

# Makine Öğrenmesi Teknikleri Yardımıyla Otizm Spektrum Bozukluğu Teşhisi

## Diagnosing Autism Spectrum Disorder Using Machine Learning Techniques

Hidayet Takçı  
Bilgisayar Mühendisliği Bölümü  
Sivas Cumhuriyet Üniversitesi  
Sivas, Türkiye  
htakci@cumhuriyet.edu.tr

Saliha Yeşilyurt  
Bilgisayar Mühendisliği Bölümü  
Sivas Cumhuriyet Üniversitesi  
Sivas, Türkiye  
salihayesilyurt@cumhuriyet.edu.tr

**Öz**—Otizm, dil ve iletişim bozuklukları ile karakterize edilebilen yaygın bir gelişimsel bozukluktur. Hastalığın teşhisinde sıklıkla tarama testlerinden yararlanılır. Bununla birlikte tarama testlerinin süresi ve maliyeti yüksektir. Otizm spektrum bozukluğu teşhisinde tarama testlerine yardımcı olarak son dönemde makine öğrenmesi yöntemleri kullanılmaktadır. Bu kapsamda yapılan çalışmamızda toplam sekiz farklı makine öğrenmesi algoritması dört farklı veri seti üzerinde otizm teşhisi için kullanılmıştır. Kullanılan veri setleri güncel olup kaynağını Q-CHAT, AQ-10-child ve AQ-10-adult tarama testlerinden almaktadır. Kullanılan algoritmaların değerlendirilmesinde doğruluk, hassasiyet, duyarlılık ve özgüllük metrikleri kullanılmıştır. Deneysel sonuçlara göre, en iyi sonuçlar destek vektör makinesine dayalı bir sınıflandırıcı olan C-SVC ile elde edilmiştir. C-SVC için tüm veri setlerinde sınıflandırma performans metrikleri %100 olarak elde edilmiştir. Çok değişkenli lojistik regresyon ikinci sırada yer almıştır. En düşük sonuç ise karar ağacı tabanlı bir algoritma olan C4.5 algoritması ile elde edilmiştir.

**Anahtar Sözcükler**—Otizm spektrum bozukluğu, Makine öğrenme algoritmaları, Karar ağaçları, Naive Bayes, K-en yakın komşu, Destek vektör makinesi, Yapay sinir ağları

**Abstract**—Autism is a generalized pervasive developmental disorder that can be characterized by language and communication disorders. Screening tests are often used to diagnose such a disorder; however, they are usually time-consuming and costly tests. In recent years, machine learning methods have been frequently utilized for this purpose due to their performance and efficiency. This paper employs the most eight prominent machine learning algorithms and presents an empirical evaluation of their performances in diagnosing autism disorder on four different benchmark datasets, which are up-to-date and originate from the QCHAT, AQ-10-child, and AQ-10-adult screening tests. In doing so, we also utilize precision, sensitivity, specificity, and classification accuracy metrics to scrutinize their performances. According to the experimental results, the best outcomes are obtained with C-SVC, a classifier based on a support vector machine. More importantly, in terms of C-SVC performance metrics even lead to 100% in all datasets. Multivariate logistic regression has been taken second place. On the other hand, the lowest results are obtained with the C4.5 algorithm, a decision tree-based algorithm.

**Index Terms**—Autism spectrum disorder, Machine learning algorithms, Decision trees, Naive Bayes, K-nearest neighbor, Support vector machine, Neural network

### I. GİRİŞ

Otizm, tekrarlayan davranışlar, sosyal iletişimde bozukluk ve zayıf dil becerileri ile karakterize edilebilen nörolojik bir gelişim bozukluğudur. Hastalığın kesin nedeni bilinmemekle birlikte, bazı çalışmalarda genetik ve çevresel faktörlerin hastalığı etkilemiş olabileceğinden bahsedilmektedir [1]. Otizm için bilinen bir tedavi veya ilaç olmasına rağmen, erken ve kapsamlı davranışsal müdahaleler sosyal iletişimi geliştirirken, kullanılan ilaçlar ise kaygı ve saldırganlığı azaltabilmektedir [2].

