BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ KONULARI

Editör: Dr.Öğr.Üyesi Ali PAŞAOĞLU

yaz
yayınları

Bilgisayar Bilimleri ve Mühendisliği Konuları

Editör

Dr.Öğr.Üyesi Ali PAŞAOĞLU

yaz yayınları

2025

yaz yayınları

Bilgisayar Bilimleri ve Mühendisliği Konuları

Editör: Dr.Öğr.Üyesi Ali PAŞAOĞLU

E_ISBN 978-625-5547-72-9

Mart 2025 – Afyonkarahisar

Dizgi/Mizanpaj: YAZ Yayınları

Kapak Tasarım: YAZ Yayınları

YAZ Yayınları. Yayıncı Sertifika No: 73086

M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar/AFYONKARAHİSAR

www.yazyayinlari.com

yazyayinlari@gmail.com

info@yazyayinlari.com

İÇİNDEKİLER

GENETİK ALGORİTMA KULLANILARAK EĞİTİLMİŞ ŞABLON EŞLEME YÖNTEMLERİYLE MEME BÖLGESİNİN BÖLÜTLENDİRİLMESİ VE KİTLELERİN TESPİT EDİLMESİ[1]

Serhat ÖZEKES[2]

Ali Yılmaz ÇAMURCU[3]

1. GİRİŞ

Bilgisayar destekli tespit (BDT), ileri görüntü işleme ve örüntü tanıma teknikleri kullanılarak radyoloji uzmanlarına medikal görüntülerdeki anormalliklerin tespit edilmesi çalışmalarında yardımcı olunmasıdır. BDT, teşhis aşamasında radyologlara destek verecek faydalı çıkarımlar sağlaması, karar vermeyi hızlandırması, insan hatasının teşhisteki yerini azaltması ve sağlık sektöründe maliyetleri düşürmesi gibi avantajlarından dolayı tıbbi görüntüye dayalı bilgisayar destekli teşhis teknikleri günümüzde önem kazanmış ve güncel teknolojilerden biri haline gelmiştir.

Tıp alanında kaydedilen teknolojik yeniliklerin temel olarak iki amacı vardır: insanlara sağlanan sağlık hizmetlerinin kalitesinin arttırılması ve sağlık sektöründeki maliyetlerin düşürülmesi. Bu çalışma sonucunda ortaya çıkan uygulamaların

---

[1] Bu çalışma Serhat ÖZEKES'in Tıbbi Görüntülemede Bilgisayar Destekli Tespit isimli doktora tezinden üretilmiştir.

[2] Prof.Dr., Marmara Üniversitesi, Teknoloji Fakültesi, Bilgisayar Mühendisliği Bölümü, serhat.ozekes@marmara.edu.tr, ORCID: 0000-0002-7432-0272.

[3] Prof.Dr., Fatih Sultan Mehmet Vakıf Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği, ycamurcu@fsm.edu.tr, ORCID: 0000-0003-1409-9905.

her iki konuda da tıp dünyasına önemli yararlar sağlayacağı düşünülmektedir. Mamografi, meme kanseri taramasında temel yöntemdir. Meme görüntülerinde rastlanabilen anormalliklerden bazıları kitleler, mikrokalsifikasyonlar ve yapısal distorsiyonlardır. Bu çalışma kapsamında kitlelerin yerlerinin bilgisayar aracılığıyla tespit edilmesi amaçlanmıştır.

Meme görüntülerinde karşılaşılan kitleler çeşitlilik gösterirler. Kitleler şekil, yoğunluk ve kenar gibi özellikleriyle kategorize edilirler. Şekilleri yuvarlak, oval, yuvarlak çıkıntılı ve düzensiz olabilir. Yoğunlukları yüksek yoğunluklu, düşük yoğunluklu, eşit yoğunluklu ve yağ içerikli olabilir. Kenar özellikleri de düzgün kenarlı (circumscribed) ve tırtıklı (spiküler) kenarlı olabilir. Bu kategoriler radyoloji uzmanlarına kitlelerin benign veya malign oldukları konusunda karar vermelerini kolaylaştırır (Sample, 2003).

Literatürdeki çalışmalar incelendiğinde mamografi görüntülerinde meme bölgesinin elde edilmesini sağlayan bölütleme aşaması, anormallik arama alanını daraltarak BDT'nin hızını arttırdığı görülmektedir. Kural tabanlı eşikleme yöntemi ile bölütleme, yoğunluk değişimini kontrol eden histogram tabanlı bölütleme ve bölge büyütme tabanlı bölütleme literatürde en sık karşılaşılan bölütleme yöntemlerindendir. İkinci bir aşama olarak karşımıza çıkan ilgi alanlarının belirlenmesi işlemindeki amaç anormal olmaya aday yapıların ortaya çıkarılması ve arama işleminin bu yapılar üzerinde gerçekleştirilerek sonucun en kısa sürede elde edilmesidir. Literatürde ilgi alanları tespiti amacıyla genel olarak kullanılan yöntemler gri seviyeli eşikleme, filtreleme, bulanık mantık ve kümeleme yöntemleridir. Son aşamada ilgi alanlarının sınıflandırılması ve anormalliklerin tespiti gerçekleştirilmektedir. Bu amaçla kural tabanlı yaklaşımlar, şablon eşleme, en yakın komşu kümelemesi, Markov rastgele alanları, yapay sinir ağları ve Bayes sınıflandırıcı kullanılan yöntemlerden birkaçıdır.

Brzakovic ve Neskovic meme bölgesinin bölütlendirilmesi için bulanık mantık tabanlı bir algoritma kullanmışlardır (Brzakovic vd., 1993). Petrick ve arkadaşları mamografik kitlelerin tespiti için iki basamaklı bir yoğunluk ağırlıklı parlaklık arttırma algoritması kullanmışlardır (Petrick vd., 1996). Chan ve arkadaşları uzamsal gri seviye bağımlılık matrislerinden çıkarılan doku özelliklerini kullanarak kitlelerin normal yapılardan ayrılmalarını sağlamışlardır (Chan vd., 1995). Yin ve arkadaşları kitle lezyonlarının belirlenmesi için mamografik asimetrileri tespit etmişlerdir (Yin vd., 1993).

2. MATERYAL VE YÖNTEMLER

2.1. Kullanılan Veri Seti

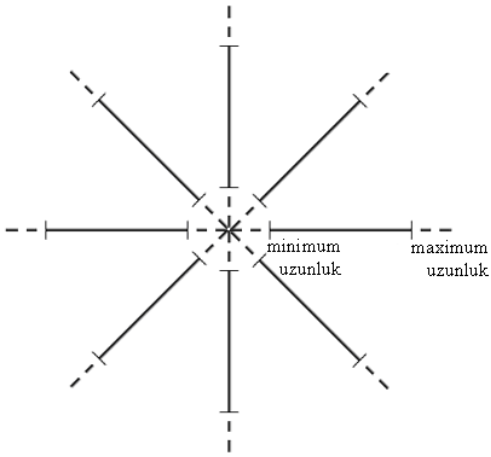
Bu çalışma kapsamında gerçekleştirilen mamografi görüntülerindeki kitlelerin tespitini gerçekleştiren BDT sisteminin geliştirilmesi ve değerlendirilmesi için MiniMIAS (Mammographic Images Analysis Society) veri seti kullanılmıştır (Suckling vd., 1994). Bu veri seti yaşları 50 ve 65 arasında değişen 161 hastanın sağ ve sol meme görüntülerini içermektedir. Tüm görüntüler 1024 X 1024 piksel çözünürlüğünde ve 8 bit gri seviyede dijital hale getirilmiştir. Tüm görüntülerdeki anormalliklerin konumları da ayrı bir metin dosyasında mevcuttur.

2.2. İlgi Alanlarının Belirlenmesi

Medikal görüntülerde bilgi verici işaretler, görüntü arka planı, kemikler, kaslar, damarlar ve ilgisiz doku alanları gibi anormal yumuşak doku yapılarıyla ilgisi bulunmayan yapılar da mevcuttur. Bu yapılar sistemin karmaşıklığını arttırmakta ve sistem duyarlılığını tehdit etmektedirler. Bu bölümde anlatılan yöntem ile karmaşıklık yaratan yapılar çıkartılarak sadece anormal olmaya aday yapılar elde edilmiştir. Bu yapılara ilgi

alanları adı verilmektedir. Böylece anormal yapılar sadece bu ilgi alanları içinde aranmakta ve sonuç çok daha kısa sürede elde edilmektedir.

Eldeki veri seti incelendiğinde mamografi görüntülerindeki kitlelerin kendi aralarında benzer fakat normal yapıların büyük bir kısmından farklı morfolojik yapıda oldukları anlaşılmıştır. Damarlar ve kemikler gibi normal yapılar ince uzun yapıdayken, anormal yapılar daha dairesel ve belirli bir büyüklüktedir. Yani anormal alanlarının büyüklüklerinin alt ve üst sınırları vardır. Bu nedenle anormal olmaya aday bölgelerinin çapları önemle incelenmelidir. Bu nedenle eşikleme önişleminden geçirilmiş olan görüntüler piksel piksel taranmış ve bir pikselin anormal aday bölgesinde olup olmadığının anlaşılması için, söz konusu piksel merkez alınıp Şekil 1'de görüldüğü gibi 8 yönlü inceleme gerçekleştirilmiştir.



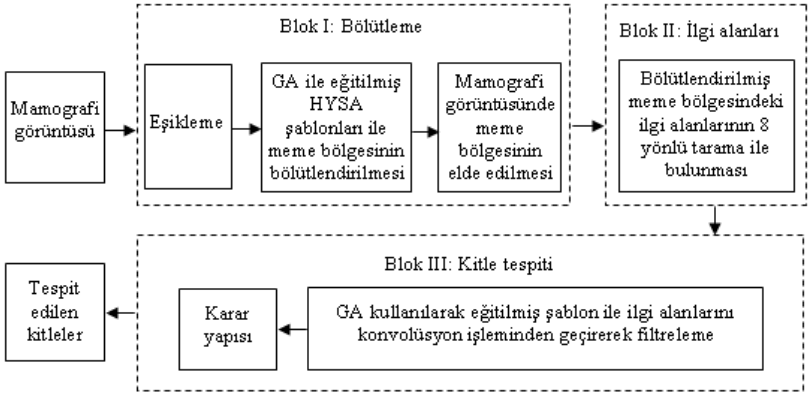
Şekil 1. 8 Yöndeki Eşik Değerleri

Burada "minimum uzunluk eşiği" alt sınırı, "maksimum uzunluk eşiği" ise üst sınırı temsil etmektedir. Bir piksel 8 yönde, "minimum uzunluk eşiği" değerinden daha az veya "maksimum uzunluk eşiği" değerinden daha fazla komşuya sahipse, söz konusu piksel anormal aday bölgesinde olamaz. Aksi halde söz

konusu piksel ilgi alanı bölgesindedir. Minimum ve maksimum uzunluk eşik değerleri bilgi verici işaretler ve damarlar gibi çok büyük veya çok küçük yapıların ilgi alanında yer almasını önlemek amacıyla kullanılmaktadır.

2.3.Genetik Algoritma Kullanılarak Eğitilmiş Şablon Eşleme Yöntemleriyle Meme Bölgesinin Bölütlendirilmesi ve Kitlelerin Tespit Edilmesi

Bu çalışmada meme bölgesinin bölütlendirilmesi ve ilgi alanlarını arama işleminin bu bölgede yapılması sağlanmıştır. Böylece ilgi alanlarının belirlenmesi işlemi hızlandırılması amaçlanmıştır. Belirlenmiş olan ilgi alanlarının sınıflandırılması amacıyla şablon eşleme yöntemi kullanılmıştır. Kullanılan şablonu oluşturan değerler genetik algoritma kullanılarak hesaplanmış ve amaca uygun en iyi şablon değerleri kullanılmıştır. Ardından hesaplanmış olan bu en iyi şablon ile eşleme yapılırken benzerlik ölçümü olarak konvolüsyon tabanlı filtreleme kullanılmıştır. Bu çalışmada kullanılan yöntem Şekil 2'de detaylı bir şekilde verilmiştir.



Şekil 2. GA ile Eğitilen Şablonlar Kullanılarak Bölütlendirme ve Kitle Tespiti Aşamalarının Gösterimi

Bu çalışmada ilk olarak meme bölgesini bölütlendirecek HYSA şablonlarının eğitimi gerçekleştirilmiştir. Bu eğitim için kullanılan genetik algoritma işlem basamakları şu şekildedir:

1. Basamak. İlk popülasyonun oluşturulması: Başlangıç popülasyon matrisi *m x n* boyutlarında rasgele oluşturulmuştur.

2. Basamak. HYSA şablonlarının oluşturulması: Kromozomlar HYSA şablonları olan A, B ve I şablonlarının ikili kodlanmasından oluşturulmuştur. Şablon değerleri ise [-1,1] değer aralığında kromozom kodlarının çözülmesiyle hesaplanmıştır. HYSA şablonları için komşuluk derecesi 1 olarak seçildiğinde A ve B şablonları 3 x 3'lük matris oluştururken, I şablonu her bir elemanı aynı olan *m x n*'lik bir matris oluşturur. Bu A ve B şablonları simetrik olarak şu şekilde verilir:

$$A = \begin{bmatrix} a_2 & a_1 & a_2 \\ a_1 & a_0 & a_1 \\ a_2 & a_1 & a_2 \end{bmatrix}, \quad \mathbf{B} = \begin{bmatrix} b_2 & b_1 & b_2 \\ b_1 & b_0 & b_1 \\ b_2 & b_1 & b_2 \end{bmatrix}$$

Bu durumda *S* ile ifade edilen değişkenlerin sayısı 7 olur ve *S*'nin her bir elemanı ikili olarak kodlanmıştır:

$$S = \begin{bmatrix} a_0 & a_1 & a_2 & b_0 & b_1 & b_2 & i \end{bmatrix}$$

burada *i* ile ifade edilen değişken I matrisinin tüm elemanlarını oluşturan değerdir.

Meme bölgesinin bölütlendirilmesi için A ve B şablonları için komşuluk derecesi olarak 5 değeri kullanılmıştır. Şablonlar simetrik tasarlanmasından dolayı her bir şablon 21 değişkenden oluşmuştur. Böylece S değişken vektörü A şablonu için 21 değişken, B şablonu için 21 değişken ve I şablonu için 1 değişken olmak üzere toplam 43 değişkene sahiptir. 11 x 11 piksel boyutlarındaki A ve B şablonları aşağıda görülmektedir. Bölütlendirme işleminin hesaplama süresini azaltmak amacıyla mamografi görüntülerinin boyutları 128 x 128 piksel değerine

düşürülmüştür. Meme bölgesinin bölütlendirilme işlemi başarıyla tamamlanmasının ardından mamografi görüntüsünün boyutları tekrar eski değeri olan 1024 x 1024 piksel değerine büyütülmüştür.

$$A=\begin{bmatrix} A0 & A1 & A2 & A3 & A4 & A5 & A4 & A3 & A2 & A1 & A0 \\ A1 & A6 & A7 & A8 & A9 & A10 & A9 & A8 & A7 & A6 & A1 \\ A2 & A7 & A11 & A12 & A13 & A14 & A13 & A12 & A11 & A7 & A2 \\ A3 & A8 & A12 & A15 & A16 & A17 & A16 & A15 & A12 & A8 & A3 \\ A4 & A9 & A13 & A16 & A18 & A19 & A18 & A16 & A13 & A9 & A4 \\ A5 & A10 & A14 & A17 & A19 & A20 & A19 & A17 & A14 & A10 & A5 \\ A4 & A9 & A13 & A16 & A18 & A19 & A18 & A16 & A13 & A9 & A4 \\ A3 & A8 & A12 & A15 & A16 & A17 & A16 & A15 & A12 & A8 & A3 \\ A2 & A7 & A11 & A12 & A13 & A14 & A13 & A12 & A11 & A7 & A2 \\ A1 & A6 & A7 & A8 & A9 & A10 & A9 & A8 & A7 & A6 & A1 \\ A0 & A1 & A2 & A3 & A4 & A5 & A4 & A3 & A2 & A1 & A0 \end{bmatrix}$$

$$B=\begin{bmatrix} B0 & B1 & B2 & B3 & B4 & B5 & B4 & B3 & B2 & B1 & B0 \\ B1 & B6 & B7 & B8 & B9 & B10 & B9 & B8 & B7 & B6 & B1 \\ B2 & B7 & B11 & B12 & B13 & B14 & B13 & B12 & B11 & B7 & B2 \\ B3 & B8 & B12 & B15 & B16 & B17 & B16 & B15 & B12 & B8 & B3 \\ B4 & B9 & B13 & B16 & B18 & B19 & B18 & B16 & B13 & B9 & B4 \\ B5 & B10 & B14 & B17 & B19 & B20 & B19 & B17 & B14 & B10 & B5 \\ B4 & B9 & B13 & B16 & B18 & B19 & B18 & B16 & B13 & B9 & B4 \\ B3 & B8 & B12 & B15 & B16 & B17 & B16 & B15 & B12 & B8 & B3 \\ B2 & B7 & B11 & B12 & B13 & B14 & B13 & B12 & B11 & B7 & B2 \\ B1 & B6 & B7 & B8 & B9 & B10 & B9 & B8 & B7 & B6 & B1 \\ B0 & B1 & B2 & B3 & B4 & B5 & B4 & B3 & B2 & B1 & B0 \end{bmatrix}$$

3. Basamak. Amaç ve uygunluk fonksiyonlarının hesaplanması: Bu aşamada eğitim görüntüsü olarak seçilen görüntü, ilk kromozoma bağlı olarak çalışan HYSA sisteminin girişine uygulanmıştır. HYSA çıkışı kararlı olduğunda elde edilen çıkış görüntüsü ile istenen görüntü arasındaki amaç fonksiyonu hesaplanmıştır. Bu işlem popülasyondaki her bir kromozoma ait olan şablon kümeleri ile tekrar edilmiştir. Bu çalışmada kullanılan amaç fonksiyon denklem, şu şekildedir (Matsumoto vd., 1990):

$$amaç(A,B,I)=\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}P_{i,j}\oplus T_{i,j}$$

Bu fonksiyonda P ve T ifadeleri sırasıyla HYSA çıkış görüntüsünü ve hedef görüntüyü belirtmektedir. $\oplus$ sembolü P ve

T 'nin pikselleri arasındaki XOR işlemi temsil etmektedir. Amaç fonksiyonunun bu şekilde hesaplanmasından sonra, her bir kromozom için uygunluk değeri şu denklem ile hesaplanmıştır:

$$uygunluk(A,B,I) = m \times n - amaç(A,B,I)$$

Durdurma kriteri olarak şu denklem kullanılmıştır:

$$durdurmakriteri = 0.99 \times m \times n$$

4. Basamak. Çaprazlama ve Mutasyon: Çaprazlama ve mutasyon ile yeni bir kuşağın oluşması sağlanmıştır. Durdurma kriteri sağlandığında ve genetik algoritma sonlandığında, en iyi A, B ve I şablonları elde edilmiştir.

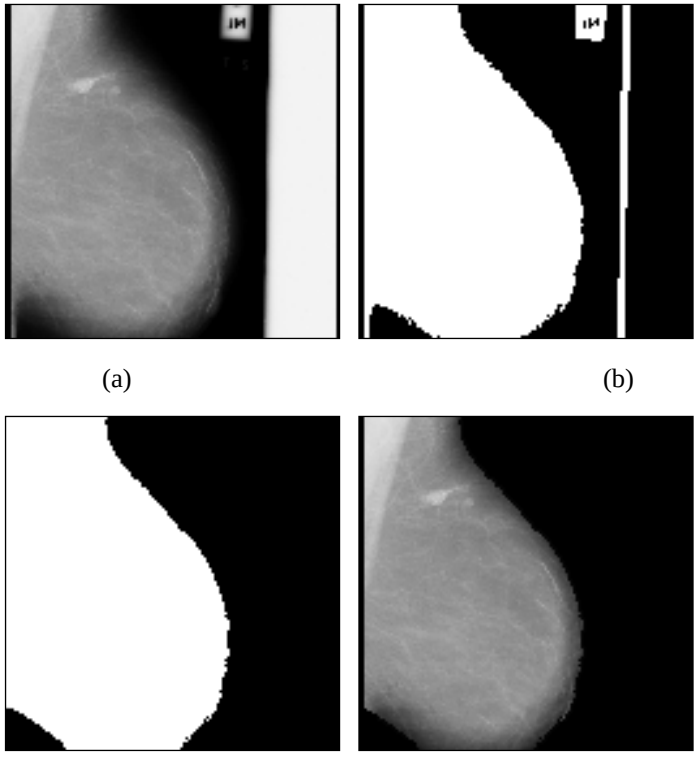
Bu çalışmada HYSA şablonlarının eniyilemesinde genetik algoritma yöntemi kullanılmıştır. Elde edilen şablonlar ile yeni bir görüntü HYSA girişine uygulandığında, bu görüntü için meme bölgesi elde edilmiştir. Orijinal mamografi görüntüsü HYSA girişine uygulanmadan önce bir eşikleme işleminden geçirilmelidir. Şekil 3(a)'da görüldüğü gibi meme kitlesi, kaslar ve dokular görüntü arka planından daha açık renktedir. Bu daha büyük yoğunluk değerlerine sahip olduklarını göstermektedir. Bu yüzden meme bölgesinin kabaca elde edilmesi için öncelikle eşikleme yöntemi uygulanmıştır. *I(x,y)* ile ifade edilen Şekil 3(a)'daki görüntüye aşağıdaki kural uygulanırsa Şekil 3(b)'deki görüntü elde edilir. Bu kuralda 1 beyaz rengi temsil ederken, 0 siyah rengi temsil etmektedir.

EĞER I(x,y) > 30 İSE

I(x,y) = 1

DEĞİLSE

I(x,y) = 0



(a) (b)

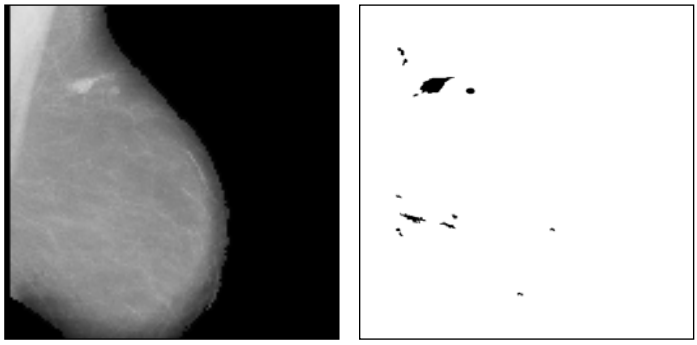
(c) (d)

Şekil 3. a) Sınırları Düzgün Kitle İçeren Orijinal Mamografi Görüntüsü, (b) HYSA'nın Girişini Oluşturan Eşikleme Kuralı ile Elde Edilen Görüntü, (c) HYSA Kullanılarak Bölütlendirilmiş Meme Bölgesi, (d) Meme Bölgesinin Orijinal Görüntüsü.

Şekil 3(b)'deki görüntü eğitilmiş A, B ve I şablonları ile çalışan HYSA girişine uygulandığında Şekil 3(c)'deki görüntü elde edilir. Şekil 3(d)'de orijinal mamografi görüntüsündeki meme bölgesini göstermektedir.

Bölütlenmiş meme görüntüsüne 8 yönlü ilgi alanı arama yöntemi uygulandığında Şekil 4(b)'de görülen ilgi alanları elde edilir. Bu noktadaki amaç ilgi alanlarının kitle veya normal yapı olarak sınıflandırılmasıdır. Önceki aşamada bulunmuş olan ilgi alanları incelendiğinde farklı yapılarda oldukları gözlemlenmiştir. Kitleler daha kalın ve dairesel iken, diğer

yapılar daha ince uzundur. Buna göre yapısal özellikleri kullanarak kitleleri normal yapılardan ayırmak için kitle şablonu kullanılmıştır. Bu şablonun değerlerinin hesaplanması da genetik algoritma ile gerçekleştirilmiştir. Şablonun eğitimi için gerçekleştirilen genetik algoritma basamakları şu şekildedir:



(a) (b)

Şekil 4. (a) Sınırları Düzgün Kitle İçeren Mamografi Görüntüsü, (b) Bulunan İlgi Alanları

1. Basamak. İlk popülasyonun oluşturulması: Başlangıç popülasyon matrisi *m x n* boyutlarında rastsal olarak oluşturulmuştur.

