

EEG Sinyallerinin Sınıflandırılmasında Evrimsel Öznitelik Seçim Metotlarının Kullanılması

Using Evolutionary Feature Selection Methods in Classification of EEG Signals

Ferda ABBASOĞLU¹, Ayla GÜLCÜ², Ulvi BAŞPINAR³

¹ Gebze Teknik Üniversitesi, Bilgisayar Mühendisliği Bölümü, 41400, Kocaeli, Türkiye

² Fatih Sultan Mehmet Vakıf Üniversitesi, Bilgisayar Mühendisliği Bölümü, 34445, İstanbul, Türkiye

³ Marmara Üniversitesi, Teknoloji Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü, 34722, İstanbul, Türkiye

Öz

Elektroensefalografi beyindeki elektriksel akımın ölçülmesi ile elde edilen sinyallerdir. Bu sinyallerin sınıflandırılması özellikle beyin sinyalleri ile ilgili rahatsızlıkların teşhis, tanı ve tedavisine katkı sağladığı için önemlidir. Bu çalışmada bu alanda epilepsi hastalığının tanısı için en çok kullanılan veri kümesi olan Bonn Üniversitesi veri kümesi kullanılmıştır. Beş farklı denekten alınan sinyallerden oluşan bu veri kümesinden anlamlı sonuçlar elde edebilmek için öncelikle veri temizleme, öznitelik çıkarma ve öznitelik seçme yöntemleri kullanılmıştır. Daha sonra bu yöntemler sınıflandırma başarısına katkıları açısından kıyaslanmıştır. İlk olarak filtrelenen veriden Ayrık Dalgacık Dönüşümü metodu ile istatistiksel özellikler çıkarılmış, ardından Diferansiyel Evrim Algoritması kullanılarak en iyi sınıflandırma sonucunu veren öznitelik alt kümesi seçilmiştir. Seçilen özniteliklere sahip veri kümesinin sınıflandırma başarısı Destek Vektör Makineleri ile test edilmiştir. Kullanılan yöntem ile bazı sınıfların ayrılmasında literatürdeki benzer çalışmalardan daha iyi sonuçlar elde edilmiştir. Bazı ikili ve üçlü kümelerin sınıflandırılmasında sırasıyla 0,98 ve 0,94 doğruluk oranları elde edilmiştir.

Anahtar Kelimeler: Elektroensefalografi sinyal analizi, Diferansiyel Evrim Algoritması, Öznitelik Çıkarımı, Öznitelik Seçimi.

Abstract

Electroencephalography signals are obtained by measuring the electrical current in the brain. The classification of these signals are especially important, as they contribute to the diagnosis, and treatment of disorders related to brain signals. In this study, the data set of the University of Bonn, which is the most widely used data set for the diagnosis of epilepsy, was used in this field. In order to obtain meaningful results from this data set consisting of signals from five different subjects, firstly, data filtering, feature extraction and feature selection methods have been used first. Later, these methods were then compared in terms of their contribution to classification success. First, statistical properties were extracted from the filtered data by the Discrete Wavelet Transform method, and then the subset of the features that gave the best classification result was selected using the Differential Evolution Algorithm. The classification success of the data set with the selected features has been tested with the Support Vector Machines. With the method used, better results were obtained than similar studies in separating some classes. In the classification of some double and triple sets, accuracy rates of 0.98 and 0.94, respectively, were obtained.

Keywords: Electroencephalography signal analysis, Differential Evolution Algorithm, Feature Extraction, Feature Selection.

I. GİRİŞ

Beyin çalışması esnasında elektriksel akım üretmektedir. Beyin Bilgisayar Arayüzü (BBA) ile beyinde üretilen elektriksel akım ölçülebilir ve bilgisayarlar tarafından yorumlanabilir. BBA sayesinde felçli dolayısıyla kas hareketlerini sağlayamayan kişilerin yaşamlarını kolaylaştırabilecek araçlar geliştirilebilmektedir.

Elektroensefalogram (EEG), beyindeki elektriksel akımın oluşturduğu sinyallerin kişinin kafasına bağlanan elektrotlar yardımıyla bilgisayar ortamı gibi sayısal bir ekrana aktarılmasıyla elde edilen bir BBA türüdür. EEG kullanılarak geliştirilen çeşitli cihazlarla kişinin zihinsel elektrik akımı ölçülmektedir. Ölçülen bu akımla o anda gerçekleştirilmek istenen hareket tespit edilmektedir. Bu tespitle birlikte mekanik kol, mekanik ayak gibi insan yaşamını kolaylaştıran aletler üretilebilmektedir.

EEG aynı zamanda kişinin nörolojik yapısının anlaşılması ile ilgili bilgiler de içermektedir. EEG verisi kullanılarak kişinin uyku bozukluğuna sahip olup olmadığı belirlenebilmekte, psikolojik rahatsızlıklarının ve epilepsi gibi beyin kaynaklı hastalıklarının tanı ve tedavisi yapılabilmektedir.

Bu çalışmada EEG sinyallerinin sınıflandırma başarısını arttırabilmek için Bonn Üniversitesi'nin veri kümesi [1] kullanılmıştır. Bu veri kümesinde 5 adet denekten alınmış veriler bulunmaktadır. Farklı fiziksel özelliklere sahip

bu deneklerden elde edilen sinyaller farklı sınıflara ait veri kümesini oluşturmaktadır. Sağlıklı denekten gözleri açık iken kayit edilen EEG verisi A sınıfını, sağlıklı ancak gözleri kapalı olan denekten alınan sinyaller B sınıfını oluşturmaktadır. Epilepsi hastası olduğu bilinen ancak nöbet halinde olmayan iki farklı denekten alınan veriler de C ve D sınıfını oluşturmaktadır. C ve D sınıfının farklılığı EEG sinyalleri okunurken elektrotların kafada takıldıkları alanın farklı olmasıdır. Son olarak epilepsi nöbeti geçiren kişiden alınan sinyaller E sınıfını oluşturmaktadır. Bu veri kümesine öncelikle öznitelik çıkarım yöntemleri, ardından öznitelik seçim yöntemleri uygulanmıştır. Bu yöntemler sonucu boyutu küçültülen veri kümesi sınıflandırıcı yardımıyla sınıflandırılmış ve başarı oranı ölçülmüştür.

