



Comparative classification performances of filter model feature selection algorithms in EEG based brain computer interface system

Cem Bulut^{1*}, Tuğçe Ballı², E. Fatih Yetkin²

¹Department of Computer Engineering, Istanbul University-Cerrahpasa, 34320, Istanbul, Türkiye

²Management Information Systems Department, Kadir Has University, 34083, Istanbul, Türkiye

Highlights:

- Classification of EEG recordings obtained from a BCI system
- Feature extraction with band power
- Impact of filter-based feature selection algorithms on classification performance

Keywords:

- BCI
- EEG
- Band Power
- Feature Selection
- Classification

Article Info:

Research Article
Received: 04.08.2021
Accepted: 10.12.2022

DOI:

10.17341/gazimmfd.978895

Correspondence:

Author: Cem Bulut
e-mail:
bulutcem22@gmail.com
phone: +90 538 592 7275

Graphical/Tabular Abstract

In this study, EEG signals obtained from a neuroprosthetic device designed for a brain computer interface system are used. These signals were recorded during resting state versus imagination of right-hand movement. Then, two feature sets are created for classification of EEG segments. The first set includes EEG band power features extracted from six main frequency bands (delta (1.0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), high-beta (25-30Hz) and gamma (30-50 Hz)). The second set contains ten frequency sub-bands (delta (1-4 Hz), theta (4-8 Hz), alpha1 (8-10 Hz), alpha2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), beta3 (18-25 Hz), gamma1 (30-35 Hz), gamma2 (35-40 Hz), gamma3 (40-50 Hz)).

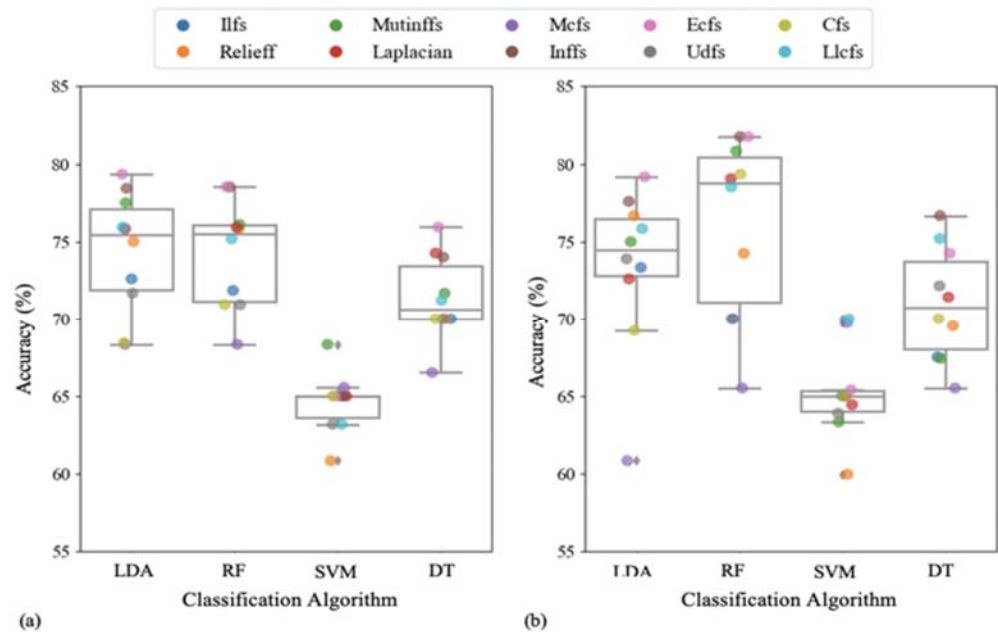


Figure A. Classification accuracies for (a) main frequency bands and (b) sub frequency bands

Purpose: The impact of feature selection algorithms on classification of EEG segments and the effect of EEG feature extraction from six main frequency bands versus ten frequency sub-bands are investigated.

Theory and Methods: We applied ten filter-based feature selection algorithms including Ilfs, Relief, Mutinffs, Laplacian, Mcfs, Inffs, Ecfs, Udfs, Cfs, Llcs along with four classification methods, namely: Linear Discriminant Analysis, Random Forests, Support Vector Machines and Decision Trees.

Results: The results indicate that feature selection methods leads to a higher classification accuracy and eigenvalue centrality (Ecfs) and infinite feature selection (Inffs) methods have consistently provided higher accuracy rates as compared to rest of the feature selection methods (Figure A).

Conclusion: It is shown that the optimal selection of feature extraction, feature selection and classification methods have a positive impact on the classification performances of EEG segments. The findings are quite promising and will shed light to future studies that will bring BCI systems to clinical practice.



Filtre modellenli öznelik seçim algoritmalarının EEG tabanlı beyin bilgisayar arayüzü sistemindeki karşılaştırmalı sınıflandırma performansları

Cem Bulut^{1*}, Tuğçe Ballı², E. Fatih Yetkin²

¹İstanbul Üniversitesi-Cerrahpaşa, Bilgisayar Mühendisliği Bölümü, 34320, İstanbul, Türkiye

²Kadir Has Üniversitesi, Yönetim Bilişim Sistemleri Bölümü, 34083, İstanbul, Türkiye

Ö N E Ç İ K A N L A R

- BBA sisteminden elde edilmiş EEG kayıtlarının sınıflandırılması
- Bant gücü yöntemi ile öznelik çıkarılması
- Filtre tabanlı öznelik seçim algoritmalarının sınıflandırma performansına etkisi

Makale Bilgileri

Araştırma Makalesi

Geliş: 04.08.2021

Kabul: 10.12.2022

DOI:

10.17341/gazimmfd.978895

Anahtar Kelimeler:

BBA,
EEG,
bant gücü,
öznelik seçme,
sınıflandırma

ÖZ

Beyin bilgisayar arayüzleri (BBA), beyin elektriksel aktivitelerini kontrol komutlarına çevirerek bilgisayar veya nöroprostatik kol gibi yardımcı teknolojilerin kullanılmasını sağlayan sistemlerdir. Bu çalışmada filtre tabanlı öznelik seçim yöntemlerinin farklı sınıflandırma algoritmaları ile birlikte kullanılmasının BBA sistemlerine getirebileceği kazanımlar araştırılmıştır. Bu çerçevede nöroprostatik bir cihazın kontrolü için tasarlanan BBA sisteminden elde edilmiş EEG kayıtları analiz edilmiştir. EEG kayıtlarının analizi için delta (1.0-4 Hz), teta (4-8 Hz), alfa (8-12 Hz), beta (12-25 Hz), yüksek-beta (25-30Hz) ve gama (30-50 Hz) frekans bantlarından ve delta (1-4 Hz), teta (4-8 Hz), alfa1 (8-10 Hz), alfa2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), beta3 (18-25 Hz), gama1 (30-35 Hz), gama2 (35-40 Hz), gama3 (40-50 Hz) alt frekans bantlarından bant gücü öznelikleri çıkarılmıştır. Elde edilen iki veri seti öznelik seçimi uygulamadan ve öznelik seçimi uygulayarak sınıflandırılmıştır. Çalışmada toplam 10 adet filtre tabanlı öznelik seçimi yöntemi ile birlikte, doğrusal ayırt eden analizi, rassal ormanlar, karar ağaçları ve destek vektör makinaları sınıflandırma algoritmaları kullanılmıştır. Çalışma sonucunda EEG kayıtlarının sınıflandırılması için öznelik seçme algoritmalarının uygulanmasının daha yüksek başarılı sonuçlar verdiği ve bu çalışmada ele alınan öznelik seçme yöntemlerinden, özdeğer merkezliyetine göre öznelik seçimi (Ecfs) ve sonsuz öznelik seçimi (Inffs) yöntemlerinin filtre tabanlı yaklaşımlar arasında en iyi sonuçları verdiği gözlenmiştir.

Comparative classification performances of filter model feature selection algorithms in EEG based brain computer interface system

H I G H L I G H T S

- Classification of EEG recordings obtained from a BCI system
- Feature extraction with band power
- Impact of filter-based feature selection algorithms on classification performance

Article Info

Research Article

Received: 04.08.2021

Accepted: 10.12.2022

DOI:

10.17341/gazimmfd.978895

Keywords:

BCI,
EEG,
band power,
feature selection,
classification

ABSTRACT

Brain-computer interface (BCI) systems enable individuals to use a computer or assistive technologies such as a neuroprosthetic arm by translating their brain electrical activity into control commands. In this study, the use of filter-based feature selection methods for design of BCI systems is investigated. EEG recordings obtained from a BCI system designed for the control of a neuroprosthetic device are analyzed. Two feature sets were created; the first set was band power features from six main frequency bands (delta (1.0-4 Hz), theta (4-8 Hz), alpha (8-12 Hz), beta (12-25 Hz), high-beta (25-30Hz) and gamma (30-50 Hz)) and the second set was band power features from ten frequency sub-bands (delta (1-4 Hz), theta (4-8 Hz), alpha1 (8-10 Hz), alpha2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), beta3 (18-25 Hz), gamma1 (30-35 Hz), gamma2 (35-40 Hz), gamma3 (40-50 Hz)). Ten filter-based feature selection methods are investigated along with linear discriminant analysis, random forests, decision tree and support vector machines algorithms. The results indicate that feature selection methods leads to a higher classification accuracy and eigen value centrality (Ecfs) and infinite feature selection (Inffs) methods have consistently provided higher accuracy rates as compared to rest of the feature selection methods.