2. Basamak. Kitle şablonunun oluşturulması: Kromozomlar kitle şablonu olan *T* şablonunun ikili kodlanmasından oluşturulmuştur. Şablon değerleri ise [-5, 5] değer aralığında kromozom kodlarının çözülmesiyle hesaplanmıştır. 16 x 16 piksel boyutlarında olan *T* kitle şablonu şu şekildedir:

$$T=\begin{bmatrix}
t_{28} & t_{29} & t_{30} & t_{31} & t_{32} & t_{33} & t_{34} & t_{35} & t_{35} & t_{34} & t_{33} & t_{32} & t_{31} & t_{30} & t_{29} & t_{28}\\
t_{29} & t_{21} & t_{22} & t_{23} & t_{24} & t_{25} & t_{26} & t_{27} & t_{27} & t_{26} & t_{25} & t_{24} & t_{23} & t_{22} & t_{21} & t_{29}\\
t_{30} & t_{22} & t_{15} & t_{16} & t_{17} & t_{18} & t_{19} & t_{20} & t_{20} & t_{19} & t_{18} & t_{17} & t_{16} & t_{15} & t_{22} & t_{30}\\
t_{31} & t_{23} & t_{16} & t_{10} & t_{11} & t_{12} & t_{13} & t_{14} & t_{14} & t_{13} & t_{12} & t_{11} & t_{10} & t_{16} & t_{23} & t_{31}\\
t_{32} & t_{24} & t_{17} & t_{11} & t_{6} & t_{7} & t_{8} & t_{9} & t_{9} & t_{8} & t_{7} & t_{6} & t_{11} & t_{17} & t_{24} & t_{32}\\
t_{33} & t_{25} & t_{18} & t_{12} & t_{7} & t_{3} & t_{4} & t_{5} & t_{5} & t_{4} & t_{3} & t_{7} & t_{12} & t_{18} & t_{25} & t_{33}\\
t_{34} & t_{26} & t_{19} & t_{13} & t_{8} & t_{4} & t_{1} & t_{2} & t_{2} & t_{1} & t_{4} & t_{8} & t_{13} & t_{19} & t_{26} & t_{34}\\
t_{35} & t_{27} & t_{20} & t_{14} & t_{9} & t_{5} & t_{2} & t_{0} & t_{0} & t_{2} & t_{5} & t_{9} & t_{14} & t_{20} & t_{27} & t_{35}\\
t_{35} & t_{27} & t_{20} & t_{14} & t_{9} & t_{5} & t_{2} & t_{0} & t_{0} & t_{2} & t_{5} & t_{9} & t_{14} & t_{20} & t_{27} & t_{35}\\
t_{34} & t_{26} & t_{19} & t_{13} & t_{8} & t_{4} & t_{1} & t_{2} & t_{2} & t_{1} & t_{4} & t_{8} & t_{13} & t_{19} & t_{26} & t_{34}\\
t_{33} & t_{25} & t_{18} & t_{12} & t_{7} & t_{3} & t_{4} & t_{5} & t_{5} & t_{4} & t_{3} & t_{7} & t_{12} & t_{18} & t_{25} & t_{33}\\
t_{32} & t_{24} & t_{17} & t_{11} & t_{6} & t_{7} & t_{8} & t_{9} & t_{9} & t_{8} & t_{7} & t_{6} & t_{11} & t_{17} & t_{24} & t_{32}\\
t_{31} & t_{23} & t_{16} & t_{10} & t_{11} & t_{12} & t_{13} & t_{14} & t_{14} & t_{13} & t_{12} & t_{11} & t_{10} & t_{16} & t_{23} & t_{31}\\
t_{30} & t_{22} & t_{15} & t_{16} & t_{17} & t_{18} & t_{19} & t_{20} & t_{20} & t_{19} & t_{18} & t_{17} & t_{16} & t_{15} & t_{22} & t_{30}\\
t_{29} & t_{21} & t_{22} & t_{23} & t_{24} & t_{25} & t_{26} & t_{27} & t_{27} & t_{26} & t_{25} & t_{24} & t_{23} & t_{22} & t_{21} & t_{29}\\
t_{28} & t_{29} & t_{30} & t_{31} & t_{32} & t_{33} & t_{34} & t_{35} & t_{35} & t_{34} & t_{33} & t_{32} & t_{31} & t_{30} & t_{29} & t_{28}
\end{bmatrix}$$

Görüldüğü gibi *T* şablonunun simetrik olmasından dolayı *S* ile ifade edilen değişkenlerin sayısı 36 olur ve *S*'nin her bir elemanı aşağıda görüldüğü gibi ikili olarak kodlanmıştır:

$$S=\begin{bmatrix}t_0 & t_1 & t_2 & . & . & . & t_{33} & t_{34} & t_{35}\end{bmatrix}$$

3. Basamak. Amaç ve uygunluk fonksiyonlarının hesaplanması: Bu aşamada eğitim görüntüsü olarak seçilen görüntü, ilk kromozoma bağlı olarak çalışan *T* şablonu ile konvole edilmiştir. Bu konvolüsyon işleminin amacı giriş görüntüsündeki normal yapılardan kitleyi ayırmaktır. *n x m* boyutlarındaki şablonun *T(x, y)* ile ve *M x N* boyutlarındaki giriş görüntüsünün *I(x, y)* ile temsil edildiği düşünülürse, *T*'nin *I* ile konvolüsyonu denklem şu şekilde ifade edilir (Chua vd., 1988):

$$C(x,y)=T*I(x,y)=\sum_{i=0}^{n-1}\sum_{j=0}^{m-1}T(i,j)I(x-i,y-i)$$

Burada * işareti konvolüsyon işlemini temsil etmektedir. Konvolüsyon işlemi sonucunda şablona benzeyen yapılar daha güçlenirler. Bu yapıların piksel değerleri yüksek pozitif değerler

alırken, şablona benzemeyen yapıların piksel değerleri yüksek negatif değerler alır. Bu yüzden konvolüsyon işleminden geçirilmiş *C* görüntüsünden kitleye benzer yapıları elde etmek amacıyla şu karar kuralı kullanılmıştır:

EĞER $C(x,y) > 1$ *İSE*

$C(x,y) = 1$

DEĞİLSE

$C(x,y) = -1$

Konvolüsyon ve karar kuralının ardından çıktı görüntüsü olan *C* ve hedef görüntü *A* arasındaki amaç fonksiyonu hesaplanmıştır. Bu işlem popülasyondaki her bir kromozoma ait olan şablon kümeleri ile tekrar edilmiştir. Bu çalışmada kullanılan amaç fonksiyonu şu şekildedir (Matsumoto vd., 1990):

$$amaç(T) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} C_{i,j} \oplus A_{i,j}$$

Bu fonksiyonda $\oplus$ sembolü *C* ve *A* 'nın pikselleri arasındaki XOR işlemi temsil etmektedir. Amaç fonksiyonunun bu şekilde hesaplanmasından sonra, her bir kromozom için uygunluk değeri şu denklem ile hesaplanmıştır:

$$uygunluk(T) = M \times N - amac(T)$$

Durdurma kriteri olarak şu denklem kullanılmıştır:

$$durdurmakriteri = 0.99 \times M \times N$$

4. Basamak. Çaprazlama ve Mutasyon: Çaprazlama ve mutasyon ile yeni bir kuşağın oluşması sağlanmıştır. Durdurma kriteri sağlandığında ve genetik algoritma sonlandığında, en iyi *T* kitle şablonu elde edilmiştir.

Gerçek meme kitlelerinin morfolojileri kullanılarak normal yapılardan ayrılması için bu çalışmada şablon eşleme

yöntemi kullanılmıştır. Mamografi görüntüsü kitle şablonu ile taranmış ve şablona benzeyen yapılar aranmıştır. Böylece çok küçük, çok ince veya çok uzun yapılar elenmiştir. Kitle şablonunun değerlerinin hesaplanması için genetik algoritma eniyileme yöntemi olarak kullanılmıştır. En iyi *T* kitle şablonunun bulunmasının ardından, ilgi alanlarından oluşan görüntü *T* şablon ile konvolüsyon işlemine tutulmuş ve gerçek kitlelerin bulunması için karar kuralı uygulanmıştır.

3. SONUÇLAR

Bu çalışmada gerçekleştirilen BDT yazılımı sınırları düzgün kitle içeren 22 görüntü, spiküler kitle içeren 19 görüntü ve 11 adet normal görüntü olmak üzere toplam 52 adet görüntü üzerinde test edilmiştir. Meme bölgesinin bölütlendirilmesi için genetik algoritma ile eğitilmiş HYSA şablonları kullanılmıştır. Eğitim işlemi sonucunda elde edilen genetik algoritma parametreleri Tablo 1'de görülmektedir. A ve B şablonları da aşağıda görülmektedir. I şablonun değeri de -4.1373 olarak elde edilmiştir.

Bölütlendirilmiş meme bölgesi üzerinde 8 yönlü tarama yapılarak 273 adet ilgi alanı bulunmuştur. Bu ilgi alanlarının sınıflandırılması için boyutları 16 x 16 piksel olan T kitle şablonu kullanılmıştır. T şablonu değerlerinin hesaplanması için genetik algoritma kullanılmıştır. Eğitim işlemi sonucunda elde edilen genetik algoritma parametreleri Tablo 2'de ve kitle şablonunun değerleri de Tablo 3'de görülmektedir.

Tablo 1. Meme Bölgesinin Bölütlendirilmesinde Kullanılan HYSA Şablonlarının GA ile Eğitim Parametreleri

Parametreler:	**Değerler**
Popülasyon başına düşen kromozom sayısı	100
Değişken başına düşen Bit sayısı	8
Değişken sayısı	43
Kromozom uzunluğu	344
Popülasyondaki toplam bit sayısı	34400
Üreme için kullanılan çaprazlama olasılığı	70%
Mutasyon olasılığı	1%
Kuşak farkı	98%
Şablon parametreleri değer aralığı	[-5, 5]

$$A=\begin{bmatrix}
-2.49 & 3.43 & -0.84 & 0.10 & 1.24 & 4.45 & 1.24 & 0.10 & -0.84 & 3.43 & -2.49 \\
3.43 & 2.14 & 4.06 & 3.78 & 1.86 & 1.78 & 1.86 & 3.78 & 4.06 & 2.14 & 3.43 \\
-0.84 & 4.06 & -1.12 & 2.25 & 3.08 & 3.12 & 3.08 & 2.25 & -1.12 & 4.06 & -0.84 \\
0.10 & 3.78 & 2.25 & -4.61 & 2.88 & -4.53 & 2.88 & -4.61 & 2.25 & 3.78 & 0.10 \\
1.24 & 1.86 & 3.08 & 2.88 & 4.37 & -3.24 & 4.37 & 2.88 & 3.08 & 1.86 & 1.24 \\
4.45 & 1.78 & 3.12 & -4.53 & -3.24 & 2.45 & -3.24 & -4.53 & 3.12 & 1.78 & 4.45 \\
1.24 & 1.86 & 3.08 & 2.88 & 4.37 & -3.24 & 4.37 & 2.88 & 3.08 & 1.86 & 1.24 \\
0.10 & 3.78 & 2.25 & -4.61 & 2.88 & -4.53 & 2.88 & -4.61 & 2.25 & 3.78 & 0.10 \\
-0.84 & 4.06 & -1.12 & 2.25 & 3.08 & 3.12 & 3.08 & 2.25 & -1.12 & 4.06 & -0.84 \\
3.43 & 2.14 & 4.06 & 3.78 & 1.86 & 1.78 & 1.86 & 3.78 & 4.06 & 2.14 & 3.43 \\
-2.49 & 3.43 & -0.84 & 0.10 & 1.24 & 4.45 & 1.24 & 0.10 & -0.84 & 3.43 & -2.49
\end{bmatrix}$$

$$B=\begin{bmatrix}
-0.88 & -0.29 & -3.63 & 2.61 & -3.63 & 2.69 & -3.63 & 2.61 & -3.63 & -0.29 & -0.88 \\
-0.29 & -3.27 & 0.49 & -0.18 & -1.55 & -4.02 & -1.55 & -0.18 & 0.49 & -3.27 & -0.29 \\
-3.63 & 0.49 & -3.35 & 1.75 & 0.88 & 3.39 & 0.88 & 1.75 & -3.35 & 0.49 & -3.63 \\
2.61 & -0.18 & 1.75 & 2.69 & 1.16 & 1.08 & 1.16 & 2.69 & 1.75 & -0.18 & 2.61 \\
-3.63 & -1.55 & 0.88 & 1.16 & -2.33 & 4.92 & -2.33 & 1.16 & 0.88 & -1.55 & -3.63 \\
2.69 & -4.02 & 3.39 & 1.08 & 4.92 & 4.61 & 4.92 & 1.08 & 3.39 & -4.02 & 2.69 \\
-3.63 & -1.55 & 0.88 & 1.16 & -2.33 & 4.92 & -2.33 & 1.16 & 0.88 & -1.55 & -3.63 \\
2.61 & -0.18 & 1.75 & 2.69 & 1.16 & 1.08 & 1.16 & 2.69 & 1.75 & -0.18 & 2.61 \\
-3.63 & 0.49 & -3.35 & 1.75 & 0.88 & 3.39 & 0.88 & 1.75 & -3.35 & 0.49 & -3.63 \\
-0.29 & -3.27 & 0.49 & -0.18 & -1.55 & -4.02 & -1.55 & -0.18 & 0.49 & -3.27 & -0.29 \\
-0.88 & -0.29 & -3.63 & 2.61 & -3.63 & 2.69 & -3.63 & 2.61 & -3.63 & -0.29 & -0.88
\end{bmatrix}$$

Tablo 2. Kitle Şablonu Eğitimi İçin Kullanılan Genetik Algoritma Parametreleri

Parametreler:	**Değerler**
Popülasyon başına düşen kromozom sayısı	100
Değişken başına düşen Bit sayısı	8
Değişken sayısı	36
Kromozom uzunluğu	288
Popülasyondaki toplam bit sayısı	28800
Üreme için kullanılan çaprazlama olasılığı	70%
Mutasyon olasılığı	1%
Kuşak farkı	98%
Şablon parametreleri değer aralığı	[-1, 1]

Tablo 3. Genetik Algoritma ile Hesaplanmış Kitle Şablonu Değerleri

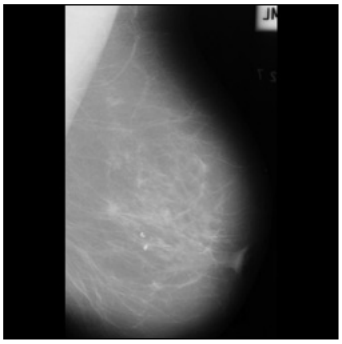
Parametre	**Değerler**	**Parametre**	**Değerler**
T_0	0.92941	t_{18}	0.05098
T_1	0.17647	t_{19}	-0.42745
T_2	0.98431	t_{20}	-0.85098
T_3	0.12157	t_{21}	-0.16863
T_4	0.1451	t_{22}	-0.066667
T_5	0.56863	t_{23}	0.69412
T_6	0.81176	t_{24}	0.41961
T_7	0.95294	t_{25}	-0.48235
T_8	0.23922	t_{26}	-0.45098
T_9	-0.15294	t_{27}	0.92157
T_{10}	0.62353	t_{28}	0.16863
T_{11}	0.51373	t_{29}	0.11373
T_{12}	0.17647	t_{30}	0.0039216
T_{13}	0.019608	t_{31}	-0.42745
T_{14}	0.43529	t_{32}	-0.67059
T_{15}	0.69412	t_{33}	-0.80392
T_{16}	0.38039	t_{34}	0.2
T_{17}	-0.2549	t_{35}	0.74118

İlgi alanlarından oluşan görüntü ile T şablonunun konvolüsyonu sonucunda ilgi alanları kural tabanlı sistem tarafından sınıflandırılmıştır. Elde edilen sonuçlar Tablo 4'de görülmektedir.

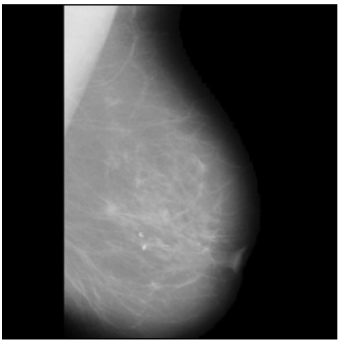
Tablo 4. Genetik Algoritma Kullanılarak Eğitilmiş Şablon Eşleme Yöntemleriyle Kitlelerin Tespit Edilmesi Çalışmasının Sonuçları

		Gerçek Bulgu	
		Hasta (H^+)	*Hasta Değil (H^-)*
BDT Gözlemi	*Pozitif (T^+)*	41 adet Doğru pozitif	48 adet Yanlış pozitif
	Negatif (T^-)	2 adet Yanlış negatif	182 adet Doğru negatif
Duyarlılık		% 95.3	
Görüntü başına düşen YP oranı		0.92	

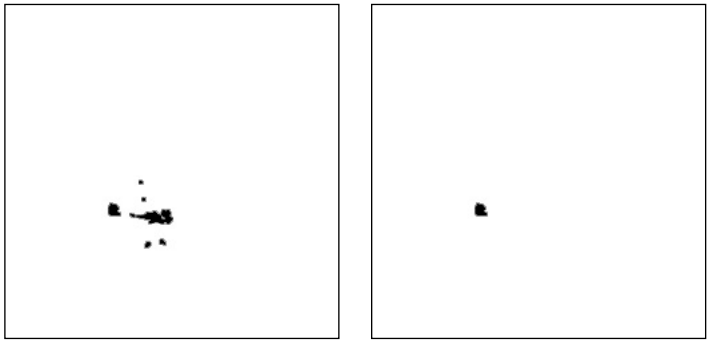
Şekil 5'de meme bölgesinin bölütlendirilmesi, ilgi alanlarının bulunması ve kitlelerin tespiti ile ilgili bir örnek görülmektedir. Şekil 5(a) spiküler kitle içeren bir mamogramı göstermektedir. Bu mamogramdan elde edilen meme bölgesi Şekil 5(b)'de, meme bölgesinden çıkarılan ilgi alanları da Şekil 5(c)'de görülmektedir. Bu ilgi alanları şablon ile karşılaştırılır ve benzer olanlar kitle olarak sınıflandırılmış ve Şekil 5(d)'de gösterilmiştir. Orijinal mamografi görüntü üzerindeki kitlenin yeri Şekil 5(e)'de görülmektedir.



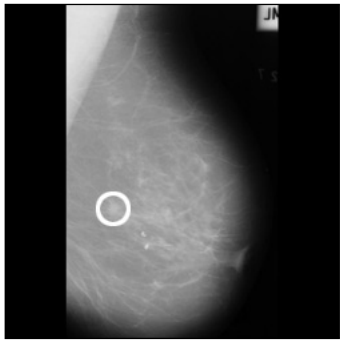
(a)



(b)



(c) (d)



(e)

Şekil 5a) Spiküler Kitle İçeren Mamografi Görüntüsü, (b) Bölütlendirilmiş Meme Bölgesi, (c) Meme Bölgesinde Belirlenmiş İlgi Alanları, (d) Şablon Eşleme Yöntemi İle Kitle Olarak Sınıflandırılmış İlgi Alanı, (e) Mamografi Görüntüsü Üzerinde Tespit Edilen Kitle

4. TARTIŞMA

Bu çalışmada tasarlanan BDT yazılımları sayesinde mamografi görüntülerindeki kitlelerin otomatik tespit edilmesi sağlanmıştır. Medikal görüntülerin analizinde bölütleme işlemi sonucu arama yapılacak alan daraltılmış olur. Bölütlenmiş görüntüde anormal yapı özelliği gösteren ilgi alanlarının belirlenmesi için her piksel için yoğunluk değerlerinin

sorgulandığı 8 yönlü tarama gerçekleştirilmiştir. Anormallik tespitinde kullanılan şablon eşleme tekniğinde, kullanılan şablon değerlerinin hesaplanması için genetik algoritma yöntemi kullanılmıştır. Benzerlik ölçümü olarak konvolüsyon yönteminin kullanıldığı bu çalışmada şablon değerlerinin optimizasyonu sağlanmıştır. Elde edilen sonuçlar literatürdeki benzer çalışmalar ile karşılaştırıldığında gerek duyarlılık gerekse de görüntü başına düşen yanlış pozitif oranlarıyla başarılı oldukları görülmektedir. İlerleyen çalışmalarda derin öğrenme yöntemlerinin medikal görüntülerindeki anormalliklerin tespitinde kullanılmasıdır.

KAYNAKÇA

Brzakovic, D. ve Neskovic, M. (1993). Mammogram screening using multiresolution-based image segmentation. *International Journal of Pattern Recognition and Artificial Intelligence, 7*(5), 1437-1460.

Chan, H., Wei, D., Helvie, N., Sahiner, B., Adler, D., Goodsitt, M. ve Petrick, N. (1995). Computer-aided classification of mammographic masses and normal tissue: Linear discriminant analysis in texture feature space. *Physics in Medicine and Biology, 40*(5), 857-876.

Chua, L. O. ve Yang, L. (1988). Cellular neural networks: Applications. *IEEE Transactions on Circuits and Systems, 35*(10), 1273-1290.

Matsumoto, T., Chua, L. O. ve Yokohama, T. (1990). Image thinning with a cellular neural network. *IEEE Transactions on Circuits and Systems, 37*(5), 638-640.

Petrick, N., Chan, H. P. ve Wei, D. (1996). An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection. *IEEE Transactions on Medical Imaging, 15*(1), 59-67.

Sample, J. T. (2003). *Computer assisted screening of digital mammogram images* (Doctoral dissertation). Louisiana State University.

Suckling, J., Parker, J., Dance, D., Astley, S., Hutt, I. ve Boggis, C. (1994). The mammographic images analysis society digital mammogram database. *Exerpta Medica, 1069*, 375-378.

Yin, F., Giger, M., Vyborny, C., Doi, K. ve Schmidt, R. (1993). Comparison of bilateral-subtraction and single-image processing techniques in the computerized detection of

mammographic masses. *Investigative Radiology*, *28*(8), 473-481.

PERFORMANCE COMPARISONS OF CENTRALITY ALGORITHMS WITH GREEDY MINIMUM APPROACH ON NP GRAPH PROBLEMS

Furkan ÖZTEMİZ[1]

1. INTRODUCTION

In this book chapter, we analyze the solution of maximum independent set and maximum matching, which are important NP problems in graph theory, using the greedy-min approach. Greedy-min (Gmin) approach is a very popular solution approach in the literature. In graph theory, this approach is usually realized by selecting nodes with minimum priority value. When the studies in the literature are examined, it is seen that centrality methods are used extensively in solving matching and independent set problems. In this study, we will investigate how both popular and newly introduced centrality methods will produce results with the Gmin approach. The centrality methods to be used in the analysis include Differential Malatya Centrality, Malatya Centrality, Pagerank, Betweenness, Degree, Eigenvector algorithms. Each of these centrality methods are important methods that can directly contribute to the solution of real world problems

[1] Asst. Prof. Dr., Inonu University, Faculty of Engineering, Software Engineering Department, furkan.oztemiz@inonu.edu.tr, ORCID: 0000-0001-5425-3474.

1.1. Related Works

There are many studies on centrality methods in the literature. Many studies on Pagerank, Degree, Betweenness, Eigenvector, Differential Malatya Centrality and Malatya Centrality are given below.

Centrality metrics in graph theory and network analysis have been extensively studied for determining node importance rankings. The PageRank algorithm and its derivatives have been applied in areas such as the two-layer PageRank model (Tortosa et al., 2021) with urban street networks (Bowater & Stefanakis, 2023), real-time temporal networks (Lv et al., 2019), and network-based feature selection algorithms (Hashemi et al., 2020). The efficient hybrid PageRank model (Shen et al., 2025) offers an effective approach for large-scale and layered networks. PageRank is also used for link prediction in directed temporal networks (Lv et al., 2022), in weighted and directed networks (Zhang et al., 2022), and in fully weighted graph structures (Hashemi et al., 2020). The Eigenvector Centrality metric has been applied to social network analysis (Parand et al., 2016), complex networks (Xu et al., 2023), and large datasets, developing effective node identification methods with fuzzy logic-based computations for platforms such as Facebook, Epinions, and Slashdot-Zoo. The Malatya Centrality metric has been applied to DIMACS benchmark graphs (Öztemiz & Yakut, 2024) and molecular graph structures (Yakut, 2024), and approaches have been developed to solve problems such as maximum independent set (Yakut et al., 2023), minimum vertex cover (Karci et al., 2022), and maximum clique (Yakut & Öztemiz, 2024). Malatya Centrality-based models have also been proposed for text summarization methods (Yakut & Bakan, 2023). Malatya Centrality-based methods have been developed for independent set, maximum clique, and vertex cover problems (Öztemiz, 2025). The Betweenness Centrality metric has been

applied to multi-agent pathfinding (Ewing et al., 2022), large-scale complex networks (Barthélemy, 2004), delay-sensitive networks (Zheng et al., 2023), and social opportunistic IoT (Social OppIoT) networks (Nigam et al., 2022), proving to be important for optimizing inter-node flows within networks. Eigenvector Centrality has been used to assess connectivity in large-scale road networks (Ando et al., 2020). Furthermore, research on hypergraph-based centrality models (Tudisco & Higham, 2021), social IoT networks (Nigam et al., 2022), anomaly detection in traffic networks (Lin et al., 2024), and edge domination-based centrality metrics (Amadeo et al., 2023) has provided new perspectives in centrality analysis. All of these studies demonstrate that centrality metrics have a broad range of applications in social network analysis, urban planning, biological networks, optimization problems, and graph theory-based computations.

2. MATERIAL AND METHOD

The presented study is summarized in three stages. In Stage 1, the graph to be analyzed is transformed into the appropriate format. In Stage 2, the centrality algorithms to be used in the comparative analysis are presented. During this stage, the centrality values produced by each centrality algorithm on the graph are calculated.

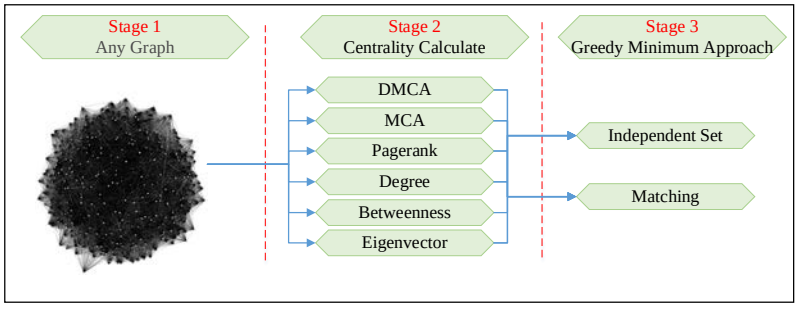


Figure. 1. Graphical Abstract

In Stage 3, the greedy minimum approach is applied, and the node with the lowest centrality value for each algorithm is selected as a member of the independent set. This node and its neighboring vertices are removed from the graph. The next iteration is then performed, and the process continues until no vertices remain in the graph. Another analysis performed at this stage is the identification of the maximum matching members. To identify the matching members, the line graph of the graph from Stage 1 is created. The line graph represents the edges of the graph as nodes. By doing so, the independent set members are used to determine the maximum independent edge set (maximum matching) on the line graph (Hemminger & Beineke, 1978).

Independent Set: It is a fundamental optimization problem. In graph theory, given a graph G = (V, E), a subset S of the vertex set V is called an independent set if no pair of vertices in S is connected by an edge in the edge set E. In other words, no vertex in the subset S is adjacent to another vertex in S (Yakut et al., 2023).

Matching: In graph theory, given a graph G = (V, E), a subset M of the edge set E is called a matching if no pair of edges in M shares a common vertex. In other words, all the edges in the set M are independent of each other [Carrabs et al., 2009].

2.1.Centrality Algorithms

For the comparison in the study, the following algorithms were chosen: Differential Malatya Centrality (DMC), Malatya Centrality (MC), PageRank Centrality (PC), Degree Centrality (DC), Betweenness Centrality (BC), and Eigenvector Centrality (EC). While most of these algorithms are widely used in the literature, methods such as DMC and MC have been newly introduced to the literature.

The DMC algorithm is a method developed to determine the dominance value of nodes. When calculating the centrality

value of the relevant node, its own node degree is subtracted individually from the degrees of its neighboring nodes, and these values are then summed to calculate the centrality. The centrality calculation for the DMC algorithm is performed as shown in Equation 1 (Öztemiz, 2025).

$$\Psi^{\partial}(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i) - d(v_j)}{d(v_j)}, 1 \leq i \leq |V| \wedge 1 \leq i \leq |V|, i \neq j \quad (1)$$

v_i signifies the vertex whose centrality value is to be determined, whereas v_j denotes its neighbors. d denotes the degree of the relevant vertex. $|V|$ indicates the total number of vertices in the graph.

