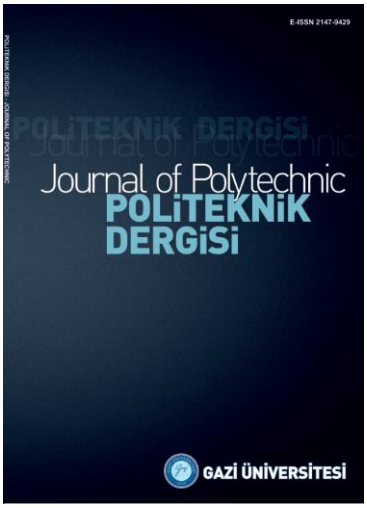


POLİTEKNİK DERGİSİ

JOURNAL of POLYTECHNIC

ISSN: 1302-0900 (PRINT), ISSN: 2147-9429 (ONLINE)

URL: http://dergipark.org.tr/politeknik

CLARIVATE ANALYTICS
EMERGING SOURCES CITATION INDEX
INDEXED

Epileptik nöbet tespiti için destek vektör regresyon temelli yeni bir sınıflandırma yaklaşımı

A new classification approach based on support vector regression for epileptic seizure detection

***Yazar(lar) (Author(s))**: Esra Betül KINACI[1], Hasan BAL[2], Harun KINACI[3]*

ORCID[1]: 0000-0002-4263-148X

ORCID[2]: 0000-0003-0570-8609

ORCID[3]: 0000-0002-8572-1143

ERKEN GÖRÜNÜM

Bu makaleye şu şekilde atıfta bulunabilirsiniz(To cite to this article): Kınacı E. B., Bal H. ve Kınacı H., "Epileptik nöbet tespiti için destek regresyon temelli yeni bir sınıflandırma yaklaşımı", ***Politeknik Dergisi***, *(*): *, (*).

Erişim linki (To link to this article): http://dergipark.org.tr/politeknik/archive

DOI: **10.2339/politeknik.1055549**

Epileptik Nöbet Tespiti İçin Destek Regresyon Temelli Yeni Bir Sınıflandırma Yaklaşımı

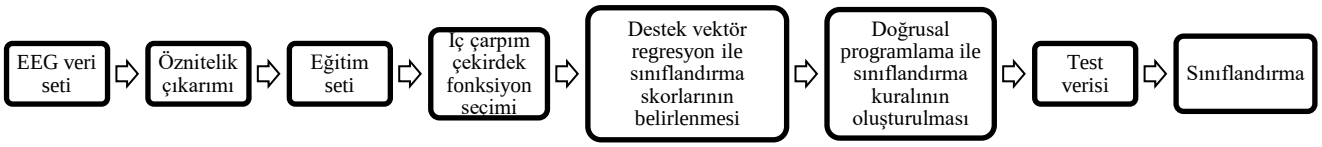
A New Classification Approach Based On Support Vector Regression For Epileptic Seizure Detection

Önemli noktalar (Highlights)

- *Doğrusal olmayan sınıflandırma problemleri için yeni bir yaklaşım / A new approach for nonlinear classification problems*
- *Destek vektör regresyon analizi ile sınıflandırma skorlarının elde edilişi / Obtaining classification scores by support vector regression analysis*
- *Matematiksel programlama ile sınıflandırma kuralı belirlenmesi / Determination of classification rule with mathematical programming*
- *Farklı öznitelikler ile epileptik nöbet tespiti / Detection of epileptic seizures with different features*

Grafik Özet (Graphical Abstract)

Öznitelikleri çıkarılan veri seti için uygun çekirdek fonksiyonu belirlenmektedir. İki aşamalı sınıflandırmanın ilk aşamasında sınıflandırma skorları elde edilir ve ikinci aşamada sınıflandırma kuralı oluşturulmaktadır. /The appropriate kernel function is determined for the data set whose features are extracted. In the first step of the two-stage classification, classification scores are obtained, and in the second step, the classification rule is created



Şekil. Önerilen Yöntem Akış Diyagramı /**Figure.** Flow chart of the proposed method

Amaç (Aim)

Bu çalışma doğrusal olmayan sınıflandırma problemleri için yeni bir sınıflandırma algoritması önermektedir. / This study proposes a new classification algorithm for nonlinear classification problems.

Tasarım ve Yöntem (Design & Methodology)

Önerilen yöntem destek vektör regresyon ve matematiksel programlama yaklaşımlarının birleşiminden oluşmaktadır. / The proposed method consists of a combination of support vector regression and mathematical programming approaches.

Özgünlük (Originality)

Çalışma literatüre matematiksel programlama yaklaşımlarının doğrusal olmayan problem türlerinde kullanılmasında katkı sunmaktadır. / The study contributes to the literature in the use of mathematical programming approaches in nonlinear problem types.

Bulgular (Findings)

Önerilen yöntem ile, EEG veri setinde yer alan grupların yüksek oranda doğru sınıflandırması gerçekleştirilmiştir. / With the proposed method, a highly accurate classification of the groups in the EEG data set was achieved.

Sonuç (Conclusion)

Önerilen yöntem çok gruplu ve doğrusal olmayan sınıflandırma problemleri için alternatif bir yaklaşımdır. / The proposed method is an alternative approach for multigroup and nonlinear classification problems

Etik Standartların Beyanı (*Declaration of Ethical Standards*)

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler. / The author(s) of this article declare that the materials and methods used in this study do not require ethical committee permission and/or legal-special permission.

Epileptik Nöbet Tespiti İçin Destek Regresyon Temelli Yeni Bir Sınıflandırma Yaklaşımı

Araştırma Makalesi / Research Article

Esra Betül KINACI[1*], Hasan BAL[1], Harun KINACI[2]

[1]Fen Fakültesi, İstatistik Bölümü, Gazi Üniversitesi, Türkiye
[2]İktisadi İdari Bilimler Fakültesi, İşletme Bölümü, Erciyes Üniversitesi, Türkiye

(Geliş/Received : 09.01.2022 ; Kabul/Accepted : 17.08.2022 ; Erken Görünüm/Early View : 27.11.2022)

ÖZ

Sınıflandırma problemi araştırmacılar tarafından uzun zamandır incelenen bir konu olmasına rağmen güncelliğini hala korumaktadır. Özellikle görüntü işleme ve hastalık tanısının belirlenmesi problemleri güncel uygulama alanlarından bazılardır. Bu çalışma destek vektör regresyon ve matematiksel programlamaya dayalı yeni bir veri sınıflandırma yöntemi sunmaktadır. Önerilen yöntem iki aşamalı hibrit bir yapıdan oluşmaktadır. İlk aşamada, destek vektör regresyon denklemi ile her bir birim için sınıflandırma skoru elde edilirken ikinci aşamada ise birimlerin sınıflandırma skorları kullanılarak bir matematiksel model yardımıyla sınıflandırma kuralı oluşturulur ve birimlerin sınıflandırılması sağlanır. Önerilen yöntem geleneksel yöntemlere alternatif bir yenilik sunmaktadır. Geleneksel matematiksel programlamaya dayalı yöntemler sınıfları doğrusal bir fonksiyon ile ayırır. Bu durum ise matematiksel programlamaya dayalı algoritmalarının kullanımını kısıtlar. Önerilen yöntem, doğrusal veya doğrusal ayrılamayan veri yapılarının tamamında kullanılabilir olmasının yanı sıra ikiden fazla grup sayısının olduğu problem türlerine de kolaylıkla dönüştürülebilmektedir. Model önce simülasyon ile irdelenmiş sonrasında Elektroensefalograf (EEG) sinyallerinin sınıflandırılması probleminde uygulanmış ve sınıflandırma performansı mevcut yöntemlerle karşılaştırılmıştır. Elde edilen sonuçlar tablolarda verilmiş ve önerilen modelin mevcut algoritmalara alternatif olabileceğini gösterilmiştir.

Anahtar Kelimeler: Destek vektör makineleri, destek vektöe regresyon, sınıflandırma, bonn veri seti.

A New Classification Approach Based On Support Vector Regression For Epileptic Seizure Detection

ABSTRACT

Although the classification problem is a subject that has been studied by researchers for a long time, it is still up-to-date. Especially the problems that image processing and diagnosis of disease are some of the most current application topics. This study presents a new data classification method based on support vector regression and mathematical programming. The proposed method consists of a two-stage hybrid structure. In the first step, the classification score is obtained for each unit with the support vector regression equation. In the second stage, using the classification scores of the units, a classification rule is created with the help of a mathematical model and the classification of the units is provided. The proposed method offers an alternative innovation to traditional methods. Methods based on traditional mathematical programming separate classes with a linear function. This situation limits the use of algorithms based on mathematical programming. The proposed method can be used in all linear or non-linearly separable data structures, as well as easily transforming into problem types with more than two groups. The model is first examined with simulation and then applied to the classification problem of Electroencephalograph (EEG) signals and the classification performance was compared with existing methods. The results obtained are given in the tables and it is shown that the proposed model can be an alternative to the existing algorithms

Keywords: Support vector machines, support vector regression, classification, bonn dataset.

ERKEN GÖRÜNÜM

1. GİRİŞ (INTRODUCTION)

Epilepsi, beyin fonksiyon bozukluğuna sebep olan nörolojik bir hastalıktır. Dünya Sağlık Örgütü'ne göre dünya çapında elli milyon epilepsi hastası vardır [1]. Epilepsi hastalığında nöronlarda ani ve kontrolsüz boşalmalar meydana gelir. Bunun sonucunda bilinç veya sezgi kaybı ve diğer bazı bilişsel bozukluklar meydana gelir [2]. Nöbetlerin belirsiz zamanlarda oluşması, epileptik hastalarda nöbet sırasında düşme, boğulma gibi ölüme veya ciddi yaralanmalara sebep olabilir. Nöbet zamanının tanınması ve müdahale epilepsi hastalarının yaşayacağı riskli durumlarını azaltabilir.

EEG, beyindeki elektriksel dalgalanmaların cihaz yardımıyla ölçülmesini sağlayan bir araçtır. EEG ile epileptik nöbetlerin özellikleri incelenip, nöbetler hakkında fizyolojik sinyallerin sağlamadığı bilgilere erişilebilir [3]. EEG, serebral kortekste nöronlar tarafından üretilen spontane, ritmik potansiyel değişiklikleri kaydeder [4,5]. EEG, epileptik bozuklukların tanısında, nöbetlerin kalitatif

*Sorumlu yazar (Corresponding Author)
e-posta : esrakinaci@gazi.edu.tr

lokalizasyonunda, durum tipinin belirlenmesinde ve tedavinin etkinliğinin izlenmesinde yeri doldurulamaz bir rol oynar [6].

EEG veri toplama süreci uzun ve yüksek maliyetlidir. Aynı bireyden büyük miktarda veri elde etmek zordur, bu nedenle genellikle küçük örnek veriler elde edilir. EEG sinyalinin durağan olmayan yapısı ve denekler arasındaki bireysel değişkenlik, diğer deneklerden alınan verilerin doğrudan kullanılmasını imkansız hale getirir [7]. Epileptik hastalarla çalışmanın zorluğu ve nöbet uyarısı olmaması nedeniyle örneklem hacmi genellikle küçüktür. Mevcut epilepsi teşhisi, EEG sinyalini analiz eden ve hastanın anormal deşarjlar ve nöbetler yaşayıp yaşamadığını belirlemek için bunu hastanın konvülsiyonları ile birleştiren doktorlar tarafından konur. Ancak bu yöntem sadece zaman alıcı olmayıp aynı zamanda hekimin aşırı subjektifliği nedeniyle yanlış tanıya da yatkındır [8]. Bu nedenle, ani başlangıçlı nöbetler için gerçek zamanlı, doğru bir tanı tanımlama yöntemi tasarlamak gereklidir.

Epileptik nöbetlerin, EEG kaydında belirlenmesi bir sınıflandırma problemi olarak ele alınmaktadır. EEG kaydının alınması ve işlenebilir forma getirilmesi ilk aşama olarak değerlendirilebilir. Ardından öznitelik çıkarımı ve nihai olarak bu özniteliklerin sınıflandırılması ile aşamalar tamamlanır. Epileptik nöbet tespiti, bir girdi sinyalinin, epileptik nöbet sinyali olup olmadığının sınıflandırılması olarak düşünülebilir [9,10].

EEG nöronlar tarafından üretilen beynin elektriksel aktivesinin ayrıntılarını içeren bir sinyaldir [11]. Bu sinyaller zaman serisi verileridir ve zaman alanından öznitelikler çıkarılarak EEG modeli analiz edilir [12]. Minimum, maksimum, ortalama gibi genel istatistiksel özellikler daha az hesaplama süresi gerektiren ve uygulaması kolay özniteliklerdir [13]. Ayrıca entropi gibi frekans alanı özelliği de öznitelik olarak kullanılır [14].

