

# EPİLEPTİK EEG İŞARETLERİNİN SINIFLANDIRILMASINDA KARAR KURALLARI VE KARAR AĞAÇLARININ KULLANILMASI

Yılmaz KAYA<sup>1\*</sup>; Ö.Faruk ERTUĞRUL<sup>2</sup>; Ramazan TEKİN<sup>3</sup>

<sup>1</sup>*Siirt Üniversitesi, Müh. Mim. Fakültesi, Bilgisayar Mühendisliği Bölümü, 56100, Siirt*

<sup>2</sup>*Batman Üniversitesi, Müh. Mim. Fakültesi, Elektrik-Elektronik Mühendisliği Bölümü,  
72100, Batman*

<sup>3</sup>*Batman Üniversitesi, Müh. Mim. Fakültesi, Bilgisayar Mühendisliği Bölümü, 72100,  
Batman*

\*Yazışılan yazar e-posta:yilmazkaya1977@gmail.com

**Özet:** Epilepsi, genellikle nöbetler şeklinde kendini gösteren beyinde bir grup nöronun anormal aktivitesi sonucu dengesiz vücut hareketleri, bilinç kaybı veya kas kısılmalarına neden olan sinirsel bir hastalıktır. EEG işaretleri epilepsi hastalığının teşhisinde yaygın bir şekilde kullanılmaktadır. Bu çalışmada Ayrık Dalgacık Dönüşümü ile EEG işaretlerinin spektral çözümlemesi yapılmış ve elde edilen her bir spektral bileşenin istatistiksel bazı özellikleri baz alınarak karar ağaçları (ADTree, Functional Tree, J48, NBTree,) ve karar kuralları (Furia, DTNB, Jrip, PART, Ridor) yöntemleri ile sınıflandırma işlemi gerçekleştirilmiştir. Sonuçlara bakıldığında karar ağaçları ve karar kuralları epileptik EEG işaretlerin sınıflandırılmasında yüksek performans gösterdiği görülmüştür. Farklı kombinasyonlardaki EEG veri kümelerinin sınıflandırılmasında elde edilen sınıflandırma başarı oranları %96,6 ile %99,70 arasında değişim göstermiştir.

**Anahtar Kelimeler:** Epilepsi, EEG, karar ağaçları, karar kuralları.

## Epileptic EEG Detection Using Classifiers Decision Trees and Decision Rules

**Abstract :** Epilepsy is a well defined chronic neurological disorder usually characterized by seizures which are transient symptoms of abnormal or hyper synchronous neuronal activity in the brain. Electroencephalography (EEG) is one of the most common methods in diagnosis of epilepsy. In this study, decision trees (ADTree, Functional Tree, J48, NBTree) and decision rules (Furia, DTNB, Jrip, PART, Ridor) classifier algorithms were implemented for statistical features of EEG signal that obtained with discrete wavelet transformation. The obtained results show that decision rules and trees classifier algorithms performed a good performance and high prediction on epileptic EEG signals. For various combinations of epileptic EEG data sets the accuracy was ranged from 96.6% to 99.70%.

**Key Words:** Epilepsy, EEG, decision trees, decision Rules.

## 1. GİRİŞ

Epilepsi, beyinde bir grup nöronun anormal aktivitesi sonucu bilinç kaybı, dengelessiz vücut hareketleri, anormal duygusallık veya kas kasılması şeklinde kendini gösteren beynin önemli hastalıklarından biridir [1, 2]. Bu hastalığın teşhisinde EEG işaretlerinin analizi en iyi yöntemlerden biridir. EEG, spike ve keskin dalgalar gibi epileptik bölgenin belirlenmesi için gerekli önemli bilgiler içerebilir[3, 4]. EEG işaretlerini daha iyi analiz etmek ve bu analiz sonuçlarına göre bir sınıflandırılma ya da yorum yapmak için doğrusal ve doğrusal olmayan bir takım teknikler geliştirilmiştir. Spektral analiz EEG işaretlerinin karakteristiğini frekans düzlemlerinde incelemek amacıyla kullanılan en kolay yöntemlerden biridir. Epileptik EEG işaretlerinin frekans içeriği nöbet öncesinde ve nöbet sonrasında farklı olduğundan, frekansı zaman ölçeğinde analiz yapan Dalgacık Dönüşümü (DD) epileptik EEG işaretinin karakteristiğinin tespitinde ayrıcalıklı bir öneme sahiptir. Son yıllarda EEG işaretlerin sınıflandırılmasında yapay zekâ tekniklerinden faydalandığı görülmektedir. Epilepsi hastalığın teşhisi için kullanılan ve yüksek başarı gösteren makine öğrenmesi yöntemler aşağıdaki gibi listelenebilir.

