Güncel Bilimsel Analizler:
Teknik ve İstatistiksel Yaklaşımlar

Editör
Sadi ELASAN

BİDGE

BİDGE Yayınları

Güncel Bilimsel Analizler: Teknik ve İstatistiksel Yaklaşımlar

Editörler: Doç. Dr. Sadi ELASAN

ISBN: 978-625-372-230-2

1. Baskı
Sayfa Düzeni: Gözde YÜCEL
Yayınlama Tarihi: 25.06.2024
BİDGE Yayınları

Sertifika No: 71374
Yayın hakları © BİDGE Yayınları
www.bidgeyayinlari.com.tr - bidgeyayinlari@gmail.com
Krc Bilişim Ticaret ve Organizasyon Ltd. Şti.
Güzeltepe Mahallesi Abidin Daver Sokak Sefer Apartmanı No: 7/9 Çankaya / Ankara

ÖNSÖZ

Bilimsel araştırmalarda kullanılan bazı teknik ve istatistiksel yöntemlerin kapsamlı bir incelemesini sunan bu kitabın amacı, araştırmacılara ve akademisyenlere çeşitli alanlarda karşılaşılan problemleri çözmek için kullanabilecekleri bazı yöntemleri tanıtmak ve bunların uygulama örneklerini göstermektir.

Bu kitaptaki her bir bölüm, kendi alanında uzman olan yazarlar tarafından kaleme alınmış olup, okurlara hem teorik bilgiler hem de pratik uygulamalar sunmaktadır. Bu eserin, bilimsel çalışmalara katkı sağlaması ve okuyuculara faydalı olması ümit edilmektedir.

Editör
Doç. Dr. Sadi ELASAN

İÇİNDEKİLER

BÖLÜM I

Türkiye'de Havalimanlarının Etkinliğinin Panel Stokastik Sınır Analizi ile İncelenmesi

Umut AYDIN[1]

Giriş

Günümüzde havalimanları, küresel ulaşım ağlarının temel taşları olarak önemli bir role sahiptir. Hava taşımacılığı, küresel anlamda insanları, mal ve hizmetleri hızlı ve güvenli bir şekilde taşımanın yanı sıra ekonomik, sosyal ve kültürel etkileşimi artırma potansiyeliyle de büyük önem taşır. Bu bağlamda, Türkiye gibi stratejik bir konuma sahip ülkelerde havalimanlarının etkinliği, uluslararası ticaret, turizm ve ekonomik büyüme için kritik bir faktördür. Türkiye, coğrafi konumuyla Avrupa ile Asya arasında köprü görevi görmektedir. Bu nedenle, Türkiye'nin hava taşımacılığı sektörü sadece ulusal ekonomisine değil, aynı zamanda bölgesel ve

[1] Dr. Öğr. Üyesi, Bandırma Onyedi Eylül Üniversitesi, Ömer Seyfettin Uygulamalı Bilimler Fak., Uluslararası Ticaret ve Lojistik Bölümü, Balıkesir/Türkiye, Orcid: 0000-0003-4802-8793, uaydin@bandirma.edu.tr.

küresel ticaretin ve turizmin gelişimine de önemli katkılar sağlamaktadır. Türkiye'deki havalimanlarının etkinliği, ülkenin ekonomik büyüme potansiyelini artırabilir, istihdamı teşvik edebilir ve bölgesel kalkınmayı destekleyebilir.

Türkiye'deki havalimanlarının etkinliğini etkileyen faktörler arasında çeşitli ekonomik, sosyal ve altyapısal faktörler bulunmaktadır. Literatürde, bu faktörler ulaşım altyapısı, ekonomik büyüme, turizm, hava taşımacılığı politikaları, teknolojik gelişmeler ve güvenlik başlıkları altında incelenmektedir. Havalimanlarının etkinliğini belirleyen önemli bir faktör, ülkenin genel ulaşım altyapısıdır. İyi bir karayolu ve demiryolu ağı, havalimanlarına erişimi kolaylaştırarak yolcu ve yük taşımacılığını artırabilir. Bir ilin ekonomik büyümesi ise, hava taşımacılığı sektörünün gelişmesini ve havalimanlarının etkinliğini doğrudan etkileyebilir, hava yolu seyahatini teşvik edebilir ve havalimanlarının kullanımını artırabilir.

Türkiye gibi turistik bir destinasyonda, turizm sektöründeki büyüme havalimanlarının etkinliğini artırabilir. Turizm, hava taşımacılığının büyümesine ve havalimanlarının yolcu trafiğine katkıda bulunmaktadır. Diğer taraftan, hükümet politikaları, hava taşımacılığı sektörünün büyümesini ve havalimanlarının etkinliğini etkilemektedir. Özellikle düzenlemeler ve teşvikler hava taşımacılığı sektörünü şekillendirmektedir. Son olarak, hava taşımacılığındaki teknolojik gelişmeler, daha güvenli ve verimli uçuş sistemleri ve yolcu hizmetleri, havalimanlarının daha etkin bir şekilde çalışmasını sağlayabilir. Ek olarak, güvenlik önlemlerinin artırılması, havalimanlarının daha güvenli ve daha etkin bir şekilde işlemesini sağlayabilir.

Literatürde, Türkiye'deki havalimanlarının etkinliğini etkileyen faktörlerin geniş bir yelpazede incelendiği ve bu faktörlerin hava taşımacılığı sektöründeki gelişmeler üzerinde önemli bir rol oynadığı görülmektedir. Bu faktörlerin dikkatle yönetilmesi, Türkiye'nin hava taşımacılığı sektöründe daha rekabetçi bir konuma gelmesine ve havalimanlarının etkinliğinin artmasına yardımcı olabilir.

Bu çalışma ile 2015-2021 yılları arasında Türkiye'deki en fazla yolcu trafiğine sahip ilk 10 havalimanının etkinliğinin zaman içindeki değişiminin panel veri analizi ile incelenmesi amaçlanmıştır. Bu bağlamda, çalışmada panel stokastik sınır analizi (SSA) yöntemi tercih edilmiştir. Analiz sonucunda Türkiye'deki yolcu taşımacılığı bakımından önemli havalimanlarının etkinliğini artırmak için yapılması gerekenler üzerine önerilerde bulunulmuştur. Sonuç olarak, Türkiye'nin hava taşımacılığı sektöründeki potansiyelini değerlendirmek ve havalimanlarının etkinliğini artırmak için yapılması gerekenlerin belirlenmesi, ülkenin ekonomik kalkınması ve uluslararası rekabet gücü açısından büyük önem taşımaktadır. Bu çalışma, Türkiye'nin hava taşımacılığı alanındaki mevcut durumunu analiz ederek, gelecekteki stratejiler için bir çerçeve sunmayı amaçlamaktadır.

Literatür taraması

Tablo 1'de SSA ile havalimanlarının etkinliklerini inceleyen güncel çalışmalara yer verilmiştir. Tablo incelendiğinde çalışmaların bazılarının verilerinde birden fazla veri yılı kullandığından panel veri analizine dayandığını, bazılarının ise (Fageda & Voltes-Dorta, 2012; Pavlyuk, 2013; Lin & ark., 2013; Oliveira-Neto & ark., 2014; Yalçın, 2018; Matulova & Rejentova, 2021 gibi) tek bir yıl için yatay

kesit verisi düzeyinde ilgili ülkedeki havalimanlarının etkinliğini inceledikleri tespit edilmiştir. Bazı çalışmalar (Diana, 2011; Lin & ark., 2013; Matulova & Rejentova, 2021) ise parametrik olan SSA'nın etkinlik sonuçları ile parametrik olmayan yöntemlerden olan veri zarflama analizinin (VZA) etkinlik sonuçlarını karşılaştırmıştır. Bunlara ek olarak havalimanları arası komşuluk etkilerinin havalimanlarının etkinliğini nasıl incelediğini inceleyen çalışmalar da mevcuttur (Pavlyuk, 2013).

Tablo 1: Havalimanlarının etkinliğini SSA kullanarak inceleyen güncel çalışmalar.

Çalışma	Ülke	Zaman	Değişkenler	Yöntem	Sonuç
Pels & ark. (2003)	Avrupa'daki 33 havalimanı	1995-1997	Yolcu sayısı, uçuş sayısı, alan, pist sayısı, check-in sayısı, bagaj alım noktası	SSA	Havalimanlarının neredeyse yarısı teknik olarak yeterlidir.
Martín & Voltes-Dorta (2007)	Avrupa, Kuzey Amerika, Asya, Avustralya'daki 41 havalimanı	1991-2005	Uçuş sayısı, yolcu sayısı, taşınan kargo, gelir, gürültü seviyesi, sermaye, malzeme, personel, zaman, diğer sabit varlıklar	SSA	Londra'daki Luton havalimanının teknik etkinliği en yüksektir.
Oum & ark. (2008)	Dünya çapında 109 havalimanı	2001-2004	Yolcu sayısı, uçuş sayısı, havacılık dışı çıktı, pist sayısı, terminal büyüklüğü, ücret oranı, toplam havalimanı trafiği içinde kargo trafiğinin yüzdesi, uluslararası yolcu yüzdesi,	SSA	Devlet kontrolündeki kurumların sahip olduğu/işlettiği havalimanları, çoklu havalimanı pazarlarında tekli

			zaman ve bölge değişkenleri, yüzde olarak havalimanı sahipliği		havalimanı pazarlarına kıyasla önemli ölçüde daha düşük etkinliğe sahiptir.
Diana (2011)	ABD'deki 30 havalimanı	2000, 2007, 2009	Havalimanı etkinlik oranı, belirli bir çeyrek saat boyunca varış sayısı, kapıya varış gecikmesinin ortalama dakikası, kapı kalkış gecikmesinin ortalama dakikası, ortalama taksi çıkış gecikme dakikası, ortalama taksi girişi gecikme dakikası, havalimanının toplam kullanılabilir kapasitesinin kullanılan yüzdesi, hava koşulları, pist konfigürasyonları, havadaki gecikme, blok gecikmesi	SSA, VZA	Modeldeki değişkenlerin çoğu, Haziran 2000 ve 2007'deki tüm varışlar için ortalama gecikme dakikaları dışında %95 güven düzeyinde anlamlıdır.

Colombi & ark. (2011)	İtalya'daki 38 havalimanı	2005-2008	Yıllık uçak hareketi sayısı, pist sayısı, toplam alan, taşınan yolcu sayısı, check-in sayısı, bagaj alım sayısı, uçak park yeri sayısı	SSA	Uzun ve kısa dönemli etkinsizlik kaynaklarının firmaların performansları üzerindeki etkisini araştırmaya olanak tanıyan dört rastgele bileşene sahip yeni bir model tanıtılmıştır.
Scotti & ark. (2012)	İtalya'daki 38 havalimanı	2005-2008	Yıllık uçak hareketleri, yolcu hareketleri, yük sayıları, altyapılar (pist kapasitesi, saat başına izin verilen maksimum uçuş sayısı, uçak park pozisyonları, terminal yüzey alanı, check-in sayısı, bagaj sayısı, çalışan sayısı	SSA	Kamu işletmesine sahip havalimanları özel ve karma işletmeye sahip havalimanlarına göre daha etkindir.
Fageda & Voltes-Dorta (2012)	İspanya'daki 47 havalimanı	2010	İşgücü, sermaye, gelirler, yurtiçi-Schengen ve uluslararası-sınır ötesi yolcu sayısı, hava taşımacılığı hareketleri, ortalama iniş süresi, maksimum kalkış ağırlığı,	SSA	İspanya'daki küçük havalimanlarının yaşadığı önemli kayıplar büyük ölçüde gelir etkinsizliğinden kaynaklanmakta.

			metrik ton kargo, m^2 cinsinden brüt taban alanı, yolcu terminal binalarının sayısı, pist uzunluğu, kapı sayısı, check-in sayısı, zaman, tam zamanlı çalışan sayısı, hirschmann-herfindhal havayolu trafik payları endeksi, charter uçuşlarının payı, düşük maliyetli uçuşların payı		
Pavlyuk (2013)	Avrupa'daki 112 havalimanı	2009	Taşınan yolcu sayısı (gidiş/geliş), pist sayısı, check-in sayısı	Mekânsal SSA	Havalimanı etkinlik seviyelerinde pozitif mekânsal ilişki vardır.
Lin & ark. (2013)	ABD ve Kanada'daki 62 havalimanı	2006	Yolcu sayısı, uçuş sayısı, havacılık dışı gelir, çalışan sayısı, yumuşak maliyetli girdi, tıkanıklık gecikmesi, havalimanı çıktı ölçeği, ortalama uçak büyüklüğü, uluslararası trafik yüzdesi, hava kargo yüzdesi,	SSA, VZA	Havacılık dışı gelirler, yolcu hacmi, ortalama uçak büyüklüğü, uluslararası ve aktarmalı trafik yüzdeleri havalimanı etkinlik tahminini etkilemektedir.

			transit yolcu yüzdesi, hub taşıyıcı pazar payı		
Oliveira-Neto & ark. (2014)	Brezilya'daki 63 havalimanı	2006	Uçak sayısı, yolcu sayısı, kargo tonu, toplam havalimanı alanı, uçak park alanı, yolcu terminali kapasitesi, yolcu terminali alanı, yolcu park yeri kapasitesi	SSA	Brezilya havalimanları yüksek teknik etkinliğe sahiptir.
Nwaogbe & ark. (2018)	Nijerya'daki 4 havalimanı	2005-2015	Yolcu hareketi, uçuş sayısı, çalışan sayısı, toplam varlık, toplam maliyet	SSA	Nijerya havalimanının ortalama etkinlik puanı %64'tür. Söz konusu havalimanlarında etkinlik düzeyi iyileştirilebilir.
Yalçın (2018)	Türkiye'deki 55 havalimanı	2013, 2014, 2015	Yolcu sayısı, ticari uçak sayısı, personel sayısı, check-in sayısı	SSA	Analiz edilen tüm yıllarda uçak sayısı ve personel sayısının parametre katsayıları istatistiksel olarak anlamlıdır. İncelenen tüm

					yıllarda Isparta Süleyman Demirel Havalimanı ilk sıradadır.
Zhang & ark. (2019)	Çin'deki Guangzhou havalimanı	2012-2017	Guangzhou'nun yıllık havalimanı gayrisafi yurtiçi hasılası, çalışan sayısı, işletme maaliyeti, havalimanı altyapı inşaatı, bölgesel coğrafi konum, bölgesel ekonomik gelişmişlik düzeyi, bölgesel bilim ve teknoloji eğitim düzeyi	SSA	Havalimanı altyapısının inşası, bölgenin coğrafi konumu, havalimanı altyapısının seviyesi, bölgesel ekonomik kalkınma ve bölgesel bilim ve teknoloji eğitiminin Guangzhou Havalimanı'nın ekonomik etkinliğini etkilemektedir.
Nwaogbe & ark. (2021)	Nijerya'daki 6 havalimanı	2011-2018	Yolcu sayısı, uçuş sayısı, terminal kapasitesi, pist boyutu, toplam operasyon maliyeti, yer hizmetleri ekipmanı, çalışan sayısı	SSA	Üç havalimanı %50 etkinlik seviyesi ve üzerindeyken, geri kalan havalimanları %50 etkinlik seviyesinin altında

					faaliyet göstermektedir.
Matulova & Rejentova (2021)	Avrupa'daki 115 havalimanı	2018	Yolcu sayısı, uçuş sayısı, taşınan kargo, terminal sayısı, pist sayısı, kapı sayısı, park yeri sayısı, uçak park yeri sayısı	SSA, VZA	Yoğun havalimanları arasında London Heathrow havalimanının teknik etkinliği en yüksektir.
Ripoll-Zarraga (2023)	İspanya'daki 49 havalimanı	2009-2013	Yolcu sayısı, uçuş sayısı, kargo tonu, işgücü geliri, ticari gelir, işletme geliri, amortisman hava tarafı, amortisman kara tarafı	SSA	Havalimanlarının altyapı kapasitesinin ve trafiğinin birbiriyle ilişkili değildir. Bölgesel havalimanları sistem için mali bir yük haline gelmiştir.

Ulusal literatüre yönelik çalışmalardan Yalçın (2018) üç yıl için analizleri yatay kesit olarak ayrı ayrı çalıştırmış olup havalimanlarının etkinliklerindeki değişimleri incelemiştir. Tablo 1'deki çalışmalar incelendiğinde Türkiye'de havalimanlarının etkinliğinin incelenmesinde parametrik yöntemlerden olan SSA'nın panel veri yaklaşımı ile incelenmesinde eksiklik olduğu görülmektedir. Literatürdeki diğer çalışmalar incelendiğinde de genellikle havalimanı etkinlik analizlerinin parametrik olmayan yöntemlerden olan VZA ile yapıldığı görülmektedir (Kıyıldı & Karaşahin, 2006; Peker & Birdoğan, 2009; Yazgan & Karkacier, 2015; Bolat & ark., 2016 vb). Bu nedenle bu çalışma, Türkiye'deki havalimanlarının etkinliğindeki yıllara göre değişimi ve havalimanlarının etkinliğini etkileyen faktörleri panel SSA yaklaşımı ile bulması nedeniyle diğer çalışmalardan farklılaşmaktadır.