The MC algorithm shares many similarities with the DMC algorithm in terms of implementation. The difference with MC is that, when calculating the centrality value of nodes, it does not consider the difference between neighboring nodes. Instead, it simply divides the degree of the node by the degrees of its neighboring nodes and sums the results. The MC method has produced successful results in solving many graph problems. The formula is given in Equation 2 (Yakut et al., 2023).

$$\Psi(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i)}{d(v_j)}, 1 \leq i \leq |V| \wedge 1 \leq i \leq |V|, i \neq j \quad (2)$$

The PageRank algorithm is a method developed to determine the importance of web pages. Essentially, it calculates the value of a web page based on the number of incoming links (backlinks) and the authority of the sources of these links. The more links a web page receives, the more important it is considered to be. If a high-authority page links to another page, this link carries more weight. The formula for the algorithm is given in Equation 3 (Hashemi et al., 2020).

$$PR(A) = (1 - d) + d \sum_{i=1}^{N} \frac{PR(L_i)}{C(L_i)} \quad (3)$$

$PR(A)$ represents the PageRank value. d is damping factor. L_i denotes the vertices that link to vertex A. $C(L_i)$ is the total number of outgoing links from vertex L_i

Degree centrality measures the number of incoming and outgoing connections to a vertex. While the method is used to determine the popularity of individual nodes, it is also used to identify the minimum, maximum, average degrees, and standard deviation values across the entire graph (Degree Centrality, 2025).

Betweenness centrality calculates how frequently a vertex appears on the shortest paths between other vertices. The more frequently a vertex is found on the shortest paths, the more critical it is as a bridge. The formula for the algorithm is given in Equation 4 (Barthélemy, 2004).

$$BC(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}} \quad (4)$$

$BC(v)$ represents the betweenness centrality value for vertex v. $\sigma_{st}(v)$ indicates how many of the shortest paths between vertices s and t pass through vertex v. σ_{st} represents the shortest path connection between s and t.

Eigenvector centrality is a method that measures the influence and connections of nodes. Vertices with high centrality contribute more than those with low centrality. As a natural consequence of this, it indicates that a vertex with a high score is connected to another vertex with a high score. The general formula is given in Equations 5 and 6 (Bonacich, 2007).

$$Ax = x, x_i = \sum_{j=1}^{n} a_{ij} x_{j,} \; i = 1, .., n \quad (5)$$

$$c(\beta) = \sum_{k=1}^{\infty} \beta^{k-1} A^k 1 \, , |\beta| < 1/\lambda \quad (6)$$

A is the adjacency matrix of the graph, and λ is the largest eigenvector of the matrix A.

3. EXPERIMENTAL RESULTS

Test operations have been performed on a variety of graphs with different structures. The comparative Independent Set and Matching results of the algorithms have been evaluated under separate headings. The primary objective of the experimental study is to determine the success of different centrality methods in solving two distinct graph problems using the Gmin approach. Independent Set tests were conducted on 46 different graphs, and Matching tests were performed on 24 different graphs. In total, 420 different experiments were conducted on 70 different graphs.

3.1. Independent Set

For the comparative test, a variety of graphs were generated, including lattice, bipartite, multipartite, social network, and random graphs. Table 1 presents the results of the DMC, MC, DC, BC, EC, and PC algorithms. Upon examining the table, for example, the Grid(6*10) graph consists of 60 vertices and 104 edge connections. When the greedy min approach is applied based on DMC values, the number of independent set members is determined to be 30. This value has also been determined to be 30 for MC, DC, and PC. For BC and EC, the number of independent set members is determined to be 24.

Table 1. Comparative Independent Set Results

Graphs	V	E	**Greedy Min – Indepedent Set** (Öztemiz, 2025)					
			DMC	MC	DC	BC	EC	PC
Grid(6*10)	60	104	30	30	30	24	24	30
Grid(5*5*5)	125	300	63	63	63	63	63	63
King(9*9)	81	272	25	25	25	25	25	25
Hexagonal(20*20)	880	1279	440	440	440	403	403	440
Knight(15*15)	225	728	113	113	103	82	84	113
Hypercube(10)	1024	5120	512	512	512	512	512	512
Banana Tree(7*7)	50	49	42	42	42	42	42	42
Folkman	20	40	10	10	10	10	10	10
(3*4*5)	12	47	5	5	5	5	5	5
(4*8*12)	24	176	12	12	12	12	12	12
(3*5*7*9*11*13)	48	925	13	13	13	13	13	13

(4*8*12*16*20*24)	84	2,800	24	24	24	24	24	24
Zachary	34	78	20	20	20	20	20	20
Dolphin	62	159	28	27	28	27	26	27
Zebra	27	111	7	7	7	7	7	7
Complex	70	133	27	27	27	27	27	27
100(p: 0.08)	100	403	33	33	33	33	31	33
100(p: 0.15)	100	732	25	24	22	23	22	24
100(p: 0.20)	100	990	18	18	17	18	16	18
250(p: 0.08)	250	2,504	50	50	47	49	42	50
250(p: 0.15)	250	4,702	30	29	28	28	28	29
250(p: 0.20)	250	6,259	24	24	23	24	23	24
500(p: 0.08)	500	10,015	64	62	63	60	58	65
500(p: 0.15)	500	18,757	37	37	35	35	35	37
500(p: 0.20)	500	24,963	28	29	28	29	26	29
1000(p: 0.01)	1000	4,984	305	304	297	302	283	299
1000(p: 0.02)	1000	10,217	199	196	191	183	178	195
1000(p: 0.03)	1000	15,190	150	148	148	144	141	145
1000(p: 0.04)	1000	20,057	127	120	117	118	113	119
1000(p: 0.05)	1000	25,226	104	100	101	100	97	99
1000(p: 0.06)	1000	29,603	93	92	89	91	87	89
1000(p: 0.07)	1000	35,714	81	78	76	77	74	75
1000(p: 0.08)	1000	40,088	76	74	72	70	66	74
1000(p: 0.10)	1000	49,997	61	59	58	57	57	62
1000(p: 0.15)	1000	74,465	43	41	42	43	41	40
1000(p: 0.18)	1000	91,098	40	36	34	35	32	36
1000(p: 0.20)	1000	99,912	33	32	33	31	31	32
2000 (p:0.20)	2000	400070	35	35	35	36	35	35
3000 (p:0.20)	3000	899799	38	38	38	37	35	37
4000 (p:0.20)	4000	1601099	41	39	39	38	38	39
5000 (p:0.20)	5000	2500051	40	39	38	41	38	41
6000 (p:0.20)	6000	3599454	43	42	40	42	38	42
7000 (p:0.20)	7000	4900863	43	40	41	43	42	40
8000 (p:0.20)	8000	6398801	44	43	44	44	40	43
9000 (p:0.20)	9000	8101085	44	43	43	43	41	43
10000(p:0.20)	10000	9991966	46	45	43	43	41	45

When the Zachary Karate Club graph is examined, it is observed that all methods identify the independent set as 20. All methods have produced optimal results for this graph. When 9000 randomly generated graphs (p: 0.20) are examined, it is seen that they have 9000 vertices and 8,101,085 edge connections. The value of p here indicates the approximate density of the generated graph. The Erdős–Rényi model was used for the randomly generated graphs. After calculating the centrality methods, the independent set member selection was made using the Gmin approach. According to these results, DMC formed the largest independent set with a value of 44. Among the other methods,

MC, DC, BC, and PC identified the independent set as 43, while EC identified it as 41.

When the table is examined in general, it is observed that DMC with Gmin usually produces better or equal results compared to other methods. Subsequently, MC, PC, DC, and BC methods generally produce better results. Most of the time, the worst results are produced with EC values. The DMC and MC methods generally yield the largest independent set sizes. This suggests that these two methods employ a more aggressive strategy in independent set selection.

In the study, it is observed that the size of the independent set changes depending on the number of edges. As the edge density increases, the size of the independent set decreases. For example:

1000(p:0.01) → 4,984 edges → 305-283 independent vertex

1000(p:0.10) → 49,997 edges → 61-57 independent vertex

1000(p:0.20) → 99,912 edges → 31-33 independent vertex

This trend indicates that the size of the independent set is inversely proportional to the edge density.

3.2. Matching

There are significant similarities in terms of application between the Matching tests and the Independent Set tests. The only difference in Matching tests is that they are performed by determining the independent set clusters on line graphs (Hemminger & Beineke, 1978). A line graph is created by representing the edges of the original graph as vertices. Another well-known name for the Matching problem is the independent edge set. Table 2 presents the matching results generated with the values of six different centrality methods.

Table 2. Comparative Matching Results

Line Graphs	V	E	Greedy Min - Matching					
			DMC	MC	DC	BC	EC	PC
Grid(8*8)	64	112	32	32	32	32	31	32
Grid(3*4*5)	60	133	30	30	30	24	27	30
King(7*8)	56	181	28	28	28	28	28	28
Hexagonal(4*4*4)	96	132	48	48	48	48	48	46
Knight(9*9)	81	272	40	40	40	40	40	40
Hypercube(7)	128	448	64	64	64	64	64	64
Banana Tree(6*6)	37	36	7	7	7	7	7	7
Folkman	20	40	10	10	10	10	9	10
Zachary	34	78	13	13	13	13	13	13
Dolphin	62	159	29	29	29	30	29	29
Zebra	27	111	13	13	13	13	13	13
Complex	70	133	35	34	34	34	34	34
50(p: 0.1)	50	121	25	25	25	25	25	25
50(p: 0.2)	50	251	25	25	24	25	24	25
50(p: 0.3)	50	370	25	25	25	25	25	25
100(p: 0.05)	100	267	50	50	50	50	47	50
100(p: 0.1)	100	485	50	50	50	50	48	50
100(p: 0.2)	100	1008	50	50	50	50	50	50
200(p: 0.01)	200	203	81	81	79	81	78	81
200(p: 0.02)	200	383	98	97	96	98	90	97
200(p: 0.03)	200	577	99	99	99	99	95	99
500(p: 0.01)	500	1290	247	247	247	248	236	248
500(p: 0.02)	500	2424	250	250	250	249	244	250
500(p: 0.03)	500	3754	250	250	250	250	247	250

When the table is examined, it is observed that the Grid (8*8) graph has 64 vertices and 112 edge connections. The DMC, MC, DC, BC, and PC algorithms have been applied to the line graph of Grid (8*8). In the selections made according to the Gmin approach, all algorithms identified 32 as the matching number. However, in the selection based on EC values, the matching number was determined to be 31. Another example, the random 200(p:0.01) graph, has 200 vertices and 203 edge connections. The matching results determined as 81 with DMC, MC, BC, and PC values, were found to be 79 with DC and 78 with EG. When the table is examined as a whole, it is evident that the matching members selected with DMC values produced better results than the other methods. While MC, PC, BC, and DC algorithms produced similar results, the worst results were generally obtained with EC values.

In irregular graphs (Folkman, Zachary, Dolphin, Zebra, Complex), small differences in matching sizes are observed. For example, in the Folkman graph, the PC method (9 matchings) produced somewhat lower results compared to the other methods. In the ranges of 200(p:0.01) - 200(p:0.03) and 500(p:0.01) - 500(p:0.03), it was observed that as the number of edges increased, the matching size also increased, but no significant differences were found between the different methods. This suggests that in large-scale random graphs, the methods produce similar results. In the 500(p:0.01), 500(p:0.02), and 500(p:0.03) graphs, the matching size remained almost constant (around 250 matchings), and all methods produced similar results. However, the PC method generally found a few units lower matching compared to the other methods (e.g., 236 matchings for 500(p:0.01)). This suggests that the PC method might show slightly lower performance compared to other methods in large-scale random graphs.

4. RESULTS

In this study, the solutions produced by different centrality methods for the independent set and matching problems using the greedy minimum approach have been compared and analyzed. For the analysis, the Differential Malatya (DMC), Malatya (MC), Degree (DC), PageRank (PC), Betweenness (BC), and Eigenvector (EC) centrality algorithms were selected. Additionally, to make the comparison more comprehensive, various types of graphs, including regular, social, random, multipartite, and others, were used for testing.

In regular graphs, the methods generally produce similar results for the independent set, while small differences are observed in irregular and random graphs. As edge density increases, the size of the independent set decreases, and the

differences between methods become more pronounced. DMISA and MISA generally produce the best results, while EC and BC methods tend to produce smaller independent set sizes. In large-scale random graphs, the differences between methods decrease, but DMISA and MISA appear to be more advantageous. The effect of edge density on the size of the independent set has been clearly observed, and this trend can be studied more thoroughly through statistical analysis.

Another important analysis presented in the study is the matching members produced by the algorithms. When the test operations conducted on different graph types are examined, it is observed that the difference between the results is smaller than the differences in the independent set results.

While no significant differences are observed between the methods in regular graphs, small-scale differences are found in irregular and random graphs. Specifically, the PC method produces a lower number of matchings in some cases, while the other methods generally show similar performance in large-scale graphs. This analysis can provide a foundation for future studies aimed at determining the most suitable method for a particular type of graph. In particular, it could be examined through more detailed statistical analysis whether some methods provide advantages or disadvantages in irregular and real-world graphs. Furthermore, in random graphs, how the performance differences between methods change as edge density increases could be explored in more depth in future research.

REFERENCES

Amadeo, M., Ruggeri, G., Campolo, C., & Molinaro, A. (2023). Content-driven closeness centrality based caching in softwarized edge networks. IEEE International Conference on Communications (ICC), 3264-3269. https://doi.org/10.1109/ICC45041.2023.10278928.

Ando, H., Bell, M., Kurauchi, F., Wong, K. I., & Cheung, K. F. (2020). Connectivity evaluation of large road network by capacity-weighted eigenvector centrality analysis. Transportmetrica A: Transport Science, 17(4), 648–674. https://doi.org/10.1080/23249935.2020.1804480.

Barthélemy, M. (2004). Betweenness centrality in large complex networks. European Physical Journal B, 38, 163–168. https://doi.org/10.1140/epjb/e2004-00111-4.

Bonacich, P. (2007). Some unique properties of eigenvector centrality. Social Networks, 29(4), 555-564. https://doi.org/10.1016/j.socnet.2007.04.002.

Bowater, D., & Stefanakis, E. (2023). Extending the Adapted PageRank Algorithm centrality model for urban street networks using non-local random walks. Applied Mathematics and Computation, 446, 127888. https://doi.org/10.1016/j.amc.2023.127888.

Carrabs, F., Cerulli, R., & Gentili, M. (2009). The labeled maximum matching problem. Computers & Operations Research, 36(6), 1859-1871. https://doi.org/10.1016/j.cor.2008.05.012.

Degree Centrality. (2025). Neo4j Graph Algorithms Documentation. https://neo4j.com/docs/graph-algorithms/current/labs-algorithms/degree-centrality/. (Erişim tarihi: 06.03.2025)

Ewing, E., Ren, J., Kansara, D., Sathiyanarayanan, V., & Ayanian, N. (2022). Betweenness centrality in multi-agent path finding. AAMAS '22: International Foundation for Autonomous Agents and Multiagent Systems, 400–408.

Hashemi, A., Dowlatshahi, M. B., & Nezamabadi-pour, H. (2020). MGFS: A multi-label graph-based feature selection algorithm via PageRank centrality. Expert Systems with Applications, 142, 113024. https://doi.org/10.1016/j.eswa.2019.113024.

Hemminger, R. L., & Beineke, L. W. (1978). Line graphs and line digraphs. In L. W. Beineke & R. J. Wilson (Eds.), Selected Topics in Graph Theory (pp. 271–305). Academic Press Inc.

Karci, A., Yakut, S., & Öztemiz, F. (2022). A new approach based on centrality value in solving the minimum vertex cover problem: Malatya Centrality Algorithm. Computer Science, 7(2), 81-88. https://doi.org/10.53070/bbd.1195501.

Lin, W., Li, C., Xu, L., & Xie, K. (2024). Enhanced network traffic anomaly detection: Integration of tensor eigenvector centrality with low-rank recovery models. IEEE Transactions on Services Computing, 17(6), 3597-3612. https://doi.org/10.1109/TSC.2024.3433580.

Lv, L., Bardou, D., Hu, P., Liu, Y., & Yu, G. (2022). Graph regularized nonnegative matrix factorization for link prediction in directed temporal networks using PageRank centrality. Chaos, Solitons & Fractals, 159, 112107. https://doi.org/10.1016/j.chaos.2022.112107.

Lv, L., Zhang, K., Zhang, T., Bardou, D., Zhang, J., & Cai, Y. (2019). PageRank centrality for temporal networks.

Physics Letters A, 383(12), 1215-1222. https://doi.org/10.1016/j.physleta.2019.01.041.(Metin içinde: Lv et al., 2019)

Nigam, R., Sharma, D. K., & Jain, S. (2022). A local betweenness centrality-based forwarding technique for social opportunistic IoT networks. Mobile Networks and Applications, 27, 547–562. https://doi.org/10.1007/s11036-021-01820-7.

Öztemiz, F., & Yakut, S. (2024). Analysis of the Malatya Centrality-Based Clique Method on DIMACS Benchmarks and Random Graphs. NATURENGS, 5(2), 63-69. https://doi.org/10.46572/naturengs.1591530.

Öztemiz, F. (2025). A greedy approach to solve maximum independent set problem: Differential Malatya independent set algorithm. Engineering Science and Technology, an International Journal, 63, 101995. https://doi.org/10.1016/j.jestch.2025.101995.

Parand, F.-A., Rahimi, H., & Gorzin, M. (2016). Combining fuzzy logic and eigenvector centrality measure in social network analysis. Physica A: Statistical Mechanics and its Applications, 459, 24-31. https://doi.org/10.1016/j.physa.2016.03.079.

Shen, Z. L., Jiao, Y. H., Wei, Y. K., Wen, C., & Carpentieri, B. (2025). Efficient hybrid PageRank centrality computation for multilayer networks. Chaos, Solitons & Fractals, 192, 116018. https://doi.org/10.1016/j.chaos.2025.116018.

Tortosa, L., Vicent, J. F., & Yeghikyan, G. (2021). An algorithm for ranking the nodes of multiplex networks with data based on the PageRank concept. Applied Mathematics and Computation, 392, 125676. https://doi.org/10.1016/j.amc.2020.125676.

Tudisco, F., & Higham, D. J. (2021). Node and edge nonlinear eigenvector centrality for hypergraphs. Communications Physics, 4, 201. https://doi.org/10.1038/s42005-021-00704-2.

Xu, Q., Sun, L., & Bu, C. (2023). The two-steps eigenvector centrality in complex networks. Chaos, Solitons & Fractals, 173, 113753. https://doi.org/10.1016/j.chaos.2023.113753.

Yakut, S., Öztemiz, F., & Karci, A. (2023). A new approach based on centrality value in solving the maximum independent set problem: Malatya Centrality Algorithm. Computer Science, 8(1), 16-23. https://doi.org/10.53070/bbd.1224520.

Yakut, S., & Bakan, C. T. (2023). Development of text summarization method based on graph theory and Malatya Centrality Algorithm. Computer Science, IDAP-2023: International Artificial Intelligence and Data Processing Symposium, 90-102. https://doi.org/10.53070/bbd.1350971

Yakut, S., & Öztemiz, F. (2024). An effective and robust approach based on Malatya Centrality Algorithm for interpreting cheminformatics graphs using maximum clique. Journal of Physical Chemistry and Functional Materials, 7(2), 192-199. https://doi.org/10.54565/jphcfum.1590385.

Yakut, S. (2024). An effective approach based on the Malatya Centrality Algorithm for determining the maximum independent set and minimum vertex cover in molecular graphs. 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP), 1-6. https://doi.org/10.1109/IDAP64064.2024.10710745

Zhang, P., Wang, T., & Yan, J. (2022). PageRank centrality and algorithms for weighted, directed networks. Physica A: Statistical Mechanics and its Applications, 586, 126438. https://doi.org/10.1016/j.physa.2021.126438.

Zheng, Z., Du, B., Zhao, C., & Xie, P. (2023). Path merging-based betweenness centrality algorithm in delay tolerant networks. IEEE Journal on Selected Areas in Communications, 41(10), 3133-3145. https://doi.org/10.1109/JSAC.2023.3310071.

MERKEZİYETSİZ OTONOM ORGANİZASYONLAR: BLOK ZİNCİR ÜZERİNDE ÇALIŞAN YENİ NESİL YÖNETİŞİM MODELİ

Bora ASLAN[1]

Füsun YAVUZER ASLAN[2]

1. GİRİŞ

DAO (Decentralized Autonomous Organization-Merkeziyetsiz Otonom Organizasyon), blok zinciri üzerinde akıllı kontratlarla çalışan ve merkezi otorite olmaksızın ortak karar almayı sağlayan organizasyon yapılarıdır (Hassan & De Filippi, 2021). Bir DAO'nun kuralları ve işlemleri blok zincire akıllı kontrat kodu olarak tanımlanır. Bu kontrat tüm üyelerce denetlenebilir ve dış müdahalelere karşı korumalıdır. DAO'lar genellikle bir yönetim token'ı çıkararak üyelik ve oy hakkı dağıtır. Token sahipleri, sahip oldukları pay oranında organizasyonun kararlara katılım sağlarlar. DAO'lar gerçek hayattaki geleneksel şirketler ve kooperatiflere benzetilirler. Fakat bu yapılardan farklı olarak, merkezi bir yönetim yapısına ihtiyaç duymadan çalışabilen, blok zinciri teknolojisi üzerine kurulu organizasyonlardır. Geleneksel şirketlerde kararlar genellikle üst yönetim veya belirli bir yönetim kurulu tarafından alınırken, DAO'larda yönetişim topluluğa dağılmış durumdadır. DAO'lar, şirketlerin hisse sahipliği modeline benzer şekilde

[1] Sorumlu Yazar Dr. Öğr. Üyesi, Kırklareli Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği, bora.aslan@klu.edu.tr, ORCID: 0000-0002-8069-8204.

[2] Dr. Öğr. Üyesi, Kırklareli Üniversitesi, Mühendislik Fakültesi, Yazılım Mühendisliği, fusunyavuzer@klu.edu.tr, ORCID: 0000-0001-7096-3425.

token tabanlı bir yönetişim sistemi kullanır, ancak burada kararlar merkezi bir yönetici yerine topluluk üyelerinin oylarıyla belirlenir (Han vd., 2025). Bu, katılımcılara doğrudan yönetim gücü vererek daha demokratik ve şeffaf bir yönetişim modeli sunar. Kooperatiflerle karşılaştırıldığında ise DAO'lar, benzer şekilde üyelerine karar alma süreçlerine doğrudan katılma hakkı tanır. Ancak kooperatiflerde veya şirketlerde yönetişim genellikle "bir üye, bir oy" ilkesiyle işlerken, DAO'larda çoğunlukla "bir token, bir oy" sistemi uygulanır. Bu durum, DAO'ların finansal güce dayalı bir karar alma mekanizmasına sahip olmasına yol açabilir. Öte yandan, kooperatifler belirli bir coğrafi bölgede faaliyet gösterme eğilimindeyken, DAO'lar küresel ölçekte çalışan, tamamen dijital topluluklar olarak faaliyet gösterebilir. DAO'lar ayrıca akıllı kontratlar sayesinde belirli görevleri otomatikleştirerek bürokratik süreçleri ortadan kaldırırken, kooperatiflerde kararlar genellikle toplantılar ve insan etkileşimi yoluyla alınır.

2. DAO ÇALIŞMA PRENSİPLERİ

Bu bölümde, DAO'ların temel çalışma prensipleri başlıklara ayrılmış ve detaylandırılarak anlatılmıştır.

2.1. Akıllı Kontratlar ve Teknik İşleyiş

DAO'ların kalbinde, blok zinciri üzerinde otonom olarak çalışan akıllı kontratlar bulunur(Chambefort & Chaudey, 2024). Bu kontratlar, organizasyonun tüzüğü gibi davranarak DAO'nun kurallarını uygular ve fonları tutar. Kararlar, akıllı kontratların öngördüğü koşullar yerine geldiğinde otomatik olarak yürütülür. Örneğin bir bütçe harcaması, üyelerin oy çokluğu ile onaylandığında akıllı kontrat tarafından otomatik olarak gerçekleştirilebilir. Tüm işlemler dağıtık defterde kayıtlı olduğu için şeffaflık sağlanır ve kayıtlar değiştirilemez.

"Otonom" terimi, bazı fonksiyonların insan müdahalesi olmadan kodla yürütülmesini ifade eder; ancak DAO'lar tamamen kendi kendine çalışmaz. İnsan katılımcılar yani DAO üyeleri oy kullanma veya öneri sunma gibi adımlarda aktif olarak yer alır. DAO yapıları ilk olarak Ethereum platformunda ortaya çıkmıştır fakat günümüzde farklı blok zincirlerinde de benzer yapılar oluşmuştur.

2.2. Token Tabanlı Yönetişim

Birçok DAO, yönetim mekanizması olarak token tabanlı oylama sistemini kullanır. DAO kurulduğunda bir yönetim token'ı basılır ve proje destekçileri, kullanıcılar veya yatırımcılara dağıtılır. Üyeler ellerindeki token sayısına orantılı oy hakkına sahip olur ve önerilen işlemler yada işlevler bu oylarla karara bağlanır (Fritsch vd., 2024). Örneğin bir protokol parametresinin değiştirilmesi veya hazine fonlarının kullanımı gibi konular, token sahiplerinin oyuna sunulur. Tipik olarak daha fazla tokene sahip olanın oylamadaki ağırlığı da fazladır; bu durum her ne kadar sermaye oranında söz hakkı verse de, "birikimin çokluğu = güç" prensibi bazı dezavantajlar da doğurabilir. Yine de bu model, geleneksel şirketlerdeki tek merkezli yönetime kıyasla daha katılımcı ve tabandan yukarıya bir yönetim biçimi sunar.