Öznitelik çıkarımı, EEG işaretlerinin karakteristikleri çıkarılarak elde edilir. Elde edilen öznitelikler farklı sınıflar arasında ayrışmayı en iyi sağlayacak şekilde seçilmeye çalışılır. Sınıflandırma başarısını etkileyecek önemli unsurlardan biride öznitelik seçimidir [15–19]. Bu bağlamda öznitelik çıkarımı üzerine kurulu birçok çalışma bulunmaktadır. [20]' da eşit genişlikte ayrıklaştırma (EWD) yöntemine dayalı yeni bir öznitelik çıkarma yaklaşımı önerilmiş ve bu yaklaşımla elde edilen istatistiksel öznitelikleri, çok katmanlı algılayıcı sinir ağı (MLPNN) modelinin girdileri olarak epileptik nöbetlerin tespitinde kullanılmıştır. EWD yöntemine dayalı farklı bir çalışma [21]'de verilmiştir. Burada EEG sinyalleri EWD yöntemi kullanılarak ayrıklaştırıldıktan sonra her bir aralıktaki veri noktalarının sayısına göre sinyallerin olasılık yoğunlukları hesaplanmıştır. Epileptik nöbet geçiren ve geçirmeyenler için iki farklı olasılık yoğunluk fonksiyonu tanımlanmış ve daha sonra bu iki fonksiyon için ortalama kare hata kriteri kullanılarak nöbet saptama başarısı elde edilmiştir. Diğer bir çalışma ise [22] olarak verilebilir. Bu çalışmada EEG sinyallerinin sınıflandırılmasında kullanılmak üzere ayrıklaştırma tabanlı entropi adı verilen yeni bir öznitelik çıkarımı önermiştir. Çalışmada EEG sinyalleri ayrık dalgacık dönüşümü (DWT) kullanılarak istenilen sayıda aralıklara ayrıştırılır, ayrıştırılmış alt bantların Shonnan entropi değerleri hesaplanır ve bunlar daha sonra Uyarlanabilir Nöro-Bulanık Sınıflandırıcısının (ANFIS) girdileri olarak kullanılır. Bu çalışmalara ek olarak, daha önce literatürde kullanılan bir boyutlu yerel ikili örüntü (1B-YİÖ) yönteminden yola çıkarak, sinyalin yerel değişimlerine daha duyarlı, uç noktaların etkisini de değerlendiren bir boyutlu medyan yerel ikili örüntü temelli öznitelik çıkarımı yöntemi [9]' da önerilmiş ve bu çalışmada *k*-En Yakın Komşu (*k*-EYK) sınıflandırıcısı ile sınıflandırma başarısı değerlendirilmiştir. [2]'de motif algoritması isimli yeni bir yaklaşım öne sürülmüş ve motif öznitelikleri elde edildikten sonra sınıflama aşamasında *k*-EYK , Yapay Sinir Ağları, Destek Vektör Makinaları (DVM) gibi farklı sınıflandırma algoritmaları kullanılmıştır. EEG sinyalleri ile epileptik nöbet tespiti çok zengin bir literatüre sahiptir. Yukarıda verilen çalışmaların yanı sıra, permutasyon entropi [23], kesikli harmonik arama [24], dalgacık paket entropisi [25], hızlı fourier dönüşümü [26], korelasyon tabanlı öznitelik [27] seçimi gibi farklı birçok çalışma mevcuttur.

Matematiksel programlama tabanlı yöntemlerin sınıflandırma problemlerinde kullanımı çoğunlukla iki gruplu durumlar için geliştirilmiş olsa da çok gruplu problemler içinde geliştirilen yeni yöntemler bulunmaktadır. Fakat bu çalışmaların tamamı doğrusal durumdaki problemler için uygundur. Varsayım gerektirmeyen ve bu sayede geniş kullanım alanına sahip matematiksel programlama yaklaşımlarının sadece belirli bir veri türü için kullanılabilir olduğu görülmektedir. Diğer yandan bir istatistiksel modelleme olan regresyon analizi, birimlerin sınıflandırma skorlarının elde edilmesi için kullanılan bir araçtır. İlk olarak [28]'de her birimin sınıflandırma skorunu doğrusal regresyon problemi olarak ele almış ve en küçük kareler yöntemi ile skorları elde etmiştir. Fakat bu çalışmanın devamında birimlerin nasıl gruplara atanacağı konusunda net bir kural oluşturulmamıştır. [29]'de ise [30] 'da önerilen matematiksel model ile doğrusal regresyon yaklaşımını birlikte kullanarak hibrit bir model ortaya konulmuştur. Yalnız bu hibrit modelde sadece doğrusal yapıdaki durumlar için kullanılmakta ve doğrusal olmayan durumlardaki literatür eksikliğini gidermemektedir.

Bu çalışmada farklı bir regresyon analizi yaklaşımı olarak Destek Vektör Regresyon (DVR) kullanılmıştır. Önerilen yöntem regresyon analizi için aykırı değere karşı en küçük kareler yönteminden daha avantajlı olmasının yanı sıra iç çarpım çekirdek fonksiyonlarının kullanımı ile doğrusal olmayan durumlarda kullanım imkanı sağlamaktadır. Bu yönüyle önerilen yöntem [29] ve [30] 'daki yaklaşımlardan farklı bir regresyon bakış açısı katarak doğrusal olmayan veri yapılarında kullanılabilmesini bakımından katkı sunmaktadır.

ERKEN GÖRÜNÜM

Bölüm 2' de önerilen yöntemin dayandırıldığı iki sınıflandırma modeline ait temel bilgiler verilmiş, Bölüm 3'de ise önerilen model tanıtılmıştır. Bölüm 4 'te modelin irdelenmesi için simülasyon çalışmasına yer verilmiştir. Bölüm 5'te kullanılan veri seti ve veri setinin uygulamaya hazırlanması için kullanılan öznitelik çıkarım aşamalarına yer verilmiştir. Ayrıca uygulamada karşılaştırılması yapılan sınıflandırma yöntemlerinden bahsedilmiş, yöntemlerin kıyaslanması için kullanılan performans kriterleri tanıtılmıştır. Bölüm 6 ve Bölüm 7'de ise sırasıyla yapılan analizlere ilişkin araştırma bulguları ile sonuç ve tartışma kısmı yer almaktadır

2. MATERYAL VE METOD (MATERIAL and METHOD)

Bu çalışmada sınıflandırma problemlerinde kullanılan iki farklı yöntemin güçlü yanları birleştirilerek yeni bir sınıflandırma algoritması ileri sürülmüştür. Bunlardan birinci Destek Vektör Regresyon (DVR), diğeri ise Matematiksel Programlamadır (MP).

2.1. DESTEK VEKTÖR MAKİNELERİ (SUPPORT VECTOR MACHİNE)

Bu bölümde DVM'nın teorik alt yapısından ve takiben regresyon için standart DVM formülasyonu sunulmaktadır. DVM, istatistiksel öğrenme teorisinde güçlü bir teorik yapıya sahip bir sınıflandırma tekniğidir[31,32]. [33]'da eğitim modelleri ve karar sınırı arasındaki marjı en üst düzeye çıkaran bir algoritma ileri sürülmüştür. Bu tekniğin teorik alt yapısı Vapnic-Chervonenkis (VC) teorisine dayanmaktadır ve VC-boyut olarak bilinen bir karmaşıklık ölçütünün sınıflandırma algoritmasının genelleştirme hatasının tahmini için analitik bir sınır oluşturulur [34,35]. Negatif olmayan bir tam sayı olan VC boyutu, öğrenme makinesi tarafından gerçekleştirilen sınıflandırma fonksiyonları ailesinin ifade gücünü ölçer [36]. $\{(x_i, y_i), i = 1, \ldots, N\}$ sonlu bir eğitim örneği olsun. Burada x_i, n boyutlu uzayda $(i.e., x_i = [x_{1i}, \ldots, x_{1n}] \in \mathbb{R}^n)$ i-inci örneği, y_i ise aynı örneğe karşılık gelen sınıf etiketini $(y_i \in \{-1, +1\})$ temsil eder. Makine öğrenmesi bir dizi olası $x \rightarrow f(x; w)$ eşleşmesine izin verir. Burada $f(\cdot)$, belirli bir $\boldsymbol{x}$ ve ayarlanabilir $\boldsymbol{w}$ $(w \in \mathbb{R}^{\boldsymbol{n}})$ paremetresi için her zaman aynı çıktıyı veren deterministik bir fonksiyondur. Makine öğrenmesinde eğitim aşaması $\boldsymbol{w}$ parametrelerinin ayarlanmasını içerir. Makine öğrenmesinin eğitim hatası $v(\boldsymbol{w})$ ile gösterilsin. Yapısal risk minimizasyonu temel ilkesine göre test kümesindeki sınıflandırma hatalarının olasılığının bir sınırı vardır, $P(\boldsymbol{w})$, ve şu şekilde ifade edilir.

$$P(\boldsymbol{w}) \leq v(\boldsymbol{w}) + \varepsilon_1(N, h, \eta, v) \quad (1)$$

Burada;

$$v(\boldsymbol{w}) = \frac{1}{2N}\sum_{i=1}^{N}|y_i - f(x_i, \boldsymbol{w})| \quad (2)$$

$$\varepsilon_1(N, h, \eta, v) = 2\varepsilon_0^2(N, h, \eta)\left(1 + \sqrt{1 + \frac{v(\boldsymbol{w})}{\varepsilon_0^2(N,h,\eta)}}\right) \quad (3)$$

olmak üzere, Eşitlik (3) ifadesi için de;

$$\varepsilon_0(N, h, \eta) = \sqrt{\frac{h}{N}\left[ln\left(\frac{2N}{h}\right) + 1\right] - \frac{1}{N}ln\left(\frac{\eta}{4}\right)} \quad (4)$$

biçimindedir. Eşitsizlik (1), $(1 - \eta)$ olasılıkla geçerlidir. $\varepsilon_0(N, h, \eta)$ ifadesi makine öğrenmesinin kapasitesini temsil eder ve VC boyutunun güven aralığı olarak adlandırılır. Bu güven aralığının değeri, eğitim örnekleri N 'in sayısına, makine öğrenmesinin VC boyutuna ve η değerine bağlıdır. Sıfıra yakın eğitim hatası $v(\boldsymbol{w})$ Eşitsizlik (1) 'den daha basit olan Eşitsizlik (5) ifadesine, tersi olarak büyük eğitim hatası $v(\boldsymbol{w})$ ise Eşitsizlik (6) ifadesine dönüşür[37].

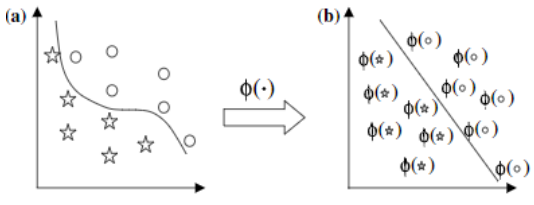
$$\mathrm{P(w)} \lesssim v(\boldsymbol{w}) + 4\varepsilon_0^2(N, h, \eta) \quad (5)$$

$$P(w) \lesssim v(\boldsymbol{w}) + \varepsilon_0(N, h, \eta) \quad (6)$$

Yukarıdaki ifadeler incelendiğinde, sadece eğitim hatasını, $v(\boldsymbol{w})$' nin minimize edilmesi iyi bir genelleme performansı elde etmek için yeterli olmadığı aşikardır. Çünkü $v(\boldsymbol{w})$ 'deki düşüş tek başına test hatasında düşüşü garanti etmez. DVM, yapısal risk minimizasyonu ilkesini kullanarak $v(\boldsymbol{w})$ ile VC güven aralığı $\varepsilon_0(N, h, \eta)$, arasında bir denge gözeterek genelleme hatası üzerindeki üst sınırı en aza indirmeye çalışır [37].

2.2. DESTEK VEKTÖR REGRESYON (SUPPORT VECTOR REGRESSİON)

Öncelikle DVM' nın temel prensipleri doğrultusunda fonksiyon yaklaşımı ele alınmaktadır. (x_i, y_i), sonlu eğitim modelini temsil etsin. Burada $x_i \in \mathbb{R}$ eğitim modellerinden oluşan (yani, $x = [x_{1,\ldots,}x_N]$) girdi vektörünün bir değeri, $y_i \in \mathbb{R}$ ise bu girdiye karşılık gelen çıktı değeridir. Giriş uzayını daha yüksek boyutlu bir özellik uzayı olan $\mathbb{R}^{n_h}$ ile eşleştirmek için doğrusal olamayan bir dönüşüm fonksiyonu $\phi(.)$ tanımlanmıştır. Bu dönüşümün matematiksel gerekçesi Cover teoremine dayanır. Bu teoreme göre karmaşık bir yapıya sahip bir sınıflandırma problemi yüksek boyutlu bir uzayda doğrusal olmayacak bir şekilde oluşmuş ise düşük boyutlu uzaydakine kıyasla daha yüksek bir olasılıkla doğrusal olarak ayrılabilir [35]. Şekil 1.'de çekirdek hilesi olarak adlandırılan yüksek boyutlu uzaya haritalama işlemi görselleştirilmiştir. Dönüşümle elde edilen girdi ve çıktı arasındaki ilişki aşağıdaki şekilde ifade edilir.



