



EEG Sinyallerinden Trigonometrik Dönüşüm Esaslı Çıkarılan Öznitelikler ile Farklı Stres Kaynaklarının Sınıflandırılması

Classification of Different Stress Sources by Extracting Features Based on Trigonometric Transformation from EEG Signals

Müslüm Serhat ÜNVER, Önder AYDEMİR

Elektrik Elektronik Mühendisliği
Karadeniz Teknik Üniversitesi
serhatunver@ktu.edu.tr, onderaydemir@ktu.edu.tr

Özet

Stres günümüzde olumsuz birçok olay ile birlikte tetiklenerek insanların psikolojik ve fiziksel tepkiler gösterdiği bir durum olarak kabul edilmektedir. İnsanlarda stres seviyesinin erken tespitinin yapılmadığı durumlarda uyku bozuklukları, depresyon vb. hastalıklara sebebiyet vermektedir. Yapılan çalışmalar incelendiğinde strestin belirlenmesi için fizyolojik sinyallerin kullanımı oldukça yaygındır. Ancak çalışmaların çoğunda stres yaklaşımlarının gevşeme durumlarıyla sınıflandırıldığı görülmektedir. Bu çalışmada diğer çalışmalara göre daha zor bir problem olan farklı stres yaklaşımlarının birbirleriyle sınıflandırılması yapılmıştır. 40 denekten farklı stres durumları oluşturularak elde edilen EEG verilerinin trigonometrik dönüşüm, dalgacık dönüşümü, medyan, toplam, çarpıklık ve standart sapma yöntemleriyle öznitelikleri çıkarılmıştır. Trigonometrik dönüşümler ve trigonometrik dönüşümler olmadan elde edilen öznitelikler 4 farklı sınıflandırıcı ile sınıflandırılmış ve birbirleriyle kıyaslaması yapılmıştır.

Abstract

In cases where the stress level in humans is not detected early, it causes diseases such as sleep disorders, depression, etc. When studies are examined, it is quite common to use physiological signals to determine stress. However, in most of the studies, it is seen that stress approaches are classified by relaxation states. In this study, different stress approaches were classified with each other, which is a more difficult problem than other studies. 40 subjects' EEG data were obtained by creating different stress states and their features were obtained by trigonometric transformation, wavelet transform, median, sum, skewness and standard deviation methods. The features obtained with and without trigonometric transformations were classified with 4 different classifiers and compared.

1. Giriş

Stres; 17. Yüzyılın başlarında tanımlanmaya başlanmış ve günümüzde ailede yaşanan sağlık durumları, çalışma hayatındaki zorluklar, uyku problemleri, yeme içme düzeninde gerçekleşen değişiklikler, alkol ve madde bağımlılığı gibi durumlar ile birlikte tetiklenerek insanların psikolojik ve fiziksel tepkiler gösterdiği

bir durum olarak kabul edilmektedir. Uzman hekimlerin strese maruz kalan kişilere uyguladıkları anketler, ikili görüşmeler, davranış gözlemleri vb. teknikler stresin tespit edilmesi için sıklıkla kullanılan yöntemlerdir. Ancak Fernandez ve arkadaşlarının yaptığı çalışmalar sonucunda bazı kişilerin stresli durumlarda hiçbir belirti veya farkedilmesi zor belirtiler gösterdikleri görülmüştür [1]. Stresin erken tespitinin yapılamadığı durumlarda strese maruz kalan kişilerde inme, uyku bozuklukları, depresyon, kalp krizi, işlerde performans düşüklüğü, hata yapma olasılığının artması vb. birçok hastalığın görülme riski meydana gelmektedir. Bu sebeple stresin objektif açıdan belirlenip erken tespiti ve tedavisinin yapılması önemli bir faktör olarak görülmektedir.