Literatürde EEG sinyallerinin sınıflandırılması üzerine birçok bilimsel çalışma görülmektedir. Bu çalışmalarda en yaygın olarak tercih edilen veri kümesi Bonn Üniversitesi'nin veri kümesidir[1]. 2020 yılında Zhang G. ve arkadaşlarının yaptığı çalışmada [2] Bonn Üniversitesi'nin verisi kullanılmıştır. Bu çalışmada EEG sinyal tespiti sağlamak için çok ölçekli yerel olmayan bir ağ önerilmiştir. Ağ sınıflandırma performansını iyileştirmek için iki özel katman içeren 1D evrişimli sinir ağı kullanılmıştır. Bütün kümeler E kümesiyle karşılaştırarak sınıflandırılmış; A ve E kümesi sınıflandırması için %99,93 doğruluk bulunmuştur.

2019 yılında D. Lu ve arkadaşlarının yaptığı çalışmada [3] Bonn Üniversitesi'nin verisine [1] ilk olarak filtreleme işlemi uygulanmıştır. Daha sonra sinyal evrişimli sinir ağı yapısından geçirilerek sınıflandırılmıştır. Ancak sınıflandırma işlemi için veri kümesi 3 sınıfa ayrılmıştır. Sağlıklı kişilerden alınan A ve B sınıfları 0 ile, hasta kişilerden alınan C ve D sınıfları 1 ile nöbet halindeki kişilerden alınan E sınıfı ise 2 ile etiketlenmiştir. Bu şekilde oluşturulan üç sınıf için sınıflandırma %99 doğruluk oranı ile elde edilmiştir.

2018 yılında D. Kaya ve arkadaşlarının yaptığı çalışmada [4] yine Bonn Üniversitesi'nin verisi kullanılmıştır. Bu çalışmada temel bileşen analizinin sinyalin sınıflandırma başarısına etkisi incelenmiştir. Sınıflandırma işleminde k en yakın komşu algoritması uygulanmış, deneyler sonucunda %87,5 doğruluk oranı elde edilmiştir.

2019'da S. Ramakrishnan ve ekibi ilki Bonn Üniversitesi'nin veri kümesi, [1] ikincisi ise Boston Çocuk Hastanesi verisi [5] olmak üzere iki farklı veri kümesi üzerinde çalışma [6] yapmışlardır. Her iki veri kümesinden dalgacık katsayısı elde edebilmek için Ayrık Dalgacık Dönüşümü (ADD) uygulanmıştır. Dalgacıklar dördüncü seviyeye kadar ayrılmıştır. Ayrım sonucunda elde edilen her bir alt banttan katsayılar elde edilmiştir. Bu katsayılar kullanılarak veriyi temsil eden

istatistiksel özellikler çıkarılmıştır. Daha sonra sınıflandırma yapabilmek için verinin yarısı eğitim yarısı test için kullanılmıştır ve sonuçta %95 doğruluk oranı elde edilmiştir.

2017'de Z. Lasefr ve arkadaşları da Bonn Üniversitesi'nin verisini kullanarak bir çalışma yapmışlardır [7]. Bu çalışmada sağlıklı gözü açık denek ile (A) epilepsi nöbeti geçiren (E) denekten elde edilen sinyaller kıyaslanmıştır. Bunun için veriye ilk olarak Chebyshev filtresi uygulanmış, daha sonra filtrelenen veriye ADD uygulanmıştır. ADD ile veri alt bantlarına ayrılmıştır ve her alt banttan katsayılar elde edilmiştir. Bu katsayılar kullanılarak dalgalara ait enerji değerleri hesaplanmıştır. Son olarak, destek vektör makineleri (DVM) ve yapay sinir ağları (YSA) ile veri sınıflandırılmış ve sınıflandırma başarısı ölçülmüştür. Buna göre, en yüksek doğruluk oranı %98 olarak YSA ile edilmiştir. DVM ile en fazla %96 oranı elde edilmiştir.

2017'de Md. Rashid ve ekibi yaptıkları çalışmada [8] sinyal verisine ADD uygulamıştır. Bunun için Daubechies dalgacık ailesinden Daubechies-4 (db4) dalgacığı kullanılmıştır. Dalgacıklar 5. Seviyeye kadar ayrılmıştır. Bu ayırmadan elde edilen dalgacık katsayıları kullanılarak 10 adet istatistiksel özellikler çıkarılmıştır. Sınıflandırma yapılırken veri kümesi bir sınıfa karşı diğer sınıf kıyaslanarak ikili ve üçlü kümeler olarak düşünülüp ölçülmüştür. En yüksek %100 en düşük %79,3 doğruluk oranları elde edilmiştir.

2017'de A. Ahmadi ve arkadaşları sinyali 17 alt parçaya bölerek bir çalışma [9] yapmışlardır. Bölünme sonucu elde edilen her parça ADD ile dalgacık katsayılarına dönüştürülmüştür. Dönüşümde bior2.4, rbio2.2, Daubechies4, bior1.1, sym4, db6 ve Daubechies2 gibi dalgacık aileleri kullanılmıştır. Veriyi temsil eden bu dalgacık katsayıları kullanılarak sınıflandırma işlemi yapılmıştır ve sonuçlar kıyaslanmıştır. Sınıflandırma işlemi iki şekilde yapılmıştır. İlki bir sınıfla öteki sınıf kümelenecek toplamda iki sınıfın kıyaslanması (AxE), diğeri ise bir sınıfa karşı diğer bütün sınıfların kümelenebileceği iki sınıfın karşılaştırılmasıdır (ABCDxE). DVM ile farklı sayılarda k-katlamalı çapraz doğrulama ile sınıflandırılmıştır.

2017 yılında M.Baig ve arkadaşlarının yaptığı çalışmada [10] EEG verisine [11] ortak uzamsal örüntüler uygulanmış ve elde edilen özniteliklere diferansiyel evrim algoritması uygulanarak en yüksek doğruluk oranını veren öznitelik kümeleri seçilmeye çalışılmıştır. Sınıflandırma için k en yakın komşu, DVM kullanılmış ve sonuçlar kıyaslanmıştır.

2014'de N. Ahammad ve arkadaşlarının yaptığı çalışmada [12] Bonn verisi kullanılmıştır. Bu veriye dalgacık dönüşümü uygulanmıştır. Dalgacık dönüşümü için ADD kullanılmıştır. Dalgacıklar dördüncü

seviyeye kadar ayrılmıştır. Elde edilen alt bantlardan ilgili alt bant alt kümesi seçilmiştir (D3, D4 ve A4). Bu alt bantlardan veriyi temsil eden istatistiksel özellikler çıkarılmıştır. Bu özellikler kullanılarak kümeler sınıflandırılmıştır. D ve E sınıfları kıyaslanmış ve başarı oranı %84,2 olarak bulunmuştur.