*Sorumlu Yazar/Yazarlar / Corresponding Author/Authors : *bulutcem22@gmail.com, tugce.balli@khas.edu.tr, fatih.yetkin@khas.edu.tr /
Tel: +90 538 592 7275

1. Giriş (Introduction)

Beyin bilgisayar arayüzü (BBA) sistemleri, felçli bireylerin veya Amiyotrofik Lateral Sklerozis (ALS) hastalarının beyin elektriksel aktivitelerini kontrol komutlarına çevirerek bir bilgisayarı, elektromekanik bir kolu veya çeşitli nöroprotezler gibi yardımcı teknolojileri kullanmalarını ve yönlendirmelerini sağlayan sistemler olarak tanımlanabilir. BBA sistemleri bu bireylerin hayatlarına ışık tutabilmesi ve bu hastaların hayatlarını kolaylaştırılması açısından oldukça önemli bir teknolojidir [1].

BBA uygulamalarında beyin elektriksel aktivitelerini ölçmek için elektroensefalografi (EEG), fonksiyonel manyetik rezonans görüntüleme, yakın kızılötesi spektroskopisi, manyetoensefalografi gibi invazif olmayan ve elektrokortikografi, yerel alan potansiyelleri, tek hücre kayıtlamaları, çok elektrotlu diziler gibi invazif yöntemler kullanılmaktadır. İnvazif yöntemler, elektrotların ameliyat ile yerleştirilme gerekliliği ve zaman içerisinde sinyal kalitesinin düşme riski gibi dezavantajlara sahiptir. Bu nedenle invazif olmayan yöntemler daha sık kullanılmaktadır. İnvazif olmayan yöntemler arasından EEG; taşınabilir, kolay ulaşılabilir ve yüksek zamansal çözünürlüğe sahip olması sebebi ile sıklıkla tercih edilmektedir [2].

EEG tabanlı BBA sistemlerini tasarlamak için öncelikle deneysel paradigma belirlenmelidir. Bu çerçevede sıklıkla kullanılan yaklaşımlar P300, durağan durum uyarılmış potansiyelleri ve mental görev tabanlı sistemlerdir [2-5]. Deneysel paradigma belirlendikten sonra, EEG kayıtları deneysel prosedüre göre kaydedilerek sinyal işleme ve sınıflandırma modeli oluşturma aşamalarına geçilir. EEG kayıtlarının işlenmesinde uygulanan temel yöntemler; önişleme, öznitelik çıkarma, öznitelik seçme ve sınıflandırma yöntemleridir.

Önişleme, sinyallerin paradigmaya bağlı olarak EEG kayıtlarının ilgili frekans aralığında filtrelenmesi, enerji hattı kaynaklı gürültünün ve kas/göz hareketlerinden kaynaklı artefaktların giderilmesini içeren aşamadır. Öznitelik çıkarma, EEG kayıtlarının bant gücü, sinyal karmaşıklığı, fonksiyonel bağlantısallık, otoregresif model katsayıları gibi karakteristik özelliklerini çıkararak kayıtların daha kompakt bir şekilde ifade edildiği aşamadır. Öznitelik seçimi, ayırıcı olmayan öznitelikleri eleyerek sınıflandırıcı eğitim ve test zamanlarını kısaltmayı aynı zamanda sınıflandırma performanslarını artırmayı mümkün kılmaktadır. Son aşamada seçilmiş olan öznitelikler sınıflandırma algoritmalarının eğitim, test ve doğrulama aşamalarında kullanılarak sınıflandırma modeli oluşturulmaktadır. Bu aşamada kullanılacak olan yöntemler BBA sisteminin performansını belirleyen en önemli etkenlerden biridir [4].

Sadiq vd. [5] yapmış oldukları çalışmada öncelikle ampirik mod ayrışımı, ampirik dalgacık dönüşümü, varyasyonel mod ayrışımı yöntemleri ile EEG kayıtlarındaki her bir elektrot verisi 5 içsel mod fonksiyonuna dönüştürmüştür. Daha sonra elde edilen kayıtlardan zaman tabanlı, üst basamaklı istatistik tabanlı, enerji ve entropi tabanlı, frekans tabanlı ve doğrusal olmayan yöntemlerle öznitelikler çıkarılmıştır. Elde edilen öznitelik vektörlerine çok değişkenli (doğrusal regresyon, korelasyon tabanlı, çok kümelili öznitelik seçimi, denetimsiz ayırt edici öznitelik seçimi) ve tek değişkenli (Laplace skoru, Relief, karşılıklı bilgi -mutual information) öznitelik seçme yöntemleri uygulanarak motor görev tabanlı BBA sistemlerinde hangi sinyal ayrıştırma yöntemleri ve öznitelik seçme yöntemlerinin birlikte kullanıldığında sınıflandırma performansının arttıracağı incelenmiştir. Sınıflandırma için ileri beslemeli yapay sinir ağları ve art arda bağlanmış ileri beslemeli yapay sinir ağları kullanılmıştır. Sonuçlar, çok değişkenli öznitelik seçme yöntemlerinin tek değişkenli yöntemlere kıyasla daha yüksek performansla sahip olduğu ve varyasyonel mod ayrışımı, doğrusal regresyon ve art arda sıralı yapay

sinir ağlarının birlikte kullanıldığı zaman diğer yöntem kombinasyonlarına kıyasla daha yüksek sınıflandırma performansına sahip olduğunu göstermiştir. Gupta vd. [6] tarafından yapılmış benzer yaklaşıma sahip bir başka çalışmada ise mental görev tabanlı BBA sisteminin tasarımı için kaydedilmiş EEG verilerinin güç spektral yoğunluğu tabanlı özniteliklerle birlikte, tek ve çok değişkenli öznitelik seçim yöntemlerinin performansları incelenmiştir. Çok değişkenli öznitelik seçimi için Bhattacharya uzaklığı, dağılım matrislerinin oranı, minimum fazlalık maksimum alaka yöntemi; tek değişkenli öznitelik seçimi için Fisher ayırt eden oranı, korelasyon, karşılıklı bilgi ve Wilcoxon sıralama toplamı yöntemleri kullanılmıştır. Çalışma sonuçları, çok değişkenli ve filtre tabanlı öznitelik seçim yöntemlerinin, tek değişkenli yöntemlere kıyasla daha yüksek performansa sahip olduğunu göstermiştir.

Malan & Sharma [7], motor görev tabanlı BBA sisteminin tasarımı için toplanmış olan EEG kayıtlarından çift ağaçlı karmaşık dalgacık dönüşümü yöntemi ile öznitelik çıkarmış ve bu öznitelik vektörleri üzerinde komşuluk katsayısı analizi, genetik algoritmalar, temel bileşen analizi, Relief yöntemleri ile öznitelik seçimi uygulayarak destek vektör makinaları (DVM) yöntemi ile sınıflandırılmıştır. Sonuç olarak filtre tabanlı bir öznitelik seçim yöntemi olan komşuluk katsayısı analizi yönteminin öznitelik sayısını en aza indirmekle birlikte sınıflandırma performansını maksimize ettiği belirtilmiştir. Rezaei & Shalhaf [8] tarafından yapılmış bir çalışmada ise filtre tabanlı öznitelik seçme yöntemleri arasından; Relief, Fisher, lokal öğrenme tabanlı kümeleme ve Laplace yöntemleri, doğrusal ayırt eden analizi (DAA) ve DVM sınıflandırma algoritmaları ile birlikte motor görev tabanlı BBA tasarımı için toplanmış olan EEG kayıtlarından çıkarılan transfer entropi özniteliklerinin seçilmesi ve sınıflandırılmasında kullanılmıştır. Çalışma sonuçları Relief ve DVM yöntemlerinin en iyi sınıflandırma performansını verdiğini göstermiştir. Joadder vd. [9] yaptıkları çalışmada, filtre tabanlı bir öznitelik seçme yöntemi geliştirerek motor görev tabanlı EEG kayıtlarından çıkarılmış zaman, frekans, dalgacık dönüşümü, fraktal boyut tabanlı ve istatistiksel öznitelikleri çıkararak oluşturdukları öznitelik havuzuna uygulamışlardır. Boyutu indirgenmiş olan öznitelik vektörleri DAA, DVM, K-en yakın komşu ve karar ağaçları (KA) algoritmaları ile sınıflandırılmıştır. Sonuçlar, geliştirilen öznitelik seçimi yönteminin katlımcı bağımsız bir şekilde oluşturulmuş olan sınıflandırma modellerinin performansını arttırdığını aynı zamanda hesaplama zamanını indirdiğini göstermiştir.