Otizm teşhisi konusunda bazı öncü çalışmalar yapılmıştır. Bu çalışmalardan birinde Otizm Spektrum Bozukluğu (OSB) ile motor becerilerindeki gerilik arasında bir bağlantı olduğu tespit edilmiştir [3]. McCoy ve arkadaşları otizmlili çocukların obez olma olasılıklarının daha yüksek ve fiziksel olarak daha az aktif olduğunu bulmuşlardır [4]. Bir başka çalışmada otizm rahatsızlığı olan ergenlerin (13-18 yaş) fiziksel aktivite bakımından akranlarından geri olduğu gözlenmiştir [5]. Singh ve diğerleri normalin altında ağırlıkla (<1500 g) doğan çocukların otizm olma olasılıklarının diğerlerinden 3,2 kat daha fazla olduğunu bulmuşlardır [6]. Yates ve Couture tarafından yapılan çalışmada sosyal etkileşimdeki zorlukların ve gecikmelerin göz ardı edilmesine rağmen sıklıkla OSB ile ilişkili olduğunu göstermişlerdir [7]. Bu seçeneğin değerlendirilmesi daha zordur çünkü sosyal etkileşimdeki gecikmeleri tanımlamak bazen zor ve görecelidir.

Otizm belirtileri bir yaşından itibaren ortaya çıkar. Araştırmalar, otizm spektrum bozukluğunun iki yaşından küçük çocuklarda 2 aşamalı tarama ile güvenilir bir şekilde tespit edilebildiğini göstermektedir [8]. Teşhis başarısı için en önemli işlem, rutin gelişim taramalarının tamamlanmasıdır [9]. Otizm teşhisi birçok alandan uzmanların oluşturduğu bir ekip tarafından konur. Kesin teşhis için en çok tercih edilen yöntem Revize edilmiş Otizm Tanısal Gözlem Programı (ADOS-R) olarak bilinir. Çocuğun geçmişi ve davranışları hakkında detaylı sorular içeren bir anket olan Otizm Tanısal Görüşme Revize (ADI-R) yöntemi bu alandaki diğer bilinen yöntemdir.

Bu iki yönteme ek olarak, Otizm Spektrum Katsayısı (AQ) ve Sosyal İletişim Anketi (SCQ) gibi hastanın kendisine veya ebeveynine uygulanan başka anketler de bulunmaktadır. Ayrıca, bebeklere yönelik Q-CHAT [10], çocuklar için AQ-10-child ve yetişkinler için AQ-10-adult tarama testleri bulunmaktadır.

Otizm teşhisinde kullanılan mevcut yöntemler maliyetli yöntemlerdir. Teşhis maliyetlerini düşürmek ve teşhis süresini kısaltmak için tarama testlerine makine öğrenmesi sistemleri destek olurlar. Makine öğrenmesi yöntemleri, yapılandırılmış veriler yardımıyla sağlık alanında karar destek sistemleri oluşturmak için sıklıkla kullanılmaktadır. Kalp krizi erken teşhisi [11], meme kanseri erken teşhisi [12] gibi birçok alanda çalışmalar yapılmıştır. Makine öğrenmesi kullanımının mümkün olduğu hastalıklardan biri de OSB teşhisidir. Makine öğrenmesi, davranış bilimlerinde tanı ve müdahale çalışmalarını zenginleştirme potansiyeline sahip yenilikçi bir yöntemdir. Otizm spektrum bozukluğu teşhisinde makine öğrenme yöntemlerinin kullanılması, sağlık hizmetlerine daha hızlı erişim sağlamak, teşhis performansını arttırmak ve en değerli özellikleri belirleyerek girdi verilerinin boyutunu azaltmak yoluyla teşhis süresini kısaltmak için bugüne kadar kullanılmıştır [13], [14].

Bu çalışmada otizm spektrum bozukluğu teşhisi için makine öğrenmesi yöntemlerinden yararlanılacaktır. Çalışmamızın amacı otizm teşhisinde daha düşük maliyetli ve daha hızlı makine öğrenmesi yöntemlerinin tespit edilmesidir. Yeni doğan, çocuk, ergen ve yetişkinler veri seti üzerinde deneyler yapılarak her bir algoritma doğru tanıma, hassasiyet, duyarlılık ve özgüllük metrikleriyle sunulmuştur.