Şekil 1. İç Çarpım Çekirdek Fonksiyonu (Dot Product Kernel Function)[38]

Şekil 1'deki $\phi(\cdot)$, (a) orijinal düşük boyutlu girdi uzayındaki doğrusal olmayan bir problemi daha yüksek boyutlu bir özellik uzayındaki (b) doğrusal problemine

ERKEN GÖRÜNÜM

dönüştürmek için tanımlanan bir doğrusal olmayan dönüşüm fonksiyondur.

$$\hat{y} = f(x) = \boldsymbol{w^T}\phi(x) + b \qquad (7)$$

Burada $\hat{y}$ model çıktısını, $\boldsymbol{w}$ ve $\boldsymbol{b}$ ayarlanabilir model parametrelerini ifade etmektedir. DVM yapısal riski minimize etmeyi amaçlamaktadır ve bu risk aşağıdaki şekilde tanımlanır.

$$R_{emp} = \frac{1}{N}\sum_{1=1}^{N}|y_i - \hat{y}_i|_\varepsilon \qquad (8)$$

DVM yaklaşımı uygun kayıp fonksiyonu belirlenmesiyle regresyon (DVR) problemlerinde kullanılmaktadır[39]. Eşitlik (8)' deki $|y_i - \hat{y}_i|_\varepsilon$ ifadesi ε-duyarsız kayıp fonksiyonu olarak adlandırılır [30,33] ve Eşitlik (9)'daki gibi ifade edilmektedir.

$$|y_i - \hat{y}_i|_\varepsilon = \begin{cases} 0 & eğer\ |y_i - \hat{y}_i|_\varepsilon \le \varepsilon \\ |y_i - \hat{y}_i|_\varepsilon & d.d \end{cases} \qquad (9)$$

Bu aşamada, düzenlileştirme teorisine [37], göre $\boldsymbol{w}$ ve $\boldsymbol{b}$ parametreleri amaç fonksiyonu $\psi_\varepsilon(w, \xi, \xi^*)$ minimize edecek şekilde tahmin edilir.

$$\psi_\varepsilon(w, \xi, \xi^*) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{N=1}^{N}(\xi_i + \xi_i^*)$$
$$y_i + \hat{y}_i \le \varepsilon + \xi_i \qquad i = 1,2,\dots,N$$
$$-y_i + \hat{y}_i \le \varepsilon + \xi_i^* \qquad i = 1,2,\dots,N \qquad (10)$$
$$\xi_i \ge 0 \qquad i = 1,2,\dots,N$$
$$\xi_i^* \qquad i = 1,2,\dots,N$$

Burada ξ_i ve ξ_i^* pozitif tanımlı artık değişkenleridir. Artık değişkenler, her zaman (x_i, y_i) çiftlerine ε kesinliğinde yakınsayan bir f fonksiyonu bulunamayacağından bir takım hataya izin vererek modeli esnetmek amacıyla kullanılmıştır. Amaç fonksiyonun ilk terimi ağırlık boyutunu düzenlemek ve yüksek ağırlık değerlerine engel olmak için kullanılır. İkinci terimde yer alan pozitif tanımlı $\boldsymbol{C}$ sabiti artıklar üzerinde bir kontrol mekanizması olarak çalışmaktadır ve sapmaların gözardı edilebileceği miktarı belirler. Burada ε'ın üzerindeki sapmalar ξ_i ile altında kalan sapmalar ise ξ_i^* ile temsil edilmektedir.

Eşitlik (10) bir karesel optimizasyon problemidir ve Lagrange çarpanları kullanılarak çözüme ulaşılır. Burada temel fikir amaç fonksiyonundan Lagrange fonksiyonu oluşturmaktır. Sonuçta $\boldsymbol{w}$ parametresi Eşitlik (11)'deki gibi elde edilir [41].

$$\boldsymbol{w} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\phi(x_i) \qquad (11)$$

Burada α_i ve α_i^* pozitif tanımlı gerçek sabitlerdir ve Lagrange çarpanlarıdır. Sıfırdan farklı α_i ve α_i^* 'lara karşılık gelen veri noktalarına destek vektörleri denir. Son olarak DVR 'yi temsilen Eşitlik (7) doğrusal olmayan fonksiyon tahmini için Eşitlik (12)'deki şeklini almaktadır.

$$\hat{y} = f(x) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)K(x_i, x) + b \qquad (12)$$

Burada $K(x_i, x)$, Mercer teoremine göre tanımlanan iç çarpım çekirdek fonksiyonu, b ise fonksiyonun yanıdır [42,43] ve Eşitlik (13)'deki gibi ifade edilmektedir.

$$K(x_i, x) = \phi(x_i)^T\,\phi(x_j) \qquad (13)$$

Doğrusal, polinom, sigmoid, eğri ve radyal temel işlev (gaussian) olmak üzere çekirdek işlevinin seçimi için çeşitli fonksiyonlar bulunmaktadır. Çalışmada Radyal Temel İşlev (RBF) iç çarpım çekirdeği kullanılmıştır. RBF çekirdeğinin avantajı eğitim verilerini, muhtemel sonsuz boyutlu bir uzaya doğrusal olmayan bir şekilde eşlemesidir. Bu sayede tahmin ediciler ile tahmin arasındaki ilişki doğrusal olmadığında daha efektif bir sonuç elde edilir [38].

2.3. SINIFLANDIRMA PROBLEMLERİNDE MATEMATİKSEL PROGRAMLAMA YAKLAŞIMLARI (MATHEMATİCAL PRGRAMMİNG APPROACHES TO CLASSIFICATION PROBLEMS)

Matematiksel Programlamaya (MP) dayalı modeller sınıflama problemleri için istatistiksel yöntemlere alternatif olacak şekilde ortaya atılmıştır. Herhangi bir istatistiksel varsayım gerektirmeyen bu yöntemler uygulama kolaylığı sağlamaktadır. Bu konuda ilk dikkat çekici ve öncü niteliğindeki çalışma [44,45]'de doğrusal programlamaya (DP) dayalı önerilmiş olan modeldir. Önerilen bu model (14) ile verilmiştir. Burada, sapmalar toplamını minimize etmeyi amaçlanmaktadır. İkili sınıflandırma problemi için p, sınıflandırmada kullanılan değişken sayısı olmak üzere, X, G_1 ve G_2gruplarından alınan n çaplı bir örneğin değişken değerlerinin matrisi olsun. Değişken katsayıları $\alpha_1, \alpha_2 \cdots \alpha_p$ ile ve j birime ait sınıflandırma skoru

$S_j = \sum_{i=1}^{p}\alpha_i X_{ij} (j = 1,2,\cdots n)$ olmak üzere,

$$\min \sum_{j\in G_1} S_j^+ + \sum_{j\in G_2} S_j^-$$
$$\cdots$$
$$\sum_{i=1}^{p}\alpha_i x_{ij} + S_j^+ \ge c\,, j \in G_1 \qquad (14)$$
$$\sum_{i=1}^{p}\alpha_i x_{ij} - S_j^- < c\,, j \in G_2$$

DP modellerinde, minimum sapmanın maksimizasyonu ve sapmaların toplamının minimizasyonu en sık kullanılan iki optimizasyon kriteridir. [44,45]'deki önerilerin ardından pek çok bu yaklaşıma dayanan sınıflandırma yöntemi geliştirilmiştir. [46]'de kümeleme analizine benzer bir anlayışla aynı grup içindeki birimlerin daha benzer olması gerektiği düşüncesine dayalı olarak DP tabanlı bir sınıflandırma yöntemi geliştirilmiştir. [47]'de geliştirilen iki gruplu sınıflandırma modeli çok gruplu duruma genişletilmiştir. Burada her bir sınıflandırma skorunun kendi sınıflandırma skorundan sapmasını minimize edecek bir doğrusal model oluşturulmuştur. Modelin ilk aşaması (15) ile verilmiştir

Bu modelde $\bar{x}_i^r = \frac{1}{n_r}\sum_{j\in G_r} X_{ij}$ ve $r = 1,\cdots,k$ olarak tanımlanır. Burada n_r ve $G_r$grubundaki birim sayısıdır. n ise $n = n_1 + \cdots + n_k$ olmak üzere toplam birim sayısıdır. $u = 1,\cdots,k-1$ ve $v = u+1,\cdots,h$ olmak üzere, her bir (u, v) çifti için model geliştirilerek çok gruplu duruma genişletilmiş olur. Burada, $\alpha_i, (i = 1,\cdots,p)$ serbest değişkenler ve $j \in G_v$ için $d_j \ge 0$ dır.

ERKEN GÖRÜNÜM

$\min \sum_{j\epsilon G_u, j\epsilon G_v}^{n} d_j$

...

$$\sum_{i=1}^{p} \alpha_i \left(X_{ij} - \bar{x}_i^u\right) + d_j \geq 0, j\epsilon G_u$$

$$\sum_{i=1}^{p} \alpha_i \left(X_{ij} - \bar{x}_i^v\right) - d_j \leq 0, j\epsilon G_v$$

$$\sum_{i=1}^{p} \alpha_i (\bar{x}_i^u - \bar{x}_i^v) \geq 1$$

Modeldeki kısıtlar *r.* gruptaki birimlerin sınıflandırma skorlarını *r.* grubun ortalama sınıflandırma skoruna mümkün olduğunca yaklaştırma yarar $(r = u, v)$. Bu model yardımıyla her (u, v) grup çifti için elde edilen $\alpha_i$sayesinde G_u ve G_v'deki birimlerin S_j sınıflandırma skorları elde edilir. Ardından da $, u = 1, \cdots, k-1$ ve $v = u+1, \cdots, k$ olmak üzere grupları ayırmada kullanılacak olan c_{uv} değerlerini elde etmek için ikinci bir model oluşturulmuştur ve Model (16)'da verilmiştir.

$$\sum_{j\in G_u} \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} d_{juv} + \sum_{j\in G_v} \sum_{u=1}^{k-1} \sum_{v=u+1}^{k} d_{juv} \tag{16}$$

$$S_{juv} + d_{juv} \geq c_{uv},$$
$$u = 1, \dots, k-1, v = u+1, \dots, k, j \in G_u$$
$$S_{juv} - d_{juv} \leq c_{uv},$$
$$u = 1, \dots, h-1, v = u+1, \dots, h, j \in G_v$$

Model (15)'in çözülmesiyle her birim için sınıflandırma skorları elde edilmiş olur. Model (16)'nın çözülmesiyle ise birimleri sınıflara atama kuralı olarak kullanılacak kesme değerleri elde edilir. c_{uv} serbest değerli değişkenlerdir ve eş zamanlı olarak tahmin edilirler.d_{juv} ise negatif olmayan değişkenlerdir. Literatürde farklı DP modelleri ve hedef programlamaya dayalı yöntemler bulunmaktadır [48,49]. Ayrıca doğrusal programlama modelinin genişletilmesiyle karma tam sayılı modellerde kullanılmaya başlanmıştır. Bu yöntemlerde örneğin yanlış veya doğru sınıflandırıldığını göstermek için iki değerli tam sayı değişkenleri kullanılır [50–52]. Daha yakın tarihlerde DEA-DA olarak adlandırılan iki gruplu ve çok gruplu parametrik olmayan diskriminant analizi yaklaşımları da önerilmiştir [49,53,54]. Bu modeller verinin özel bir yapıda olmasını gerektirmektedir. Özel bir yapıdan kastedilen gruplardaki birimlerin sıralı olmasıdır. Bir başka model ise [55]' de önerilen SGP modelidir. Bu çalışmada, [54]'deki DEA-DA modeli ile [56]'de önerilen çok gruplu sınıflandırma modelinin güçlü yanlarını birleştirip yeni birçok gruplu sınıflandırma modeli önerilmiştir. Önerilen model verinin özel bir şekilde sıralı olmasını gerektirmezken aynı zamanda negatif değerli değişkenlerin kullanımına da olanak sağlar. Farklı bir bakışı açısı taşıyan diğer bir matematiksel model ise hiper-kutu sınıflandırıcılarıdır [57–59]. [58]'de ileri sürülen hiper-kutu yaklaşımında modeldeki kesikli değişkenler arasındaki ilişkiler Boolean Cebir kullanılarak eşdeğer tamsayı kısıtlamalarına dönüştürülmüştür. Son olarak farklı bir yaklaşım ise [28]'de verilmiştir. Bu çalışma, literatürde güçlü yönleriyle dikkat çeken iki çoklu sınıflandırma modelini birleştirmiş ve yeni hibrit bir yaklaşım ortaya koymuştur. Bu model, iki aşamadan oluşmaktadır. Modelin ilk aşaması [27]'deki regresyon analizi tabanlı parçacık sürü optimizasyonu teknikleri kullanılarak çözümlenen modelinden oluşur. Modelin ikinci aşaması ise sınıflandırma skorlarına dayalı eşik değer sınaması ile sınıflandırma yapılmasını sağlayan [29]'da önerilen modeldir.