Veri seti ve yöntem

Çalışmada bağımlı değişken olarak kullanılan taşınan yolcu verisi ve bağımsız değişkenlerden olan ticari uçuş sayısı verisi Devlet Hava Meydanları İşletmesi'nin (DHMİ) internet sitesinden alınmış olup, diğer bağımsız değişkenler olan havalimanının bulunduğu illere yönelik kişi başına düşen gayri safi yurtiçi hasıla (GSYİH) verisi (2009 bazlı) Türkiye İstatistik Kurumu'ndan (TÜİK) elde edilmiştir. Havalimanlarının operasyonel özelliklerini yansıtan pist sayısı, check-in noktası sayısı, yolcu kapısı sayısı, araç park yeri sayısı ve çalışan sayısı gibi değişkenler ise Eurostat veri tabanından elde edilmiştir. Eurostat veritabanından elde edilen veriler Türkiye'deki 10 havalimanı için 2015-2021 yılları arasında tam olarak sağlanabildiğinden, diğer bir deyişle veri bulunabilirliği kısıdı

nedeniyle çalışma 2015-2021 veri döneminde 10 (sırasıyla Adana, Ankara Esenboğa, Antalya, Gaziantep, Muğla Dalaman, Muğla Milas-Bodrum, Trabzon, 2018 ve öncesi İstanbul Atatürk, 2018 sonrası İstanbul, İstanbul Sabiha Gökçen, İzmir Adnan Menderes) havalimanını kapsamaktadır. İlgili 10 havalimanına yönelik istatistikler incelendiğinde bu havalimanlarının yıllara göre Türkiye'de toplam taşınan yolcu sayısının yaklaşık %90'ını taşıdığı görülmüştür. Değişkenlere göre tanımlayıcı istatistiklere Tablo 2'de yer verilmiştir.

Tablo 2: Tanımlayıcı istatistikler

Değişken	**Ortalama**	**Std. Sapma**	**Min.**	**Maks.**
Taşınan yolcu sayısı	14787089.0	17265363.5	1390784	68441989
Ticari uçuş sayısı	100025.6	117611.4	13114	448124
GSYİH	54759.3	26465.8	21477	140864
Pist sayısı	1.7	0.8	1	3
Check-in noktası sayısı	150.3	147.5	19	635
Yolcu kapısı sayısı	37.3	38.8	6	172
Araç park yeri sayısı	3808.8	7081.0	328	40000
Çalışan sayısı	807.1	1369.2	213	7182

N=70, n=10, t=7

Tablo 2 incelendiğinde minimum ve maksimum değerlerden taşınan yolcu sayısının farklılaştığı görülmektedir. Tablodan ayrıca, ticari uçuş sayısının 10 havalimanında 7 yıl için ortalamasının yaklaşık 100 bin civarında olduğu, pist sayısının genelde 1 veya 2 olduğu, ortalama check-in noktası sayısının 150 civarında olduğu, ortalama yolcu kapısı sayısının 37 olduğu, ortalama çalışan sayısının

ise 807 civarında olduğu anlaşılmaktadır. Kişi başına GSYİH verisi de yine iller arasında farklılık gösteren değişkenlerdendir.

Stokastik sınır modelleri Aigner & ark. (1977) ve Meeusen & Van den Broeck (1977) tarafından tanıtılmış olup literatürde popüler bir modelleme yaklaşımı haline gelmiştir. Stokastik sınır modelleri etkinsizlik teriminin farklı özelliklerine göre üretim ve maliyet fonksiyonları olmak üzere iki farklı stokastik sınır modeli türüne uygulanabilmektedir.

Bir üreticinin $f(z_{it},\beta)$ üretim fonksiyonuna sahip olduğu varsayılırsa hata veya etkinsizliğin olmadığı durumda bir firma t zamanında $q_{it}=f(z_{it},\beta)$ kadar üretim yapacaktır. Ancak, SSA'nın temel unsuru her bir firmanın, bir dereceye kadar etkinsizlik nedeniyle potansiyel olarak üretebileceğinden daha az üretebileceğini varsaymasıdır. Bu durumda $qit=f(z_{it},\beta)*\xi_{it}$ olarak tanımlanmaktadır. Burada ξ_{it}, i firmasının t zamanındaki etkinlik düzeyi olup (0,1] aralığında dağılmaktadır. Eğer $\xi_{it}=1$ ise, firma $f(z_{it},\beta)$ üretim fonksiyonunda somutlaşan teknoloji ile optimal çıktıya ulaşmaktadır. $\xi_{it}<1$ olduğunda, firma $f(z_{it},\beta)$ üretim fonksiyonunda yer alan teknoloji ile zit girdilerinden en iyi şekilde yararlanamamaktadır. Çıktının kesinlikle pozitif olduğu varsayıldığından ($q_{it}>0$) teknik etkinlik derecesinin kesinlikle pozitif olduğu varsayılmaktadır ($\xi_{it}>0$).

Ek olarak çıktının rassal şoklara maruz kaldığı varsayılmaktadır, bu durumda $q_{it} = f(z_{it},\beta)*\xi_{it}*\exp(v_{it})$ olmaktadır. Her iki tarafın doğal logaritması alındığında $\ln(q_{it})=\ln f(z_{it},\beta)+\ln(\xi_{it})+v_{it}$ elde edilmektedir. K girdi olduğu ve üretim

fonksiyonunun logaritmik olarak doğrusal olduğu varsayılırsa, u_{it}=-ln(ξ_{it}) tanımlaması,

$$\ln(q_{it})=\beta_0+\sum_{j=1}^{k}\beta_j \ln(z_{jit})+v_{it}-u_{it} \quad (1)$$

denklemini vermektedir. Burada $u_{it}\geq 0$ kısıtı, yukarıda da belirtildiği gibi $0 < \xi_{it} \leq 1$ anlamına gelmektedir.

Daha önce de belirtildiği gibi, bir stokastik sınır modelindeki hata terimin iki bileşeni olduğu, bileşenlerden birinin kesinlikle negatif olmayan bir dağılıma sahip olduğu, diğer bileşenin ise simetrik bir dağılıma sahip olduğu varsayılmaktadır. Literatürde, negatif olmayan bileşen genellikle etkinsizlik terimi, simetrik dağılıma sahip bileşen ise kendine özgü hata terimi olarak adlandırılmaktadır. Denklem (1), v_{it}'nin kendine özgü hata ve u_{it}'nin zamanla değişen panel düzeyinde bir etki olduğu panel veri modelinin bir çeşididir. Bu modelle ilgili literatürün çoğu, u_{it} teriminin farklı özellikleri için tahminciler türetmeye odaklanmıştır. Kumbhakar ve Lovell (2000) bu literatürün bir incelemesini sunmaktadır. Bu çalışmada ise etkinsizlik teriminin zamana göre değiştiği varsayılmıştır. $N+(\mu,\sigma^2)$ ortalaması sıfır ve varyansı σ^2 olan kesilmiş (truncated) normal dağılım olduğu ve etkinsizlik terimi zaman etkilerini içerdiği durumda, zamanın belirli bir fonksiyonu ile çarpılan kesilmiş normal bir rassal değişken olarak modellenmektedir (Battese ‐ Coelli, 1992). Zamanla değişen spesifikasyonunda $u_{it}=\exp(-\eta (t-T_i))*u_i$ olmaktadır. Burada T_i birinci paneldeki son dönem, η ise bozunma parametresidir. $u_i \sim N+(\mu,\sigma^2_u)$, $v_{it} \sim N(0,\sigma^2_v)$ ve u_i ve v_{it} birbirlerinden ve modeldeki ortak değişkenlerden bağımsız olarak dağılmaktadır.

Analiz sonuçları

SSA için Denklem 1'de bağımlı değişken olarak ilgili havalimanında ilgili yılda taşınan yolcu sayısı kullanılmış olup bağımsız değişkenler olarak sırasıyla ilgili havalimanında ilgili yıldaki ticari uçuş sayısı, ilgili ilin ilgili yıldaki kişi başına GSYİH'sı (2009 bazlı), havalimanının pist sayısı, check-in noktası sayısı, yolcu kapısı sayısı, araç park yeri sayısı ve çalışan sayısı bağımsız değişken olarak kullanılmıştır. Model çalıştırıldıktan sonra GSYİH, check-in noktası sayısı, araç park yeri sayısı ve çalışan sayısı istatistiksel olarak anlamsız edildiği için modelden çıkarılmıştır. Nihai modelde elde edilen sonuçlar Tablo 3'te sunulmuştur.

Modelde sigma2 toplam hata teriminin varyansını, lnsigma2 ise doğal logaritmasını göstermektedir. Rassal hata teriminin (v_{it}) varyansı sigma_v2 ile ve teknik etkinsizlik hata teriminin (u_{it}) varyansı sigma_u2 ile gösterilmektedir. Gamma katsayısı sigma_u2/sigma2'yi göstermektedir. Bu değer ne kadar büyükse, hata terimindeki varyansın o kadar büyük bir kısmı etkinsizlik terimi tarafından açıklanmakta anlamına gelmektedir. 0.80'lik bir açıklama oranı ise yeterince yüksektir. Lgtgamma, gamma katsayısının logit değerini göstermektedir. Eta (η), etkinliğin/etkinsizliğin zaman içinde değişip değişmediği hakkında bilgi vermektedir (Barros, 2005). Modelde Eta (η) katsayısının istatistiksel olarak anlamlı olması, havalimanlarının etkinliklerinin veya etkinsizliklerinin yıllar içinde değiştiğini göstermektedir. Tablo 3'te ticari uçuş sayısındaki %1'lik artışın taşınan yolcu sayısını ortalama olarak %0.82, yolcu kapısı sayısındaki %1'lik artışın taşınan yolcu sayısını ortalama olarak % 0.10, pist sayısındaki %1'lik artışın yolcu sayısını %0.08 arttırdığı elde edilmiştir ($p<0.10$) (ceteris paribus). Zamana göre

incelendiğinde ise taşınan yolcu sayısının 2015 yılına kıyasla 2017, 2018, 2019 yılında arttığı gözlemlenmektedir. 2020 yılında ise pandemi etkisiyle bir azalma görülmüştür.

Tablo 3: Panel SSA sonuçları.

ln(Yolcu_sayısı)	**Katsayı**	**Std. Hata**	**Z**	**Anlamlılık**
ln(Ticari_uçuş_sayısı)	0.852	0.040	21.480	0.000
ln(Yolcu_kapısı_sayısı)	0.101	0.052	1.960	0.050
ln(Calısan_sayısı)	0.034	0.048	0.700	0.481
ln(Pist_sayısı)	0.084	0.048	1.740	0.082
Zaman				
2016	-0.001	0.034	-0.040	0.967
2017	0.091	0.036	2.500	0.012
2018	0.134	0.040	3.300	0.001
2019	0.163	0.048	3.360	0.001
2020	-0.302	0.066	-4.580	0.000
2021	0.063	0.082	0.770	0.440
Sabit	6.159	0.439	14.020	0.000
/mu	0.230	0.141	1.630	0.103
/eta	-0.166	0.061	-2.730	0.006
/lnsigma2	-3.546	0.677	-5.240	0.000
/lgtgamma	1.424	0.884	1.610	0.107
sigma2	0.029	0.020		
gamma	0.806	0.138		
sigma_u2	0.023	0.020		
sigma_v2	0.006	0.001		

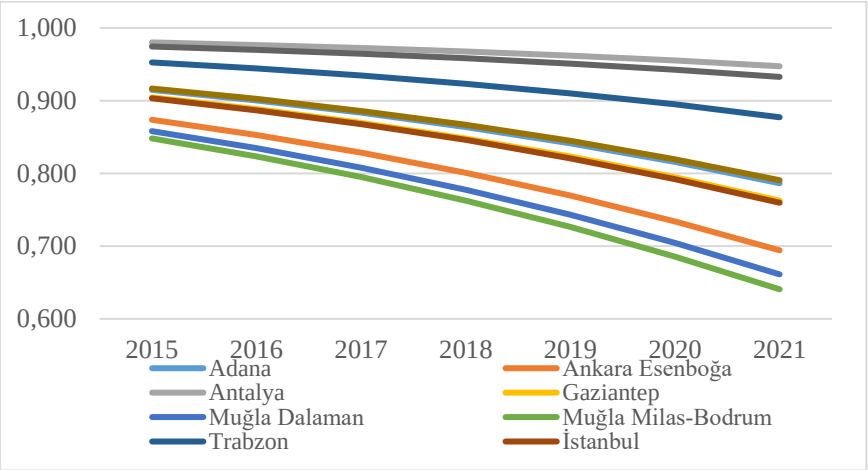
Tablo 4: Havalimanlarının yıllara göre teknik etkinlik değerleri.

Havalimanları	Zaman	Etkinlik	Havalimanları	Zaman	Etkinlik
Adana	2015	0.915	Muğla Milas-Bodrum	2015	0.848
Adana	2016	0.900	Muğla Milas-Bodrum	2016	0.823
Adana	2017	0.883	Muğla Milas-Bodrum	2017	0.795
Adana	2018	0.864	Muğla Milas-Bodrum	2018	0.763
Adana	2019	0.841	Muğla Milas-Bodrum	2019	0.726
Adana	2020	0.816	Muğla Milas-Bodrum	2020	0.686
Adana	2021	0.786	Muğla Milas-Bodrum	2021	0.641
Ankara Esenboğa	2015	0.874	Trabzon	2015	0.953
Ankara Esenboğa	2016	0.853	Trabzon	2016	0.944
Ankara Esenboğa	2017	0.828	Trabzon	2017	0.935
Ankara Esenboğa	2018	0.801	Trabzon	2018	0.923
Ankara Esenboğa	2019	0.769	Trabzon	2019	0.910
Ankara Esenboğa	2020	0.734	Trabzon	2020	0.895
Ankara Esenboğa	2021	0.694	Trabzon	2021	0.877
Antalya	2015	0.980	İstanbul	2015	0.903
Antalya	2016	0.977	İstanbul	2016	0.887
Antalya	2017	0.972	İstanbul	2017	0.868
Antalya	2018	0.967	İstanbul	2018	0.846

Antalya	2019	0.962	İstanbul	2019	0.821
Antalya	2020	0.955	İstanbul	2020	0.792
Antalya	2021	0.947	İstanbul	2021	0.760
Gaziantep	2015	0.904	İstanbul Sabiha Gökçen	2015	0.974
Gaziantep	2016	0.888	İstanbul Sabiha Gökçen	2016	0.970
Gaziantep	2017	0.869	İstanbul Sabiha Gökçen	2017	0.965
Gaziantep	2018	0.848	İstanbul Sabiha Gökçen	2018	0.958
Gaziantep	2019	0.823	İstanbul Sabiha Gökçen	2019	0.951
Gaziantep	2020	0.795	İstanbul Sabiha Gökçen	2020	0.943
Gaziantep	2021	0.763	İstanbul Sabiha Gökçen	2021	0.933
Muğla Dalaman	2015	0.858	İzmir Adnan Menderes	2015	0.917
Muğla Dalaman	2016	0.835	İzmir Adnan Menderes	2016	0.902
Muğla Dalaman	2017	0.808	İzmir Adnan Menderes	2017	0.886
Muğla Dalaman	2018	0.777	İzmir Adnan Menderes	2018	0.867
Muğla Dalaman	2019	0.743	İzmir Adnan Menderes	2019	0.845
Muğla Dalaman	2020	0.704	İzmir Adnan Menderes	2020	0.819
Muğla Dalaman	2021	0.661	İzmir Adnan Menderes	2021	0.791

Panel tabanlı SSA'dan elde edilen teknik etkinlik skorları Tablo 4'te sunulmuştur. Bu tablo şu sonuçları göstermektedir: 1) Zamansal açıdan bakıldığında, havalimanlarının ortalama etkinliği 0.856'dır. Etkinlik değerleri, Eta katsayısının da gösterdiği gibi yıldan yıla farklılık göstermekte ve yıllara göre azalmaktadır. Şekil 1'de bu durum daha net görülmektedir. Başlangıçta havalimanlarının etkinlik değerleri birbirine yakınken 2021 yılının sonunda etkinlik değerlerinin birbirinden farklılaştığı görülmektedir. 2) Havalimanlarının yıllara göre etkinlik ortalaması alındığında ortalama değerin altında Adana, Ankara Esenboğa, Gaziantep, Muğla Dalaman, Muğla Milas-Bodrum ve İstanbul havalimanlarının bulunduğu tespit edilmiştir. G

enel olarak Antalya, İstanbul Sabiha Gökçen, Trabzon ve İzmir Adnan Menderes havalimanları diğer havalimanlarına göre daha etkin bulunmuştur. 3) Tablo 5'ten anlaşıldığı üzere etkinlik değerleri 2015 yılından 2021 yılına en hızlı değişen havalimanları Muğla-Dalaman ve Muğla Milas-Bodrum havalimanları olarak elde edilmiştir. Yıl bazında ise etkinliklerde en çok azalma 2020 ve 2021 yıllarında görülmektedir. Bu durum pandeminin güçlü etkisini doğrulamaktadır.