Elektroensefalogram (EEG), elektrokardiyogram (EKG), kalp atış hızı değişkenliği, elektromiyogram, tıbbi görüntüleme teknikleri gibi fizyolojik sinyaller kullanılarak stresin erken tespitinin yapılması mümkündür [2-5]. Yapılan çalışmalar ile sinyallerin, stresin altında yatan fizyolojik durumunun ortaya çıkarılmasıyla erken tespiti amaçlanmaktadır [6]. EEG sinyalleri; kayıt aşamasında kapsamlı bir donanıma ihtiyaç duymaması, düşük maliyetleri ve yüksek doğruluk oranları vermesi sebebiyle stres seviyelerinin tespitinde sıklıkla tercih edilen fizyolojik sinyallerdir. Şimdiye kadar aritmetik test, stroop renk - kelime testi, yansıyan ayna görüntüleri testi ve Montreal görüntüleme stres testi (MIST) gibi birçok birbirinden farklı stres yaklaşımı üzerinde çalışmalar yapılmıştır. Gonzalez-Carabarin ve arkadaşları stres seviyelerini belirlemek için, EEG ve EKG sinyallerine dayalı çoklu algılama sistemi önermişlerdir. 24 kişiden topladıkları, EEG ve EKG verileri eşzamanlı olarak kaydedilirken kişiler rahatlatma, aritmetik ve sol-sağ el kaldırma testlerine maruz bırakılır [7]. Bir diğer çalışmada ise, Altaf ve arkadaşları 30 katılımcıdan oluşan grubun aritmetik test soruları çözümü ve gevşeme durumunda kaydedilen EEG sinyallerine dayalı olarak iki seviyeli insan stresini sınıflandırmak için diğer özelliklerle birleştirildiğinde alfa/beta ve teta/beta oranlarının bant gücünün etkisini araştırmışlardır [8]. Başka bir EEG tabanlı çalışmada ise, Tsai ve arkadaşları 12 masa tenisi oyuncusu üzerinde rahatlatma dururumu, Stroop renk-kelime testi ve zihinsel aritmetik testleri ile stresli / stressiz durumlar oluşturmuşlar ve EEG verilerini toplayarak lojistik regresyon, destek vektör makinesi (DVM) ve karar ağacı C4.5 sınıflandırma algoritmaları ile karşılaştırmalar yapmışlardır [9]. Diğer bir çalışmada Perez-Valero ve arkadaşları 23 katılımcıdan oluşan

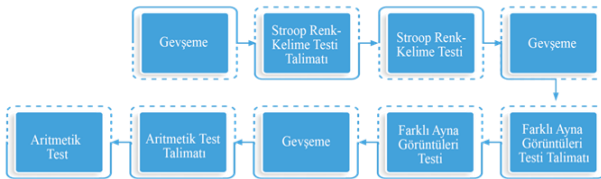
denek grubunun MIST ve ardından 360 derecelik bir sanal gerçeklik gevşeme deneyimi ile topladıkları EEG verilerini kullanarak EEG ve regresyon algoritmalarına dayalı nicel stres değerlendirilmesi için bir alternatif önermişlerdir [10]. Benzer EEG tabanlı çalışmada Kaminska ve arkadaşları 28 sağlıklı yetişkinden oluşan katılımcı grubu üzerinde katılımcıların stres düzeylerini sınıflandırmak için sanal gerçeklik kullanarak gevşeme ve Stroop renk-kelime testleri ile stresli ve stres olmayan durumları oluşturarak EEG sinyallerinin kullanımlarını araştırmışlardır [11]. Bu ve benzeri çalışmalarda görüldüğü üzere; literatürde çoğunlukla gevşeme durumuna karşın stres durumları sınıflandırılmıştır. Bu çalışmada ise farklı stres durumları yaratan görevler esnasında kaydedilen EEG verilerinin birbirleri ile sınıflandırılması üzerine çalışmalar yapılmıştır. Dolayısıyla farklı stres yaklaşımlarının birbirleriyle sınıflandırılması ile geçmişte üzerine çalışılan gevşeme durumu ve stres yaklaşımının birbirleri ile sınıflandırılması problemine göre daha zor olan bir problem bu çalışmada ele alınmıştır. Bu çalışmanın amacı ayna görüntüleri ve stroop renk-kelime testi ile oluşturulan stres seviyelerinin trigonometrik dönüşümler ile öznitelikleri çıkartılarak iki farklı stres yaklaşımının sınıflandırılmasıdır. Bu amaçla; 40 kişiden elde edilen EEG verileri normalleştirilerek trigonometrik dönüşümleri alınmış ve elde edilen sinyallerin 16 farklı ana dalgacık ile öz nitelikleri çıkarılmıştır. Öznitelikler k-en yakın komşuluk (k-EYK), Bayes, karar ağacı (KA) ve DVM ile sınıflandırılmıştır.

2. Materyal ve Yöntem

Materyal ve Yöntem bölümünde; çalışmada kullanılan veri seti, öznitelik çıkarma yöntemleri ve sınıflandırma yöntemlerine ait kısımlar detaylı olarak açıklanmıştır.