3. SINIFLANDIRMA PROBLEMLERİ İÇİN DESTEK VEKTÖR REGRESYON TABANLI İKİ AŞAMALI YENİ BİR YAKLAŞIM (A NEW TWO-STAGE APPROACH BASED ON SUPPORT VECTOR REGRESSION FOR CLASSIFICATION PROBLEMS)

Sınıflandırma problemlerinde kullanılan MP yaklaşımlarında her birim için bir sınıflandırma skoru elde edilmektedir. Regresyon analizi sınıflandırma problemlerinde birimlerin sınıflandırma skorlarını hesaplamak için kullanılmaktadır. Regresyon modelindeki Y bağımlı değişkeni grup numaralarını belirten kategorik bir değişkendir [28]

Önerilen yaklaşımın ilk aşamasında her bir birim için sınıflandırma skoru ε-duyarsız kayıp fonksiyonu kullanılarak oluşturulan DVR ile tahmin edilir. Bu aşamada doğrusal ve doğrusal olmayan modeller için regresyon tahmini yapılabilmektedir. Doğrusal olmayan modeller için literatürdeki çekirdek fonksiyonları kullanılarak yüksek boyutlu uzayda doğrusal bir yapı oluşturulur.

S_j, G_u, ve G_v'deki birimlerin sınıflandırma skor değerleri, $\hat{w}$ regresyon parametre tahminleri olmak üzere $S_j = \boldsymbol{w^T}\phi(x) + b$ eşitliği ε-duyarsız kayıp fonksiyonlu destek vektör regresyon modeli ile çözülür.

$$\psi_\varepsilon(w, \xi, \xi^*) = \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{N=1}^{N}(\xi_i + \xi_i^*)$$

$$\begin{aligned} y_i + \hat{y}_i &\leq \varepsilon + \xi_i & i &= 1,2,\dots,N \\ -y_i + \hat{y}_i &\leq \varepsilon + \xi_i^* & i &= 1,2,\dots,N \\ \xi_i &\geq 0 & i &= 1,2,\dots,N \\ \xi_i^* & & i &= 1,2,\dots,N \end{aligned} \tag{17}$$

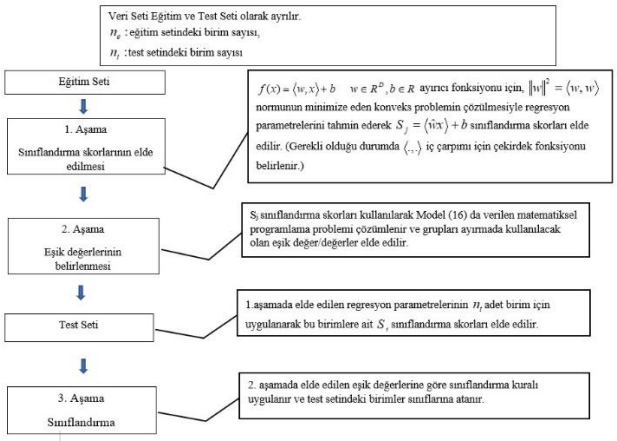
İkinci aşamada ise [47] tarafından kullanılan sınıflandırma skorlarına dayalı eşik değer sınaması ile sınıflandırma yapılmaktadır. DVR ile her birim için sınıflandırma skoru elde edildikten sonra $u = 1, \dots, k-1$ ve $v = u+1, \dots, k$ olmak üzere grupları ayırmada kullanılacak olan c_{uv} değerleri Model 16'in çözülmesiyle elde edilir.

Önerilen modelde sınıflandırma skorları verilerin doğrusal ayrılmadığı durumlarda uygun çekirdek fonksiyonu belirlenerek elde edilir. Bu sayede önerilen model hem doğrusal hem de doğrusal olmayan problem türlerinde kolaylıkla kullanılabilmektedir. Önerilen

ERKEN GÖRÜNÜM

model ile DVM sınıflandırma algoritmalarının temel farkı sınıfları ayırmada kullanılan düzlemdir. DVM sınıflandırıcı algoritmaları çekirdek fonksiyonu ile yüksek boyutlu uzaya haritalandırma yaptıklarında belirledikleri karar sınırları doğrusal değildir ve ayırma düzlemi de yüksek boyutlu uzayda yapılmaktadır. Önerilen yöntemde ise çekirdek fonksiyonları yalnızca sınıflandırma skorlarının elde edilmesinde kullanılmış ve karar kuralı olarak kullanılan eşik değer yine doğrusal bir matematiksel programlama problemi ile edilmiştir.

Önerilen yöntem için sınıflandırma sürecini içeren akış şeması Şekil 2. de verilmiştir.



Şekil 2. Önerilen Yöntemin Sınıflandırma Akış Şeması (Classification Flow Chart of the Proposed Method)

4. SİMÜLASYON ÇALIŞMASI (SIMULATION STUDY)

Çalışmanın bu bölümünde önerilen modelin diğer sınıflandırıcılarla doğru sınıflandırma performansı karşılaştırılmıştır. Karşılaştırma için kullanılan yöntemlere ilişkin temel düzeyde açıklayıcı bilgiler Çizelge 1.'de verilmiştir. Ek olarak önerilen yöntem destek vektörlerini kullandığı için karşılaştırılacak yöntemlere DVM de eklenmiştir

Çizelge 1. Yöntemlere ilişkin temel bilgiler (Basic information on methods)

Sınıflandırıcı	Yaklaşım
DVM	*Eğitim verilerindeki herhangi bir noktadan en uzak olan iki sınıf arasında bir karar sınırı belirler* [60]
Fisher Doğrusal Diskriminant Fonksiyonu (FDDF)	*Değişkenlerin doğrusal birleşimini bularak veriyi sınıflara ayırır* [61]
Naive Bayes (NB)	*Gözlemlerin sınıflarda bulunma olasılıklarını hesaplar ve en yüksek olasılığı göre sınıflandırır* [62].
k-En Yakın Komşu (*k*-EYK)	*Nesneler arasındaki uzaklığı dikkate alır. Birbirine yakın olan nesnelerin aynı sınıfa atanır* [63].

Simülasyon çalışmasında modellere ait doğruluk performansları farklı veri yapılarına göre test edilmiştir. Burada farklı veri yapısından kasıt grupların doğrusal bir ayırıcı fonksiyon ile ayrılabildiği durumlar ve gruplar arası ayrımın doğrusal ayırıcı fonksiyon ile yapılamadığı, başka bir ifade ile çekirdek fonksiyonlarının kullanımına ihtiyaç duyulduğu durumlardır. Çalışmada, üç, dört ve beş grup olması durumları dikkate alınmıştır. Ayrıca gruplardaki birim sayıları eşit olacak şekilde sırasıyla 10, 25, 50 ve 100 gözlemden oluşacak şeklide belirlenmiştir

4.1. DOĞRUSAL FONKSİYON İLE AYRILAN GRUPLAR (GROUPS SEPARATED BY LINEAR FUNCTION)

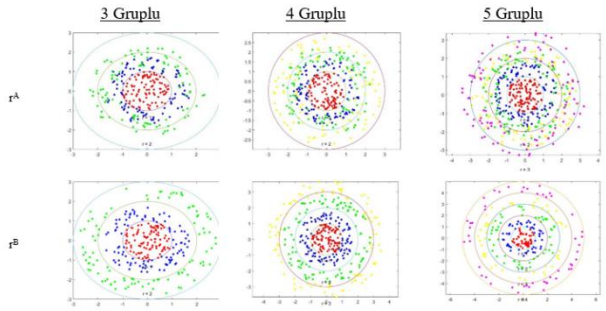
Bu bölümde doğrusal bir ayırıcı fonksiyon ile ayrılabilecek gruplar incelenmiştir. Burada gruplarda yer alan birimler düzgün ve normal dağılımdan veri üretilerek oluşturulmuştur. Veri üretiminde ortalama ve varyanslara bağlı olarak gruplar arası değişkenliğin oluşması sağlanmıştır. Simülasyon çalışmasında bir birimi hariç tutma (Leave one out -LOO-) ve 5-kat çapraz doğrulama tekniği kullanılmıştır. 5-kat çapraz doğrulama tekniğinde tüm gözlemlerin farklı kombinasyonlarda kullanımını sağlamak için çapraz doğrulama 100 tekrar ile gerçekleştirilmiştir. LOO çapraz doğrulama yöntemi *k*-kat çapraz doğrulama tekniğinin mantıksal uç noktasıdır ve her birim bir kere test verisi olarak kullanılmaktadır. Bu sebeple bu yöntem kullanılırken çapraz doğrulama sırasında tekrar gerçekleştirilmemiştir. Gruplardaki gözlem sayısı birbirine eşit olup dört farklı gözlem sayısı olacak şeklinde oluşturulmuştur. Burada farklı gözlem sayılarında yöntemlerin performanslarındaki değişimin gözlenmesi amaçlanmıştır. 5-kat çapraz doğrulama için 100 tekrardan elde edilen ortalama doğruluk performansları ve standart sapmaları ile bir birimi hariç tutma doğruluk performansları 3, 4, ve 5 grup olması durumları için sırasıyla Çizelge 2, 3 ve 4'te verilmiştir

Çizelge 2'de yer alan sonuçlara göre düzgün dağılım ve normal dağılım durumunda sınıflandırıcı performansları birbirine yakın değerler almıştır. Bununla birlikte özellikle düzgün dağılım durumunda NB sınıflandırıcısının diğer modellere göre daha iyi performans sergilediği görülmüştür. Önerilen yöntem ise Normal dağılım ve gruplardaki birim sayısının az olduğu durumların bir kısmında en iyi performansı sergileyen yöntemlerden biri olmuştur. Çizelge 3'e göre, yine NB sınıflandırıcının hemen hemen tüm teknik ve gözlem sayılarında diğerlerinden daha yüksek doğrulama performansına sahip olduğunu görülmektedir. Çizelge 4'e göre ise yöntemlerin başarı performansları nispeten düşmüştür ve düzgün dağılım durumunda NB sınıflandırıcısının en iyi doğru sınıflandırma performansına sahip olduğu görülmektedir. Çizelge 2, 3 ve 4 sırasıyla "Ekler" bölümünde verilmiştir.

ERKEN GÖRÜNÜM

4.2. DOĞRUSAL FONKSİYON İLE AYRILAMAYAN GRUPLAR (GROUPS THAT CANNOT SEPARATED BY LINEAR FUNCTION)

Simülasyon çalışmasının bu bölümünde gruplara ait birimler çember şeklinde bir bölgede olacak şekilde üretilmiştir. Üretilen birimlerin iki karakteristiği bulunmaktadır. Bunlardan ilki yarı çap, diğeri ise açıdır. Düzgün dağılımdan üretilen rasgele sayının karekökü çemberin yarı çapını ifade eder. Ardından düzgün dağılımdan üretilen bu sayı t açısını belirlemede kullanılmaktadır. Böylece birimler, $(r \times \cos(t), r \times \sin(t))$ konumuna getirilmiştir. Gözlemlerin birbirlerine yakınlıkları bakımından yarıçap oluşturulurken iki farklı durum (r^A, r^B) dikkate alınmış ve yöntemlerin doğruluk performanslarının nasıl etkilendiği gözlemlenmiştir. A senaryosunda gözlemler birbirlerine oldukça yakın ve gözlemlerin doğru sınıflandırılması daha zor olacak şekilde tasarlanmıştır. Örnek teşkil etmesi bakımından tasarlanan senaryolara uygun olarak üretilen verilere ait bir saçılım grafiği Şekil 3'te verilmiştir.