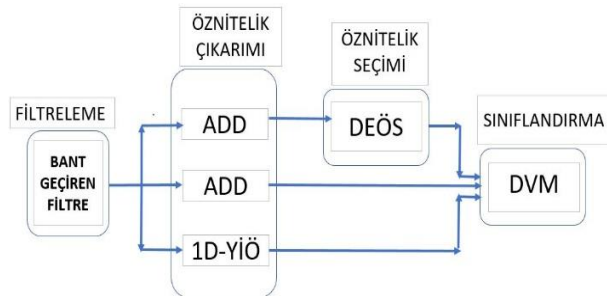
2014 yılında O. Salem ve arkadaşlarının yürüttüğü çalışmada [13] yine Bonn verisi kullanılmıştır. Veriye dalgacık dönüşümü uygulanmıştır. Elde edilen alt bantlardan özellikler çıkarılıp Karınca Kolonisi Algoritması kullanılarak sınıflandırma yapılmıştır. Sınıflandırma işleminde diğer deneklere karşı nöbet halindeki denek kümesi oluşturulmuştur. Hata oranı %9 olarak bulunmuştur.

2011 yılında Bonn verisinin kullanıldığı K. Mahajan ve arkadaşlarının yürüttüğü çalışmada [14] veri alt bantlarına ayrılmıştır. Verinin boyutunu azaltmak ve sınıflandırma başarısını arttırabilmek için temel bileşen analizi ve bağımsız bileşen analizi yöntemleri kullanılmıştır. Epilepsi rahatsızlığı olan deneklerle sağlıklı denekler kıyaslanmış ve yapay sinir ağları ile sınıflandırılmıştır.

II. MATERYAL VE YÖNTEM

Bu çalışmada kullanılan yöntemler temel olarak 4 kısma ayrılmıştır. İşlem adımları Şekil 1'deki gibidir. İlk olarak veri ön işleme tabii tutularak filtreleme işlemi gerçekleştirilmiştir. Sonra filtrelenen veriden veriyi temsil edebilecek özellikler oluşturabilmek için veriye öznitelik çıkarım yöntemleri uygulanmıştır. Bu çalışmada literatürde sıklıkla kullanılan Ayrık Dalgacık Dönüşümü (ADD) ve Bir Boyutlu Yerel İkili Örüntü (1D-YİÖ) öznitelik çıkarımı yöntemleri kullanılmıştır. Öznitelik çıkarımı sonucu elde edilen öznitelik matrisinden sınıflandırma doğruluk oranını arttıracak öznitelikler seçilmek istenmiştir. Bunun için öznitelik seçim yöntemlerinden diferansiyel evrim tabanlı öznitelik seçimi (DEÖS) kullanılmıştır. Son olarak elde edilen öznitelik matrisi DVM ile test edilmiş ve doğruluk oranları karşılaştırılmıştır.

Çalışmada yapılan deneyler için Python 3 programlama dilinde yazılmış Numpy (versiyon 1.16.4), Pandas (versiyon 0.24.2) ve SciKit-Learn (versiyon 0.21.1) kütüphaneleri kullanılmıştır.



Şekil 1. Kullanılan yöntemler

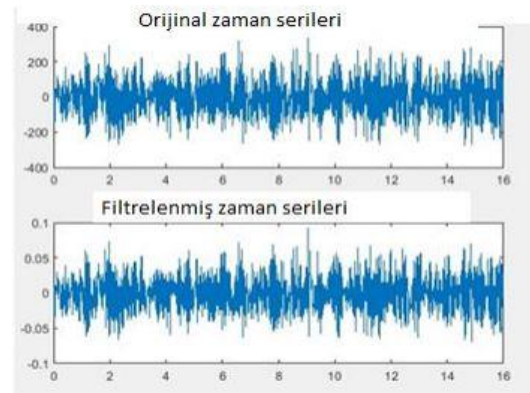
2.1. Veri Kümesi ve Ön İşleme Adımları

Bu aşamada kullanılacak verilerin detaylı tanımlanması yapılmıştır ve verilere uygulanan ön işlemler açıklanmıştır. Bu çalışmada Bonn Üniversitesi'nin Epilepsiyi inceleyen biriminde yayınlanan kullanılabilir halka açık verisi [1] kullanılmıştır. Bu veri seti 5 denekten alınan EEG sinyallerinden oluşmaktadır. 5 denek 5 sınıfı temsil etmektedir. A sınıfı sağlıklı ve gözleri açık haldeyken beyin sinyalleri kaydedilen denegi temsil etmektedir. B sınıfı ise A gibi sağlıklı fakat gözleri kapalı haldeyken sinyalleri kaydedilen veriyi temsil etmektedir. C ve D epilepsi hastalarının verileridir. C'nin D'den farkı EEG sinyalleri kaydedilirken elektrotların beyin farklı bölgelerine takılması ile elde edilmiş olmasıdır. E ise hem epilepsi hastası hem de hasta nöbet geçirirken kaydedilen veriyi temsil etmektedir. Tablo 1'de veri sınıflarının açıklamaları gösterilmiştir.

Tablo 1. Bonn Verisi Özeti

Denek	Sınıf İsimleri	Kişilerin Durumu	Elektrot Yerleştirme
Sağlıklı Denek	A	Gözler Açık	Uluslararası 10-20 sistem
	B	Gözler Kapalı	Uluslararası 10-20 sistem
Hasta Denek	C	Hasta ve Nöbet Halinde değil	Epileptojenik bölgenin karşısı
	D	Hasta ve Nöbet Halinde değil	Epileptojenik bölge
	E	Hasta ve Nöbet Halinde	Epileptojenik bölge

Bu verilere literatürde de benzer çalışmalar da rastlanan bant geçiren filtresi uygulanmıştır. Bu filtre kullanılarak gürültülü verinin elenmesi amaçlanmıştır. Şekil 2'de Filtrelemeden önce ve filtrelemeden sonra elde edilen sinyallerin frekans aralığı gösterilmiştir.