2.1. Veri seti tanımı

Bu çalışmada açık kullanıma sunulmuş stroop renk-kelime testi, aritmetik görev ve ayna görüntüleri kullanılarak kaydedilen EEG tabanlı veri seti kullanılmıştır [12]. Veri seti 14'ü kadın 26'sı erkek olmak üzere toplamda 40 deneğin 32 kanallı bir Emotiv Flex jel kiti ile birlikte EEG verilerinin kaydedilmesiyle oluşmuştur. Deney prosedürü şekil 1'de verilmiştir. Her bir denek için deney süresince toplamda 3 deneme kaydedilmiştir. Denemenin ilk aşamasında denekten 25 saniye boyunca gevşemesi istenir. Daha sonra Stroop renk-kelime testi talimatları 10 saniye boyunca deneye gösterilir ve 25 saniye boyunca testi yapması beklenir. Devamında denekten 5 saniye boyunca gevşemesi istenir. Bir sonraki aşamada deneye farklı ayna görüntüleri testi talimatları 10 saniye boyunca gösterilir ve 25 saniye boyunca testi yapması beklenir. Devamında denekten 5 saniye boyunca gevşemesi istenir. Son aşamada ise deneye 10 saniye boyunca aritmetik problem testi talimatları gösterilir ve 25 saniye boyunca testi yapması beklenir.



Şekil 1: Veri seti deney düzeneği

2.2. Öznitelik çıkarma

2.2.1. Trigonometrik dönüşüm

Bu çalışmada Min-Max normalizasyon yöntemi kullanılarak EEG verileri 0 – 1 aralığında normalize edilmiştir. Min-Max normalizasyon yöntemi, verilerdeki tüm ilişkileri tam olarak koruma avantajına sahiptir [13]. Maximum verinin alabileceği en yüksek değeri, minimum verinin alabileceği en düşük değeri ifade etmektedir. Min-Max normalizasyonunun matematiksel ifadesi Denklem 1'de verilmiştir. İlgili denklemde verilen x EEG verilerini tanımlamaktadır.

$$X' = \frac{x - \min(x)}{\max(x) - \min(x)} \quad (1)$$

0-1 aralığında normalize edilen EEG verileri matlab ortamında 0° ile 360° aralığına karşılık gelen $6.28(2\pi)$ ile normalize edilmiştir. Normalize edilen verilerin sinüs ve cosinüs dönüşümleri alınarak EEG verilerinin trigonometrik dönüşümleri elde edilmiştir. Trigonometrik dönüşümlerin matematiksel ifadeleri Denklem 2, Denklem 3 ve Denklem 4'te verilmiştir.

$$X' = X' * 6.28 \quad (2)$$

$$K' = \cos(X') \quad (3)$$

$$K'' = \sin(X') \quad (4)$$

2.2.2. Dalgacık dönüşümü

Bu çalışmada EEG verilerinin öznitelikleri dalgacık dönüşümü yöntemi ile hesaplanmıştır. Dalgacık dönüşümü; zaman fonksiyonu ile ölçeklendirme ve kaydırma işlemlerini yaparak durağan olmayan sinyallerde alçak ve yüksek frekans bileşenlerinin hesaplanmasını sağlayan bir yöntem olarak ön plana çıkmaktadır [14]. İlgili teknik düşük frekans seviyelerinde geniş, yüksek frekans seviyelerinde dar pencereler aralığında verilerin daha net bir şekilde incelenmesini sağlamaktadır. Dalgacık dönüşümünün matematiksel ifadesi Denklem 5 ve Denklem 6'da verilmiştir. İlgili denklemde verilen t zaman parametresini, s ölçek(frekans) parametresini, τ öteleme parametresini tanımlamaktadır.

$$\Psi_{\tau,s} = \Psi\left(\frac{t-\tau}{s}\right) \quad (5)$$

$$DD_x^\Psi = \frac{1}{\sqrt{s}} \int x(t) \Psi * \left(\frac{t-\tau}{s}\right) dt \quad (6)$$

2.2.3. Standart sapma

Standart sapma; dağılım varyasyonlarının karekökü olarak ifade edilmektedir [15]. Denklem 7'de standart sapmanın matematiksel gösterimi verilmiştir.

$$SS(x) = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2} \quad (7)$$

2.2.4. Medyan

Medyan; Olasılık teorisinde verilerin üst yarısı ile alt yarısını ayıran bir sayı olarak tanımlanmaktadır. Bir dağılımın, üstünde ve altında eşit sayıda değer bulunan orta değer olarak ifade edilmektedir [16].

