil 2. vs1 veri seti ikili özellikler arası ilişki*

Maliyet fonksiyonun minimum değeri hesaplanana kadar, Eş. 3'teki  $\theta_0$  ve  $\theta_1$  parametreleri bir iterasyon ile elde edilir.

$$(\theta_0, \theta_1) \leftarrow (\theta_0, \theta_1) - \alpha \frac{\partial J(\theta_0, \theta_1)}{\partial} \quad (3)$$

Bu çalışmada giriş özelliklerinin çıkış değişkenine katkısını araştırmak amacıyla her bir özellik için doğrusal regresyon oluşturulmuş ve en optimum  $y(x) = \theta_1 x + \theta_0$  belirlenmiştir. Her özelliğe ait başarımlar için TGDD yöntemindeki F-skoru hesaplanmıştır. Özellik sayısı  $k$  olmak üzere,  $\bar{X}[k]$  ve  $\bar{Y}[k]$  ortalama değer dizileri,

$$\bar{X}[k] = \frac{1}{m} \sum_{i=1}^m X_k[i] \text{ ve } \bar{Y}[k] = \frac{1}{m} \sum_{i=1}^m Y_k[i] \quad (4)$$

olarak Eşit. 4'teki gibi verilebilir. Varyans dizileri  $S_x^2[k]$  ve  $S_y^2[k]$  Eşit. 5.'te verilmiştir.

$$S_x^2[k] = \frac{1}{m-1} \sum_{i=1}^m (X_k[i] - \bar{X}[k])^2 \text{ ve } S_y^2[k] = \frac{1}{m-1} \sum_{i=1}^m (Y_k[i] - \bar{Y}[k])^2 \quad (5)$$

F-skoru Eşit. 6'daki gibi hesaplanır. Bu dizi, giriş özelliklerinin her biri için F-skor değerini vermektedir ve Tablo 1'de bu değerler üçüncü sütunda verilmektedir.

$$F[k] = \frac{S_x^2[k]}{S_y^2[k]} \quad (6)$$

## B. KARAR DESTEK MAKİNESİ REGRESYONU (SVR)

SVR çekirdek-tabanlı dönüşümlerde, giriş verileri yüksek boyutlu özellik uzaylarına taşınır ve bu uzayda doğrusal regresyon gerçekleştirilir.

Giriş verisi  $\{(x_i, y_i)\}_{i=1}^N$  olmak üzere, SVR yönteminin amaç ve kısıt ifadesi Eşit. 7 ile ifade edilir.

$$\underset{w, C, \varepsilon_i}{\operatorname{argmin}} \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*) \quad (7)$$

$$\text{Kısıt: } \begin{cases} y_i - \langle w, x_i \rangle - \theta_0 \leq \varepsilon + \varepsilon_i \\ \langle w, x_i \rangle + \theta_0 - y_i \leq \varepsilon + \varepsilon_i^* \\ \varepsilon_i, \varepsilon_i^* \geq 0 \end{cases}$$

Amaç fonksiyonundaki  $C$  regülasyon parametresi,  $w = (\theta_1, -1)$ ,  $\varepsilon$  esnek kenar payı,  $\varepsilon_i$  ve  $\varepsilon_i^*$  ise esnek değişkenlerdir. İki kısıtı olan doğrusal olmayan bir fonksiyonun çözümü için dual-çözüm önerilmektedir [36]. Dual çözüm için kullanılan Lagrange parametreleri, Eşit. 7'ye  $(\alpha, \beta)$  parametrelerinin eklenmesiyle oluşan lagrange fonksiyonu Eşit. 8'deki gibi elde edilmektedir.

$$L := \frac{1}{2} \langle w, w \rangle + C \sum_{i=1}^n (\varepsilon_i + \varepsilon_i^*) - \sum_{i=1}^n (\beta_i \varepsilon_i + \beta_i^* \varepsilon_i^*) - \sum_{i=1}^n \alpha_i (\varepsilon + \varepsilon_i - \langle w, x_i \rangle + \theta_0 - y_i) - \sum_{i=1}^n \alpha_i^* (\varepsilon + \varepsilon_i^* + y_i - \langle w, x_i \rangle - \theta_0) \quad (8)$$

L fonksiyonunu en küçükleyen  $w, \theta_0, \varepsilon$  ve  $\varepsilon_i^*$  parametreleri için kısmi türevlerin alınıp, Eşit. 6'da yerine yazılırsa amaç ve kısıt fonksiyonu Eşit. 7'deki gibi oluşur.

$$\begin{cases} \max_{\alpha_i, \alpha_i^*} L = \frac{1}{2} \sum_{i,j=1}^n (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle - \varepsilon \sum_{i=1}^n (\alpha_i + \alpha_i^*) + y_i \sum_{i=1}^n (\alpha_i - \alpha_i^*) \\ \text{kısıt fonksiyonu: } \sum_{i=1}^n (\alpha_i - \alpha_i^*) = 0 \text{ ve } \alpha_i \alpha_i^* \in [0, C] \end{cases} \quad (9)$$

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) x_i$$

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + \theta_0 \quad (11)$$

Eş. 10 ve Eş. 11 ile verilen denklemler doğrusal SVR olarak adlandırılmaktadır. Ancak, Eş. 11 doğrusal olmayan problemlerin çözümünde uygun olmayan sonuçlar üretebilmektedir. Bu nedenle giriş özellikleri bir üst uzaya taşınmaktadır ve üst uzayda uygulanacak olan lineer regresyonlar daha başarılı sonuçlar vermektedir.  $\Phi: R^2 \rightarrow R^3$  haritalamayı ve çekirdek hilesi olan  $K(x_i, x) = \langle \Phi(x_i), \Phi(x) \rangle$  olmak üzere, çekirdeğin Eş. 10 ve Eş. 11'de yerine yazılmasıyla doğrusal olmayan problemlerinde başarılı çıktılar üreten SVR regresyonu elde edilmektedir.

$$w = \sum_{i=1}^n (\alpha_i - \alpha_i^*) \Phi(x_i) \quad (12)$$

$$f(x) = \sum_{i=1}^n (\alpha_i - \alpha_i^*) k(x_i, x) + \theta_0 \quad (13)$$

Bu çalışmada sıkça kullanılan ve doğrusal olmayan problemlerde başarılı sonuçlar veren Radial temelli fonksiyon (RBF) kullanılmıştır. RBF kerneli  $k(x_i, x) = \exp(-\gamma \|x_i - x_j\|^2)$  olarak tanımlanmıştır.