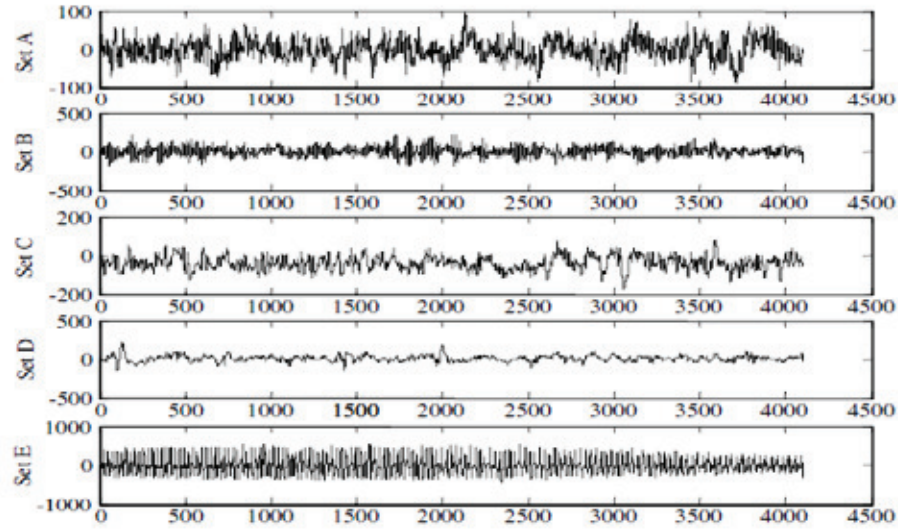
- Yapay Sinir Ağları [5]
- Karar Ağaçları [6]
- Mutlak Öğrenme Makinesi [7]
- İstatistiksel Yöntemler [8]
- Karar Destek Vektörleri [9]

Bu çalışmada EEG işaretlerinin sınıflandırılması için karar ağaçları (ADTree, BFTree, FT, J48, LADTree and LMT) ve karar kuralları (FURIA, DTNB, Jrip, OneR, PART, Ridor) yöntemleri kullanılmıştır. Ele alınan EEG veri setleri (A, B, C, D ve E) her biri 100 satırdan oluşmaktadır. A ve B setleri sırayla gözleri açık ve kapalı durumlarda sağlıklı gönüllülerden elde edilen EEG veri setleridir. C kümesi, hasta olan gönüllülerden nöbet öncesinde hipokampal oluşumlu karşı yarım kürenden ve D kümesi epileptojenik bölgeden elde edilmiştir. E kümesi ise hasta gönüllülerin kriz sırasındaki ölçümleridir. Elde edilen sonuçlara göre karar kuralları ve ağaçları EEG işaretlerinin sınıflandırılmasında önemli bir başarı gösterdiği görülmüştür. Karar ağaçları elde edilen başarı oranları %96,6 ile %99,70 arasında değişim gösterirken, karar kuralları ile elde edilen başarı oranları %95,3 ile %98,00 arasında değişim göstermiştir. Karar ağaçlarının karar kurallarına göre Epileptik işaretlerin sınıflandırılmasında daha başarılı olduğu görülmüştür.