Grafik 1: Havalimanı etkinlik değerlerinin yıllara göre değişimi

Tablo 5: Havalimanlarının teknik etkinliklerinin bir önceki yıla göre % değişimi.

	2016	**2017**	**2018**	**2019**	**2020**	**2021**
Adana	-1.6	-1.9	-2.2	-2.6	-3.1	-3.6
Ankara Esenboğa	-2.4	-2.8	-3.3	-3.9	-4.6	-5.4
Antalya	-0.4	-0.4	-0.5	-0.6	-0.7	-0.8
Gaziantep	-1.8	-2.1	-2.5	-2.9	-3.4	-4.0
Muğla Dalaman	-2.7	-3.2	-3.8	-4.4	-5.2	-6.1
Muğla Milas-Bodrum	-2.9	-3.4	-4.1	-4.8	-5.6	-6.6
Trabzon	-0.9	-1.0	-1.2	-1.4	-1.7	-2.0
İstanbul	-1.8	-2.1	-2.5	-3.0	-3.5	-4.1
İstanbul Sabiha Gökçen	-0.5	-0.5	-0.6	-0.8	-0.9	-1.1
İzmir Adnan Menderes	-1.6	-1.8	-2.2	-2.5	-3.0	-3.5

Sonuç

Bu araştırma, Türkiye'nin hava taşımacılığı sektöründe önemli bir rol oynayan havalimanlarının etkinliğini incelemek için panel SSA kullanarak yapılmıştır. Çalışmanın odak noktası, havalimanlarının taşınan yolcu sayısını etkileyen faktörleri ve havalimanlarının zaman içindeki etkinlik değişimlerini incelemektedir. Sonuçlar, ticari uçuş sayısındaki artışın taşınan yolcu sayısını olumlu yönde etkilediğini, aynı şekilde yolcu kapısı sayısı ve pist sayısındaki artışların da benzer etkiler gösterdiğini göstermektedir. Bununla birlikte, çalışma kişi başına düşen GSYİH, check-in noktası sayısı, araç park yeri sayısı ve çalışan sayısı gibi bazı faktörlerin istatistiksel olarak anlamsız olduğunu tespit etmiştir. Çalışmada ayrıca, havalimanlarının etkinliklerinin zaman içinde değiştiği ve pandemi gibi beklenmedik olayların bu değişimleri nasıl etkileyebileceği de gözlemlenmiştir. Özellikle, 2020 yılında pandemi etkisinin yolcu sayısında ciddi bir düşüşe neden olduğu ve havalimanlarının etkinliklerinde de belirgin bir azalmaya yol açtığı görülmüştür. Bu durum, hava taşımacılığı sektöründe beklenmedik olaylara karşı hazırlıklı olmanın ve kriz durumlarında etkin bir şekilde müdahale etmenin önemini vurgulamaktadır.

Çalışma bazı havalimanlarının diğerlerine göre daha etkin olduğunu ortaya koymaktadır. Bu durum, hava taşımacılığı sektöründe rekabetin arttığı ve havalimanlarının daha etkin olmaları için stratejik planlamalar yapılması gerektiğini göstermektedir. Özellikle, Antalya, İstanbul Sabiha Gökçen, Trabzon ve İzmir Adnan Menderes havalimanları gibi havalimanlarının diğerlerine göre daha etkin olduğu belirlenmiştir. Sonuçlar, havalimanlarının

etkinliklerini artırmak için altyapı yatırımlarının ve havayolu işbirliklerinin önemini vurgulamaktadır.

Sonuç olarak, bu çalışma Türkiye'nin hava taşımacılığı sektöründe havalimanlarının etkinliğini anlamak ve geliştirmek için önemli sonuçlar sunmaktadır. Elde edilen bulgular, hava taşımacılığı sektöründe stratejik kararlar alırken havalimanlarının etkinliğinin göz önünde bulundurulmasının önemini vurgulamaktadır. Ayrıca, pandemi gibi beklenmedik olaylara karşı hazırlıklı olmanın ve kriz durumlarında etkin bir şekilde müdahale etmenin önemini de ortaya koymaktadır. Bu nedenle, Türkiye'nin hava taşımacılığı sektöründe havalimanlarının etkinliğini artırmak için altyapı yatırımları, havayolu işbirlikleri ve teknolojik yenilikler gibi politikaların uygulanması önemlidir. Bu politikaların başarısı, Türkiye'nin hava taşımacılığı sektöründe daha rekabetçi ve etkin bir konuma gelmesine yardımcı olabilir.

Kaynaklar

Aigner, D., Lovell, C. K., & Schmidt, P. (1977). Formulation and estimation of stochastic frontier production function models, *Journal of Econometrics*, *6*(1), 21-37.

Barros, C. P. (2005). Performance measurement in tax offices with a stochastic frontier model, *Journal of Economic Studies*, *32*(6), 497-510.

Battese, G. E., & Coelli, T. J. (1992). Frontier production functions, technical efficiency and panel data: with application to paddy farmers in India, *Journal of Productivity Analysis*, 3, 153-169.

Bolat, B., Temur, G. T., & Gürler, H. (2016). Türkiye'deki havalimanlarının etkinlik tahmini: Veri zarflama analizi ve yapay sınır ağlarının birlikte kullanımı, *Ege Academic Review*, 16.

Colombi, R., Martini, G., & Vittadini, G. (2011). A stochastic frontier model with short-run and long-run inefficiency random effects. *Working Papers 1101*, Department of Management, Information and Production Engineering, University of Bergamo.

Diana, T. (2011). Measuring and benchmarking airport efficiency: An application of data envelopment analysis (DEA) and stochastic frontier analysis (SF). Connor R. Walsh (Ed.), In *Airline Industry: Strategies, Operations and Safety*, (pp. 143-159), New York, USA.

Fageda, X., & Dorta, A. V. (2012). Efficiency and profitability of Spanish airports: a composite non-standard profit function approach, *Working Paper*, Universitat de Barcelona. Spain, 1-21.

Kıyıldı, R., & Karaşahin, M. (2006). Türkiye'deki havaalanlarının veri zarflama analizi ile altyapı performansının değerlendirilmesi, *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, *10*(3), 391-397.

Kumbhakar, S. C., & Lovell, C. K. (2003). *Stochastic frontier analysis*. Cambridge University Press.

Lin, Z., Choo, Y. Y., & Oum, T. H. (2013). Efficiency benchmarking of North American airports: Comparative results of productivity index, data envelopment analysis and stochastic frontier analysis, *Journal of the Transportation Research Forum*, *52*(1), 47-67.

Martin, J. C., & Voltes-Dorta, A. (2007). Marginal cost estimations in airports. Multiproductive cost functions and stochastic frontiers: An international airports case study, *48th Annual Transportation Research Forum 2007,* 15-17 March 2007, Boston, Massachusetts, USA.

Matulová, M., & Rejentová, J. (2021). Efficiency of European airports: Parametric versus non-parametric approach, *Croatian Operational Research Review*, 1-14.

Meeusen, W., & van Den Broeck, J. (1977). Efficiency estimation from Cobb-Douglas production functions with composed error, *International Economic Review*, 435-444.

Nwaogbe, O. R., Ayinla, A. I., Omoke, V., Ojekunle, J. A., & Wokili-Yakubu, H. (2021). Analysis of airport operational performance in selected airports of Northern Nigeria, *LOGI–Scientific Journal on Transport and Logistics*, *12*(1), 111-122.

Nwaogbe, O. R., Ejem, E. A., Ogwude, I. C., Idoko, F. O., & Pius, A. (2018). Performance appraisal of Nigerian airports: Stochastic frontier analysis, *Transport & Logistics*, *18*(44).

Oliveira-Neto, F., Pontes, P., & Wichmann, B. (2014). A benchmarking analysis of the Brazilian airport infrastructure an application of the expected minimum input frontier. *Journal of Transport Economics and Policy*, *48*(2), 297-313.

Oum, T. H., Yan, J., & Yu, C. (2008). Ownership forms matter for airport efficiency: A stochastic frontier investigation of worldwide airports, *Journal of Urban Eeconomics*, *64*(2), 422-435.

Pavlyuk, D. (2013). Distinguishing between spatial heterogeneity and inefficiency: Spatial stochastic frontier analysis of European airports. *Transport and Telecommunication Journal*, *14*(1), 29-38.

Peker, İ., & Birdoğan, B. (2009). Veri zarflama analizi ile Türkiye havalimanlarında bir etkinlik ölçümü uygulaması, *Çukurova Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *18*(2), 72-88.

Pels, E., Nijkamp, P., & Rietveld, P. (2003). Inefficiencies and scale economies of European airport operations, *Transportation Research Part E: Logistics and Transportation Review*, *39*(5), 341-361.

Ripoll-Zarraga, A. E. (2023). Airports' public infrastructure and sources of inefficiency, *Journal of Economics, Finance and Administrative Science*, *28*(55), 176-196.

Scotti, D., Malighetti, P., Martini, G., & Volta, N. (2012). The impact of airport competition on technical efficiency: A stochastic frontier analysis applied to Italian airport, *Journal of Air Transport Management*, 22, 9-15.

Yalçın, E. (2018). Stokastik sınır analizi ile havalimanlarının etkinliklerinin ölçülmesi: Türkiye örneği, *Mustafa Kemal Üniversitesi Sosyal Bilimler Enstitüsü Dergisi*, *15*(42), 82-105.

Yazgan, A., & Karkacier, O. (2015). Veri zarflama analizi ile etkinlik ölçümleri ve havalimanı işletmeciliği sektöründe bir uygulama, *Uluslararası Alanya İşletme Fakültesi Dergisi*, *7*(2).

Zhang, Z., Yang, J., Wan, Y., Pan, R., & Kuang, M. (2019). Economic efficiency measurement of Guangzhou airport based on stochastic frontier analysis model. *1st International Conference on Business, Economics, Management Science (BEMS 2019)*, 20-21 April 2019, Hangzhou, China, (pp. 649-656). Atlantis Press.

BÖLÜM II

Evaluating Resolution Reduction Techniques in Convolutional Neural Networks: A Comparison with Max-Pooling

Emre GÜNGÖR[1]

INTRODUCTION

Convolutional neural networks (CNNs) have revolutionized the field of computer vision, enabling unprecedented advancements in image recognition, classification, and segmentation tasks. In the field of CNNs, the efficient reduction of input resolution is a critical step that directly impacts both computational efficiency and model performance. Among the various techniques employed for this purpose, max-pooling has emerged as a widely adopted method due to its simplicity and effectiveness in reducing spatial dimensions while preserving salient features. However, alternative downscaling

[1] Assistant Prof. Emre GÜNGÖR, Kütahya Health Sciences University, Faculty of Engineering and Nature Sciences, Computer Engineering Department, Kütahya/TÜRKİYE, Orcid: 0000-0003-4278-6294, emre.gungor@ksbu.edu.tr

methods such as nearest neighbor downscaling and Gaussian downscaling (also known as pyramid downsampling) offer distinct approaches that could potentially enhance or degrade from the performance of CNNs in various applications.

This study aims to provide a comprehensive comparison of these resolution reduction methods and examine their impact on the performance of convolutional neural networks. Specifically, we will evaluate the effectiveness of max-pooling, nearest neighbor downscaling, and Gaussian downscaling against a baseline where no downscaling is applied. Through experiments and analysis, we seek to uncover the strengths and weaknesses of each method, thereby providing insights into their practical usage and guiding the selection of appropriate techniques for better generalized performance.

Understanding the nuances of these downscaling methods is essential for optimizing CNN architectures, particularly in scenarios where preserving important features and minimizing information loss are paramount. Through this comparative study, we aim to contribute to the field of CNN optimization, providing valuable guidance for researchers and practitioners seeking to enhance the efficiency and accuracy of their models.

In the following sections, we systematically explore the various aspects of our study. Section 2 provides a comprehensive literature review, summarizing related studies to establish the theoretical background and identify application areas for resolution reduction methods in CNNs. Section 3 details the methodology, describing the creation of our models and the network architecture, along with the implementation of the different resolution reduction

methods. In Section 4, we present the results and discussion, analyzing the impact of image reduction techniques on CNN performance. This section includes a thorough examination of training outcomes, training times, and memory requirements for each method. Finally, Section 5 concludes with a summary of the study's key findings and their implications.

LITERATURE REVIEW

The development and optimization of convolutional neural networks (CNNs) have been profoundly influenced by various resolution reduction methods. Among these, pooling methods stands out as a foundational technique widely used to decrease the spatial dimensions of feature maps, thereby reducing computational load and mitigating overfitting. Introduced by LeCun et al. in the 1990s, pooling (subsampling) has since become a staple in CNN architectures, as demonstrated in seminal works such as the AlexNet and VGG networks (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014).

Max-Pooling

Pooling has been a cornerstone in the architecture of CNNs since its introduction. LeCun et al. (1998) with LeNet-5 architecture were among the first to implement pooling (average-pooling) in the context of image recognition tasks. Average-pooling takes an average of block region whereas max-pooling assumes features lies in the maximum pixel value in the region. So, the technique of max-pooling involves partitioning the input image into a set of non-overlapping rectangles and, for each such sub-region, outputting the maximum value. This approach not only reduces the spatial

dimensions but also retains the most salient features, thereby helping in translational invariance (Ranzato et al.,2007).

Krizhevsky et al. (2012) popularized max-pooling in their groundbreaking work on AlexNet, which won the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) 2012. They demonstrated that max-pooling layers effectively reduce the size of the network, making training more computationally feasible while maintaining high accuracy. Subsequent architectures, including VGGNet (Simonyan & Zisserman, 2014), further cemented the role of max-pooling in deep learning.

However, max-pooling is not without its criticisms. Springenberg et al. (2014) argued that max-pooling can discard valuable spatial information, suggesting that convolutional layers with increased stride could serve as a viable alternative. Despite these critiques, max-pooling remains widely used due to its simplicity and effectiveness.

Nearest Neighbor Downscaling

Nearest neighbor downscaling is a straightforward and computationally efficient method for resizing images. It involves assigning the value of the nearest pixel in the original image to the corresponding pixel in the resized image. This method is particularly appealing due to its simplicity and speed, making it suitable for real-time applications and scenarios where computational resources are limited. However, this approach can lead to aliasing artifacts, which manifest as jagged edges and blocky patterns in the resized image. These artifacts could affect the performance of convolutional neural

networks (CNNs) by introducing noise and reducing the accuracy of feature extraction.

In the context of CNNs, nearest neighbor interpolation has been used in various image preprocessing tasks. For instance, studies such as by Odena et al. (2016) in the context of generative adversarial networks (GANs) have utilized nearest neighbor interpolation for upsampling to process images for training. This method ensures that the dimensions of input images match the requirements of the network without adding significant computational overhead. It is used to fix checkerboard artifacts.

Despite its limitations, nearest neighbor downscaling remains a viable option in scenarios where computational efficiency is prioritized over image quality. Its primary advantage lies in its ability to perform rapid downscaling operations, which can be critical in real-time applications such as video processing and online learning systems.

Gaussian Downscaling (Pyramid Downsampling)

Gaussian downscaling, or pyramid downsampling, involves smoothing an image with a Gaussian filter before downsampling. This technique, developed by Burt and Adelson (1983), aims to reduce aliasing and preserve the structural integrity of the image. It has been shown to improve the robustness of feature detection and extraction in various computer vision tasks.

Lowe (2004) employed Gaussian downscaling in the development of the Scale-Invariant Feature Transform (SIFT), demonstrating its effectiveness in maintaining feature consistency across scales. Gaussian downsampling, while not always explicitly

stated, plays a role in multi-scale processing within CNN architectures. This technique is often utilized to create image pyramids or feature pyramids, which are integral to many state-of-the-art models in computer vision.

Comparative Studies

Several studies have conducted comparative analyses of these downscaling methods. Lin et al. (2013) introduced the Network in Network architecture, highlighting the effectiveness of max-pooling and also exploring alternatives like average pooling. They found that while max-pooling provides a good balance between complexity and performance in MNIST dataset, alternative methods such as NIN and dropout usage offer improvements in other tests.

Zeiler and Fergus (2014) provides an in-depth analysis of different network architectures, including the effects of max-pooling and other methods, and visualizes the learned features to understand the impact of these methods. Using switches, they use unpooling operation for reconstruction. So, their work shows a novel way to visualize activity within the model.

Despite these insights, there is a lack of consensus in the literature regarding the optimal downscaling method for CNNs. Factors such as the specific application, the dataset, and the network architecture all play crucial roles in determining the most suitable approach. This study aims to address selection of generalized downscaling method by providing a comprehensive comparison of max-pooling, nearest neighbor downscaling, Gaussian downscaling,

and the absence of downscaling, thereby contributing to a more nuanced understanding of their relative advantages and limitations.