### C. RASTGELE ORMAN AĞAÇLARI (RF) İLE REGRESYON

Bir karar ağacı, hiyerarşik olarak organize edilmiş ve bir kökten ağacın bir uç düğümüne veya yaprağına arka arkaya uygulanan bir dizi kısıtlama veya koşulu temsil etmektedir [37]. Karar ağacı, bir kök düğümü ile başlar. Bu düğümde, karar ağacı öğrenme algoritmasına göre bölünmeler gerçekleşir. Nihai sonuç, her dalın olası bir karar senaryosu ve sonucunu temsil ettiği bir karar ağacıdır. RF yöntemi ise, birden fazla karar ağacını oluşturur ve daha doğru ve istikrarlı bir tahmin elde etmek için karar ağaçlarını birleştirilir [38]. RF ile bir  $x$  değerine ait tahmini hesaplamak için, her bağımsız ağaç için Eş. 14 kullanılır.

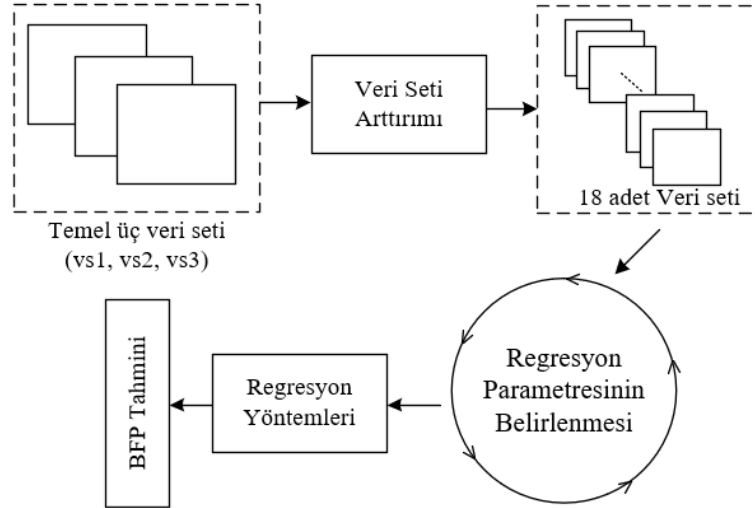
$$\widehat{f}_{rf}^B(x) = \frac{1}{B} \sum_{b=1}^B T(x; \Theta_b) \quad (14)$$

$\Theta_b$ , b inci rasgele orman ağacını, bölünmüş değişkenler, düğümlerdeki kesme noktalarını ve B ağaç sayısını ifade eder.

## III. DENEYSEL ÇALIŞMALAR ve SONUÇLARI

Bu çalışmada BFP değerinin regresyon yöntemleriyle tahmini için gerçek bir veri seti (vs1) kullanılmaktadır. Bu veri setinden ilk önce vs2 ve vs3 oluşturulmuş, ardından TGDD ve PCA yöntemleri ile BFP değeri ile yüksek ilintili belli sayıda özellikler seçilerek ve temel bileşenler belirlenerek 15 adet veri seti daha üretilmiştir. Veri setlerinin özeti Tablo 2'de açıklanmıştır. Tablo 2'de verilen 18 veri setine SVR ve RF regresyon yöntemleri uygulanmaktadır. Örnek olarak, S3 adlı veri seti; 35 özellikli bir veri setidir. S11 adlı veri seti; vs1 veri setine PCA yöntemi uygulanarak elde edilen iki temel bileşen özelliğine sahip veri seti ve S15 ise vs2 veri setindeki F-skor değeri en iyi olan ilk 6 değişken özellikli bir veri setidir.

BFP tahmini için oluşturulan regresyon çalışmasının blok akış şeması Şekil 3'te verilmiştir. Regresyon yöntemlerinde eğitim parametrelerinin değerleri önemlidir. Bu nedenle en uygun regresyon parametrelerinin belirlenmesi için bir yöntem gereklidir. Bu amaçla, bir döngü oluşturularak (grid arama) regresyonlara ait en iyi sonucu veren parametreler belirlenmiş ve regresyon yöntemleri 20 kez koşturularak, sonuçların ortalaması alınmıştır.



Şekil 3. Yöntemlere ait akış diyagramı

Tablo 2. Regresyon yöntemlerinin uygulandığı 18 veri setinin adları ve açıklamaları

Veri Seti Adı	Uygulama Veri Seti	Özellik Sayısı	Veri Setinin Açıklaması
S1	vs1	13	Orijinal veri seti
S2	vs2	25	İstatiksel yeni özellikler
S3	vs3	38	vs1+vs2 veri seti
S4	vs1	2	PCA bileşenleri
S5	vs2	2	PCA bileşenleri
S6	vs3	2	PCA bileşenleri
S7	vs1	4	PCA bileşenleri
S8	vs2	4	PCA bileşenleri
S9	vs3	4	PCA bileşenleri
S10	vs1	2	F-skor ile elde edilen özellikler
S11	vs1	4	F-skor ile elde edilen özellikler
S12	vs1	6	F-skor ile elde edilen özellikler
S13	vs2	2	F-skor ile elde edilen özellikler
S14	vs2	4	F-skor ile elde edilen özellikler
S15	vs2	6	F-skor ile elde edilen özellikler
S16	vs3	2	F-skor ile elde edilen özellikler
S17	vs3	4	F-skor ile elde edilen özellikler
S18	vs3	6	F-skor ile elde edilen özellikler

Son adımda antropometrik ölçümler ile BFP arasındaki ilişki istatistiksel olarak analiz edilmiştir. Bu şekilde orijinal özelliklerin, türetilen özelliklerin, her iki özellik setinin birleştirilmesiyle oluşturulan toplam özelliklerin, PCA ile elde edilen temel bileşenlerin ve BPF ile ilintisi yüksek olan özellikler ile oluşturulan tüm veri setinin başarımları karşılaştırılmıştır.