Literatürde yapılan çalışmalar mental görev tabanlı BBA sistemlerinin tasarımında öznitelik seçim yöntemlerinin sınıflandırma performansını arttırdığına ilişkin sonuçlar sunmaktadır. Ancak hangi yöntemin bu amaç için en uygun olduğu halen açık bir problemdir. Bununla birlikte filtre tabanlı yöntemlerin diğer yöntemlere kıyasla daha iyi performans gösterebildikleri literatürde gözlenmiştir. Bu çalışma, farklı öznitelik seçim yöntemlerinin farklı sınıflandırma algoritmaları ile kullanılmasının BBA sistemlerine getirebileceği kazanımların analizi üzerinedir. Dolayısıyla, bu çalışmada yüksek spinal kord lezyonlu bir bireyin sağ üst eklemine takılan nöroprostatik bir cihazı kontrol etmesi için tasarlanan BBA sistemi üzerinden toplanan EEG kayıtlarının analiz edilerek; (1) delta (1.0-4 Hz), teta (4-8 Hz), alfa (8-12 Hz), beta (12-25 Hz), yüksek-beta (25-30Hz) ve gama (30-50 Hz) frekans bantlarından çıkarılan bant gücü öznitelik vektörleri ile delta (1-4 Hz), teta (4-8 Hz), alfa1 (8-10 Hz), alfa2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), beta3 (18-25 Hz), gama1 (30-35 Hz), gama2 (35-40 Hz), gama3 (40-50 Hz) alt frekans bantlarından çıkarılan bant gücü öznitelik vektörlerinin sınıflandırma performanslarının karşılaştırılması; (2) bant gücü yöntemi ile elde edilen öznitelik vektörlerinin öznitelik seçimi uygulayarak ve uygulamadan elde edilen sınıflandırma performanslarının

karşılaştırılması; (3) 2.3. bölümde açıklanmış olan 10 adet öznelik seçim yönteminin, DAA, DVM, rassal ormanlar (RO) ve karar ağacı (KA) sınıflandırma algoritmaları ile birlikte kullanıldığı zaman elde edilen sınıflandırma performanslarının karşılaştırmalı olarak incelenmesi ele alınmıştır.

2. Materyal ve Metot (Materials and Method)

Bu çalışmada uygulanan veri işleme sürecinin akış şeması Şekil 1’de sunulmuştur. EEG kayıtlarının analizi için öncelikle delta (1-4 Hz), teta (4-8 Hz), alfa (8-12 Hz), beta (12-25 Hz), yüksek-beta (25-30 Hz) ve gama (30-50 Hz) bantlarından çıkarılan bant gücü öznelik vektörleri ile delta (1-4 Hz), teta (4-8 Hz), alfa1 (8-10 Hz), alfa2 (10-12 Hz), beta1 (12-15 Hz), beta2 (15-18 Hz), beta3 (18-25 Hz), gama1 (30-35 Hz), gama2 (35-40 Hz), gama3 (40-50 Hz) alt frekans bantlarından çıkarılan bant gücü öznelik vektörleri oluşturulmuştur. Belirtilen frekans bantları “NeuroGuide” el kitabında tanımlanmış olan aralıklarına göre seçilerek [10], ikinci aşamada öznelik seçimi uygulanmadan sınıflandırma yapılmış ve son aşamada 10 adet öznelik seçim yöntemi uygulanarak; DAA, RO, DVM ve KA sınıflandırma algoritmaları ile elde edilen sonuçlar değerlendirilmiştir. Bu çalışmadaki tüm hesaplamalar MATLAB 2019a sürümü ile Intel(R) Core (TM) i5 işlemcili bir sistemde gerçekleştirilmiştir.

2.1. Veri Seti (Data Set)

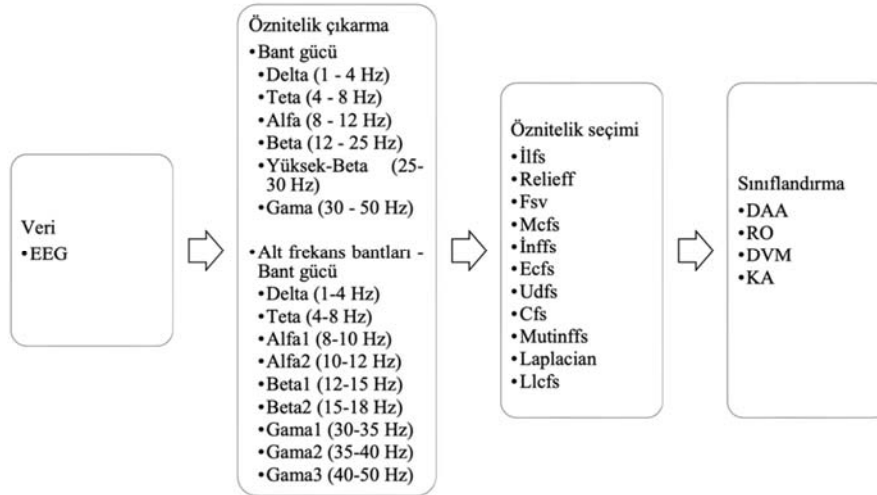
Bu çalışmada kullanılan EEG veri seti yüksek spinal kord lezyonlu bir bireyin sağ üst eklemine takılan nöroprostatik bir cihazı kontrol etmesi için tasarlanan BBA sistemi üzerinden toplanmıştır. EEG verileri uluslararası 10/20 sistemine uygun bir şekilde, F3, T3, C3, P3, and Cz elektrotlarını kullanarak kaydedilmiştir (Çalışmada kullanılan

veri setine <http://bnci-horizon-2020.eu/database/data-sets> adresinden erişilebilir.). Kayıt için referans olarak FCz elektrotu, toprak olarak AFz elektrotu kullanılmıştır (Acti-cap® and BrainAmp®, BrainProducts GmbH, Gilching, Germany). Veriler 200 Hz örnekleme frekansı ile 0.4-70 Hz aralığında bant geçiren süzgeç uygulanarak kaydedilmiştir. BBA sistemi iki temel mental görevden oluşmaktadır: “Sağ eli hareket ettirme düşüncesi” ve “dinlenme/hareketsiz kalma”. Kayıt sırasında katılımcıya rastgele kırmızı ve yeşil renkten oluşan uyaranlar gösterilmiştir.

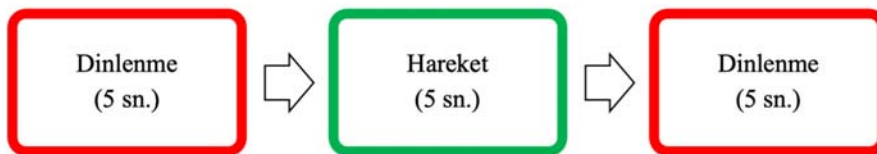
Katılımcıdan yeşil renkli uyaran ile karşılaştığında “sağ eli hareket ettirme”, kırmızı ile karşılaştığında “dinlenme/hareketsiz kalma” mental görevini yerine getirmesi istenmiştir (Şekil 2). Her iki görev katılımcıya 4-6 saniye aralıklarla 24 kez verilmiştir. Her uyaran katılımcıya 5 saniye boyunca gösterilmiştir. Veriler üç ayrı oturumda toplanmıştır [11, 12].

2.2. Öznelik Çıkarma (Feature Extraction)

EEG verileri 0.4-70 Hz. aralığında bant geçiren süzgeç uygulanarak kaydedilmiştir [11, 12]. Göz hareketi, kas hareketi gibi gürültüleri gidermek adına veriler gözden geçirilmiştir. Öznelik çıkarma işlemi için her bir elektrottan elde edilen EEG kayıtları 1 saniyelik üst üste gelmeyen segmentlere (epoklara) ayrılarak, her bir segmente bant gücü yöntemi uygulanmıştır [13]. Bant gücü öznelikleri, EEG kayıtlarının güç spektral yoğunluğu göz önüne alınarak, ilgili frekans aralıklarındaki ortalama gücün hesaplanmasıyla elde edilmiştir. 5 EEG elektrotundan 6 ana frekans aralığındaki (delta, teta, alfa, beta, yüksek-beta ve gama) öznelikler hesaplandığı zaman her 1 saniyelik EEG segmentinden 30 adet öznelik elde edilmiştir. Böylelikle, toplam 120 saniyelik EEG segmenti ve 30 öznelikten oluşan 120x30 boyutunda bir veri seti oluşturulmuştur. İkinci aşamada EEG kayıtları



Şekil 1. Veri İşleme Akış Şeması (Data Processing Flow Chart)



Şekil 2. Deneysel Tasarım (Experimental Design)

aynı şekilde 1 saniyelik üst üste gelmeyen parçalara ayrılarak alt frekans bantlarından (delta, teta, alfa1, alfa2, beta1, beta2, beta3, gama1, gama2, gama3), 10 adet bant gücü özneliği çıkarılmıştır; 5 EEG elektrotu ve 10 bant gücü olmak üzere 50 adet öznelilik ile 120x50 boyutunda bir veri seti elde edilmiştir. Oluşturulan öznelilik vektörlerinin analizleri 4 farklı durum altında yapılmıştır: (1) 120x30 boyutlu verinin tüm öznelilikleri kullanılarak sınıflandırılması, (2) 120x30 boyutundaki verinin öznelilik seçim yöntemleri ile işlendikten sonra sınıflandırılması, (3) Alt frekans bantları ile elde edilen 120x50 boyutundaki yeni verinin sınıflandırılması, (4) 120x50 boyutundaki verinin, öznelilik seçim yöntemleri ile işlendikten sonra sınıflandırılması. Sınıflandırma algoritması olarak DAA, RO, DVM ve KA algoritmaları kullanılmıştır. Sınıflandırma yapılırken 10 katmanlı çapraz doğrulama (cross validation) yöntemi kullanılmış olup sonuçların doğruluğunu arttırmak için her sınıflandırma işlemi 10 kez tekrar edilip doğruluk oranlarının ortalaması alınmıştır. Veri setinde yer alan üç oturma için yukarıdaki belirtilen işlemler ayrı ayrı yapılmıştır.