2.3. Öneri ve Karar Süreçleri

DAO'larda üyeler genellikle karar alınmasını istedikleri konularda öneriler hazırlar. Bu öneriler, DAO'nun web sitesi veya Discord gibi herkese açık bir platformunda tartışmaya açılır. Tartışma sonucunda olgunlaşan öneri, oylamaya sunulur. Oylama, DAO'nun tercihlerine göre zincir üzerinde veya zincir dışı gerçekleşebilir. Zincir üzerinde oylama, bir akıllı kontrat aracılığıyla tüm oyların blok zincirine işlemler halinde kaydedilmesiyle yapılır; böylece oylama sonucu ve oy dağılımı tamamen şeffaftır ve akıllı kontrat sonucu otomatik olarak

uygulayabilir. Ancak zincir üstü oylamada her oy bir blok zinciri işlemi olduğundan, Ethereum gibi ağlarda gas ücreti maliyeti doğar. Bu durum DAO üyelerinin bir ücret ödeyerek oy kullanması anlamına geldiğinden oylamaya katılımı azaltabilir. Bu sorunu aşmak için pek çok DAO, Snapshot gibi zincir dışı oylama araçları kullanır. Snapshot sisteminde token bakiyelerinin anlık görüntüsü alınarak zincir dışındaki bir sistemde üyelerin mesaj imzalayarak oy kullanması sağlanır. Böylelikle oy vermek için işlem ücreti ödenmez (Chainlink, 2022). Ardından oylama sonucu çoklu imza cüzdanı gibi bir mekanizma ile zincir üzerinde uygulanır. Bazı DAO'lar hibrit bir yaklaşım benimser: Kararlar önce toplulukla gayriresmî olarak tartışılıp Snapshot üzerinden oylanır, kararın alınmasının ardından DAO'da seçilmiş bir çoklu imza grubu tarafından oylama sonucunu zincire işler. Çoklu imza grupları, belirlenmiş belirli sayıdaki kişinin imzasını gerektiren akıllı kontrat cüzdanlardır. Bu durum tek bir kişinin DAO hazinesini yani birikimini kontrol etmesini engeller. Bu karar alma sürecine multi-sig denir (Chainlink, 2022).

2.4. Yönetişim Mekanizmaları ve Araçları

DAO'ların karar alma süreçlerinde kullandığı çeşitli araçlar vardır. En yaygın olanı yukarıda bahsedilen token ağırlıklı oy mekanizmasıdır. Bunun yanı sıra, daha adil bir katılım dağılımı için orantısız oy (quadratic voting) gibi yöntemler de tartışılmaktadır. Orantısız oy, her ek oy için katlanarak artan maliyet prensibi ile büyük token sahiplerinin etkisini sınırlandırmayı amaçlar (Dimitri, 2022). Bazı DAO'lar ise itibar puanı tabanlı yönetişim kullanır. Bu DAO'lar çekirdek katkıcılarına başka kişilere transfer edilemez itibar token'ları vererek, oylamada sadece finansal güce değil katkı veya katılım geçmişine de ağırlık tanırlar. Bunun dışında hemen her DAO, forumlar ve çeşitli iletişim kanalları üzerinden iletişimde kalır. Bu sosyal katman, topluluğun önerileri tartıştığı ve fikir birliği oluşturmaya çalıştığı kritik bir bileşendir. Yönetişim süreçlerinin

verimli işlemesi için, DAO'lar genellikle asgari katılım eşiği ve oy çoğunluğu barajı gibi kurallar koyar. Örneğin bir teklifin geçmesi için toplam tokenların en az %5'inin oylamaya katılması ve katılanların en az %50+1'inin "Evet" demesi akıllı kontrata şart olarak koyulabilir. Ayrıca, bazı DAO'lar zaman kilidi (timelock) mekanizmaları kullanarak onaylanan kararların yürürlüğe girmesini birkaç gün geciktirir; böylece beklenmedik veya zararlı bir karar çıkarsa, topluluk veya çekirdek geliştiriciler müdahale edebilsin (Monteiro & Correia, 2023). Örneğin Uniswap protokolünde bir yönetim kararı geçtiğinde, akıllı kontrat değişiklikleri uygulanmadan önce 2 gün bekleme süresi bulunmaktadır. Bu süre zarfında topluluk tepki verip karara itiraz edebilir yada karar iptal edilebilir. DAO bazı acil durumlarda zaman kilidi olmadan karar alabilir.

2.5. Güvenlik ve Ölçeklenebilirlik

DAO'ların çoğu zaman büyük hazine fonlarına sahiptirler bu sebeple teknik açıdan en çok odaklanılması gereken konu güvenliktir. Tüm kuralların kodda olması, koddaki bir hatanın felakete yol açabileceği anlamına gelir. 2016'da kurulan ilk büyük DAO projesi "The DAO", akıllı kontratındaki bir açıktan faydalanan saldırganlarca 3,6 milyon ETH'nin çalınmasıyla sarsılmıştır (Dhillon vd., 2017). Bu olay Ethereum tarihinde kritik bir dönemeç olmuş ve blok zincirin daha başında "hard fork" ile ayrılmasına dahi yol açmıştır. Bu gibi olaylar, akıllı kontrat denetlenmesinin ve güvenlik testlerinin önemini ortaya koymaktadır. DAO'lar, sadece kod güvenliği değil, yönetişim güvenliği konusunda da dikkatli olmak zorundadır. Güvenlik için birçok DAO, hazine anahtarlarını birden fazla kişiye yaymak adına multi-sig cüzdanları kullanır (Yu vd., 2023). Ancak burada da denge önemlidir; örneğin 2/3 imzalı bir multi-sig, hızlı karar alma açısından pratik olsa da güvenlik açısından risklidir. Bu sebeple daha büyük imza çoğunluğu oluşturulmalı yada ek güvenlik katmanları kurulmalıdır.

DAO'ların bir diğer teknik sınaması ölçeklenebilirlik konusudur. Popüler bir DAO binlerce hatta yüz binlerce üyeye sahip olabilir; herkesin oy kullandığı, sık sık karar alındığı durumlarda Ethereum gibi ana ağlar üzerinde bu oyların ve işlemlerin yürütülmesi pratik olmayabilir. Örneğin her bir oy vermenin 1-100$ gibi bir gas ücreti olduğunu düşünülürse, küçük paya sahip üyeler oy kullanmaktan çekinebilir (Rock'n'Block, 2024). Bu yüzden, az önce de değinildiği gibi birçok DAO zincir dışı oylama mekanizmalarına yönelmektedir. Ancak zincir dışı oylar tam güvence sağlamadığı için, ölçeklenme açısından uzun vadede Layer-2 çözümleri veya daha verimli zincirler değerlendirilmektedir (Schmitten vd., 2023). Öte yandan işlem hızı da önemli bir faktördür. Bir oylamanın tamamlanması tartışma süresi, oylama süresi, uygulama süresi ile genellikle birkaç gün hatta bir hafta sürebilir. Acil durumlarda bu yavaşlık sorun olabilir; bunun için bazı DAO'lar acil durum komiteleri veya hızlı karar yolları tanımlamıştır. Spam ve yönetim yükü de ölçeklenebilirlik kapsamında bir zorluktur. Eğer oylama zincir dışında maliyetsiz ise, kötü niyetli kişiler binlerce gereksiz öneri vererek oylama sistemini kilitlemeye çalışabilir. Ayrıca, DAO'larda üye ilgisizliği ciddi bir sorundur; birçok kişi oy vermeye vakit ayırmaz veya konuların teknik detayıyla ilgilenmez. Bu durumda birkaç aktif kişi tüm kararları alabilir. Bunu aşmak için temsilci atama ve önemli konularda kullanıcıları teşvik etme gibi yöntemler uygulanmaktadır.

3. FARKLI SEKTÖRLERDE DAO KULLANIMI

DAO'ların esnek yapısı, finansal protokollerden sanat dünyasına, oyun sektöründen tedarik zincirine ve sosyal yardım projelerine dek geniş bir yelpazede uygulanmaktadır. Aşağıda çeşitli sektörlerde DAO kullanım alanları ve etkileri incelenmiştir.

3.1. Merkeziyetsiz Finans Sektöründe DAO'lar

Merkeziyetsiz finans (DeFi), DAO modelinin en yoğun kullanıldığı alanlardan biridir. DeFi protokolleri, yönetimi topluluğa devretmek amacıyla DAO yapısına geçmiştir. Örneğin, Aave (https://aave.com/), MakerDAO (https://makerdao.com/), Uniswap (https://app.uniswap.org/) gibi platformlar, topluluk oylamalarıyla yönetilir ve kritik kararları kullanıcılarıyla birlikte alır. MakerDAO, DAI stabilcoin'ini 1 USD seviyesinde tutmayı başararak merkeziyetsiz finansın en güçlü örneklerinden biri olmuştur. DeFi DAO'ları, kullanıcılarını hem katılımcı hem karar verici konumuna getirerek çıkar birliği sağlar ve inovasyonu hızlandırır. Compound (https://compound.finance) gibi projeler, yönetim token'larını dağıtarak topluluk kontrollü yönetime geçişi teşvik etmiştir. Ancak büyük token sahiplerinin etkisi ve düzenleyici belirsizlikler gibi zorluklarla da karşı karşıyadırlar. Yine de Aave, MakerDAO, Uniswap, Compound gibi örnekler, merkezi aracılara ihtiyaç duymadan şeffaf ve topluluk odaklı bir finans ekosisteminin mümkün olduğunu kanıtlamaktadır.

3.2. Sanat ve NFT Dünyasında DAO'lar

Blok zinciri tabanlı sanat ve NFT ekosistemi, DAO'ların önemli kullanım alanlarından biridir. Koleksiyoner DAO'lar, üyelerinin kaynaklarını birleştirerek nadir NFT'leri satın almasını ve mülkiyetin tokenlar aracılığıyla paylaşılmasını sağlar. Flamingo DAO (https://flamingodao.xyz/) ve PleasrDAO (https://pleasr.org/), sanat eserlerini topluluk adına satın alıp yönetirken, PleasrDAO, "Doge NFT"yi alarak parçalı mülkiyet modeliyle binlerce kişiye sahiplik imkânı sunmuştur. DAO'lar yalnızca koleksiyon yönetimiyle sınırlı kalmayıp, sanatçıların doğrudan topluluk tarafından desteklenmesini de sağlar. Bazı NFT platformlarında sanatçılar kendi DAO'larını oluşturarak fon toplayabilir, satış gelirleri sanatçı DAO'larına aktarılabilir. Ayrıca DAO yapısı, sanat dünyasında küratörlük süreçlerine

uygulanarak topluluk üyelerinin hangi eserlerin sergileneceğine karar vermesine olanak tanır. DAO'lar sanat piyasasında şeffaflık, katılımcılık ve demokratik yönetim sağlayarak sanatçılar ve koleksiyoncular için yeni finansman modelleri sunmakta ve değerli sanat eserlerinin daha geniş kitleler tarafından sahiplenilmesini mümkün kılmaktadır.

3.3. Oyun Sektöründe ve "Play-to-Earn" Ekosisteminde DAO'lar

Blok zinciri tabanlı oyna-kazan (play-to-earn) oyunlar, DAO modelini benimseyerek oyuncuların oyun ekonomisinde söz sahibi olmasını sağlamaktadır. Axie Infinity (https://axieinfinity.com), merkeziyetsiz yönetime geçiş yaparak AXS token sahiplerine oyun hazinesi ve karar mekanizmaları üzerinde kontrol yetkisi tanımıştır. Bu sayede oyuncular, oyunun yönetiminde aktif paydaşlar haline gelmiştir. Oyun loncaları (gaming guilds) da DAO modelini kullanmaktadır. Yield Guild Games (YGG https://www.yieldguild.io), NFT oyun varlıklarına yatırım yaparak üyelerine bu varlıkları kiralar ve gelir paylaşımı modeli sunar. YGG, belirli oyunlara veya coğrafi bölgelere göre alt DAO'lara ayrılarak uzmanlaşmış topluluklar oluşturmuştur. Benzer şekilde Decentraland ve The Sandbox (https://www.sandboxdao.com), MANA ve SAND token sahiplerine sanal dünyaların yönetiminde söz hakkı vererek DAO modelini uygulamaktadır. DAO'lar, oyun ekosisteminde topluluk katılımını artırarak oyuncuların sadece tüketici değil, yönlendirici olmasını sağlamaktadır. DAO sayesinde oyuncular, oyun içi varlıklara sahip olabilir, karar süreçlerine katılabilir ve oyun ekonomisine uzun vadeli katkıda bulunabilir.

3.4. Sosyal Yardım ve Hayır Kurumlarında DAO'lar

DAO'lar, bağış ve sosyal yardım süreçlerini daha şeffaf, izlenebilir ve katılımcı hale getirerek geleneksel yardım kuruluşlarına alternatif sunmaktadır. Big Green DAO

(https://dao.biggreen.org), bağışçıları ve yardım alanları bir araya getirerek fon dağıtımını topluluk tarafından yönetilen bir sisteme dönüştürmüştür. Gitcoin DAO (https://gitcoin.co), açık kaynak projelerini desteklemek için karesel fonlama (quadratic funding) modelini kullanarak topluluk desteğini en çok alan projelere ek fon sağlarken, UkraineDAO, Ukrayna'ya destek için NFT açık artırmaları düzenleyerek milyonlarca dolarlık bağış toplamıştır. DAO'lar sayesinde bağışlar akıllı kontratlar ile otomatik yönetilebilir, suistimal riski azalır ve idari maliyetler düşer. Topluluk üyeleri, fonların nasıl harcanacağını oylayarak sürece aktif katılım sağlayabilir. Ancak düzenleyici belirsizlikler ve kripto erişim zorlukları, bu sistemlerin daha geniş çapta benimsenmesini sınırlayabilir. Şuana kadar olan örnekler, DAO'ların sosyal yardımlarda güçlü bir alternatif olabileceğini ve gelecekte geleneksel STK'larla entegre edilebileceğini göstermektedir.

4. DAO'LARIN AVANTAJLARI

Bu bölümde DAO'ların avantajları üzerinden fırsatlar tartışılmıştır.

4.1. Şeffaflık

DAO'ların en belirgin avantajı, işlemlerinin ve kurallarının açık olarak blok zincirine yazılması sayesinde tam şeffaflığın sağlamasıdır (Diallo vd., 2018). Tüm oy kayıtları, bütçe harcamaları, akıllı kontrat kodları isteyen herkesçe görülebilir. Geleneksel kapalı kurum yönetimlere kıyasla DAO'larda gizli kararlar ile ilgili risk çok düşüktür. Kararlar topluluk tartışmalarıyla olgunlaşır ve oylama sonucunda otomatik uygulanır. Örneğin bir DAO'nun hazinesinden 10 ETH bağış yapılmışsa, bunun hangi oylama ile gönderildiği, hangi adrese gittiği ve ne zaman gittiği zincir üzerinde görülebilir. Ayrıca akıllı kontrat kodları eğer şifrelenmediyse açık kaynak olduğu için,

sistemin nasıl çalıştığı tüm paydaşlarca incelenip denetlenebilir. Bu şeffaflık, üyeler arası güveni artırır ve dış denetimi kolaylaştırır. Hatta birçok DAO, aylık raporlarını veya önemli karar özetlerini halka açık biçimde yayınlayarak hesap verebilirlik sergiler. İletişim genellikle discord, DAO forumları gibi üyelere açık sistemlerde yapıldığı için şeffaflık büyük ölçüde sağlanır.

4.2. Merkeziyetsizlik ve Katılımcılık

DAO'lar gücü tek bir elde toplamak yerine üyeler arasında dağıtmaya odaklanır. Kararlar kolektif akılla alındığı için, tek bir liderin yanlı kararları veya hataları sistemin kaderini belirlemez. Bu da sistemi dirençli kılar; bir veya birkaç kişinin ayrılması, DAO'dan ayrılması, hesabın saldırganlar tarafından ele geçirilmesi durumunda bile DAO varlığını sürdürür. Özellikle küresel katılımcı tabanına sahip DAO'lar, coğrafi ve politik risklere karşı da dayanıklıdır. Merkezi bir otorite olmadığından kapatılması, yasaklanması daha zordur. Üyeler sahip oldukları token'lar oranında oy hakkı elde ederek yönetime doğrudan katılır, bu da geleneksel şirketlerdeki hissedarlık yapısına benzemekle birlikte genellikle daha tabana yayılan bir yapıdır. Örneğin bir DAO'da %1 payı olan yüzlerce küçük yatırımcı, toplamda %10 paya sahip tek bir üyeyi oy çokluğuyla yenebilir . Bu katılımcı yönetim, demokratik bir yönetişim modeli sunar . Bazı DAO'lar bir üyenin hak sahibi olacağı oranı bile akıllı kontrat ile kısıtlarlar. Bu da kısmen eşitliği getirir. Her üye forumlarda görüş bildirerek veya öneri sunarak sesini duyurabilir. Böylece DAO'lar, üyelerine gerçek anlamda söz hakkı ve mülkiyet hissi verir. Birçok üye, DAO yönetişimine katkı sağladıkça proje ile duygusal bir bağ kurar ve başarıya ulaşması için gönüllü çaba gösterir. Bu, klasik şirketlerdeki çalışan-müşteri-hissedar ayrımını bulanıklaştırarak daha bütünleşik bir topluluk yapısı yaratır.

4.3. Verimlilik ve Otonomi

Doğru tasarlanmış bir DAO, bürokrasiyi azaltarak verimliliği artırabilir (Wulf A. Kaal, 2021). Akıllı kontrat otomasyonu sayesinde birçok karar önceden tanımlanmış koşullara bağlanabilir. Örneğin bir fon toplama hedefi akıllı kontratta belirtilir; hedefe ulaşıldığında otomatik olarak anlaşmalı hesaba transfer olur. Arada insanların elle onay vermesi gerekmez, bu da hem zaman kazandırır hem hataları azaltır. Öte yandan özellikle uluslararası aktarımda hem zaman hem de maddi tasarruf yapılır. Benzer şekilde, akıllı kontratlar aracıları ortadan kaldırabilir. DAO yapısı sayesinde finansal işlemler için bankalara, karar onayı için yönetim kurullarına ihtiyaç duymadan, kod içinde tanımlanan yetki ve onay mekanizmalarıyla süreçler işler. Bu durum özellikle küresel ölçekte operasyon yürüten yapılarda ciddi maliyet avantajı ve hız sağlar. Örneğin uluslararası bir para transferi, DAO hazinesinden dakikalar içinde gerçekleştirilebilirken, geleneksel sistemde günler alabilir. DAO'lar ayrıca 7/24 çalışır; insan tatilinde veya mesai saatleri dışında diye işlem aksamaz, kontratlar sürekli tetiklenebilir durumdadır.

4.4. Yenilikçilik ve Esneklik

DAO ekosistemi henüz yeni olduğundan, sürekli yeni yönetişim modelleri ve araçlar geliştirilmekte ve uygulanmaktadır. Bu da organizasyon yapılarında inovasyonu teşvik eder. Örneğin "soulbound token"lar (transfer edilemeyen itibar puanı token'ları) veya "likit demokrasi" (oy haklarını istenildiğinde devredip geri alabilme modeli) gibi kavramlar DAO'lar sayesinde gerçek dünyada test edilmeye başlanmıştır (TokenMinds, 2024). Geleneksel kurumlarda görülmeyen bu denemeler gelecekte şirketlerin ve/veya kooperatiflerin modellerine ışık tutabilir. Ayrıca DAO'lar yazılım odaklı olduğu için ölçeklenmesi de esnektir. Üye sayısı birden bine çıkabilir,

kod uygunsa bu büyümeyi kaldırabilir. Öte yandan geleneksel şirket veya kooperatif yapılarında ise bu kadar hızlı büyüme muhtemel bir kaosla sonuçlanabilir.

4.5. Küresel Erişim ve Kapsayıcılık

DAO'lar internet erişimi ve kripto cüzdanı olan herkese açıktır. Dünyanın herhangi bir yerinden bir yazılımcı, bir sanatçı veya bir yatırımcı ilgi duyduğu bir DAO'ya katılıp katkı sunabilir. Coğrafi engelleri ortadan kaldırdığı için DAO'ların üye havuzu küreseldir. Bu da hem daha zengin bir fikir çeşitliliği sağlar, hem de fırsat eşitliğini artırır. Örneğin Türkiye'deki bir yazılım geliştiricisi, Amerika merkezli bir DAO projesine katkı yapıp token kazanabilir. Bu durum geleneksel şirket hiyerarşilerinde ve ülkeler arası çalışma durumlarında pek mümkün olmayan bir durumdur. DAO'lar özellikle bankacılık sistemi dışında kalan kişilere finansal katılım imkanı da sunabilir. Örneğin Axie Infinity'nin oyuncu kitlesinin %25'inin banka hesabının olmadığı fakat kripto cüzdanıyla gelir elde edebildiği bilinmektedir.

5. DAO'LARDAKİ ZORLUKLAR

Bu bölümde DAO'ların karşılaştığı zorluklar ve alınabilecek önlemler tartışılmıştır.

5.1. Kanuni ve Hukuki Statü

DAO'lar yapıları gereği mevcut hukuk sistemlerine uymakta zorlanır çünkü ne tüzel kişilikleri ne de geleneksel bir idari sorumluluk yapıları vardır. Bir DAO'nun yasal olarak şirket mi, ortaklık mı yoksa bir yatırım fonu mu olduğu pek çok ülkede belirsizdir. Bu durum büyük bir zorluk oluşturmaktadır. Örneğin ABD'de 2022'de Emtia Vadeli İşlemler Ticaret Komisyonu (CFTC), bir DeFi platformu işlettiği iddiasıyla Ooki DAO adlı bir organizasyona dava açmış ve ilk kez bir DAO'yu doğrudan

sorumlu tutmaya çalışmıştır (Marta Piekarska, 2022). CFTC, DAO üyelerini geleneksel bir ortaklığın üyeleri gibi değerlendirerek, kurallara aykırı işlemlerden kişisel olarak sorumlu olabileceklerini öne sürmüştür. Bu durum DAO topluluğunda şok etkisi yaratmış, "anonim olarak internette toplanan token sahipleri bir genel ortaklık mıdır?" sorusunu gündeme getirmiştir. Belirsizlik nedeniyle birçok DAO, riskleri azaltmak için kendini yasal bir yapıyla dönüştürme arayışına başlamıştır. Buna karşın ABD'nin Wyoming eyaleti, 2021'de çıkardığı bir yasayla DAO'ların limitet şirketi olarak tescil edilebilmesini mümkün kılmıştır (Lom & Browndorf, 2021). Bu sayede bir DAO, Wyoming'de kayıtlı bir şirket haline gelip üyelerine sınırlı sorumluluk koruması sağlayabilir. Benzer şekilde İsviçre, Singapur, Malta gibi bazı ülkeler kripto vakıf/dernek yapıları altında DAO'lara yasal kimlik kazandırmaya çalışmaktadır. Buna rağmen birçok ülkede DAO'lar halen gri alandadır. Bu da yatırımcıların ve büyük katılımcıların çekincelerine yol açar, zira belirsiz yasal statü hem vergisel sorunlar hem de olası davalar açısından risk demektir. Bu sebeple bir çok tüzel şirket ve kurum DAO'lardan resmi olarak uzak durmaktadır. Ayrıca bazı durumlarda DAO'ların aldığı kararlar yerel yasalara aykırı olabilir. Bu durumda sorumlunun hukuki olarak kim olacağı muğlaktır. Tüm bu nedenlerle, kanuni belirsizlik DAO'ların büyümesinin önündeki önemli engellerden biridir. Ancak son yıllardaki gelişmeler olumlu yöndedir. Wyoming'in ardından yakın zamanda Amerika'daki bazı eyaletler de DAO'ların yasal durumları ile ilgili yasalar oluşturmaya başlamıştır (Finger & Cramer, 2023). Ayrıca. Yakın gelecekte bu yasal hazırlıkların dünya geneline yaygınlaşması olasıdır.

5.2. Güvenlik Açıkları

Siber güvenlik DAO'lar için devamlı bir endişe kaynağıdır. Yukarıda bahsedilen The DAO saldırısı bunun en

somut örneğidir. Akıllı kontrat hataları, bilinmeyen bug'lar veya yeni siber saldırılar bir DAO'nun hazinesinin boşaltılmasına veya kontrolünün saldırganlar tarafından ele geçirilmesine yol açabilir. 2016'daki olaydan çıkarılan derslerle, bugün büyük DAO'lar kodlarını yazılım denetim firmalarına inceletmekte, yapay zeka destekli sistemler ile hataları yakalamaya çalışmaktadır. Ancak yazılım geliştirme dünyasında "hatasız kod" veya %100 güvenlik garantisi yoktur. En iyi ekiplerin yazdığı akıllı kontratlarda bile öngörülmeyen açıklar bulunabilir. Son yıllarda DeFi protokollerinde milyarlarca dolarlık hack vakaları yaşanmıştır. DAO'lar da akıllı kontratlara dayandığı için benzer riskleri taşır. Örneğin 2022'de bir NFT DAO'su olan Beanstalk protokolü, akıllı kontrat açığından yararlanan bir manipülasyonla $180 milyon kaybetmiştir. Saldırganlar, anlık olarak oy gücünü ele geçirip hazinedeki stablecoin'leri kendi hesabına onaylatmıştır. Eğer DAO kontratı tasarımında önlem alınmadıysa yada yönetişim mantığı düzgün kurulmadıysa bu tür saldırılar gerçekleşebilir. Ayrıca "geri dönüş kapısı" olarak bilinen ve tamamen değiştirilemez kontratlar yerine, acil durumlar için arka kapılar veya upgrade anahtarları bırakan kontrat tasarımları da çok tehlikelidir. Eğer bu anahtarlar iyi korunmaz veya fark edilmezse, kötü niyetli biri veya grup bunları kullanarak sistemi suistimal edebilir. Kısacası güvenlik, DAO'lar için en önde gelmesi gereken bir konudur. Bu alanda, çok imzalı doğrulama, zaman kilidi, izleme ajanlar gibi çeşitli tedbirler alınmaktadır. Ayrıca kritik protokollerde para ödülleri ile hatalar bulunması teşvik edilmektedir.

5.3. Merkeziyet Riski ve Güç Yoğunlaşması

Her ne kadar DAO'ların amacı merkeziyetsiz bir yönetim modeli olsa da pratikte birçok DAO'da oy gücü küçük bir azınlıkta toplanmıştır. Token dağılımı adil değilse veya zamanla belirli ellerde yığılırsa, DAO yönetimi ve karar alma süreçleri fiilen merkezi bir grubun kontrolüne girebilmektedir. Bu durum

DAO'ların önemli bir açmazını ortaya koymaktadır. Özellikle ilk dağıtımda DAO kurucu ekibi ve ilk yatırımcılar yüksek oranda token aldıysa, topluluğa yayılma uzun zaman alabilir ya da hiç gerçekleşmeyebilir. Örneğin bir protokolde 10 cüzdanın oyların çoğunluğunu kontrol ettiği biliniyorsa, diğer küçük token sahipleri kararlarda çok az söz sahibi olabileceği kaygısı ile DAO'ya katılım göstermeyebilir. Bu riskleri azaltmak için, DAO'larda zamanla token dağıtımını tabana yayma, oylamalarda çok düşük bir katılım varsa kararların geçmemesi için çekimserlik sınırı uygulaması, bir kişinin alabileceği maksimum token sınırı veya oy gücü için belirli bir süredir DAO'ya üye olma koşulu işlemi için kilitlenme süreli oy gibi önlemler geliştirilmektedir. Yine de güç yoğunlaşması, merkeziyetsizlik idealine gölge düşürmektedir.