METHODOLOGY

In the absence of a theoretical foundation, the inner workings of artificial neural networks and their derivatives can appear vague. For practitioners and researchers alike, an understanding of the underlying processes and the capacity to shape them according to specific needs are crucial. While many models are typically constructed through a process of trial and error, it is essential to assess the impact of each layer or structure that is added or removed on the model. This is a fundamental step in the development of more sophisticated and effective systems. In this study, the effect of the Max-Pooling feature, a frequently utilized component in convolutional neural networks, is examined in comparison to other downsampling methods.

The application of resolution downscaling methods is a fundamental aspect of digital image processing, as it enables the more efficient utilization of large visual data. By reducing the resolution, the data set can be made more manageable for recording, processing, and analysis without losing its inherent attributes. In particular, with regard to the size of the data, it is of great importance to determine the optimal resolution in order to render it more manageable in both temporary memory management (RAM or VRAM) and in the storage of data on hard disks. At this juncture, the degree of feature richness inherent to digital visual data becomes a pivotal parameter. Consequently, data points containing continuity and rich visual information can be stored with magnetic storage media, whereas the resolution may be reduced in data that is rapidly

consumed. So, it is essential to define the source image data, the intended area of use, and the desired features in advance. In circumstances where the potential for needed feature loss is minimal, reducing resolutions and employing lossy compression techniques can be advantageous in terms of efficiency. For instance, reducing the resolution of an image can result in a reduction of the image data, which in turn can facilitate the production of solutions with low memory requirements by accelerating the processing times for image analysis operations. It is crucial to strike a balance between the hardware cost and the performance cost. This is particularly relevant in the context of visual broadcasts, remote gaming sessions, and visual processing in autonomous vehicles, where real-time processing requirements are paramount.

The reduction of resolution in convolutional neural networks (CNNs) represents a fundamental design and functional step in the development of network architectures. The objective of this process is to reduce the spatial dimensionality of feature maps following convolution operations in the layers of the architecture. The objective is to retain the most pertinent attributes while eliminating less significant information. As the situation varies from data set to data set and is relative when it is not clearly defined, and as the data-target relationship in different images is not fully defined, it is necessary to make comparisons with commonly used data sets and models in order to generalize the performance of the processes. In this study, we address this issue and examine the performance of general resolution reduction operations in CNN networks. The primary rationale for reducing the resolution in CNN layers is to minimize the number of parameters and the number of requisite

operations, such as activation functions, by reducing the amount of data. In this manner, both the memory load and the processing power requirement will be diminished. Concurrently, this process will preclude the network from relying on memorization and will facilitate its acquisition of the capacity to perform more general inferences. Another method that can be employed to enhance the generalization capability is to introduce disparate levels of noise to the input image as a preprocessing step or to eliminate specific pixel data with a dropout layer.

In the context of resolution reduction, the values that will represent the new image pixel can be either a pixel value in the original image or a value affected by the surrounding data. In this study, we examine the effects of three commonly used resolution reduction methods on neural networks: Nearest Interpolation, the Pyramid Downsampling method using Gaussian values, and the Max-Pooling method, which is frequently used in intermediate layers in CNNs.

In the resolution downscaling process, a representative pixel is obtained as a result of the output of the operations applied to a region with the help of samples taken from the images or a function. This pixel represents a point in the reduced resolution. In this way, a lower resolution image will be obtained as a result of the operations applied to the whole image. It is important to note that the resolution reduction method is valid for digital images, that is, for images defined as discrete and not continuous. This approach allows for the examination of resolution differences and their effects through the use of a dimensional definition of our finite number of pixel data. In the Nearest Neighbor interpolation algorithm, it is crucial to consider

the nearest neighboring pixels when lowering or increasing the resolution. This allows for the determination of a pixel value by selecting the closest pixels. While this method is relatively fast, it is important to note that it can result in the occurrence of the aliasing effect, which refers to the sharp transitions between pixels. In the Pyramid Downscaling method using Gaussian, the pixels will exhibit smoother transitions. However, this approach also entails the loss of edge information.

In the Nearest Neighbor interpolation method, the number of neighborhoods to be handled is directly proportional to the number of operations to be performed. Gaussian resolution reduction is a form of interpolation that employs a Gaussian function kernel in accordance with the values of radius 3 and sigma = 1.5/3.0. The max-pooling technique is based on the logic of selecting the highest pixel value by dividing the image into sections according to the filter size. In this manner, the resolution of the image is diminished in accordance with the filter size, and the network is trained on lower-resolution images, thereby yielding results. The objective is typically to enhance the efficiency of the process by eliminating redundant data, reducing the memory size requirements, and creating lower-dimensional models with fewer artificial neural network parameters. In all resolution reduction algorithms, the ratio in the study is defined and coded in accordance with the process of halving the resolution.

Table 1 presents a general methodological comparison of the resolution reduction methods and absence of scaling utilized in the study. The data in the table are presented in general terms, and their detailed numerical values are analyzed in the Results and Discussion section. The primary considerations in the selection of the methods

were the distinction between functional operations and the ability to maintain the information level of the results.

Table 1: Comparison of information conservation, complexity and utilization of resolution reduction methods or lack thereof in artificial neural network layers.

Method	**Information Conservation**	**Computational Complexity**	**Use Cases**
Max-Pooling:	Medium	Low	CNNs, Object Detection
Pyramid Downsampling:	High	High	Multi-scale analysis, image processing
Nearest Interpolation:	Low	Veri Low	Real-time applications, simple resizing
No Scaling:	Maximum	Very High	Simple neural networks, memorization tasks

Artificial Neural Network Models

Neural network models are composed of cells that contain the activation function, which determines the transition to the subsequent layer based on the incoming data. Additionally, they comprise parameters that form the connections between layers. A basic neural network model can be represented by a structure similar to that depicted in Figure 1. The data stored in the inter-layer connections defines the network defined in training. Since the communication channels between the layers hold the multiplier weights for the data, it is these channels that define the information

in the network. These intercellular connections are therefore defined as parameters and carry the information in the model. When the model is saved, the weights of the model trained with the model architecture are also saved and stored in the file.

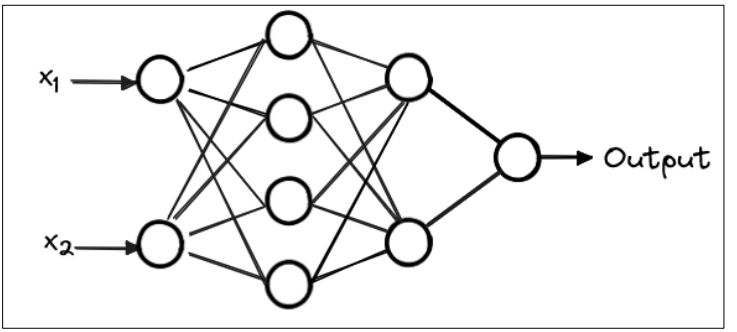


Figure 1: Basic Neural Network Model

The concept of layers is of vital importance in the context of neural networks. The most salient distinction between deep and basic neural networks is the presence of a considerable number of hidden layers. As the number of layers increases, so do the issues encountered during learning and the requirement for larger memory capacities. This has highlighted the necessity of employing distinct functional methods between layers. These requirements include methods such as reducing the data size, transferring the features without losing them, and normalization operations to transfer the data to the networks more accurately through the activation functions of the networks. In terms of advancement in activation functions, Vaca-Rubio et al. (2024) developed an alternative network named as Kolmogorov-Arnold Networks (KANs), wherein activation functions are adaptively defined within the network. Concurrently, these networks employ linear weights through the

replacement of such weights with univariate parametric spline functions.

As the data size increases in Artificial Neural Network architectures, the processing power required to train the information in the network also increases. At the same time, the feature extraction required for data processing is of particular importance. Especially in artificial neural networks containing images, layers containing convolution operations are used. In this way, the patterns in the image are revealed and fed to the neural network. Neural networks that perform these operations are called CNNs (Convolutional Neural Networks). CNNs are a popular neural network structure that combines convolutional filtering with neural networks. They are frequently used for object detection and classification, especially in 2D images.

In CNN models, a new two-dimensional image is obtained by applying convolution to each block in the image. In our study, the convolution block size is defined as a matrix of [3,3]. If we need to apply convolution to continuous data we need to define a continuous convolution process, we can define it as given in Equation 1.

$$(f * g)(t) = \int_{-\infty}^{\infty} f(\tau) g(t - \tau) d\tau \qquad \text{(Eq. 1)}$$

The equation defined here describes the result of the convolution of an incoming time-dependent function g with the filter f. The combination of these operations applied at each $d\tau$ time defines the convolution. However, since images are two-dimensional, our function Im(x,y) and our convolution filter will be two-dimensional. Also, since images are pixel-based matrices, our

convolution process will be a discrete convolution process. In this case, our equation can be defined as Equation-2 (Olga et al., 2016).

$$Conv2D(x,y) = \sum_{i=i_{min}}^{i_{max}} \sum_{j=j_{min}}^{j_{max}} f(i,j)Im(x-i, y-j) \qquad \text{(Eq. 2)}$$

In Equation 2 we describe the convolution result of a two-dimensional filter of dimensions *[i,j]* applied to a two-dimensional image when applied to an image of dimensions *[x,y]*. Since our data are two-dimensional pixel-based images, the results are obtained and processed by discrete convolution.

Since the memory size required to hold the parameters increases after convolution, resolution reduction is usually applied to reduce the number of parameters in the network. The most commonly used resolution reduction process in CNN networks is Max-Pooling. Other resolution reduction methods have been incorporated instead of Max-Pooling layer and models are built using this new layer replacement. Figure 2 depicts the layers and methods utilized in the CNN architectures in the study. It is also used as valid for the other resolution reduction methods mentioned in this network.

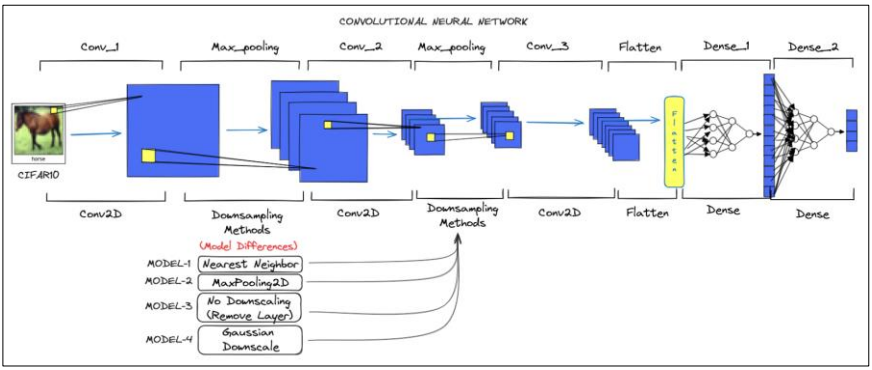


Figure 2: CNN model architectures used for comparison.

When we need to define the CNN model shown in Figure 2 in Python language, we can define it using more than one method. In this study, neural network operations were defined using the Tensorflow Keras library. Since we use custom defined functions in this study, we can define our code as in Figure 3 by creating a model class.

```
class MyModel(keras.Model):
    def __init__(self, num_classes=10):
        super(MyModel, self).__init__()
        self.conv1 = layers.Conv2D(32, (3, 3), activation='relu', name='conv1')
        self.maxPool1 = layers.MaxPooling2D((2, 2), name='maxpool1')
        self.conv2 = layers.Conv2D(64, (3, 3), activation='relu', name='conv2')
        self.maxPool2 = layers.MaxPooling2D((2, 2), name='maxpool2')
        self.conv3 = layers.Conv2D(64, (3, 3), activation='relu', name='conv3')
        self.flatten = layers.Flatten(name='flatten')
        self.dense1 = layers.Dense(64, activation='relu', name='dense1')
        self.dense2 = layers.Dense(num_classes, name='dense2')
```

Figure 3: Class and layer definitions that determine the CNN architecture.

In order to call our own functions in our class, we will need to define the following definition with the name *call* method in the class as shown in Figure 4. Here we can run the special or general methods we want in the layers by associating the input and output

data. The element to consider here is to determine the layer operations in the *call* method according to the architectural layers defined in the *__init__* method in the class definition.

```
def call(self, input_tensor, training=False):
    a=tf.nn.relu(self.conv1(input_tensor))
    #b=self.maxPool1(a)
    b=resDown2(a)
    #b=cv2.pyrDown(a)
    c=tf.nn.relu(self.conv2(b))
    #d=self.maxPool2(c)
    d = resDown2(c)
    e=tf.nn.relu(self.conv3(d))
    f=self.flatten(e)

    x=tf.nn.relu(self.dense1(f))
    return self.dense2(x)
```

Figure 4: Methods called and arguments passed according to the defined architecture.

Since the training data in the models are kept in the connections between cells, these connections determine how much memory we will need. For this reason, the memory requirements of the models are directly proportional to the number of these connections, which is, the number of parameters. We can add the necessary code for parameter counting as a method to our class as shown in Figure 5.

```
def count_params(self):
    self.build((None, 32, 32, 3))  # Adjust input shape as needed
    trainable_count = sum([tf.keras.backend.count_params(w) for w in self.trainable_weights])
    non_trainable_count = sum([tf.keras.backend.count_params(w) for w in self.non_trainable_weights])
    return trainable_count + non_trainable_count
```

Figure 5: Method definition used in MyModel class for parameter counting.

In order to calculate the number of parameters, we need to build and compile our model. The *self.build* code in Figure 5 will be

enough for us to build the model, as we can do the build process outside the model and call it after the compile process.

The resDown2 method, which we define specially in the class, refers to a function that we define as a custom method. We can define this method in the class or we can define it as a function before the class definiton. Since our codes will be run on the graphical processing unit, it should be noted that our data should be defined and used as tensor. The custom function defined in Tensorflow is shown in Figure 6.

```
@tf.function
def resDown2(input_tensor):
    H, W = input_tensor.get_shape()[1], input_tensor.get_shape()[2]
    return tf.image.resize(input_tensor, [int(H / 2), int(W / 2)], preserve_aspect_ratio=True, method='nearest')
```

Figure 6: Function definition to be used in the custom layer.

The content of the function traced on Figure 6 is compiled to create a callable Tensorflow graph (Tensorflow, 2018). In this way, a callable polymorphic function is created. Apart from *tf.function*, *tf.lambda* function, customized *@component* definitions and customized layers can be used. The class we created in this study is an example of customized layers. As seen in Figure 6, when we write our own class, we can define and use our own method as a tensorflow function. However, it should be noted here that since operations are performed using tensors, we should pay attention to tensor transformations in our in-function definitions.

Other important definitions used in CNNs are activation function and optimizer. *ReLu* activation function given in Equation 3 is used in the layers of CNN models for the study.

$$f(x) = \begin{cases} x, & x > 0 \\ 0, & x \leq 0 \end{cases} \quad [0, \infty) \qquad \text{(Eq. 3)}$$

In the definition of activation functions, as the number of hidden layers increases in multilayer architectures, it is seen that the artificial neural network cannot reflect the errors in training to the change of weights very well. This problem is called the vanishing gradient problem. Vanishing gradients means that while updating the weights in the model, the values that should be updated towards the input layers approach zero or become zero, that is, changes in the weights towards the input layers cannot be made and therefore the training cannot progress as desired. To overcome this problem, different activation functions such as *Leaky ReLu* are used instead of the *ReLu* function. In the *Leaky ReLu* function, a positive number such as *0.01x*, which is greater than zero but relatively small, is added instead of *0* if *x* is less than or equal to zero in Equation 3. The data *x* in the equation here represents the input variable to the neuron.

Another important feature in our network is the choice of optimizer, which is an important feature for training. Adam optimization is used in all models in our study (Kingma & Ba, 2014). Adam optimization is defined as a stochastic gradient descent method based on adaptive estimation of first-order and second-order moments. As noted by Kingma et al., its computational efficiency with low memory requirements makes it worthwhile to choose this method, while its invariance to diagonal rescaling of the gradients is advantageous in terms of parameters.

The high number of parameters in the architecture defined in artificial neural networks increases the memory requirement as well as the number of operations to be performed. Since the same

operation needs to be applied to the data repeatedly, processing on the graphics card (GPU) provides much more advantages than processing on the processor (CPU). For this reason, especially in new neural network architectures, high GPU and RAM requirements can be seen. In GPU architecture, a single instruction set can be applied to multiple data stored in the graphics card memory. In this way, neural network training provides relatively higher speed compared to CPUs. Although there are different options to run the code on different graphics cards, most used way for trainining is performed on graphics cards with the CUDA architecture. As for the CUDA architecture, the video card driver must be compatible with the Python and libraries used in the versions of the CUDA library required to run the CUDA cores. Version compatibility is an important factor especially if you are going to run the code on your own hardware. Therefore, it is best to review Tensorflow's CUDA compatibility list and make installations accordingly (Tensorflow, 2018).