## II. GEÇMİŞ ÇALIŞMALAR

OSB tarama sürelerini kısaltmak, maliyetlerini düşürmek, teşhis sürecinin doğruluğunu ve hassasiyetini arttırmak için makine öğrenmesine ihtiyaç vardır. OSB teşhisi problemi, geçmiş veriler yardımıyla yeni vakalara otizm tanısı koymaya çalışan sınıflandırma problemidir. Makine öğrenmesi yöntemleri tek başına bir çözüm olmaktan çok bir tarama aracı ile uyumlu olmak zorundadır. Geçmiş çalışmalarda ADOS ve ADI-R tarama süreçleriyle birlikte birçok makine öğrenmesi yöntemi kullanılmıştır.

Otizm tanısal tarama süreleri uzun olup Wall ve arkadaşları [15], tarama süresini azaltmak ve ASD özelliklerini daha hızlı saptayabilmek için 16 farklı sınıflayıcı kullanmışlardır. En iyi sonucu ADOS module 1 verisi üzerinde %100 doğru tanıma oranı ile Alternatif Karar Ağacı (ADTree) kullanarak elde etmişlerdir.

Hauck ve Kliwer [16] ADOS (Otizm Teşhis Gözlem Programı) ve ADI-R (Otizm Teşhis Röportajı Revize) taramalarıyla ilgili önemli tarama sorularını belirlemeye çalışmışlardır. Ayrıca yazarlar, tarama yöntemleri ve ADI-R ve ADOS tarama testlerinin birlikte kullanıldığında daha iyi çalışabileceğini göstermişlerdir. Yapılan çalışmada, 2500

örnekli veri seti üzerinde RBF kernel kullanan SVM algoritması ile %85,6 ile %94,3 arasında sonuç elde edilmiştir. Özgüllük aralığı ise %80,9 ile %94,3 arasında meydana gelmiştir.

Heinsfeld [17] Otizm Görüntüleme Veri Değişimi (ABIDE I) veri setinde yer alan beyin görüntüleme verileri yardımıyla OSB hastalarını tanımlamak için derin öğrenme algoritması ve sinir ağı uygulamış. %66 ila %71 doğruluk aralığında ortalama %70 sınıflandırma doğruluğu elde etmiştir, detaylarda ise; SVM algoritması ile %65, Rastgele orman algoritması ile %63 doğruluk elde edilmiştir.

Liu [18] tarafından yapılan çalışma diğerlerinden farklılık göstermektedir. Bu çalışmanın amacı yüz tarama modelleri yardımıyla otizm tanısı konulup konulamayacağı üzerinedir. Bunun için bir göz hareketi veri seti kullanılarak sınıflandırma yapılmış ve sonuçlar doğru tanıma oranı olarak sunulmuştur. Yapılan çalışma sonucuna göre %88.51 doğru tanıma oranıyla göz hareketlerinden otizm tanısı konabilmiştir. Otizm teşhisinde makine öğrenmesi tekniklerinin sıklıkla kullanıldığı görülmektedir [19].

## III. DENEYSEL ÇALIŞMA

Bu bölümde, önce deneylerde kullanılacak veri seti açıklanacak ardından performans metrikleri sunulacaktır. Veriler üzerinde makine öğrenmesi algoritmaları çalıştırıldıktan sonra elde edilen sonuçlar performans metriklerine dayalı tablolar halinde sunulacaktır. Her bir tabloya ait değerler kısaca yorumlanarak bölüm sonlandırılacaktır. Bu bölümdeki deneysel çalışmaların amacı otizm veri setinde en başarılı algoritmaların belirlenmesidir.

### A. Veri Seti ve Kullanılan Algoritmalar

Deneylerimizde Thabtah [20] tarafından derlenen veriler kullanılmış olup bu veriler Thabtah tarafından geliştirilen ASDTests isimli mobil uygulama yardımıyla toplanmıştır. Testler Q-CHAT-10, AQ-10-child ve AQ-10-adult tarama yöntemlerine dayalı olarak bebekler, çocuklar, ergenler ve yetişkinler için geliştirilmiştir.