Şekil 2. Orijinal ve filtrelenmiş sinyalin frekans aralığı

Şekil 2'de filtreleme uygulandıktan sonra değişen frekans aralıkları gösterilmiştir. Bant geçiren filtre kullanılarak belirlenen üst ve alt limit haricindeki diğer

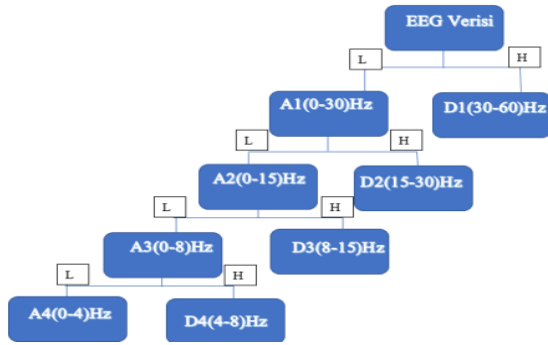
sinyaller bastırılmıştır. Bütün veri kümesine (A, B, C, D, E) üst sınırı 40 Hz ve alt sınırı 0,53 Hz olarak belirlenen bant geçiren filtresi uygulanmıştır.

2.2. Ayrık Dalgacık Dönüşümü

Sinyal verisinden anlamlı bilgiler çıkarabilmek için zaman frekans araçlarına başvurulmuştur. Dalgacık analizi bu işlemlerden biridir. Fourier yöntemleri durağan olmayan sinyaller için uygun olmadığından dalgacık dönüşümü metotları alternatif olarak kullanılmaya başlanmıştır [15]. Ayrık dalgacık dönüşümü sinyale belirlenen seviye kadar tekrarlı analiz filtresinin uygulanmasıdır [16]. Dalgacık dönüşümü matematiksel olarak aşağıdaki gibi ifade edilmiştir.

$$W(a, b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t) \Psi\left(\frac{t-b}{a}\right) dt \quad (1)$$

Eşitlik (1)'de a değeri dalgacığın ölçekleme parametresini; b dönüşüm parametresini; ψ ana dalgacığı; $W(a, b)$ işaretin sürekli dalgacık dönüşümünü belirtir. Dalgacık dönüşümü uygulanırken dalgacığın kaçınıcı seviyeye kadar ayrıldığı Şekil 3'te gösterilmektedir.



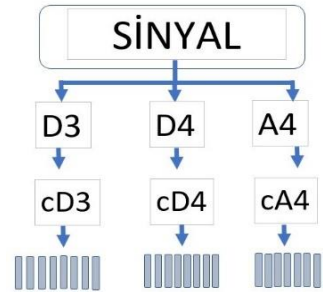
Şekil 3. EEG 4. seviye dalgacık ayrışması

Şekil 3'te gösterildiği gibi sinyal verisine hem alçak hem yüksek geçiren filtre uygulanmıştır. Sinyale alçak geçiren filtre uygulandıktan sonra elde edilen değerler sinyalin önemli özelliklerini taşımaktadır. Sinyale yüksek geçiren filtre uygulandıktan sonra elde edilen değerler sinyalin gürültülü verisini taşımaktadır. Dolayısıyla alçak geçiren filtre uygulanan kısma tekrar filtre uygulanıp dördüncü seviyeye kadar inilmektedir. Her uygulanan filtre bir seviyeyi temsil etmektedir. Dördüncü seviyeye inildiğinde elde edilen katsayılar; D1, D2, D3, D4 ve A4 'dür. Sınıflandırma için literatürde de kullanılan dalgalar alfa, teta ve delta olarak belirlenmiştir [6] [7].

Tablo 2. Dalgaların frekans aralıkları

Dalga Adı	Frekans Aralığı
Delta	0-4
Teta	4-8
Alfa	8-16
Beta	16-32
Gama	32-64

Bu dalgaları temsil eden katsayılar bulunmaktadır. Tablo 2'de gösterildiği üzere Deltayı A4, alfayı D3 ve tetayı D4 temsil etmektedir. Dalgacık dönüşümü üçüncü seviyeye kadar inseydi, A4 yani delta aralığı net elde edilemeyecekti. Beşinci seviyeye inseydi A4 aralığına tekrar filtre uygulanacaktı. Delta iki farklı dalgacık katsayısına dağılmış olacaktı. Bu sebeple literatürde de yaygın olarak kullanılan dördüncü seviye dalgacık ayrımı uygulanmıştır. Ayrıca literatürde S. Ramakrishnan ve ekibinin yaptığı çalışmada [6], N. Ahammad ve ekibinin yaptığı çalışmada [12] ve O. Salem ve ekibinin yürüttüğü çalışmada [13] dalgacıkların dördüncü seviyeye kadar ayrımı gözlenmiştir.

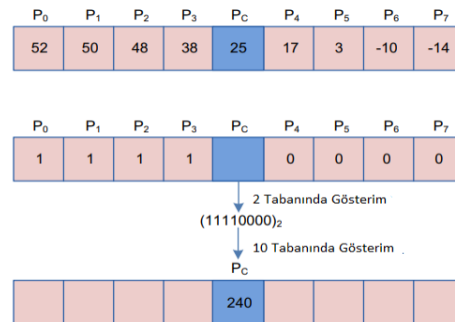


Şekil 4. ADD ile öznelik çıkarımı

Şekil 4'de gösterildiği gibi seçilen alt bantların (D3, D4, A4) her birisinden alt bant katsayıları elde edilmiştir. Bu katsayılar kullanılarak dalgayı tanımlayan istatistiksel özellikler çıkarılmıştır. Bu çalışmada her banttan çıkarılan istatistiksel özellikler; standart sapma, maksimum, varyans, çarpıklık, enerji, basıklık, minimum ve ortalamadır. 1 alt banttan 8 istatistiksel özellik çıkarılmıştır. Toplamda veriyi temsil eden 24 öznelik elde edilmiştir.

2.3. 1D-Yerel İkili Örüntü

1D-Yerel İkili Örüntü (1D-YİÖ) bu çalışmada öznelik çıkarım metoduna alternatif olarak eklenmiştir. YİÖ Ojala ve arkadaşları [17] tarafından bulunan yöntemdir. 1D-YİÖ yerel komşularda genel bir doku tanımından türetilen gri ölçekli bir doku ölçüsüdür [18]. 1D-YİÖ'da her pikselin çevredeki 8 komşusuyla kıyaslanması sonucu bir değer almasıyla oluşturulur.