RESULTS & DISCUSSIONS

This section presents and discusses the results of a comparative study on various resolution reduction methods in convolutional neural networks (CNNs). The primary objective is to evaluate the impact of max-pooling, nearest neighbor downscaling, Gaussian downscaling (pyramid downsampling), and the absence of downscaling on CNN accuracy and efficiency. By conducting a systematic examination of these methods on the CIFAR10 dataset and corresponding network architecture models, we aim to identify the strengths and weaknesses of each approach. The results are discussed in detail, highlighting key findings, statistical significance,

and practical implications. Additionally, we provide insights into how these methods influence the learning process, timings, and memory requirements for CNNs, offering an in-depth analysis for generalized usage.

As a dataset, CIFAR10 is used in the study. The CIFAR-10 dataset is a widely used benchmark in the field of machine learning and computer vision. The dataset was released in 2009 and is a labeled subset of the 80 million tiny images dataset (Krizhevsky & Hinton, 2009). CIFAR-10 consists of 60,000 32x32 color images in 10 different classes, with 6,000 images per class. The dataset is divided into 50,000 training images and 10,000 testing images. Each class represents common objects, such as airplanes, automobiles, birds, cats, and dogs, making it an excellent dataset for image classification tasks.

The CIFAR-10 dataset has been used in academic research to assess the performance of various machine learning and deep learning algorithms. It serves as a standard dataset for evaluating the efficacy of new architectures and techniques in image recognition. Significant works include "Deep Residual Learning for Image Recognition" by Kaiming He et al., which introduced ResNet, a deep residual network architecture that achieved state-of-the-art results on CIFAR-10 (He et al., 2016). Another notable paper is "Wide Residual Networks" by Sergey Zagoruyko and Nikos Komodakis in 2016, which proposed modifications to the ResNet architecture, resulting in a wider network that performed better on CIFAR-10 without a significant increase in computational complexity (Zagoruyko& Komodakis, 2016). Another one would be the paper "DenseNet: Densely Connected Convolutional Networks" by Gao

Huang et al. (2017), presented DenseNet, a new architecture that connects each layer to every other layer in a feed-forward fashion, significantly improving parameter efficiency and performance on CIFAR-10. These papers have contributed significantly to advancing the field of deep learning by providing benchmarks and improvements in network design.

With given architecture in Metodology section and using CIFAR10 dataset four similar model is created with only difference is resolution reduction methods. So that comparative analysis can be done in an objective manner. Analysis results and observations are discussed in this section. For more generalized results Google Colab environment is used to determine timings (Bisong, 2019). Codes are also tested in local offline computer in regards to ensure coherence of results in different environments.

Visual Impact of Decreasing Resolution in Digital Images

In Figure 7, Gaussian pyramid resolution reduction and Nearest Interpolation resolution reduction methods are applied on the original Lenna image with a resolution of *1280x720* along with the Max-Pooling method. The aim here is to provide a better understanding of the differences between different resolution reduction methods and the Max-Pooling method. In this way, it is aimed to reduce the loss of time and resources by better examining the performance of different methods in CNN networks and to gain higher achievements with the right methods. In the rest of the study, since there are two resolution reduction layers in the CNN architecture, by calling it Stage-2 it can be better observed how the resolution reduction on the image is affected when the resolution reduction operations in the study are continued as Stage-3 and Stage-

4. In this subsection, instead of the CIFAR10 images, the Lenna image is used to observe the effects of downsampling methods visually.

Figure 7: Stage-2 resolution reduction at 320x180 pixel size.

The effects of Stage-3 and Stage-4 resolution reduction were also examined by applying the same methods again on the images obtained from Stage-2. Especially in Figure 8, the Max-Pooling process was repeated in the 3rd stage of resolution reduction and the resolutions were obtained in size of *160x90*. Here, the effects of the processes on the image and the differences between the methods in terms of results started to become more understandable. As the most prominent features, it is seen that the Gaussian method gives us a softer (smoother) image when looking at the edges, while the Nearest Interpolation method gives the sharpest image. At this point, the Max-Pooling method presents a similar image between the other two methods, but the difference in the eye area can be seen compared to the other methods. In order to understand the differences between

the methods, a resolution reduction was performed again as Stage-4 operation.

Figure 8: Stage-3 resolution reduction at 160x90 pixel size.

As a result of the resolution reduction process applied in Stage-4 shown in Figure 9, the images become 80x45 pixels in size. At this size, the success of the methods in preserving visual features can be seen more clearly. While the loss rates of edge information are high in all three algorithms, the Nearest Interpolation method has the hardest transitions. It can be seen that the Max-Pooling method loses its advantage in this level of reduction and has difficulty in providing many feature distinctions such as brightness and transitions according to human perception.

Figure 9: Stage-4 resolution reduction at 80x45 pixel size.

Comparison of Training Results

The CNN architecture in the models used for training consists of three convolution layers, with 2 layers of maxpooling2D in between, replaced by Nearest Neighbor and Gaussian downscaling, respectively. After that, there is a dense layer that forms after flatten operation (the fully connected layer) and another dense layer for 10 outputs that determines the output layer. The results obtained by applying convolution without any resolution reduction after that are also analyzed. In this way, the effect of different resolution reduction approaches on the network performance was observed. The summary of the network model architecture that uses max-pooling (Model-2) is given in Table 2. Summary consists layer information with output dimensions and their parameter count in the network. Model is fed with 32x32 pixel images from the CIFAR10 dataset.

Table 2: Sequential Model Layers and Parameters for Max-Pooling (Summary)

Layer (type)	**Output Shape**	**Parameter Count**
Conv2D	(None, 30, 30, 32)	896
MaxPooling2D	(None, 15, 15, 32)	0
Conv2D	(None, 13, 13, 64)	18496
MaxPooling2D	(None, 6, 6, 64)	0
Conv2D	(None, 4, 4, 64)	36928
Flatten	(None, 1024)	0
Dense	(None, 64)	65600
Dense	(None, 10)	650
Total params:	122,570	

According to the model training results given in Figure 10, it is seen that the model trained without any resolution reduction (Model-3) is the model with the lowest loss rate of *0.2* at epoch 6, showing a successful and fast training. In the Max-Pooling method, the same value (loss rate of *0.2*) was reached at epoch 33, while the other resolution reduction methods (Nearest Neighbor and Gaussian) reached these values at epoch 25. If we look at the timings, the effect of resolution reduction does not seem to make much sense when we look at the size/time tradeoff. Especially when we examine the timings, it seems logical not to use the resolution reducing method for the training, which lasts between *10-11* seconds. This method only has downside in terms of memory limits. In this case, reducing the *batch_size* size appears as another solution. Increasing the striding window in the convolution is also another way to introduce

reduction. However, according to the results data loss inversely effect performance and timing.

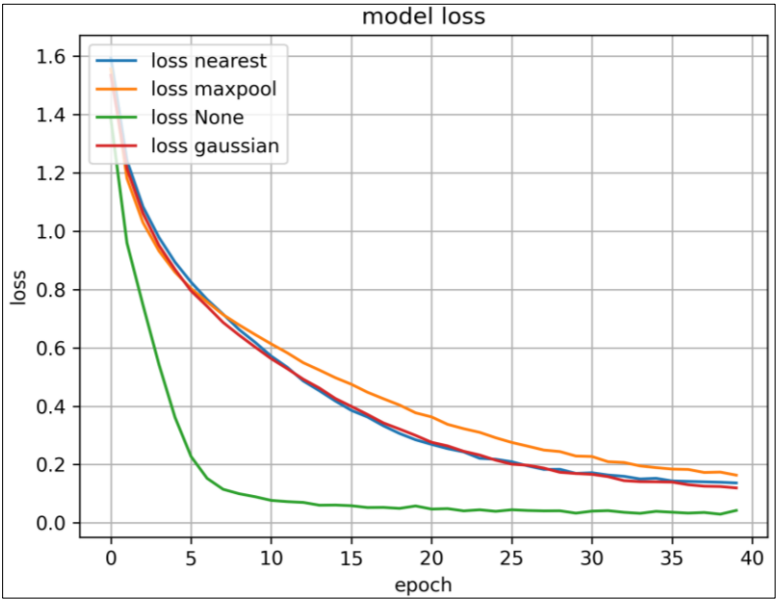


Figure 10: Train results in terms of model losses for resolution reduction methods.

When we look at Max-Pooling as a method to reduce resolution, we see that it is about 8 epochs behind the other loss reduction approaches. Of course, while the simplicity of the operation is an advantage for Max-Pooling, the time taken by the filtering process should also be taken into account. In the comparisons using Google colab T4 GPU, it was observed that the training using the Nearest Neighbor method took 8 seconds per epoch, while Max-Pooling took the same amount of time. One epoch training time, which was around *10-11* seconds without resolution reduction, was around *16-17* seconds with the Gaussian resolution reduction method. Considering these results, it is seen that it is more

advantageous not to reduce the resolution. However, if it is necessary to reduce the resolution, it is more advantageous to reduce the resolution by using the Nearest Neighbor interpolation process in terms of both training time and performance.

4.3. Analysis of Accuracy Results

When we examine the accuracy results, the performance differences can be seen in Figure 11. In particular, the fastest and highest performance is achieved with only convolution layers in the original image. In terms of performance, resolution reduction methods have relative performance compared to Max-Pooling. However, when the training times are analyzed, the Nearest Interpolation method turns out to be more advantageous. It should be noted that while resolution reduction shows similar performance in the first five epochs, there is a certain difference in terms of performance as training time increases. At this point, it would be more logical to use nearest interpolation instead of Max-Pooling, especially for large data sets and complex problems.

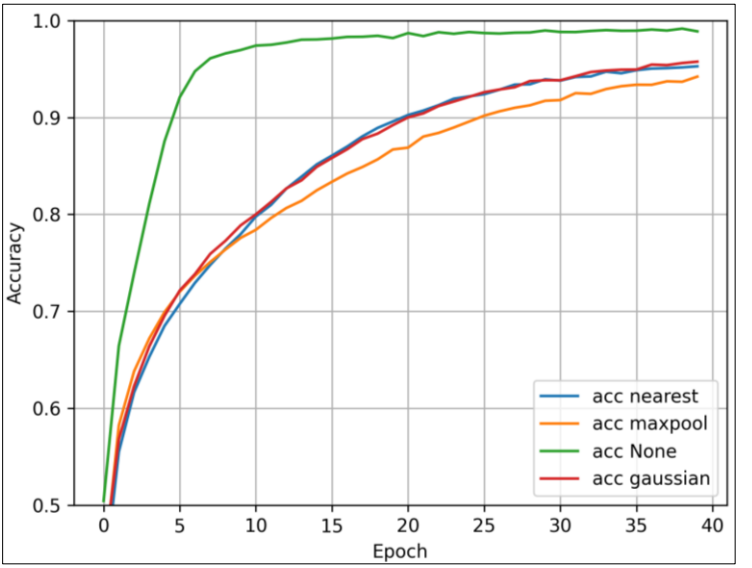


Figure 11: Resolution reduction methods Accuracy graph.

Training Time Analysis

On Table 3, the differences between reducing the resolution in the intermediate layers on the image with different methods and without any resolution reduction process (Model 3) are analyzed. In terms of timing, it can be seen that the resolution reduction with Nearest Interpolation process gives the biggest time advantage with a time of *8.38* seconds per epoch in terms of shortening the training time the most. On the other hand, it is noteworthy that the model that provides the fastest convergence in training is created only with the convolution layers of the original images and the neural network layers without downsampling layers. In particular, it is one of the biggest factors that it reaches a loss of *0.2* within *6* epochs and achieves the highest training accuracy compared to the others. Another noteworthy element in the table is that the Gaussian-based

resolution reduction method requires almost twice the training time compared to Model-1. The complexity of the downscaling method used and its performance on the GPU is an important factor affecting the training time.

Table 3: Training time and performance comparison of models.

Total Epoch (50)	Model-1 Nearest	Model-2 Max Pooling	Model-3 No Downscale	Model-4 Gaussian Downscale
Elapsed Time for Fit (sec)	**418.76**	445.32	565.07	838.48
Epoch Avg. (sec)	**8.38**	8.91	11.30	16.77
Loss 0.2 (Epochs)	25	33	**6**	25
Loss 0.2 Epoch Time (seconds)	209.38	293.91	**67.81**	419.24
Train Acc.	0.9633	0.9587	**0.9914**	0.9672
Test Acc.	0.6269	0.6764	0.6251	**0.6807**

Although the Epoch amounts seen in Table 3, which are used to determine the time to reach the amount of loss set for comparison, vary from training to training, it is observed that there is a maximum difference between 1-3 epochs in the tests performed.

Parameters and Memory Requirements

In particular, the timing and performance values in Table 3 show the effects of resolution reduction in convolutional neural

network models. Although the tests show the effect of no resolution reduction on performance, hardware constraints often require the use of methods such as data reduction and *batch_size* reduction in order for the models to run in lower memory. Although the model structure is generally the same, there is a layer difference between reducing the resolution or applying convolution over the original resolution. For this reason, if we examine the memory requirements of the models, as seen in Table 4, except for Model-3, which does not apply any resolution reduction process, the other three models have the same number of parameters since there is only a difference in the method applied, so the memory requirements of these three networks are the same number.

For Model 3, the reason why the network requires more parameters and hence more memory is that there are more pixels due to the fact that the total resolution is not reduced and as a result, more connections are needed between neurons. The memory requirement is calculated by keeping the values in 4-byte variables in float32 format as standard. It should be noted that memory requirement is not directly linked with processing needs as there are clearly differences between memory requirements in Table 4 and process timings in Table 3.

Table 4. *Comparison of Parameter Numbers and Memory Requirements of Models*

	Model-1: Nearest Neighbor	**Model-2: Max Pooling 2D**	**Model-3: None No Downscale**	**Model-4: Gaussian Downscale**
Parameter Count	122570	122570	2825930	122570
Required Memory (byte)	490280	490280	11303720	490280

The comparison Table 4 shows the parameter counts and memory requirements for four distinct models that employ disparate downscaling methodologies. The first model, which employs nearest interpolation, has a parameter count of 122,570 and a memory requirement of 490,280 bytes. Similarly, Model 2, which employs Max Pooling 2D, also has 122,570 parameters and requires 490,280 bytes of memory. In contrast, Model-3, which does not apply any downscaling method, has a significantly higher parameter count of 2,825,930 and a corresponding memory requirement of 11,303,720 bytes. Finally, Model 4, which employs Gaussian downscale, exhibits the same parameter count and memory requirement as Models 1 and 2, with 122,570 parameters and 490,280 bytes of memory. This table demonstrates that while Model-3 (without downscaling) significantly increases both the parameter count and memory requirement, the other three models (Nearest Interpolation, Max Pooling 2D, and Gaussian Downscale) maintain the same lower levels of parameters and memory usage.

CONCLUSION

While there are discrepancies in terms of resolution downscaling, the impact of the methodologies on performance is largely analogous, with the exception of the training time necessitated by Gaussian-based resolution downscaling. Despite requiring more memory, the most pronounced distinction is observed in the network achieved through Model-3, which does not employ any downscaling layer. Given that both the time required to attain the targeted performance and the training performance should be the primary determinants of preference, it is evident that further investigation is necessary.

The small size of the 32x32x3 images in the dataset is another factor that should be considered. By examining the difference at different resolution levels (up to 100 pixels, up to 1000 pixels, etc.) with different data sets of complexity, it will be possible to reveal the effects of intermediate layer resolution reduction at different resolution sizes in a clearer manner.

The study examines the impact of resolution reduction methods in convolutional neural network layers using training sessions with low-resolution visual inputs. It elucidates the discrepancies between these approaches and demonstrates how resolution layers should be utilized in model formation, with the objective of aligning with the specific criteria of the applications. It is evident that each layer, designated as feature formation, must be meticulously examined to ensure that the system does not introduce any form of data discrimination due to data loss.

In future studies, it is planned to investigate the effect of input resolution dimensions in the intermediate layers and the structural and consequential effects of the changes of different layer operations on the networks. Investigating boundary detections and the effect of latent space layers, which can be defined as the space of feature vectors, will also be an important resource for designing new network models.

References

LeCun, Y., Boser, B., Denker, J., Henderson, D., Howard, R., Hubbard, W., & Jackel, L. (1989). Handwritten digit recognition with a back-propagation network. *Advances in neural information processing systems*, *2*.

LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.

Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, *25*.

Simonyan, K., & Zisserman, A. (2014). Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*.

Ranzato, M. A., Boureau, Y. L., & Cun, Y. (2007). Sparse feature learning for deep belief networks. *Advances in neural information processing systems*, *20*.

Springenberg, J. T., Dosovitskiy, A., Brox, T., & Riedmiller, M. (2014). Striving for simplicity: The all convolutional net. *arXiv preprint arXiv:1412.6806*.

Odena, A., Dumoulin, V., & Olah, C. (2016). Deconvolution and checkerboard artifacts. *Distill*, *1*(10), e3. https://distill.pub/2016/deconv-checkerboard/

Burt, P. J., & Adelson, E. H. (1983). The Laplacian Pyramid as a Compact Image Code. IEEE *Transactions on Communications*, *31*(4), 532-540.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International journal of computer vision*, *60*, 91-110.