TABLO I  
OTİZM SPEKTRUM BOZUKLUĞU VERİ KÜMELERİ

	Veri Kümesi	Yaş Aralığı	Özellik Adedi	Kayıt Sayısı
Bebek	Toddler Autism Dataset – Version 2	0-3 yaş	18	1054
Çocuk	Children Autism Dataset – Version 2	4-11	24	509
Ergen	Adolescent Autism Dataset version 2	12-16	24	248
Yetişkin	Adults Autism Dataset version 2	17+	24	1118

Veri setlerinin birisi 18 diğerleri 24 adet özellik içermektedir. Özelliklerden bazıları boş değerler içerdiği için bazıları da analiz değeri olmayan veriler içerdiği için bir elemeye geçirilmiştir. Bu eleme sırasında; testin neden yapıldığı, vakanın hangi milletten olduğu, hangi ülkede yaşadığı, tarama testini kimin yaptığı, daha önce test yapıp yapılmadığı, testin dili, A1-A10 arasındaki sorulara verilen cevapların skorunu

veren 'Score' alanı gibi özellikler elenmiştir. Bununla birlikte, bütün veri setlerindeki tarama testlerinde, önemli 10 sorunun cevabını veren A1-A10 arası özellikler, ailede otizm tanısı olan birinin olup olmadığı, vakanın sarılık geçirip geçirmediği gibi bilgiler önemli bulunarak bırakılmıştır.

Otizm teşhisi konusunda kullanılacak algoritmalar karar verirken en önemli dayanak geçmiş çalışmalar ve makine öğrenme algoritmalarının popüleritesi olmuştur. Örneğin, karar ağaçları bilinen en eski sınıflandırıcılardan biridir, anlaşılabilir ve yorumlanabilir oldukları için hala makine öğrenme sistemlerinde sıklıkla kullanılmaktadır [22]. Bu çalışmada karar ağacı algoritmalarından C4.5 ve RndTree tercih edilmiştir. Makine öğrenmesi çalışmalarında yine sıklıkla kullanılan algoritmalarından birisi Naive Bayes algoritmasıdır. Bayes sınıflayıcılar, bir dizi koşullu olasılık içeren yapısal bir modele sahiptir [22]. Bu çalışmada Naive Bayes algoritmasının sürekli verilere uygun sürümü olan Naive Bayes Continuous kullanılmıştır. İstatistiksel tahmin ve örüntü tanıma için sıklıkla kullanılan K-NN algoritması da bu çalışmada tercih edilmiştir. K-NN sınıflayıcı komşuluk bilgilerine göre sınıflandırma yapmakta olup komşuları bulmak için sıklıkla benzerlik veya mesafe bilgisinden faydalanır [23]. Makine öğrenmesi algoritmaları arasında yapay sinir ağlarının ayrı bir yeri vardır. Bu çalışmada çok katmanlı algılayıcı algoritması (MLP) da otizm teşhisi için kullanılmıştır. Ayrıca regresyon analizi tekniklerinden çok değişkenli lojistik regresyon (MLR), istatistiksel bir yöntem olan doğrusal ayırt edici analizi (LDA) ve başarılı bir destek vektör makinesi algoritması olan C-SVC algoritması kullanılmıştır. Algoritmaların otizm veri setleri üzerinde uygulanmasında Tanagra isimli makine öğrenme aracından yararlanılmıştır [24]. Seçilen algoritmalar ve algoritmalarla ait parametre değerleri Tablo II'de verilmiştir.

### B. Model Değerlendirme

Hangi algoritmanın daha başarılı sonuç verdiğini tespit edebilmek için yapılan çalışmaya model değerlendirme denir. Model değerlendirme için bu çalışmada; sınıflandırma doğruluğu (accuracy), hassasiyet (precision), duyarlılık (sensitivity) ve özgüllük (specifity) değerlerinden yararlanılmıştır. Ayrıca elde edilen sonuçlardaki tesadüfiliği ortadan kaldırmak için 10 kat çapraz doğrulama (10-fold cross validation) yapılmıştır. Otizm spektrum verileri, ASD tanısı olan ve olmayan şeklinde iki sınıftır. Bu atamaya göre oluşan karışıklık matrisi Tablo III'de verildiği gibi olacaktır. Karışıklık matrisine dayanarak, doğru tanıma ve diğer ölçümler Tablo IV'deki gibi yazılabilir.

Doğru tanıma (accuracy) tek başına yeterli bir ölçüm yöntemi değildir çünkü veri kümesi sınıflar arasında dengeli dağılmadığında tek başına yeterli bir metrik olmaktan uzaktır. Bu nedenle diğer metriklere ihtiyaç duyulmuştur. Hassasiyet, duyarlılık ve özgüllük ölçümlerinin birlikte ifade edildiği ölçüm, model karşılaştırması sırasında daha değerli sonuçlar verecektir.

