



ŞİZOFREN HASTALARININ TESPİTİNDE YAPAY ZEKA TABANLI PERFORMANS ANALİZİ

Ali NARİN^{1*}, Burhan SEZEN¹

¹Zonguldak Bülent Ecevit University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 67100, Zonguldak, Türkiye

Özet: Şizofreni (SZ), algıda ve davranışlarda olumsuzlukların yaşandığı, seslerin ya da yaşananların gerçek veya hayal ayrımının yapılamadığı zihinsel bir hastalıktır. Halüsinasyonlar, hezeyanlar, konuşma ve duyguyu ifade etmede zorluk yaşama şizofreni belirtileri arasında yer almaktadır. Bu gibi kalıcı problemlerin önüne geçilmesi için SZ'nin erken tespiti hayati öneme sahiptir. Tanı amaçlı çoğunlukla ön muayene ile birlikte kan tahlili, beyin görüntüleme ve elektroensefalogram (EEG) sinyalleri kullanılmaktadır. EEG oldukça düşük genliğe sahip olmasının yanı sıra çok farklı frekanslar barındırması nedeniyle SZ'nin tespit performansı değişebilmektedir. Bu çalışmada EEG sinyalleri kullanılarak SZ hastalarının tespiti gerçekleştirilecektir. 39 sağlıklı, 45 SZ hastasına ait EEG verileri üzerinden çalışma sonuçları elde edilmiştir. Farklı normalize edilmiş EEG işaretleri üzerinden zaman ve frekans alanı ölçümleri ile birlikte geriye doğru eleme yöntemi de uygulanarak tespit edilen özniteliklerin başarımları kayıt altına alınmıştır. Çalışmada k-en yakın komşu (k-NN) algoritması kullanılmıştır. Sonuç olarak z-skor normalizasyonlu ve 13 öznitelikle k-NN (k=1) için %83,6 doğruluk değeri elde edilmiştir. Elde edilen bu sonuçlar SZ hastalarının oldukça yüksek başarımla tespit edilebileceğini göstermektedir. Çalışmanın nöroloji uzmanlarına da karar destek sistemi olarak fayda vereceği düşünülmektedir.

Anahtar kelimeler: Şizofreni, EEG, Öznitelik çıkarımı, Geriye doğru eleme, k-NN

Artificial Intelligence Based Performance Analysis in Detection of Schizophrenia

Abstract: Schizophrenia (SZ) is a mental illness in which there are negative perceptions and behaviors, and it is not possible to distinguish between real and imaginary sounds or experiences. Hallucinations, delusions, difficulty in speaking and expressing emotions are among the symptoms of schizophrenia. Early detection of SZ is of vital importance in order to prevent such permanent problems. Blood tests, brain imaging and EEG signals are mostly used together with preliminary examination for diagnosis. Since EEG has a very low amplitude and contains very different frequencies, SZ detection performance may vary. In this study, SZ patients will be detected using EEG signals. The study results were obtained from EEG data of 39 healthy and 45 SZ patients. The performance of the detected features was recorded by applying backward elimination method together with time and frequency domain measurements on different normalized EEG signals. The k-nearest neighbor (k-NN) algorithm was used in the study. As a result, 83.6% accuracy value was obtained for k-NN (k=1) with z-score normalization and 13 features. These results show that SZ patients can be detected with very high accuracy. It is also thought to be useful as a decision support system for neurology specialists.

Keywords: Schizophrenia, EEG, Feature extraction, Backward elimination, k-NN

*Corresponding author: Zonguldak Bülent Ecevit University, Faculty of Engineering, Department of Electrical and Electronics Engineering, 67100, Zonguldak, Turkey.

E mail: alinarin@beun.edu.tr (A. NARİN)

Ali NARİN  <https://orcid.org/0000-0003-0356-2888>

Burhan SEZEN  <https://orcid.org/0000-0002-7286-740X>

Gönderi: 12 Kasım 2024

Kabul: 21 Mart 2025

Yayınlanma: 15 Mayıs 2025

Received: November 12, 2024

Accepted: March 21, 2025

Published: May 15, 2025

Cite as: Narin A, Sezen B. 2025. Artificial intelligence based performance analysis in detection of schizophrenia patients. BSJ Eng Sci, 8(3): 695-703.

1. Giriş

Son yıllarda, psikiyatrik bozukluklar oldukça yaygın hale gelmiştir ve bu rahatsızlıkların gelişimini etkileyen faktörler önemli bir araştırma konusu olmuştur. Yapılan araştırmalar, genetik, biyolojik, olumsuz yaşantıların ve travmaların ruhsal bozuklukların gelişim sürecinde belirleyici bir rol oynadığını ortaya koymaktadır (Akcan ve Taşören, 2020, Aktay ve Hızlı Sayar, 2021). Özellikle, çocukluk çağındaki maruz kalınan fiziksel, cinsel ve duygusal ihmal ile istismar türlerinin, bireyin ileriki yaşamında psikiyatrik rahatsızlıkların gelişimine zemin hazırladığı vurgulanmaktadır (Karal ve Atak, 2022). Psikiyatrik bozukluklar, bireylerin günlük yaşam aktivitelerini, mesleki performanslarını ve sosyal

ilişkilerini önemli ölçüde olumsuz etkilemektedir. Bu bozukluklara sahip kişiler, günlük işlevselliıklarını yerine getirmede, temel ihtiyaçlarını karşılamada ve sosyal etkileşimlerini sürdürmede zorluklarla karşılaşabilmektedir. Bunun yanı sıra, toplumsal düzeyde psikiyatrik hastalıklara yönelik olumsuz ve damgalayıcı tutumların varlığı, bu bireylerin topluma yeniden entegrasyonunu ve sosyal destek sistemlerine erişimini güçleştirmektedir (Boz vd., 2020). Yaygın olarak görülen psikiyatrik hastalıklar arasında anksiyete bozuklukları, dikkat eksikliği ve hiperaktivite bozukluğu, uyku bozuklukları, yeme bozuklukları, obsesif-kompulsif bozukluk, şizofreni (SZ), sosyal anksiyete bozukluğu ve alkol ile madde kullanımına bağlı gelişen bozukluklar



