

Alkol Bağımlılığına Genetik Yatkınlığın EEG Verileriyle Analizi

Mohamed Hamdo^{1*}, Ahmad Akkach^{2*}

¹Necmettin Erbakan Üniversitesi, Mühendislik Fakültesi, Bilgisayar Mühendisliği Bölümü. Konya/TÜRKİYE

7goffy@gmail.com, ahmadalamer1999@gmail.com

Öz: Bu çalışma, alkol bağımlılığına genetik yatkınlığı olan bireylerle sağlıklı kontrol grubunun EEG (Elektroensefalogram) sinyallerini karşılaştırarak, alkolizmle ilişkili nörofizyolojik farklılıkları incelemeyi amaçlamaktadır. EEG-Alcohol veri kümesi kullanılarak gerçekleştirilen analizlerde, farklı uyaran türlerinin (tek, eşleşen ve eşleşmeyen görsel uyaranlar) beyin aktivitesi üzerindeki etkileri detaylı şekilde değerlendirilmiştir. Makine öğrenmesi algoritmaları (Random Forest, Logistic Regression, Gradient Boosting) ile yapılan sınıflandırma sonuçları, özellikle oksipital ve parietal bölgelerde alkolik bireylerde belirgin farklılıklar olduğunu göstermiştir. En başarılı performans %85.25 doğruluk oranı ile Random Forest modeli tarafından elde edilmiştir. Ayrıca, uyarılara verilen EEG yanıtlarının zamansal desenleri ve beyin bölgeleri arası korelasyonlar, alkolik bireylerde sağlıklı bireylere kıyasla anlamlı derecede farklılık göstermektedir. Elde edilen bulgular, alkol bağımlılığı riskinin erken dönemde belirlenmesi ve kişiselleştirilmiş müdahale stratejilerinin geliştirilmesi açısından önemli katkılar sunmaktadır.

Anahtar Kelimeler:

Alkol bağımlılığı, EEG analizi, genetik yatkınlık, makine öğrenmesi, Random Forest, görsel uyaran, beyin dalgaları, sınıflandırma, nörofizyolojik desenler, erken tanı.

1. Giriş

Alkol bağımlılığı, bireylerin sağlığı üzerinde yarattığı ciddi olumsuz etkilerin yanı sıra toplumsal yaşamı ve ekonomik kaynakları da önemli ölçüde tehdit eden karmaşık bir sorundur. Alkolizmin altında yatan genetik yatkınlık faktörlerinin belirlenmesi ve bu yatkınlığın nörofizyolojik temellerinin anlaşılması, erken tanı ve etkili koruyucu önlemler açısından büyük önem taşımaktadır. Bu bağlamda, genetik yatkınlığı bulunan bireyler ile sağlıklı kontrol grubunun EEG sinyalleri incelenerek, alkolizm riskine dair önemli ipuçlarının elde edilmesi hedeflenmektedir.

Çalışmada kullanılacak EEG-Alcohol veri kümesi, katılımcıların beyin dalgalarının, 64 elektrot aracılığıyla, 256 Hz örnekleme frekansında ve 1 saniyelik kayıtlar şeklinde elde edilmiş ölçümlerinden oluşmaktadır. Bu ölçümlerde katılımcılara tek veya çift görsel uyaran (S1 ve S2) sunulmuş, ikinci uyaran (S2) ise birinci uyaran ile eşleşen (matched) ya da eşleşmeyen (non-matched) koşullarda gösterilmiştir.

Bu çalışmanın temel amaçları; (1) alkolik ve kontrol gruplarının EEG kayıtları arasındaki farklılıkları belirlemek, (2) makine öğrenmesi yöntemleri kullanılarak iki grup arasındaki ayrımı gerçekleştirebilmek ve (3) sunulan uyaran türlerinin (tek uyaran, eşleşen ya da eşleşmeyen uyaran) beyin aktivitesi üzerindeki etkilerini detaylı olarak incelemektir.