## 2. MATERYAL VE YÖNTEM

### 1.1. Materyal

Veri seti olarak çalışmada kullanılan EEG işaretleri; Bonn Üniversitesi Epileptoloji bölümü veri tabanından alınmıştır [1]. Tüm veri, 5 kümeden oluşmaktadır (A-E) ve her küme 23.6 saniye süreli 100 tek kanal EEG segmenti içermektedir. Aşağıda Şekil 1’te beş kümeden elde edilen örnek işaretler görülmektedir. A ve B kümeleri sırasıyla gözler açıkken ve kapalıken sağlıklı olan 5 gönüllüden elde edilen yüzey EEG kayıtlarını içermektedir. C kümesi, hasta olan gönüllülerden nöbet öncesinde hipokampal oluşumlu karşı yarım kürenden ve D kümesi epileptojenik bölgeden elde edilmiştir. E kümesi ise hasta gönüllülerin kriz sırasındaki ölçümleridir. İşaretler 12 bit analog sayısal dönüştürücü ile dönüştürüldükten sonra bilgisayar ortamına 173,61 Hz örnekleme frekansı ile (toplamda 4097 örnek) aktarılmıştır. Epileptik özellikler kendini 30-40 Hz altındaki frekans bantlarında gösterdiğinden spektral aralığı 0.5-85 Hz olan işaretlere 0.53-40 Hz bant geçiren filtre uygulanmıştır.



Şekil 1. Beş farklı hasta grubuna ait işaret kayıtları.

## 1.2. Yöntem

### 1.2.1. Dalgacık Dönüşümü

EEG işaretleri durağan olmayan işaret türündendir. Bu tür işaretlerin spektral analizinde Fourier dönüşümünün iyi sonuç vermediği bilinmektedir. Bazı çalışmalarda kısa-süreli Fourier dönüşümü (KSFD) kullanılsa da özellikle seçilen pencere boyutunun sabit olmasından dolayı bu metot hem zaman hem de frekans boyutlarında *Uncertainty principle* gereği eşit bir çözünürlük ve dolayısıyla çok iyi bir sonuç vermeyebilir. Problemi kısmen gidermek için KSFD yerine Dalgacık Dönüşümü (DD) kullanılmıştır. DD yüksek frekanslar için küçük ve düşük frekanslar için büyük boyutlu pencere kullandığından zaman ve frekans açısından en uygun çözünürlüğü sağlamaya çalışır [10, 11]. Aşağıda sürekli dalgacık dönüşüm denklemi verilmiştir,

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

Burada işaretin kendisi, dalgacık dönüşümü, ana dalgacık fonksiyonu,  $a$  ölçek ve  $b$  zamanda kaydırma parametreleridir.

anadalgacık  $a$  ve  $b$  ölçeklerine göre elde edilmektedir:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi \left( \frac{t-b}{a} \right) \quad (2)$$

ana dalgacık hem frekans hemde zaman çözünürlüğünün uygun olmasını sağlar.  $a$  ölçeği büyük bir değer seçildiğinde ana dalgacık genişler ve düşük frekanslı detaylar elde edilirken,  $a$  ölçeği küçük bir değer seçildiğinde ise ana dalgacık daralır ve yüksek frekanslı detaylar elde edilir.

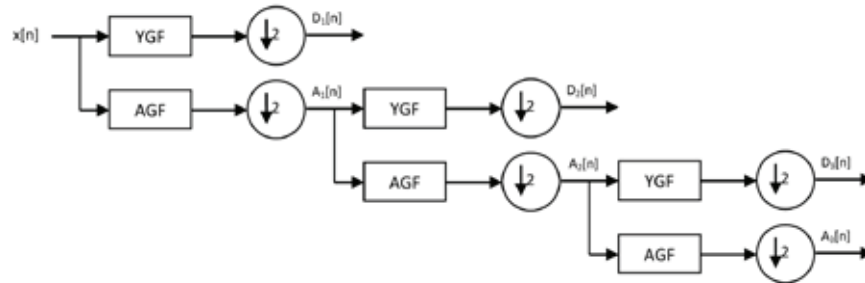
Dalgacık dönüşümünün sürekli veya ayrık formları kullanılabilir. Ancak bu çalışma için Sürekli dalgacık dönüşümü sinyali tanımlamak için gerekenden fazla katsayı üretir ve sinyalin yeniden oluşturulmasında tüm bu parametreler gerekli olduğundan aşırı hesaplama zorluğuna neden olur. Bu yüzden dönüşüm ve ölçek değişimini kısıtlayan Ayrık Dalgacık Dönüşümünü (ADD) kullanmak uygulamada daha büyük kolaylık sağlamaktadır. Sürekli dalgacık dönüşümünde ölçek ve kaydırma parametreleri  $a=2^j$  ve  $b=k.2^j=k.a$  ile ayrık hale getirilerek ADD temel ifadesi aşağıdaki gibi ifade edilebilir:

$$W(j, k) = \sum_j \sum_k x(n) 2^{-j/2} \psi(2^{-j}n - k) \quad (3)$$

Burada  $x(n)$  işaretin kendisi ve ana dalgacıdır. Logaritmik ve dengeli yapıya sahip çok frekanslı süzgeç banklarıyla AD analizi yapılabilir. ADD ile işaret yüksek geçiren filtreye sokularak ayrıntı (detail) katsayıları ve alçak geçiren filtreye sokularak yaklaşık (Approximate) katsayıları elde edilir. Bu işlemler elde edilen ayrıntı veya yaklaşık frekans bantlarında istenilen çözünürlük elde edilinceye kadar Şekil 1 de gösterildiği gibi tekrar edilir. Şekilde de görüldüğü gibi ayrıntı  $x[n]$  sinyali yüksek-geçiren filtreden (YGF) geçerek detay katsayıları ( $D_i[n]$ ) ve alçak-geçiren filtreden (AGF) geçerek yaklaşık katsayıları ( $A_i[n]$ ) elde edilir. Bu ayrıştırma işleminin zaman alanındaki matematiksel ifadesi  $g[n]$  ve  $h[n]$  sırasıyla yüksek-geçiren ve alçak-geçiren filtreler olmak üzere aşağıdaki gibi ifade edilir:

$$\begin{aligned} h_{i+1}(k) &= \sum_n h_i[n] h[2k - n] \\ g_{i+1}(k) &= \sum_n h_i[n] g[2k - n] \end{aligned} \quad (4)$$

Her ayrıştırma seviyesinde yarım-band filtreler frekans bandının yarısını içeren sinyaller oluşmasını sağlar.



Şekil 1. Ayrık Dalgacık Dönüşümünde Yüksek-Geçiren Filtre (YGF) ve Alçak-Geçiren Filtre ile alt bantlara ayrıştırma.

### 1.2.2. Karar Ağaçları ve Karar Kuralları

Karar ağaçları karar düğümleri, dalları ve yapraklarından oluşan olayların sonuçlarını bir ağacın yaprakları gibi göstermeyi sağlayan sınıflama modelleri arasında en çok kullanılan tekniklerden birisidir. Kolay anlaşılabilen

kurallar oluşturması sayesinde popülerlik kazanan tekniklerdir. Yapılarının kuruluşlarının ucuz olması, farklı sistemlerine kolayca entegre edilebilmeleri ve yorumlanmasının kolay olması sebebiyle yaygın olarak kullanılan karar ağaçları ve karar kuralları yaygın olarak aşağıdaki alanlarda kullanılmaktadır[12, 13].

- Sınıflara ayırma (segmentation) sayesinde belirli bir sınıfın üyelerinin belirlenmesi tespitinde
- Alt kümelere ilişkin ilişkilerin tanımlanmasında
- Gelecekteki olayların tahmin edilebilmesi için gerekli kuralların oluşturulmasında

Karar ağacı oluşturulduktan sonra bu ağaçtan karar kuralları türetilir ve test verisi üzerinde denir. Olumlu sonuç elde edilirse yeni gözlemleri sınıflandırmak için bu kurallar kullanılır.

Özellikle EEG işaretlerin sınıflandırılması, hangi değişkenin hedef özelliği etkilediği, önemli değişkenlerin belirlenmesi gibi alanlarda sıklıkla başvurduğumuz karar ağaçları insan kaynaklarından, tıba kadar birçok sektörde kullanılmaktadır. Birçok karar ağacı ve kuralı tekniği bulunmakla birlikte bu çalışmada, epileptik EEG işaretlerinin sınıflandırılmasında ADTree, BFTree, FT, J48, LADTree LMT karar ağaçları ve FURIA, DTNB, Jrip, OneR, PART ve RIDOR karar kuralları teknikleri kullanılmıştır.