## A. REGRESYON PARAMETRELERİ VE KARŞILAŞTIRMA METRİKLERİ

Parametre ayarları ve doğrulama için, sırasıyla çapraz doğrulama ile grid arama teknikleri bir arada kullanılmıştır [39]. Regresyon yöntemlerinin başarımları için literatürde sıklıkla kullanılan karşılaştırma metrikleri kullanılmıştır.

Sonuçlara göre, SVR ve RF yöntemlerine ait parametreler Tablo 3 ve Tablo 4'deki gibi belirlenmiştir:

*Tablo 3. SVR yöntemi sonuçlarının alınmasında kullanılan parametreler*

Veri Seti Adı	Regülasyon Parametresi	SVR Çekirdeği	Gamma Parametresi
vs1	1100	RBF	0,1
vs2	2000	RBF	0,0001
vs3	0.1	RBF	0,001

*Tablo 4. RF yöntemi sonuçlarının alınmasında kullanılan parametreler*

Veri Seti Adı	Ağaç sayısı	Max. Ağaç Derinliği	Min. Düğüm Bölücü Sayısı	Yaprak Düğümde Min. Örnek Sayısı
vs1	391	50	5	1
vs2	500	50	5	4
vs3	64	100	2	2

SVM ve RF yöntemlerinin başarımları belirlemek amacıyla literatürde sıklıkla kullanılan, *ortalama mutlak yüzde hatası* (MAPE), *ortalama karesel hata* (RMSE), *medyan mutlak hata* (MAE), *determinasyon katsayısı* (R<sup>2</sup>), *korelasyon katsayısı* (R) kullanılmaktadır. Ayrıca veri setinin %80'i eğitim için, %20'si test ve skorlar için ayrılmıştır.

Deneylerimizde kullanılan vücut yağ veri setlerinin ölçeği dikkate alınarak, karşılaştırılan regresyon modellerinin performansını değerlendirmek için beş kat çapraz doğrulama uygulanmaktadır. Beş kat çapraz doğrulamada ise N numuneli bir veri seti verildiğinde, her setin N / 5 numunesi test için ve geri kalanı eğitim için kullanılmıştır. Veriler rastgele veri setinden seçilmiş ve her deney 20 kez tekrarlanmıştır. Ardından nihai sonuç için ortalama alma işlemi gerçekleştirilmiştir.

R<sup>2</sup>, model tarafından görünmeyen örneklerin ne kadar iyi tahmin edilebileceğinin bir ölçüsünü sağlar. n veri sayısı,  $\bar{y}$  gerçek çıkış değerlerinin aritmetik ortalaması ve  $\hat{y}_i$ , i. çıkış örneği  $y_i$ 'nin tahmini olmak üzere R<sup>2</sup>, Eşit. 15.'deki denklem ile ifade edilir.

$$R^2(y, \hat{y}) = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (15)$$

Eşit. 16 ile formüle edilen MAE, mutlak hata kaybının veya l1-norm kaybının beklenen değerine karşılık gelen bir risk ölçütüdür.

$$MAE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (16)$$

MSE ve RMSE hata veya kaybın beklenen değerine karşılık gelen bir risk ölçütüdür ve sırasıyla Eşit. 17 ve Eşit. 18 ile hesaplanmaktadır.

$$MSE(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (17)$$

$$RMSE(y, \hat{y}) = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (18)$$

## B. DENEYSEL SONUÇLAR

Bu çalışmada, kişilerden elde edilen antropometrik veri seti ile BFP oranının RF ve SVR yöntemleri ile tespit edilmesi amaçlanmıştır. 13 özellikli bir veri setinin, başarımlarını arttırabilmek adına özellik sayısında bir arttırıma gidilmesi hedeflenmiştir. Ancak yeni eklenecek özelliklerin, orijinal özellikleri doğru temsil edebilmesi ve ilişkileri daha iyi ortaya göstermesi gerekmektedir. Bu nedenle orijinal özellikler ve önerilen istatistiksel yöntemler kullanılarak (Tablo 1 [X14-X38]) yeni bir veri seti oluşturulmuştur. Bu veri seti orijinal özellikleri içermemekte ve tamamen istatistiksel yöntemler kullanılarak elde edilmektedir. Orijinal veri seti, oluşturulan yeni veri seti ve her iki veri setinin birleşiminden oluşturulan üç veri seti ile RF ve SVR testleri gerçekleştirilmiştir. Ayrıca her veri setine özellik azaltma ve seçme teknikleri kullanılarak deneyler gerçekleştirilmiştir. Bu amaçla iki yöntem seçilmiştir. Birincisi PCA ve ikinci TDDG analizidir. PCA ile tüm özellikleri barındıracak temel bileşenler belirlenerek, TDDG ile BFP oranı ile ilintili olan en iyi özellikler seçilerek deneyler gerçekleştirilmiştir. RF ve SVR yöntemine ait doğru parametreler, çapraz doğrulama ve grid arama teknikleri kullanılarak elde edilmiştir.