Problem Tanımı: Problem, alkol bağımlılığı bulunan bireylerin EEG verilerinin sağlıklı kontrol grubuna göre ne ölçüde farklılık gösterdiğini belirlemek ve bu farklılıkları otomatik olarak saptayabilecek bir model geliştirmektir. Bu problemin çözümü, alkol bağımlılığının erken teşhisi ve genetik yatkınlığı olan bireylerde koruyucu girişimlerin zamanında uygulanabilmesi açısından önem taşımaktadır. EEG verilerinde ortaya çıkan anlamlı örüntülerin tespiti, gelecekte kişiselleştirilmiş tedavi yaklaşımlarının geliştirilmesine de zemin hazırlayabilir.

Araştırma Soruları:

1. Alkolik ve Kontrol Gruplarının EEG Örüntüleri Nasıl Farklılık Gösteriyor?
Bu soru, temel olarak iki grubun beyin dalgalarında (frekans bantları, zaman alanı özellikleri vb.) istatistiksel veya görsel anlamda belirgin bir ayrışma olup olmadığını araştırır. Bu ayrışmanın hangi elektrot bölgelerinde daha baskın olduğunu da inceleyebiliriz.
2. Makine Öğrenmesi Modelleri, Alkolik Bireyleri Kontrol Grubundan Ne Kadar Başarıyla Sınıflandırabilir?
Sınıflandırıcı algoritmalar (ör. SVM, Random Forest, Yapay Sinir Ağları) kullanılarak, EEG verileri üzerinden grup ayrımı yapmak mümkündür. Başarı oranlarını ölçerek, hangi algoritmanın veya hangi özellik setinin (frekans-temelli, zaman-temelli vb.) en iyi performansı verdiğini değerlendirebiliriz.
3. Uyarın Türü (Tek Uyarın, Eşleşen, Eşleşmeyen) Beyin Dalgalarını Nasıl Etkiliyor ve Bu Etki Alkolik/Kontrol Grupları Arasında Farklılık Gösteriyor mu?
İki uyarının birbirine eşleştiği veya eşleşmediği koşullarda beyin dalgalarının nasıl değiştiğini analiz etmek mümkündür. Alkolik grupta bu uyarılara verilen tepkilerin kontrol grubuna kıyasla daha yavaş veya daha farklı bir desen (pattern) içerip içermediği incelenebilir.

1.1. **Motivasyon:**

Alkol bağımlılığı dünya genelinde yaygın görülen bir rahatsızlıktır ve bireyin fiziksel, zihinsel ve sosyal yaşantısını olumsuz yönde etkiler. Genetik yatkınlığı olan bireylerin, beyindeki elektriksel aktiviteyi yansıtan EEG sinyallerinde belirgin ipuçlarına sahip olabileceği öngörülmektedir. Dolayısıyla, erken dönemde bu bulguların tespiti, bireye özgü tedavi yöntemlerinin geliştirilmesinde ve koruyucu önlemlerin alınmasında kilit rol oynayabilir.

Nörobilim alanında daha önce yapılan çalışmalarda, EEG dalgalarının farklı bozukluklar (örn. epilepsi, dikkat eksikliği, uyku bozuklukları) için tanıda yardımcı olduğu gösterilmiştir. Alkolizm gibi kronik rahatsızlıklarda da EEG analizinin erken uyarı ve tanı araçları arasında sayılabilecek potansiyele sahip olduğu bilinmektedir. Özellikle makine öğrenmesi yöntemlerinin yaygınlaşmasıyla, büyük boyutlu EEG verilerinde saklı kalmış desenleri otomatik olarak çıkarmak daha olanaklı hâle gelmiştir.

Bu nedenle, bu projede seçtiğimiz EEG-Alcohol veri kümesini analiz ederek, alkolizm ile ilişkili olabilecek özellikleri (features) belirlemenin hem bilimsel hem de toplumsal yönden yüksek öneme sahip olduğunu düşünüyoruz. Eğer EEG sinyallerindeki farklılıklar güvenilir şekilde tespit edilip sınıflandırılabilirse, gelecekte erken müdahale stratejileri veya kişiye özel tedavi yaklaşımları geliştirilebilir.