Şekil 5. 1D sinyale YİÖ uygulanması [18]

Şekil 5'te görüldüğü gibi ortadaki piksel değeri sekiz komşusunun ortadaki değeri ile kıyaslanır. Eğer

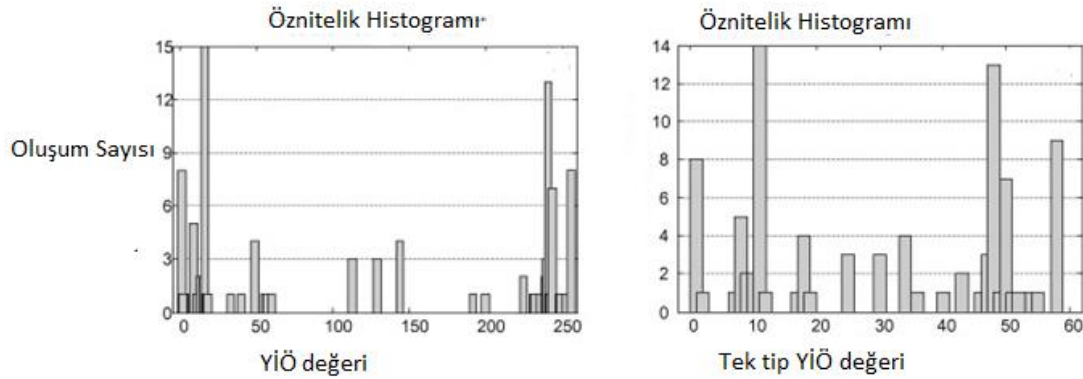
ortadaki komşudan büyük değerler ise o değeri temsilen 1 yazılır. Eğer ortadaki komşusundan küçükse o değeri temsilen 0 yazılır. Bu dönüşüm sonucunda 1 ve 0'lardan oluşan bir sayı bulunmuş olur. (x_i) merkezdeki değer i 'nci komşusu ($i = 1, 2, \dots, 8$), $G(x)$ ise merkezdeki değeri ifade eder. Anlatılan dönüşüm aşağıdaki gibi hesaplanmıştır.

$$t = G(x_i) - G(x) \quad (2)$$

$$YiÖ(x) = \sum_{i=0}^p s(t). 2^i, \quad s(t) = \begin{cases} 1, & t \geq 0 \\ 0, & t < 0 \end{cases} \quad (3)$$

1 ve 0'lerden oluşan sayı elde edebilmek için **Eşitlik (2)** ve **(3)** kullanılır. Her bir kanal için 256 özellik elde edilir. Bir kanal için 256 özelliğin fazla olması sebebiyle veri boyutu düşürülür. Veri boyutunun azaltılması için elde edilen 8 bitlik sayının düzenli olup

olmamasına göre ayırım yapılır. 8 bitlik sayının düzenli olup olmaması şu şekilde hesaplanır; eğer sayıda 1'ler görülüp 0'a rastlanırsa bu bir değişimdir. Aynı şekilde 0'lar varken 1'e rastlanması da bir başka değişimdir. Bir sayıda değişim sayısı 2'den fazla ise düzensizdir. 2'den az ise düzenlidir. Bu işlem sonunda 58 adet düzenli sayı bulunmuştur. Her bir düzenli sayı 0-58 aralığına yerleştirilmiştir. Bulunan bütün düzensiz sayılar da 59. sütuna eklenmiştir. Daha önce 256 olan veri boyutu 59'a düşmüştür. Kullandığımız veri tek kanallı olduğu için öznitelik sayısı 59, veri boyutu da 100 örnek için 100x59 olmuştur. Veri boyutunun düşürülmesi veriyi daha anlamlı hale dönüştürmüştür ve karmaşıklığını azaltmıştır. 1D-YiÖ kullanılan başka çalışmalarda veri boyutunun azaltıldığı gözlenmiştir [19] [20].



Şekil 6. Benzer çalışmada öznitelik oluşturulup veri boyutunun küçültülmesi [19]

Literatürdeki çalışmada sinyal verisi 8 bitlik alınıp, 1ve 0'lara dönüştürülmüştür. Şekil 6'daki gibi 256 öznitelik oluşturulmuştur. Veri boyutunun büyüklüğü sebebiyle düzenli ve düzensiz olarak ayrılmıştır. 58 adet düzenli bulunmuştur. Diğer düzensizler bir sütuna eklenmiştir. Bu sayede 59 öznitelik bulunmuştur. 2014'de Y. Kaya ve arkadaşlarının yaptığı çalışmada [19] veri boyutunun azaltılmasının oluşturulan kümelerde doğruluk oranını arttırdığı görülmüştür. Doğruluk oranını arttırdığı için bu çalışmada da öznitelik boyutu azaltılmıştır.

2.4. Diferansiyel Evrim Algoritması ile Öznitelik Seçimi

ADD ile elde edilen 24 öznitelikten sınıflandırma başarısını olumlu yönde etkileyecek özellikler Diferansiyel Evrim Algoritması ile Öznitelik Seçimi (DEÖS) yöntemi ile belirlenmiştir. Kullanılan Diferansiyel Evrim algoritmasında amaç fonksiyonu olarak Destek Vektör Makineleri (DVM) ile elde edilen sınıflandırma doğruluk oranı kullanılmıştır. Bu sayede Diferansiyel Evrim (DE) metodu ile tüm öznitelikler içinden en yüksek sınıflandırma başarısı verecek öznitelik alt kümesinin belirlenmesi hedeflenmiştir. Diferansiyel Evrim (DE) algoritması Storn ve Price tarafından bulunmuştur [21]. Bu algoritmada

popülasyon NP adet bireyden oluşmaktadır. Popülasyon içerisinde ise her bir birey D gerçek değerli parametre içermektedir. Diferansiyel evrim algoritması NP tane D boyutlu vektörden oluşmaktadır. Popülasyon **Eşitlik (4)** ve **(5)**'deki gibi tanımlanmıştır.

$$P_{x,g} = (x_i, g), i = 0, 1, \dots, NP-1 \quad g = 0, 1, \dots, g_{max} \quad (4)$$

$$x_{i,g} = (x_{j,i,g}), j = 0, 1, \dots, D-1 \quad (5)$$

Eşitlik (4) ve **(5)**'de $g = 0, 1, \dots, g_{max}$ vektörün hangi nesilden geldiğini belirtmektedir. $i = 0, 1, \dots, (NP-1)$ popülasyon içindeki bireyi temsil eder. $j = 0, 1, \dots, (D-1)$ birey içerisinde hangi parametrenin kullanıldığını belirtmektedir. D boyutlu b_j^L ve b_j^U olarak tanımlanan iki vektör için L alt sınırı U ise üst sınırı göstermektedir. Vektörlerin parametrelerine **Eşitlik (6)**'daki gibi üst ve alt sınırlar arasında değerler atanılmıştır.