Tablo 5 ve Tablo 6'da deney sonuçları verilmiştir. SVR ile yapılan 18 eğitimden, orijinal veri seti S1 ve S1'den türetilen S2 ve S3 veri setleriyle yapılan regresyonlar, BFP değeri ile yüksek ilintili özelliklerle eğitilmiş deneylere oranlara daha düşük başarımlar ürettiği gözlemlenmiştir. S1 ve S2, S3 özellik sayısı en fazla olan veri setleridir. Ancak barındıkları bazı özellikler BFP ile zıt korelasyona sahiptir. Bu nedenle BFP ile doğrusal olan özellikler ile yapılan deneyler doğru tahmin değerleri vermektedir. Öncelikle sonuç tablolarındaki SVR1 deneyi, orijinal özellikleri içeren deney seti iken, SVR3 deneyi ise en büyük özellik sayısına sahip veri setinden gerçekleştirilen deneylerdir. SVR18 adlı deney, sadece en iyi 6 F-skor değerine sahip özelliklerle gerçekleştirilen deney olmasına rağmen en iyi SVR sonucunu vermektedir. Benzer şekilde SVR 17 adlı deney, en iyi ikinci başarılı sonucu ürettiği gözlemlenebilir. Sırasıyla SVR 17 ve SVR18 deneyleri, 38 özellikli veri setlerinden elde edilen 4 ve 6 özellikli veri setleri olmasına karşın, sadece 13 (vs1) özellikli orijinal veri setinden elde edilen en iyi 6 ve 4 F-skor değerine sahip SVR12 ve SVR11 adlı deneylerde başarılı sonuçlar ürettiği tespit edilmiştir. Verilen açıklamaların sayısal olarak örneklendirmek istersek en iyi RMSE sonuçları SVR18, SVR17 ve SVR12 deneyleri için sırasıyla, 3,1741, 3,4366 ve 3,5378'dir, benzer bir şekilde R2 sonuçları SVR18, SVR17 ve SVR11 deneyleri için sırasıyla, 0,823, 0,7925 ve 0,7763 olmaktadır. SVR sonuçlarından, en düşük eğitim sonucunu veren deneyler, orijinal veri seti ile yapılan SVR1 deneyi ile türetilmiş veri seti vs2'ye PCA yöntemi uygulanarak temel 2 bileşen içeren SVR8 adlı deneydir. Ancak SVR2 deney sonuçları için bir parantez açmak gerekir. SVR2 deneyi, orijinal veri setinden hiçbir özellik içermemektedir. Buna karşın, deneyde ulaşılan sonuçlar diğer deneylere nazaran tatmin edicidir.

Tablo 5'den elde edilen SVR sonuçlarından faydalanılarak BFP değerinin tespitinde SVR ile yapılan eğitimler için çıkarılabilecek sonuç; orijinal veri seti ile yapılan eğitimler, kendisinden türetilen yeni özellikli test verileri ile yapılan eğitimlerden daha düşük verebilmektedir. Ayrıca hem orijinal veri setinden hem de türetilen veri setinden elde edilen belli sayıda en iyi F-skor değerine sahip özelliklerle yapılan deneyler en başarılı sonuçları vermektedir. Ayrıca PCA yöntemi ile elde edilen özelliklerle yapılan eğitimlerin başarımlarının, bazı metrikler bakımından diğer deneylerden düşük olduğu tespit edilmiştir.

RF yöntemi ile yapılan deneylerin başarımları Tablo 6 ile verilmiştir ve SVR yöntemi ile yapılan eğitimlere göre çok farklı sonuçlar göstermektedir. Örneğin RF yöntemi ile yapılan deneylerde, temel

bileşenler ile F-skor değeri yüksek olan özellikler seçilerek yapılan eğitimler, orijinal özelliklerle yapılan eğitimden başarılı sonuç vermemiştir. Bu durum SVR ile yapılan deneylerde tersi olduğu görülmektedir. Örnek olarak, RF1, RF2 ve RF3 adlı deneylerde yapılan tüm eğitimler, diğer deneylerden daha başarılı çıktılar üretmiştir. Ancak RF2'nin başarımı dikkat çekmektedir. S2 veri setindeki özellikler, orijinal özelliklerin hiçbirini içermemektedir ancak orijinal verilerden elde edilen bir eğitim veri setiydi. Niceliksel olarak bu durum örneklenir ise, en yüksek ve en düşük tahmin doğruluğunu veren deneyler sırasıyla RF2 ve RF10 adlı deneylerdir. Sırasıyla RMSE'de 2,2519 ile 5,4905, MAE'de 2,006 ile 4,7847, R2'de 0,9267 ile 0,2097 gibi sonuçlar elde edilmiştir. En iyi üç sonucu veren deneyler dışında, orijinal veri setinin F-skor değeri en yüksek olan 2 özelliğini kullanan S10 veri seti de dikkat çekmektedir. Dolayısıyla RF yöntemi için üretilebilecek yeni veri setlerinin, orijinal veri setine göre daha başarılı sonuçlar verebileceği kanaatine varılabilir. Örneğin, RMSE değeri en düşük olan ilk 4 deney RF1, RF2, RF3 ve RF10 için sırasıyla 2,2519, 2,4389, 2,4974 ve 2,4982 değerlerine ulaşmıştır.