Şekil 3. Üretilen verilere ilişkin saçılım grafiği (Scatter plot of the generated data)

Şekil 3'te görüldüğü üzere gruplarda yer alan veriler dairesel şekilde simetriktir. Bu sebeple bu grupları doğrusal bir sınıflandırıcı kullanarak ayırmak anlamlı olmayacaktır. Simülasyon çalışmasının bu bölümünde önerilen yöntem ve DVM 'de Gaussian Çekirdek fonksiyonu kullanılmıştır. Bir önceki çapraz doğrulama teknikleri ile değişen birim sayısına göre performansların karşılaştırılması için yine dört farklı gözlem sayısında testler gerçekleştirilmiştir. Çizelge 5, 6 ve 7 de sırasıyla 3, 4 ve 5 grup olması durumundaki sınıflandırma performansları yer almaktadır.

Çizelge 5'de ilk olarak üç gruplu bir sınıflandırma problemi ele alınmıştır. Çizelge 5'e göre, senaryo A durumunda her iki çapraz doğrulama tekniğini için önerilen model yüksek başarı performansına sahip olduğu görülmektedir. Yalnızca birimlerin 100'er adet olduğu LOO tekniğinde önerilen yöntem en iyi değildir. Aynı şekilde senaryo B durumunda da çoğunlukla önerilen yöntemin diğer sınıflandırıcılara karşı üstünlüğü görülmektedir. Ayrıca önerilen yöntemin başarı performansının üstünlüğü diğer sınıflandırıcılardan özellikle daha az gözleme sayısına sahip olduğu durumlarda çok daha belirgindir. Bunun yanında gözlem sayısındaki değişimin diğer sınıflandırıcıların başarı performanslarının değişiminde daha fazla rol oynadığı görülmektedir.

Çizelge 6'da grup sayısının dört olduğu durum incelendiğinde, yine senaryo A'da önerilen yöntemin diğer sınıflandırıcılardan üstünlüğü olduğu görülmektedir. Senaryo A için, gruplardaki birimlerin birbirine çok daha yakın olduğu ve yanlış sınıflandırma olmasının daha muhtemel olduğu bir problem yapısına sahip olduğu düşünüldüğünde önerilen yöntem böyle bir problem yapısında karşılaştırıldığı diğer sınıflandırıcılardan daha yüksek başarı performansı göstermiştir. Aynı grup sayısında senaryo B de ise yine çoğunluk olarak önerilen yöntem iyi olmakla beraber bazı örnek hacimlerinde DVM sınıflandırıcısının gerisinde kalmıştır. Çizelge 8'e göre ise önerilen yöntem her aşamada en iyi sınıflandırıcı performansına sahiptir.

Çizelge 6, 7 ve 8 "Ekler" kısmında verilmiştir.

5. GERÇEK VERİ UYGULAMASI (APPLICATION)

Çalışmada epileptik nöbet tespiti özelliğine sahip, çok yaygın olarak kullanılan bir veri setinin işlenmiş ve yeniden yapılandırılmış bir versiyonu kullanılmıştır. Veri setinin orijinal şekli, Almanya, Bonn Üniversitesi Epilepsi Bölümü'nün halka açık BONN veri tabanında yer almaktadır. EEG veri seti 5 setten oluşur ve her set tek kanallı 100 bölütten oluşmaktadır. Her bölüt 23,6 sn'lik bir zaman uzunluğuna sahiptir ve 4.096 örnekleme noktası içerir. El veya göz hareketlerinden kaynaklanan artefaktların, verileri oluşturan kişi tarafından manuel olarak silindiğini, dolayısıyla verileri önceden işlemeye gerek duyulmamaktadır. Tüm EEG sinyallerinin örnekleme frekansı 173,61 Hz'dir [64].

Her 4.097 veri noktası bir saniyelik bir süreye karşılık gelen 23 parçaya bölünmüş ve karıştırılmıştır. Her parça bir saniyeye karşılık gelen 178 veri noktası içermektedir ve her veri noktası farklı bir zaman noktasındaki EEG kaydının değerini ifade eder. Sonuçta veri seti 23x500=11.500 parça bilgi (satır) ve her bilgi bir saniye için 178 veri noktası (sütun) içerir. Buradaki son sütun ise $y = 1,2,3,4,5$ sınıf etiketlerini temsil etmektedir [65]. Çizelge 8'de veri kümesinin özelliklerini içeren bir tablo verilmiştir

ERKEN GÖRÜNÜM

Çizelge 8. EEG veri kümesinin özellikleri (EEG dataset features)

Veri Kümesi	Ölçüm Özelliği
1	5 hastadan nöbet sırasında intrakraniyal olarak kaydedilmiştir.
2	5 hastadan nöbet öncesinde, epileptojenik bölgeden intrakraniyal olarak kaydedilmiştir.
3	5 hastadan nöbet öncesinde, hipokampal oluşumlu karşı yarım küreden intrakraniyal olarak kaydedilmiştir.
4	5 sağlıklı gönüllüden gözler kapalı kaydedilmiş yüzey EEG kayıtları.
5	5 sağlıklı gönüllüden gözler açıkken kaydedilmiş yüzey EEG kayıtları.

5.1. ÖZNİTELİK ÇIKARIMI (FEATURE EXTRACTİON)

Öznitelik çıkarımı büyük veri kümesinin açıklamak için gereken kaynak miktarının azaltmayı amaçlar. Karmaşık bir veri üzerinde analiz yapılırken ortaya çıkan problemlerden biri sınıflandırma algoritmalarının eğitim kümesine aşırı uymasına (over fitting) sebep olması ve tahmin performansını düşürmesidir. Ayrıca EEG verileri hastalık, verinin alınma bölgesi ve beyin aktivasyonu bakımından birbirlerinden farklı gönüllü bireylerden alınmıştır. EEG kaydının sınıflara göre zaman bakımından değerlendirilmesi amaçlanmaktadır. Farklı grupta gönüllülerden oluşan beynin elektriksel aktivasyonu zaman bakımından farklılık gösterecektir. Bu sebeple veri bölütlerinin saniye açısından sinyal enerji yoğunluğu, bu sinyallerdeki ortalamaya göre değişimin değeri gibi EEG kayıtlarına ilişkin karakteristikleri belirleyecek çeşitli istatistiksel öznitelikler çıkarılmıştır.

Çıkarılan öznitelikler, ortalama sinyal gücünü belirlemek için kareler ortalaması, varyasyona ait bilgi için standart sapma, serinin düzenliliğini ölçen ve sistemin zaman içindeki değişimini gösteren entropi değerleri hesaplanmış ve sınıflandırma algoritmasında kullanılmıştır. Bu sayede epileptik nöbet tespiti için bir çıkarım elde etmenin yanı sıra hangi özniteliğin veriyi daha iyi karakterize ettiği ve ayrıştırabildiği gözlenmiştir.

5.2. SINIFLANDIRMA PERFORMANSI İÇİN BAŞARI ÖLÇÜTLERİ (SUCCESS CRITERIA FOR CLASSIFICATION PERFORMANCE)

Çalışmada kullanılan sınıflandırma algoritmalarının karşılaştırılması literatürde yer alan başarı ölçütlerine göre yapılmıştır. Bu ölçütler Çizelge 9'da verilen ifadelerden elde edilen karışıklık matrisinden elde edilmektedir.

Çizelge 9. Karışıklık matrisi ve başarı ölçütleri (Confusion matrix and success measures))

		Tahmin Edilen Sınıf	
		Pozitif	**Negatif**
Gerçek Sınıf	**Pozitif**	Doğru-Pozitif (True Positive:TP)	Yanlış-Negatif (False Negative:FN)
	Negatif	Yanlış-Pozitif (FalsePositive:FP)	Doğru-Negatif (True Negative:TN)

Karışıklık matrisinden yöntemlerin sınıflandırma performanslarının kıyaslanması için birçok çıkarım yapılabilir. Burada; *Doğruluk* $\left(\frac{TP+TN}{TP+TN+FP+FN}\right)$: Genel olarak, sınıflayıcının ne sıklıkta doğru tahmin ettiğinin ölçüsüdür.

Yanlış sınıflandırma $\left(\frac{FP+FN}{TP+TN+FP+FN}\right)$: Genel olarak sınıflayıcının ne sıklıkla yanlış tahmin ettiğinin ölçüsüdür. Hata oranı olarak da bilinir.

Özgünlük $\left(\frac{TN}{FP+TN}\right)$: Sınıflayıcının ne kadar gerçek negatif değeri doğru tahmin ettiğinin bir ölçüsüdür. Seçicilik olarak bilinir.

Hassaslık $\left(\frac{TP}{FP+TP}\right)$: Sınıflar arasından doğru olarak ne kadar tahmin edildiğinin ölçüsüdür. Pozitif tahmin edici değer olarak da bilinir.

Duyarlılık $\left(\frac{TP}{TP+FN}\right)$: Sınıflayıcının ne kadar gerçek pozitif değeri doğru tahmin ettiğinin ölçüdür. İsabet oranı olarak bilinir.

F *skoru* $2 \times \left(\frac{Hassaslık \times Duyarlılık}{Hassaslık+Duyarlılık}\right)$: Gerçek pozitif oranının ve hassasiyetin harmonik ortalamasıdır. Sınıflandırıcının ne kadar iyi performans gösterdiğinin bir ölçüsüdür ve sınıflandırıcıları karşılaştırmakta sıklıkla kullanılır.

6. BULGULAR (FINDINGS)

Çalışmada yeni bir, çok gruplu sınıflandırma yöntemi önerilmiştir. Önerilen yöntem literatürde sıklıkla konu edilen Bonn veri setine uygulanmış ve sınıflandırma performansı diğer yöntemlerle karşılaştırılmıştır. Performans karşılaştırılmasında LOO ve 5-kat çapraz doğrulama yöntemi kullanılmış ve elde edilen bulgular Çizelge 10, 11 ve 12'de verilmiştir. En iyi doğruluk performansı LOO için sonuçlar koyulaştırılmış, 5-kat çapraz doğrulama tekniği için yıldız işareti ile belirtilmiştir. Değişen miktarların büyüklüğünün ölçülmesinde kullanılan istatistiksel bir yöntem olan ortalama karekök özniteliği özellikle sinyal analizlerinde sıklıkla kullanılan bir özelliktir [66–68]. Bu öznitelik dikkate alınarak yapılan ve LOO tekniğinin kullanıldığı çapraz doğrulama sınıflandırma performansına göre NB ve önerilen yöntemin çok yüksek bir olasılıkla veriyi doğru sınıflandırdığı görülmektedir. Her iki yönteminde

ERKEN GÖRÜNÜM

tüm başarı ölçütleri birbirlerine çok yakın değere sahiptir. Özellikle önerilen yöntemin DVM sınıflandırıcısından tüm başarı ölçütlerine göre daha iyi sonuca sahip olduğu görülmektedir. Diğer çapraz doğrulama tekniği ile elde edilen sonuçlar incelendiğinde ise yine NB modelinin doğru sınıflandırma performansının diğer yöntemlerden iyi olduğu görülmekle birlikte önerilen yöntemin doğruluk ve buna bağlı ölçülen diğer metriklerde performansının düştüğü gözlenmiştir.

Çizelge 10. Ortalama karekök özniteliği için sınıflandırma performansları (Correct classification ratio for the root mean square feature)

		Önerilen Model	DVM	FDDF	k -EYK	NB
Doğruluk	LOO	**0.9913**	0.7565	0.7130	0.8957	**1**
	5 kat	*0.76*	*0.78*	*0.75*	*0.90*	1*
Yanlış Sınıflandırma	LOO	**0.0087**	0.2435	0.2870	0.1043	**0**
	5 kat	*0.2400*	*0.2200*	*0.25*	*0.10*	0*
Özgünlük	LOO	**0.9891**	0.6957	0.6413	0.8696	**1**
	5 kat	*0.4778*	*0.7250*	*0.6879*	*0.8750*	1*
Hassaslık	LOO	**0.9583**	0.4510	0.4107	0.6571	**1**
	5 kat	-	*0.4798*	*0.4489*	*0.6867*	1*
Duyarlılık	LOO	**1**	**1**	**1**	**1**	**1**
	5 kat	*0*	*1*	*1*	*1*	*1**
***F* Skoru**	LOO	**0.9787**	0.6216	0.5823	0.7931	**1**
	5 kat	-	*0.4462*	*0.6183*	*0.8089*	1*

Çizelge 11. Standart Sapma özniteliği için sınıflandırma performansları (Correct classification ratio for standard deviaton feature)

	Önerilen Model	DVM	FDDF	k -EYK	NB
Doğruluk	**0.8609**	0.6957	0.3304	0.8522	0.6957
	0.73	*0.73*	*0.31*	*0.84**	*0.70*
Yanlış Sınıflandırma	**0.1391**	0.3043	0.6696	0.1478	0.3043
	0.27	*0.27*	*0.69*	*0.16*	*0.30*
Özgünlük	**0.9239**	0.6304	0.3043	0.8261	0.6304
	0.7125	*0.6750*	*0.2875*	*0.8125**	*0.6375*
Hassaslık	**0.6667**	0.3929	0.1351	0.5789	0.3929
	0.3721	*0.4239*	*0.1213*	*0.5619**	*0.4070*
Duyarlılık	0.6087	**0.9565**	0.4348	**0.9565**	**0.9565**
	0.80	*0.95**	*0.40*	*0.55*	*0.95**
***F* Skoru**	0.6367	0.5570	0.2062	**0.7213**	0.5570
	0.5053	*0.58*	*0.1854*	*0.7042**	*0.5664*