$$X_{j,i,g} = rand(0,1) (b_j^U - b_j^L) + b_j^L \quad (6)$$

Eşitlik (6)'da rastgele sayı üretimi temsil edilmektedir. Mutasyon işlemi için r_0 temel olarak alınan vektörü, r_1 ve r_2 ise rastgele seçilmiş vektörleri temsil etmektedir.

$$v_{i,g} = x_{r_0,g} + F * (x_{r_1,g} - x_{r_2,g}) \quad (7)$$

Eşitlik (7)'deki F ifadesi $[0,1)$ arasında bir değer almaktadır. Yeni popülasyon aşağıdaki eşitlikteki gibi ifade edilmektedir.

$$\begin{aligned} P_{v,g} &= (v_{i,g}), \quad i = 0,1, \dots, N_{p-1}, \\ g &= 0,1, \dots, g_{max}, \quad v_{i,g} = v_{i,g} = (v_{j,i,g}), \\ j &= 0,1, \dots, D - 1 \end{aligned} \quad (8)$$

Eşitlik (8)'deki ara popülasyon $P_{v,g}$, NP adet mutasyon sonucu oluşmuş vektör $v_{i,g}$ 'yi içermektedir. Daha sonra Diferansiyel evrim operatörlerinden biri olan çaprazlama işlemine geçilmektedir. Bu işlem sonucu deneme vektörü oluşturulmaktadır. Bu vektörlerin parametrelerinin mutasyon sonucu oluşmuş vektörden mi yoksa popülasyondan mı geleceğine karar verilmektedir.

$$u_{j,i,g} = \begin{cases} V_{j,i,g}, & \text{if } (rand_j(0,1) \leq C_r \wedge j = j_{rand}) \\ X_{j,i,g}, & \text{if } (rand_j(0,1) > C_r \vee j = j_{rand}) \end{cases} \quad (9)$$

Eşitlik (9)'da gösterilen C_r çaprazlama olasılığını temsil etmektedir. Eşitlikte gösterildiği gibi rastgele üretilen sayı çaprazlama faktörüyle kıyaslanmaktadır. Eğer çaprazlama faktörü büyükse deneme faktörünün parametresi olarak mutasyon sonucu oluşan vektörün parametresi alınmaktadır. Çaprazlama faktörü küçükse popülasyonun parametresi alınmaktadır. Bu işlem sonucu deneme vektörleri oluşturulmuştur. Bu aşamadan sonra hangisinin seçileceği amaç fonksiyonuna en büyük katkının hangisi tarafından yapıldığına bağlı olarak değişmektedir.

DEÖS parametrelerinden biri olan iterasyon sayısı için 100, 1000 ve 10000 kullanılmıştır. 24 öznitelik içinden doğruluk oranını arttıran 10, 15 ve 20 elemanlı öznitelik kümeleri oluşturulmuştur. Çaprazlama oranı (C_r) değeri için 0,7;0,5 ve 0,3 kullanılmıştır.

2.5. EEG Verisinin Sınıflandırılması

Literatürde EEG verisini sınıflandırmak için farklı yöntemler tercih edilmiştir. Zhang G. ve arkadaşlarının yaptığı çalışmada [2] ve D. Lu ve arkadaşlarının yaptığı çalışmada [3] evrişimli sinir ağıyla sınıflandırma başarısı ölçülmüştür. O. Salem ve ekibinin yürüttüğü çalışmada [13] Karınca Kolonisi Algoritması kullanılarak sınıflandırma yapılmıştır. D. Kaya ve arkadaşlarının yaptığı çalışmada [4] sınıflandırma işleminde k en yakın komşu algoritması uygulanmıştır. Z. Lasefr ve arkadaşları yaptığı çalışmada [7] veriyi Destek Vektör Makineleri (DVM) ve Yapay Sinir Ağları (YSA) kullanarak sınıflandırmışlardır. A. Ahmadi ve ekibinin yaptığı çalışmada [9] DVM ile farklı sayılarda k-katlamalı çapraz doğrulama ile sınıflandırılmıştır.

DVM'nin tarihi 1936 yılında örüntü tanıma problemlerine çözüm arayan R.A.Fisher'a dayanmaktadır [22]. 1950 de ise Aronszajn DVM'nin

parçası olan çekirdeği ortaya çıkarmıştır [23]. 1963'e gelindiğinde Vapnik ve Lerner DVM'nin temeli sayılabilecek algoritmayı üretmişlerdir [24]. 1992'de Vapnik ve ekibi maksimum marjlı hiperplanlara çekirdek kuralı uygulayarak lineer olmayan bir sınıflandırıcı ürettiler [25]. 1995 marj sınıflandırıcı geliştirilmiştir [26].

DVM büyük boyutlu sınıflandırma problemlerinde kullanılan önemli bir öğrenme algoritmalarından biridir. DVM ile sınıflandırmada amaç sınıflara ait örneklerin, eğitim verisi ile elde edilen bir karar fonksiyonu yardımıyla birbirinden ayrılmasıdır. Söz konusu karar fonksiyonu kullanılarak eğitim verisini en uygun şekilde ayırabilecek hiper-düzlem bulunmaktadır. Bahsedilen sınıfları ayırabilecek birçok hiper düzlem çizilebilir fakat DVM'nin amacı kendisine en yakın noktaların arasındaki uzaklığı maksimuma çıkaran hiper düzlemi bulmaktır.

Bu çalışmada ayrıca literatürde sıklıkla kullanılan gözetimli öğrenme algoritmalarından Doğrusal Diskriminant Analizi ve DVM kullanılmıştır, ancak daha iyi sonuçlar elde edildiğinden sadece DVM sonuçları raporlanmıştır.