2.3. Öznelilik Seçim Yöntemleri (Feature Selection Methods)

Öznelilik seçimi, farklı gruplara ait veri setlerinde istatistiksel olarak en ayırıcı altkümenin seçilmesi işlemi olarak tanımlanabilir. Diğer bir deyişle öznelilik seçimi, kullanılan algoritmaya göre veri setindeki n adet öznelilik arasından en ayırıcı k adedi seçerek kullanma işlemidir [14]. Öznelilik seçimi yöntemleri sıralanan avantajları sunmaktadır [15]: 1) Öznelilik kümesinin boyutunu düşürerek analiz hızını artırır, 2) Gürültülü ve ilgisiz veriyi azaltarak veri kalitesini artırır, 3) Oluşturulan sınıflandırma modelinin başarısını artırır, 4) Kullanılan bellek miktarının azalmasını sağlar. Öznelilik seçme algoritmaları literatürde filtre, sarmal ve gömülü modelli olmak üzere üç temel başlıkta incelenmektedir. Filtre tabanlı öznelilik seçme algoritmaları, belirli bir istatistiksel metrik kullanarak (korelasyon, ki-kare vb.) özneliliklerin önem skorlarının hesaplanması ve ilgisiz özneliliklerin filtrelenmesi prensibiyle çalışır. Filtre tabanlı yöntemlerin, sarmal tabanlı yöntemlere göre avantajı daha hızlı ve kolay uygulanabilir olmaları ve aynı zamanda herhangi bir sınıflandırma algoritmasına bağlı çalışmadıkları için aşırı öğrenme (overfitting) problemleri olmamasıdır. Sarmal modelli öznelilik seçme algoritmaları ise en uygun öznelilik kümesini sınıflandırma algoritmasında test ederek bulmaya dayanmaktadır. Bu modelin başarılı olmasının nedeni algoritmanın çalışma prensibine göre seçilmiş farklı öznelilik alt kümelerini sınıflandırma algoritmasında da test ederek sınıflandırma performansını maksimize edecek alt kümeyi seçmesidir. Sarmal tabanlı yöntemlerin filtre tabanlı yöntemlere kıyasla avantajı, sınıflandırma performanslarının görece daha yüksek olmasıdır. Gömülü modelli öznelilik seçme algoritmaları sınıflandırma algoritmasında gömülü bulunan algoritmalarıdır, bu yöntemler herhangi bir ek öznelilik seçimi yapmaksızın sınıflandırma algoritmasının içerisinde daha ayırıcı özneliliklerin sınıflandırmaya olan katkısının daha yüksek olması prensibiyle çalışırlar [3]. Bu çalışmada filtre modelli öznelilik seçme yöntemleri üzerinde durulmuştur. KA ve RO sınıflandırma algoritmaları gömülü öznelilik seçme yöntemlerinden biri olsa da bu algoritmaların sınıflandırma performansları bu çerçevede değerlendirilmemiştir. [16]. Bu çalışmada kullanılan 10 adet öznelilik seçim yöntemi aşağıda özetlenmiştir:

Ilfs (Infinite Latent Feature Selection- Sonsuz Gizli Özellik Seçimi) yöntemi, öznelilikler arasındaki ilişkiyi PLSA (probabilistic latent semantic analysis - Olasılıksal Gizli Semantik Analiz) yöntemi ile modelleyerek en ayırıcı öznelilik altkümelerini seçmektedir [17]. Relieff yöntemi iki sınıflı öznelilik kümelerinde, özneliliklerin benzerlik/farklılıklarına bağlı olarak, önem katsayılarını hesaplayıp öznelilikleri sıralamaya ve seçmeye dayalı bir yöntemdir [18]. Mutinffs (Mutual Information Feature Selection- Karşılıklı Bilgi

Özellik Seçimi), öznelilik seçimini öznelilikler ve sınıf etiketleri arasındaki karşılıklı bilgiye dayanan maksimum istatistiksel bağımlılık kriterlerine göre yapmaktadır [19]. Laplacian (Laplacian Score for Feature Selection- Laplacian Puanı ile Özellik Seçimi) yönteminde her öznelilik için en yakın komşuluk grafiği oluşturularak bir Laplace skoru elde edilir. Laplacian algoritmasına göre, iki veri noktasının birbirine yakınlığı, o noktaların benzer özelliklere sahip olduğuna işaret eder [20]. Mcfs (Unsupervised Feature Selection for Multi-Cluster Data- Çoklu Küme Verileri için Denetimsiz Özellik Seçimi) yönteminde sınıf etiketi bilgileri olmadan, spektral analiz yöntemleri ile öznelilikler arasındaki korelasyonları ölçerek öznelilik seçimi yapılır [21]. Inffs (Infinite Feature Selection- Sonsuz Öznelilik Seçimi), özneliliklerin arasındaki ilişkiyi ağırlıklandırılmış şekilde gösteren bir grafik oluşturup farklı öznelilik altkümelerini değerlendiren ve sonuç olarak her bir özneliliğin önem katsayısını hesaplayarak seçim yapan bir yöntemdir [22]. Ecfs (Features Selection via Eigenvector Centrality- Özdeğer Merkeziliğine göre Öznelilik Seçimi) öznelilikler arasındaki ilişkiyi ifade eden grafiğin merkezilik ölçüsüne göre özneliliklerin önem sıralamasını yapmaya dayalı bir yöntemdir [23]. Udfs (Regularized Discriminative Feature Selection for Unsupervised Learning- Denetimsiz Öğrenme ile Düzenli Ayırt Edici Öznelilik Seçimi) yöntemi L2, 1-normlu optimizasyon problemini diklik kısıtı olmadan minimize etmeye çalışarak ayırt edici öznelilik altkümelerini toplu olarak seçen denetimsiz öğrenmeye dayalı bir yöntemdir [24]. Cfs (Correlation-based Feature Selection - Korelasyon Tabanlı Öznelilik Seçimi) yöntemi öznelilikleri sınıf etiketleriyle ve birbirleri arasındaki korelasyonlarına bağlı olarak seçen sezgisel bir fonksiyon yardımı ile yapmaktadır. Seçilen öznelilikler sınıf etiketiyle yüksek korelasyona, birbirleri arasında düşük korelasyona sahiptirler [25]. Llcf (Feature Selection and Kernel Learning for Local Learning-Based Clustering- Yerel Öğrenmeye Dayalı Kümeleme için Özellik Seçimi ve Çekirdek Öğrenmesi) algoritmasında, yerel öğrenme tabanlı kümeleme mantığı ile öznelilik seçimi veya çekirdek öğrenme vasıtasıyla uygun veri temsili elde edilmesi amaçlanmaktadır [26].

2.4. Sınıflandırma Algoritmaları (Classification Methods)

EEG kayıtlarının değerlendirilmesi aşamasında literatürde sıklıkla tercih edilen dört sınıflandırma algoritması kullanılmıştır: DAA (Doğrusal Ayırteden Analizi-Linear Discriminant Analysis), RO (Rastgele Orman-Random Forest), DVM (Destek Vektör Makineleri-Support Vector Machines) ve KA (Karar Ağacı- Decision Tree). DAA temel olarak klasik Fisher algoritmasının geliştirilmiş versiyonu olarak kabul edilebilir ve sınıf ayrımlarını en iyi şekilde gösteren öznelilik doğrusal birleşimlerinin hesaplanmasına dayanır [27]. KA yöntemi hem sınıflandırma hem de regresyon analizi için kullanılabilir parametrik olmayan denetimli bir öğrenme yöntemidir. Temel yaklaşımı, verinin özneliliklerinden yararlanarak hedef değişkeni tahmin edecek bir kurallar bütünü çıkartmaktır. Bu yaklaşım temel olarak, bir ağaç yapısı kurgulanarak bu karar ağacının her bir düğüm noktasında öznelilik üzerinde sorgulanmasına dayanır [28]. Öte yandan RO algoritması, veri kümesinin çeşitli alt kümelerinde rastgeleliğin avantajlarını kullanarak bir dizi karar ağacı sınıflandırıcısını bir arada kullanır. Böylelikle eşzamanlı yapılan çeşitli KA hesaplamalarının ortalamaları ile tahmin doğruluğu iyileştirilir ve veriden kaynaklanabilecek aşırı öğrenme problemleri kontrol altına alınır [29]. DVM ise temelde eğitim verileri için sınıflar arasında ayırıcı bir hiper-düzlem tespit edilmesine dayalıdır. Bu yöntemde, olası tüm hiper-düzlemler arasında en büyük marjine sahip olanı belirlemek için kurgulanmış bir eniyileme problemi çözümlenerek sınıflandırma işlemi gerçekleştirilir. Farklı öznelilik seçim yöntemleri ile daha yüksek doğruluklu sonuçlar verdiği literatürde gösterilmiştir [30]. Bu çalışmada literatürde EEG analizinde çoğunlukla kullanılmakta olan bu dört yöntem üzerinde durulmuş ve bu yöntemlerin öznelilik sayısına ve öznelilik seçme yöntemlerine karşı olan duyarlılıkları analiz edilmiştir.