TABLO II  
KULLANILAN MAKİNE ÖĞRENME ALGORİTMALARI VE PARAMETRELERİ

	Algoritma	Parametreleri
C4.5	C4.5	Min size of leaves=5 Confidence level=0,25
NB	Naïve Bayes Continuous	Lambda=0
KNN	K-nearest Neighbors	K=5 Distance method=HEOM(Wilson -Martinez,JAIR'97)
MLP	Multilayer Perceptron	Number of neurons=10 Learning rate=0,15 Validation set proportion=0,20 Stopping rule (Max iteration=100 or error rate threshold=0,01)
RT	Rnd Tree	Selected attributes=-1
MLR	Multinomial Logistic regression	Varsayılan değerler
LDA	Linear discriminant analysis	Matrix inversion=Exact
CSVC	C-SVC from LIBSVM	Kernel type=Linear Degree of kernel function=1 Gamma=0 Coef 0=0 Complexity=1 Transformation=Normalization Epsilon for tolerance=0,001

TABLO III  
KARIŞIKLIK MATRİSİ

	Tahmin Edilen Sınıf	
Gerçek Sınıf	ASD	Non-ASD
ASD	TP(a)	FN(b)
Non-ASD	FP(c)	TN(d)

### C. Deneyisel Sonuçlar

Otizm erken teşhisi için farklı yaş gruplarına dayalı olarak elde edilen dört farklı veri seti üzerinde sekiz farklı makine öğrenmesi algoritması çalıştırılmıştır. Sonuçların tamamı tek bir tabloda bir araya getirilerek sonuçların izlenmesi kolaylaştırılmıştır. Tablo V, bu çalışmada elde edilen çok sayıda deneyin bir özeti şeklinde hazırlanmıştır.

Tablo V'te dört farklı veri seti ve sekiz farklı makine öğrenmesi yöntemi için doğru tanıma, hassasiyet, duyarlılık ve özgüllük değerlerine göre toplu sonuçlar elde edilmiştir. Tablo V'te elde edilen veriler yardımıyla veri setinden bağımsız olarak her bir sınıflayıcının ortalama performans değerleri Tablo VI'da verilmiştir. Ayrıca algoritmadan bağımsız olarak her bir veri setinin ortalama performans değerleri elde edilmiş ve Tablo VII'de verilmiştir.

TABLO IV  
DOĞRULUK, HASSASİYET VE DUYARLILIK ÖLÇÜMLERİ

Doğru tanıma (accuracy)	$(TP+TN)/(TP+FP+FN+TN)$
Hassasiyet (precision)	$TP/(TP+FP)$
Duyarlılık (sensitivity)	$TP/(TP+FN)$
Özgüllük (specifity)	$TN/(TN+FP)$

TABLO V  
DÖRT FARKLI VERİ SETİ ÜZERİNDE SEKİZ FARKLI MAKİNE ÖĞRENMESİ ALGORİTMASI İÇİN DÖRT FARKLI PERFORMANS METRİĞİNE GÖRE SONUÇLARIN  
SUNULMASI

	Metrik	C4.5	NB	K-NN	MLP	RT	MLR	LDA	C-SVC
Bebek	Accuracy	0,93	0,94	0,95	0,99	0,94	1,00	0,95	<b>1,00</b>
	Precision	0,92	0,92	0,94	0,98	0,92	1,00	0,94	<b>1,00</b>
	Sensitivity	0,92	0,96	0,95	0,99	0,93	1,00	0,96	<b>1,00</b>
	Specifity	0,92	0,94	0,95	0,99	0,92	1,00	0,95	<b>1,00</b>
Çocuk	Accuracy	0,87	0,95	0,89	0,98	0,89	1,00	0,96	<b>1,00</b>
	Precision	0,87	0,95	0,90	0,98	0,89	1,00	0,96	<b>1,00</b>
	Sensitivity	0,87	0,95	0,89	0,98	0,89	1,00	0,96	<b>1,00</b>
	Specifity	0,87	0,95	0,90	0,98	0,89	1,00	0,96	<b>1,00</b>
Ergen	Accuracy	0,79	0,89	0,88	0,98	0,85	0,99	0,93	<b>1,00</b>
	Precision	0,79	0,90	0,90	0,98	0,85	0,99	0,94	<b>1,00</b>
	Sensitivity	0,79	0,89	0,88	0,97	0,85	0,99	0,94	<b>1,00</b>
	Specifity	0,79	0,89	0,89	0,98	0,85	0,99	0,94	<b>1,00</b>
Yetişkin	Accuracy	0,94	0,94	0,94	0,99	0,94	1,00	0,96	<b>1,00</b>
	Precision	0,93	0,93	0,93	0,99	0,93	1,00	0,95	<b>1,00</b>
	Sensitivity	0,93	0,94	0,94	0,99	0,921	1,00	0,96	<b>1,00</b>
	Specifity	0,93	0,94	0,94	0,99	0,93	1,00	0,95	<b>1,00</b>