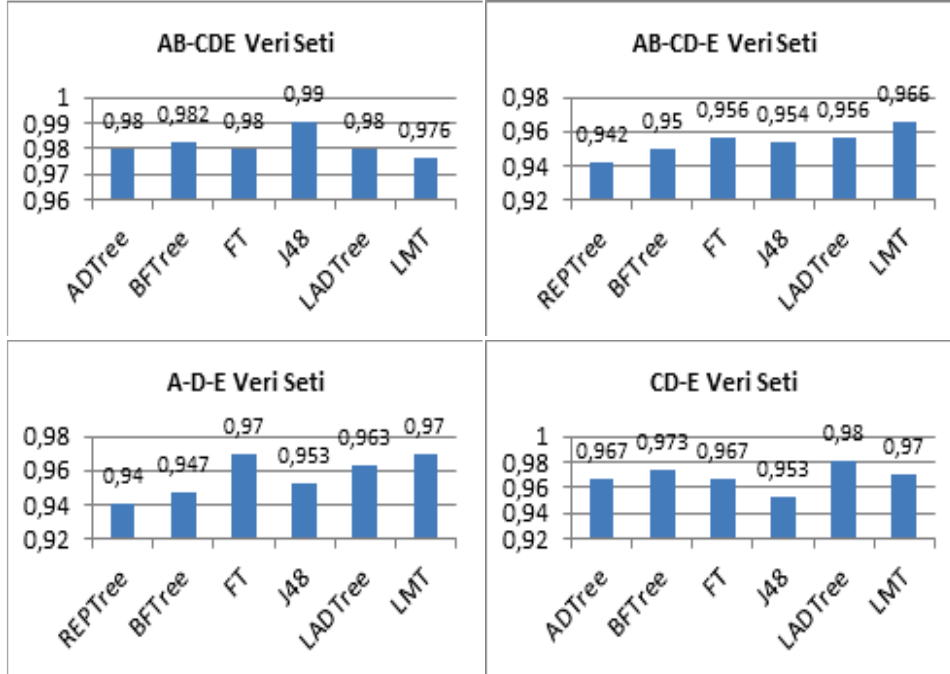
### 3. BULGULAR

Karar kuralları yöntemleri veri kümesinden *If-Then* kalıpları şeklinde bir dizi kural oluşturan yöntemlerdir. Karar ağaçları ise dallar ve yapraklardan oluşan olayların sonuçlarını bir ağacın yaprakları gibi göstermeyi sağlayan sınıflama modelleri arasında en çok kullanılan tekniklerdir. Bu çalışmada A,B,C,D ve E epileptik EEG veri setlerinin farklı kombinasyonlarından oluşan setler sınıflandırılmıştır. Karar ağaçları ile sınıflandırma işlemleri 10 katlı çapraz geçerlilik testine göre gerçekleştirilmiş ve elde edilen başarı oranları Tablo 1 ve Şekil 3'de gösterilmiştir.

**Tablo1.** Karar ağaçlarına ait performans değerleri.

Veri Seti	Karar Ağacı	Başarı Oranı %	ROC %	Duyarlılık %
AB-CDE	ADTree	98,00	99,50	98,00
	BFTree	98,20	98,90	98,20
	FT	98,00	98,60	98,00
	<b>J48</b>	<b>99,00</b>	<b>98,70</b>	<b>99,00</b>
	LADTree	98,00	99,40	98,00
	LMT	97,60	99,50	97,60
AB-CD-E	REPTree	94,20	97,30	94,30
	BFTree	95,00	96,60	95,00
	FT	95,60	97,70	95,60
	J48	95,40	96,60	95,40
	LADTree	95,60	98,90	95,60
	<b>LMT</b>	<b>96,60</b>	<b>99,20</b>	<b>96,60</b>
A-D-E	REPTree	94,00	96,50	94,00
	BFTree	94,70	96,10	94,80
	<b>FT</b>	<b>97,00</b>	<b>99,30</b>	<b>97,00</b>
	J48	95,30	95,60	95,40
	LADTree	96,30	98,50	96,50
	<b>LMT</b>	<b>97,00</b>	<b>99,70</b>	<b>97,00</b>
CD-E	ADTree	96,70	98,80	96,70
	BFTree	97,30	97,70	97,30
	FT	96,70	96,30	96,70
	J48	95,30	94,30	95,30
	<b>LADTree</b>	<b>98,00</b>	<b>98,70</b>	<b>98,00</b>
	LMT	97,00	99,30	97,00