*Tablo 5. SVR performans sonuçları*

Deney Adı	Uygulama Veri Seti	RMSE	MSE	MAE	R2	R
SVR1	S1	4,5806	20,9818	3,793	0,532	0,7757
SVR2	S2	3,5711	12,7528	2,9465	0,7748	0,8755
SVR3	S3	3,5593	12,6685	2,9239	0,7763	0,8911
SVR4	S4	4,1497	17,2201	3,4903	0,6773	0,774
SVR5	S5	4,1871	17,532	3,4533	0,6904	0,8639
SVR6	S6	4,2826	18,3407	3,5566	0,6761	0,8637
SVR7	S7	3,8922	15,1492	3,2427	0,7325	0,8863
SVR8	S8	5,2748	27,8231	4,2458	0,3674	0,7506
SVR9	S9	4,6506	21,6279	3,7928	0,5231	0,7548
SVR10	S10	3,6788	13,5334	3,1146	0,761	0,881
SVR11	S11	3,5589	12,6654	3,015	0,7763	0,8962
SVR12	S12	3,5378	12,5159	2,9994	0,779	0,8772
SVR13	S13	6,1539	37,8704	4,8844	0,132	0,5222
SVR14	S14	4,6244	21,385	3,764	0,5285	0,7273
SVR15	S15	3,8982	15,1956	3,308	0,697	0,8226
SVR16	S16	3,6314	13,1872	2,8897	0,7683	0,8948
SVR17	S17	3,4366	11,8099	2,7079	0,7925	0,8932
SVR18	S18	3,1741	10,0747	2,4865	0,823	0,8756

*Tablo 6. RF performans sonuçları*

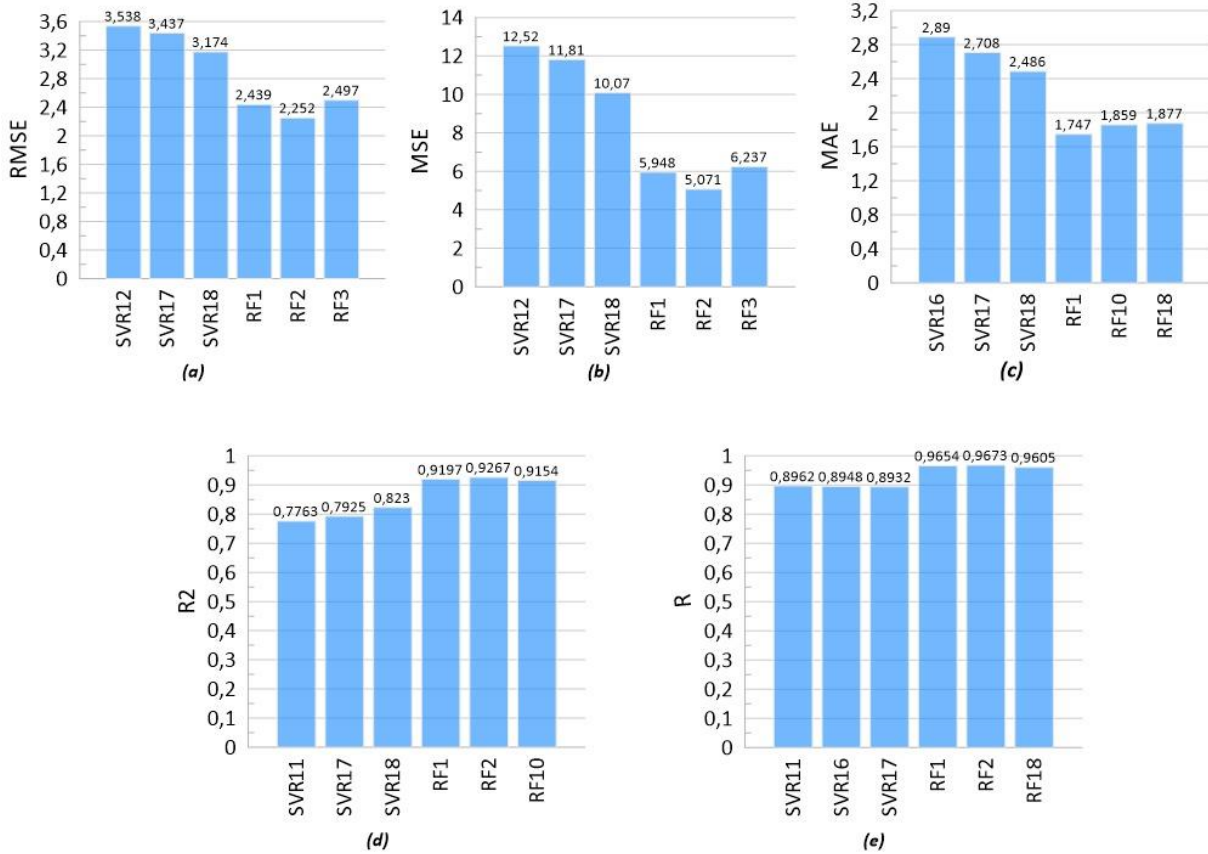
Deney Adı	Uygulama Veri Seti	RMSE	MSE	MAE	R2	R
RF1	S1	2,4389	5,9483	1,747	0,9197	0,9654
RF2	S2	2,2519	5,0712	2,006	0,9267	0,9673
RF3	S3	2,4974	6,237	2,226	0,9099	0,968
RF4	S4	3,6101	13,0325	2,873	0,7964	0,8586
RF5	S5	3,1271	9,779	2,7175	0,8106	0,9308
RF6	S6	3,2379	10,484	2,6975	0,8156	0,8909
RF7	S7	2,9029	8,4268	2,206	0,8779	0,886
RF8	S8	3,3129	10,9754	2,748	0,7949	0,8525

Tablo 6. (devam) RF performans sonuçları

Deney Adı	Uygulama Veri Seti	RMSE	MSE	MAE	R2	R
RF9	S9	2,7738	7,6941	2,2515	0,8069	0,9252
RF10	S10	2,4982	6,2411	1,859	0,9154	0,9489
RF11	S11	2,6986	7,2823	2,2325	0,8808	0,9203
RF12	S12	2,7571	7,6017	2,4592	0,8756	0,9398
RF13	S13	5,4905	30,1455	4,7847	0,2097	0,5656
RF14	S14	2,9058	8,4439	2,3745	0,8599	0,9206
RF15	S15	3,0469	9,2837	2,369	0,8658	0,9183
RF16	S16	2,6897	7,2347	1,9757	0,8456	0,8887
RF17	S17	2,8125	7,91	2,2205	0,8857	0,9153
RF18	S18	2,4808	6,1544	1,8765	0,911	0,9605