5.4. Katılım Problemleri ve Organizasyonel Zorluklar

DAO'lar insanların yönetime katılma modeline göre şekillenmiştir. Fakat birçok DAO'da üyelerin küçük bir kısmı aktif olarak yönetime katılırken, büyük çoğunluk sessiz izleyici konumundadır (Rikken vd., 2019). Üye sayısı arttıkça bu oran genelde düşer. Binlerce kişinin bulunduğu DAO'larda birkaç düzine kişi tüm tartışmaları yapıp kararları şekillendirir. Aslında bu durum mevcut şirket veya kooperatif yapılarında da bu şekildedir. Fakat bu durum hem adil temsil açısından sorunlu hem de aktif üyeler için iş yükü açısından yorucudur. Sürekli her konuda oy kullanmak üyelerde yönetişim açısından yorgunluk yaratabilir. Bu nedenle ister istemez proje ilerledikçe genellikle bir tür hiyerarşi geri gelmeye başlar. Çalışma grupları, komiteler, komisyonlar veya kurullar kurulmaya başlanır. Bu aslında DAO'ların başlangıç vizyonundaki tam yatay yapının pratikte sürdürülemediğini gösterir. Örneğin bir protokolde teknik geliştirmeler için bir çekirdek ekip yetkilendirilebilir veya pazarlama harcamaları için topluluğun seçtiği bir kurul oluşturulabilir. Bu alt gruplar kararları hazırlayıp genel oya sunar.

Bu, verimliliği artırsa da yine merkezi bir unsur getirir ve grupların DAO'yu yönlendirmesini sağlar.

5.5. Teknik Ölçeklenebilirlik ve Kullanım Kolaylığı

Teknik açıdan, Ethereum, Bitcoin gibi ana blok zincirlerinin yüksek gas ücretleri, sınırlı işlem hızı gibi kısıtları DAO'ları maddi ve organizasyonel olarak olumsuz etkilemektedir. Bu sorunlar kısmen Base, Polygon gibi Layer-2 ağlara geçiş veya zincir dışı işlemler ile aşılmaya çalışılsa da halen ideal bir çözüm bulunamamıştır. DAO'ların oluşturdukları sistemlerdeki cüzdan kullanımının karmaşıklığı veya doğrudan kripto cüzcan ile varlıklarık yönetimindeki zorluklar düşünüldüğünde ortalama bir kullanıcının DAO'ya katılımı zordur. Kripto cüzdanı kurma, token alma, yönetim arayüzüne bağlanma gibi adımlar teknik bilgi ve tecrübe gerektirir. Ayrıca kriptopara dünyasına karşı olan güvenlik kaygıları gibi olumsuzluklar da buna eklendiğinde birçok potansiyel katılımcı dışarıda kalmaktadır. Bu nedenle, gelecekte daha kullanıcı dostu DAO platformlarına ihtiyaç vardır. Örneğin tek tıkla mobil uygulama üzerinden oy kullanma veya kripto cüzdan destekli sosyal medya hesabıyla DAO'ya katılma gibi deneyimler geliştirilmelidir.

5.6. Güven ve Sorumluluk Sorunları

Tam merkeziyetsiz yapılarda bir şeyler ters gittiğinde sorumluluk almak ve kriz yönetimi zor olabilir. Örneğin bir saldırı sonucu çalınan veriler/varlıklar olduğunda veya yanlış bir karar alındığında, "sorumlu kim" sorusu belirsizdir. Bu durumda bazen topluluk içinde suçlamalar, bölünmeler yaşanabilir. Ayrıca yeni katılanlar için, anonim insanların oylarıyla yönetilen bir sisteme güvenmek başlangıçta çok zor olabilir. Geleneksel yapılarda insanlar yöneticiye veya markaya güvenirken, DAO'da akıllı kontrat koduna ve topluluk iradesine güvenmek gerekir. Bu güveni inşa etmek zaman alır ve genellikle iyi sonuçların kendini

kanıtlamasıyla mümkün olur. Fakat özellikle kriptoparaların 15-25 yaş aralığındaki gençlerin ilgi alanında olduğu, blokzincir projelerinin ve DAO girişimlerinin bu kitle tarafından domine edildiği düşünüldüğünde gerekli sorumluluk ve güvenin sağlanması zor olacaktır. Bu açıdan henüz yeni doğan konumundaki DAO'ların biraz daha zamana ihtiyaç duyduğu bir gerçektir. Öte yandan yakın gelecekte başarılı gerçek dünya örneklerinin artması, ülkelerin ilgili yasaları yürürlüğe koyması ve kurumsal şirketlerin DAO'lara yatırım yapması ile bu kaygılar olumlu bir şekilde değişecektir.

6. SONUÇLAR

DAO'lar, geleneksel organizasyon yapılarına kıyasla merkeziyetsiz, şeffaf ve otomatikleştirilmiş bir yönetişim modeli sunarak organizasyonel süreçlerde devrim niteliğinde değişiklikler getirmiştir. Blok zinciri ve akıllı kontratlar sayesinde karar alma süreçleri daha demokratik hale gelmiş, yönetişimde aracılar ortadan kaldırılmış ve oylamalar doğrudan zincir üzerinde yürütülebilir hale gelmiştir. Finans, sanat, oyun, tedarik zinciri ve sosyal yardım gibi birçok sektörde kullanılan DAO'lar, katılımcıların karar alma süreçlerine doğrudan dahil olmasını sağlayarak, merkezi yönetim modellerinin sınırlarını zorlamaktadır.

DAO'ların sunduğu avantajlar arasında şeffaflık, merkeziyetsizlik, otomasyon, küresel erişim ve kolektif yönetişim öne çıkmaktadır. Ancak, DAO'ların halen karşı karşıya olduğu önemli zorluklar bulunmaktadır. Yasal belirsizlikler, güvenlik riskleri, oy gücünün az sayıda kişinin elinde yoğunlaşması, düşük topluluk katılımı ve teknik ölçeklenebilirlik sorunları, DAO ekosisteminin tam anlamıyla olgunlaşmasının önündeki başlıca engellerdir. Özellikle akıllı kontratların

güvenliği ve yönetişim modellerinin sürdürülebilirliği, DAO'ların uzun vadeli başarısı için kritik öneme sahiptir.

DAO'ların geleceği, daha etkin yönetişim mekanizmalarının geliştirilmesi, zincir dışı oylama çözümlerinin yaygınlaşması ve yasal çerçevenin netleşmesi gibi faktörlere bağlı olacaktır. DAO yapıları, merkezi olmayan finans protokollerinden oyun ekosistemlerine, sanat dünyasından sosyal yardımlara kadar birçok alanda kendini kanıtlamaya devam etmektedir. Önümüzdeki yıllarda DAO'ların kurumsal dünyada daha fazla benimsenmesi, büyük teknoloji şirketleri ve devlet kurumları tarafından araştırılması ve hibrit yönetişim modelleriyle daha geniş çapta uygulanması beklenmektedir.

Sonuç olarak, DAO'lar organizasyonların çalışma biçiminde yeni bir paradigma sunarken, beraberinde getirdiği teknik ve yönetişimsel zorlukların aşılmasıyla daha geniş ölçekli adaptasyon görebilir. Gelecekte DAO'lar, daha esnek ve güvenli yapılarla bireylerin ve kuruluşların finansal ve yönetişim süreçlerine daha fazla dahil olmasını sağlayarak küresel ölçekte daha adil ve demokratik ekonomik sistemlerin temel taşlarından biri olabilir.

KAYNAKÇA

Chainlink. (2022). *DAOs and the Complexities of Web3 Governance*. https://blog.chain.link/daos/#:~:text=%2A%20Off,websites%20like%20Discourse%20or%20a

Chambefort, C., & Chaudey, M. (2024). Blockchain, tokens, smart contracts, and "decentralized autonomous organization": Expanding and renewing the mechanisms of governance? *European Management Review*, *21*(3), 511-515. https://doi.org/10.1111/emre.12677

Dhillon, V., Metcalf, D., & Hooper, M. (2017). The DAO Hacked. Içinde *Blockchain Enabled Applications* (ss. 67-78). Apress. https://doi.org/10.1007/978-1-4842-3081-7_6

Diallo, N., Shi, W., Xu, L., Gao, Z., Chen, L., Lu, Y., Shah, N., Carranco, L., Le, T.-C., Surez, A. B., & Turner, G. (2018). eGov-DAO: a Better Government using Blockchain based Decentralized Autonomous Organization. *2018 International Conference on eDemocracy & eGovernment (ICEDEG)*, 166-171. https://doi.org/10.1109/ICEDEG.2018.8372356

Dimitri, N. (2022). Quadratic Voting in Blockchain Governance. *Information*, *13*(6), 305. https://doi.org/10.3390/info13060305

Finger, J., & Cramer, P. (2023). *With New DAO Law on the Books, Utah Joins Race with Wyoming and Tennessee to Become U.S. "Crypto Capital"* . https://www.blockchainandthelaw.com/2023/05/part-ii-with-new-dao-law-on-the-books-utah-joins-race-with-wyoming-and-tennessee-to-become-u-s-crypto-capital/

Fritsch, R., Müller, M., & Wattenhofer, R. (2024). Analyzing voting power in decentralized governance: Who controls DAOs? *Blockchain: Research and Applications*, *5*(3), 100208. https://doi.org/10.1016/j.bcra.2024.100208

Han, J., Lee, J., & Li, T. (2025). A review of DAO governance: Recent literature and emerging trends. *Journal of Corporate Finance*, *91*, 102734. https://doi.org/10.1016/j.jcorpfin.2025.102734

Hassan, S., & De Filippi, P. (2021). Decentralized Autonomous Organization. *Internet Policy Review*, *10*(2). https://doi.org/10.14763/2021.2.1556

Lom, A., & Browndorf, R. (2021). *Wyoming to Recognize DAOs as LLCs*. Global Regulation Tomorrow. https://www.regulationtomorrow.com/us/wyoming-to-recognize-daos-as-llcs/

Marta Piekarska. (2022). *The State of DAO Security*. https://consensys.io/blog/the-state-of-dao-security

Monteiro, F., & Correia, M. (2023). Decentralised Autonomous Organisations for Public Procurement. *Proceedings of the 27th International Conference on Evaluation and Assessment in Software Engineering*, 378-385. https://doi.org/10.1145/3593434.3593519

Rikken, O., Janssen, M., & Kwee, Z. (2019). Governance challenges of blockchain and decentralized autonomous organizations. *Information Polity*, *24*(4), 397-417. https://doi.org/10.3233/IP-190154

Rock'n'Block. (2024). *Key Challenges in Launching a DAO & How to Overcome Them*. medium. https://rocknblock.medium.com/key-challenges-in-launching-a-dao-how-to-overcome-them-b130aa36e8ca

Schmitten, J.-P., Augart, G., & Hüsig, S. (2023). Decentralized Blockchain Governance and Transaction Costs in Digital Transformation: The Case of the DAO Revisited. *2023 Portland International Conference on Management of Engineering and Technology (PICMET)*, 1-14. https://doi.org/10.23919/PICMET59654.2023.10216831

TokenMinds. (2024). *What is Liquid Democracy? A New Model for Agile Decision-Making in DAOs*. https://tokenminds.co/blog/knowledge-base/what-is-liquid-democracy

Wulf A. Kaal. (2021). Blockchain-based corporate governance. *Stanford Journal of Blockchain Law & Policy*.

Yu, G., Wang, Q., Bi, T., Chen, S., & Xu, X. (2023). Leveraging Architectural Approaches in Web3 Applications - A DAO Perspective Focused. *2023 IEEE International Conference on Blockchain and Cryptocurrency (ICBC)*, 1-6. https://doi.org/10.1109/ICBC56567.2023.10174988

NESNELERİN İNTERNETİ AĞLARINDA SALDIRI TESPİTİ İÇİN FEDERE ÖĞRENME

Nesibe YALÇIN[1]

Semih ÇAKIR[2]

1. GİRİŞ

Nesnelerin İnterneti (Internet of Things, IoT) güvenliğinde temel amaç, IoT ekosistemi içindeki kullanıcıların, cihazların ve altyapının güvenliğini sağlamak, sunulan hizmetlerin erişilebilirliğini devam ettirmek ve verilerin gizliliğini korumaktır (Sasi vd., 2024). Saldırı tespit sistemleri, günümüzde giderek daha karmaşık hale gelen siber saldırılar ile başa çıkmak için kritik güvenlik bileşenleridir. Özellikle makine öğrenmesi teknikleri ile desteklenerek ağ ve sistem güvenliği daha etkili hale gelmektedir. IoT bağlamında, standart makine öğrenmesi tabanlı saldırı tespit sistemleri, farklı IoT cihazlarından toplanan eğitim verilerinin (özellikler ve örneklerin) merkezi bir sunucuya veya bulut ortamına yüklendiği, analiz edildiği ve model eğitiminin gerçekleştiği bir yaklaşım kullanır (Campos vd., 2022). Federe öğrenme (federated learning) yaklaşımında ise eğitim verileri merkezden uzak tutularak (gerçek verileri açığa çıkarmadan) model eğitimi merkezi olmayan bir şekilde gerçekleşir. Böylece veri mahremiyeti ve gizliliğin korunmasının yanı sıra iletişim daha

[1] Dr. Öğr. Üyesi, Erciyes Üniversitesi Mühendislik Fakültesi Bilgisayar Mühendisliği Bölümü, nesibeyalcin@erciyes.edu.tr, ORCID: 0000-0003-0324-9111.

[2] Dr. Öğr. Üyesi, Zonguldak Bülent Ecevit Üniversitesi Karadeniz Ereğli Meslek Yüksekokulu Bilgisayar Teknolojileri Bölümü, semih.cakir@beun.edu.tr, ORCID: 0000-0003-3072-9532.

verimli hale gelir, hesaplama yükü azaltılır ve daha düşük gecikme sağlanır.

Merkezi yaklaşım ile ilgili güvenlik endişelerini azaltmak amacıyla federe öğrenme yaklaşımı, son yıllarda sağlık (Schneble ve Thamilarasu, 2019; Huang and Liu, 2019; Li vd., 2019; Yuan vd., 2020, Xu vd., 2020), finans (Zhao vd., 2019), doğal dil işleme (McMahan vd., 2017; Hard vd., 2018), otonom araçlar (Samarakoon vd., 2018; Al Mallah vd., 2021), siber güvenlik (Nguyen vd., 2020; Khramtsova vd., 2020; Friha vd., 2022; Doriguzzi-Corin ve Siracusa, 2024) gibi birçok alanda ilgi görmüştür (Şekil 1). Yuan vd. (2020), sağlık IoT cihazları için bir federe öğrenme çerçevesi önermişlerdir. Önerdikleri çerçeve ile IoT cihazları üzerindeki hesaplama yükünü ve IoT cihazları ile merkezi sunucu arasındaki iletişim yükünü önemli ölçüde azaltmışlardır. Ayrıca çalışmalarında, gerçek dünya aritmi tespit görevlerinde çok küçük bir doğruluk kaybı gözlemlendiği belirtilmiştir. Li vd. (2019) tarafından beyin tümörü segmentasyonu için bir federe öğrenme sistemi önerilmiş ve federe model paylaşımının çeşitli pratik yönleri incelenmiştir. Hasta veri gizliliğinin korunmasına vurgu yapılarak güçlü bir diferansiyel mahremiyet garantisi sağlandığı belirtilmiştir. Mao vd. (2022) tarafından farklı çiftlikler arasında dağıtılmış verilere dayalı federe öğrenme modeli ile otomatik hayvan aktivitesi tanıma işlemi gerçekleştirilmiştir. Akıllı çiftlik uygulaması kapsamında yapılan başka bir çalışmada (Idoje vd., 2023), federe öğrenme mahsul türü sınıflandırmasında kullanılmış ve merkezi olmayan modellerin hız ve doğruluk açısından daha iyi performans gösterdiği kanıtlanmıştır. Mahsul verimi tahmini için yapılan çalışmada (Manoj vd., 2022) ise farklı istemcilerde dağıtılmış veri kümeleri üzerinde model eğitimi için federe öğrenmeden yararlanılmıştır. Federe öğrenmenin enerji sistemleri alanındaki yük tahmininde (Fekri vd., 2022; Gao vd., 2021), enerji yönetiminde (Lee ve Choi, 2020) ve tüketicilerin sosyo-

demografik özelliklerinin belirlenmesinde (Wang vd., 2021) başarılı uygulamaları da mevcuttur.

Federe öğrenme tabanlı çözümler, özellikle kullanıcı gizliliğinin kritik olduğu uygulamalarda (örneğin, IoT cihazlarının davranışsal verileri kullanılıyorsa) büyük bir avantaj sağlamaktadır. Zhao vd. (2019) çalışmalarında, IoT cihaz üreticilerinin müşterilerine daha iyi ürün ve hizmet sunabilmelerine yardımcı olacak federe öğrenme tabanlı bir sistem tasarlamışlardır. Tasarlanan sistemde ayrıca federe öğrenme sırasında tüm müşterilerin güncellemeleri blok zincir ve diferansiyel mahremiyet (differential privacy) ile denetlenerek/doğrulanarak kötü niyetli müşteriler veya üreticiler tespit edilebilmektedir.

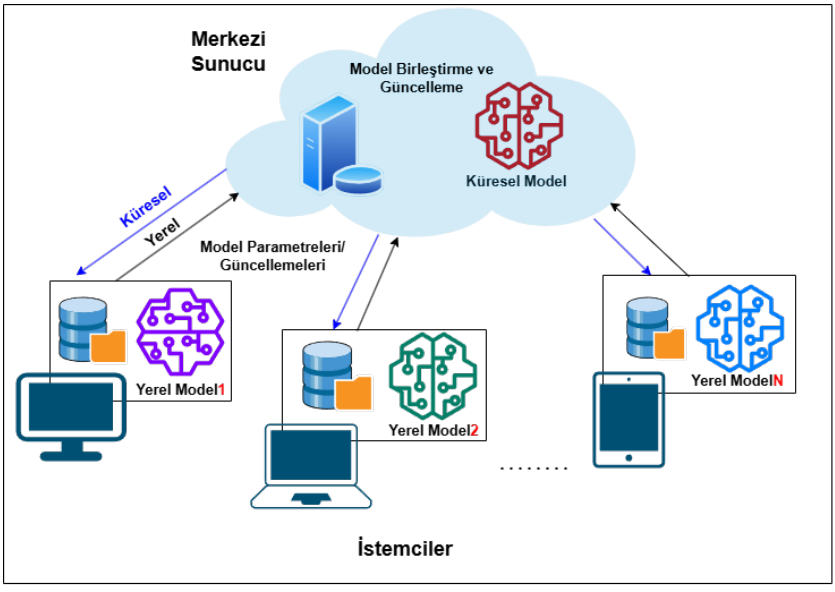
Şekil 1. Federe öğrenme uygulama alanları

Dağıtık ve gizlilik odaklı çözümler sunan federe öğrenmenin siber saldırı tespiti için kullanımı henüz yaygın olmamakla birlikte giderek daha fazla önem kazanmaktadır. Özellikle IoT ağlarında büyük avantajlar sunar: merkezi sistemlere bağımlılık azalır, kullanıcı verileri cihazlarda kalarak veri gizliliği korunur, farklı ağ mimarileri (heterojen ağ yapıları)

ile uyum sağlanır ve kötü niyetli faaliyetler dağıtık olarak tespit edilebilir. Bununla birlikte kötü niyetli istemciler, model güncellemeleri üzerinden hassas bilgilerin çıkarılması gibi potansiyel riskler de içermektedir. Bu konu, dördüncü bölümde daha detaylı tartışılmıştır.

2. FEDERE ÖĞRENME

İlk olarak 2016 yılında Google tarafından önerilen (McMahan vd., 2016) ve dağıtık bir makine öğrenmesi yaklaşımı olan federe öğrenme, merkezi bir modelin iş birlikçi bir şekilde eğitildiği/iyileştirildiği yinelemeli bir süreçtir (Zang vd., 2022; Canbay ve Büyüknacar, 2021). Şekil 2'de mimari yapısı gösterilen federe öğrenme, üç temel adımda gerçekleştirilir: 1) küresel model başlatma ve dağıtımı, 2) yerel model eğitimi ve 3) model birleştirme ve güncelleme.



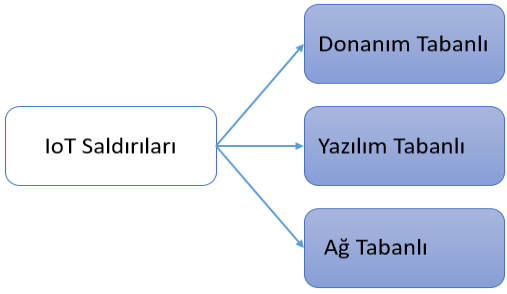
Şekil 2. Federe öğrenme mimarisi

İlk olarak merkezi bir sunucudaki küresel model, başlangıç parametreleri ile eğitilir ve ortamdaki tüm istemcilerle (IoT cihazları, mobil telefonlar, vb.) paylaşılır. Her bir istemci, kendi yerel verilerini kullanarak modelini öğrenir, model üzerinde yerel bir güncelleme hesaplar ve daha sonra model güncellemeleri/parametreleri, modelleri birleştiren ve öğrenme sürecini düzenleyen merkezi sunucuya iletir. Sunucu, kendisine iletilen güncellemeleri toplar, toplu bir model güncellemesi hesaplar ve hesaplanan toplu güncellemeyi kullanarak küresel modeli günceller. Küresel model parametreleri yeni bir yineleme için tekrar istemcilerle paylaşılır (Alazab vd., 2022). İletişim verimliliğini artırmak için model birleştirme süreci, kayıplı sıkıştırmayı içerebilir (Kairouz vd., 2021; Li vd., 2020; Žalik ve Žalik, 2023). Federe öğrenme ortamına yeni istemciler katılabilir, ancak bu ek öğrenme yinelemeleri gerektirmektedir (Žalik ve Žalik, 2023). Bir diğer husus istemci tarafı eğitim sürecinin sonunda modelin, yerel eğitim örneklerini aşırı öğrenmiş ve ezberlemiş olabileceğidir. Bu modeli paylaşmak, eğitim verilerini ifşa etme riski taşıyabilir. Bu riski azaltmak ve aşırı öğrenmeyi önlemek için seçici parametre paylaşım yöntemleri kullanılarak bir istemcinin paylaştığı bilgi miktarı sınırlandırılabilir (Li vd., 2019).

Veri kümesinin gizliliği ve iletişim gereksinimleri gibi avantajları nedeniyle son zamanlarda birçok alanda federe öğrenme tabanlı çözümler geliştirilmiştir. Federe öğrenme, özellikle hassas verilere erişim ve analiz sürecini düzenlemek, kullanıcı gizliliğini/veri sahipliğini korumak için yürürlüğe giren Sağlık Sigortası Taşınabilirlik ve Sorumluluk Yasası (Health Insurance Portability and Accountability Act, HIPAA), Genel Veri Koruma Tüzüğü (General Data Protection Regulation, GDPR) ve Kişisel Verilerin Korunması Kanunu (KVKK) gibi çeşitli mevzuatlara uyum için uygun çözümler sunmaktadır.

3. NESNELERİN İNTERNETİ VE GÜVENLİĞİ

IoT ağları, her geçen zaman diliminde internete bağlı cihazların sayısındaki artış ile büyümeye devam eden bir ağ türüdür. IoT ağları, zengin çeşitliliği ile farklı üreticilere ait donanım ve yazılıma sahip cihazları içerebilmektedir. Tutarsız kalan güvenlik standartlarının yol açtığı zafiyetler nedeniyle saldırı yüzeyi genişleyen IoT ağlarında; kimlik sahtekarlığı (spoofing), ortadaki adam saldırısı (Man in The Middle, MiTM), Dağıtık Hizmet Reddi (Distributed Denial of Service, DDoS) ve yan kanal (side channel) saldırıları gibi farklı türde siber saldırılara karşı cihazlar savunmasız hale gelir ve önemli bir güvenlik zorluğu oluşturur. IoT cihazlarına yönelik gerçekleştirilen saldırıları, yöntemlerine göre Şekil 3'te verildiği gibi donanım tabanlı, yazılım tabanlı ve ağ tabanlı saldırılar olmak üzere üç temel gruba ayırmak mümkündür (Sasi vd., 2024).



Şekil 3. Yöntemlerine göre IoT saldırıları

- Donanım tabanlı saldırılar, Açık Sistemler Bağlantısı (Open Systems Interconnection, OSI) katmanlarından fiziksel katman ile veri bağı katmanlarında haberleşme ve alt yapı amaçlı kullanılan cihazlara yetkisiz erişimle başlayan ve cihazların özelliklerine bağlı olarak güvenlik açıklarını ve gömülü sistem bileşenlerini çeşitli yöntemlerle hedef alan saldırılardır. En bilinen donanım tabanlı saldırı türü, yan kanal saldırılarıdır. Fiziksel yan

alanları analiz ederek, şifreleme anahtarlarını ya da diğer hassas verileri elde etmek amacıyla gerçekleştirilirler. Donanımsal arka kapı (back door), firmware (donanım yazılımları) saldırıları, Evrensel Seri Veri Yolu (Universal Serial Bus, USB) saldırıları ve sahte çip klonlama saldırıları diğer önemli saldırı türleridir.

- Yazılım tabanlı saldırılar, IoT cihazlarının yazılımı ve/veya aygıt yazılım ile ilişkili güvenlik açıkları kullanılarak gerçekleştirilen saldırılardır (Neshenko vd., 2019). Saldırganlar, IoT cihazlarında çalışan yazılımların güvenlik açıklarından faydalanarak kötü amaçlı yazılım (malware) bulaştırabilirler, arka kapı kullanarak sisteme sızabilirler ya da IoT cihazlarının kontrollerini ele geçirdikten sonra botnet ağının bir parçası olarak kullanılmalarını sağlayarak DDoS saldırıları gerçekleştirebilirler. Günümüzde büyük veriyi oluşturan önemli elemanlardan olan IoT cihazlarının kullanımı ve eylemleri hakkındaki bilgiler, kullanıcılarının davranış ve alışkanlıklarının profillenmesine olanak tanımanın yanında kullanıcı gizliliğinin ihlal edilmesine neden olabilmektedir (Nguyen vd., 2020). Cihazlardan toplanan kişisel verilerin ele geçirilmesi veri ihlalleri ile verinin gizliliği ilkesine de aykırılık teşkil etmektedir.
- Ağ tabanlı saldırılar, IoT cihazlarının bağlı olduğu ağ üzerinden iletişim trafiğini hedef alan saldırılardır. MiTM, Hizmet Reddi (Denial of Service, DoS), DDoS, oltalama (phishing) saldırıları, web site veri tabanlarına yönelik gerçekleştirilen SQL enjeksiyonu saldırıları ve dinleme (eavesdropping) saldırıları örnek olarak verilebilir. IoT cihazları internet aracılığı ile birbirleriyle haberleşerek günümüz teknolojilerinde büyük kolaylık ve imkanlar sunmaktadır. Ancak bu cihazların hızla yaygınlaşması özellikle ağ tabanlı saldırılarda ciddi artışa neden

olmaktadır. Ağ tabanlı saldırılarda, enerji, sağlık, endüstri, akıllı şehirler gibi kritik alanlarda kullanılan gömülü sistemler en zayıf alan olarak görülmektedir.