Tablo 1'e göre, genel anlamda veri kümelerin sınıflandırılmasında yüksek başarı elde edildiği görülmektedir. Elde edilen sınıflandırma başarı oranları %96,6 ile %99,70% arasında değişim gösterdiği görülmektedir.



Şekil 3.Farklı epileptik EEG veri setleri için karar ağaçları performans değerleri.

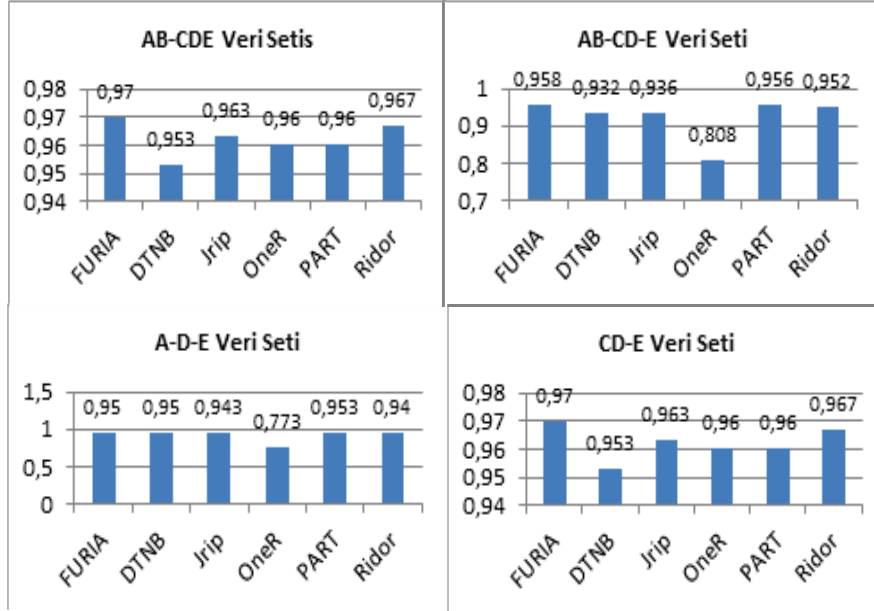
Karar kuralları ile elde edilen performans değerleri Tablo 2 ve Şekil 4’de gösterilmiştir.



**Tablo2.** Karar kurallarına ait performans değerleri.

Veri Seti	Karar Kuralı Modeli	Başarı Oranı %	ROC %	Duyarlılık%
AB-CDE	<b>FURIA</b>	<b>97,00</b>	<b>97,90</b>	<b>97,00</b>
	DTNB	95,30	98,50	95,30
	Jrip	96,30	96,30	96,30
	OneR	96,00	96,00	96,10
	PART	96,00	97,00	96,00
	Ridor	96,70	96,50	96,70
AB-CDE-E	<b>FURIA</b>	<b>95,80</b>	<b>98,00</b>	<b>95,80</b>
	DTNB	93,20	98,90	93,20
	Jrip	93,60	96,30	93,70
	OneR	80,80	84,30	81,80
	PART	95,60	96,70	95,60
	Ridor	95,20	96,30	95,20
A-D-E	FURIA	95,00	97,30	95,00
	DTNB	95,00	98,40	95,00
	Jrip	94,30	96,20	94,40
	OneR	77,30	83,00	77,80
	<b>PART</b>	<b>95,30</b>	<b>95,70</b>	<b>95,40</b>
	Ridor	94,00	95,50	94,00
CD-E	<b>FURIA</b>	<b>97,00</b>	<b>97,90</b>	<b>97,00</b>
	DTNB	95,30	98,50	95,30
	Jrip	96,30	96,30	96,30
	OneR	96,00	96,00	96,10
	PART	96,00	97,00	96,00
	Ridor	96,70	96,50	96,70

Tablo 2'ye bakıldığında karar kuralları ile elde edilen performans değerlerin %95,3 ile %97,00 arasında değişim gösterdiği görülmektedir. Karar kurallarına ait performanslar Şekil 4'te verilmiştir.