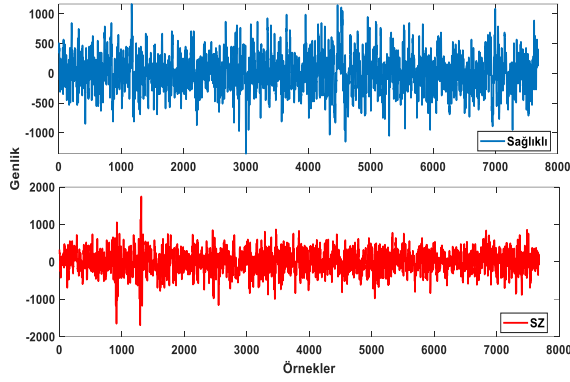
bulunmaktadır. Bu rahatsızlıkların içerisinde SZ, halüsinasyonlar, düzensiz konuşma, düşünmede zorluk ve motivasyon eksikliği gibi belirtiler gösteren zihinsel bir hastalıktır (Rizvi vd., 2024). SZ hastalığının dünya nüfusunun yaklaşık %1'ini etkilediği bilinmektedir (Birnbaum, 2023). SZ dünya çapında önde gelen 25 engellilik nedeninden biri olarak görülmektedir (Jin ve Mosweu, 2017). Burada en önemli konu SZ hastalarının doğru ve kesin bir şekilde tespit edilebilmesidir. Genel anlamda SZ hastalarıyla birlikte diğer psikolojik hastalıklarda da hezeyan, halüsinasyon gibi zihinsel bozukluklar olabileceğinden kesin tanı konulamamaktadır (Larson vd., 2010). SZ tespitinde hastanın klinik geçmişi, kan tahlili, beyin görüntüleme ve elektroensefolagram (EEG) sinyallerinden yararlanılır. Bunlar arasında daha düşük maliyetli EEG sinyallerini kullanarak literatürde sınırlı sayıda SZ hastalarının otomatik tespiti üzerine yapılmış çalışma bulunmaktadır. Literatüre bakıldığında, Kim vd. (2015) çalışmalarında SZ hastalarından ve aynı sayıda sağlıklı kontrol grubundan elde edilen EEG sinyallerinde delta, teta, alfa, beta ve gama gibi beş frekans bandını analiz etmişler, bu frekans bantlarının spektral değerlerini Hızlı Fourier Dönüşümü ile hesaplamışlardır. Delta alt frekans bandında %62,2 doğruluk elde etmişlerdir. Piryatinska vd. (2017) SZ tespiti için rastgele orman sınıflandırıcısını kullanan klasik bir özellik mühendisliği modeli uygulamış ve %83,6 oranında doğruluk değeri elde etmiştir. Dvey-Aharon vd. (2015) Stockwell tabanlı zaman-frekans temsili ve özellik optimizasyon yöntemi ile SZ hastalarına ait EEG sinyallerini başarıyla analiz etmiş ve %88,7 doğruluk elde etmiştir. Bu yöntem ile, EEG sinyallerindeki zaman-frekans değişimleri ayrıntılı olarak değerlendirilmiştir. Santos-Mayo vd. (2016) tarafından yapılan çalışmada ise 157 özellik çıkarılmış ve özellik seçimi için istatistiksel bir yaklaşım olan student's t-test uygulanmıştır. Çeşitli sınıflandırıcıların denendiği bu çalışmada, destek vektör makinesi (SVM) ile en iyi sınıflandırma performansını %93,4 doğruluk değeri ile elde etmişlerdir. Das ve Pachori (2021) iç mod fonksiyonları ve Hjorth parametrelerini kullanarak SVM sınıflandırıcı algoritması ile %98,9 doğruluk elde etmiştir. Bir başka çalışmada Türk vd. (2024) EEG sinyallerinin dalgacık alt bantlarını iki boyutlu görüntüye dönüştürerek SZ teşhisini gerçekleştirmek için ResNet tabanlı bir model sunmaktadır. %92,9 doğruluk değeri elde eden sistem, hastalığa özgü özellikleri görselleştirmeyi sağlamaktadır. Bu çalışmalar ışığında, EEG sinyallerini temel alan algoritmalarla SZ tanısında önemli gelişmeler elde edildiği gösterilmektedir. Bunlara ek olarak EEG işaretleri üzerindeki normalizasyon işlemlerinin sınıflandırma başarımlarını ele alan çalışmalar da vardır (Hosseini & Naghibi-Sistani, 2011; Bashivan et al., 2016). Fakat, SZ hastalarının tespitine yönelik farklı normalizasyon tekniklerinin sınıflandırma başarımlarının analizi detaylı olarak incelenmemiştir. Ayrıca, EEG işaretlerinin farklı hastalıkları sınıflandırma çalışmalarında, k-NN algoritmasında k değerinin model performansı üzerindeki etkisi bazı çalışmalarda sunulmuştur (Subasi, 2007; Lotte et al., 2018). Burada da,

SZ teşhisi özelinde detaylı bir çalışma bulunmadığı görülmektedir. Bu sebeple k değerinin optimizasyonu ve uzaklık ölçütleri ile kombinasyonu üzerine daha çok incelemeye ihtiyaç duyulmaktadır. Bu çalışmada, Moskova Devlet Üniversitesi tarafından halka açık şekilde sunulan EEG veri seti kullanılmıştır. Veri seti içinde 39 sağlıklı ve 45 SZ'li olmak üzere etiketlenmiş toplam 84 adet EEG kaydı bulunmaktadır. Normalize edilmemiş ham EEG işareti, z-normalizasyonlu ve min-max normalizasyonlu EEG işaretleri üzerinden zaman ve frekans alanı ölçümleri kullanılarak öznitelik çıkarılmıştır. Elde edilen özniteliklere geriye doğru eleme yöntemi uygulanmış ve bu şekilde yüksek performans gösteren öznitelikler seçilerek belirlenmiştir. Bu öznitelikler ile k- en yakın komşu (k-NN) algoritması için k değerinin 1-20 arasındaki değerleri için performans sonuçlarına bakılmıştır. Bu çalışmanın özgün değerini özetleyecek olursak maddeler halinde şu şekilde ifade edebiliriz: (i) Normalizasyon işlemlerinin performansa etkileri analiz edildi, (ii) Literatürde sıklıkla tercih edilen temel algoritmalar olan k-NN algoritmasının k değerinin performansa olan etkileri ve uzaklık ölçütünün performansları incelendi, (iii) Temel öznitelikler olan zaman ve frekans alanı ölçümlerinin Şizofren hastalarını tespitindeki başarımları analiz edildi, (iv) Son olarak geriye doğru eleme yöntemi ile elde edilen özniteliklerin performans analizleri yapılmıştır. Çalışmanın devamında, kullanılan veri seti, gerçekleştirilen ön işlem adımlarından z-skor normalizasyon, min-max normalizasyon, zaman ve frekans alanı ölçümleri, k-NN algoritması, performans ölçütleri, bulgular ve tartışma ve sonuç başlıklarına yer verilecektir.

2. Materyal ve Yöntem

2.1 Veri Seti

Çalışmada http://brain.bio.msu.ru/eeg_schizophrenia.htm web sitesinden ulaşılan Moskova Devlet Üniversitesi tarafından erişime açılmış EEG veri seti kullanılmıştır (Borisov and Kaplan, 2005). Veri seti içinde 39 adet sağlıklı ve 45 adet SZ hastasına ait veri bulunmaktadır. Tüm EEG kayıtları, tüm deneklerin gözleri kapalı bir şekilde rahat ve uyanık durumda olduğu sırada elde edilmiştir. Tüm EEG işaretleri 10-20 elektrot sistemine göre yerleştirilen 16 kanaldan oluşmaktadır (O1, O2, P3, P4, Pz, T5, T6, C3, C4, Cz, T3, T4, F3, F4, F7). Her bir EEG sinyali 128 Hz hızında örneklenmiştir. Şekil 1'de sağlıklı ve SZ hastalarına ait EEG kayıtları gösterilmektedir. Çalışmada tüm EEG kanalları kullanılmıştır.