Lin, M., Chen, Q., & Yan, S. (2013). Network in network. *arXiv preprint arXiv:1312.4400*.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13* (pp. 818-833). Springer International Publishing.

Vaca-Rubio, C. J., Blanco, L., Pereira, R., & Caus, M. (2024). Kolmogorov-arnold networks (kans) for time series analysis. *arXiv preprint arXiv:2405.08790*.

Olga Russakovsky, Andras Ferencz, Kyle Genova and Riley Simmons-Edler (2016). COS 429 - Computer Vision Course Presentations. *Princeton University*. https://www.cs.princeton.edu/courses/archive/fall17/cos429/notes/fall2016/cos429_f16_lecture03_filtering.pdf

TensorFlow, Developers (2018). TensorFlow. *Site oficial*. Available Online. Accessed Date: 06.2024. https://www.tensorflow.org/api_docs/python/tf/keras/layers/Lambda

https://www.tensorflow.org/tfx/guide/custom_function_component?hl=en

https://www.tensorflow.org/tutorials/customization/custom_layers?hl=en

https://www.tensorflow.org/api_docs/python/tf/function
https://www.tensorflow.org/guide/intro_to_graphs?hl=en

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*. https://www.tensorflow.org/api_docs/python/tf/keras/optimizers/Adam

TensorFlow, D. (2018). Cuda Compatibility. Build from source. TensorFlow. *Site oficial*. Available Online. Accessed Date: 06.2024. https://www.tensorflow.org/install/source?hl=en#gpu

Krizhevsky, A., & Hinton, G. (2009). Learning multiple layers of features from tiny images. https://www.cs.toronto.edu/~kriz/learning-features-2009-TR.pdf

He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

Zagoruyko, S., & Komodakis, N. (2016). Wide residual networks. *arXiv preprint arXiv:1605.07146*.

Huang, G., Liu, Z., Van Der Maaten, L., & Weinberger, K. Q. (2017). Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4700-4708).

Bisong, E (2019). Google colaboratory. Building machine learning and deep learning models on google cloud platform: a comprehensive guide for beginners, 59-64.

BÖLÜM III

The Assesment of The Risk of Fire at Woodworking Plants Via Fmea-Gra Approaches

Neylan KAYA[1]

Introduction

The flames that originate from a specific source and distribute by itself with no help are called fire (Güvel & Güvel, 2002). The risk of fire at woodworking enterprises is high. The burning rate of the raw materials and auxiliary materials that are used at plants is high. Therefore, it is quite important for the safety management of enterprises to assess the risks and to set a risk assessment matrix.

It is hard to assess the risks for fire safety at woodworking enterprises. The risks in this study were determined by risk experts.

[1] Dr. Öğr. Üyesi, Akdeniz Üniversitesi, İktisadi ve İdari Bilimler Fakültesi, İşletme Bölümü, Antalya/Türkiye, Orcid: 0000-0003-2645-3246, neylankaya@akdeniz.edu.tr

10 risk experts assessed the 14 risks about fire which were encountered at woodworking plants.

In order to measure the priority numbers (RPN) of fire risks, FMEA was utilized. Since the probability, seriousness and determinability rate of risk depend on expert knowledge on account of RPN in FMEA, these data display subjectivities. To overcome those deficiencies of FMEA, the risk factors were solved via Grey Relational Analysis through setting up two different models which were determined by expert opinions. The fact that Grey Relational Analysis provides a safer and more flexible way, helps the sector develop strategies towards the future, ranks the risk, which is uncertain through risk priority numbers, and the fact that the enterprises take safety precautions against unacceptable risks are the contribution of the method that was utilized to the study.

When the studies which have been carried out via FMEA method recently are investigated, Li and Chen (2019) presented a new FMEA (GPRM) which integrates fuzzy and grey method to the conventional risk priority point calculations. They used GPRM to overcome the limitations in conventional FMEA and to make a wider assessment by integrating the criteria to the weights. Nie et. al. (2019) suggested Bayesian Fuzzy Assessment Number (BFAN) to get a further explanation to the various opinions of the experts and a new FMEA method by using extended grey relational analysis method like TOPSIS method.

Lastly, they utilized the extended GRA-TOPSIS method to give priorities to failure modes. In order to confirm the effectiveness of the suggested approach and to display the advantages against the

current FMEA methods, they presented a simulation study. Wang et. al. (2018) proposed an enhanced FMEA approach which depends on Dempster-Shafer Theory (DST) and TOPSIS in order to increase the efficiency of the conventional FMEA. They applied the new approach to a real-world problem, made comparisons and found out that the suggested method was more applicable.

Wang et. al. (2018) proposed a new risk-based FMEA with HoR based rough VIKOR. In order to manipulate the subjectivity and fuzziness, they used rough numbers. They displayed the efficiency and practicality of the suggested model by application the model to the transmission system of vertical processing center. Baynal et. al. (2018) determined the priorities of automotive production failures via GRA approach and minimized them via FMEA. The results showed that production process was improved with the help of the suggested method. Ayaz and Testik (2018) provided FMEA automatization for the computer users who utilize log data through grey relational analysis. They utilized grey relational analysis to automatize the process and to overcome the subjectivity.

Hu et. al. (2018) presented a new approach that can overcome the disadvantages of FMEA method. Considering the fuzziness in the assessments in the failure modes in FMEA, they proposed two-dimensional fuzzy linguistic variables in order to define the reliability of the assessment result. They applied GRA-TOPSIS in order to determine the risk rankings of specified failure modes. Liu (2017) proposed a FMEA approach which integrates PROMETHEE with the cloud model in order to deal with the fuzziness and randomness. They displayed the efficiency of the model by applying

the proposed approach in the risk analysis of health services. Zhao et. al. (2017) applied interval valued intuitionistic fuzzy (IVIF) entropy in order to weigh the risk factors and proposed FMEA-IVIF-MULTIMOORA approach by utilizing IVIF-MULTIMOORA method to determine the risk priority rank of failure modes. Alinezhad et. al. (2017) aimed to investigate the most significant risks in new product developing. For this aim, the risks were prioritized via COPRAS-G method.

It is aimed in this study to overcome the deficiencies on the subject in national literature. In this study, the grey modelling of failure mode and effect analysis which was developed by Chang et.al in 2001 was utilized in the prioritization of fire risks at woodworking plants.

2. Method

In this study, the valuation analysis report items of insurance companies and the potential risk modes which woodworking enterprises face were determined in accordance with the views of 10 fire experts. The experts graded each failure mode's probability, criticalness and determinability according to the scale on Table 1. The average points that 10 experts determined for each mode were calculated.

The grey modelling of failure mode and effect analysis approach which was developed by Chang et. al in 2001 was utilized in the prioritization of fire risks at woodworking plants. In the final phase, the results of the failure mode and effect analysis and two grey models were compared.

2.1. Failure Mode and Effect Analysis

Failure mode and effect analysis (FMEA) is a proactive method which prevents risks, failures and faults which may occur in the system, production and enterprise. It is used as a planning instrument in developing the processes and services.

FMEA depends on three factors: the probability of the risk (O), the criticalness of the risk (S) and the risk of not being determinable (D). Risk priority number is calculated by multiplying the probability, criticalness and determinability of the risk. With this number used as risk priority number (RPN), the risks are ranked. The level of the probability, criticalness and determinability of the risk is graded by using a 10-point scale. The case that the point is high shows that the effects of parameters in the system are undesirable in the system. In order to solve the problems in the system, the risk modes which have the highest RPN are needed to be focused on (Zhang and Chu, 2011).

The scale on which each risk mode was assessed according to three risk factors is on Table 1.

Table 1. Qualitative Assessment Scale for FMEA

Probability	**Score**	**Definition**
Almost no	1	There is no certain probability
Little	2	Rare failure
Weak	3	Very few failures
Low	4	A few failures
Mid low	5	Occasional failures
Middle	6	Medium range failure
Mid high	7	Medium range high failure
High	8	High failure
Very high	9	Very high failure
Almost certain	10	Almost certain failure

Different O, S, D combinations may give the same RPN value however the case that the risk modes which have the same RPN correspond to the different risk factors and do not take the indirect relations between the factors into account is one of the deficiencies of the method (Khasha et.al., 2013). It was aimed in the literature to review the deficiencies of failure mode and effect analysis through fuzzy or grey theory approaches and overcome them. In this study, the prioritization of fire risks at wood working enterprises with the help of grey FMEA modelling which Chang et. al. developed in 2001.

2.2. Failure Mode and Effect - Grey Relational Analysis

Grey system theory was developed by Deng Julong in 1982 in order to solve the problems in which fewer data is available and sampling distribution is unknown (Deng, 1982). The systems in which there is fuzziness or a lack of knowledge on system structure, system limitations, system behavior and parameters are called grey systems. Human body, agriculture and finance may be listed as examples of grey systems (Deng, 1989).

The purpose of grey system theory is to propose theory, technique, notions and ideas in order to solve fuzziness problems with deficient or incomplete knowledge. The basic content of grey system theory consists of grey relational analysis, grey forecasting, grey modelling, grey decision making, grey theory, grey mathematics and grey control (Deng, 1989).

Chang et. al. put forward in their study titled failure mode effects analysis using grey theory in 2001 that since the risk factors provide the characteristics of grey system theory which enables to

analyze the relationship between the current, expandable, countable, independent intermittent quantitative and qualitative series, it is possible to apply grey theory to FMEA.

The aim of grey relational analysis is to determine the rate of the relationship between each factor and the factor which is taken as a reference for comparison. The grade of effect between the factors is called grey relational grade and the function which defines grey relational grade is supposed to provide normality, dual symmetry, wholeness and approachability axioms (Yıldırım, 2018).

The steps of grey relational analysis in the failure mode and effect analysis may be listed as below (Chang et. al., 2001).

Step 1. Comparative series are determined.

$$X_i' = \left(X_i'(1), X_i'(2), \dots, X_i'(K)\right) \in X \quad \text{i=1,2,…,m} \tag{1}$$

X_i(k) shows the k factor of X_i . If n knowledge series is comparable, X matrix is defined as below.

$$\begin{bmatrix} X_1 \\ X_2 \\ \cdot \\ \cdot \\ \cdot \\ X_n \end{bmatrix} = \begin{bmatrix} X_1(1) & X_1(2) \cdots & X_1(k) \\ \vdots & \ddots & \vdots \\ X_n(1) & X_n(1) \cdots & X_n(1) \end{bmatrix} \tag{2}$$

Step 2. Reference series are determined.

In failure mode and effect analysis, small point means small risk. Thus, reference series are ranked by taking the smallest values of all risk factors.

Reference series are defined as

$$X_0 = \{X_0(1), X_0(2), \dots, X_0(k)\} = \{1,1, \dots,1\} \quad (3)$$

Step 3. Fuzzy relationship grade is determined by finding the difference between comparative series and reference series.

Step 4. Grey relationship coefficient is calculated.

$$\Upsilon\big(X_0(k), X_i(k)\big) = \frac{\Delta_{min} + \zeta\Delta_{max}}{\Delta_{0j}(k) + \zeta\Delta_{max}} \quad (4)$$

Here j=1,2,…,m and k=1,2,…,n

$X_0(k)$: reference series

$X_j(k)$: comparative series

$$\Delta_{min} = min\forall j \in i\ min\forall\ k\left\|X_0(k) - X_j(k)\right\| \quad (5)$$

$$\Delta_{max} = max\forall j \in i\ max\forall\ k\left\|X_0(k) - X_j(k)\right\| \quad (6)$$

is a determinative coefficient between $\boldsymbol{\zeta}\in(0,1)$. The studies express that $\boldsymbol{\zeta}$ value does not affect the ranking which will be formed after grey relational grade. In order to decrease the effect of the maximum value to the relationship coefficient, $\boldsymbol{\zeta}$=0,5 is accepted (Sofyalıoğlu, 2011; Kurt, 2008; Kuo et.al., 2008; Wu, 2007).

Step 5. Grey relational grade is determined.

The degree of how comparative series is similar to reference series is defined as grey relational degree (Üstünışık, 2007)

$$\Upsilon(X_o, X_i) = \frac{1}{n}\sum_{k=1} \Upsilon\left(X_0(k), X_i(k)\right) \quad (7)$$

On the condition that there are weigh values of the priority grades of criteria, grey relational grade can be found by multiplying the weigh value of the criterion and the grey relational coefficient.

$$\Upsilon(X_o, X_i) = \frac{1}{n}\sum_{k=1} \Upsilon\left(X_0(k), X_i(k)\right) * W_i(k) \quad (8)$$

Step 6. Risk priorities are ranked.

In a decision making problem, the best decision alternative is the one which has the highest grey relational degree (Sofyalıoğlu, 2011; Kuo et. al., 2008; Wu, 2002). The higher the relational grade is, the lower the effect of the failure source is.

3. Findings

10 fire risk experts graded each failure mode's probability, criticalness and determinability according to the scale on Table 1. The average points that 10 experts determined for each mode were calculated. The results obtained by assessing the fire risks in woodworking business via FMEA are displayed on Table 2.

Table 2. The Results of Failure Mode and Effect Analysis Application

Risk No	Potential Risk Mode	O	D	S	RPN	Rank No
1	Wood shavings left for piling up	8	8	8	512	5
2	Storing combustible material close to the heat sources of the plant (shaving burning stove)	6	6	5	180	13
3	Uncontrolled smoking	7	9	6	378	8
4	Improper using or storing inflammable materials during painting process	5	9	5	225	12
5	Disrepair of electrical installation	7	9	5	315	10
6	Uncontrolled storage of wastes (fiber etc.)	7	6	8	336	9
7	Disorganizations seen in general storage	3	9	4	108	14
8	Problems caused by heating installation	6	6	8	288	11
9	Risks arising from the structural characteristics of the enterprise building such as building style, building date, construction materials	7	7	8	392	7
10	Insufficient number of fire extinguishers	8	9	10	648	4
11	Lack of a lightning rod installation	6	9	9	486	6
12	The location or the lack of fire-escape stairs	8	10	9	720	3
13	Staff's lack of fire training.	9	9	9	729	2
14	The lack of fire/safety team	9	10	9	810	1

The information series matrix about the risk factor points on Table 2 is displayed on (9).

$$\begin{bmatrix} X_1(1) & X_1(2) & X_1(3) \\ X_2(1) & X_2(2) & X_2(3) \\ X_3(1) & X_3(2) & X_3(3) \\ X_4(1) & X_4(2) & X_4(3) \\ X_5(1) & X_5(2) & X_5(3) \\ X_6(1) & X_6(2) & X_6(3) \\ X_7(1) & X_7(2) & X_7(3) \\ X_8(1) & X_8(2) & X_8(3) \\ X_9(1) & X_9(2) & X_9(3) \\ X_{10}(1) & X_{10}(2) & X_{10}(3) \\ X_{11}(1) & X_{11}(2) & X_{11}(3) \\ X_{12}(1) & X_{12}(2) & X_{12}(3) \\ X_{13}(1) & X_{13}(2) & X_{13}(3) \\ X_{14}(1) & X_{14}(2) & X_{14}(3) \end{bmatrix} = \begin{bmatrix} 8 & 8 & 8 \\ 6 & 6 & 5 \\ 7 & 9 & 6 \\ 5 & 9 & 5 \\ 7 & 9 & 5 \\ 7 & 6 & 8 \\ 3 & 9 & 4 \\ 6 & 6 & 8 \\ 7 & 7 & 8 \\ 8 & 9 & 10 \\ 6 & 9 & 9 \\ 8 & 10 & 9 \\ 9 & 9 & 9 \\ 9 & 10 & 9 \end{bmatrix} \quad (9)$$

The differences matrix is on (10).

$$\begin{bmatrix} \Delta_{01}(1) & \Delta_{01}(2) & \Delta_{01}(3) \\ \Delta_{02}(1) & \Delta_{02}(2) & \Delta_{02}(3) \\ \Delta_{03}(1) & \Delta_{03}(2) & \Delta_{03}(3) \\ \Delta_{04}(1) & \Delta_{04}(2) & \Delta_{04}(3) \\ \Delta_{05}(1) & \Delta_{05}(2) & \Delta_{05}(3) \\ \Delta_{06}(1) & \Delta_{06}(2) & \Delta_{06}(3) \\ \Delta_{07}(1) & \Delta_{07}(2) & \Delta_{07}(3) \\ \Delta_{08}(1) & \Delta_{08}(2) & \Delta_{08}(3) \\ \Delta_{09}(1) & \Delta_{09}(2) & \Delta_{09}(3) \\ \Delta_{010}(1) & \Delta_{010}(2) & \Delta_{010}(3) \\ \Delta_{011}(1) & \Delta_{011}(2) & \Delta_{011}(3) \\ \Delta_{012}(1) & \Delta_{012}(2) & \Delta_{012}(3) \\ \Delta_{013}(1) & \Delta_{013}(2) & \Delta_{013}(3) \\ \Delta_{014}(1) & \Delta_{014}(2) & \Delta_{014}(3) \end{bmatrix} = \begin{bmatrix} 7 & 7 & 7 \\ 5 & 5 & 4 \\ 6 & 8 & 5 \\ 4 & 8 & 4 \\ 6 & 8 & 4 \\ 6 & 5 & 7 \\ 2 & 8 & 3 \\ 5 & 5 & 7 \\ 6 & 6 & 7 \\ 7 & 8 & 9 \\ 5 & 8 & 8 \\ 7 & 9 & 8 \\ 8 & 8 & 8 \\ 8 & 9 & 8 \end{bmatrix} \quad (10)$$