## IV. SONUÇ

BFP oranının tahmininde, RF yönteminin SVR yönteminden karşılaştırma metrikleri bakımından daha üstün olduğu tespit edilmiştir. Şekil 4'te SVR ve RF yöntemlerinin verilen metriklere göre her birisine ait en iyi sonuç üreten ilk 3 sonuç gösterilmektedir. Şekil 4 (a) ve Şekil 4 (b)'de RMSE ve MSE metriklerinde RF2 deneyi, Şekil 4 (c)'de MAE metriğinde RF1 deneyi, Şekil 4 (d)'de R2 metriğinde RF2 deneyi ve Şekil 4 (e)'de R metriğinde RF2 deneyi BFP değerini en iyi tahmin eden deneyler olduğu görülmektedir.



Şekil 4. SVR ve RF deneylerindeki en iyi 3 performansın karşılaştırılması

Şekil 4.a'da verilen RMSE kıstası bakımından, istatikselsel olan özellikler (önerilen istatistiki yöntemler ile oluşturulan vs2 veri seti, 25 özellikli) ile gerçekleştirilen RF deneyi 2,252 ile en iyi tahmin skoruna sahip iken, SVR ile yapılan deneylerde 38 özellikli (vs1+vs2) veri setinden elde edilen en iyi 6 F-skor değerine sahip özellikler ile gerçekleştirilen deney 3,174 skor değerine ulaşılmıştır. MSE kriteri bakımından başarımlar değerleri, RMSE başarımlar değerleri ile benzer olacağı aşıkardır.

Şekil 4.c'de MAE değerleri için, RF1 uygulanmasında, orijinal veri setinde (vs1) en iyi skor değerine ulaşılmıştır. Ancak orijinal veri setinden elde edilen ve sadece en iyi 2 F-skor değeri olan özellikler ile gerçekleştirilen RF10 deneyi, neredeyse RF1 deneyi kadar etkili olduğu görülmektedir. RF1 ve RF10 için sırasıyla 1,747 ve 1,859 skor değerleri elde edilmiştir. MAE için, SVR tahminlerinde, en geniş veri setinde (vs3 veri seti, 35 özellikli) F-skor değerleri için yapılmış deneylerin tahmin değerleri göze çarpmaktadır. En iyi 6, 4 ve 2 sayıda F-skor değerine sahip özellikler ile yapılan deneylerde sırasıyla 2,486, 2,708 ve 2,89 değerlerine ulaşılmıştır.

RF deneylerinde Şekil 4.d'de verilen R2 kıstasına göre, orijinal ve türetilen yeni veri seti ile gerçekleştirilen tahminlerin (RF1 ve RF2 deneyleri) başarımlar oranlarının yüksek olduğu tespit edilmiştir. Ancak orijinal veri setine ait en iyi 2 F-skor değeri ile elde edilen özellikler ile yapılan deney (RF10) neredeyse orijinal veri seti ile yapılan deneye eşdeğerdir. SVR deneyleri için ise, 35 özellikli (vs3 yani orijinal ve türetilen veri setlerinin birleşimi) veri setlerinden elde edilen en iyi 4 ve 6 özellikli veri setlerinin başarımlar diğer SVR deneylerinden yüksektir. R kriteri bakımından başarımlar değerleri, R2 başarımlar değerleri ile benzerdir.

Önerilen yöntemlerin, literatürde yer alan diğer çalışmalar ile başarımlar karşılaştırması Tablo 7'de verilmiştir. Tabloda verilen tüm çalışmalarda, [31] çalışmasından elde edilen antropometrik veri seti üzerinde yöntemler uygulanmıştır. SVR yöntemine yeni bir parametre eklenerek yeni bir yöntem olarak sunulan IRE-SVM [23] ile yapılan BFP tahmin sonuçları Tablo 7'de ilk sırada verilmiştir. Tablonun 2 ile 6 arasındaki sonuçlar, hibrit makine öğrenmesi yöntemleri uygulanarak elde edilmiştir [27]. 7 ve 8 satırdaki sonuçlar, standart RF ve SVM parametreleri ile elde edilmiştir. 9 ile 13 arasındaki sonuçlar, bu çalışmadan elde edilen en iyi tahmin sonuçlarının ulaşıldığı deneylerdir. Tabloda boş olarak verilen sonuç değerleri ilgili çalışmalardan elde edilememektedir.

*Tablo 7. Literatür ile karşılaştırma*

Sıra	Literatür çalışmaları ve önerilen yöntemler	Uygulama Veri Seti	RMSE	MAE	R	R2
1	[23]-IRE-SVM	S1	4,4334	3,6261	-	-
2	[27]-MLFNN	S1	8,535	-	0,451	0,203
3	[27]-MLFNN+DT	S1	5,832	-	0,656	0,43
4	[27]-MLFNN+SVM	S1	5,662	-	0,631	0,398
5	[27]-DT+SVM	S1	4,817	-	0,74	0,547
6	[27]-MLFNN+DT+SVM	S1	11,33	-	0,717	0,514
7	Standart SVR	S1	5,32	4,17	0,69	0,47
8	Standart RF	S1	4,218	3,826	0,7652	0,548
9	SVR1	S1	4,5806	3,793	0,7757	0,532
10	SVR18	S18	3,174	2,486	0,8756	0,823
11	RF1	S1	2,4389	<b>1,747</b>	0,9654	0,9197
12	RF2	S2	<b>2,2519</b>	2,006	0,9267	<b>0,9673</b>
13	RF18	S18	2,4808	1,8765	0,9605	0,911