TABLO VI  
VERİ SETİNDEN BAĞIMSIZ SINIFLAYICILAR İÇİN ORTALAMA DOĞRU  
TANIMA ORANLARI

C4.5	NB	K-NN	MLP	RT	MLR	LDA	C-SVC
0,8835	0,9315	0,9164	0,9851	0,9023	0,9979	0,9508	<b>1,00</b>

TABLO VII  
ALGORİTMADAN BAĞIMSIZ VERİ KÜMELERİ İÇİN ORTALAMA DOĞRU  
TANIMA ORANLARI

Bebek	Çocuk	Ergen	Yetişkin
0,9633	0,9428	0,9136	<b>0,9644</b>

Tablo V’te topluca verilen, Tablo VI ve Tablo VII’de özetlenerek verilen sonuçlara göre; en başarılı sınıflayıcı C-SVC algoritması olmuştur. C-SVC algoritması bir tür destek vektör makinesi algoritmasıdır. Literatürde olduğu gibi çalışmamızda da en iyi sonuç, destek vektör makinesi tabanlı algoritma ile elde edilmiştir. Algoritma, makine öğrenmesi aracının varsayılan parametreleri ile işleme alınmıştır. Kullanılan kernel tipi Linear kernel’dir. Diğer kernel tiplerinde bu başarının elde edilemediği görülmüştür. C-SVC algoritmasına oldukça yakın değer veren diğer algoritma Multinomial Logistic regression algoritması olmuştur. Çok değişkenli lojistik regresyon algoritması da başarılı sınıflayıcılardan birisidir. Bu çalışmada da veri setlerinin üçünde tam doğru tanıma oranı vererek başarısını göstermiştir. Çalışmamızda en düşük sınıflama doğruluğunu veren algoritma C4.5 ve RndTree algoritması olmuştur. En başarısız bir diğer algoritma K-NN algoritmasıdır. Algoritmalar üzerinde parametre optimizasyonu yapıldığında da sonuç değişmemiştir.

Veri setine göre yapılan deneylerde en yüksek sonuç, yetişkinler (adult) veri seti ile elde edilmiştir. Yetişkinler veri setini diğerlerinden ayıran en önemli özellik yapılan tarama testinde muhatabın bizzat hastanın kendisi olmasıdır. Diğer veri setlerinde kimi zaman aile veya sağlık çalışanı hastanın yerine tarama sorularını cevaplamaktadır. Ayrıca yaş ilerledikçe hastalığın sınırlarının daha fazla açığa çıkması da

bu sonuçta etkili olmuştur.

#### IV. SONUÇ VE DEĞERLENDİRME

Otizm spektrum bozukluğu her geçen gün artan ciddi hastalıklardan biridir. Şu anda %1-2 civarında olan otizm yoğunluğu, elektronik cihazların kullanımıyla artmaktadır. Çevresel faktörlerden etkilenen otizmin her geçen gün sosyal etkileşimi zayıflatan sanal ortamlardan etkilenmemesi mümkün değildir. Sosyal iletişimin azaldığı bir dünyada gözlerini hayata açan çocuklar nispeten otizme daha eğilimli olabilirler.

Hastalığın kesin nedenleri bilinmemekle birlikte, teşhisi üzerine bazı çalışmalar yapılmıştır. Özellikle otizm tanısında anket ve görüntüleme teknikleri sıklıkla kullanılmaktadır. Anket bazen hastaya, bazen akrabaya veya bakıcıya yapılmakta ve hastanın en iyi durumu belirlenmeye çalışılmaktadır. Görüntüleme teknikleri, bozukluğun nörolojik yönünü ortaya çıkarmakla daha ilgilidir.