Şekil 1. Sağlıklı ve SZ EEG kayıt örnekleri.

2.2. Normalizasyon

Farklı ölçeklerde bulunan verilerin ortak bir ölçeğe indirgenmesi sürecine normalizasyon adı verilmektedir. Veri dağılımının homojen hale getirmeyi amaçlayan bu işlem verilerin analiz süreçlerine önemli katkılar sunmaktadır. Normalizasyon işleminin veriler arasındaki düzensizliği ve karmaşıklığı minimize ederek model performansını iyileştirdiği bilinmektedir (Singh ve Singh, 2020). Bu çalışmada iki farklı normalizasyon yöntemi kullanılmıştır. Bunlar min-max normalizasyon ve z-skor normalizasyon yöntemleridir.

2.2.1. Min-max normalizasyon

Min-max normalizasyon yöntemi, bir özneliğe ait verilerin en küçük ve en büyük değerlerinin kullanılmasıyla, ilgili öznelikteki tüm verilerin belirli bir aralığa, genellikle $[0, 1]$ veya $[-1, 1]$ aralığına, ölçeklenmesini sağlayan bir tekniktir. Bu yöntem, her bir öznelik için ayrı ayrı uygulanır ve verilerin farklı aralıklarda olmasından kaynaklanan aşırı sayısal farkların ortadan kaldırılmasını hedefler. Böylelikle, tüm öznelikler belirli bir standart aralıkta normalize edilerek, modelleme ve analiz süreçlerinin performansı iyileştirilir ve daha tutarlı sonuçlar elde edilmesi sağlanır. Min-max normalizasyonu, makine öğrenmesi ve veri madenciliği gibi alanlarda veri ön işleme aşamasında sıklıkla kullanılan etkili bir yöntemdir (Narin, 2020). Min-max normalizasyon yönteminin matematiksel gösterimi (eşitlik 1),

$$X_{min-max(i)} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (1)$$

şeklinde. Burada, $X_{min-max(i)}$, normalize edilmiş i. veriyi, X_i , normalize edilecek i. veriyi, X_{min} , aynı öznelikteki en küçük değeri, X_{max} , aynı öznelikteki en büyük değeri ifade etmektedir.

2.2.2. Z-Skor normalizasyon

Z-skor normalizasyon işleminde veriler ortalamadan uzaklıklarına göre standart hale getirilir. Yani, veri noktaları ortalamadan ne kadar uzakta ise o nispette bir z-skoru alır. Bu normalizasyon işlemi verilerin birim varyansa sahip olduğu ve ortalamalarının sıfır olduğu bir ölçeklendirmeye olanak sağlar. Matematiksel formülü şu şekildedir (eşitlik 2) (Sürücü vd., 2023):

$$Z - skor(i) = \frac{X_i - \mu}{\sigma} \quad (2)$$

Burada, $Z - skor(i)$, z-skor normalizasyonu yapılarak elde edilen yeni değeri, X_i normalizasyon yapılacak i. veri noktasını, μ ortalamayı ve σ standart sapmayı temsil eder. Bu yöntem özellikle makine öğrenimi ve istatistiksel analizlerde çok yaygın olarak kullanılır.

2.3. Öznelik Çıkarımı

Öznelik çıkarımı, ham veya işlenmiş verilerin altında yatan özellikleri elde etme sürecidir. Bir diğer ifade ile bir veri kümesindeki sınıflara veya ayrımlara katkıda bulunan temel özellikleri belirleme sürecidir. Bu aşamada ham veri boyutları azalırken makine öğrenmesi algoritmalarının da model performansları artmaktadır. Bu çalışmada, Tablo 1'de yer alan özneliklerden yararlanılmıştır.

Tablo 1. Elde edilen öznelikler ve sayıları

Öznelik	Ölçümler	Adet
Zaman Analizi	Ortalama, Varyans, Standart Sapma, En Büyük Değer, En Küçük Değer, Basıklık (Kurtosis), Çarpıklık (Skewness) ve Hurst Üssü	8
Frekans Analizi	Delta, Teta, Alfa, Beta ve Gama bantları ile Teta/Alfa, Beta/Alfa, Teta + Alfa/Beta, Teta/Beta, Teta + Alfa/Beta + Alfa, Gama/Delta, Gama + Beta/Delta + Alfa	12
Toplam		20

2.3.1. Zaman analizi ölçümleri

Zaman analizi ölçümleri kapsamında toplamda 8 adet öznelik çıkarılmıştır. Çıkarılan bu öznelikler EEG verisinin zaman boyutundaki ölçümleridir. Bunlar, ortalama, varyans, standart sapma, en büyük değer, en küçük değer, çarpıklık, basıklık, hurst üssü ölçümleridir. Sırasıyla matematiksel formülleri şu şekildedir (eşitlikler 3, 4, 5, 6, 7 ve 8):

$$\text{Ortalama (Mean, } \mu) = \frac{1}{N} \sum_{t=1}^N x(t) \quad (3)$$

Burada N değeri veri nokta sayısını, $x(t)$ ortalaması alınacak sinyalin zaman boyutundaki gösterimini ifade eder.

$$\text{Varyans (Variance, } \sigma^2) = \frac{1}{N} \sum_{t=1}^N (x(t) - \mu)^2 \quad (4)$$

Standart Sapma (Standard Deviation) (eşitlik 5)

$$\sigma = \sqrt{\frac{1}{N} \sum_{t=1}^N (x(t) - \mu)^2} \quad (5)$$

En Büyük Değer (Maximum, x_{max}) = $\max(x(t))$

En Küçük Değer (Minimum, x_{min}) = $\min(x(t))$

$$\text{Çarpıklık (Skewness, } Y1) = \frac{\frac{1}{N} \sum_{t=1}^N (x(t) - \mu)^3}{\sigma^3} \quad (6)$$

$$\text{Basıklık (Kurtosis, } Y3) = \frac{\frac{1}{N} \sum_{t=1}^N (x(t) - \mu)^4}{\sigma^4} - 3 \quad (7)$$

$$\text{Hurst Üssü (Hurst Exponent, H)} = \lim_{N \rightarrow \infty} \frac{\log(R/\sigma)}{\log(N)} \quad (8)$$

Burada, R, zaman serisinin kümülatif aralığını ifade eder.