1.2. **Literatür Araştırması:**

Alkol bağımlılığının genetik temellerini inceleyen literatür taraması kapsamında yapılan çalışmalar incelendiğinde, genetik faktörlerin alkol bağımlılığı üzerinde önemli etkileri olduğu ortaya konmuştur. Edenberg ve arkadaşları, genetik varyasyonların özellikle aldehid dehidrojenaz genindeki bir nokta mutasyonun Uzak Doğu popülasyonlarında alkol bağımlılığı riskini önemli ölçüde artırdığını belirtmiştir [1]. Bu mutasyon, enzim inaktivasyonuna neden olmakta ve dominant bir kalıtım yoluyla popülasyonu alkol bağımlılığı riski yüksek ve düşük olarak iki gruba ayırmaktadır. Ayrıca dopamin D2 (DRD2) geni ve alkol metabolizmasında rol alan diğer genlerin de Avrupa ve Uzak Doğu popülasyonlarında alkol tüketimi üzerinde etkileri olduğu ifade edilmiştir [1].

Alkol bağımlılığının multifaktöriyel bir hastalık olduğunu vurgulayan çalışmalar, genetik faktörlerin bağımlılığa yakalanma riskinin %50-60'ını oluşturduğunu ortaya koymuştur [2]. Bu

genetik yatkınlık, alkolün farmakokinetik ve farmakodinamik süreçlerine ilişkin genlerle ilişkilendirilmiş ve impulsivite, disinhibisyon, duyuşsal arayış gibi nörofizyolojik yanıtları düzenleyen genlerin de alkol bağımlılığı üzerinde önemli etkileri olduğu gösterilmiştir. Ayrıca, alkol bağımlılığının diğer psikiyatrik hastalıklarla ve bağımlılıklarla ortak genetik temelleri paylaştığı ve komorbidite gösterdiği vurgulanmıştır [2].

İkiz çalışmaları temel alınarak yürütülen araştırmalar, alkol bağımlılığındaki genetik yatkınlığın kadın ve erkeklerde benzer oranlarda olduğunu (0.5-0.6) ortaya koymuştur [3]. Bu çalışmalar, bağımlılığa yatkınlık oluşturan genetik varyasyonların anlaşılması için özellikle kaygı, impulsivite gibi öncü davranışların incelenmesinin önemini vurgulamaktadır. Son dönemde gerçekleştirilen genetik bağlantı analizleri, alkolizme yatkınlıkla ilişkili spesifik gen bölgelerinin tanımlanmasına olanak sağlamıştır [3].

Sonuç olarak, alkol bağımlılığının genetik boyutunu irdeleyen literatür, genetik ve çevresel faktörlerin etkileşiminin anlaşılması için daha fazla kapsamlı ve detaylı çalışmalara ihtiyaç olduğunu göstermektedir. Bu tür çalışmaların geliştirilmiş tedavi yöntemlerine ve kişiselleştirilmiş tedavilere zemin hazırlayacağı düşünülmektedir.

2. Materyal ve Metot

2.1. Materyal

Bu çalışmada kullanılan veri seti, EEG (Elektroensefalogram) sinyalleri üzerinden alkol bağımlılığı ile genetik yatkınlık arasındaki ilişkileri incelemek amacıyla oluşturulan geniş çaplı bir araştırmadan elde edilmiştir. Veri seti, deneklerin kafatasına yerleştirilen 64 elektrottan alınan ölçümleri içermekte olup, her bir elektrot 256 Hz frekansında (yaklaşık 3.9 ms aralıklarla) 1 saniyelik süre boyunca veri toplamıştır.