Tablo 1. DAA ile sınıflandırma sonuçları (The clasification accuracies with LDA)

	Birinci oturum			İkinci oturum			Üçüncü oturum					
	Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.				
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
-	30	55,56	50	55,56	30	75,00	50	60,00	30	70,00	50	60,00
	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.		
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
İlfs	8	70,00	2	75,00	13	77,78	3	75,00	3	70,00	37	70,00
Relieff	2	70,00	11	75,00	3	75,00	14	75,00	2	80,00	7	80,00
Mutinfss	2	75,00	5	70,00	6	87,50	10	75,00	7	70,00	8	80,00
Laplacian	4	70,00	4	70,00	15	87,50	8	77,78	3	70,00	3	70,00
Mcfs	3	60,00	2	60,00	26	75,00	12	62,50	7	70,00	6	60,00
Inffs	6	77,78	25	75,00	7	87,50	12	77,78	5	70,00	9	80,00
Ecfs	6	77,78	15	70,00	6	87,50	7	87,50	6	72,73	7	80,00
Udfs	13	70,00	17	66,67	21	75,00	25	75,00	19	70,00	18	80,00
Cfs	22	66,67	3	60,00	11	75,00	27	75,00	18	63,64	14	72,73
Llcs	4	70,00	5	70,00	13	77,78	9	87,50	2	80,00	3	70,00

(Ösy: Öznitelik seçim yöntemi, Bir.dur.: Birinci durum, İki.dur.: İkinci durum, Üç.dur.: Üçüncü durum, Dör.dur.: Dördüncü durum, Ö.say.: Öznitelik sayısı, Doğ.: Doğruluk)

Tablo 2. RO ile Sınıflandırma Sonuçları (The clasification accuracies with RF)

	Birinci oturum			İkinci oturum			Üçüncü oturum					
	Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.				
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
-	30	77,78	50	75,00	30	75,00	50	77,78	30	60,00	50	60,00
	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.		
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
İlfs	2	77,78	14	75,00	12	77,78	8	75,00	7	60,00	29	60,00
Relieff	8	80,00	23	75,00	12	87,50	3	75,00	3	60,00	3	72,73
Mutinfss	5	77,78	15	75,00	12	77,78	10	87,50	4	72,73	14	80,00
Laplacian	5	80,00	5	76,92	3	75,00	5	87,50	3	72,73	3	72,73
Mcfs	26	70,00	9	61,54	26	75,00	51	75,00	3	60,00	11	60,00
Inffs	6	77,78	15	77,78	6	77,78	8	87,50	5	80,00	4	80,00
Ecfs	5	77,78	4	77,78	5	77,78	11	87,50	5	80,00	4	80,00
Udfs	15	77,78	21	75,00	21	75,00	23	75,00	13	60,00	16	60,00
Cfs	10	77,78	22	77,78	5	75,00	44	87,50	3	60,00	9	72,73
Llcs	7	75,00	3	77,78	4	77,78	4	77,78	4	72,73	3	80,00

(Ösy: Öznitelik seçim yöntemi, Bir.dur.: Birinci durum, İki.dur.: İkinci durum, Üç.dur.: Üçüncü durum, Dör.dur.: Dördüncü durum, Ö.say.: Öznitelik sayısı, Doğ.: Doğruluk)

Tablo 3. DVM ile Sınıflandırma Sonuçları (The clasification accuracies with SVM)

	Birinci oturum			İkinci oturum			Üçüncü oturum					
	Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.				
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
-	30	60,00	50	66,67	30	75,00	50	75,00	30	60,00	50	60,00
	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.	İki.dur.	Dör.dur.		
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
İlfs	17	60,00	2	58,33	16	75,00	28	50,00	19	60,00	22	60,00
Relieff	8	60,00	12	58,33	3	62,50	36	61,54	12	60,00	12	60,00
Mutinfss	16	60,00	48	55,00	12	75,00	22	75,00	14	70,00	9	60,00
Laplacian	23	60,00	23	58,33	8	75,00	12	75,00	3	60,00	17	60,00
Mcfs	6	60,00	17	69,23	26	66,67	51	70,00	4	70,00	10	70,00
Inffs	11	60,00	44	60,00	14	75,00	14	75,00	3	60,00	16	60,00
Ecfs	5	60,00	50	66,67	11	75,00	13	75,00	12	60,00	12	54,55
Udfs	17	60,00	46	66,67	19	75,00	24	75,00	1	54,55	3	50,00
Cfs	24	60,00	13	70,00	18	75,00	46	75,00	3	60,00	36	60,00
Llcs	5	60,00	12	75,00	8	75,00	8	75,00	19	54,55	16	60,00

(Ösy: Öznitelik seçim yöntemi, Bir.dur.: Birinci durum, İki.dur.: İkinci durum, Üç.dur.: Üçüncü durum, Dör.dur.: Dördüncü durum, Ö.say.: Öznitelik sayısı, Doğ.: Doğruluk)

3. Deneysel Sonuçlar (Experimental Results)

Bu çalışmada seçilmiş olan bir EEG verisi özelinde; öznitelik seçimi algoritmalarının farklı sınıflandırma algoritmaları ile kullanılmasının doğruluk performansını üzerindeki etkileri incelenmiştir. EEG verilerinin çok boyutluluğu göz önünde bulundurulduğunda öznitelik seçme yöntemlerinin kullanılması önem taşımaktadır. Bu nedenle,

öznitelik seçme yöntemleri hakkında uygun bir karşılaştırma yapabilmek amacıyla 4 farklı sınıflandırma yöntemi üzerinde 10 farklı öznitelik seçme yöntemi uygulanmış ve sonuçlar karşılaştırılmıştır. Yapılan çalışma sonucunda elde edilen her bir sınıflandırma algoritmasının farklı öznitelik seçim yöntemleri ile farklı oturum ve durumlarda verdiği başarımlara ilişkin detaylı sonuçlar Tablo 1, Tablo 2, Tablo 3 ve Tablo 4'te sunulmuştur. Çalışma sonucunda tüm

durumlardaki her bir oturum için elde edilen en yüksek doğruluk oranları Tablo 5, Tablo 6, Tablo 7 ve Tablo 8’de verilmektedir. Bu sonuçlara göre her bir oturum için birinci duruma kıyasla ikinci durumun; üçüncü duruma kıyasla dördüncü durumun sınıflandırma algoritmalarının doğruluk oranlarını arttırdığı ve buradan yola çıkarak

öznitelik seçme algoritmalarının EEG kayıtlarının sınıflandırılması probleminde başarımı arttırdığı gözlenmiştir. Öznitelik seçim yöntemleri kullanılan ikinci ve dördüncü durumlarda her bir oturum için en yüksek doğruluk oranlarının elde edildiği öznitelikler Tablo 9’da gösterilmiştir. Sonuçlar çoğunlukla delta bant gücünün

Tablo 4. KA ile Sınıflandırma Sonuçları (The classification accuracies with DT)

	Birinci oturum			İkinci oturum			Üçüncü oturum					
	Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.		Bir.dur.	Üç.dur.				
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)		
-	30	70,00	50	66,67	30	75,00	50	75,00	30	60,00	50	60,00
	İki.dur.		Dör.dur.		İki.dur.		Dör.dur.		İki.dur.		Dör.dur.	
Ösy.	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)	Ö.say.	Doğ.(%)
Ilfs	3	75,00	16	69,23	2	75,00	2	66,67	2	60,00	29	66,67
Relieff	5	75,00	2	69,23	2	75,00	1	66,67	5	60,00	9	72,73
Mutinffs	4	80,00	1	63,64	2	75,00	2	75,00	6	60,00	3	63,64
Laplacian	5	77,78	4	69,23	11	75,00	1	75,00	4	70,00	4	70,00
Mcfs	26	75,00	8	61,54	26	70,00	9	75,00	3	54,55	8	60,00
Inffs	3	75,00	2	75,00	2	75,00	2	75,00	5	72,00	4	80,00
Ecfs	5	77,78	3	75,00	2	75,00	7	75,00	4	75,00	4	72,73
Udfs	25	75,00	24	77,78	19	75,00	4	75,00	17	60,00	20	63,64
Cfs	17	75,00	31	66,67	10	75,00	11	75,00	7	60,00	2	60,00
Llcfs	4	75,00	3	77,78	1	75,00	5	77,78	4	63,64	2	70,00

(Ösy: Öznitelik seçim yöntemi, Bir.dur.: Birinci durum, İki.dur.: İkinci durum, Üç.dur.: Üçüncü durum, Dör.dur.: Dördüncü durum, Ö.say.: Öznitelik sayısı, Doğ.: Doğruluk)