Çizelge 12. Entropi (l^p norm $1 \leq p$) özniteliği için sınıflandırma performansları (Correct classification ratio for entropi feature)

	Önerilen Model	DVM	FDDF	k -EYK	NB
Doğruluk	**0.9130**	0.6348	0.2870	0.8000	0.7391
	0.78	*0.72*	*0.39*	*0.79**	*0.78*
Yanlış Sınıflandırma	**0.0870**	0.3652	0.7130	0.2000	0.2609
	0.22	*0.28*	*0.62*	*0.21**	*0.22*
Özgünlük	**0.9022**	0.5543	0.2391	0.7609	0.6848
	*0.7750**	*0.6625*	*0.35*	*0.75*	*0.7375*
Hassaslık	**0.7097**	0.3492	0.1358	0.5000	0.4314
	0.4365	*0.4298*	*0.1648*	*0.5026**	*0.4861*
Duyarlılık	**0.9565**	0.9565	0.4783	0.9565	0.9565
	0.80	*0.95**	*0.50*	*0.95**	*0.95**
***F* Skoru**	**0.8148**	0.5116	0.2115	0.6567	0.5946
	0.6496	*0.5876*	*0.2477*	*0.653*1*	*0.6395*

LOO tekniği ile elde edilen sonuçlar dikkate alındığında, doğruluk, yanlış sınıflandırma, özgünlük ve hassaslık ölçütleri bakımından en başarılı sınıflandırıcı önerilen sınıflandırma yöntemi olmuştur. Özellikle önerilen yöntem ile benzer çalışma prensibine sahip DVM sınıflandırıcısı arasında başarı ölçütlerinin performansları arasında ciddi farklar bulunmaktadır. Modellerin, 5-kat çapraz doğrulama tekniği bakımından doğru sınıflandırma performansları incelendiğinde ise *k*-EYK 'nın en başarılı model olduğu görülmektedir. FDDF yöntemi hariç diğer modellerin başarı performansı ise birbirlerine yakındır. FDDF yönteminin sınıflandırma

ERKEN GÖRÜNÜM

performansı oldukça düşüktür ve bu durum standart sapma özniteliği veri yapısının doğrusal bir ayırıcı fonksiyon ile ayrılmaya uygun olmadığı göstermektedir.

Çizelge 12'ye göre, çalışmada kullanılan bir diğer değişkenlik metriği, bir zaman serisinin düzenliliğini veya tahmin edilebilirliğini ölçen ve sistemin zaman içinde nasıl geliştiğini gösteren entropidir [69,70]. Bu özniteliğe göre yapılan sınıflandırmaya göre ilk çapraz doğrulama tekniğinde önerilen yöntem diğer tüm yöntemlerden daha iyi performans sergilemiştir. Önerilen yöntemin doğruluk, yanlış sınıflandırma, özgünlük, duyarlılık ve *F* skoru başarı ölçütlerinin tamamında diğer sınıflandırıcılardan fark edilir seviyede üstün olduğu görülmektedir. Diğer çapraz doğrulama tekniği ile elde edilen sonuçlarda ise FDDF modelinin yine çok düşük başarı oranına sahip olduğu görülmekte ve bu yöntemin bu sınıflandırma probleminde kullanımının çok uygun olmadığı sonucunu ortaya çıkarmaktadır. Önerilen model ile diğer sınıflandırma yöntemlerinin doğru sınıflandırma performansları ise birbirine yakın değerler almaktadır.

7. **SONUÇLAR VE TARTIŞMA** (RESULTS AND DISCUSSION)

DVR, regresyon parametrelerinin elde edilişinde alternatif bir yöntem sunmaktadır. Bu yöntem aykırı değerlere karşı en küçük karelerden daha avantajlı olmasının yanı sıra çekirdek fonksiyonlarının kullanımı ile doğrusal olamayan durumlarda da kullanım kolaylığı oluşturmaktadır. Önerilen yöntem destek vektör makine regresyon ile [46]' de önerilen matematiksel sınıflandırma modelinin ikinci aşamasının bir birleşimi şeklindedir. DVM modeli iki gruplu durumlar için tasarlanmış olup grup sayısı arttığında ikili karşılaştırma yöntemleriyle sınıflandırma işlemini gerçekleştirir. Bu durum performans ve zaman kaybına neden olmaktadır [58]. Önerilen yöntemde teorik olarak grup sayısı önemli değildir. Yapılan simülasyon çalışmasında, grupların doğrusal bir düzlem ile ayrılamadığı problem türlerinde ortaya konan yeni yaklaşım, karşılaştırılan diğer sınıflandırıcılardan çok daha yüksek başarı performansına sahip olduğu görülmüştür. Bunun yanı sıra yine aynı tür problemlerde yöntemin doğru sınıflandırma performansının gözlem değerlerinin değişiminden fazlaca etkilenmediği ancak diğer sınıflandırıcıların grup sayılarının değişiminden etkilendiği, birim sayısı azaldıkça doğru sınıflandırma performanslarının azaldığı görülmüştür. Bu açıdan ortaya koyulan yeni yaklaşımın diğer sınıflandırıcılara kıyasla bu tür problem türlerinde daha güvenilir bir karar kuralı oluşturmaktadır.

Çalışmada önerilen yöntem gerçek bir veri seti için de uygulanmıştır. Literatürde Bonn veri seti ile olarak bilinen bu veri seti için, öncelikle çeşitli öznitelikler kullanılarak sinyallerin ayrıştırılması sağlanmıştır. Önerilen yöntem ve dört farklı sınıflandırıcının öznitelik veri setlerinin sınıflandırması iki farklı çapraz doğrulama tekniği ile gerçekleştirilmiştir. Sınıflandırıcı algoritmalarının performanslarını karşılaştırmak için literatürdeki başarı ölçütleri elde edilmiştir. Elde edilen sonuçlara göre LOO çapraz doğrulama tekniği ile elde edilen sonuçlarda önerilen yöntem tüm öznitelik veri setlerinde benzer bir sınıflandırma prensibine sahip olan DVM sınıflandırıcısından daha iyi başarı ölçütüne sahiptir. Karekök ortalama ve standart sapma özniteliği veri setlerinde önerilen yöntemin tüm başarı ölçütleri oldukça yüksek olduğu ve çoğu yöntemden daha iyi bir performansa sahip olduğu görülmüştür. Bunun yanında entropi özniteliği veri seti sınıflandırmasında önerilen yöntem tüm başarı ölçütlerinde en iyi performans sonuçlarına sahiptir. Diğer çapraz doğrulama tekniğinde ise önerilen yöntemin başarı performansı düşerken diğer yöntemlerin performansları çok fazla değişmemiş buna bağlı olarak da önerilen yöntemin başarı sıralaması bu teknikde değişmiştir. Bunun sebebi, LOO tekniğinde çok daha fazla gözlemin eğitim seti içerisinde yer alması ve modelin eğitimine katılıyor olmasından kaynaklanmaktadır. Veri setinde yeterince gözlem sayısı bulunmadığında *k*-kat çapraz doğrulama tekniğinde rasgele seçilen birimlere bağlı olarak sonuçların daha yanlı olması beklenirken *k* değeri arttıkça, rasgele satır atamalarından kaynaklanan sapma riski azalacaktır.

Literatürde Bonn veri seti ile yapılan onlarca çalışma bulunmaktadır. Bu çalışmaların birçoğu ikili karşılaştırmalardan oluşmaktadır. Veri setindeki kümeler hasta veya sağlıklı bireylerin durumuna göre ikili karşılaştırmalar ile sınıflandırma gerçekleştirilmiştir. Bu çalışmada diğer çalışmalardan farklı olarak tüm sınıflar dikkate alınmış ve sınıflandırma tüm sınıflar için yapılmıştır. Fakat çalışma benzer bir şekilde diğer durumlar içinde genişletilebilir ve epileptik nöbet ve değil şeklinde ikili karşılaştırmalar da yapılabilir.

EEG sinyal kayıtları elde edilmesi birçok bakımından zor olan bir veri türüdür ve elde edilen veriler doğrusal ayrılabilen bir veri yapısına sahip değildir. Bu haliyle önerilen sınıflandırma yönteminin EEG sinyallerinin sınıflandırılması problemlerinde etkin bir şekilde kullanılabileceği görülmektedir. Başka bir açıdan önerilen yöntem farklı öznitelikler ile kullanılabilir ve bununla alakalı farklı çalışmalar yapılabilir.

Önerilen yöntem birçok açıdan sınıflandırma algoritmalarına katkı sunmaktadır. Matematiksel programlamaya dayalı sınıflandırma algoritmalarının doğrusal durumdaki problemler için geliştirilmesi bu algoritmaların kullanım alanını kısıtlamaktadır. Gerçek hayat problemlerinde veri gruplarının her zaman doğrusal ayrılması mümkün olmamaktadır. Bu açıdan önerilen yöntem doğrusal olmayan veri yapılarında kullanılabilmesi açısından matematiksel programlama yaklaşımlarına yenilik kazandırmaktadır. Çalışmada sınıflandırma skorlarının destek vektör regresyon yaklaşımı ile elde edilişi, önerilen modelin doğrusal olmayan durumlarda kullanımın yanı sıra aykırı gözlemlere karşı daha dayanıklı bir model ileri sürülmesini sağlamıştır. Bu açıdan yapılan çalışmadan ayrı olarak, regresyon ağaçları ya da gaussian regresyon

ERKEN GÖRÜNÜM

süreci gibi farklı regresyon yaklaşımları kullanılarak veri yapılarına uygun yeni modeller geliştirilebilir.

ETİK STANDARTLARIN BEYANI (DECLARATION OF ETHICAL STANDARDS)

Bu makalenin yazar(lar)ı çalışmalarında kullandıkları materyal ve yöntemlerin etik kurul izni ve/veya yasal-özel bir izin gerektirmediğini beyan ederler.

YAZARLARIN KATKILARI (AUTHORS' CONTRIBUTIONS)

Esra Betül KINACI: Literatür taraması, modelin ortaya konması ile kodun yazılması ve analizlerin yapılıp yorumlanmasını sağlamıştır. enabled the model to be revealed, the code to be written, and the analyzes to be made and interpreted

Hasan BAL: Problemin öne sürülmesi, son kontrollerin yapılması gerçekleştirmiştir.

Harun KINACI: Hesaplamaların ve analizlerin yapılması ve makalenin yazım işlemini gerçekleştirmiştir.

ÇIKAR ÇATIŞMASI (CONFLICT OF INTEREST)

Bu çalışmada herhangi bir çıkar çatışması yoktur.

KAYNAKLAR (REFERENCES)

[1] T.S. Kumar, V. Kanhangad, R.B. Pachori, "Classification of seizure and seizure-free EEG signals using local binary patterns", ***Biomedical Signal Processing and Control,*** 15: 33–40, (2015).

[2] Y. Kaya, T. Ramazan, "Epileptik EEG işaretlerinin sınıflandırılması için yeni bir öznitelik çıkarım yöntemi", ***Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi,*** 22:(2018).

[3] J. Jing, X. Pang, Z. Pan, F. Fan, Z. Meng, "Classification and identification of epileptic EEG signals based on signal enhancement", ***Biomedical Signal Processing and Control***, 71:(2022).

[4] X. Ma, Z. Zhuo, L. Wei, Z. Ma, Z. Li, H. Li, "Altered temporal organization of brief spontaneous brain activities in patients with alzheimer's Disease", ***Neuroscience,*** 425: 1–11, (2020).

[5] Y. Fu, X. Xiong, C. Jiang, B. Xu, Y. Li, H. Li, "Imagined hand clenching force and speed modulate brain activity and are classified by NIRS combined with EEG", ***IEEE Transactions on Neural Systems and Rehabilitstion Engineering,*** 25: 1641–1652, (2017).

[6] R. W. Peng, "Classification method of EEG based on emprical mode decomposition and SVM", ***Comp. Meas. Cont.,*** 28: 189–194, (2020).

[7] W. Klonowski, "Everything you wanted to ask about EEG but were afraid to get the right answer", ***Nonlinear Biomedical Physics***, 3: 1–6, (2009).

[8] M. Heyden, "***Classification of EEG data using machine learning techniques"***, Lund University Press, (2016).