Katılımcılar iki gruba ayrılmıştır: alkolik bireyler ve kontrol grubu. Her bir deneye, ya tek bir uyaran (S1 objesi) ya da iki farklı uyaran (S1 ve S2) gösterilmiştir. Bu uyaranlar, Snodgrass ve Vanderwart'ın 1980 yılında hazırladığı nesne resimleri setinden seçilmiştir. İki uyaran sunulan deneklerde, S1 ve S2 aynıysa “eşleşen koşul (S2 match)”, farklıysa “eşleşmeyen koşul (S2 nomatch)” olarak etiketlenmiştir.

Veri setindeki her bir deneme (trial), aşağıdaki sütunlardan oluşmaktadır:

- trial number (deneme numarası)
- sensor position (elektrot konumu)
- sample num (örnek numarası: 0-255)
- sensor value (mikrovolt cinsinden ölçüm)
- subject identifier (alkolik = 0, kontrol = 1)
- matching condition (S1 obj, S2 match, S2 nomatch)
- channel number (0-63)
- subject name (kişiyeye özel seri kodu)
- time (saniye cinsinden zaman değeri)

Veri seti Henri Begleiter tarafından, State University of New York Health Center at Brooklyn'deki Neurodynamics Laboratuvarı'nda toplanmıştır. Veri setine UCI Machine Learning Repository üzerinden erişilebilir ve herhangi bir kullanım kısıtlaması bulunmamaktadır.

Veri kümesi boyutu:

- Eğitim verisi: 7,667,712 satır
- Test verisi: 7,864,320 satır

-
- Özellik sayısı: 9 sütun (ham haliyle)

2.2. Metot

Bu çalışmada, EEG sinyallerinden elde edilen veri kümesine çeşitli veri ön işleme adımları uygulanmıştır. Bu adımlar, eksik veri analizi, kategorik verilerin işlenmesi, normalizasyon, özellik mühendisliği ve görselleştirme süreçlerini kapsamaktadır.

1. Eksik Veri Analizi:

Veri kümesinde yapılan ilk incelemelerde herhangi bir eksik veri (NaN) bulunmadığı tespit edilmiştir. Bu nedenle herhangi bir doldurma veya silme işlemi uygulanmamıştır.

2. Kategorik Veri İşlemleri:

Veri kümesinde yer alan bazı sütunlar kategorik özellikler içermektedir: 'sensor position', 'subject identifier' ve 'matching condition'. Bu değişkenler, makine öğrenmesi modellerinde kullanılabilir hale getirilmek için etiketleme yöntemi (label encoding) ile sayısal formata dönüştürülmüştür.

3. Veri Normalizasyonu:

Özelliklerin farklı ölçeklerde olması nedeniyle, modellerin daha kararlı öğrenebilmesi için StandardScaler yöntemi kullanılarak veriler normalize edilmiştir. Bu işlemde, ortalaması 0 ve standart sapması 1 olacak şekilde tüm sayısal özellikler yeniden ölçeklendirilmiştir.

4. Özellik Seçimi:

Veri seti içerisinde, her bireyin ve sensörün ölçüm istatistiklerini (ortalama, minimum, maksimum, standart sapma) çıkarmak için grup bazlı toplulaştırma işlemleri yapılmıştır. Bu sayede zaman serisi halindeki EEG verileri sabit uzunlukta özellik vektörlerine dönüştürülmüştür. Her satır artık bir birey/sensör kombinasyonunu temsil etmektedir.

5. Görselleştirme:

Veri keşfi ve analiz sürecinde çeşitli görselleştirme tekniklerinden faydalanılmıştır. Özellikle veri dağılımını, aykırı değerleri ve korelasyonları analiz edebilmek için şu yöntemler kullanılmıştır:

- sns.histplot ile veri dağılımı incelenmiştir.
- sns.boxplot ile aykırı değerlerin tespiti yapılmıştır.
- heatmap ve correlation_matrix kullanılarak özellikler arasındaki ilişkiler görselleştirilmiştir.
- Ayrıca model sonrası Confusion Matrix ve Feature Importance grafikleri ile model performansı analiz edilmiştir.

Bu ön işleme adımları sonrasında, veri kümesi sınıflandırma algoritmalarıyla eğitmeye hazır hale getirilmiştir.