Karşılaştırma metrikleri bakımından, istatistiki yöntemler kullanılarak elde edilen yeni veri seti ile yapılan RF deneyi (Tablo 7, satır 12) en başarılı performansı göstermiştir. Benzer şekilde hem orijinal hem de türetilen veri setlerinin birleşimden oluşturulan veri setinden elde edilen sadece 6 özellikli veri setinde (Tablo 7, satır 13) başarılı tahmin sonuçlarına ulaşılmıştır. En iyi performans sonuçları olarak RMSE, MAE, R2 başarımlar kriterleri bakımından sırasıyla RF2 deneyinde 2,2519, RF1 deneyinde 1,747 ve RF2

deneyinde 0,9673 deęerleri elde edilmiřtir. SVR yntemlerinde, [23] alıřmasında RMSE deęeri 4,4334, standart SVR deneyinde 5,32 ve SVR18 deneyinde 3,174 deęerlerine ulařılmıřtır. SVR18 deneyi, 38 zellikli tretilmiř veri setinden elde edilen en iyi 6 F-skor deęerine sahip zellikler ile oluřturulmuř veri setine SVR uygulanmıř bir deneydir.

BFP deęerinin, SVR ve RF yntemleriyle tahmininde olduka bařarılı sonulara ulařılmaktadır. 13 zellikli 252 veriye sahip bir veri setinde regresyon yntemleri ile BFP deęeri yksek doęrulukta tahmin edilebildięi grlmektedir. Ancak RF yntemi, SVR yntemlerine nazaran daha doęru tahmin deęerlerine ulařtıęı sonulardan grlmektedir. Ayrıca bir veri setinden retilbilecek yeni veri setlerinin ve veri setlerine uygulanabilecek zellik seme yntemleri ile oluřturulacak yeni veri setlerinin bařarılı ıktılar reteceęi bu alıřmada gsterilmiřtir.

## **V. KAYNAKLAR**

- [1] F. McLellan, "Obesity rising to alarming levels around the world," *The Lancet*, c. 359, s. 9315, ss. 1412, 2002.
- [2] C. L. Edelman, C. L. Mandle ve E. C. Kudzma, *Health Promotion Throughout the Life Span-E-Book*, 9. baskı, Missouri, United States of America: Elsevier Health Sciences, 2017, bl. 2, ss. 23-24.
- [3] I. G. Polat, "Effect of Er stress and Sik2 Reciprocal relationship on human precursor fat cell (LiSa-2) differentiation," Doktora Tezi, Gebze Teknik niversitesi, Kocaeli, Trkiye, 2017.
- [4] F. Ortega, C. Lavie ve S. Blair, "Obesity and cardiovascular disease," *Circulation Research*, c. 118, s. 11, ss. 1752-1770, 2016.
- [5] C. Lavie, A. Schutter, P. Parto, E. Jahangir, P. Kokkinos, F. Ortega, R. Arena ve R. Milani, "Obesity and prevalence of cardiovascular diseases and prognosis—the obesity paradox updated," *Progress in Cardiovascular Diseases*, c. 58, s. 5, ss. 537-547, 2016.
- [6] A. Keys, F. Fidanza, M. Karvonen, N. Kimura ve H. Taylor, "Indices of relative weight and obesity," *Journal of Chronic Diseases*, c. 25, s. 6-7, ss. 329-343, 1972.
- [7] R. Huxley, S. Mendis, E. Zheleznyakov, S. Reddy ve J. Chan, "Body mass index, waist circumference and waist:hip ratio as predictors of cardiovascular risk," *Obesity and Metabolism*, c. 8, s. 1, ss. 69-69, 2011.
- [8] C. Lee, R. Huxley, R. Wildman ve M. Woodward, "Indices of abdominal obesity are better discriminators of cardiovascular risk factors than BMI: A meta-analysis," *Journal of Clinical Epidemiology*, c. 61, s. 7, ss. 646-653, 2008.
- [9] B. Srdi, B. Obradovi, G. Dimitri, E. Stoki ve S. Babovi, "Relationship between body mass index and body fat in children—age and gender differences," *Obesity Research & Clinical Practice*, c. 6, s. 2, ss. 167-173, 2012.
- [10] A. Kupusinac, E. Stoki ve R. Doroslovaki, "Predicting body fat percentage based on gender, age and BMI by using artificial neural networks," *Computer Methods and Programs in Biomedicine*, c. 113, s. 2, ss. 610-619, 2014.