Tablo 5. Birinci durumda en yüksek doğruluk oranları (Highest accuracy rates in the first case)

İşlem	Birinci oturum	İkinci oturum	Üçüncü oturum
Sınıflandırma algoritması	RO	RO	DAA
Öznitelik seçim yöntemi	-	-	-
Seçilen öznitelik sayısı	30	30	30
Doğruluk (%)	77,78	75,00	70,00
Duyarlılık (%)	75,00	75,00	80,00
Özgüllük (%)	80,00	75,00	60,00
F1 skoru (%)	75,00	75,00	72,73
Kesinlik (%)	75,00	75,00	66,67
Duyarlılık (%)	75,00	75,00	80,00

Tablo 6. İkinci durumda en yüksek doğruluk oranları (Highest accuracy rates in the second case)

İşlem	Birinci oturum	İkinci oturum	Üçüncü oturum
Sınıflandırma algoritması	KA	DAA	DAA
Öznitelik seçim yöntemi	Mutinffs	Mutinffs	Relieff
Seçilen öznitelik sayısı	4/30	6/30	2/30
Doğruluk (%)	80,00	87,50	80,00
Duyarlılık (%)	80,00	75,00	80,00
Özgüllük (%)	80,00	97,56	80,00
F1 skoru (%)	80,00	84,51	80,00
Kesinlik (%)	80,00	96,77	80,00
Duyarlılık (%)	80,00	75,00	80,00

Tablo 7. Üçüncü durumda en yüksek doğruluk oranları (Highest accuracy rates in the third case)

İşlem	Birinci oturum	İkinci oturum	Üçüncü oturum
Sınıflandırma algoritması	RO	RO	RO
Öznitelik seçim yöntemi	-	-	-
Seçilen öznitelik sayısı	50	50	50
Doğruluk (%)	75,00	77,78	60,00
Duyarlılık (%)	75,00	75,00	60,00
Özgüllük (%)	75,00	80,00	60,00
F1 skoru (%)	75,00	75,00	60,00
Kesinlik (%)	75,00	75,00	60,00
Duyarlılık (%)	75,00	75,00	60,00

seçildiğini göstermiştir. Oturumların ortalama sınıflandırma performansı ile tüm durumlarda elde edilen doğruluk dağılımları, sınıflandırma algoritması bazında Şekil 3'te detaylandırılmıştır. Bu şekiller incelendiğinde, öznelik seçiminin DAA, KA ve RO sınıflandırma algoritmalarının doğruluk oranlarına olumlu bir etki gösterdiği görülmüştür. Bununla birlikte ikinci ve dördüncü durumlarda Ecfs ve Inffs öznelik seçim yöntemlerinin DAA, KA ve RO algoritmaları için tutarlı bir şekilde en yüksek doğruluk performansını veren öznelikleri seçtikleri görülmüştür. Öznelik seçim yöntemlerinin değerlendirilmesini yapabilmek adına her bir sınıflandırma algoritması ve öznelik seçim yönteminin tüm oturumlardaki ortalamasını hesaplayarak elde edilen en yüksek doğruluk oranları Tablo 10'da gösterilmiştir. Tabloya göre en yüksek ortalama doğruluk oranları RO ve DAA algoritmaları ile elde edilmiştir. Öte yandan öznelik seçme uygulanan ikinci ve dördüncü durumlarda Ecfs ve Inffs yöntemlerinin diğer yöntemlere göre doğruluğu daha çok arttırdığı gözlemlenmiştir. Tablo 10 ve Şekil 3'ten çıkartılabilecek bir diğer sonuç ise öznelik sayısının artırılmasının (bant gücü hesaplanmasında alt frekans bantlarının

kullanılması ile) tüm öznelik seçimi ve sınıflandırma yöntemleri için doğruluğu anlamlı bir şekilde arttırmadığı olmuştur. Veri setinde yer alan üç oturumun ortalamaları alınarak sınıflandırma algoritmasına göre tek bir sonuca indirildikten sonra öznelik seçim yöntemlerinin kullanıldığı ikinci (30 özneliğe öznelik seçim yöntemlerinin uygulanması) ve dördüncü (50 özneliğe öznelik seçim yöntemlerinin uygulanması) durumlar için Şekil 4 oluşturulmuştur. İkinci durumda en yüksek doğruluklar; DAA (%79,34), RO (%78,52) ve KA (%75,93) sınıflandırma algoritmalarında Ecfs yöntemi ile elde edilmiştir. DVM (%68,33) algoritmasında ise Mutinffs ile elde edilmiştir. Dördüncü durumda ise en yüksek performansa sahip sınıflandırma algoritması ve öznelik seçim yöntemi kombinasyonları; RO ve Inffs-Ecfs (%81,76), DAA ve Ecfs (%79,17), KA ve Inffs (%76,67), DVM ve Mcfs (%69,74), olarak gözlemlenmiştir. Bulgular DVM algoritmasının ortalama doğruluk oranlarının RO, DAA ve KA algoritmalarına kıyasla daha düşük olduğunu ve öznelik seçim yöntemlerinin DVM algoritmasının, doğruluk oranlarına tutarlı bir katkı sağlamadığını göstermiştir.

Tablo 8. Dördüncü durumda en yüksek doğruluk oranları (Highest accuracy rates in the fourth case)

İşlem	Birinci oturum	İkinci oturum	Üçüncü oturum
Sınıflandırma algoritması	RO	DAA	RO
Öznelik seçim yöntemi	Llcf	Llcf	Llcf
Seçilen öznelik sayısı	3/50	9/50	3/50
Doğruluk (%)	77,78	87,50	80,00
Duyarlılık (%)	80,00	75,00	80,00
Özgüllük (%)	75,00	100,00	80,00
F1 skoru (%)	80,00	85,71	80,00
Kesinlik (%)	80,00	100	80,00
Duyarlılık (%)	80,00	75,00	80,00

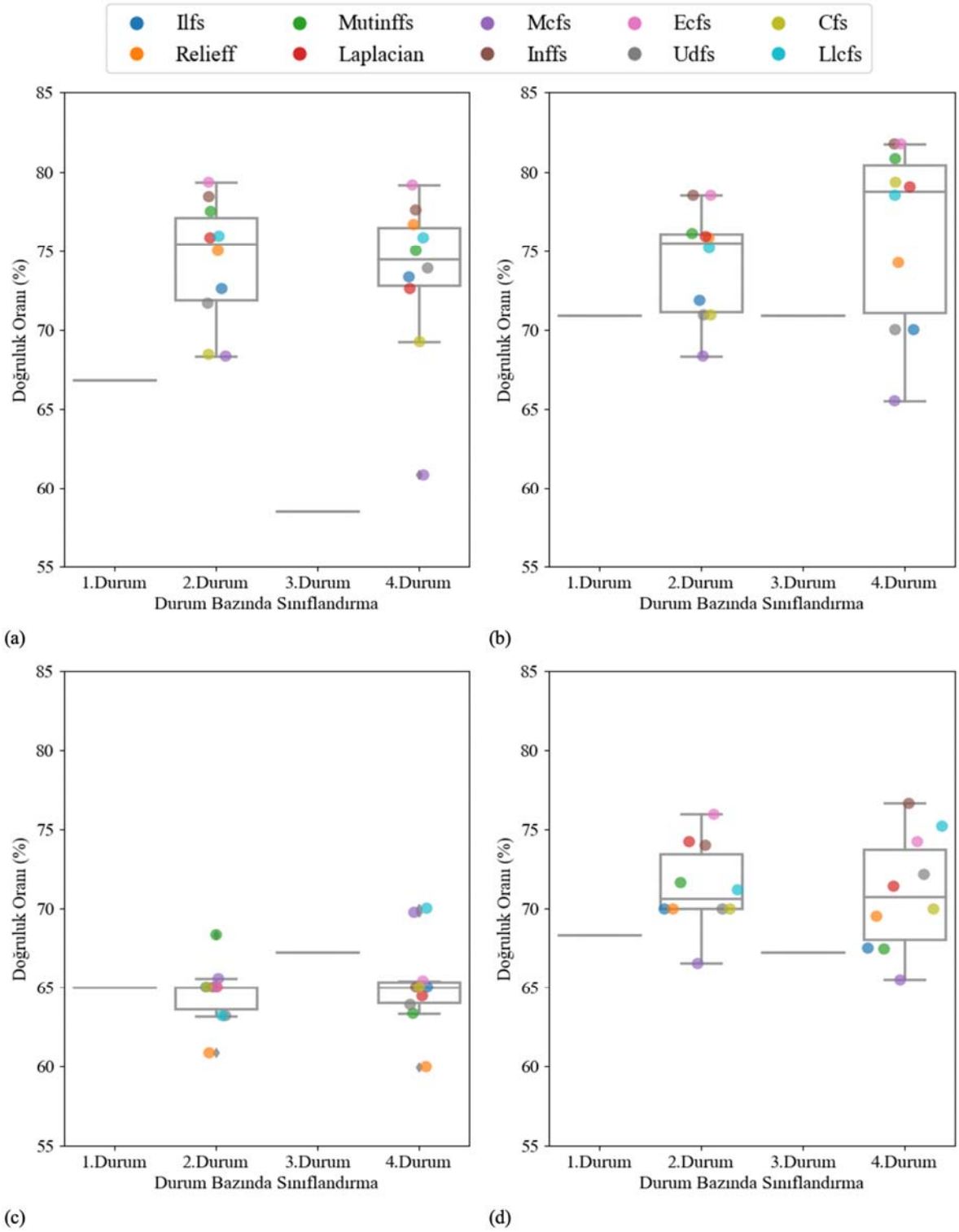
Tablo 9. İkinci ve dördüncü durumlarda her bir oturum için en yüksek doğruluk elde edilen sınıflandırma algoritması, öznelik seçim yöntemi ve seçilen öznelikler