2.3.2. Frekans analizi ölçümleri

Frekans analizi, zaman serisi verilerinin spektral bileşenlerini inceleyerek sinyalin zamana bağlı değişimlerini analiz etmeye olanak sağlayan önemli bir yöntemidir. EEG sinyallerinin frekans analizi, beyin aktivitesini karakterize etmek ve nörolojik durumları anlamlandırmak açısından büyük bir öneme sahiptir. EEG sinyalleri genellikle delta (0.5-4 Hz), teta (4-8 Hz), alfa (8-12 Hz), beta (12-30 Hz) ve gama (30 Hz ve üstü) olmak üzere beş ana frekans bandından oluşmaktadır. Bu çalışmada, EEG verilerinin barındırdığı frekans bileşenlerinin elde edilmesi için Sürekli Dalgacık Dönüşümü (SDD) kullanılmıştır. SSD, zaman-frekans analizi yaparak sinyalin zamana bağlı değişen frekans bileşenlerini belirlemekte güçlü bir yöntemdir. Bu dönüşüm sayesinde EEG sinyallerinin spektral dağılımı çıkarılmış ve her bir frekans bandının spektral güç yoğunluğu hesaplanmıştır.

Elde edilen 12 adet frekans alanı ölçümü aşağıdaki gibidir:

- Delta (0.5-4 Hz) güç değeri
- Teta (4-8 Hz) güç değeri
- Alfa (8-12 Hz) güç değeri
- Beta (12-30 Hz) güç değeri
- Gama (30 Hz ve üstü) güç değeri
- Teta/Alfa oranı - Zihinsel yorgunluk ve dikkat süreçleri ile ilişkilendirilmiştir.
- Beta/Alfa oranı - Bilişsel aktivite ve dikkat düzeyiyle bağlantılıdır.
- (Teta+Alfa)/Beta oranı - Bilişsel süreçlerde dikkat ve gevşeme dengesini temsil eden bir parametredir.
- Teta/Beta oranı - Dikkat eksikliği ve hiperaktivite bozukluğu (DEHB) araştırmalarında önemli bir ölçümdür.
- (Teta+Alfa)/(Alfa+Beta) oranı - Bilişsel yük ve gevşeme durumlarını analiz etmek için kullanılan bir parametredir.
- Gama/Delta oranı - Beyin fonksiyonlarının entegrasyonu ve bilinç düzeyi ile ilişkilendirilen bir ölçümdür.
- (Ga

ma+Beta)/(Delta+Alfa) oranı - Nörobilişsel süreçlerin değerlendirilmesinde kullanılan bir bileşik parametredir. Bu parametrik özellikler, beyin aktivitesinin bilişsel ve nörolojik durumlarını daha iyi anlamak amacıyla literatürde yaygın olarak kullanılmaktadır (Eoh vd., 2005; Jap vd., 2009; da Silveira vd., 2016).

2.4. K-En Yakın Komşu

k-En Yakın Komşu (k-NN) algoritması, en basit ve en çok kullanılan sınıflandırma algoritmalarından biridir. Parametrik olmayan denetimli öğrenme kapsamında yer alan k-NN algoritması, sınıflandırma ve regresyon problemlerinde sıklıkla kullanılır (Şenyer Yapıcı ve Uzun Arslan, 2024). Bu algoritmada ayrıca bir eğitim süreci yoktur. Test aşamasında yeni bir veri noktasının sınıfını ya da değerini belirlerken, veri kümesindeki en yakın k adet komşusuna bakar. Komşu kavramı, verilerin birbiri arasındaki uzaklık ve benzerlik olarak tanımlanır. Sınıflandırma problemlerinde sınıfı bilinmeyen bir veri

girdi olarak verildiğinde mevcut veri noktaları ile arasındaki uzaklıklar hesaplanır. En küçük mesafeye sahip k adet komşu seçilir. Seçilen k adet komşunun çoğunlukta olduğu sınıf değerine göre, yeni veri noktasının sınıfı olarak atanır. Bu sayede yeni verinin en yakın komşularına göre karar verilir. Regresyon problemleri için kullanılıyorsa, seçilen k adet komşunun ortalama değerine göre yeni veri üretilir. k-NN algoritmasında en önemli parametre test sürecinde kullanılan k değeridir. Probleme özgü k değerlerinin belirlenmesi önemlidir. Her probleme göre farklı k değerlerinde algoritma farklı performans gösterir. Örneğin, k değerinin çok küçük seçilmesi, modelin aşırı uyum yapmasına yani her veriye çok fazla duyarlı hale gelmesine neden olabilir. Yüksek k değeri ise modelin genellemeyi artırarak daha dengeli tahminlerde bulunmasını sağlayabilir. Bazen yüksek k değerleri karar sınırlarının bulanıklaşmasına yol açarak başarımları düşürebilmektedir. Performansı etkileyen bir diğer parametre uzaklık ölçütüdür. "Euclidean", "Cosine", "Hamming", "Minkowski", "Chebyshev" ve "Cityblock" gibi bir çok uzaklık ölçütü bulunmaktadır. Bu çalışmada dört farklı uzaklık ölçütüne göre çalışma sonuçları elde edildi. Bunlar (eşitlikler 9, 10, 11 ve 12):

$$d_{\text{Euclidean}} = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (9)$$

$$d_{\text{Minkowski}} = \sum_{i=1}^n (|x_i - y_i|^p)^{\frac{1}{p}} \quad (10)$$

$$d_{\text{Chebyshev}} = \max_{i=1}^n |x_i - y_i| \quad (11)$$

$$d_{\text{Cityblock}} = \sum_{i=1}^n |x_i - y_i| \quad (12)$$

şeklinde.

2.5. Geriye Doğru Eleme ile Öznitelik Seçimi

Geriye doğru eleme yöntemi küçük veri setleri ve öznitelikleri bir özellikleri az olan veri setleri için model performansını arttırmak için kullanılan öznitelik seçme yöntemidir. Bu yöntem gereksiz ya da önemsiz özniteliklerin veri setinden elenmesini hedefler. Yöntem olarak tüm öznitelikler ile başlanılan başarımların performanslarının kademeli olarak çıkarılan her bir öznitelik performansını düşürüp düşürmediğine bakılır. Eğer başarımlar düşüyorsa o öznitelik veri seti için önemlidir. Çıkarıldığında başarımlar yükseliyorsa o öznitelik veri setinde istenmez ve elenir. Bu sayede yalnızca modelin performansını arttırmaya katkıda bulunan önemli özniteliklerin veri setinde kalması sağlanır (Farahdiba vd., 2023).