Grey relational coefficients are calculated with the help of differences matrix values and decisiveness coefficient. Grey relational coefficients calculated by using $\Delta_{min} = 2, \Delta_{max} = 9$ and $\boldsymbol{\zeta}$=0,5 values are displayed on (11).

$$\begin{bmatrix} \gamma_{01}(1) & \gamma_{01}(2) & \gamma_{01}(3) \\ \gamma_{02}(1) & \gamma_{02}(2) & \gamma_{02}(3) \\ \gamma_{03}(1) & \gamma_{03}(2) & \gamma_{03}(3) \\ \gamma_{04}(1) & \gamma_{04}(2) & \gamma_{04}(3) \\ \gamma_{05}(1) & \gamma_{05}(2) & \gamma_{05}(3) \\ \gamma_{06}(1) & \gamma_{06}(2) & \gamma_{06}(3) \\ \gamma_{07}(1) & \gamma_{07}(2) & \gamma_{07}(3) \\ \gamma_{08}(1) & \gamma_{08}(2) & \gamma_{08}(3) \\ \gamma_{09}(1) & \gamma_{09}(2) & \gamma_{09}(3) \\ \gamma_{010}(1) & \gamma_{010}(2) & \gamma_{010}(3) \\ \gamma_{011}(1) & \gamma_{011}(2) & \gamma_{011}(3) \\ \gamma_{012}(1) & \gamma_{012}(2) & \gamma_{012}(3) \\ \gamma_{013}(1) & \gamma_{013}(2) & \gamma_{013}(3) \\ \gamma_{014}(1) & \gamma_{014}(2) & \gamma_{014}(3) \end{bmatrix} = \begin{bmatrix} 0{,}545455 & 0{,}545455 & 0{,}545455 \\ 0{,}666667 & 0{,}666667 & 0{,}750000 \\ 0{,}600000 & 0{,}500000 & 0{,}666667 \\ 0{,}750000 & 0{,}500000 & 0{,}750000 \\ 0{,}600000 & 0{,}500000 & 0{,}750000 \\ 0{,}600000 & 0{,}666667 & 0{,}545455 \\ 1{,}000000 & 0{,}500000 & 0{,}857145 \\ 0{,}666667 & 0{,}666667 & 0{,}545455 \\ 0{,}600000 & 0{,}600000 & 0{,}545455 \\ 0{,}545455 & 0{,}500000 & 0{,}461538 \\ 0{,}666667 & 0{,}500000 & 0{,}500000 \\ 0{,}545455 & 0{,}461538 & 0{,}500000 \\ 0{,}500000 & 0{,}500000 & 0{,}500000 \\ 0{,}500000 & 0{,}461538 & 0{,}500000 \end{bmatrix} \quad (11)$$

Two grey models were set. In the first model, risk factors were accepted as equally weighted and the grey relational degree of each risk was calculated as on the equation (7). In the second model, fire risk experts of insurance firm weighted the priority levels of risk factors. The weight coefficients were determined as W_O=0.3, $W_D = 0.3, W_S = 0.4$ according to this. With the help of the equation (8), grey relational grade on the condition that risk factors had different weights was recalculated. The failure mode and effect analysis of the potential risks which enterprises in woodworking business may encounter, equally weighted risk factor models, different weighted risk factors model and priority ranking were displayed on Table 3.

Table 3. The comparison of FMEA, Equally weighted Grey and Different Weighted Grey Models

Risk No	Rank No	HMEA RPN	Rank No	Risk Factors Equally Weighted Grey RPN	Rank No	Risk Factors Different Weighted Grey RPN
1	5	512	5	0,546	5	0,545
2	13	180	13	0,694	13	0,700
3	8	378	8	0,589	8	0,597
4	12	225	12	0,667	12	0,675
5	10	315	10	0,617	11	0,63
6	9	336	9	0,604	9	0,598
7	14	108	14	0,786	14	0,793
8	11	288	11	0,626	10	0,618
9	7	392	7	0,582	7	0,578
10	4	648	3	0,502	2	0,498
11	6	486	6	0,556	6	0,550
12	3	720	4	0,502	4	0,502
13	2	729	2	0,500	3	0,500
14	1	810	1	0,487	1	0,489

When Table 3 is viewed, the risks of wood shavings left for piling up, storing combustible material close to the heat sources of the plant (shaving burning stove), uncontrolled smoking, improper using or storing inflammable materials during painting process, uncontrolled storage of wastes (fiber etc.), disorganizations seen in general storage, lack of a lightning rod installation, the lack of fire/safety team displayed the same results for three applications. However, there were slight differences in the ranking of risks of disrepair of electrical installation, problems caused by heating installation, insufficient number of fire extinguishers, staff's lack of fire training. It was found out that the highest risks for the enterprises in woodworking business were insufficient number of fire

extinguishers, lack of a lightning rod installation, the location or the lack of fire-escape stairs, staff's lack of fire training and the lack of fire/safety team risks which were the ones about the inadequacy of fire precautions of the enterprises. Disorganization seen in general storage had the lowest risk.

4. Conclusion

There are numerous risks and risky areas which may cause a fire at woodworking plants. It is aimed in this study to prioritize the fire risks at woodworking enterprises. Failure mode and effect analysis and grey theory approaches were utilized in the process of prioritizing the risks.

In this study, 14 possible risks for woodworking plants were determined by 10 risk experts. After all the risks subjected to this study were assessed by the experts according to their probability, criticalness and determinability, the findings of failure mode and effect analysis weighted and unweighted grey relational analysis were compared at the end of the study.

According to the results obtained, woodworking enterprises are supposed to take primary precautions against the risks of insufficient number of fire extinguishers, lack of a lightning rod installation, the location or the lack of fire-escape stairs, staff's lack of fire training and the lack of fire/safety team risks.

In this study, the applicability of grey relational analysis which was proposed in literature in order to overcome the deficiencies of failure mode and effect analysis was also tried. In the methods utilized, the case that the experts make different weighting

changes the ranking of risk priority numbers. Therefore, the knowledge and experience of the experts are of the essence.

Through this application on woodworking enterprises, possible risks of fire, the effects and the priorities of these risks were determined. The risks about fire are not supposed to be neglected. Enterprises may review their safety management and use their limited resources more efficiently with the help of this study.

References

Alinezhad, A., Amini, A., Rahmani, M. (2016). New product development risk assessment in the core banking using FMEA combined with COPRAS method and Grey Relations, *Journal of Money and Economy*, 10(3), 87-121.

Ayaz, H. İ., Testik, M. C. (2018). Automation of FMEA for computer servers using log data with grey relation analysis, 2nd. International Conference on Computer Science and Engineering, UBMK'17, 616-620.

Baynal, K., Sarı, T., Akpınar, B. (2018). Risk management in automotive manufacturing process based on FMEA and grey relation analysis: A case study, Advances in production engineering & management, 13(1), 69-88.

Chang, C. L., Liu, P. H., Wei, C. C. (2001). Failure mode and effect analysis using grey theory, *Integrated Manufacturing Systems*, 12(3), 211-216.

Deng, J. L. (1982). Control problems of grey systems, *Systems&Control Letters*, 1(5), 288-294.

Güvel, E. A., Güvel, A.Ö. (2002). *Sigortacılık,* Seçkin Yayıncılık, Ankara.

Hu, Y., You, X., Wang, L., Liu, H. (2018). An integrated approach for failure mode and effect analysis based on uncertain linguistic GRA–TOPSIS method, *Soft Computing*, 1-14.

Khasha, R., Sepehri, M. M., Khatibi, T. (2013). A fuzzy FMEA approach to prioritizing surgical cancellation factors, *International Journal of Hospital Research*, 2(1), 17-24.

Li, Z., Chen, L. (2019). A novel evidential FMEA method by integrating fuzzy belief structure and grey relational projection method, *Engineering Applications of Artificial Intelligence*, 77, 136–147.

Liu, H. (2017). Failure mode and effect analysis using cloud model theory and PROMETHEE Method, *IEEE Transactions on Reliability*, 66(4), 1058-1072.

Nie, W., Liu, W., Wu, Z., Chen, B., Wu, L. (2019). Failure mode and effect analysis by integrating Bayesian fuzzy assesment number and extended gray relational analysis-technique for order preference by similarity to ideal solution method, *Quality and Reliability Engineering*, 1-21.

Sofyalıoğlu, Ç. (2011). Süreç hata modu etki analizini gri değerlendirme modeli, *Ege Akademik Bakış*, 11(1), 155-164.

Wang, Z., Gao, J., Wang, R., Chen, K., Gao, Z. (2018). Failure mode and effect analysis using Dempster-Shafer theory and TOPSIS method: application to the gas insulated metal enclosed transmission line (GIL), *Applied Soft Computing*, 70, 633-647.

Wang, Z., Gao, J., Wang, R., Chen, K., Gao, Z., Zheng, W. (2018). Failure mode and effect analysis by using the house of reliability-based rough VIKOR approach, *IEEE Transactions on Reliability*, 67(1), 230-248.

Yıldırım, B. F. (2018). Gri ilişkisel analiz, yönetsel ve stratejik problemlerin çözümünde çok kriterli karar verme yöntemleri, Ed. Bahadır Fatih Yıldırım, Emrah Önder, Dora Yayıncılık, 3. Baskı, Bursa.

Zhao, H., You, J., Liu, H. (2017). Failure mode and effect analysis using MULTIMOORA method with continuous weighted entropy under interval-valued intuitionistic fuzzy environment, *Soft Computing*, 21, 5355-5367.

BÖLÜM IV

The Future of Data Science and Statistics

Yusuf Dilbilir[1], Sadi ELASAN[2]

Introduction

Data science and statistics have become critical disciplines, whose importance is increasing with the rapid advancement of technology today and causing radical changes in various industries. These disciplines, together with the integration of innovative technologies such as big data analytics, artificial intelligence and machine learning, are transforming the strategic decision-making processes of businesses, the healthcare sector, financial institutions and governments.

[1] Lecturer Dr., Hakkari University, Vocational School of Health Serv., Türkiye, Orcid: 0000-0002-4569-251X, yusufdilbilir@hakkari.edu.tr
[2] Assoc. Prof. Dr., Yuzuncu Yil University, Faculty of Medicine, Department of Biostatistics, Van/Türkiye, Orcid: 0000-0002-3149-6462, sadielasan@yyu.edu.tr

Data science can be defined as a process that aims to extract meaningful information from large data sets. This process includes the stages of data collection, cleaning, analysis and interpretation, and thanks to these stages, businesses and institutions can make more accurate and informed decisions by understanding their data. Statistics plays a critical role in making data meaningful by providing the scientific methods that form the basis of these processes.

Today, data science and statistics are applied in a wide range of areas, from marketing strategies to patient treatment, from energy management to city planning. In-depth analyzes and insights provided in these areas enable institutions to gain competitive advantage and better manage societies.

In this section, we will examine the current situation in the field of data science and statistics, future development trends, potential societal impacts, and innovations in these fields. Understanding the advancements of advanced technologies and methodologies in these disciplines will allow us to discover new opportunities and solutions in both academic and applied fields.

1. Current Situation in Data Science and Statistics

Today, it is seen that rapidly advancing technologies and methods are used extensively. Techniques such as big data analytics, machine learning and deep learning have an important place in data science applications. While the use of statistical methods and data analysis increases in the business world and scientific research, ethical issues such as data security and privacy are also gaining more importance. In the future, the role of data science and statistics will

expand further with artificial intelligence, industry 4.0 and digital transformation processes, which will bring new opportunities and challenges (Anderson, 2008; Hastie et al., 2009, Manyika et al., 2016).

1.1 Definition and Importance of Data Science

Data science refers to the process of extracting meaningful information from large data sets. This process is carried out using data analysis tools and methods provided by modern technology. The data science discipline covers the stages of data collection, data cleaning, data analysis and finally extracting meaningful information from the data.

During the data collection phase, large amounts of data are collected from various sources. This data can be structured or unstructured and comes in different formats and sources. The data collection process is a critical step in the success of data science projects because the analysis process cannot be carried out successfully without creating accurate and comprehensive data sets (Davenport and Patil, 2012).

During the data cleaning phase, the quality of the collected data is improved and errors, omissions or inconsistencies in the data set are eliminated. Data science practitioners generally perform operations such as data standardization, filling in missing data values, and handling outlier data points at this stage. This step is vital for the reliability and accuracy of data analysis (McAfee and Brynjolfsson, 2012).

During the data analysis phase, cleaned data sets are examined using statistical methods, machine learning algorithms, or

other analytical techniques. The goal at this stage is to discover patterns, relationships, and trends in the data set. A well-structured data analysis process enables meaningful insights to be obtained from data and makes decision-making processes informed (Provost and Fawcett, 2013).

The data interpretation phase is the process of extracting meaningful and applicable information based on the results of the analyzed data. At this stage, the information and insights necessary for businesses to make strategic decisions and make trends and forecasts are provided. At this stage, data science is used in many industries such as finance, health, and marketing with the deep analyzes and predictions it provides and transforms data-based decision-making processes in these industries (Kitchin, 2014).

Today, the discipline of data science has a wide range of applications, from financial analysis to patient treatment methods, from analysis of customer behavior to product recommendations. Businesses and institutions can gain competitive advantage, use their resources more effectively and increase their operational efficiency with the right data science strategies (Manyika et al., 2011).

1.2 The Role of Statistics and the Problem of Proper Use

Statistics is a critical discipline that underpins data science. Using correct statistical methods in data collection and analysis processes is of great importance for the reliability and meaningfulness of the results obtained. Statistical methods are necessary to make data meaningful and to derive statistically reliable results.

However, applying statistical methods accurately and effectively is not an easy process. Errors or incorrect methods in statistical analysis can lead to incorrect conclusions and faulty decision-making processes. For example, an incorrect assumption or incorrect choice of model can seriously affect the reliability of conclusions drawn from the analyzed data.

Another important problem is that statistical analyzes are often not done by experts. In data science projects, the data collection and cleaning stages are usually carried out by data engineers or analysts, while the statistical modeling and analysis steps are usually not done by statisticians or data scientists who specialize in this field. This can lead to incomplete or misunderstanding of statistical information and, as a result, inaccurate analyses.

Correct understanding and correct use of statistical information is vital for the reliability and accuracy of data science applications. Proper selection and application of statistical methods ensures that data analysis is based on a solid foundation and provides reliable information in decision-making processes.

In summary, the lack or misuse of statistical information in data science projects can negatively affect the success of the project. Therefore, it is important to include professionals specialized in statistics in data science teams and for these experts to take an active role in data analysis processes. In this way, accurate and reliable results can be obtained and data-based decision-making processes can be made more effective.

2. Future Trends, Technologies and Innovations

2.1 Big Data and Data Analytics

Big data analysis is the process of processing and analyzing the complex and high-volume data sets that businesses encounter today. These analyzes are often performed by integrating advanced techniques such as data mining, machine learning and artificial intelligence. Thanks to these techniques, big data analytics enables businesses to obtain meaningful insights from their data assets (Jordan and Mitchell, 2015).

The term big data is often used to describe data sets that are too large, diverse, and rapidly changing to be processed by traditional data management tools. These datasets can consist of millions or billions of records and can be obtained from different sources (such as sensors, social media, web traffic). Big data analytics requires specifically designed algorithms and systems to manage, store, process and analyze these voluminous data sets.

Data mining is one of the fundamental steps in big data analytics and is used to discover hidden patterns, relationships, and trends from data sets. For example, large retail companies can use data mining techniques to understand customer purchasing habits and optimize inventory management accordingly. Data mining applications often include various machine learning techniques such as pattern recognition, classification, regression analysis, and identification of clusters (Wu et al., 2014).

Machine learning and artificial intelligence (AI) are one of the most important components of big data analytics. These technologies are used to detect complex patterns in large data sets

and make accurate predictions. In particular, techniques such as deep learning and neural networks are used to more precisely identify hidden relationships and trends within big data. Artificial intelligence applications play a critical role in big data analytics to improve automated decision-making processes and enable businesses to gain a competitive advantage.

Big data analytics is a powerful tool, especially used in market analysis and customer behavior prediction. Thanks to big data analysis, companies can better understand customer preferences, trends and shopping habits. This information can be used to optimize marketing strategies and increase customer satisfaction.

Big data analytics has an important place in today's business world and helps businesses gain competitive advantage. With the integration of techniques such as data mining, machine learning and artificial intelligence, big data analytics processes become more effective and efficient. With the advancement of these technologies, the importance and prevalence of big data analytics will increase in the future (Obermeyer and Emanuel, 2016).