Kaynak kısıtlı olan IoT cihazlarının sınırlı işlem gücü ve bellek kapasiteleri (Ioulianou vd., 2018) sebebiyle alınacak güvenlik önlemleri ifade edilen bu saldırılara karşı oldukça sınırlıdır. Kullanıcılar ve kurumlar için ciddi tehdit durumunda olan IoT saldırılarına karşı proaktif güvenlik önlemleri almak ve güvenlik stratejilerini güncel tutmak hayati öneme sahiptir. Saldırı tespit sistemleri, IoT'ye yönelik kimlik sahtekarlığı saldırıları, DoS/DDoS saldırıları, MiTM saldırıları, host tabanlı ve ağ tabanlı saldırılar gibi çeşitli saldırı türlerinin etkili bir şekilde tespit edilmesini sağlayan proaktif bir önleme yöntemidir. Saldırı tespit sistemleri ile IoT cihazları ve ağları, siber saldırılara karşı dayanıklı hale getirilerek daha etkili güvenlik sağlanır.

4. FEDERE ÖĞRENME TABANLI SALDIRI TESPİTİ

IoT ağlarını kötü amaçlı saldırılardan korumak için önerilen birçok araştırma vardır. Sahu ve Mukherjee (2020) çalışmalarında, akıllı ev IoT cihazlarındaki anormallikleri tespit etmek amacıyla makine öğrenmesi tabanlı bir tespit sistemi önermişlerdir. Cakir vd. (2021), kaynak kısıtlı IoT cihazlarına yönelik merhaba taşkını (hello flooding) saldırısının tespitinde Geçitli Tekrarlayan Birim (Gated Recurrent Unit, GRU) yöntemini kullanarak yüksek başarım elde etmişlerdir. Diwan vd. (2021) tarafından IoT ağlarındaki kötü amaçlı ağ trafiğinin tespiti için makine öğrenmesine dayalı bir çerçeve önerilmiş ve en yüksek %99,96 doğruluk değeri bildirilmiştir. IoT ortamında botnet saldırılarını tespit etmek ve sınıflandırmak için yapılan çalışmada ise (Alissa vd., 2022) %94 doğruluğa ulaşılmıştır. Yalçın vd. (2024) çalışmalarında, endüstriyel IoT güvenliğine

odaklanmışlar ve farklı kategorilerde makine öğrenmesi yöntemleri kullanarak saldırı tespit modelleri geliştirmişlerdir. Tüm modeller ile %96,82'yi aşan test doğruluk oranlarına ulaşmışlardır. Kavitha ve Ramalakshmi (2024), DDoS saldırılarının tespiti ve önlenmesine yönelik makine öğrenmesi yöntemlerine dayalı etkili bir yaklaşım sunmuşlar ve %99,99 doğruluk elde etmişlerdir. Makine öğrenmesi kullanımının saldırıları doğru bir şekilde tespit etmede çok etkili olduğu açıktır. Makine öğrenmesi tabanlı saldırı tespit sistemleri, IoT güvenliği için güçlü araçlar olsalar da veri gizliliğini garanti etmezler ve eğitim sürecinde genellikle büyük veri kümeleri gerektirirler. Ayrıca veri ön işleme, veri birleştirme, hesaplama maliyeti ve özellikle IoT ağlarındaki gibi kaynak karmaşıklığı gibi sınırlamalar nedeniyle dikkatli tasarlanmalıdırlar.

Federe öğrenme, eğitim mekanizmasının sağladığı avantajlar nedeniyle IoT ağlarında saldırı tespiti için etkili ve güçlü çözümler sunmaktadır. IoT cihazlarının saldırıları yerel olarak öğrenmesini ve tespit etmesini sağlar. Akıllı ev ağlarında kötü amaçlı yazılım ile enfekte olmuş IoT cihazlarını tespit etmek için Nguyen vd. (2019) tarafından federe öğrenme tabanlı bir yaklaşım sunulmuştur. Önerilen yaklaşım ile ayrıca yanlış alarmlar en aza indirilmiştir. Schneble ve Thamilarasu (2019), tıbbi siber fiziksel sistemlerin güvenliğini iyileştirmek ve daha sağlam bir saldırı tespiti sağlamak için federe öğrenme yaklaşımının uygulanabilirliğini araştırmışlardır. Tek bir makine öğrenmesi modelinin kullanımına kıyasla federe öğrenme yaklaşımı benimsemelerinin, daha iyi eğitim süresi sağladığını, tespit doğruluğunu ve yanlış pozitif oranını da anlamlı bir oranda iyileştirdiğini belirtmişlerdir. Friha vd. (2022), tarımsal IoT altyapılarına yönelik siber saldırıları azaltmak/önlemek için federe derin öğrenme tabanlı bir saldırı tespit sistemi önermişlerdir. Önerilen sistemin, IoT cihazlarının verilerinin gizliliğini korumada klasik/merkezi sürümlerinden daha iyi

performans gösterdiği ve saldırıları tespit etmede en yüksek doğruluğu sunduğu ifade edilmiştir. IoT ağlarındaki izinsiz girişi proaktif olarak tanımak için federe öğrenme tabanlı anomali tespitinin önerildiği çalışmada (Mothukuri vd. 2022), kullanıcı verilerinin gizliliğini güvence altına almada klasik makine öğrenmesinden daha iyi performans elde edilmiş ve saldırı tespitinde optimum bir doğruluk oranı sağlanmıştır. DDoS saldırı tespiti için önerilen uyarlanabilir bir federe öğrenme yaklaşımının (Doriguzzi-Corin ve Siracusa, 2024), dengesiz veri kümeleri arasında birleşme süresi ve doğruluk açısından daha iyi performans gösterdiğini görülmüştür. IoT için farklı saldırıların tespitine yönelik Campos vd. tarafından gerçekleştirilen çalışmada (Campos vd., 2022) ise çok sınıflı bir sınıflandırıcıya dayalı federe öğrenme tabanlı saldırı tespit yaklaşımı değerlendirilmiştir. Değerlendirmede farklı veri dağıtımları, eğitim turlarını ve toplama yöntemleri dikkate alınmıştır.

Yine de federe öğrenme tabanlı saldırı tespit sistemleri, özellikle arka kapı saldırıları olmak üzere zehirlenme (poisoning) saldırılarına karşı savunmasızdır. Bu saldırı yönteminde saldırgan, saldırı ya da anormallik tespit sisteminden kaçınmak için veri setine (normal etiketli) kötü niyetli trafiği gizlice enjekte ederek eğitim verilerini zehirler. Böylece, ortaya çıkan model kötü niyetli trafiği yanlış bir şekilde "normal" olarak sınıflandırır ve bu tür saldırı trafiği kalıpları için bir alarm oluşturmaz. Bu tür saldırılara karşı savunmayı güçlendirmek için kötü amaçlı trafik enjeksiyonunu belirleme, istemci tarafında zehirli verileri filtreleme veya tolere etme gibi yöntemlere başvurulabilir (Nguyen vd., 2020; Fung vd., 2018; Shen vd., 2016). Diğer önemli bir husus da IoT cihazların heterojenliğinden dolayı, bazıları yerel eğitimi birkaç milisaniyede gerçekleştirebilirken, diğer cihazların modeli güncellemek için daha uzun bir süreye ihtiyaç duyabilmesidir (örneğin, kaynak kısıtlamaları nedeniyle). Bu durum genel federe eğitimini yavaşlatabilir ve saldırı tespit

sistemi açısından değerlendirildiğinde, belirli bir saldırının tespit edilmesinde daha uzun bir gecikmeye yol açabilir ve dolayısıyla ağın genel siber güvenliği üzerinde ciddi sonuçlar doğurabilir (Nishio ve Yonetani, 2019; Campos vd., 2022).

5. DEĞERLENDİRME VE ÖNERİLER

Federe öğrenme ile IoT verileri merkezi bir sunucuya yüklenmeden işlenerek veri gizliliği korunur, iletişim genel olarak daha verimli hale gelir ve veriler yerel cihazlarda tutulduğu (başka taraflarla paylaşmayı gerektirmediği) için de veri sızıntısı riski azaltılır. Bununla birlikte, performansı kablosuz kanalların koşullarına bağlıdır ve birçok son kullanıcının katılımı nedeniyle çeşitli siber tehditlere ve gizlilik sorunlarına karşı hala savunmasız olabilir. Verileri ifşa etme riskini azaltmak için seçici parametre paylaşım yöntemleri kullanılabilir ve böylece bir istemcinin paylaştığı bilgi miktarı sınırlandırılmış olur. Federe öğrenme ile dağıtılmış IoT cihazlarını etkili bir şekilde bağlamak mümkündür, ancak büyük miktarda verimli bir şekilde işleyebilmek için optimizasyon algoritmalarına ihtiyaç vardır. Bununla birlikte model sıkıştırma yöntemleri de verimliliği artırmak için kullanılabilir. Ek olarak federe öğrenmenin sunduğu avantajları en üst düzeye çıkarmak için sistemler tasarlanırken kriptografi yöntemleri (homomorfik şifreleme, Güvenli Çok Taraflı Hesaplama (Secure Multi-Party Computation, SMPC) gibi), blok zincir ve diferansiyel mahremiyet teknolojileri entegre edilebilir. Böylece model güncellemelerinin güvenilirliği artırılabilir ve daha güçlü bir koruma sağlanabilir.

KAYNAKÇA

Alazab, M., RM, S. P., P. M, Maddikunta, P. K. R., Gadekallu T. R., Pham, Q. -V. 2022. Federated Learning for Cybersecurity: Concepts, Challenges, and Future Directions, IEEE Transactions on Industrial Informatics, 18(5):3501-3509, doi: 10.1109/TII.2021.3119038.

Alissa, K., Alyas, T., Zafar, K., Abbas, Q., Tabassum, N., Sakib, S. 2022. Botnet Attack Detection in IoT Using Machine Learning, Computational Intelligence and Neuroscience, 2022(1), 4515642.

Al Mallah, R., Badu-Marfo, G., Farooq, B. 2021. Cybersecurity Threats in Connected and Automated Vehicles based Federated Learning Systems, IEEE Intelligent Vehicles Symposium Workshops (IV Workshops), Nagoya, Japan, 13-18.

Campos, E. M., Saura, P. F., González-Vidal, A., Hernández-Ramos, J. L., Bernabé, J. B., Baldini, G., Skarmeta, A. 2022. Evaluating Federated Learning for Intrusion Detection in Internet of Things: Review and Challenges, Computer Networks, 203, 108661.

Canbay Y., Büyüknacar Y. 2021. Federe Öğrenme ve Veri Mahremiyeti, Yapay Zekâ ve Büyük Veri Çalışmaları, Siber Güvenlik ve Mahremiyet (Editörler: Ş. Sağıroğlu, M. U. Demirezen), Cilt 3, Nobel Yayınevi.

Cakir, S., Toklu, S., Yalcin, N. 2020. RPL Attack Detection and Prevention in the Internet of Things Networks Using a GRU Based Deep Learning, IEEE Access, 8:183678-183689, doi: 10.1109/ACCESS.2020.3029191.

Diwan, T. D., Choubey, S., Hota, H. S., Goyal, S. B., Jamal, S. S., Shukla, P. K., Tiwari, B. 2021. Feature Entropy Estimation (FEE) for malicious IoT Traffic and Detection

Using Machine Learning, Mobile Information Systems, 2021(1):8091363.

Doriguzzi-Corin, R., Siracusa, D. 2024. FLAD: Adaptive Federated Learning for DDoS Attack Detection, Computers & Security, 137, 103597.

Fekri, M. N., Grolinger, K., Mir, S. 2022. Distributed Load Forecasting Using Smart Meter Data: Federated Learning with Recurrent Neural Networks, International Journal of Electrical Power & Energy Systems, 137:107669.

Friha, O., Ferrag, M. A., Shu, L., Maglaras, L., Choo, K. K. R., Nafaa, M. 2022. FELIDS: Federated Learning-Based Intrusion Detection System for Agricultural Internet of Things, Journal of Parallel and Distributed Computing, 165, 17-31.

Fung, C., Yoon, C. J. M., Beschastnikh, I. 2018. Mitigating Sybils in Federated Learning Poisoning, CoRR, abs/1808.04866.

Gao, J., Wang, W., Liu, Z., Billah, M. F. R. M., Campbell, B. 2021. Decentralized Federated Learning Framework for the Neighborhood: A Case Study on Residential Building Load Forecasting, 19th ACM Conference on Embedded Networked Sensor Systems, 453-459.

Hard, A., Rao, K., Mathews, R., Beaufays, F., Augenstein, S., Eichner, H., Kiddon, C., Ramage, D. 2018. Federated Learning for Mobile Keyboard Prediction, arXiv:1811.03604.

Huang, L., Liu, D. 2019. Patient Clustering Improves Efficiency of Federated Machine Learning to Predict Mortality and Hospital Stay Time Using Distributed Electronic Medical Records, arXiv:1903.09296.

Idoje, G., Dagiuklas, T., Iqbal, M. 2023. Federated Learning: Crop Classification in A Smart Farm Decentralised

Network, Smart Agricultural Technology, 5:100277, doi: 10.1016/j.atech.2023.100277.

Ioulianou, P., Vasilakis, V., Moscholios, I., Logothetis, M. 2018. A Signature-Based Intrusion Detection System for the Internet of Things, Information and Communication Technology Form.

Kairouz, P., McMahan, H. B., Avent, B., Bellet, A., Bennis, M., Bhagoji, A. N., Bonawit, K., Charles, Z., Cormode, G., Cummings, R., et al. 2021. Advances and Open Problems in Federated Learning,Foundations and Trends® in Machine Learning, 14(1-2):1-210.

Khramtsova, E., Hammerschmidt, C., Lagraa, S., State, R. 2020. Federated Learning for Cyber Security: SOC Collaboration for Malicious URL Detection, IEEE 40th international conference on distributed computing systems (ICDCS), 1316-1321.

Lee, S., Choi, D. H. 2020. Federated Reinforcement Learning for Energy Management of Multiple Smart Homes with Distributed Energy Resources, IEEE Transactions on Industrial Informatics, 18(1):488-497.

Li, T., Sahu, A. K., Talwalkar, A., Smith, V. 2020. Federated learning: Challenges, Methods, and Future Directions, IEEE Signal Processing Magazine, 37(3):50-60.

Li, W., Milletarì, F., Xu, D., Rieke, N., Hancox, J., Zhu, W., et al. 2019. Privacy-Preserving Federated Brain Tumour Segmentation, arXiv:1910.00962.

Manoj, T., Makkithaya, K., Narendra, V. G. 2022. A Federated Learning-Based Crop Yield Prediction for Agricultural Production Risk Management, IEEE Delhi Section Conference (DELCON), New Delhi, India, 1-7.

Mao, A., Huang, E., Gan, H., Liu, K. 2022. FedAAR: A Novel Federated Learning Framework for Animal Activity Recognition with Wearable Sensors, Animals, 12:2142.

McMahan, H. B., Moore, E., Ramage, D., Hampson, S., Arcas, B. A. 2016. Communication-Efficient Learning of Deep Networks from Decentralized Data, arXiv:1602.05629.

McMahan, H. B., Ramage, D., Talwar, K., Zhang, L. 2017. Learning Differentially Private Recurrent Language Models, arXiv:1710.06963.

Mothukuri, V., Khare, P., Parizi, R. M., Pouriyeh, S., Dehghantanha, A., Srivastava, G. 2022. Federated-Learning-Based Anomaly Detection for IoT Security Attacks, IEEE Internet of Things Journal, 9(4):2545-2554, doi: 10.1109/JIOT.2021.3077803.

Neshenko, N., Bou-Harb, E., Crichigno, J., Kaddoum, G., Ghani, N. 2019. Demystifying IoT Security: An Exhaustive Survey on IoT Vulnerabilities and a First Empirical Look on Internet-Scale IoT Exploitations, IEEE Communications Surveys & Tutorials, 21(3):2702-2733.

Nguyen, T. D., Marchal, S., Miettinen, M., Fereidooni, H., Asokan N., Sadeghi, A.-R. 2019. DÏoT: A federated Self-Learning Anomaly Detection System for IoT, 39th International Conference on Distributed Computing Systems (ICDCS 2019), 756-767.

Nguyen, T. D., Rieger, P., Miettinen, M., Sadeghi, A. R. 2020. Poisoning Attacks on Federated Learning-Based IoT Intrusion Detection System, Workshop on Decentralized IoT Systems and Security (DISS), 79, doi: 10.14722/diss.2020.23003.

Nishio, T., Yonetani, R. 2019. Client Selection for Federated Learning with Heterogeneous Resources in Mobile Edge,

2019 IEEE International Conference on Communications (ICC), 1-7.

Sahu N. K. ve Mukherjee I. 2020. Machine Learning Based Anomaly Detection for IoT Network: (Anomaly Detection in IoT Network), 4th International Conference Trends in Electronics and Informatics (ICOEI), 787-794.

Samarakoon, S., Bennis, M., Saad, W., Debbah, M. 2018. Federated Learning for Ultra-Reliable Low-Latency V2V Communications, 2018 IEEE Global Communications Conference (GLOBECOM), 1-7.

Sasi, T., Lashkari, A. H., Lu, R., Xiong, P., Iqbal, S. 2024. A Comprehensive Survey on IoT Attacks: Taxonomy, Detection Mechanisms and Challenges, Journal of Information and intelligence, 2(6):455-513.

Schneble, W., Thamilarasu, G. 2019. Attack Detection Using Federated Learning in Medical Cyber-Physical Systems, 28th EEE International Conference on Computer Communications and Networks (ICCCN), 29, 1-8.

Shen, S., Tople, S., Saxena, P. 2016. Auror: Defending Against Poisoning Attacks in Collaborative Deep Learning Systems, 32nd Annual Conference on Computer Security Applications (ACSAC ’16), 508-519, doi: 10.1145/2991079.2991125.

Xu, J., Xu, Z., Walker, P., Wang, F. 2020. Federated Patient Hashing, Proceedings of the AAAI Conference on Artificial Intelligence, 3(4):6486-6493, doi: 10.1609/aaai.v34i04.6121.

Wang, Y., Bennani, I. L., Liu, X., Sun, M., Zhou, Y. 2021. Electricity Consumer Characteristics Identification: A Federated Learning Approach, IEEE Transactions on Smart Grid, 12(4):3637-3647.

Yalçın, N., Çakır, S., Ünaldı, S. 2024. Attack Detection Using Artificial Intelligence Methods for SCADA Security, IEEE Internet of Things Journal, 11(24):39550-39559, doi: 10.1109/JIOT.2024.3447876.

Yuan, B., Ge, S., Xing, W. 2020. A Federated Learning Framework for Healthcare IoT Devices, arXiv:2005.05083.

Žalik, K. R., Žalik, M. 2023. A Review of Federated Learning in Agriculture, Sensors, 23(23):9566.

Zang, L., Zhang, X., Guo, B. 2022. Federated Deep Reinforcement Learning for Online Task Offloading and Resource Allocation in WPC-MEC Networks, IEEE Access, 10:9856-9867.

Zhao, Y., Zhao, J., Jiang, L., Tan, R., Niyato, D. 2019. Mobile Edge Computing, Blockchain and Reputation-Based Crowdsourcing IoT Federated Learning: A Secure, Decentralized and Privacy-Preserving System, arXiv:1906.10893.

FEATURE ENGINEERING VERSUS AUTOMATIC FEATURE EXTRACTION METHODS: NECESSITY AND ROLE

Emre DELİBAŞ[1]

1. INTRODUCTION

Advances in machine learning and deep learning have also transformed data processing and feature extraction techniques. Especially with the increase in big data and computational capacity, traditional feature engineering approaches are increasingly being replaced by automatic feature extraction methods. However, despite this development, questioning the importance and necessity of traditional feature engineering remains a critical issue on the agenda.

1.1. Definition and purpose of feature engineering

Feature engineering is the general name of a series of transformation, selection and improvement techniques that aim to produce meaningful and effective representations from raw data. In order for a model to make successful predictions, the data must be processed appropriately. In this process, making features meaningful, making statistical inferences and removing excess noise play a critical role (Guyon, Elisseeff, & Kaelbling, 2003).

Traditional feature engineering means that data scientists optimize the obtained data for the model using domain knowledge. This process includes techniques such as statistical

[1] Assistant Professor, Sivas Cumhuriyet University, Faculty of Engineering, Department of Computer Engineering, edelibas@cumhuriyet.edu.tr, ORCID: 0000-0001-7564-5020.

transformations, feature selection, dimensionality reduction and data normalization. However, the time-consuming and sometimes subjective nature of traditional methods has revealed the need for automation in this field.

1.2. The role of features in machine learning and deep learning

Machine learning models are algorithms that can generalize by extracting meaningful representations from data. Traditional machine learning models (support vector machines, decision trees, logistic regression, etc.) require the determination of appropriate features for the model to be successful (Bengio, Courville, & Vincent, 2013). When the wrong features are selected, the generalization ability of the model decreases and performance is lost.

Deep learning approaches differ in their ability to learn features directly from data. Convolutional Neural Networks (CNN) have an automatic feature extraction mechanism in image processing, Recurrent Neural Networks (RNN) in sequential data, and Transformer architecture in language models (Lecun, Bengio, & Hinton, 2015) Therefore, one of the fundamental differences between traditional machine learning and deep learning is whether the features are extracted manually or automatically.

1.3. The emergence of automatic feature extraction methods

Automatic feature extraction has come to the fore with the ability of deep learning models to work with big data and determine complex structural features. In particular, the concept of representation learning has entered the literature as an approach that aims to create meaningful representations from data (Bengio et al., 2013).

The reasons for the emergence of these methods are as follows:

- **The need for big data:** Automatic feature extraction has become critical in order to use more data effectively.
- **The limitation of manually designed features:** Manual feature extraction is sometimes insufficient in complex structured data.
- **Increased learning capacity of the model:** Deep learning models have reached the capacity to learn more complex representations.

1.4. The role of feature engineering in this process

Although deep learning-based approaches have the ability to automatically extract features, traditional feature engineering still maintains its importance. Especially in areas such as medicine, finance and law, where explainability is critical, it is preferred that manually determined features are interpretable (Dhaygude et al., 2023a). In addition, traditional feature engineering improves model performance in cases where data is scarce.

2. FEATURE ENGINEERING AND AUTOMATIC FEATURE EXTRACTION

The success of machine learning models depends not only on the complexity of the algorithm but also on how well the data is represented. While feature engineering consists of a set of techniques applied to make raw data modelable, automatic feature extraction allows direct learning of features from data through deep learning-based approaches. In this section, the differences, advantages, and limitations between traditional feature

engineering and automatic feature extraction methods will be discussed.

2.1. Traditional feature engineering

Traditional feature engineering is based on the process of predetermining features developed by data scientists and domain experts. In these methods, statistical, mathematical, and heuristic techniques are used to extract meaningful features from data (Y. Zhu, Zhong, Lu, & Yang, 2013). Especially in applications with small data, traditional feature engineering usually works more efficiently.

The basic steps of traditional feature engineering are as follows:

- **Feature transformations:** Changing the statistical properties of the data. For example, the distributions of variables can be normalized by applying logarithmic or square root transformations.
- **Feature selection:** Removing unnecessary or irrelevant features to improve model performance (Hamdard & Lodin, 2023).
- **Dimensionality reduction:** Reducing the dimensionality of the data using methods such as principal component analysis (PCA) or t-SNE (Shen, 2023).

2.1.1. Advantages and disadvantages

Advantages:

- **Using domain knowledge:** Domain experts can add important information about the data.
- **Low data dependency:** It can perform well with little data.

- **Explainability:** It is easier to make sense of the model's decisions.

Disadvantages:

- **Manual labor requirement:** Data scientists need to spend a lot of time.
- **Risk of bias:** Human-determined features may be biased.
- **Problems fitting complex data:** It may be ineffective in large and multidimensional data.

2.2. Automatic feature extraction

Automatic feature extraction is based on deep learning models directly learning meaningful features in data. These methods are used by deep neural networks (DNN), Convolutional Neural Networks (CNN) and Transformer-based models, especially those that work with large data sets (Dhaygude et al., 2023b).

2.2.1. Advantages and disadvantages

Advantages:

- **Minimal human intervention:** Automatically learns features from data.
- **Effective work with big data:** Deep learning models can process more data effectively.
- **Overall performance improvement:** Often performs better than manually designed features.

Disadvantages:

- **Requirement of big data and high computational power:** Requires large amounts of data and high computational capacity.

- **Explainability problem:** It may be difficult to understand the outputs of the model.

Comparison of traditional and automatic methods is given in Table 1.

Table 1. Comparison of traditional and automatic methods

Criteria	Traditional Feature Engineering	Automatic Feature Extraction
Data Requirement	Can work with little data	Requires large data
Domain Knowledge	Highly required	Does not require
Computational Cost	Lower	High
Explainability	High	Low
Performance	Can be medium/low	Usually higher

3. WHEN IS FEATURE ENGINEERING NECESSARY?

Although deep learning and automatic feature extraction methods have largely automated traditional feature engineering, traditional feature engineering continues to play a critical role in certain scenarios. This section will discuss the situations where traditional feature engineering is still necessary.

3.1. When working with small data

Deep learning models are usually dependent on large amounts of data and may lose their ability to generalize when there is not enough data. In scenarios where small data is used, traditional feature engineering becomes critical in terms of extracting maximum information from the data(Kanjilal & Uysal, 2021).

Advantages of traditional feature engineering when working with small data:

- **Making sense of data:** In cases with low data volume, extracting statistical features can increase the success of the model.
- **Preventing overfitting:** While models with many parameters memorize very quickly in small data, well-designed features can increase the generalization ability of the model (Zeng, Liu, Lu, Zhang, & Lu, 2023).

3.2. In Cases of noisy and incomplete data

Real-world data can often be incomplete, incorrect, or noisy. While deep learning models can work directly with raw data, their ability to deal with incomplete and incorrect data may be limited. At this point, traditional feature engineering with data cleaning and preprocessing techniques comes into play (Bala & Behal, 2024).

The role of traditional feature engineering in case of noisy data:

- **Imputation of missing data:** Missing parts of the data can be filled with mean, median, or regression-based imputation methods.
- **Cleaning of outliers:** The effect of noise can be reduced with techniques such as anomaly detection and winsorization.
- **Summarizing the data with dimensionality reduction:** Data can be made more compact with PCA or manifold learning techniques.