2.6. Performans Ölçütleri

Performans ölçütleri, sınıflandırma algoritma ve parametrelerinin ne kadar doğru olduğunu gösteren göstergelerdir. En çok kullanılan performans metrikleri Doğruluk (Accuracy, ACC), Duyarlılık (Recall, REC), Kesinlik (Precision, PRE) ve F1-Skor (F1) dur (Kamble vd., 2022). Tüm bu metrikler hata matrisi üzerinde bulunan TP, TN, FP, FN değerleri ile hesaplanır. Bir sınıflandırma probleminde, gerçekten hasta olup sınıflandırıcı tarafından doğru bir şekilde hasta olarak tanımlanan

örneklerin sayısı 'TP' (True Positive) olarak adlandırılır. Gerçekte hasta olan ancak sınıflandırıcı tarafından yanlış bir şekilde sağlam olarak belirlenen örneklerin sayısı ise 'FN' (False Negative) olarak ifade edilir. Benzer şekilde, gerçekten sağlam olup sınıflandırıcı tarafından doğru bir şekilde sağlam olarak tanımlanan örneklerin sayısı 'TN' (True Negative) ve aslında sağlam olan bireylerin yanlışlıkla hasta olarak sınıflandırıldığı durumlar ise 'FP' (False Positive) olarak tanımlanır. ACC değeri, tüm örnekler içindeki doğru sınıflandırma oranını temsil eder (eşitlikler 13, 14, 15 ve 16) (Erdoğan ve Narin, 2021).

$$ACC = \frac{TP+TN}{TP+TN+FP+FN} \quad (13)$$

REC değeri, modelin gerçek pozitifleri ne kadar doğru tahmin ettiğini temsil eder. Özellikle "pozitif" sınıfa ait örneklerin önemli olduğu durumlarda sıklıkla tercih edilir.

$$REC = \frac{TP}{TP+FN} \quad (14)$$

PRE değeri, modelin pozitif olarak sınıflandırdığı örneklerin ne kadarının gerçekten pozitif olduğunu temsil eder.

$$PRE = \frac{TP}{TP+FP} \quad (15)$$

F1 değeri ise REC ve PRE değerlerinden elde edilen bir değerdir. Harmonik ortalama kullanılarak hesaplanır.

$$F1 = 2x \frac{PRE \times REC}{PRE + REC} \quad (16)$$

Çalışma sonuçlarını elde ederken verilerin test ve eğitim olarak ayrılmasında 10-fold çapraz doğrulama yöntemi kullanılmıştır. Burada verilerden bir tanesi test diğer kalan dokuz tanesi eğitim için kullanılır. Her bir parça test sürecinden geçirilinceye kadar devam eder. Tüm parçalar bitince ortalama değer hesaplanır.

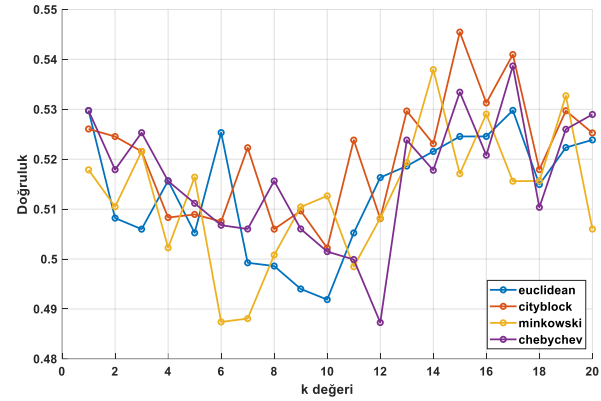
3. Bulgular ve Tartışma

Bu çalışmada, verilen ön işlem süreçleri, öznetelik çıkarımı, öznetelik seçimi ve k-NN algoritmasının performansının elde edilmesinde MATLAB programı kullanılmıştır. Çalışma sonuçları iki kısma ayrılmıştır. Birinci kısımda elde edilen zaman ve frekans alanı ölçümlerinin tamamı kullanılmıştır. k-NN algoritması için farklı k değerleri, dört farklı uzaklık ölçütü ve iki farklı normalizasyon işlemi için sonuçlar hesaplanmıştır. İkinci kısımda, elde edilen özneteliklere geriye doğru eleme yöntemi uygulanmış ve seçilen öznetelikler kullanılarak k-NN algoritması için farklı k değerleri, dört farklı uzaklık ölçütü ve iki farklı normalizasyon işlemi ile sonuçlar tekrarlanmıştır. Tüm çalışmada veriler 10-fold çapraz geçerlilik yöntemi kullanılarak test edilmiştir.

3.1. Tüm Öznetelikler ile Performans Sonuçları

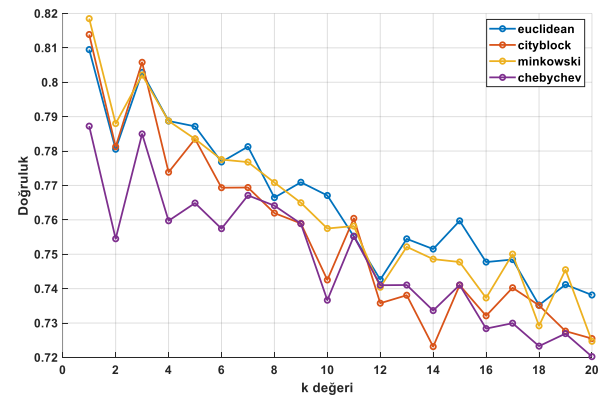
Bu kısımda elde edilen 20 adet öznetelik kullanılmıştır. Bu özneteliklerin 8 tanesi zaman alanı 12 tanesi frekans alanına aittir. k-NN algoritmasındaki uzaklık ölçütleri 'Euclidean', 'Cityblock', 'Minkowski', 'Chebychev' ile k değerinin 1-20 arasındaki değerleri için Şekil 2'de doğruluk değerleri verilmiştir. Normalizasyon

uygulanmamış ve tüm öznetelikler ile k=15 için en yüksek %54,6 oranında doğruluk değeri bulunmuştur. En yüksek doğruluk değeri 'Cityblock' uzaklık ölçütü ile elde edilmiştir. En kötü performans değeri 'Chebychev' uzaklık ölçütü ile k=12 değerinde %48,8 olarak kaydedilmiştir.

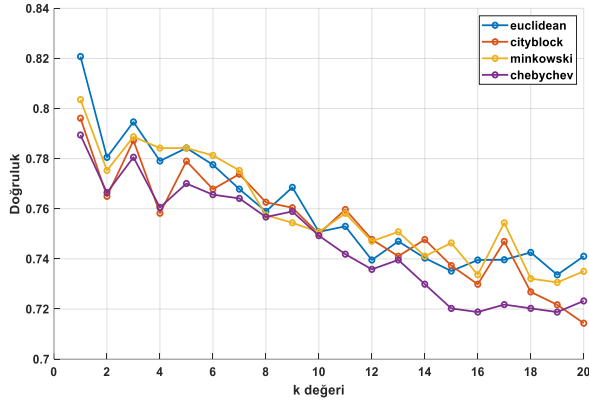


Şekil 2. Farklı uzaklık ölçütlerine göre normalize edilmemiş veriler için k değerlerine göre doğruluk grafiği.

Verilere z-skor normalize işlemi uygulandığında elde edilen veriler üzerinden çıkarılan tüm öznetelikler kullanıldığında sonuçlar Şekil 3'de sunulmuştur. Z-skor ile normalize edilen verilerde kayda değer bir yükseliş görülmektedir. k=1 değeri için 'Minkowski' uzaklık ölçütü ile %81,9 doğruluk değerine ulaşıldığı görülmektedir. K değerinin artması ile genel doğruluk değerinin düştüğünü de görmekteyiz. En kötü sonucun k=20 için 'Chebychev' uzaklık ölçütünde %72 olduğu açıkça görülmektedir. Şekil 4'de min-max normalizasyon işlemi uygulanmasıyla elde edilen veriler üzerinden çıkarılan tüm öznetelikler kullanıldığında sonuçlar detaylı bir şekilde verilmiştir. Min-max normalizasyon işlemi z-skor normalizasyon işlemine göre az da olsa performans artışı sergilemiştir. Sonuçlara bakıldığında, k=1 değeri için 'Euclidean' uzaklık ölçütü ile %82,1 oranında doğruluk değerine ulaşıldığı görülmektedir. Bu sonuçlarda da k değerinin azalması genel performansı düşürmektedir. En kötü sonucun k=20 için 'Minkowski' uzaklık ölçütünde %71,2 olarak elde edildiği görülmektedir.