3. Bulgular ve Tartışma

Bu çalışmada, sınıflandırma problemi için üç farklı makine öğrenmesi algoritması karşılaştırılmıştır: Random Forest, Logistic Regression ve Gradient Boosting.

1. Kullanılan Algoritmalar:

- Random Forest
- Logistic Regression
- Gradient Boosting

2. Hiperparametre Ayarlamaları (GridSearchCV Sonuçları):

- Random Forest: {'max_depth': 10, 'min_samples_split': 2, 'n_estimators': 100} – Doğruluk: 0.9247

- Logistic Regression: {'C': 0.01, 'penalty': 'l2'} – Doğruluk: 0.9528
- Gradient Boosting: {'learning_rate': 0.05, 'max_depth': 3, 'n_estimators': 100} – Doğruluk: 0.8883

3. İstatistiksel Test:

Algoritmalar arasında anlamlı fark olup olmadığını ölçmek amacıyla McNemar Testi kullanılmıştır. Bu test, aynı veri kümesi üzerinde farklı iki modelin tahmin performanslarını çiftler halinde karşılaştırarak, aralarındaki farkın istatistiksel olarak anlamlı olup olmadığını test eder.

- Random Forest vs Logistic Regression → p-değeri: 0.54883 → fark anlamlı değildir.
- Random Forest vs Gradient Boosting → p-değeri: 0.00813 → fark anlamlıdır.
- Logistic Regression vs Gradient Boosting → p-değeri: 0.00091 → fark anlamlıdır.

4. En İyi Model Seçimi:

GridSearch doğrulukları ve istatistiksel testler dikkate alındığında, ****en iyi algoritma Logistic Regression**** olarak belirlenmiştir. Ancak test verileri üzerinde en yüksek doğruluk ****Random Forest modeli**** tarafından elde edilmiştir (Model doğruluğu: 0.8525). Bu nedenle son model olarak Random Forest kullanılmıştır.

5. Performans Değerleri (Random Forest):

- Doğruluk (Accuracy): 0.85
- Precision: Alkollü (0): 0.91, Kontrolcü (1): 0.81
- Recall: Alkollü (0): 0.75, Kontrolcü (1): 0.94
- F1-score: Alkollü (0): 0.83, Kontrolcü (1): 0.87

Karışıklık Matrisi (Confusion Matrix):

```
[[43 14]
 [ 4 61]]
```

Bu sonuçlara göre, Random Forest modeli hem istatistiksel hem de performans açısından güvenilir sonuçlar vermektedir. Literatürde bu veri setiyle yapılmış benzer çalışmalar bulunmadığından doğrudan karşılaştırma yapılamamıştır.

4. Sonuç ve Öneriler

Bu çalışma, EEG sinyalleri aracılığıyla alkol bağımlılığına genetik yatkınlığın tespit edilmesini ve alkolik bireyler ile sağlıklı bireyler arasındaki nörofizyolojik farklılıkların analiz edilmesini amaçlamıştır. Yapılan veri işleme adımları, filtreleme teknikleri ve makine öğrenmesi algoritmaları sayesinde, elde edilen verilerden anlamlı sonuçlar çıkarılmıştır. Araştırma sorularına ilişkin bulgular şu şekilde özetlenebilir:

Araştırma Sorusu 1: Alkolik ve Kontrol Gruplarının EEG Örüntüleri Nasıl Farklılık Gösteriyor?

Yapılan analizler, EEG sinyallerinin istatistiksel özellikleri (ortalama, standart sapma, minimum, maksimum) açısından iki grup arasında belirgin farklar olduğunu ortaya koymuştur. Alkolik bireylerin sinyallerinde düzensizliklerin daha baskın olduğu ve bazı elektrot bölgelerinde (özellikle oksipital ve parietal alanlarda) bu farkların daha belirgin hale geldiği gözlemlenmiştir.

Araştırma Sorusu 2: Makine Öğrenmesi Modelleri, Alkolik Bireyleri Kontrol Grubundan Ne Kadar Başarıyla Sınıflandırabilir?