2.2.5. Toplam

Toplam işaretlerin genliklerini toplamak için kullanılmaktadır. Denklem 8'de toplama ait matematiksel denklem verilmiştir.

$$K(x) = \sum_{i=1}^N X_i \quad (8)$$

2.2.6. Çarpıklık

Çarpıklık; skew(x) değerinin pozitif, negatif veya sıfır olmasına bağlı olarak değerin pozitif, negatif çarpık veya çarpık olup olmadıklarına karar vermektedir. Denklem 9'da çarpıklığa ait matematiksel gösterim verilmiştir. İlgili denklemde verilen σ standart sapmayı, μ ortalamayı ifade etmektedir.

$$K(x) = \frac{1}{N} \sum_{i=1}^N \frac{(x_i - \mu)^2}{\sigma} \quad (9)$$

2.3. Sınıflandırma

2.3.1. K-en yakın komşuluk

k-EYK algoritması sınıflandırma problemlerini çözen denetimli bir öğrenme algoritmasıdır. k-EYK yöntemi; kolay kullanılabilir olması ve üstün performansı sayesinde başarılı sonuçlar vermektedir. Veri sınıflandırması yapılırken mevcut veriler arasındaki mesafe ve benzerlikler dikkate alınarak sınıflandırma işlemi gerçekleştirilir. k-EYK algoritmasında parametreler arasındaki mesafe, ağırlıklandırma ve komşu sayısı(k) performans etki eden unsurlardır. Mesafe hesaplanmasında Öklid, Manhattan, Minkowski, Chebyshev ve Dilca en sık tercih edilen yöntemlerdir.

2.3.2. Bayes

Bayes; sınıflama etiketleri ile belirli bir özelliği kullanarak olasılık temellerine dayandıran sınıflandırma yöntemidir [17]. Bayes yöntemi basit ve yüksek verimli olması sebebiyle sınıflandırma problemlerinde sıklıkla tercih edilen bir yöntemdir.

2.3.3. Karar ağacı

KA algoritması; öğrenilen fonksiyonu bir karar ağacı şeklinde temsil eden ayrık değerli hedef fonksiyona yaklaşmak için kullanılan bir tekniktir. KA algoritması ile elde edilen örneklerin özellik değerlerine göre kökten yaprak düğümlerine doğru sınıflandırma işlemi gerçekleştirir. Yapısında, bir sınıfla etiketlenmiş yaprak düğümünden ve iki veya daha fazla alt ağaca bağlı bir test düğümünden oluşmaktadır. Algoritmada düğüm seviyelerinde nasıl en yüksek sınıflandırma sonucunun elde edileceğine odaklanır [18].

2.3.4. Destek vektör makineleri

DVM hem regresyon hem de örüntü tanıma işlemleri için kullanılan bir öğrenme algoritmasıdır. DVM; giriş verilerini yüksek boyutlu bir özellik alanında doğrusal olarak eşlemek için bir aygıt kullanılmaktadır. Karar fonksiyonları destek makinelerinin ağırlıkları, sayısının yanında destek vektör çekirdekleri ile alakalıdır. Sınırlı örneklem bilgisine göre model karmaşıklığı ve öğrenme yeteneği arasında optimal bir denge bulmaya çalışan DVM

yüksek boyutlu problemlerin çözümündeki avantajları sayesinde birçok alanda uygulanmaktadır [19].

3. Sonuçlar

Bu çalışmada ayna görüntüleri ve stroop renk-kelime testi ile oluşturulan iki farklı stres yaklaşımının sınıflandırılması için çalışmalar yapılmıştır. Sinyal işaretlerinin öznelikleri trigonometrik dönüşüm, dalgacık dönüşümü, medyan, toplam, standart sapma ve çarpıklık yöntemleri ile belirlenmiş ve öznelikleri çıkarılan sinyal işaretleri k-EYK, Bayes, KA ve DVM kullanılarak sınıflandırılmıştır. Ayna görüntüleri ve stroop renk-kelime testi verileri için yapılan trigonometrik dönüşümlerin 16 farklı dalgacık ile elde edilen sonuçları Çizelge 1'de verilmiştir. Sinüs dönüşümü için en yüksek sınıflandırma doğruluğu DVM ve Gaus3 ana dalgacığı kullanımları ile %72.50 olarak hesaplanmıştır. Gaus1 ana dalgacığında k-EYK %70.83, Gaus3 ana dalgacığında Bayes %68.33 ve Gaus 3 ana dalgacığında KA %65.83 ile en yüksek doğruluk oranlarına ulaşılmıştır. Cosinüs dönüşümü için en yüksek sınıflandırma doğruluğu k-EYK ve Gaus1 ana dalgacığı kullanımları ile %70.83 olarak hesaplanmıştır. Morlet ana dalgacığında Bayes %61.67, Gaus2 ana dalgacığında KA %67.50 ve Coif1 ana dalgacığında DVM %61.67 ile en yüksek doğruluk oranlarına ulaşılmıştır.