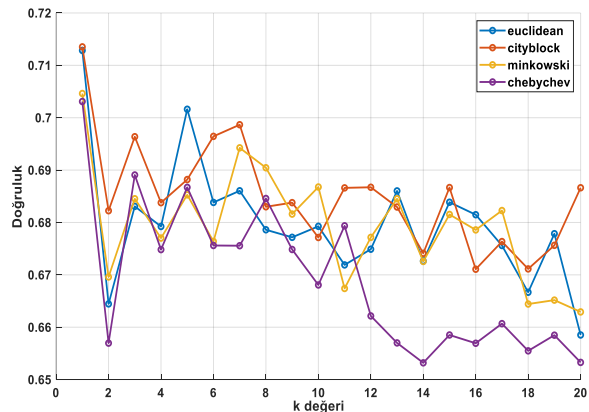
Şekil 3. Farklı uzaklık ölçütlerine göre z-skor ile normalize edilmiş veriler için k değerlerine göre doğruluk grafiği.



Şekil 4. Farklı uzaklık ölçütlerine göre min-max ile normalize edilmiş veriler için k değerlerine göre doğruluk grafiği.

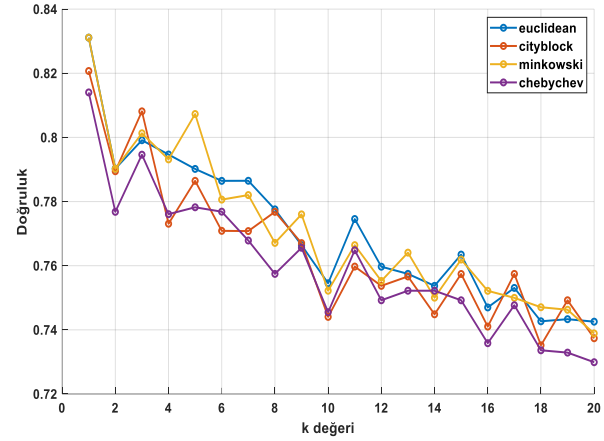
3.2. Geriye Doğru Eleme Yöntemi Kullanılarak Seçilen Özniteliklerin Performans Sonuçları

Çalışmanın bu aşamasında, elde edilen tüm öznitelikler geriye doğru eleme yöntemi ile seçilen öznitelikler kullanılarak çalışma sonuçları sunulmuştur. Normalizasyonlu ve normalizasyon olmadan seçilen özniteliklerin sayısı aynıdır ve toplam 13 adettir. Zaman alanı ölçümlerinden Ortalama, Standart Sapma, Basıklık, Çarpıklık, Hurst üssü olmak üzere 5 adet öznitelik seçilmiştir. Frekans alanı ölçümlerinde ise Delta, Teta, Gama, (Teta/Alfa), (Teta/Beta), (Teta + Alfa)/(Beta + Alfa), (Gama/Delta) ve (Gama+Beta)/(Delta+Alfa) olmak üzere 8 adet öznitelik seçilmiştir. Öznitelik seçiminin sonuçlara olumlu etkileri olmuştur. Normalize edilmemiş veriler üzerinden çıkarılan 13 öznitelik ile elde edilen sonuçlar Şekil 5'de verilmiştir. Genel doğruluk değerinin %54,6'dan %71,4'e çıktığı görülmektedir. En yüksek sonuç k değerinin 1 olduğu ve 'Cityblock' uzaklık ölçütünde %71,4 olarak elde edildiği görülmektedir. Fakat normalize edilmemiş verilerden elde edilen sonuçların hala düşük olduğu görülmektedir. K değerinin artması ile birlikte genel doğruluk değerinin düştüğü açıkça ifade edilebilir. En kötü sonucun k değerinin 20 olduğu 'Chebychev' uzaklık ölçütündedir. Burada elde edilen sonuç %65,4'tür.



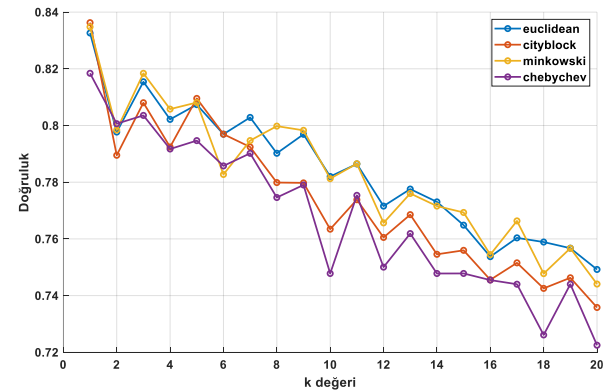
Şekil 5. Farklı uzaklık ölçütlerine göre normalize edilmemiş veriler üzerinden seçilen öznitelikler için k değerlerine göre doğruluk grafiği.

Min-max yöntemi sonucunda elde edilen özniteliklere yapılan seçim ile kullanılan 13 özniteliklerin performansı Şekil 6'da verilmiştir. Z-skor yöntemine göre az daha düşük çıkan doğruluk değeri %83,2'dir. Bu sonuç k=1 için 'Minkowski' uzaklık ölçütünde elde edilmiştir. Öznitelik seçiminin sonuca pozitif yansıdığı görülmektedir.



Şekil 6. Farklı uzaklık ölçütlerine göre min-max ile normalize edilmiş veriler üzerinden seçilen öznitelikler için k değerlerine göre doğruluk grafiği.

Son olarak, Z-skor ile normalize edilmiş veriler üzerinden elde edilen özniteliklerden seçilen 13 adet ölçümden elde edilen sonuçlar Şekil 7'de sunulmuştur. Sonuçların az da olsa arttığı söylenebilir. Öznitelik seçimi yapılmadığında %81,9 olan doğruluk değeri burada %83,6 olarak elde edilmiştir. %1,7 oranında bir artış gözlenmiştir. En yüksek sonucun k=1 için 'Cityblock' uzaklık ölçütünde %83,6 olarak elde edildiği görülmektedir. Burada da k değerine göre doğruluk değerinin azaldığını açıkça görülmektedir.



Şekil 7. Farklı uzaklık ölçütlerine göre z-skor ile normalize edilmiş veriler üzerinden seçilen öznitelikler için k değerlerine göre doğruluk grafiği.