III. BULGULAR VE TARTIŞMA

Bu çalışmada epilepsi verisinin sınıflandırma başarısını arttırmak için farklı öznitelik çıkarım yöntemleri ve evrimsel tabanlı öznitelik seçim yöntemi uygulanmıştır. İlk olarak bu veri kümelerine bant geçiren filtresi uygulanmıştır. Filtrelenen verilerden ADD VE 1D-YİÖ ile öznitelik çıkarımı yapılmıştır. Daha sonra DEÖS ile öznitelik seçimi yapılmıştır. Seçilen özniteliklerin doğruluk oranları ölçülmüştür. Daha önceki çalışmalar incelenerek Bonn verisinin sınıfları arasında farklı kümeler oluşturulmuştur. Bu farklı kümelerin sınıflandırma doğruluk oranları denenmiştir. Bu kümeler ikili (AxB, AxC, AxD, AxE) ve üçlü (AxBxC, AxBxD, AxCxE, BxDxE) olmak üzere sınıflandırılmaya çalışılmıştır. Kümelerin başarı oranları kıyaslanmıştır. Bütün kümelemeler denenmiştir. Bu çalışmada ise 2'li 3'lü ve 5'li kümelerin sonuçları gösterilmektedir. Sonuçları gösterilen kümeler açıklamalarıyla Tablo 3'de verilmiştir.

Tablo 3. Oluşturulan Kümeler ve Açıklamaları

Küme Adı	Açıklaması
AxE	Sağlıklı ve Gözler Açık x Epileptik Nöbet
BxE	Sağlıklı ve Gözler Kapalı x Epileptik Nöbet
AxBxC	Sağlıklı ve Gözler Açık x Sağlıklı ve Gözler Kapalı x Hasta Nöbet Öncesi
AxBxD	Sağlıklı ve Gözler Açık x Sağlıklı ve Gözler Kapalı x Hasta Nöbet Öncesi
AxBxE	Sağlıklı ve Gözler Açık x Sağlıklı ve Gözler Kapalı x Epileptik Nöbet
AxDxE	Sağlıklı ve Gözler Açık x Hasta Nöbet Öncesi x Epileptik Nöbet

AxCxE	Sağlıklı ve Gözler Açık x Hasta Nöbet Öncesi x Epileptik Nöbet
BxDxE	Sağlıklı ve Gözler Kapalı x Hasta Nöbet Öncesi x Epileptik Nöbet
BxCxE	Sağlıklı ve Gözler Kapalı x Hasta Nöbet Öncesi x Epileptik Nöbet

Tablo 4’de ikili kümelerin sınıflandırma sonuçları gösterilmektedir.

Tablo 4’de görüldüğü üzere AE kümesi için literatürde 1D-YİÖ ile ulaşılabilen değere bu çalışmada ADD kullanarak ulaşılmıştır. Tablo 5’te üçlü kümelerin sınıflandırma sonuçları ve literatürdeki en iyi sonuçlar

gösterilmektedir.

Tablo 5’de gösterildiği üzere AxBxC ile AxBxE’nin doğruluk oranları karşılaştırıldığında E sınıfı içeren kümenin içermeyene oranla daha yüksek doğruluk oranına sahip olduğu anlaşılmaktadır. Çünkü en farklı sınıf E (epilepsi nöbeti halindeki denekten alınmış veri) sınıfıdır. Bu durumda bir kümede E deneğinin olması sınıflandırmada oluşturulan diğer kümelere göre daha yüksek doğrulukla sınıflandırma yapılmasına olanak vermektedir. Bütün sınıfların bulunduğu kümenin başarı oranı tek bir tabloda kıyaslanmıştır. Tablo 6’da sonuçlar gösterilmektedir.

Tablo 4. İkili kümelerin sınıflandırma doğruluk oranı

Çalışma	Öznitelik Seçim Yöntemi	Sınıflayıcı	Küme	Doğruluk Oranı	Özgüllük Oranı	Duyarlılık Oranı
Bu çalışma	ADD	DVM	AxE	0,9875	0,98	1,0
	1D-YİÖ		AxE	0,8	0,8	1,0
	ADD		BxE	0,8125	0,81	0,88
			BxE	0,9812	0,99	1,0
Y.Kaya	1D-YİÖ	DVM	AxE	0,985	-	-

Tablo 5. Üçlü kümelerin sınıflandırma doğruluk oranları

Çalışma	Öznitelik Seçim Yöntemi	Küme	Doğruluk Oranı	Özgüllük Oranı	Duyarlılık Oranı
Bu çalışma	ADD (4. Seviye, db2)	AxBxC	0,9041	0,92	0,92
		AxBxD	0,891	0,91	0,90
		AxBxE	0,9166	0,90	0,92
		AxDxE	0,92	0,92	0,95
		AxCxE	0,929	0,92	0,95
		BxDxE	0,9458	0,95	0,97
		BxCxE	0,95	0,92	0,97
Md. Rasih v.d.	ADD (db4)	AxCxE	0,80	-	-
		BxCxE	0,787	-	-
		AxDxE	0,80	-	-
		BxDxE	0,793	-	-
S.Ramakrishnan v.d.	ADD (4. Seviye)	AxDxE	0,98	-	-

Tablo 6. Beşli kümelerin sınıflandırma doğruluk oranları

Öznitelik Seçimi	Küme	Doğruluk Oranı	Kesinlik	Duyarlılık
ADD	AxBxCxDxE	0,7975	0,80	0,82
1D-YİÖ		0,555	0,52	0,59
ADD+DEÖS		0,8225	0,82	0,85

Tablo 6’dan elde edilen sonuçlara göre 5’li küme oluşturulduğunda en yüksek doğruluk ADD ile elde edilen öznitelik kümesinden DEÖS ile sınıflandırma başarısını arttıran özniteliklerin seçilmesi sonucu elde edilmiştir. Sınıflandırma için DVM kullanılmıştır.

Bu kısımda çalışmamızda bulunan sonuçlar kendi içerisinde kıyaslanmış ve açıklanmıştır. Bonn verisi özellikle tek kanallı olması sebebiyle işlenmesi oldukça kolay bir veri kümesidir. Bu sebeple literatürde bu

veriyle ilgili pek çok çalışma vardır. Elde edilen doğruluk oranları hem literatürde hem de bu çalışmada oldukça yüksektir.

ADD ile üçlü kümelerin varlığında yapılan çalışmada ise bilinen birçok çalışmanın doğruluk oranlarının geçildiği görülmüştür. Literatürde görülen AxCxE kümelerinin kıyaslanmasında elde edilen en yüksek doğruluk oranı %80 iken bu çalışmada %92’lere kadar çıkmıştır. Aynı şekilde bir başka üçlü küme olan

BxCxE ile yapılan deneyde literatürde elde edilen en yüksek doğruluk oranı %78,7 iken bu çalışmada %95 gibi yüksek bir doğruluk oranı bulunmuştur. Ayrıca BxDxE kümesi için bilinen doğruluk oranı %79,3 iken bu çalışmadan %94 elde edilmiştir.