Ayna görüntüleri ve stroop renk-kelime testi verilerinin trigonometrik dönüşümleri alınmadan 16 farklı dalgacık ile çıkarılan özneliklerinin sınıflandırılması Çizelge 2'de verilmiştir. Trigonometrik dönüşümler olmadan elde edilen en yüksek sınıflandırma doğruluğu k-EYK, Coif2 ve Db3 ana dalgacıkları ile %72.50 olarak hesaplanmıştır. Sym3 ve Sym4 ana dalgacıkları ile Bayes %70.83, Gaus2 ana dalgacığı ile KA %71.67, Sym3 ve Sym4 ana dalgacıkları ile DVM %71.67 oranlarıyla en yüksek sınıflandırma doğrulukları elde edilmiştir.

Çizelge 2: Trigonometrik dönüşümler olmadan elde edilen sınıflandırma doğrulukları

Ana Dalgacık	k-EYK	Bayes	KA	DVM
Mor1	67.50	66.67	60.83	66.67
Mexh	66.67	65.00	62.50	63.33
Haar	65.83	69.17	58.33	66.67
Meyr	66.67	67.50	60.83	68.33
Db1	69.17	64.17	69.17	69.17
Db2	70.00	67.50	67.50	70.00
Db3	72.50	70.00	62.50	65.00
Sym2	71.67	69.17	57.50	70.83
Sym3	69.17	70.83	63.33	71.67
Sym4	70.00	70.83	60.83	71.67
Coif1	68.33	69.17	58.33	68.33
Coif2	72.50	63.33	60.83	63.33
Coif3	69.17	69.17	62.50	68.33
Gaus1	66.67	68.33	61.67	68.33
Gaus2	71.67	64.17	71.67	63.33
Gaus3	71.67	62.50	68.33	64.17

Çizelge 1: Trigonometrik dönüşümler ile elde edilen sınıflandırma doğrulukları

Ana Dalgacık	k-EYK		Bayes		KA		DVM	
	Sin	Cos	Sin	Cos	Sin	Cos	Sin	Cos
Mor1	63.33	65.83	55.00	61.67	55.00	61.67	51.67	50.83
Mexh	66.67	67.50	62.50	58.33	56.67	65.83	62.50	58.33
Haar	61.67	65.83	65.00	58.33	57.50	60.83	62.50	51.67
Meyr	65.00	65.83	60.83	58.33	62.50	59.17	57.50	51.67
Db1	62.50	63.33	66.67	60.00	59.17	65.00	64.17	52.50
Db2	68.33	62.50	60.83	51.67	62.50	65.83	64.17	52.50
Db3	62.50	64.17	68.33	55.00	58.33	55.00	68.33	60.00
Sym2	65.00	66.67	60.00	54.17	55.83	55.83	62.50	52.50
Sym3	65.83	66.67	65.00	55.00	57.50	54.17	66.67	57.50
Sym4	65.83	65.00	63.33	52.50	55.00	55.00	65.00	51.67
Coif1	65.00	65.00	63.33	60.83	59.17	59.17	64.17	61.67
Coif2	66.67	66.67	65.83	50.00	59.17	60.83	65.00	51.67
Coif3	65.00	63.33	63.33	60.83	56.67	56.67	70.83	51.67
Gaus1	67.50	70.83	53.33	63.33	58.33	63.33	52.50	60.83
Gaus2	70.83	67.50	67.50	60.00	61.67	67.50	65.83	60.83
Gaus3	70.00	68.33	68.33	58.33	65.83	63.33	72.50	60.83

4. Bulgular

Bu çalışmada EEG sinyallerinden trigonometrik dönüşüm esaslı çıkarılan öznitelikler ile farklı stres kaynaklarının sınıflandırılması yapılmıştır. Farklı stres kaynakları olarak belirlenen stroop renk-kelime testi ve yansıyan ayna görüntüleri testi verilerinin birbirleriyle trigonometrik dönüşüm esaslı sınıflandırılması sonucunda çıkan en yüksek sınıflandırma doğruluğu %72.50 olarak hesaplanmıştır. Önerilen yöntemle strese sebep olan farklı stres kaynaklarının tespit edilebileceği gösterilerek, stresin kaynak sebebinin belirlenebilmesi ve buna göre çözüm yollarının üretilmesinin alt yapısı gerçekleştirilmiştir.