Tablo 2'de sunulan literatür çalışmaları başta olmak üzere EEG sinyallerinin SZ teşhisindeki etkinliği konusunda pek çok çalışma mevcuttur. Bu çalışmalar, EEG'nin SZ hastalarında karakteristik değişimler içerdiğini göstermekte ve farklı öznitelik çıkarma ile sınıflandırma yöntemleri kullanarak yüksek başarı oranları elde etmektedir. Ancak, mevcut çalışmada kullanılan k-NN

algoritmasının doğruluk oranı, literatürde bildirilen derin öğrenme veya gelişmiş öznitelik çıkarma yöntemlerine kıyasla daha düşük kalmıştır. Bunun başlıca nedenleri şunlardır:

- Derin öğrenme tabanlı model ve yaklaşımlar, EEG sinyallerinden daha karmaşık ve yüksek seviyeli özellikleri otomatik olarak öğrenebilmektedir. Geleneksel öznitelik çıkarma yöntemleri belirli varsayımlara dayandığından, öğrenilen özniteliklerin veri kümesi üzerindeki ayırım gücü sınırlı kalabilmektedir.
- Daha gelişmiş öznitelik çıkarma yöntemleri, yalnızca frekans tabanlı ölçümlerle değil, zaman-frekans analizi, faz senkronizasyonu ve bağlantısallık ölçümleri gibi daha karmaşık biyomarker'lar ile de desteklenmektedir.
- k-NN algoritması, küçük veri kümelerinde etkili olabilmektedir. Ancak büyük ve yüksek boyutlu veri kümelerinde performansı düşebilir. Uzaklık temelli bir yöntem olduğu için, veri dağılımına ve gürültüye karşı oldukça hassastır.

Bu bağlamda, çalışmada kullanılan yöntemlerin geliştirilmesi için alternatif öznitelik çıkarma ve sınıflandırma yaklaşımlarına yönelmek faydalı olacaktır. Özellikle, zaman-frekans analizi, dalgacık dönüşümleri veya derin öğrenme tabanlı özellik çıkarımı tekniklerinin kullanılması, SZ teşhisinde daha yüksek doğruluk oranları sağlayabilir. Bunun yanı sıra, genelleme kabiliyetinin artırılması için daha büyük ve çeşitli veri kümeleri ile

modelin eğitilmesi gereklidir. Çalışmada kullanılan veri kümesi 84 EEG kaydından oluşmaktadır ve sınırlı bir hasta grubunu temsil etmektedir. Büyük ölçekli veri setlerinin kullanılması, modelin farklı hasta popülasyonları üzerindeki etkinliğini değerlendirmek ve klinik uygulanabilirliği artırmak için kritik öneme sahiptir. Özellikle, farklı yaş grupları, hastalık evreleri ve ilaç kullanım durumlarını içeren veri setleri, modelin sağlamlığını test etmek açısından gereklidir. Son olarak, elde edilen özniteliklerin şizofreni tanısındaki rolü değerlendirildiğinde, özellikle belirli frekans bantlarındaki değişimlerin hastalığın nörofizyolojik temelleri ile ilişkili olduğu görülmektedir. Önceki çalışmalar da (Santos-Mayo vd., 2016; Das ve Pachori, 2021) düşük frekans bantlarının (delta ve teta) SZ hastalarında belirgin şekilde değiştiğini ve bu değişimlerin tanısınal değer taşıdığını göstermektedir. Bu çalışmada da, özellikle frekans tabanlı oranların SZ teşhisinde belirli bir ayırt ediciliğe sahip olduğu gözlemlenmiştir. Ancak, klinik olarak tanı koyma sürecinde bu özniteliklerin tek başına yeterli olup olmadığı daha geniş çaplı çalışmalarla doğrulanmalıdır. Özetle, bu çalışma EEG tabanlı özniteliklerin SZ teşhisinde kullanılabilirliğini ortaya koyarken, kullanılan normalizasyon yöntemlerinin ve sınıflandırma algoritmasının parametrelerinin sonuçları değiştirdiği görülmektedir. Derin öğrenme temelli yaklaşımlar ve daha büyük veri kümeleri ile yapılacak ileri çalışmaların modelin genel başarımını artırabileceğini göstermektedir.

Tablo 2. EEG Verileri ile SZ Tespiti Yapan Çalışmaların Karşılaştırılması.

Yazar/yıl	Data	Yöntem	Doğruluk(%)
Kim vd., 2015	Sağlıklı: 90 SZ: 90	Spektral Yaklaşım, Alt bant analizi + ROC	62,20
Dvey-Aharon vd., 2015	Sağlıklı: 25 SZ: 25	Time-frequency feature optimization-based feature extraction + LDA	88,70
Piryatinska vd., 2017	Sağlıklı: 39 SZ: 45	Epsilon-complexity function + RF	85,30
Santos-Mayo vd., 2016	Sağlıklı: 31 SZ: 16	Zaman-frekans alanı analizi, Kanal kümeleme tekniği+SVM, ÇKAA	93,40
Das and Pachori, 2021	Sağlıklı: 14 SZ: 14	Hjorth moments on multivariate IMFs+SVM	98,90
Türk vd., 2024	Sağlıklı: 39 SZ: 45	Discrete Wavelet Transform + Resnet52	92,94
Bu çalışma	Sağlıklı: 39 SZ: 45	Zaman ve frekans alanı ölçümleri,geriye doğru eleme, normalizasyon+k-NN	83,62

4. Sonuç

Bu çalışmada, SZ hastalarının tespitinde EEG verileri kullanılarak elde edilen zaman ve frekans alanı ölçümlerinin farklı uzaklık ve k değerleri için k-NN algoritmasının performansları incelenmiştir. Veriler iki farklı normalizasyon işlemi ile normalize edilmiştir. Geriye doğru eleme yöntemi ile öznitelik seçimi yapıp

sonuçlar tekrarlanmıştır. Z-skor normalizasyon uygulanmış veriler ile elde edilen özniteliklere geriye doğru eleme yöntemi uygulandığında k=1 ve 'Cityblock' uzaklık ölçütü için en yüksek %83,6 doğruluk değerine ulaşılmıştır. Z-skor normalizasyonu, veriyi ortalaması sıfır ve standart sapması bir olacak şekilde dönüştürmesi ve tüm özellikleri dengeli şekilde değerlendirmesine olanak

sağladığı için daha yüksek başarımlar gösterdiği düşünülmektedir. Ayrıca 'Cityblock' uzaklık ölçütü ise mesafeyi yalnızca koordinat eksenleri boyunca hesaplayan bir yaklaşım sunması ve çıkarılan öznelitliklerin yapısına bağlı olarak daha uygun bir mesafe ölçümü sunması bu çalışma için daha uygun olduğu değerlendirilmektedir. Elde edilen bu bulguların literatüre katkı sağlayacağı düşünülmektedir. Gelecek çalışmalarda, literatüre göre daha yüksek performans elde edebilmek için, frekans alanı ölçümleri ve lineer olmayan öznelitlik sayısı arttırılacak, hesaplama yükü fazla olan algoritmalar ile sonuçlar denenecektir. Detaylı parametrik çalışmalar gerçekleştirilip sunulacaktır.