(The classification algorithm, feature selection methods, and selected features with the highest classification accuracies for the second and fourth cases)

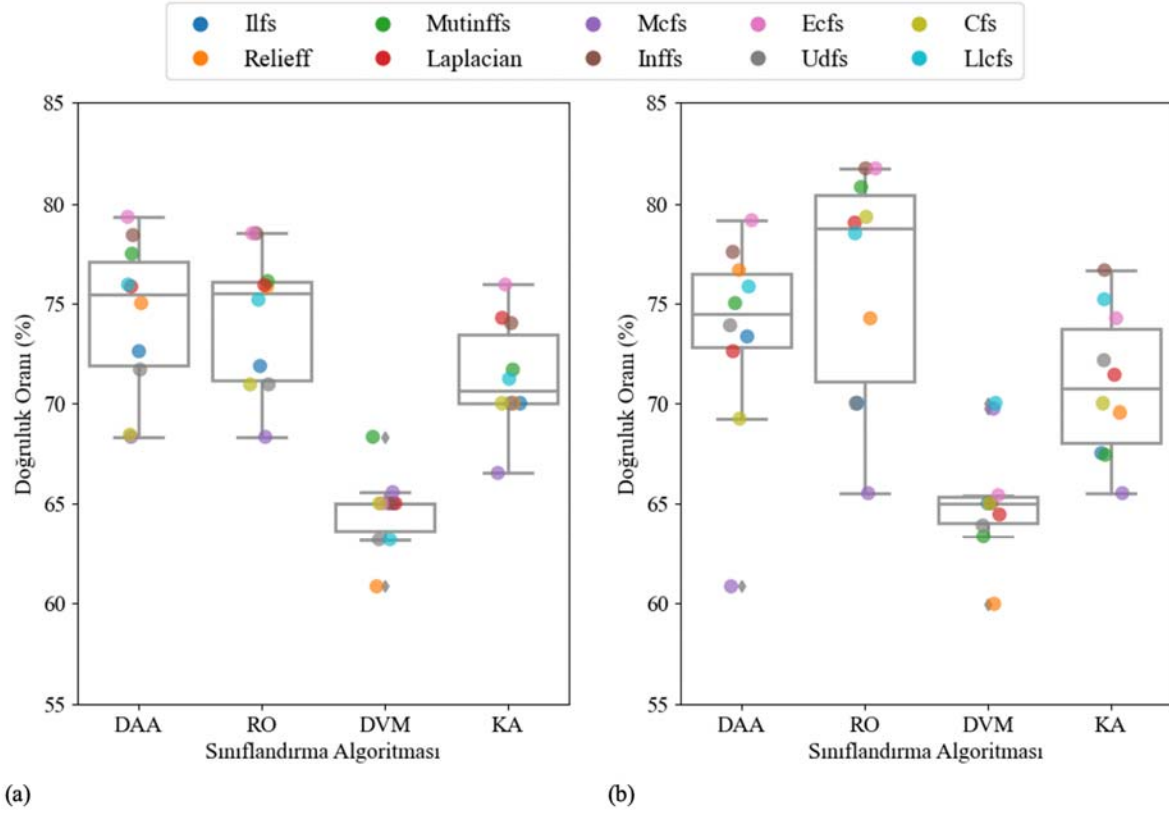
İşlem	İkinci durum			Dördüncü durum		
	Birinci oturum	İkinci oturum	Üçüncü oturum	Birinci oturum	İkinci oturum	Üçüncü oturum
Sınıflandırma algoritması	KA	DAA	DAA	RO	DAA	RO
Öznelik seçim yöntemi	Mutinffs	Mutinffs	Relieff	Llcf	Llcf	Llcf
	delta_Cz	delta_T3	delta_Cz	delta_Cz	delta_Cz	delta_Cz
	theta_F3	delta_Cz	delta_T3	delta_P3	delta_T3	delta_P3
	gamma_C3	delta_P3		delta_F3	delta_C3	delta_F3
	alpha_P3	delta_F3			delta_F3	
Seçilen öznelikler (frekansBandı_elektrot)		gamma_Cz			delta_P3	
		yüksek-beta_C3			gama1_T3	
					teta_P3	
					teta_C3	
					teta_F3	

Tablo 10. Oturumların birleştirilmesi sonucunda en yüksek doğruluk oranları (Highest accuracy rates with combined sessions)

Durumlar	Sınıflandırma algoritması	Öznelik seçim yöntemi	Doğruluk (%)
Birinci durum	RO	-	70,92
İkinci durum	DAA	Ecfs	79,34
Üçüncü durum	RO	-	70,92
Dördüncü durum	RO	Inffs / Ecfs	81,76



Şekil 3. Öznitelik seçim yöntemlerinin (a) DDA (b) RO (c) DVM ve (d) KA algoritmalarında ve tüm durumlarda doğruluk oranları (Accuracy rates of feature selection methods with (a) LDA (b) RF (c) SVM and (d) DT algorithms in all cases)



Şekil 4. (a) İkinci ve (b) dördüncü durumlarda öznelik seçim yöntemlerinin doğruluk oranları
(The accuracy rates of the feature selection methods for the (a) second and (b) fourth cases.)

4. Sonuçlar (Conclusions)

Bu çalışma kapsamında ele alınan EEG veri seti üç oturumdan oluşmaktadır. Her oturumda yer alan EEG kayıtları, dört aşamalı bir analize tabii tutulmuş ve her aşama için sınıflandırma performansları incelenmiştir. İlk aşamada EEG kaydı üzerinde delta, teta, alfa, beta, yüksek-beta, ve gama bantları için bant güçleri hesaplanıp, öznelikler çıkarılmış ve öznelik vektörleri oluşturulmuştur. Daha sonra bu öznelik vektörleri, herhangi bir öznelik seçim yöntemi uygulanmadan RA, DAA, DVM ve KA algoritmaları ile sınıflandırılmıştır. İkinci aşamada, birinci adımda elde edilen 30 öznelikten oluşan veri setine 10 farklı öznelik seçim yöntemi (Ilfs, Relief, Mutinffs, Laplacian, Mcfs, Inffs, Ecfs, Udfs, Cfs, Llcfs) uygulanmış ve elde edilen sınıflandırma sonuçları incelenmiştir. Üçüncü aşamada, EEG kayıtları alt frekans bantlarında filtrelenerek (delta, teta, alfa1, alfa2, beta1, beta2, beta3, gama1, gama2, gama3) bant gücü hesaplanmış, sonuç olarak 50 öznelik elde edilmiş ve sınıflandırılmıştır. Dördüncü ve son aşamada ise, üçüncü adımda elde edilen 50 öznelikten oluşan veri setine ikinci aşamada kullanılan öznelik seçim yöntemleri uygulanarak elde edilen sınıflandırma sonuçları incelenmiştir.

Öznelik seçim yöntemlerinin genel performansını değerlendirmek adına her bir sınıflandırma algoritması, öznelik seçim yöntemi ve durumun üç oturumdaki ortalamaları hesaplanmıştır. Tüm durumlar göz önüne alındığında en yüksek performans %81,76 doğruluk oranı ile Inffs ve Ecfs öznelik seçim yöntemlerinin, RO sınıflandırma algoritmasıyla birlikte kullanıldığı dördüncü durumda elde edilmiştir. İkinci ve dördüncü durumlar ayrı ayrı incelendiğinde de hem oturum bazında, hem de oturumların ortalama doğruluk oranları hesaplandığında Ecfs ve Inffs öznelik seçim yöntemlerinin DAA, KA ve RO algoritmaları için tutarlı bir şekilde en yüksek doğruluk

performansını veren öznelikleri seçtikleri görülmüştür. Bunlarla birlikte her bir durum ve oturum için seçilen öznelikler incelendiğinde ağırlıklı olarak delta bant gücünün seçildiği görülmüştür. Bu sonuçlar motor görev tabanlı EEG kayıtlarının sınıflandırılmasında delta bant gücünün ayırıcılığını gösteren çalışmalarla tutarlı bulunmuştur [31].