Hastalığın tedavisi bağlamında, sindirim sistemi temelli çalışmalar yürütülmektedir ve çalışmalar henüz olgunlaşma döneminde. Bağırsakların ikinci beyin olduğu düşünülmesine ve bağırsak florasındaki bozukluğun otizmi tetiklemesi ile ilgili bulgular olmasına rağmen, henüz herkes tarafından kabul edilmemiştir. Birçok akraba, hala çözüm için pediatrik psikiyatri polikliniklerine gitmektedir. Psikiyatri kliniklerinde hastaya verilen ilaçlar, hastalığa değil, hastalığa eşlik eden davranış bozukluğunu düzenleme etkisine sahiptir. Örneğin, siniri hafifleten ilaçlar, hastalığa eşlik eden sinir krizleri için verilmektedir. Benzer şekilde, dikkat eksikliğini gidermek için de ilaçlar kullanılmaktadır.

Tedavi bağlamında tam bir çözümün olmaması, erken teşhisin önemini artırmaktadır. Son zamanlarda, erken teşhis için makine öğrenme tekniklerinin kullanılması bu anlamda değerlidir. Anket ve görüntüleme tekniklerinde tamamlayıcı bir role sahip olan makine öğrenmesi yöntemleri, hastalığın tanılma doğruluğunu artıracaktır.



Bu çalışmada, daha önce başka problemlerde başarılı olan, kolayca anlaşılabilen, yorumlanabilir ve çok hızlı sonuçlar veren sınıflandırıcılar otizm teşhisi için kullanılmıştır. UCI deposundan alınan otizm hastalarının verileri üzerinde yapılan deneylerde sekiz sınıflandırıcı incelenmiştir. Sınıflandırıcıları karşılaştırmak için doğru tanıma, hassasiyet, duyarlılık ve özgüllük metrikleri kullanılmıştır. Bu problem için hangi sınıflandırıcının daha uygun olduğu araştırılmıştır. Araştırmamız sonucunda aşağıdaki bulgular elde edilmiştir:

- Erken tanının önem kazandığı hastalıklardan biri de Otizm'dir. Diğer hastalıklarda olduğu gibi, makine öğrenme yöntemleri otizmin erken teşhisinde hekimlere yardımcı olmaktadır.
- Seçilen algoritmaların çoğu, %95 ve üzeri yüksek değerlerde doğru tanıma oranlarına sahiptir. Bu nedenle, veri madenciliği yöntemleri ile otizm teşhisi yüksek doğrulukla yapılabilmektedir.
- Veri setindeki eksik verilere rağmen %100 doğru teşhis etme oranı elde edilmiştir. Bu, otizm verilerinin kolayca modellenebileceğini ve eğitilebileceğini göstermektedir.
- Ek olarak, otizm teşhisinde karar ağacı algoritmalarının en iyi sonuç olmasa bile yaklaşık %80 ve üzeri başarı göstermesi ve diğer algoritmalarla oranla daha yorumlanabilir olması uzmanlara hastalık hakkında değerli bilgiler sağlayabilir.