Şekil 4. Farklı epileptik EEG veri setleri için karar kuralları performans değerleri.

#### 4. TARTIŞMA VE SONUÇLAR

Sınıflandırma algoritmaları, tıbbi tanı sistemlerinde medikal verilerin daha kısa sürede incelenmesini sağlar. Bu çalışmada Epileptik EEG işaretlerin karar ağaçları ve karar kuralları ile tanı ve sınıflandırılması gerçekleştirilmiştir. Farklı epileptik EEG veri setleri için karar ağaçları ve karar kuralları yöntemleri ile elde edilen tanı performans değerleri %95,3 ile %99,70 arasında değişim göstermiştir. Bu çalışmada elde edilen sonuçlara göre, önerilen yöntemlerin tıbbi karar destek sistemlerinde uzmanlara karar vermede yardımcı olacağı düşünülmektedir.

#### 5. KAYNAKLAR

- [1] Andrzejak R., Lehnertz K., Mormann F., Rieke C., David P., Elger C., 2001. Indications of nonlinear deterministic and finite-dimensional structures in time series of brain electrical activity: Dependence on recording region and brain state. *Phys. Rev. E* 64, 061907.
- [2] Fisher R., van Emde Boas W., Blume W., Elger C., Genton P., Lee P., Engel J., 2005. Epileptic seizures and epilepsy: definitions proposed by the International League Against Epilepsy (ILAE) and the International Bureau for Epilepsy (IBE). *Epilepsia* 46(4): 470-2.
- [3] Ropper A., Brown R. H., Principles of neurology (8th ed.). Boston, USA: McGraw-Hill. 2005.

- [4] Altunay S., Telatar Z., Erogul O., 2010. Epileptic EEG detection using the linear prediction error energy. *Expert Systems with Applications*, 37(2): 5661-5665.
- [5] Ling G., Daniel R., Julián D., Juan R. R., Alejandro P., 2010. Automatic epileptic seizure detection in EEGs based on line length feature and artificial neural networks. *Journal of Neuroscience Methods* 191, 101-109.
- [6] Polat K., Günes S., 2007. Classification of epileptiform EEG using a hybrid system based on decision tree classifier and fast Fourier transform. *Applied Mathematics and Computation*. 187(2):1017-26.
- [7] Yuan Q., Weidong Z., Shufang L., Dongmei C., 2011. Epileptic EEG classification based on extreme learning machine and nonlinear features. *Epilepsy Research*. 96(1-2): 29-38.
- [8] Subasi A., 2006. Automatic detection of epileptic seizure using dynamic fuzzy neural networks. *Expert Systems with Applications*, 31(2):320-8.
- [9] Übeyli E. D., 2008. Analysis of EEG signals by combining eigenvector methods and multiclass support vector machines. *Computers in Biology and Medicine*, 38(1): 14-22.
- [10] Daubechies I., 1990. The wavelet transform, time-frequency localization and signal analysis, *IEEE Transactions on Information Theory*, 36(5), 961-1005.
- [11] Englehart K., Hudgins B., Parker P.A., 2001. A Wavelet-based continuous classification scheme for multifunction myoelectric control, *IEEE Trans. on Biomedical Engineering*, 48:302-311.
- [12] Frank E., Witten I. H., 1998. Generating Accurate Rule Sets Without Global Optimization. In: *Fifteenth International Conference on Machine Learning*, 144-151.
- [13] Karahoca A., Karahoca D., Aydin, N., 2009. Benchmarking the Data Mining Algorithms with Adaptive Neuro-Fuzzy Inference System in GSM Churn Management. *Data Mining and Knowledge Discovery in Real Life Applications*. pp. 438.