5. Kaynaklar

[1] Fernández, José Raúl Machado, and Lesya Anishchenko. "Mental stress detection using bioradar respiratory signals." *Biomedical signal processing and control* 43 (2018): 244-249..

[2] De La Cruz, J., Shimizu, D., George, K., "EEG and fNIRS Analysis Using Machine Learning to Determine Stress Levels." In: *2022 IEEE World AI IoT Congress (AIoT)*. IEEE, 2022. p. 318-322.

[3] Shim, M., et al. "Enhanced Performance by Interpretable Low-Frequency Electroencephalogram Oscillations in the Machine Learning-Based Diagnosis of Post-traumatic Stress Disorder." *Frontiers in neuroinformatics* 16 (2022).

[4] Pourmohammadi, Sara, and Ali Maleki. "Stress detection using ECG and EMG signals: A comprehensive study." *Computer methods and programs in biomedicine* 193 (2020): 105482..

[5] Reinertsen, Erik, et al. "Heart rate-based window segmentation improves accuracy of classifying posttraumatic stress disorder using heart rate variability measures." *Physiological measurement* 38.6 (2017): 1061.

[6] Giannakakis, Giorgos, et al. "Review on psychological stress detection using biosignals." *IEEE Transactions on Affective Computing* 13.1 (2019): 440-460.

[7] Gonzalez-Carabarin, Lizeth, et al. "Machine Learning for personalised stress detection: Inter-individual variability of EEG-ECG markers for acute-stress response." *Computer Methods and Programs in Biomedicine* 209 (2021): 106314.

[8] ALTAF, Hunain, et al. Machine Learning Approach for Stress Detection based on Alpha-Beta and Theta-Beta Ratios of EEG Signals. In: *2021 13th International Conference on Information & Communication Technology and System (ICTS)*. IEEE, 2021. p. 201-206.

[9] Tsai, Yu-Hung, et al. "Analyzing Brain Waves of Table Tennis Players with Machine Learning for Stress Classification." *Applied Sciences* 12.16 (2022): 8052.

[10] Perez-Valero, Eduardo, et al. "Quantitative assessment of stress through EEG during a virtual reality stress-relax session." *Frontiers in Computational Neuroscience* (2021): 64.

[11] Kamińska, D., Smółka, K. and Zwoliński G., "Detection of Mental Stress through EEG Signal in Virtual Reality Environment." *Electronics* 10.22 (2021): 2840.

[12] Ghosh, R., et al. "SAM 40: Dataset of 40 subject EEG recordings to monitor the induced-stress while performing Stroop color-word test, arithmetic task, and mirror image recognition task." *Data in Brief* 40 (2022): 107772.

[13] Jayalakshmi, T., and A. Santhakumaran. "Statistical normalization and back propagation for classification." *International Journal of Computer Theory and Engineering* 3.1 (2011): 1793-8201.

[14] Polikar, R., The engineer's ultimate guide to wavelet analysis-the wavelet tutorial. available at <http://www.public.iastate.edu/~rpolikar/WAVELETS/WTtutorial.html>. (1996):

[15] Mutlag, W.K., et al. "Feature extraction methods: a review." In: *Journal of Physics: Conference Series*. IOP Publishing, 2020. p. 012028.

[16] Yang, J., Zhang, D., Yang, J., "Median LDA: a robust feature extraction method for face recognition." In: *2006 IEEE International Conference on Systems, Man and Cybernetics*. IEEE, 2006. p. 4208-4213.

[17] Tang, B., Kay, S., and He, H., "Toward optimal feature selection in naive Bayes for text categorization." *IEEE*

transactions on knowledge and data engineering 28.9
(2016): 2508-2521.

- [18] Quinlan, J. R., "Learning decision tree classifiers," *ACM Computing Surveys (CSUR)*, vol. 28(1), pp. 71-72, 1996
- [19] Zhang, L., et al. "Support vectors pre-extracting for support vector machine based on K nearest neighbour method." In: *2008 International Conference on Information and Automation*. IEEE, 2008. p. 1353-1358.