[9] Ö. Türk, M.S. Özerdem, "Epileptik EEG sinyallerinin sınıflandırılması için bir boyutlu medyan yerel ikili örüntü temelli öznitelik çıkarımı", ***GU Journal of Sciences, Part C***., 5: 97–107, (2017).

[10] E. Tuncer, E.D. Bolat, "Destek vektör makineleri ile EEG sinyallerinden epileptik nöbet sınıflandırması Epileptic seizure classification from eeg signals with support vector machines", ***Politeknik Dergisi,*** 25(1): 239–249, (2022).

[11] S. Ramakrishnan, A.S. Muthanantha Murugavel, P. Sathiyamurthi, J. Ramprasath, "Seizure detection with local binary pattern and CNN classifier", ***Journal of Physsics: Conference Series***, 1767: (2021).

[12] V. Srinivasan, C. Eswaran, A.N. Sriraam, "Artificial neural network based epileptic detection using time-domain and frequency-domain features", ***Journal of Medical Systems,*** 29: 647–660, (2005).

[13] E.D. Übeyli, "Decision support systems for time-varying biomedical signals: EEG signals classification", ***Expert Systems withApplications***. 36(2): 2275–2284, (2009).

[14] H. Ocak, "Automatic detection of epileptic seizures in EEG using discrete wavelet transform and approximate entropy", ***Expert Systems withApplications***. 36(2):2027–2036, (2009).

[15] Y. Li, P.P. Wen, "Clustering technique based least square support vector machine for EEG signal classification", ***Computer Methods and Programs in Biomedicine***. 104: 358–372, (2011).

[16] S. Chandaka, A. Chatterjee, S. Munshi, "Cross-correlation aided support vector machine classifier for classification of EEG signals", ***Expert Systems withApplications,*** 36(2): 1329–1336, (2009).

[17] I. Guler, E.D. Ubeyli, "Multiclass support vector machines for EEG-signals classification", ***IEEE Transactions on Information Technology Biomedicine,*** 11(2): 117–126, (2007).

[18] L. Guo, D. Rivero, J. Dorada, J.R. Rabunal, A. "Pazos, Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks", ***Journal of Neuroscience Methods,*** 191(1): 101–109, (2010).

[19] L. Guo, D. Rivero, A. Pazos, "Epileptic seizure detection using multiwavelet transform based approximate entropy and artificial neural networks", ***Journal of Neuroscience Methods,*** 193(1): 156–163, (2010).

[20] U. Orhan, M. Hekim, M. Özer, "Epileptic seizure detection using artificial neural network and a new feature extraction approach based on equal width discretization", ***Journal of Faculty Engineering and Architecture of Gazi University,*** 26(3): 575–580, (2011).

[21] U. Orhan, M. Hekim, M. Ozer, "Epileptic seizure detection using probability distribution based on equal frequency discretization", ***Journal of. Medical Systems,*** 36(4): 2219–2224, (2012).

[22] M. Hekim, "The classification of EEG signals using discretization-based entropy and the adaptive neuro-fuzzy inference system", ***Turkish Journal of Electrical Engineering & Computer Sciences*** 24: 285–297, (2016).

[23] N. Nicolaou, J. Georgiou, "Detection of epileptic electroencephalogram based on permutation entropy and support vector machines", ***Expert Systems with Applications***, 39: 202–209, (2012).

[24] T.K. Gandhi, P. Chakraborty, G.G. Roy, B.K. Panigrahi,

ERKEN GÖRÜNÜM

"Discrete harmony search based expert model for epileptic seizure detection in electroencephalography", ***Expert Systems with Applications,*** 39: 4055–4062, (2012).

[25] D. Wang, D. Miao, C. Xie, "Best basis-based wavelet packet entropy feature extraction and hierarchical EEG classification for epileptic detection", ***Expert Systems with Applications,*** 38: 14314–14320, (2011).

[26] K. Polat, S. Güneş, "Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform", ***Applied Mathematics and Computation,*** 18: 1017–1026, (2007).

[27] M. Mursalin, Y. Zhang, Y. Chen, N. V. Chawla, "Automated epileptic seizure detection using improved correlation-based feature selection with random forest classifier", ***Neurocomputing.*** 241: 204–214, (2017).

[28] S.C. Satapathy, J.V.R. Murthy, P.V.G.D. Prasad Reddy, B.B. Misra, P.K. Dash, G. Panda, "Particle swarm optimized multiple regression linear model for data classification", ***Applied Soft Computing,*** 9(2): 470–476, (2009).

[29] M.I. Doğan, A. Orman, M. Örkcü, H.H. Örkcü, "A new approach based on regression analysis and mathematical programming to multigroup classification problems", ***Journal of the Faculty of Engineering and Architecture of Gazi University,*** 34(4): 1939–1955, (2019).

[30] K.. Lam, J.. Moy, "An experimental comparison of some linear programming approaches to the discriminant problem", ***Computers & Operations Research,*** 24(7): 593–599, (1997).

[31] A. Tekerek, "Support vector machine based spam SMS detection", ***Politeknik Dergisi***, 22(2): 779–784, (2019).

[32] U. Köse, "Zeki optimizasyon tabanlı destek vektör makineleri ile diyabet teşhisi", ***Politeknik Dergisi,*** 22(3): 557–566, (2019).

[33] B.E. Boser, I.M. Guyon, V.N. Vapnik, "A Training Algorithm for Optimal Margin Classifiers", ***Proceedings Annual Conferences of Computational Learning Theory,*** 144–152, (1992).

[34] V. Cherkassky, F. Mulier, "***Learning From Data: Concepts, Theory and Methods***", John Wiley, New York, (1998).

[35] V. Cherkassky, X. Shao, F. Mulier, V. Vapnik, "Model Complexity control for regression using VC generalization bounds", ***IEEE Transactions Neural Networks***, 10: 1075–1089, (1999).

[36] S. Haykin, "***Neural networks: A comprehensive foundation***", 2nd ed., Prentice Hall, Upper Saddle River, New Jersey, (2001).

[37] S. Haykin, "***Neural networks: A comprehensive foundation***", 4th ed, Pearson Education, Singapore, (2003).

[38] S. Tripathi, V. V. Srinivas, R.S. Nanjundiah, "Downscaling of precipitation for climate change scenarios: A support vector machine approach", ***Journal of Hydrology,*** 330(3-4): 621–640, (2006).

[39] S.R. Gunn, "***Support Vector Machines for Classification and Regression***", (1998)

[40] S. Haykin, "***Neural networks and learning machine***", 3rd ed., Pearson Prentice-Hall, New York.

[41] A.J. Smola, B. Schölkopf, "A tutorial on support vector regression", ***Statistics and Computing,*** 14: 199–222, (2004).

[42] J. Mercer, "Functions of positive and negative type and their connection with the theory of integral equations", ***Philosoptical Transactions of the Royal Society of London,*** 209: 415–446, (1909).

[43] R. Courant, D. Hilbert, "***Methods of Mathematical Physics***", Wiley Interscience, New York, (1970).

[44] N. Freed, F. Glover, "A linear programming approach to the discriminant problem", ***Decision Sciences***, 12(1): 68–74, (1981).

[45] N. Freed, F. Glover, "Simple but powerful goal programming models for discriminant problems", ***European Journal of Operational Research,*** 44–66, (1981).

[46] K. Lam, E.. Choo, W. Moy, J, "Minimizing deviations from the group mean: A new linear programming approach for the two-group classification problem", ***European Journal of Operational Research,*** 358–367, (1996).

[47] K.. Lam, J.. Moy, "Improved linear programming formulations for the multigroup discriminant problem", ***The Journal of the Operational Research Society***, 47(12):1526–1529, (1996).

[48] H. Bal, H.H. Örkcü, S. Çelebioğlu, "An alternative model to Fisher and linear programming approaches in two-group classification problem: minimizing deviations from the group median", ***GU Journal of Sciences***, 49–55, (2006).

[49] T. Sueyoshi, "DEA-discriminant analysis in the view of goal programming", ***European Journal of Operational Research,*** 115: 564–582, (1999).

[50] A. Stam, E.A. Joachimsthaler, "A comparison of a robust mixed-integer approach to existing methods for establishing classification rules for the discriminant problem", ***European Journal of Operational Research,*** 46: 113–122, (1990).

[51] R. Pavur, C. Loucopoulos, "Examining optimal criterion weights in mixed integer programming approaches to the multi group classification problem", ***The Journal of the Operational Research Society***, 626–640, (1995).

[52] J.M. Wilson, "Integer programming formulations of statistical classification problems", ***Omega-International Journal of Management Science,*** 24(6): 681–688, (1996).

[53] T. Sueyoshi, "Mixed integer programming approach of extended DEA-Discriminant analysis", ***European Journal of Operational Research,*** 45–55: (2004).

[54] T. Sueyoshi, "DEA-Discriminant analysis: Methodological comparison among eight discriminant analysis approaches", ***European Journal of Operational Research,*** 169: 247–272, (2006).

[55] H. Bal, H.H. Örkcü, "A new mathematical programming approach to multi-group classification problems", ***Computers & Operations Research,*** 38: 105–111, (2011).

[56] W. Gochet, A. Stam, V. Srinivasan, S. Chen, "Multigroup discriminant analysis using linear programming", ***Operations Research***, 45(2): 213–225, (1997).

[57] G. Xu, L.G. Papageorgiou, "A mixed integer optimisation

ERKEN GÖRÜNÜM

model for data classification", ***Computers & Industrial Engineering,*** 56(4): 1205–1215, (2009).

[58] F. Uney, M. Türkay, "A mixed-integer programming approach to multi-class data classification problem", ***European Journal of Operational Research.***, 173: 910–920, (2006).

[59] L. Yang, S. Liu, S. Tsoka, L.G. Papageorgiou, "Sample re-weighting hyper box classifier for multi-class data classification", ***Computers & Industrial Engineering,*** 85: 44–56, (2015).

[60] C. Cortes, V. Vapnik, "Support-Vector Networks", ***Machine Learning***, 20: 273–297, (1995).

[61] R.A. Fisher, "The use of multiple measurements in taxonomic problems", ***Annals of Eugenics***. 7: 179–188, (1936).

[62] J. Lakoumentas, J. Drakos, M. Karakantza, G. Sakellaropoulos, V. Megalooikonomou, G. Nikiforidis, "Optimizations of the naïve-Bayes classifier for the prognosis of B-Chronic lymphocytic leukemia incorporating flow cytometry data", ***Computers Methods and Programs in Biomedicine***, 108(1): 158–167, (2012).

[63] S.-K. Lee, P. Kang, S. Cho, "Probabilistic local reconstruction for k-NN regression and its application to virtual metrology in semiconductor manufacturing", ***Neurocomputing,***. 131: 427–439, (2014).

[64] R.G. Andrzejak, K. Lehnertz, F. Mormann, C. Rieke, P. David, C.E. Elger, "Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state", ***Physical Review:E,*** 64: (2001)

[65] **https://www.kaggle.com/datasets/harunshimanto/epileptic-seizure-recognition.**

[66] M.D. Latt, H.B. Menz, V.S. Fung, S.R. Lord, "Acceleration patterns of the head and pelvis during gait in older people with Parkinson's disease: A comparison of fallers and nonfallers", ***Journal of Gerontology Series A,*** 64(6): 700–706, (2009).

[67] M. Sekine, T. Tamura, M. Yoshida, Y. Suda, Y. Kimura, H. Miyoshi, Y. Kijima, Y. Higashi, T. Fujimoto, "A gait abnormality measure based on root mean square of trunk acceleration", ***Journal of Neuroengineering and Rehabilitation,*** 10(118): (2013).

[68] E. de M. Mesquita, F.B. Rodrigues, A.P. Rodrigues, T.S. Lemes, A.O. Andrade, M.F. Vieira, "Discrimination capability of linear and nonlinear gait features in group classification", ***Medical Engineering & Physics,*** 93: 59–71, (2021).

[69] J.H. Ko, K.M. Newell, "Aging and the complexity of center of pressure in static and dynamic postural tasks", ***Neuroscience Letters***, 610: 104–109, (2016).

[70] S. Ramdani, B. Seigle, J. Lagarde, F. Bouchara, P.L. Bernard, "On the use of sample entropy to analyze human postural sway data", ***Medical Engineering & Physics,*** 31(8): 1023–1031, (2009).

ERKEN GÖRÜNÜM

EKLER

Ek 1.