5 sınıflı kümede (AxBxCxDxE) ADD ile elde edilen öznelik kümesine sınıflandırma başarısını arttıracak şekilde öznelik seçimi DEÖS ile uygulandığında elde edilen doğruluk oranı %79'dur. DEÖS parametrelerinde iyileştirmeler yapıldığında (çaprazlama oranı 0,7 ve iterasyon sayısı 1000 seçildiğinde) doğruluk oranının %82'ye çıktığı görülmektedir.

1D-YİÖ yöntemi bir örüntü tanıma işlemidir. Bu çalışmada ADD'ye eşdeğer bir yöntem olarak uygulanmıştır. Daha önce yapılan çalışmaya benzer sonuçlar bulunmuştur. Fakat ADD ile kıyaslanınca başarı oranını daha az arttırdığı görülmektedir. Ayrıca ADD'ye göre daha büyük öznelik kümesi elde edilmiştir. Bu da veri kümesinin daha hantal olmasına sebep olmuştur. 1D-YİÖ'nün daha etkili kullanılabilmesi için çeşitli ön işleme adımları ile 5'li kümenin üzerinde çalışılması gerektiği gelecek çalışmalar için düşünülmektedir.

KAYNAKLAR

- [1] Andrzejak, R. (2005). BONN UNIVERCITY EEG time series download page. (10 2019) tarihinde, http://epileptologie-bonn.de/cms/front_content.php?idcat=193&lang=3 adresinden alındı.
- [2] Zhang, G., Yang, L., Li, B., Lu, Y., Liu, Q., Zhao, W., & ... & Che, W. (2020). MNL-Network: A Multi-Scale Non-local Network for Epilepsy Detection From EEG Signals. *Frontiers in Neuroscience*, 14.
- [3] Lu, D., & Triesch, J. (2019). Residual deep convolutional neural network for eeg signal classification in epilepsy. *arXiv preprint arXiv:1903.08100*.
- [4] Kaya, D., Türk, M., & Kaya, T. (2018). En Yakın Komşu Algoritması Kullanılarak EEG Sinyallerine Boyut Azaltmanın Etkilerinin İncelenmesi. *El-Cezeri Journal of Science and Engineering*, s. 5.2: 591-595.
- [5] Physio Bank. (2018). 11 12, 2018 tarihinde (<http://physionet.org/cgi-bin/atm/ATM> adresinden alındı).
- [6] Ramakrishnan, S., & Murugavel, A. M. (2019). Epileptic seizure detection using fuzzy-rules-based sub-band specific features and layered multi-class SVM. *Pattern Analysis and Applications*, 16.
- [7] Lasefr, Z. A. (2017). Epilepsy seizure detection using EEG signals. In 2017 IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON) (pp. 162-167), IEEE. IEEE.
- [8] Or Rashid, M. M., & Ahmad, M. (2017). Epileptic Seizure Classification using Statistical Features of EEG Signal. In 2017 International Conference on Electrical, Computer and Communication Engineering (ECCE) (pp. 308-312). IEEE.
- [9] Ahmadi, A. S. (2017). A New Method for Epileptic Seizure Classification in EEG Using Adapted Wavelet Packets. İstanbul, Turkey: IEEE.
- [10] Baig, M. Z., Aslam, N., Shum, H. P., & Zhang, L. (2017). Differential evolution algorithm as a tool for optimal feature subset selection in motor imagery EEG. *Expert Systems with Applications*, 2017, 90: 184-195.
- [11] Intelligent Data Analysis Group Klaus-Robert Müller, B. B.-U. (2004, Aralık 12). BCI competition III dataset IVa. 12 10, 2020 tarihinde http://www.bbci.de/competition/iii/desc_IVa.html adresinden alındı
- [12] Ahammad, N. F. (2014). Detection of epileptic seizure event and onset using EEG. *BioMed research international* 2014, 2014(7).
- [13] Salem, O., Naseem, A., & Mehaoua, A. (2014). Epileptic Seizure Detection From EEG Signal using Discrete Wavelet Transform and Ant Colony classifier *Communications (ICC). IEEE International Conference on. IEEE*, 2014, 3529-3534.
- [14] Mahajan, K., Vargantwar, M. R., & Rajput, S. M. (2011). Classification of EEG using PCA, ICA and Neural Network. *International Journal of Engineering and Advanced Technology*, 1(80-83).
- [15] Siuly, S., Li, Y., & Zhang, Y. (2016, 11). EEG signal analysis and classification. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, s. 141-4.
- [16] Subasi, A. (2007). EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Systems with Applications*, s. Cilt 4, 1084-1093.
- [17] Ojala, T., Pietikäinen, M., & Mäenpää, T. (2000). Gray scale and rotation invariant texture classification with local binary patterns. *European Conference on Computer Vision*. Springer, Berlin, Heidelberg, s. 404-420.
- [18] Burçin, K., & Vasif, N. V. (2011). Down syndrome recognition using local binary patterns and statistical evaluation of the system. *Expert Systems with Applications*. 38 2011, Cilt 7, s. 8690-8695.
- [19] Kaya, Y., Uyar, M., Tekin, R., & Yıldırım, S. (2014). 1D-local binary pattern based feature extraction for classification of epileptic EEG signals. *Applied Mathematics and Computation*, s. 243: 209-219.

-
- [20] Tiwari, A. K., Pachori, R. B., Kanhangad, V., & Panigrahi, B. K. (2016). Automated diagnosis of epilepsy using key-point-based local binary pattern of EEG signals. *IEEE journal of biomedical and health informatics*, s. 21(4), 888-896.
- [21] Price, K., Storn, R. M., & Lampinen, J. A. (2006). *Differential Evolution A Practical Approach to Global Optimization*. Springer Science & Business Media.
- [22] Fisher, R. A. (1936). The use of multiple measurements in taxonomic problems. *Annals of eugenics*, 7(2), 179-188.
- [23] Aronszajn, N. (1950). Theory of reproducing kernels. *Transactions of the American mathematical society*, 68(3), 337-404.
- [24] Vapnik, V., & Lerner, A. Y. (1963). Recognition of patterns with help of generalized portraits. *Avtomat. i Telemekh*, 24(6), 774-780.
- [25] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992, July). A training algorithm for optimal margin classifiers. *In Proceedings of the fifth annual workshop on Computational learning theory* (pp. 144-152).
- [26] Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine learning*, 20(3), 273-297.