#### REFERENCES

- [1] P. Chaste and M. Leboyer, "Autism risk factors: genes, environment, and gene-environment interactions," *Dialogues Clin. Neurosci.*, vol. 14, no. 3, pp. 281–292, 2012.
- [2] M.-C. Lai, M. V. Lombardo, and S. Baron-Cohen, "Autism," *Lancet*, vol. 383, no. 9920, pp. 896–910, 2014.
- [3] P. Setoh, P. B. Marschik, C. Einspieler, and G. Esposito, "Autism spectrum disorder and early motor abnormalities: Connected or coincidental companions?," *Res. Dev. Disabil.*, vol. 60, pp. 13–15, 2017.
- [4] S. M. McCoy, J. M. Jakicic, and B. B. Gibbs, "Comparison of obesity, physical activity, and sedentary behaviors between adolescents with autism spectrum disorders and without," *J. Autism Dev. Disord.*, vol. 46, no. 7, pp. 2317–2326, 2016.
- [5] W. L. Mangerud, O. Bjerkeset, S. Lydersen, and M. S. Indredavik, "Physical activity in adolescents with psychiatric disorders and in the general population," *Child Adolesc. Psychiatry Ment. Health*, vol. 8, no. 1, p. 2, 2014.
- [6] G. K. Singh, M. K. Kenney, R. M. Ghandour, M. D. Kogan, and M. C. Lu, "Mental health outcomes in US children and adolescents born prematurely or with low birthweight," *Depress. Res. Treat.*, vol. 2013, pp. 1–13, 2013.
- [7] K. Yates and A. Le Couteur, "Diagnosing autism/autism spectrum disorders," *Paediatr. Child Health (Oxford)*, vol. 26, no. 12, pp. 513–518, 2016.
- [8] J. A. Pinto-Martin, M. Dunkle, M. Earls, D. Flidner, and C. Landes, "Developmental stages of developmental screening: Steps to implementation of a successful program," *Am. J. Public Health*, vol. 95, no. 11, pp. 1928–1932, 2005.
- [9] A. Keil, C. Breunig, S. Fleischfresser, and E. Oftedahl, "Promoting routine use of developmental and autism-specific screening tools by pediatric primary care clinicians," *WMJ*, vol. 113, no. 6, pp. 227–231, 2014.
- [10] C. Allison et al., "The Q-CHAT (Quantitative CHecklist for Autism in Toddlers): a normally distributed quantitative measure of autistic traits at 18-24 months of age: preliminary report," *J. Autism Dev. Disord.*, vol. 38, no. 8, pp. 1414–1425, 2008.
- [11] H. Takci, "Improvement of heart attack prediction by the feature selection methods," *TURK. J. OF ELECTR. ENG. COMPUT. SCI.*, vol. 26, pp. 1–10, 2018.
- [12] H. Takci, "Diagnosis of breast cancer by the help of centroid based classifiers," *Gazi Üniv. Mühendis.-Mimar. Fak. Derg.*, vol. 323, p. 330, 2016.
- [13] F. Thabtah, "Autism spectrum disorder screening: Machine learning adaptation and DSM-5 fulfillment," in *Proceedings of the 1st International Conference on Medical and Health Informatics 2017*, 2017.
- [14] D. Bone, M. S. Goodwin, M. P. Black, C.-C. Lee, K. Audhkhasi, and S. Narayanan, "Applying machine learning to facilitate autism diagnostics: pitfalls and promises," *J. Autism Dev. Disord.*, vol. 45, no. 5, pp. 1121–1136, 2015.
- [15] D. P. Wall, J. Kosmicki, T. F. Deluca, E. Harstad, and V. A. Fusaro, "Use of machine learning to shorten observation-based screening and diagnosis of autism," *Transl. Psychiatry*, vol. 2, no. 4, p. e100, 2012.
- [16] F. Hauck and N. Klierer, "Machine learning for autism diagnostics: Applying support vector classification," *Ucmss.com*. [Online]. Available: <http://csce.ucmss.com/cr/books/2017/LFS/CSREA2017/HIM3143.pdf>. [Accessed: 23-Jun-2021].
- [17] A. S. Heinsfeld, A. R. Franco, R. C. Craddock, A. Buchweitz, and F. Meneguzzi, "Identification of autism spectrum disorder using deep learning and the ABIDE dataset," *NeuroImage Clin.*, vol. 17, pp. 16–23, 2018.
- [18] W. Liu, M. Li, and L. Yi, "Identifying children with autism spectrum disorder based on their face processing abnormality: A machine learning framework," *Autism Res.*, vol. 9, no. 8, pp. 888–898, 2016.
- [19] F. Thabtah, "Machine learning in autistic spectrum disorder behavioral research: A review and ways forward," *Inform. Health Soc. Care*, vol. 44, no. 3, pp. 278–297, 2019.
- [20] F. Thabtah, "An accessible and efficient autism screening method for behavioural data and predictive analyses," *Health Informatics J.*, vol. 25, no. 4, pp. 1739–1755, 2019.
- [21] Y. Xing, J. Wang, Z. Zhao, and A. Gao, "Combination data mining methods with new medical data to predicting outcome of coronary heart disease," in *2007 International Conference on Convergence Information Technology (ICCIT 2007)*, 2007, pp. 868–872.
- [22] L. Jiang, H. Zhang, and Z. Cai, "A novel Bayes model: Hidden naive Bayes," *IEEE Trans. Knowl. Data Eng.*, vol. 21, no. 10, pp. 1361–1371, 2009.
- [23] A. Christobel, "An empirical comparison of data mining classification methods," 2011.
- [24] R. Rakotomalala, "Tanagra: a free software for research and academic purposes," in *Proceedings of EGC*, vol. 2, Paris, France, 2005, pp. 697–702.