Dolayısıyla bu çalışma çerçevesinde elde edilen üç önemli bulgu söz konusudur. İlk olarak, Şekil 3 ile gösterildiği gibi öznelik seçme yöntemlerinin tümünün sınıflandırıcı performansı üzerinde olumlu etkisi vardır. Bu durum EEG kayıtlarının ölçüm teknikleri gereği içerdikleri gürültünün fazla oluşuna ve aynı zamanda farklı frekans bantlarının ve elektrotların yapılan mental göreve bağlı olarak sınıflandırmaya katkısının olmayışına bağlanabilir. Kullanılan öznelik seçme yöntemleri yardımıyla sınıflandırma performansını düşüren özneliklerin filtrelenmesi doğal olarak sınıflandırma başarımını arttırmaktadır. İkinci önemli gözlem ise Şekil 4 ile verildiği üzere bu çalışmada kullanılan EEG kayıtları için, öznelik sayısının artırılmasının sınıflandırıcı performansı üzerinde belirgin bir etki yapmamış olmasıdır. Bu durum çalışmada kullanılan mental görev tabanlı EEG kayıtlarının sınıflandırılmasında delta bant gücünün ağırlıklı olarak seçilmesi ve alt frekans bantlarına ayrılmış olan alfa, beta, gama bantlarının ayırıcılığının düşük olmasına bağlanabilir, öte yandan bu bulgular öznelik seçme mekanizmalarının mutlaka göz önüne alınması gerektiğini göstermektedir. Üçüncü ve son gözlem olarak ise Ecfs ve Inffs öznelik seçme algoritmalarının diğer yöntemlere göre daha yüksek başarımları gösterdikleri söylenebilir. Özetle, EEG kayıtlarının analizi için öznelik çıkartma mekanizmalarından bağımsız olarak bir ön işleyici olarak öznelik seçme algoritmalarının uygulanmasının, EEG kayıtlarının analizinde çok daha yüksek başarımlı sonuçlar verdiği ve bu çalışmada ele alınan öznelik seçme yöntemleri arasında da Ecfs

ve Inffs yöntemlerinin filtre tabanlı yaklaşımlar arasında en iyi sonuçları verdiği gözlenmiştir. Bunlarla birlikte BBA sistemlerinin tasarlama ve test aşamalarında yapılan çalışmaların çoğunluğu sağlıklı bireylerin katılımıyla yapılmaktadır. Kullanılan veri setinin yüksek spinal kord lezyonlu bir bireyin katılımıyla oluşturulmuş olması, BBA sistemlerinin günlük hayatta kullanımı için uygulanabilirliğini göstermekle birlikte veri setinde sadece bir katılımcının olması çalışma sonuçlarının genelleştirilebilirliğini kısıtlamaktadır.

Bu çalışmada, nöroprostatik bir cihazın kontrol edilmesi için tasarlanan sistem ele alınmıştır. Başlıca amaç, modelin performans ve doğruluğunu iyileştirerek sınıflandırma maliyetini azaltmaktır. Ayrıca her bir öznelik seçim yönteminin sınıflandırma performansı da değerlendirilmiştir. Sonuçlar, sınıflandırma performanslarının öznelik seçim yöntemleri ile birlikte oldukça umut verici olduğunu, klinik uygulamada kullanılabilirliğini ve gelecekteki çalışmalara ışık tutabileceğini göstermiştir.

Kaynaklar (References)

- Blankertz B., Dornhege G., Krauledat M., Müller K.R., Kunzmann V., Losch F., Curio G., The Berlin Brain-Computer Interface: EEG-based communication without subject training, *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 14 (2), 147-152, 2006.
- Abiri R., Borhani S., Sellers E.W., Jiang Y., Zhao X., A comprehensive review of EEG-based brain-computer interface paradigms, *Journal of neural engineering*, 16 (1), 011001, 2019.
- Lotte F., Bougrain L., Cichocki A., Clerc M., Congedo M., Rakotomamonjy A., Yger F., A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update, *Journal of neural engineering*, 15 (3), 031005, 2018.
- Zhang W., Tan C., Sun F., Wu H., Zhang B., A review of EEG-based brain-computer interface systems design, *Brain Science Advances*, 4 (2), 156-167, 2018.
- Sadiq M. T., Y.X., Yuan Z., Aziz M. Z., Siuly S., Ding W., Toward the Development of Versatile Brain-Computer Interfaces, *IEEE Transactions on Artificial Intelligence*, 2 (4), 314-328, 2021.
- Gupta A., Agrawal R. K., Kirar J. S., Andreu-Perez J., Ding W.-P., Lin C.-T., Prasad M., On the Utility of Power Spectral Techniques With Feature Selection Techniques for Effective Mental Task Classification in Noninvasive BCI, *IEEE Transactions on Systems, Man and Cybernetics: Systems*, 51 (5), 3080-3092, 2021.
- Malan N. S., Sharma S., Feature selection using regularized neighbourhood component analysis to enhance the classification performance of motor imagery signals, *Computers in biology medicine*, 107, 118-126, 2019.
- Rezaei, E., Shalhaf, A., Classification of Right/Left Hand Motor Imagery by Effective Connectivity Based on Transfer Entropy in EEG Signal, *Basic and Clinical Neuroscience (BCN)*, 2021.
- Joadder A.M., Myszewski J., Rahman M.H., Wang I., A performance based feature selection technique for subject independent MI based BCI, *Health information science systems*, 7 (1), 1-10, 2019.
- Applied Neuroscience. NeuroGuide Help Manual. Neuro Guide. https://www.appliedneuroscience.com/PDFs/NeuroGuide_Manual.pdf, Erişim tarihi Ekim 02, 2021.
- Soekadar S., Neuroprosthetic control of an EEG-EOG BNCI system by a paralyzed patient with high spinal cord injury, <https://lampx.tugraz.at/~bci/database/002-2015/description.pdf>, Erişim tarihi Şubat 02, 2021.
- Soekadar S., W.M., Vitiello N., Birbaumer N., An EEG/EOG-based hybrid brain-neural computer interaction (BNCI) system to control an exoskeleton for the paralyzed hand, *Biomedical Engineering / Biomedizinische Technik*, 60 (3), 199-205, 2015.
- Wahid M.F., Tafreshi R., Recognition of Upper-limb Movement Using Electroencephalogram Signals with Deep Learning, 2020 IEEE 5th Middle East and Africa Conference on Biomedical Engineering (MECBME), Amman-Ürdün, 27-29 Ekim, 2020.
- Forman G., An extensive empirical study of feature selection metrics for text classification, *Journal of machine learning research*, 3, 1289-1305, 2003.
- Ladha L., Deepa T., Feature selection methods and algorithms, *International journal on computer science and engineering (IJCSSE)*, 3 (5), 1787-1797, 2011.
- Weinmann M., Jutzi B., Hinz S., Mallet C., Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers, *ISPRS Journal of Photogrammetry and Remote Sensing*, 105, 286-304, 2015.
- Roffo G., Melzi S., Features selection via eigenvector centrality, *Proceedings of new frontiers in mining complex patterns (NFMCP 2016)*, 2016.
- Kononenko I., Estimating Attributes: Analysis and Extensions of RELIEF, *European conference on machine learning (ECML-94)*, İtalya, 171-182, 6-8 Nisan, 1994.
- Roffo G., Melzi S., Castellani U., Vinciarelli A., Infinite Latent Feature Selection: A Probabilistic Latent Graph-Based Ranking Approach, *Proceedings of the IEEE International Conference on Computer Vision, Venedik-İtalya*, 1398-1406, 25-27 Ekim, 2017.
- He X., Cai D., Niyogi P., Laplacian score for feature selection, *Advances in neural information processing systems*, 18, 507-514, 2005.
- Cai D., Zhang C., He X., Unsupervised feature selection for multi-cluster data, *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining, Washington DC-ABD*, 333-342, 25-28 Temmuz, 2010.
- Roffo G., Castellani U., Vinciarelli A., Cristani M., Infinite feature selection: a graph-based feature filtering approach, *IEEE Transactions on Pattern Analysis Machine Intelligence*, 2020.
- Roffo G., Simone M., Feature selection via eigenvector centrality, *New Frontiers in Mining Complex Patterns in conjunction with ECML/PKDD, Riva del Garda-İtalya*, 2016.
- Yang Y., Shen H. T., Ma Z., Huang Z., Zhou X., L2, 1-norm regularized discriminative feature selection for unsupervised learning, *IJCAI international joint conference on artificial intelligence, Barcelona-İspanya*, 1589-1594, 19-22 Temmuz, 2011.
- Chormunge S., Jena S., Correlation based feature selection with clustering for high dimensional data, *Journal of Electrical Systems and Information Technology*, 5 (3), 542-549, 2018.
- Zeng H., Cheung Y., Feature selection and kernel learning for local learning-based clustering, *IEEE transactions on pattern analysis machine intelligence*, 33 (8), 1532-1547, 2010.
- McLachlan G.J., *Discriminant analysis and statistical pattern recognition*, John Wiley & Sons, 583, 2005.
- Kamiński B., Jakubczyk M., Szufel P.s., A framework for sensitivity analysis of decision trees, *Central European journal of operations research*, 26 (1), 135-159, 2018.
- Ho T. K., The random subspace method for constructing decision forests, *IEEE transactions on pattern analysis machine intelligence*, 20 (8), 832-844, 1998.
- Chen Y.W., Lin C.J., Combining SVMs with various feature selection strategies, *Feature extraction, Springer*, 315-324, 2006.
- Bulea T.C., Prasad S., Kilicarslan A., Contreras-Vidal J. L., Sitting and standing intention can be decoded from scalp EEG recorded prior to movement execution, *Frontiers in neuroscience*, 8, 376, 2014.