3.3. Applications where explainability is critical

In some areas, it is imperative that model decisions are understandable by humans. For example, in the fields of health, finance, and law, the reasons for the decisions made by the model must be clearly stated (Jagannathan et al., 2023). Automatic

feature extraction methods are usually a “black box” and the outputs are difficult to interpret.

In cases where explainability is important:

- It is necessary to understand why the model made a certain decision.
- It is mandatory for the decision-making mechanism to be clear due to regulations. (Credit decisions in banking, models presented as evidence in courts, etc.)
- Clinical decision support systems must be verifiable by doctors.

Therefore, in fields that prefer explainable models, manually designed features are still of critical importance.

3.4. Problems requiring specific domain knowledge

In some fields, expert knowledge is required to extract the best features from the data. Automatic feature extraction techniques are suitable for general use, but in fields such as bioinformatics, chemistry, materials science, and astronomy, specific features need to be defined manually (Bonidia et al., 2022).

Why is traditional feature engineering necessary in such cases?

- Features defined by domain experts can better reflect the physical and biological meanings of the data.
- Automatic methods may not be successful in extracting specific structural information.
- Without expert knowledge, the extracted features may be meaningless or incomplete.

3.5. Hybrid use of traditional and automatic methods

In some applications, the best results can be achieved by using both traditional and automatic feature extraction methods. For example:

- In the first stage, the data is optimized using traditional feature engineering.
- Then, additional features are learned with automatic feature extraction methods.
- In the last stage, the most meaningful features are determined by feature selection.

This hybrid approach combines the advantages of traditional and modern techniques, allowing for more effective and balanced models to be created (Gibert, Planes, Mateu, & Le, 2022).

4. HYBRID APPROACHES: FEATURE ENGINEERING AND AUTOMATIC FEATURE EXTRACTION

Traditional feature engineering and automatic feature extraction methods can be considered as complementary processes, not as replacements for each other. Especially in large-scale and complex datasets, combining both methods can improve model performance. In this section, we will discuss how hybrid approaches can be applied, their advantages, and limitations.

4.1. Fundamentals of the hybrid approach

Hybrid feature engineering is a model development strategy created by combining traditional techniques and machine learning methods (Gibert et al., 2022). This approach is shaped specifically in the following stages:

- **Preprocessing and traditional feature engineering:** Manually deriving meaningful features using missing data filling, noise removal, and domain knowledge.
- **Automatic feature extraction:** Extracting additional features from the data using deep learning models or statistical methods.
- **Feature optimization:** Evaluating manual and automatically generated features together to determine the most effective ones.
- **Model training and result analysis:** Training the model using selected features and performing performance analysis.

This process allows the model to generalize more strongly, while reducing the risk of overfitting and increasing the interpretability of the model.

4.2. Advantages of hybrid approaches

The biggest advantages of hybrid feature engineering approaches are as follows (Nimma & Uddagiri, 2024):

- **Performance Increase:** Meaningful features extracted by manual engineering can increase model accuracy when combined with deep representations produced by automatic methods.
- **Explainability:** It makes it easier to understand the features automatically extracted by deep learning models.
- **Generalization Ability:** It allows the development of more flexible and robust models for different data types and applications.

For example, in image processing applications, using CNN-based automatic feature extraction methods together with manually determined features using domain knowledge can provide advantages in terms of both accuracy and model interpretability.

4.3. Limitations of hybrid approaches

Although hybrid methods provide advantages, they may also present some challenges:

- **Computational cost:** Using both traditional and automated methods together may result in high computational requirements.
- **Feature Incompatibility:** Features extracted by different techniques may sometimes be contradictory and may negatively affect the learning process of the model.
- **Feature optimization challenges:** Determining which features provide the best results can be a complex process.

Despite these limitations, hybrid approaches offer significant benefits, especially in big data analysis and high-dimensional datasets (Srihari, Gholipour, Khoshkangini, & Orand, 2022).

4.4. Application areas of hybrid approaches

Hybrid approaches have been successfully applied in many different areas:

- **Healthcare sector:** Combining traditional medical statistics with deep learning-based image processing models.

- **Financial models:** Supporting manually determined risk factors in credit scoring systems with automatic feature extraction techniques.
- **Autonomous systems:** Processing sensor data together with traditional engineering and deep learning techniques.
- **Natural language processing (NLP):** Integrating rule-based language processing techniques with deep learning-based language models.

Using hybrid methods in these areas not only increases model accuracy, but also makes the model more robust and explainable.

4.5. Hybrid approaches in the future

With the development of machine learning and deep learning techniques, hybrid approaches are becoming more sophisticated. The following developments may come to the fore in the future:

- **Automatic feature engineering tools:** Automating manual processes with AI-supported feature engineering systems.
- **Adaptive feature selection:** Dynamically selecting the most appropriate features during the model's training process.
- **Explainable deep learning models:** Reducing the black-box nature of deep learning models and increasing interpretability.

The utilisation of hybrid approaches is gaining prevalence in practical applications. To illustrate this point, consider the domain of fraud detection. Conventional statistical indicators of fraud can be amalgamated with deep learning-based feature

extraction techniques to enhance the precision of detection. A comparable scenario can be observed in the analysis of biomedical data, wherein the integration of manual domain expertise with deep learning methodologies facilitates enhanced disease prediction.

5. THE FUTURE OF FEATURE ENGINEERING

The rapid developments in the fields of machine learning and deep learning are constantly changing the role and necessity of feature engineering. Although traditional methods are gradually being replaced by automatic feature extraction techniques, feature engineering has not completely disappeared. On the contrary, it is expected to become more sophisticated and semi-automatic in the future and work in integration with artificial intelligence-supported systems. In this section, the potential future directions of feature engineering, whether fully automatic systems are possible, and new trends in developing fields will be discussed (Q. Zhu et al., 2024).

5.1. Artificial intelligence-supported feature engineering

While traditional feature engineering processes are carried out manually, artificial intelligence-supported automatic feature engineering tools have become widespread in recent years. AutoML (Automated Machine Learning) systems develop algorithms that automatically extract features from data sets and select the best features (Truong et al., 2019).

How is Feature Engineering Evolving with Artificial Intelligence?

- **Automation of feature generation algorithms:** AI-supported systems can derive meaningful features from raw data.

- **Genetic algorithms and optimization techniques:** Evolutionary approaches are used to optimize feature selection.
- **AI-based explainability models:** New techniques are being developed to better understand which features a model uses and why (Aljalaud & Hosny, 2024).

5.2. Is it possible to eliminate feature engineering completely?

Some researchers argue that with the advancement of deep learning, feature engineering will become completely unnecessary. However, this claim faces several important obstacles:

- **Variability of data quality:** Not all datasets have robust representations that deep learning models can automatically extract. Incomplete, noisy, and imbalanced datasets require traditional feature engineering (Li, 2024).
- **Explainability and regulation:** In critical areas such as finance and healthcare, model decisions need to be explainable. It can be difficult to understand why the features produced by deep learning models are selected(Jagannathan et al., 2023).
- **Dataset-specific optimization requirement:** General-purpose deep learning models may not be sufficient for some specific datasets and may require specialized engineering by domain experts.

Considering these factors, it seems unlikely that feature engineering will completely disappear. However, it is likely to become more efficient by working in integration with AI-supported automated systems.

5.3. Future major trends

The following trends are expected to come to the fore in the field of feature engineering in the future:

- **Feature learning models becoming more effective:** The explainability and data efficiency of deep learning models will increase.
- **Domain expertise and ai collaboration:** It is expected that human experts and AI systems will work together to perform more effective feature engineering.
- **Automation of feature optimization:** Semi-automatic and fully automatic feature optimization tools are expected to become widespread(Q. Zhu et al., 2024).
- **Use of transfer learning and meta-learning with feature engineering:** Adaptation of features learned from pre-trained models to new datasets will accelerate.

Feature engineering still plays a critical role in machine learning processes despite the rise of automatic feature extraction methods. In the future, it is expected to become more efficient by combining with automatic systems supported by artificial intelligence. However, it is unlikely to disappear completely due to data quality, model explainability and specific application areas. Hybrid approaches will provide the most effective results by combining the power of both human expertise and artificial intelligence.

6. CONCLUSION AND EVALUATION

Developments in the fields of machine learning and deep learning have made automatic feature extraction methods increasingly effective. However, traditional feature engineering still plays a critical role in certain scenarios. This book chapter

has discussed the relationship between feature engineering and automatic feature extraction methods, their advantages, disadvantages, and future trends.

6.1. General evaluation

First, the basic components of traditional feature engineering and its role in machine learning models are examined. Traditional methods still provide a significant advantage, especially when working with small data sets, in situations requiring explainability, and in areas requiring domain knowledge. On the other hand, automatic feature extraction methods offer the ability to learn more powerful and generalizable representations from large data sets by minimizing human intervention thanks to deep learning models. However, the high computational cost of these methods, the need for large amounts of data, and the limitations in model explainability still pose a great challenge for some applications.

6.2. Balance of feature engineering and automatic methods

Developments in recent years have shown that traditional feature engineering and automatic feature extraction methods are complementary processes. In particular, hybrid approaches offer significant advantages in terms of optimizing model performance, increasing explainability, and adapting to different data types.

In summary, the following conclusions can be drawn:

- Traditional feature engineering is still indispensable in applications where little data is used, domain knowledge is required, and model explainability is critical.
- Automatic feature extraction methods offer effective solutions in large data sets, especially in complex

structural data such as image and natural language processing.

- Hybrid approaches can achieve the best results by combining traditional feature engineering and automatic methods.

6.3. Feature engineering in the future

The future of feature engineering is moving towards hybrid solutions where artificial intelligence-supported automated systems and human expertise are combined. Developments in areas such as AutoML and explainable artificial intelligence (XAI) provide important clues about how feature engineering will shape in the future. In the coming years, the following trends are expected to emerge:

- Development of AI-powered automated feature engineering tools
- Increased automation of feature selection and optimization processes
- Expansion of explainable feature engineering techniques
- Increased use of transfer learning and meta-learning in feature learning processes

6.4. Concluding remarks

Feature engineering will not disappear completely, but will continue to evolve. The best results will be achieved by integrating human expertise and AI in data-driven decision-making processes. The development of automated feature extraction techniques will not cause feature engineering to disappear completely, but rather will make it more powerful and flexible.

This book chapter aims to provide a framework for researchers who want to understand the balance between feature engineering and automated feature extraction, understand how and in which situations these techniques can be used, and evaluate future trends.

REFERENCES

Aljalaud, E., & Hosny, M. (2024). Counterfactual Explanation of AI Models Using an Adaptive Genetic Algorithm With Embedded Feature Weights. *IEEE Access*, *12*, 74993–75009. https://doi.org/10.1109/ACCESS.2024.3404043

Bala, B., & Behal, S. (2024). A Brief Survey of Data Preprocessing in Machine Learning and Deep Learning Techniques. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1755–1762. https://doi.org/10.1109/I-SMAC61858.2024.10714767

Bengio, Y., Courville, A., & Vincent, P. (2013). Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *35*(8), 1798–1828. https://doi.org/10.1109/TPAMI.2013.50

Bonidia, R. P., Santos, A. P. A., De Almeida, B. L. S., Stadler, P. F., Da Rocha, U. N., Sanches, D. S., & De Carvalho, A. C. P. L. F. (2022). BioAutoML: automated feature engineering and metalearning to predict noncoding RNAs in bacteria. *Briefings in Bioinformatics*, *23*(4). https://doi.org/10.1093/BIB/BBAC218

Dhaygude, A. D., Varma, R. A., Yerpude, P., Swarnkar, S. K., Kumar Jindal, R., & Rabbi, F. (2023a). Deep Learning Approaches for Feature Extraction in Big Data Analytics. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, *10*, 964–969. https://doi.org/10.1109/UPCON59197.2023.10434607

Dhaygude, A. D., Varma, R. A., Yerpude, P., Swarnkar, S. K., Kumar Jindal, R., & Rabbi, F. (2023b). Deep Learning

Approaches for Feature Extraction in Big Data Analytics. *2023 10th IEEE Uttar Pradesh Section International Conference on Electrical, Electronics and Computer Engineering (UPCON)*, *10*, 964–969. https://doi.org/10.1109/UPCON59197.2023.10434607

Gibert, D., Planes, J., Mateu, C., & Le, Q. (2022). Fusing Feature Engineering and Deep Learning: A Case Study for Malware Classification. *Expert Syst. Appl.*, *207*. https://doi.org/10.1016/J.ESWA.2022.117957

Guyon, I., Elisseeff, A., & Kaelbling, L. P. (2003). An introduction to variable and feature selection. *The Journal of Machine Learning Research*, *3*, 1157–1182. https://doi.org/10.5555/944919.944968

Hamdard, Asst. P. M. S., & Lodin, Asst. P. H. (2023). Effect of Feature Selection on the Accuracy of Machine Learning Model. *INTERNATIONAL JOURNAL OF MULTIDISCIPLINARY RESEARCH AND ANALYSIS*, *06*(09). https://doi.org/10.47191/IJMRA/V6-I9-66

Jagannathan, J., K, Dr. A. R., Labhade-Kumar, Dr. N., Rastogi, R., Unni, M. V., & Baseer, K. K. (2023). Developing interpretable models and techniques for explainable AI in decision-making. *The Scientific Temper*, *14*(04), 1324–1331. https://doi.org/10.58414/SCIENTIFICTEMPER.2023.14.4.39

Kanjilal, R., & Uysal, I. (2021). The Future of Human Activity Recognition: Deep Learning or Feature Engineering? *Neural Processing Letters*, *53*(1), 561–579. https://doi.org/10.1007/S11063-020-10400-X

Lecun, Y., Bengio, Y., & Hinton, G. (2015). Deep learning. *Nature 2015 521:7553*, *521*(7553), 436–444. https://doi.org/10.1038/nature14539

Li, Z. (2024). Feature Engineering and Data Visualization Analysis in Artificial Intelligence in Big Data Era. *International Journal of Computer Science and Information Technology*, *3*(3), 390–395. https://doi.org/10.62051/IJCSIT.V3N3.41

Nimma, D., & Uddagiri, A. (2024). Enhancing Tackle Prediction in NFL Games Through Feature Engineering and Hybrid Machine Learning Models. *2024 8th International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, 1913–1919. https://doi.org/10.1109/I-SMAC61858.2024.10714680

Shen, Z. (2023). Comparison and Evaluation of Classical Dimensionality Reduction Methods. *Highlights in Science, Engineering and Technology*, *70*, 411–418. https://doi.org/10.54097/HSET.V70I.13890

Srihari, M., Gholipour, Z., Khoshkangini, R., & Orand, A. (2022). Optimization of the Hybrid Feature Learning Algorithm. *2022 Swedish Artificial Intelligence Society Workshop (SAIS)*, 1–8. https://doi.org/10.1109/SAIS55783.2022.9833044

Truong, A., Walters, A., Goodsitt, J., Hines, K., Bruss, C. B., & Farivar, R. (2019). Towards Automated Machine Learning: Evaluation and Comparison of AutoML Approaches and Tools. *2019 IEEE 31st International Conference on Tools with Artificial Intelligence (ICTAI), 2019-November*, 1471–1479. https://doi.org/10.1109/ICTAI.2019.00209

Zeng, Z., Liu, Y., Lu, X., Zhang, Y., & Lu, X. (2023). An Ensemble Framework Based on Fine Multi-Window Feature Engineering and Overfitting Prevention for Transportation Mode Recognition. *Adjunct Proceedings of the 2023 ACM International Joint Conference on Pervasive and Ubiquitous Computing & the 2023 ACM International Symposium on Wearable Computing*, 563–568. https://doi.org/10.1145/3594739.3610756

Zhu, Q., Wang, D., Ma, S., Wang, A. Y., Chen, Z., Khurana, U., & Ma, X. (2024). Towards Feature Engineering with Human and AI's Knowledge: Understanding Data Science Practitioners' Perceptions in Human&AI-Assisted Feature Engineering Design. *Proceedings of the 2024 ACM Designing Interactive Systems Conference, DIS 2024*, 1789–1804. https://doi.org/10.1145/3643834.3661517

Zhu, Y., Zhong, E., Lu, Z., & Yang, Q. (2013). Feature engineering for semantic place prediction. *Pervasive Mob. Comput.*, *9*(6), 772–783. https://doi.org/10.1016/J.PMCJ.2013.07.004

DATA-DRIVEN INSIGHTS INTO CONSUMER BEHAVIOR SHIFTS DURING CRISIS

Burcu YILMAZEL[1]

Ali YÜREKLİ[2]

1. INTRODUCTION

Crises such as global pandemics not only threaten human health but also leave lasting impacts on economies, consumer behaviors, and daily life routines. The novel coronavirus disease 2019 (COVID-19), which emerged in December 2019 (Zhu et al., 2020), has significantly altered social and economic dynamics worldwide. In response, governments have implemented strict measures such as lockdowns, travel restrictions, social distancing policies, and quarantine regulations to mitigate the spread of the virus (Stoecklin et al., 2020; Nicola et al., 2020). While these measures aim to reduce physical contact and control infection rates, they have also profoundly transformed consumer behaviors, particularly in the retail and e-commerce sectors (Anderson et al., 2020).

As movement restrictions took effect, consumers increasingly shifted their activities from physical stores to online platforms. Studies analyzing consumer mobility trends during periods of crisis have observed a significant decline in physical visits to workplaces, shopping malls, and grocery stores,

[1] Assistant Prof. Dr., Eskişehir Technical University, Faculty of Engineering, Department of Computer Engineering, byurekli@eskisehir.edu.tr, ORCID: 0000-0001-8917-6499.

[2] Assistant Prof. Dr., Eskişehir Technical University, Faculty of Engineering, Department of Computer Engineering, aliyurekli@eskisehir.edu.tr, ORCID: 0000-0001-8690-7559.

accompanied by a sharp increase in online transactions. Such shifts have historically accelerated digital adoption, compelling businesses to expand their e-commerce infrastructure to accommodate rising demand (Ozili & Arun, 2020). The pandemic has also introduced new consumption trends, including product hoarding, shifts in essential product categories, and changing shopping time preferences (Addo et al., 2020; Paul & Chowdhury, 2021; Roggeveen & Sethuraman, 2020).

One sector experiencing substantial disruption is grocery shopping, which represents a fundamental and continuous consumer activity—even in times of crisis (Szymkowiak et al., 2020). Globally, online grocery shopping has surged, driven by safety concerns and convenience (Kirk & Rifkin, 2020). Similar trends have been observed in Turkey, where the first confirmed COVID-19 case on March 11, 2020, triggered an immediate increase in online grocery orders (Pabuççiyan, 2020). This surge in demand forced major retailers to expand their supply chains and recruit additional workforce to accommodate the rising volume of orders (Petrushevska, 2020). Beyond the increase in online transactions, the pandemic has also altered shopping basket compositions, reflecting shifts in consumer priorities toward hygiene products, household essentials, and long shelf-life food items (Aydinli et al., 2021). These behavioral changes highlight how consumers respond to uncertainty and adapt to unforeseen disruptions in supply and demand.

This study examines the immediate effects of a large-scale crisis on consumer behavior in online grocery shopping, with a specific focus on the transition to digital consumption, shifts in product preferences, and changes in shopping patterns. Using real-world transaction data from a well-established e-commerce platform that connects consumers with grocery retailers, we analyze shopping trends over two distinct 30-day periods: before and after the first officially confirmed COVID-19 case in Turkey.

Our analysis aims to explore the evolving patterns of panic buying, digital adoption, and delivery preferences by addressing the following research questions (RQ):

- *RQ1: How does a public health crisis influence consumer attitudes toward online grocery shopping?*
- *RQ2: Which product categories experience significant demand fluctuations during crisis periods?*
- *RQ3: How do consumers adjust their active shopping hours and delivery window preferences in response to supply constraints and increased demand?*

By leveraging transaction-based data rather than self-reported surveys, this study provides empirical insights into consumer adaptation under unpredictable conditions. The findings contribute to the broader discourse on digital consumer behavior, e-commerce resilience, and crisis-driven market dynamics, offering valuable implications for businesses, policymakers, and researchers navigating uncertain environments.

2. LITERATURE REVIEW

Consumption habits are highly dependent on various factors including time, location, and context (Sheth, 2020). In the era of COVID-19, with lockdown and social distancing measures implemented by governments, people have time flexibility but location rigidity, which results in inevitable changes in consumption behaviors. The early days of the pandemic were passed by witnessing how people had rushed into grocery stores and left empty shelves behind. The main reason for such an exaggerated buying trend is hoarding behavior (Frost & Gross, 1993) among consumers. In a time of uncertainty for the future availability of supply chains, people tend to stockpile essential

products for their basic needs. In the case of the COVID-19 outbreak, stockpiling has been observed in toilet papers, bread, meat, water, disinfectants, and cleaning products (Kirk & Rifkin, 2020). Among these stock-out products, equipment for personal protection (which are relatively specific to the pandemic) can be associated with the fear appeal to purchase behavior (Addo et al., 2020). Studies examining the consumption habits during the pandemic from a financial perspective also emphasize the high increase in food and grocery purchases. Utilizing transaction-level financial data, Baker et al. (2020) document the early consumption response of the households in the United States and report a decline in overall spending, but an increase in grocery spending. A vital aspect in grocery consumption during COVID-19 is the shift towards online shopping. Along with social distancing measures, the fear of contradicting the novel coronavirus makes consumers more demanding in services that do not involve face-to-face contact (Watanabe & Omori, 2020). Accordingly, online grocery shopping services are in high demand globally since the early days of the pandemic. Li et al. (2020) compare grocery shopping channel preferences of Chinese consumers (i.e., neighborhood supermarket outlets, local small shops, farmer markets, and online shopping) before and during the outbreak. The authors report a surge in online shopping, with the percentage of consumers buying groceries online increasing from 11% to 38%, which promotes online shopping as the most popular consumption channel in China during the outbreak. In the Canadian fruit and vegetable markets, prior to the spread of COVID-19, only 1.5% of groceries were sold online, which then had grown to over 9% by the end of March 2020 (Richards & Rickard, 2020). Similarly, online demand had sharply increased in the German food and grocery retails, even some potential consumers could not place orders due to the unavailability of delivery services (Dannenberg et al., 2020). Global evidence from different parts of the world confirms that the pandemic

significantly impacted consumer behavior and accelerated the digitalization of consumption services (Sheth, 2020).

3. GROCERY SALES DATA COLLECTION

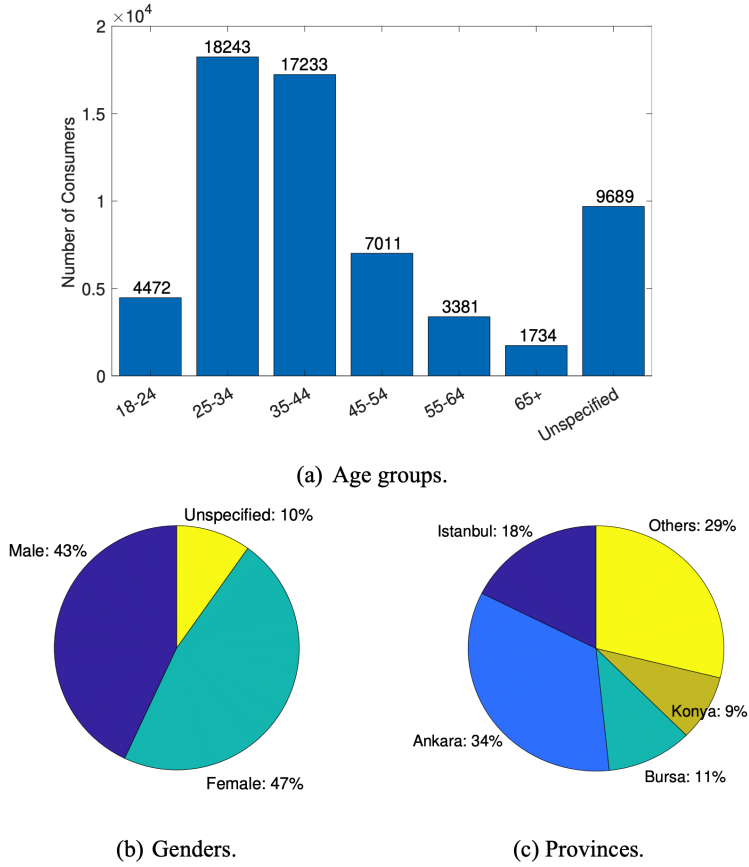
Marketyo was an e-commerce enterprise that provided omni-channel grocery sales services to more than 30 grocery chains in Turkey. Following its significant expansion during the COVID-19 pandemic, the company was later acquired by a major global player in online food and grocery delivery. This study analyzes a dataset of 116,453 consumer orders collected from the platform, spanning the period between February 11, 2020, and April 10, 2020. The dataset covers two distinct phases—before and after the first confirmed COVID-19 case in Turkey—enabling a comparative assessment of changes in consumer behavior. Further descriptive statistics regarding the dataset are presented in Table 1.

Table 1. Basic statistics of grocery sales data collection.

Property	Value	Description
Total orders	116,453	The collection includes 116,453 grocery orders.
Days covered	60	The orders cover a period of 60 days.
Grocery chains	32	The orders are placed from 32 different grocery chains.
Provinces and districts	26 - 93	The grocery chains serve in 93 districts in 26 provinces.
Consumers	61,321	The orders are placed by 61,321 different consumers.

The data collection also contains demographic information of the consumers that can be utilized when analyzing the trends in grocery sales during the COVID-19 outbreak. Figure 1 presents these overall demographics in terms of age groups, genders, and provinces where consumers reside. According to the figure, majority of the consumers are young people between the

ages of 25 and 44. It can be assumed that this age range represents the vast majority of the working population in a country. When the distribution of users is examined in terms of gender, it is seen that the number of female consumers is greater than the number of male consumers. Furthermore, the vast majority of orders belong to the consumers living in Ankara and Istanbul (covering 34% and 18% of the consumers, respectively), the two largest metropolitan cities of Turkey.



(a) Age groups.

(b) Genders.

(c) Provinces.

Figure 1. Overall demographics of the consumers by age group, genders and provinces.

4. EXPERIMENTAL ANALYSIS AND RESULTS

In this section, we present our analyses of changing consumer behaviors in online grocery shopping during the COVID-19 outbreak. The results and corresponding findings of the analyses are organized according to the three research questions given in Section 1, respectively.