- [11] P. Deurenberg ve M. Yap, "The Assessment of Obesity: Methods for measuring body fat and global prevalence of obesity," *Best Practice & Research Clinical Endocrinology & Metabolism*, c. 13, s. 1, ss. 1-11, 1999.
- [12] N. Jensky-Squires, C. Dieli-Conwright, A. Rossuello, D. Erceg, S. McCauley ve E. Schroeder, "Validity and reliability of body composition analysers in children and adults," *British Journal of Nutrition*, c. 100, s. 4, ss. 859-865, 2008.
- [13] W. Beeson, M. Batech, E. Schultz, L. Salto, A. Firek, M. Deleon, H. Balcazar ve Z. Cordero-Macintyre, "Comparison of body composition by bioelectrical impedance analysis and dual-energy X-ray absorptiometry in hispanic diabetics," *International Journal of Body Composition Research*, c. 8, s. 2, ss. 45-50, 2010.
- [14] A. M. Bongioiolo, K. Castro ve M. A. da Silva. "Bioelectrical impedance analysis: body composition in children and adolescents with Down Syndrome," *Minerva Pediatrica*, c. 69, s. 6, ss. 560-563, 2017.
- [15] D. Anblagan, R. Deshpande, N. Jones, C. Costigan, G. Bugg, N. Raine-Fenning, P. Gowland ve P. Mansell, "Measurement of fetal fat in utero in normal and diabetic pregnancies using magnetic resonance imaging," *Ultrasound in Obstetrics & Gynecology*, c. 42, s. 3, ss. 335-340, 2013.
- [16] J. Josefson, M. Nodzinski, O. Talbot, D. Scholtens ve P. Catalano, "Fat mass estimation in neonates: anthropometric models compared with air displacement plethysmography," *British Journal of Nutrition*, c. 121, s. 3, ss. 285-290, 2019.
- [17] D. Fukuda, M. Wray, K. Kendall, A. Smith-Ryan ve J. Stout, "Validity of near-infrared interactance (FUTREX 6100/XL) for estimating body fat percentage in elite rowers," *Clinical Physiology and Functional Imaging*, c. 37, s. 4, ss. 456-458, 2017.
- [18] A. Fernández-Sánchez, E. Madrigal-Santillán, M. Bautista, J. Esquivel-Soto, Á. Morales-González, C. Esquivel-Chirino, I. Durante-Montiel, G. Sánchez-Rivera, C. Valadez-Vega ve J. A. Morales-González, "Inflammation, oxidative stress, and obesity," *International Journal of Molecular Sciences*, c. 12, s. 5, ss. 3117-3132, 2011.
- [19] T. Ferenci, "Two Applications Of Biostatistics in The Analysis of Pathophysiological Processes," Doktora Tezi, Óbuda Univeristy, Budapest, Hungary, 2013.
- [20] T. Ferenci ve L. Kovács, "Predicting body fat percentage from anthropometric and laboratory measurements using artificial neural networks," *Applied Soft Computing*, c. 67, ss. 834-839, 2018.
- [21] S. Balasundaram, "On lagrangian support vector regression," *Expert Systems with Applications*, c. 37, s. 12, ss. 8784-8792, 2010.
- [22] Y. Xu ve L. Wang, "A weighted twin support vector regression," *Knowledge-Based Systems*, c. 33, ss. 92-101, 2012.
- [23] R. Chiong, Z. Fan, Z. Hu ve F. Chiong, "Using an improved relative error support vector machine for body fat prediction," *Computer Methods and Programs in Biomedicine*, c. 198, ss. 105749, 2020.

- [24] P. Deurenberg, M. Yap ve W. van Staveren, "Body mass index and percent body fat: a meta analysis among different ethnic groups," *International Journal of Obesity*, c. 22, s. 12, ss. 1164-1171, 1998.
- [25] A. Jackson P. Stanforth, J. Gagnon, T. Rankinen, A. Leon, D. Rao, J. Skinner, C. Bouchard ve J. Wilmore, "The Effect of sex, age and race on estimating percentage body fat from body mass index: the heritage family study," *International Journal of Obesity*, c. 26, s. 6, ss. 789-796, 2002.
- [26] Y. Shao, "Body fat percentage prediction using intelligent hybrid approaches," *The Scientific World Journal*, c. 2014, ss. 1-8, 2014.
- [27] M. Uçar, Z. Uçar, F. Köksal ve N. Daldal, "Estimation of body fat percentage using hybrid machine learning algorithms," *Measurement*, c. 167, ss. 108173, 2020.
- [28] K. DeGregory, P. Kuiper, T. DeSilvio, J. D. Pleuss, R. Miller, J. W. Roginski, C. B. Fisher, D. Harness, S. Viswanath, S. B. Heymsfield, I. Dungan ve D. M. Thomas, "A review of machine learning in obesity," *Obesity Reviews*, c. 19, s. 5, ss. 668-685, 2018.
- [29] M. Akman, M. K. Uçar, Z. Uçar, K. Uçar, B. Baraklı ve M. R. Bozkurt, "Determination of body fat percentage by gender based with photoplethysmography signal using machine learning algorithm," *Innovation and Research in BioMedical Engineering*, Basımda.
- [30] C. Cortes ve V. Vapnik, "Support-vector networks," *Machine learning*, c. 20, s. 3, ss. 273-297. 1995.
- [31] T. K. Ho, "Random decision forests", In: *Proceedings of 3rd International Conference on Document Analysis and Recognition*. IEEE, Montreal, QC, Canada, 1995, ss. 278-282.
- [32] R. Johnson, "Fitting percentage of body fat to simple body measurements," *Journal of Statistics Education*, c. 4, s. 1, 1996.
- [33] W. E. Siri, "body composition from fluid spaces and density: analysis of methods," *University of Michigan Library*, ss. 1-33, 1956.
- [34] X. Yan ve S. Xiaogang, "Linear regression analysis: theory and computing," *World Scientific*, ss. 1-2, 2009.
- [35] H. B. Curry, "The method of steepest descent for non-linear minimization problems," *Quart. Appl. Math.*, s. 2, ss. 258-261, 1944.
- [36] S. Boyd ve L. Vandenberghe, "*Convex Optimization*", 7. baskı, Newyork, United States of America: Cambridge University Press, 2004, böl. 5, ss. 215-216
- [37] L. Breiman, J. Friedman, C. J., Stone ve R. A. Olshen, "*Classification and Regression Trees*," 1. baskı, London, England: CRC Press, 1984, böl. 11, ss. 246-259
- [38] L. Breiman, "Random forests," *Machine Learning*, c. 45, s. 1, ss. 5-32, 2001.
- [39] B. Schölkopf, "*Statistical Learning and Kernel Methods*", In: *Data Fusion and Perception*, G. Della Riccia, HJ. Lenz, R. Kruse, *International Centre for Mechanical Sciences Book Series*, 1. baskı, Vienna, Austria :Springer, 2001, böl. 431, ss. 3-24.