Çizelge 2. 3 Grup durumu için çapraz doğru sınıflandırma performansı. (Correct classification ratio for 3 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				Bir birimi hariç tutma			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Düzgün Dağılım	Önerilen Model	0.9567 *0.0954*	0.9360 *0.0669*	0.9700 *0.0288*	0.9782 *0.0187*	0.9546 *	0.9600	0.9800 *	0.9833 *
	DVS	0.9500 *0.0969*	0.9080 *0.0623*	**0.9747** *0.0258*	0.9761 *0.0195*	0.9600 *	0.9667	0.9800 *	0.9767 *
	FLDF	0.9400 *0.0844*	0.9440 *0.0680*	0.9648 *0.0331*	0.9705 *0.0211*	0.9667 *	0.9467 *	**0.9667** *	0.9733 *
	k -EYK	0.9033 *0.0688*	0.8947 *0.0777*	0.9644 *0.0313*	0.9793 *0.0176*	0.9333 *	0.9667	0.9930 *	0.9500 *
	NB	**0.9700** *0.1053*	**0.9507** *0.0574*	0.9689 *0.0305*	**0.9815** *0.0167*	**0.9800** *	**0.9946**	0.9900 *	**0.9967**
Normal dağılım	Önerilen Model	**0.9667** *0.0745*	0.8933 *0.1011*	0.9400 *0.0641*	0.9033 *0.0617*	**0.9333** *	**0.9733** *	0.9600 *	0.9367 *
	DVS	**0.9667** *0.0725*	**0.9200** *0.0869*	**0.9467** *0.0506*	0.9333 *0.0527*	0.8667 *	**0.9733** *	0.9333 *	0.9333 *
	FLDF	0.9333 *0.0913*	0.9067 *0.1116*	0.9267 *0.0548*	0.9133 *0.0532*	0.9000 *	0.9600	0.9535	0.9267
	k -EYK	**0.9667** *0.0745*	**0.9200** *0.0548*	0.9067 *0.0435*	**0.9433** *0.0325*	0.9000 *	0.9467 *	0.9067 *	**0.9567** *
	NB	0.9333 *0.0913*	0.8933 *0.1011*	0.9267 *0.0641*	0.9233 *0.0401*	**0.9333** *	0.9700 *	**0.9667** *	0.9533 *

Ek 2.

Çizelge 3. 4 Grup durumu için çapraz doğru sınıflandırma performansı (Correct classification ratio for 4 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				Bir birimi hariç tutma			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Düzgün Dağılım	Önerilen Model	0.9155 *0.0834*	0.9590 *0.0407*	0.9653 *0,0250*	0.9532 *0.0216*	**0.9250** *	0.9300 *	0.9015 *	0.8975 *
	DVS	0.9250 0.0816	0.9574 *0.0415*	0.9567 *0.0290*	0.9550 *0.1215*	0.8750 *	0.9400 *	0.9230 *	0.9150 *
	FLDF	0.9330 0.0812	0.9560 *0.0435*	0.9602 *0.0310*	0.9472 *0.0243*	0.7750 *	0.9000 *	0.9180 *	0.9050 *
	k -EYK	0.8675 0.1002	0.9370 *0.0480*	0.9390 *0.0364*	0.9460 *0.0197*	0.8000 *	0.9100 *	0.9205 *	0.9400 *
	NB	**0.9290** *0.0837*	**0,.9690** *0.0255*	**0.9702** *0.0259*	**0.9586** *0.0211*	0.9000 *	**0.9600** *	**0.9550** *	**0.9600** *
Normal dağılım	Önerilen Model	0.9015 *0.0989*	0.9331 *0.0527*	0.9600 *0.0265*	0.9569 *0.0235*	**0.9750** *	**0.9600** *	0.9700 *	0.9425 *
	DVS	0.9055 *0.1022*	0.9679 *0.0371*	0.9663 *0.0255*	0.9696 *0.0185*	0.9500 *	0.9400 *	0.9750 *	0.9500 *
	FLDF	0.9425 *0.0815*	0.9665 *0.0394*	0.9662 *0.0279*	0.9738 *0.0182*	0.9500 *	0.9400 *	0.9750 *	0.9550 *
	k -EYK	**0.9303** *0.0877*	0.9383 *0.3426*	**0.9781** *0.0226*	0.9690 *0.0173*	0.9250 *	0.9300 *	0.9900 *	**0.9725** *
	NB	0.9130 *0.0985*	**0.9682** *0.0385*	0.9610 *0.0308*	**0.97774** *0.0168*	0.9500 *	0.9500 *	**0.9800** *	0.9625 *

ERKEN GÖRÜNÜM

Ek 3.

Çizelge 4. 5 Grup durumu için çapraz doğru sınıflandırma performansı (Correct classification ratio for 5 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				Bir birimi hariç tutma			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Düzgün Dağılım	Önerilen Model	0.7726 *0.0284*	0.7522 *0.0327*	0.7420 *0.0121*	0.7698 *0.0085*	0.7960	0.7648	0.7704	0.7624
	DVS	0.7552 *0.0304*	0.7728 *0.0165*	0.7820 *0.0133*	0.8092 *0.0770*	0.7640	0.7808	0.7976	0.7924
	FLDF	0.7982 *0.0277*	0.7286 *0.0218*	0.7750 *0.0132*	0.7892 *0.0073*	0.7808	0.7584	0.7776	0.7604
	k -EYK	0.7974 *0.0220*	0.8609 *0.0187*	0.7988 *0.0116*	0.8963 *0.0058*	0.7584	0.8272	0.85552	0.8736
	NB	**0.8780** *0.0316*	**0.8803** *0.0187*	**0.8616** *0.0059*	**0.9020** *0.0036*	**0.8272**	**0.8912**	**0.8872**	**0.8980**
Normal dağılım	Önerilen Model	0.8808 *0.0213*	0.9024 *0.0710*	0.9497 *0.006*	0.9196 *0.0035*	**0.9120**	**0.9232**	0.9192	0.9060
	DVS	0.8632 *0.0200*	0.9422 *0.0086*	**0.9712** *0.0026*	**0.9457** *0.053*	0.8800	0.9184	0.9296	0.9124
	FLDF	0.8992 *0.0184*	0.9200 *0.008*	0.9488 *0.0018*	0.9228 *0.0033*	0.8960	0.9184	0.9216	0.9112
	k -EYK	0.8852 *0.0144*	**0.9438** *0.0090*	0.9623 *0.0062*	0.9256 *0.0049*	0.9040	0.9264	**0.9304**	**0.9240**
	NB	**0.9006** *0.0100*	0.9390 *0.0050*	0.9672 *0.0026*	0.9228 *0.0024*	0.8960	0.9216	0.9280	0.9180

Ek 4.

Çizelge 5. 3 Grup durumu için çapraz doğru sınıflandırma performansı (Correct classification ratio for 3 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				Bir birimi hariç tutma			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Senaryo A	Önerilen Model	**0.8667** *0.0745*	**0.8400** *0.0365*	**0.7933** *0.0723*	**0.8233** *0.0548*	**0.7067**	**0.7893**	**0.7907**	0.8263
	DVS	0.7667 *0.1491*	0.8267 *0.0365*	0.7667 *0.0667*	0.8267 *0.0384*	0.5800	0.7160	0.7667	**0.8363**
	FLDF	0.4000 *0.1900*	0.2533 *0.1095*	0.3133 *0.0506*	0.3467 *0.0298*	0.3800	0.2853	0.2333	0.3137
	k -EYK	0.6667 *0.1179*	0.7867 *0.0989*	0.7400 *0.0454*	0.7933 *0.0435*	0,6300	0.6787	0.7440	0.7960
	NB	0.6330 *0.1394*	0.7333 *0.1054*	0.7933 *0.0494*	0.7500 *0.0656*	0.5367	0.6667	0.6913	0.7607
Senaryo B	Önerilen Model	**0.9233** *0.0419*	**0.9596** *0.0178*	**0.9672** *0.0124*	0.9819 *0.0059*	**0.8867**	**0.9387**	0.9560	0.9573
	DVS	0.7127 *0.0698*	0.9187 *0.0225*	0.9638 *0.0106*	**0.9847** *0.0065*	0.8033	0.9160	**0.9613**	**0.9693**
	FLDF	0.2207 *0.0727*	0.3059 *0.0391*	0.3923 *0.0177*	0.3387 *0.0204*	0.1800	0.3147	0.2787	0.3303
	k -EYK	0.7170 *0.0589*	0.8704 *0.0302*	0.9453 *0.0145*	0.9633 *0.0074*	0.7533	0.8827	0.9453	0.9000
	NB	0.6387 *0.0759*	0.8495 *0.0471*	0.8569 *0.0226*	0.9453 *0.0101*	0.7167	0.7973	0.8767	0.9280

ERKEN GÖRÜNÜM

Ek 5.

Çizelge 6. 4 Grup durumu için çapraz doğru sınıflandırma performansı (Correct classification ratio for 4 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				**Bir birimi hariç tutma**			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Senaryo A	Önerilen Model	**0.6000** *0.1046*	**0.6200** *0.0788*	**0.7000** *0.0771*	0.7175 *0.0288*	**0.7250**	**0.6700**	**0.6840**	**0.6760**
	DVS	0.4750 *0.1630*	0.5700 *0.1204*	0.6700 *0.1267*	**0.7325** *0.0420*	0.5300	0.5400	0.6460	0.6560
	FLDF	0.2750 *0.1630*	0.2300 *0.0975*	0.2500 *0.0694*	0.2600 *0.0285*	0.2150	0.1880	0.1710	0.2150
	k -EYK	0.5500 *0.1677*	0.5300 *0.0274*	0.7000 *0.0500*	0.6975 *0.0185*	0.6300	0.5660	0.6300	0.6340
	NB	0.4950 *0.2054*	0.5600 *0.1981*	0.6000 *0.1000*	0.6925 *0.0979*	0.5100	0.5020	0.5880	0.6140
Senaryo B	Önerilen Model	**0.8250** *0.1425*	0.8600 *0.0822*	0.9200 *0.0326*	0.9350 *0.0511*	**0.8038**	**0.8545**	0.8858	**0.9565**
	DVS	0.7000 *0.0685*	**0.8700** *0.1204*	**0.9500** *0.0177*	**0.9475** *0.0285*	0.6650	0.8430	**0.9377**	0.9555
	FLDF	0.3500 *0.1630*	0.3800 *0.0671*	0.2400 *0.0285*	0.2250 *0.0360*	0.2275	0.2030	0.2335	0.1895
	k -EYK	0.7250 *0.0559*	0.8500 *0.1541*	0.900 *0.0395*	0.9300 *0.0326*	0.6675	0.8140	0.9037	0.9375
	NB	0.6250 *0.1768*	0.7300 *0.0837*	0.8150 *0.0675*	0.7900 *0.0881*	0.5913	0.6870	0.7435	0.8455

Ek 6.

Çizelge 7. 5 Grup durumu için çapraz doğru sınıflandırma performansı (Correct classification ratio for 5 group cases)

		Çapraz Doğrulama Yöntemi							
		5-kat				**Bir birimi hariç tutma**			
	Birim Sayısı	**n=10**	**n=25**	**n=50**	**n=100**	**n=10**	**n=25**	**n=50**	**n=100**
Senaryo A	Önerilen Model	**0.6400** *0.1375*	**0.6861** *0.0730*	**0.6794** *0.0592*	0.6592 *0.0387*	**0.6960**	**0.6296**	**0.6836**	**0.6616**
	DVS	0.4720 *0.1449*	0.6107 *0.0826*	0.6390 *0.0601*	0.6661 *0.0404*	0.4840	0.5312	0.6236	0.6460
	FLDF	0.1310 *0.1695*	0.2549 *0.0728*	0.2186 *0.0446*	0.6082 *0.0314*	0.1560	0.1584	0.1704	0.1740
	k -EYK	0.6120 *0.1562*	0.5813 *0.0863*	0.6394 *0.0525*	0.6082 *0.0368*	0.5460	0.5432	0.5932	0.6192
	NB	0.2440 *0.1096*	0,5691 *0.0930*	0.5556 *0.0615*	0.5720 *0.0491*	0.4600	0.4704	0.5320	0.5730
Senaryo B	Önerilen Model	**0.8936** *0.0952*	**0.9338** *0.0484*	**0.9512** *0.0278*	**0.9684** *0.0161*	**0.9360**	**0.9696**	**0.9728**	**0.9748**
	DVS	0.6900 *0.1437*	0.8462 *0.0728*	0.9029 *0.0427*	0.9609 *0.0224*	0.6340	0.8272	0.9404	0.9686
	FLDF	0.2132 *0.1103*	0.1750 *0.0542*	0.2224 *0.0473*	0.2455 *0.0321*	0.1520	0.1560	0.1400	0.1870
	k -EYK	0.6356 *0.1413*	0.7797 *0.0796*	0.8766 *0.0413*	0.9308 *0.0253*	0.6260	0.8224	0.9112	0.9466
	NB	0.5756 *0.1588*	0.6040 *0.1087*	0.6858 *0.0873*	0.7444 *0.0568*	0.5220	0.6460	0.7248	0.7832

ERKEN GÖRÜNÜM