4.1. The Shift to Online Grocery Shopping

The novel coronavirus has changed the way we purchase products. Consumers staying at home (due to social distancing measures) have turned to online grocery services for their daily needs (Richards & Rickard, 2020). In order to present the shift towards online consumption in Turkey, we examine the following criteria in daily grocery sales during the pandemic: (i) the number of daily orders and (ii) the number of new consumers registering Marketyo. As presented in Figure 2, there has been a sharp increase in online grocery orders after the announcement of the first confirmed coronavirus case in Turkey on March 11, 2020. Prior to this date, the sales were stable; there was not much difference between daily order numbers. As of this date, the number of daily orders has increased rapidly and reached a peak point with the lockdown of major cities. When we consider all the grocery orders in the collection in two periods (as pre-pandemic and pandemic), we see that 88% of all orders have been placed during the pandemic period. In other words, COVID-19 has increased online grocery sales nearly 7 times. Such a large increase in demand clearly reveals the sudden shift in consumption trends towards online transactions.

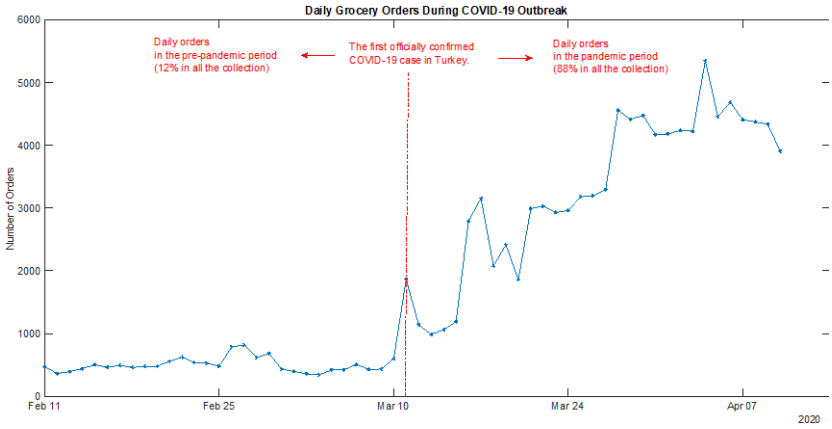


Figure 2. Daily grocery orders in the collection and the distribution of total orders in pre-pandemic and pandemic periods.

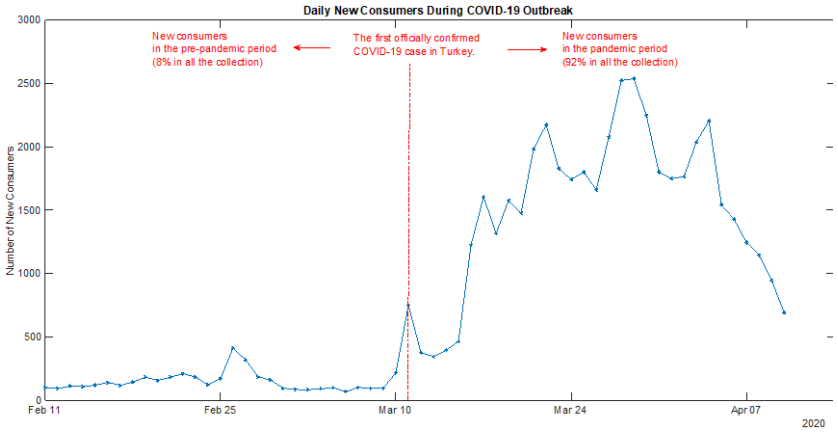


Figure 3. Daily new consumers in the collection and the distribution of total new consumers in pre-pandemic and pandemic periods.

One of the subjects where remarkable change is observed with the COVID-19 outbreak is the new registrations in the online grocery shopping platform. The increase in daily sales is not only dependent on existing users but also new consumers who sign up for the system. As illustrated in Figure 3, the number of daily new

consumers (approximately 150 new users per day) was stable before the pandemic. Since March 11, 2020, an enormous increase in the number of new registrations in the platform has been observed. Especially during the first two weeks of the pandemic (due to panic buying and hoarding behaviours (Sheth, 2020), the number of new consumers per day has reached up to 2,500. Afterwards, new registrations tend to decrease and normalize; still, more consumers register in the platform than in the pre-pandemic period. When we inspect all the new consumers in the collection, we observe that 92% of the new consumers have registered to the platform during the pandemic period.

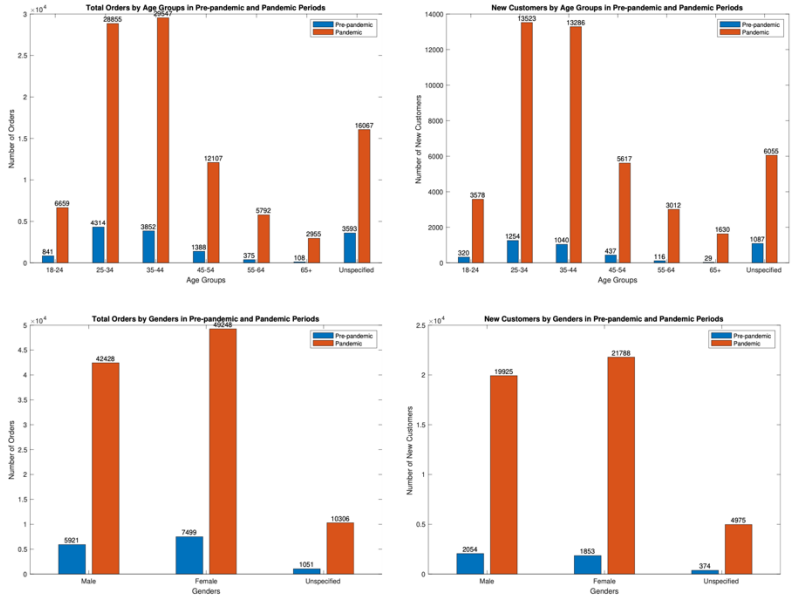


Figure 4. The changes in total grocery orders and new consumers by age groups and genders.

Further analysis on the increasing online shopping trend can be carried out by examining consumer demographics. As presented in Figure 4, the highest increase in both orders and new memberships has been observed in consumers between the ages of 25 and 44. Additionally, the demand for online shopping by

consumers over the age of 65, with whom the government imposed a curfew, has increased significantly compared to the past. Furthermore, it is also notable that female consumers have created more new memberships and placed more orders than male consumers.

4.2. Diversification of Product Preferences

The hoarding instinct leads consumers to make purchases beyond their needs, which makes consumer product preferences during COVID-19 worth to explore. In this section, we first present the most purchased products by category. Then, we show the products with demand boom compared to the pre-pandemic period.

In the data collection, products are organized according to 11 certain categories (namely, ‘Baby’, ‘Beverages’, ‘Bread & Bakery’, ‘Breakfast’, ‘Cleaning’, ‘Fruits & Vegetables’, ‘General Food’, ‘Meat & Seafood’, ‘Paper & Cosmetics’, ‘Snacks’, and ‘Others’). Using this categorization, we illustrate the distribution of product groups for the grocery orders placed during the pandemic in Figure 5. The most purchased product group is ‘Fruits & Vegetables’ with a rate of 31%. This category is followed by ‘General Food’ and ‘Breakfast’ with percentages of 18% and 14%, respectively. While the ‘General Food’ group includes long shelf-life products such as dried legumes, pasta, oil, and spices, the ‘Breakfast’ group contains daily foods such as cheese, milk, eggs, yoghurt, and jams. A list of most purchased products per each product group is presented in Table 2.

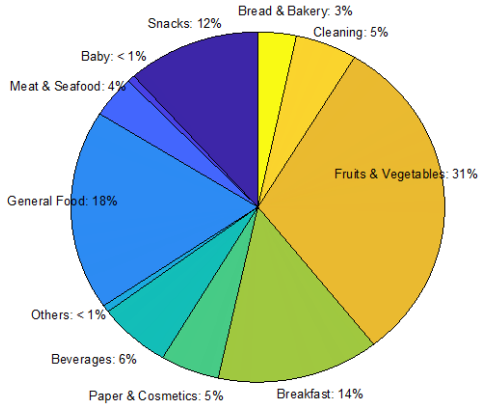


Figure 5. Distribution of product groups in grocery orders during the pandemic.

Table 2. Top-5 purchased products for each product group.

Product group	Top-5 purchased products
Baby	Diapers, baby wipes, baby biscuits, infant formula, baby shampoos.
Beverages	Cokes, tea, mineral water, coffee, bottled water.
Bread & Bakery	Bread, pastry, grissini, thin bread, desserts.
Breakfast	Cheese, milk, eggs, yoghurt, olives.
Cleaning	Bleachers, laundry detergents, kitchen & bath cleaners, food storage & thrash bags, dishwasher detergents.
Fruits & Vegetables	Carrot, banana, parsley, potato, cauliflower
General Food	Pasta, dry legumes, dry yeast, wheat flour, spices.
Meat & Seafood	Red meat, white meat, bologna, salami, sausage.
Paper & Cosmetics	Soaps, toilet papers, paper towels, wet wipes, shampoos.
Snacks	Biscuits, chocolate, dried nuts, chips, wafer.
Others	Batteries, pet food, toys, office supplies, lightbulbs.

During the outbreak, consumers have shown a demand boom for certain products. In Table 3, we list some of these items drawing attention with their degree of increase in orders. Among food products, there exists an intense demand for snacks (dried nuts, chips, crackers, chocolate, and biscuits) and ingredients to bake homemade pastries (whipped cream, dry yeast, baking powder, powdered sugar, cacao, and wheat flour). Also, in connection with baking at home, which is proposed to be a way

of coping with the pandemic by doing-it-yourself (Kirk & Rifkin, 2020), the sales of aluminum foils, baking sheets, and food storage bags have been on the rise. In terms of cleaning and hygiene products, soaps, latex gloves, colognes, cleaning towels, and cleaning vinegar have been commonly purchased.

One thing to note about the increasing sales of cologne and latex gloves is that both products were not commonly preferred in online market shopping before the pandemic, as shown in Table 3. The COVID-19 outbreak has increased sales of these products 67 times for latex gloves and 32 times for cologne. Thereby, the demand in these products can be associated with personal hygiene concerns of the consumers (Addo et al., 2020).

Table 3. Products with significantly high sales during the pandemic.

Product group	Product	Number of times product purchased		Degree of increase
		Pre-pandemic	Pandemic	
Snacks	Dried nuts	2,059	36,057	≈ 16 times
	Chips	1,946	29,216	≈ 14 times
	Crackers	949	13,613	≈ 13 times
	Chocolate	3,918	52,312	≈ 12 times
	Biscuits	4,503	57,481	≈ 12 times
General Food	Whipped cream	19	772	≈ 39 times
	Dry yeast	811	29,649	≈ 35 times
	Baking powder	38	1,127	≈ 28 times
	Powdered sugar	85	2,336	≈ 26 times
	Cacao	144	2,993	≈ 20 times
	Wheat flour	1,624	26,940	≈ 15 times
Paper & Cosmetics	Cologne	87	2,920	≈ 32 times
	Soaps	1,336	17,034	≈ 12 times
Cleaning	Latex gloves	19	1,295	≈ 67 times
	Aluminium foils and baking sheets	281	5,578	≈ 19 times
	Food storage and thrash bags	755	14,067	≈ 17 times
	Cleaning towels	172	2,652	≈ 14 times
	Cleaning vinegar	128	1,818	≈ 13 times

4.3. Changes in Shopping Hours and Delivery Preferences

In a typical online grocery shopping scenario, a consumer chooses the date and time after selecting the products she wants. The choice is made among the options - delivery windows (Agatz et al., 2011) - offered by the supermarket. This time slot booking system is directly related to the capacity of the supermarket's delivery network. In line with the online consumption trend that results in an overload in product delivery, consumers have changed their active shopping hours and delivery window preferences. Figure 6 presents a heatmap showing the intensity of active shopping hours in the collection. In the pre-pandemic period, grocery orders were usually placed between "09:00-19:59", with the highest intensity observed in "12:00-17:59". There was no tendency to shop after midnight. On the other hand, shopping continues almost every hour of the day during the pandemic. While the "12:00-17:59" interval is still very intense, shopping is done even in the early morning. Remarkably, the most frequent shopping hour is "00:00-00:59". The main reason for this difference is the desire of people to catch the closest delivery window for the next day and get the products as soon as possible.

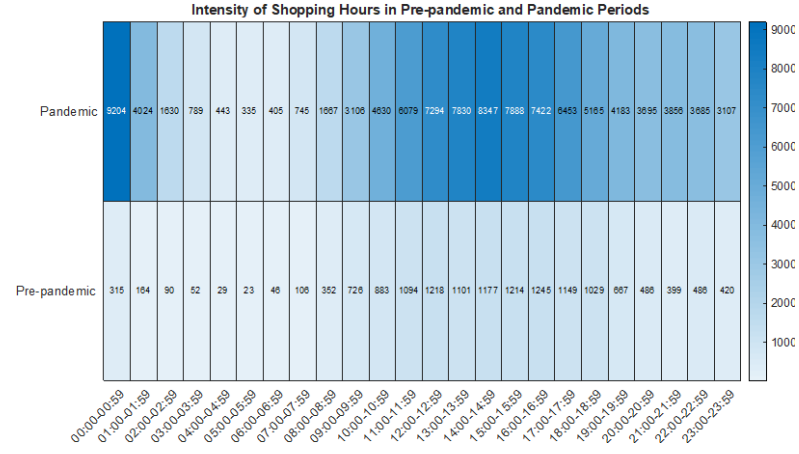


Figure 6. Shopping hour intensity before and after the pandemic.

The online shopping boom caused by COVID-19 has forced grocery chains and supermarkets to take action on the delivery issues. In order to meet the increasing demand, the grocery businesses have increased and diversified the available time slots for delivery. As shown in Table 4, the grocery chains were offering 90 time slots in total for delivery before the pandemic, and a consumer could select a delivery time for an average of 6.1 hours after the order. The grocery chains have increased available time slots by 68% to 151 during the outbreak. Despite this increase, the delivery of products has started to take longer (22.3 hours on average) when compared to the past. This delay means that consumers receive products one or two days after the order (depending on the moment of order), not within the same day. For this reason, consumers have started to prefer shopping at midnight hours (e.g., "00:00-00:59" and "01:00-01:59").

Table 4. Delivery time slot bookings before and after the pandemic.

Period	Total Grocery Orders	Available Time Slots	Average Delivery Time
Pre-pandemic	14,471	90	6.1 hours
Pandemic	101,982	151	22.3 hours

When we examine the time slots offered by the supermarkets to their consumers, we see that product delivery can be done in a wide variety of options. In Table 5, we show sample time slots for three supermarkets in the grocery sales collection. As the table presents, grocery chains may have different dynamics in product delivery in terms of both the number of sessions, preferable hours and estimated durations. This difference raises the need to normalize data when examining the effects of COVID-19 on product delivery.

Table 5. Sample delivery time slots from three different supermarkets.

Grocery A	Grocery B	Grocery C
10:30-12:30	10:00-12:00	11:00-13:00
13:00-15:00	13:00-15:00	14:00-17:00
15:00-17:00	15:00-17:00	18:00-22:00
17:00-19:00	17:00-18:30	
19:00-21:00	18:30-20:00	
21:00-23:59		

In order to represent delivery time slots on the same scale, we split each day into half-hour intervals. Then, we measure the extent to which bookings cover these intervals. For example, a grocery order requested to be delivered between “15:00-16:30” covers 3 half-hour intervals (‘15:00’, ‘15:30’, and ‘16:00’). Using the coverage of consumer bookings, we present a heatmap of the intensity of delivery hours in Figure 7. Before the COVID-19 outbreak, the most intense delivery interval used to be “17:00-19:59”, which corresponds to a period after typical working hours in Turkey. With the pandemic, this range has shifted towards “14:00-16:59”. This change can be associated with consumers staying at home due to protective measures related to self-quarantine and work at home. Conspicuously, supermarkets have started to make deliveries in the “22:00-01:59” interval, which was not much preferred before the outbreak. Increasing demand for online shopping during COVID-19 caused the products to be delivered late at night.

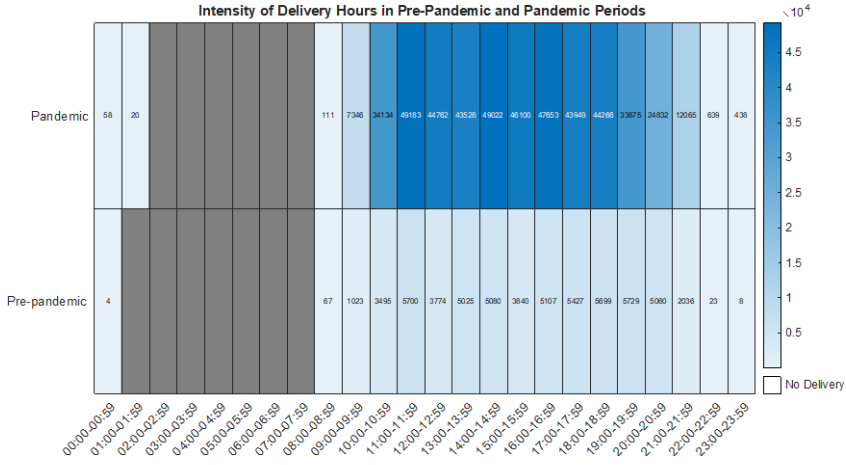


Figure 7. Delivery hour intensity before and after the pandemic.

5. DISCUSSION

Focusing on three main aspects, we conduct quantitative analyses on the impacts of the COVID-19 pandemic on Turkish consumer behaviors in grocery shopping. In this section, we provide an overview and discussion of our findings regarding the performed analyses.

While the panic and fear caused by COVID-19 had rushed people into supermarkets, many consumers have turned to online shopping platforms for their needs (Pabuççiyan, 2020). In order to illustrate this shift towards online consumption that is questioned in RQ1, we quantitatively present total grocery orders and new registrations in pre-pandemic and pandemic periods. The number of consumers registered to Marketyo has been increasing as of the first coronavirus case in Turkey, and in parallel, the number of daily orders has been rising rapidly. Additionally, our analysis on consumer demographics shows us that for every age group, more consumers prefer online shopping than before. While the most intense demand belongs to the consumers between the ages of 25 and 44, a significant increase can be observed in the

number of consumers over the age of 65, for whom the Turkish government has been imposing a curfew.

For RQ2, which questions the diversification of consumer product preferences during the pandemic, we present the categorical distribution of the products purchased after the first confirmed COVID-19 case in Turkey. Among 11 different product groups, 'Fruits & Vegetables', 'General Food', and 'Breakfast' come to the forefront as they constitute 63% of the total orders. When we deepen this analysis on a product basis, we observe that the sales of some products have increased enormously compared to the past. The presence of baking powder, dry yeast, wheat flour, and baking sheets among these products indicates that consumers tend to even make bakery and pastry food at home during the pandemic. Among cleaning products, soaps, cologne, which is rumored to be protective against coronavirus in Turkish media (DHA, 2020), and latex gloves are noteworthy. These findings are similar to the increase in personal protective equipment reported by Addo et al. (2020). Eventually, in an extreme situation like the COVID-19 outbreak, identification of situation-specific products is crucial for the continuity of product supply in the long term. Just as in the case of grocery shopping that we review, the supply of certain products may become important in other e-commerce domains.

Another important change in customer behaviors due to COVID-19 is the preference of active shopping hours and delivery windows, which is questioned in RQ3. During the pandemic, the demand for online shopping is so intense that consumers have started to shop at "00:00-00:59" hours to book earliest delivery windows and reach products quickly. Supermarkets have increased and diversified available choices (from 90 to 151 time slots) to complete deliveries shortly; deliveries were made even in the "22:00-01:59" interval that was not used much before the pandemic. Despite these efforts, it can

be said that delivery network of supermarkets cannot fully cope with the overload since the average delivery duration has increased from 6.1 hours to 22.3 hours during COVID-19. Thus, supermarkets (or their delivery partners) should review and strengthen their delivery infrastructure in order to provide better service in extraordinary situations. In addition, an important observation regarding product delivery is that the most preferred delivery window has shifted from "17:00-19:59" to "14:00-16:59" in the pandemic period. As the prior range indicates common after-hours in Turkey, it can be said that remote working policies promoted by the Turkish government have also impacts on product delivery hours.

6. CONCLUSION

Global crises, such as the COVID-19 pandemic, have profound and lasting effects across multiple domains, including health, the economy, education, and social life. One of the most affected sectors is grocery shopping, a fundamental daily necessity. This study examines how consumer behavior in grocery shopping evolved during a period of crisis, with a particular focus on the shift to online shopping, changes in product demand, and modifications in shopping and delivery preferences. Through quantitative analyses, we demonstrate the extent to which consumer demand surged, which product categories experienced the highest growth, and how shopping patterns adapted in response to external disruptions. Our findings indicate that government-imposed restrictions, panic buying, and supply chain constraints significantly influenced online grocery consumption. While consumers rapidly adapted to digital grocery platforms, the sudden surge in demand placed considerable pressure on delivery networks, revealing inefficiencies in last-mile logistics and fulfillment capacities. Beyond the immediate

effects observed during the pandemic, this study underscores the broader implications of crisis-driven consumer behavior shifts. As digital commerce continues to expand, understanding how consumers respond to uncertainty and supply chain disruptions remains essential. As e-commerce and retail logistics continue to evolve, further studies should explore how businesses and policymakers can enhance supply chain resilience and digital infrastructure to better respond to future disruptions.

ACKNOWLEDGMENTS

This study was supported by Eskişehir Technical University Scientific Research Projects Commission under grant no: 21GAP089.

REFERENCES

Addo, P. C., Jiaming, F., Kulbo, N. B., & Liangqiang, L. (2020). Covid-19: fear appeal favoring purchase behavior towards personal protective equipment. *The Service Industries Journal, 40*(7-8), 471-490.

Agatz, N., Campbell, A., Fleischmann, M., & Savelsbergh, M. (2011). Time slot management in attended home delivery. *Transportation Science, 45*(3), 435-449.

Anderson, R. M., Heesterbeek, H., Klinkenberg, D., & Hollingsworth, T. D. (2020). How will country-based mitigation measures influence the course of the COVID-19 epidemic? *The Lancet, 395*(10228), 931–934.

Aydinli, A., Lamey, L., Millet, K., ter Braak, A., & Vuegen, M. (2021). How do customers alter their basket composition when they perceive the retail store to be crowded? an empirical study. *Journal of Retailing, 97*(2), 207-216.

Baker, S. R., Farrokhnia, R. A., Meyer, S., Pagel, M., & Yannelis, C. (2020). How does household spending respond to an epidemic? Consumption during the 2020 COVID-19 pandemic. *The Review of Asset Pricing Studies, 10*(4), 834-862.

Dannenberg, P., Fuchs, M., Riedler, T., & Wiedemann, C. (2020). Digital transition by covid-19 pandemic? the german food online retail. *Tijdschrift voor economische en sociale geografie, 111*(3), 543-560.

DHA. (2020). *Toptancılarda kolonya sıkıntısı*. https://www.cnnturk.com/turkiye/ toptancilarda-kolonya-sikintisi. (Accessed: 2025-03-01)

Frost, R. O., & Gross, R. C. (1993). The hoarding of possessions. *Behaviour Research and Therapy, 31*(4), 367-381.

Kirk, C. P., & Rifkin, L. S. (2020). I'll trade you diamonds for toilet paper: Consumer reacting, coping and adapting behaviors in the covid-19 pandemic. *Journal of Business Research, 117*, 124-131.

Li, J., Hallsworth, A. G., & Coca-Stefaniak, J. A. (2020). Changing grocery shopping behaviours among chinese consumers at the outset of the covid-19 outbreak. *Tijdschrift voor economische en sociale geografie, 111*(3), 574-583.

Nicola, M., Alsafi, Z., Sohrabi, C., Kerwan, A., Al-Jabir, A., Iosifidis, C., . . . Agha, R. (2020). The socio-economic implications of the coronavirus and COVID-19 pandemic: a review. *International Journal of Surgery, 78*, 185-193.

Ozili, P. K., & Arun, T. (2020). Spillover of COVID-19: impact on the Global Economy. In *Managing inflation and supply chain disruptions in the global economy* (pp. 41-61). IGI Global.

Pabuççiyan, A. (2020). *Türkiye'de corona virüs görülmesi ile marketlerde ciro ve sipariş adedi nasıl değişti?* https://webrazzi.com/2020/03/25/corona-virus-salgini-online-market-alisverislerimizi-nasil-degistirdi/. (Accessed: 2025-03-05)

Paul, S. K., & Chowdhury, P. (2021). A production recovery plan in manufacturing supply chains for a high-demand item during COVID-19. *International Journal of Physical Distribution & Logistics Management, 51*(2), 104-125.

Petrushevska, D. (2020). *Turkish supermarket chain Migros to hire 1,000 workers amid coronavirus outbreak*. https://seenews.com/news/turkish-supermarket-chain-migros-to-hire-1000-workers-amid-coronavirus-outbreak-report-691961. (Accessed: 2025-03-04)

Richards, T. J., & Rickard, B. (2020). Covid-19 impact on fruit and vegetable markets. *Canadian Journal of Agricultural Economics/Revue canadienne d'agroeconomie, 68* (2), 189-194.

Roggeveen, A. L., & Sethuraman, R. (2020). How the covid-19 pandemic may change the world of retailing. *Journal of Retailing*, *96*(2), 169-171.

Sheth, J. (2020). Impact of covid-19 on consumer behavior: Will the old habits return or die? *Journal of Business Research, 117*, 280-283.

Stoecklin, S. B., Rolland, P., Silue, Y., Mailles, A., Campese, C., Simondon, A., ... & Levy-Bruhl, D. (2020). First cases of coronavirus disease 2019 (COVID-19) in France: surveillance, investigations and control measures, January 2020. *Eurosurveillance*, *25*(6), 2000094.

Szymkowiak, A., Gaczek, P., Jeganathan, K., & Kulawik, P. (2020). The impact of emotions on shopping behavior during epidemic. What a business can do to protect customers. *Journal of Consumer Behavior*, 2020; 1-13.

Watanabe, T., & Omori, Y. (2020). Online consumption during and after the COVID 19 pandemic: Evidence from Japan. *The impact of COVID-19 on E-commerce*, *10*, 971-978.

Zhu, N., Zhang, D., Wang, W., Li, X., Yang, B., Song, J., . . . Wenjie, T. (2020). A novel coronavirus from patients with pneumonia in China, 2019. *New England Journal of Medicine*, *382*, 727-733.

BİLGİSAYAR BİLİMLERİ VE MÜHENDİSLİĞİ KONULARI

yaz
yayınları

YAZ Yayınları
M.İhtisas OSB Mah. 4A Cad. No:3/3
İscehisar / AFYONKARAHİSAR
Tel : (0 531) 880 92 99
yazyayinlari@gmail.com • www.yazyayinlari.com