Random Forest, Logistic Regression ve Gradient Boosting algoritmaları kullanılarak gerçekleştirilen sınıflandırma sonucunda, test verisi üzerinde en başarılı performansı Random Forest algoritması göstermiştir (%85.25 doğruluk). Modelin precision ve recall değerleri dengeli olup, McNemar testi

sonuçları da bu modelin diğerlerine kıyasla istatistiksel olarak anlamlı farklılıklar sunduğunu göstermiştir.

Araştırma Sorusu 3: Uyarın Türü (Tek Uyarın, Eşleşen, Eşleşmeyen) Beyin Dalgalarını Nasıl Etkiliyor ve Bu Etki Alkolik/Kontrol Grupları Arasında Farklılık Gösteriyor mu?

Bu soruya yönelik yapılan analizlerde, her bir uyarın türü için alkolik ve kontrol grubu arasındaki EEG tepkileri istatistiksel olarak karşılaştırılmıştır [4].

S1 (tek uyarın) durumunda yaklaşık %90 oranında kanal, iki grup arasında anlamlı fark göstermiştir (~55 kanal) [4].

S2 matched (eşleşen) durumunda bu oran %77 (~47 kanal),

S2 nomatched (eşleşmeyen) durumunda ise %80 civarındadır (~50 kanal) [4].

Ayrıca, bazı beyin bölgeleri arasındaki korelasyon (örneğin FP1-FPZ, FP1-FP2) kontrol grubunda %90'ın üzerinde yüksek değerler gösterirken, alkolik grupta bu oran %20 civarında kalmıştır. Bu da, uyarınlara verilen yanıtların sadece şiddetinde değil, aynı zamanda beyin bölgeleri arasındaki eşgüdüm düzeyinde de farklılıklar olduğunu ortaya koymaktadır [4].

Genel Öneriler:

- Uyarın türlerine özgü sinyal tepkilerinin daha ayrıntılı zamansal ve frekans analizleri yapılabilir.
- EEG verileri ile birlikte genetik dizilim verileri entegre edilerek çok katmanlı modeller geliştirilebilir.
- Daha geniş yaş aralığı ve kültürel çeşitlilik içeren örneklerle çalışarak modelin genellenebilirliği artırılabilir.
- Derin öğrenme modelleri (CNN, LSTM vb.) ile klasik algoritmaların performansları karşılaştırılabilir.
- Beyin bölgeleri arası korelasyon analizleri daha kapsamlı şekilde modellenerek, fonksiyonel bağlantılar incelenebilir.

Teşekkür

Projenin pratik uygulama çalışmalarının gerçekleştirilebilmesi için vermiş oldukları katkılar adına Dr. Öğr. Üyesi Ayşe Merve ACILAR'a teşekkür ederiz. MF-11111 kodlu bu proje, Necmettin Erbakan Üniversitesi Bilimsel Araştırma Projeleri Birimi tarafından desteklenmiştir. Bu çalışmanın ortaya çıkmasında verdiği destekten ötürü Bilimsel Araştırma Projeleri Birimine teşekkür ederiz.

Kaynaklar

- [1]. Gurling, Hugh M.D.^a; Cook, Christopher C.H.^b. The genetic predisposition to alcohol dependence. *Current Opinion in Psychiatry* 12(3):p 269-275, May 1999.
- [2]. Buscemi L, Turchi C. An overview of the genetic susceptibility to alcoholism. *Medicine, Science and the Law*. 2011;51(1_suppl):2-6. doi:10.1258/msl.2010.010054
- [3]. Enoch, MA., Goldman, D. The genetics of alcoholism and alcohol abuse. *Curr Psychiatry Rep* 3, 144–151 (2001). <https://doi.org/10.1007/s11920-001-0012-3>
- [4]. R. RuslanKL, "EEG Data Analysis," Kaggle, [Online]. Available: <https://www.kaggle.com/code/ruslankl/eeg-data-analysis?kernelSessionId=9349628>. [Accessed: May 3, 2025].