Katkı Oran Beyanı

Yazarların katkı yüzdesi aşağıda verilmiştir. Yazarlar makaleyi incelemiş ve onaylamıştır.

	A.N.	B.S.
K	50	50
T	50	50
Y	60	40
VTI	40	60
VAY	50	50
KT	50	50
YZ	50	50
KI	50	50
GR	50	50
PY	50	50
FA	50	50

K= kavram, T= tasarım, Y= yönetim, VTI= veri toplama ve/veya işleme, VAY= veri analizi ve/veya yorumlama, KT= kaynak tarama, YZ= Yazım, KI= kritik inceleme, GR= gönderim ve revizyon, PY= proje yönetimi, FA= fon alımı.

Çatışma Beyanı

Yazarlar bu çalışmada hiçbir çıkar ilişkisi olmadığını beyan etmektedirler.

Etik Onay Beyanı

Bu araştırmada hayvanlar ve insanlar üzerinde herhangi bir çalışma yapılmadığı için etik kurul onayı alınmamıştır.

Kaynaklar

Akcan G, Taşören AB. 2020. Genç yetişkinlerde çocukluk çağı olumsuz yaşantıları, öz-şefkat ve duygu düzenleme becerileri depresyon belirtilerini yordar mı? *Bogazici Univ J Educ*, 37(2): 59-80.

Aktay M, Hızlı Sayar G. 2021. Psikiyatrik bozuklukların psikososyal yönü. *Kıbrıs Türk Psikiyatri ve Psikoloji Derg*, 3(1): 48-55.

Bashivan P, Rish I, Heisig S. 2016. Mental state recognition via wearable EEG. *arXiv*, arXiv:1602.00985.

Birnbaum R. 2023. Rediscovering tandem repeat variation in schizophrenia: challenges and opportunities. *Transl Psychiatry*, 13(1): 402.

Borisov SV, Kaplan AY, Gorbachevskaya NL, Kozlova IA. 2005. Analysis of EEG structural synchrony in adolescents with schizophrenic disorders. *Human Physiol*, 31: 255-261.

Boz C, Özdemir M, Çalgı B. 2020. Mental hastalıkların prevalansına göre OECD ülkelerinin çok boyutlu analizi ve MOORA yöntemi ile sıralanması. *Gümüşhane Üniv Sosyal Bilim Derg*, 11: 245-256.

da Silveira TL, Kozakevicius AJ, Rodrigues CR. 2016. Automated drowsiness detection through wavelet packet analysis of a single EEG channel. *Expert Syst Appl*, 55: 559-565.

Das K, Pachori RB. 2021. Schizophrenia detection technique using multivariate iterative filtering and multichannel EEG signals. *Biomed Signal Process Cont*, 67: 102525.

Dvey-Aharon Z, Fogelson N, Peled A, Intrator N. 2015. Schizophrenia detection and classification by advanced analysis of EEG recordings using a single electrode approach. *PLoS One*, 10(4): e0123033.

Eoh HJ, Chung MK, Kim SH. 2005. Electroencephalographic study of drowsiness in simulated driving with sleep deprivation. *Int J Ind Ergon*, 35(4): 307-320.

Erdoğan YE, Narin A. 2021. COVID-19 detection with traditional and deep features on cough acoustic signals. *Comput Biol Med*, 136: 104765.

Farahdiba S, Kartini D, Nugroho RA, Herteno R, Saragih TH. 2023. Backward elimination for feature selection on breast cancer classification using logistic regression and support vector machine algorithms. *IJCCS*, 17(4): 429-440.

Hosseini SA, Naghibi-Sistani MB. 2011. Emotion recognition method using entropy analysis of EEG signals. *Int J Image Graphics Signal Process*, 3(5): 30-38.

Jap BT, Lal S, Fischer P, Bekiaris E. 2009. Using EEG spectral components to assess algorithms for detecting fatigue. *Expert Syst Appl*, 36(2): 2352-2359.

Jin H, Mosweu I. 2017. The societal cost of schizophrenia: A systematic review. *Pharmacoecon*, 35(1): 25-42.

Kamble A, Ghare P, Kumar V. 2022. Machine-learning-enabled adaptive signal decomposition for a brain-computer interface using EEG. *Biomed Signal Process Control*, 74: 103526.

Karal E, Atak H. 2022. Çocukluk çağı ruhsal travmaları üzerine kavramsal bir çalışma. *Muş Alparslan Univ J Fac Educ*, 2(1), 82-103.

Kim JW, Lee YS, Han DH, Min KJ, Lee J, Lee K. 2015. Diagnostic utility of quantitative EEG in un-medicated schizophrenia. *Neurosci Lett*, 589: 126-131.

Larson MK, Walker EF, Compton MT. 2010. Early signs, diagnosis and therapeutics of the prodromal phase of schizophrenia and related psychotic disorders. *Expert Rev Neurother*, 10(8): 1347-1359.

Lotte F, Bougrain L, Cichocki A, Clerc M, Congedo M, Rakotomamonjy A, Yger F. 2018. A review of classification algorithms for EEG-based brain-computer interfaces: a 10 year update. *J Neural Eng*.

Narin A. 2020. Parkinson hastalarının tespitinde karınca koloni algoritması ile seçilen öznelitliklerin performansına etkisi. *Düzce Univ J Sci Technol*, 8(4): 2443-2454.

Piryatinska A, Darkhovsky B, Kaplan A. 2017. Binary classification of multichannel-EEG records based on the ϵ -complexity of continuous vector functions. *Comput Methods Programs Biomed*, 152: 131-139.

Rizvi A, Safwi SR, Usmani MA. 2024. Schizophrenia: Disability, clinical insights, and management. *Cham: Springer Nature Bern, Switzerland*, pp: 1-12.

Santos-Mayo L, San-José-Revuelta LM, Arribas JL. 2016. A computer-aided diagnosis system with EEG based on the P3b wave during an auditory odd-ball task in schizophrenia. *IEEE Trans Biomed Eng*, 64(2): 395-407.

Şenyer Yapıcı İ, Uzun Arslan R. 2024. A machine learning-based decision support system design to predict maternal health risk groups during pregnancy. *BSJ Eng Sci*, 7(3): 509-520.

- Singh D, Singh B. 2020. Investigating the impact of data normalization on classification performance. *Applied Soft Computing*, 97: 105524.
- Subasi A. 2007. EEG signal classification using wavelet feature extraction and a mixture of expert model. *Expert Syst Applicat*, 32(4): 1084-1093.
- Sürücü M, İşler Y, Kara R. 2023. Paroksizmal atriyal fibrilasyonun 30 dakikalık kalp hızı değişkenliği analizi kullanılarak teşhisinde kalp hızı ve öznitelik normalizasyon yöntemlerinin etkisi. *Karaelmas J Sci Eng*, 13(1): 191-204.
- Türk Ö, Aldemir E, Acar E, Ertuğrul ÖF. 2024. Diagnosis of schizophrenia based on transformation from EEG sub-bands to the image with deep learning architecture. *Soft Comput*, 28(9): 6607-6617.