2.2 Artificial Intelligence and Machine Learning

Artificial intelligence (AI) and machine learning (ML) are rapidly developing trends that have attracted great attention in the field of data science and statistics in recent years. These technologies play an important role, especially in areas such as discovering complex patterns in large data sets, analyzing data and making accurate predictions. Artificial intelligence and machine learning are used in big data analytics processes in integration with techniques

such as data mining, deep learning and neural networks, providing strong support for businesses in their strategic decision-making processes. The advancement of these technologies contributes to the acceleration of automation and data-driven innovations and the deepening of industrial transformation processes (Brynjolfsson and McAfee, 2011).

Fundamentals and Applications of Machine Learning

Machine learning is a branch of artificial intelligence that enables computer systems to learn from specific data sets. It can be examined under two main categories: supervised and unsupervised learning. Supervised learning focuses on predicting specific outcomes by learning on labeled data sets. For example, it can take certain symptoms as input to diagnose a patient and learn to associate those symptoms with certain diseases. Unsupervised learning, on the other hand, tries to find patterns and structures on unlabeled data sets. This method is used to identify natural groupings and relationships in the data set (Chui et al., 2016).

Deep Learning and Neural Networks

Deep learning is a branch of machine learning that performs complex tasks by creating multilayered structures using artificial neural networks (ANNs). These techniques are used to discover deep and hidden patterns in large data sets. Especially in areas such as image and voice recognition, deep learning models have achieved accuracy rates that exceed human performance. Deep learning allows the model to create more complex representations based on data during the learning process and to perform certain tasks using these representations. Therefore, deep learning plays an important

role in complex data analytics and artificial intelligence applications, especially used in discovering meaningful patterns in large data sets and making predictions with high accuracy.

2.3 Data Visualization and Interactive Analysis

Presenting complex data sets in a clear and effective way significantly simplifies the process of sharing and understanding the results of data analysis. Data visualization tools enable data to be presented with graphs, tables, and interactive displays. This helps decision makers understand data faster and make strategic decisions.

Charts and Chart Types: Data visualization is used to visually reveal patterns and relationships in data sets. Various chart types, such as line charts, column charts, scatter plots, box plots, and heat maps, are used to understand different aspects of the data.

Tables and Summaries: Data tables present large amounts of data in an organized manner, enabling detailed analysis. Summary tables visually present the basic statistical properties and summaries of the data, allowing decision makers to quickly get an overview.

Interactive Displays: Interactive visualization tools allow users to take a deeper look at data and analyze it from different perspectives. Users can zoom in, filter charts, or examine different segments of data interactively on the visualization.

Data Discovery and Recommendation Systems: Some data visualization tools automate data discovery processes and can make suggestions to users on discovering hidden patterns or relationships within the data. These features make it easier for data scientists and analysts to conduct in-depth exploration in data.

Use in the Strategic Decision Making Process: Visualization tools enable decision makers to quickly focus on data and make strategic decisions. For example, it is possible to optimize marketing strategies by visually analyzing sales trends or monitor business processes with visualization tools to increase operational efficiency.

As a result, data visualization and interactive analysis tools make complex data sets understandable, accelerating data analysis processes and giving decision makers the ability to make more robust data-driven decisions. These tools help businesses gain competitive advantage and strengthen their data-driven strategic management.

2.4. Artificial Intelligence and Automation

Artificial intelligence and machine learning enable a significant transformation in the data-driven decision-making processes of businesses and industries. Machine learning models, which are used to make important decisions such as credit risk assessments, especially in the financial sector, can make automatic decisions by processing big data analysis and past performance data.

For example, when evaluating loan applications, a bank or financial institution can make decisions such as approving or rejecting the loan application by analyzing data such as the applicant's financial history, credit score, and income status through machine learning models. These processes increase the efficiency of businesses by minimizing human errors and enabling faster decisions, while also improving the customer experience.

Artificial Intelligence and Strategic Decisions

Artificial intelligence stands out as an extremely powerful tool to support the strategic decisions of businesses. Especially in the retail industry, artificial intelligence and machine learning algorithms are used effectively to make sales forecasts and optimize inventory management. These technologies increase operational efficiency by enabling businesses to quickly respond to market demand changes (Kitchin, 2014).

For example, a retail company can predict future sales through machine learning models that analyze factors such as historical sales data, seasonal factors, the impact of promotions. These forecasts can be used to optimize stock levels and provide the right amount of product at the right time. In addition, understanding customer behavior and developing strategies accordingly becomes possible thanks to artificial intelligence.

Artificial intelligence also enables customer segmentation with big data analysis, creation of personalized marketing campaigns and increasing customer satisfaction. In this way, while businesses gain competitive advantage, they have the opportunity to make strategic decision-making processes more data-based and accurate.

Artificial intelligence also plays an important role in strategic decision-making processes in the healthcare sector. Machine learning models, especially used in medical imaging analysis and disease diagnosis, provide valuable support to doctors and healthcare professionals. For example, a hospital's medical imaging unit can detect potential abnormalities in MRI or CT scans using artificial

intelligence algorithms. In this way, early diagnoses can be increased and treatment processes can be made more effective.

Additionally, through the analysis of health data, deeper insights into patients' health status can be gained. Using information from large data sets, machine learning can make recommendations or identify risk factors for patients based on their personal medical history. This could be an important step in preventing or managing previously unanticipated health risks to patients.

In the healthcare industry, artificial intelligence and machine learning are also used in operational processes to increase the efficiency of healthcare services. For example, hospital management can predict patient density and resource needs using AI-based predictive models for optimal use of resources. In this way, patient care processes can be better planned and the quality of health services can be improved.

As a result, artificial intelligence and machine learning stand out as an important aid in strategic decision-making processes in the healthcare industry and have great potential to improve the quality of life of both healthcare professionals and patients (Yoo and Kim, 2014).

2.5. Future Developments and Expectations

Rapid advances in artificial intelligence and machine learning technologies will enable the emergence of even more advanced and effective applications in the coming years. These technologies will play a central role in digital transformation processes, especially industry 4.0 and the internet of things.

Industry 4.0 and Digital Transformation: Industry 4.0 is a transformation process characterized by the integration of digital technologies in production processes. Artificial intelligence and machine learning will play a critical role in increasing automation and efficiency in this process. By analyzing large data sets collected through sensors with machine learning algorithms, it will be possible to reduce errors in production processes, optimize maintenance and increase production efficiency (Laser et al., 2014).

The Rise of Autonomous Systems: Thanks to artificial intelligence and machine learning, automation levels will be further advanced and the development of autonomous systems will accelerate. These systems will be able to make complex decisions, optimize themselves and adapt to changing conditions. For example, the use of autonomous systems in areas such as traffic management or energy optimization will increase in smart cities.

Human-Machine Interaction and Interface Advances: AI-powered interfaces and interaction technologies will significantly improve user experience. Voice assistants will become smarter and more effective thanks to their natural language processing capabilities; It will help users do their jobs faster and more efficiently. Additionally, increasing robotic applications and human-machine collaboration will increase efficiency in the manufacturing and service sectors (Brynjofsson and McAfee, 2014).

Health and Medical Applications: In the healthcare sector, artificial intelligence and machine learning will provide great benefits in areas such as improving disease diagnoses, personalizing treatment plans and more effective management of healthcare

services. The use of artificial intelligence in medical imaging analysis, genetic analysis and patient management processes will become widespread and innovative solutions in the field of healthcare will increase.

2.6. Contributions of Statistical Analysis

Statistical methods are of great importance for the successful implementation and development of these technologies. Statistics plays a critical role in extracting meaning from large data sets, evaluating model performance, and ensuring the reliability of results. For example:

Data Analysis and Interpretation: Statistical methods are used in data collection, cleaning, analysis and interpretation processes in data science projects. These processes contribute to accurate results and reliable decision-making processes.

Evaluating Model Accuracy: Statistical metrics and tests are used to objectively evaluate the performance of machine learning and artificial intelligence models. In this way, the predictive power and generalization ability of the model can be evaluated more accurately.

Support in Decision-Making Processes: Statistical analyzes provide solid data-driven information in decision-making processes. For example, the use of statistical models when assessing credit risk in the financial sector provides decision makers with a more reliable and objective basis.

In this way, the integration of statistical analysis together with advances in artificial intelligence and machine learning will

enable the effective use of technologies and achieve results that create value for society.

3. Social Impacts and Ethical Issues

3.1 Data Privacy and Security

With the spread of big data analysis and artificial intelligence applications, data privacy and security are becoming increasingly important. Security vulnerabilities that occur during the collection, storage, processing and sharing of personal data can have serious consequences for both individuals and institutions. Protecting this data is of great importance, especially in sensitive areas such as health, finance, and trade.

Data privacy breaches can compromise individuals' privacy and undermine their trust. For example, if the security of systems that monitor user behavior or analyze health records is not ensured, serious grievances may occur as a result of misuse or leakage of this information. Therefore, strict legal regulations and ethical rules regarding data privacy must be implemented. Transparency allows users to understand what type of data is collected and how it is used and builds trust (Vayena et al., 2015).

3.2 Social Transformation and Innovation

Data science and statistics are leading to radical transformations in economic, social and scientific fields. These technologies are effective in a wide range of areas, from healthcare to education, from urban planning to environmental sustainability. For example, in the healthcare sector, it may be possible to track

disease spreads and manage healthcare resources more effectively thanks to big data analysis (Varian, 2014).

Data-based decision-making processes allow societies to be better managed and resources used more efficiently. For example, data analysis in urban planning processes can help create more sustainable and user-friendly solutions in areas ranging from traffic management to green space planning.

3.3 Statistical Ethical Issues

Ethical issues in statistical analysis are critical to the reliability and accuracy of data science and statistics. Ethical issues may arise, especially in the processes of data selection, biases and interpretation of results. For example, manipulating the selection of data sets to exclude certain groups or misinterpreting the results can mislead the public and mislead decision makers.

It is also important that algorithms are designed and implemented in accordance with the principles of equality, justice and diversity. For example, artificial intelligence systems used in recruitment processes may develop features that may discriminate based on subjectivities such as gender or ethnicity. Such situations deepen social inequalities and damage the sense of justice.

As a result, the social impacts and ethical issues of developments in data science and statistics should be carefully considered for the correct and fair use of technology. Establishing a strong ethical framework will help protect the rights of individuals and the general interests of society while ensuring maximum benefit from the potential of these technologies.

4. Conclusion

Data science and statistics are constantly evolving disciplines that are gaining increasing importance today and in the future. Technological and methodological advances in these areas enable businesses, institutions and societies to make data-driven strategic decisions. In particular, the use of innovative techniques such as big data analytics, machine learning and artificial intelligence enables the widespread adoption and application of these disciplines. However, in order to obtain maximum benefit from the potential of these technologies, it is extremely important to use them correctly and adopt approaches that are sensitive to ethical issues.

Data science and statistics are having profound and transformative impacts across a variety of industries and fields. For example, it can increase the efficiency of healthcare services by making disease predictions in the healthcare sector, reduce operational costs by optimizing stock management in the retail sector, and improve the quality of life by developing sustainable solutions in urban planning. In addition, statistical models used in risk management and market analysis in the financial sector encourage decision makers to make informed and data-driven decisions.

Important issues such as data privacy and security that arise with big data analysis and artificial intelligence applications should be addressed with ethical rules and legal regulations. Protecting personal data and using it securely is critical to ensuring the trust of individuals and societies. Fair and transparent approaches should be

adopted in the processes of data selection, design of algorithms and interpretation of results (Diebold, 2012).

Ethical issues in statistical analysis should not be ignored. The selection and manipulation of data sets can affect the accuracy and objectivity of the results. Therefore, the principles of non-bias, transparency and accuracy must be strictly adhered to in statistical analyses. Data scientists and statisticians are obliged to comply with ethical standards in data-based decision-making processes.

As a result, rapid developments in data science and statistics play a critical role in social transformation and innovation. In the future, the impact of these disciplines will increase further in industry 4.0 and digital transformation processes and will offer innovative solutions to the challenges faced by businesses, institutions and societies. However, in this process, the correct and ethical use of technology is an indispensable element for sustainable success.

Data science and statistics will continue to have a critical role in solving and advancing the complexities of the information age. Therefore, interdisciplinary collaboration and continuing education are important factors for the sustainability of advances in data science and statistics.

References

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer Science & Business Media.

Davenport, T. H., & Patil, D. J. (2012). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*, 90(10), 70-76.

McAfee, A., & Brynjolfsson, E. (2012). Big data: The management revolution. *Harvard Business Review*, 90(10), 60-68.

Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big Data*, 1(1), 51-59.

Kitchin, R. (2014). Big data, new epistemologies and paradigm shifts. *Big Data & Society*, 1(1), 2053951714528481.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011). Big data: The next frontier for innovation, competition, and productivity. *McKinsey Global Institute*, 1-156.

Vayena, E., Salathé, M., Madoff, L. C., & Brownstein, J. S. (2015). Ethical challenges of big data in public health. *PLOS Computational Biology*, 11(2), e1003904.

Jordan, M. I., & Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245), 255-260.

Obermeyer, Z., & Emanuel, E. J. (2016). Predicting the future - big data, machine learning, and clinical medicine. *New England Journal of Medicine*, 375(13), 1216-1219.

Lazer, D., Kennedy, R., King, G., & Vespignani, A. (2014). The parable of Google Flu: Traps in big data analysis. *Science*, 343(6176), 1203-1205.

Brynjolfsson, E., & McAfee, A. (2014). The second machine age: Work, progress, and prosperity in a time of brilliant technologies. *W. W. Norton & Company*.

Wu, X., Zhu, X., Wu, G. Q., & Ding, W. (2014). Data mining with big data. *IEEE Transactions on Knowledge and Data Engineering*, 26(1), 97-107.

Varian, H. R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, 28(2), 3-28.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. *SAGE Publications*.

Anderson, C. (2008). The end of theory: The data deluge makes the scientific method obsolete. *Wired Magazine*, 16(07), 16-07.

Chui, M., Manyika, J., & Miremadi, M. (2016). Where machines could replace humans and where they can't (yet). *McKinsey Quarterly*, 1-11.

Diebold, F. X. (2012). On the origin(s) and development of the term 'big data'. *The Oxford Handbook of Economic Forecasting*, 2016.

Yoo, S., & Kim, M. (2014). Big data analytics in the public sector: A case study of the Korea health insurance review and assessment service. *Government Information Quarterly*, 31, S25-S40.

Manyika, J., Lund, S., Chui, M., Bughin, J., Woetzel, J., Batra, P., & Ko, R. (2016). Digital America: A tale of the haves and have-mores. *McKinsey Global Institute*, 3, 2016.

Brynjolfsson, E., & McAfee, A. (2011). Race against the machine: How the digital revolution is accelerating innovation, driving productivity, and irreversibly transforming employment and the economy. *Digital Frontier Press*.

BÖLÜM V

İlaç Yönetiminde Risk Analizi Yöntemiyle Ramak Kala Olayların (risklerin) Belirlenmesi: Balikesir Özel Sevgi Hastanesi Merkezi Eczane Birimi Örneği

Tuğba ERMİŞLER[3]
Vahide BAYRAKAL[4]
A. Hüseyin BASKIN[5]

1.Giriş

Hasta ve çalışan güvenliği, sağlıkta kalite geliştirme ve akreditasyon çalışmalarında esas teşkil eder. Çalışan, hasta ve hasta

[3] Ecz., Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Sağlıkta Kalite Geliştirme ve Akreditasyon Anabilim Dalı, İzmir; Büyükçekmece Mimar Sinan Devlet Hastanesi, İstanbul; tugba_yesil@hotmail.com , ORCID: 0009-0006-4993-4922
[4] Öğr. Gör. Dr., Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Sağlıkta Kalite Geliştirme ve Akreditasyon Anabilim Dalı, İzmir; Dokuz Eylül Üniversitesi Hastanesi, Adli Mikrobiyoloji ve Biyolojik Savunma Laboratuvarı, İzmir; vahide.bayrakal@deu.edu.tr , ORCID: 0000-0003-0670-2575
[5] Prof. Dr., Dokuz Eylül Üniversitesi, Tıp Fakültesi Temel Tıp Bilimleri Bölümü Tıbbi Mikrobiyoloji Anabilim Dalı, İzmir; Dokuz Eylül Üniversitesi Sağlık Bilimleri Enstitüsü Sağlıkta Kalite Geliştirme ve Akreditasyon Anabilim Dalı, İzmir; Dokuz Eylül Üniversitesi Hastanesi, Adli Mikrobiyoloji ve Biyolojik Savunma Laboratuvarı, İzmir; huseyin.baskin@deu.edu.tr , ORCID: 0000-0003-3237-267X

yakını güvenliğini tehdit edebilecek, gerçekleşmek üzereyken son anda gerçekleşmeyen, istenmeyen olaylar 'ramak kala olay' (risk) olarak tanımlanmaktadır. Sağlık alanında risk analizi; (i) ilgili birim çalışanları ile risklerin belirlenmesi, (ii) risklerin ortaya çıkma olasılığı ve zarar verme etkisi ile risklerin sınıflandırılması (iii) kök-neden analizi yöntemi risklerin nedenlerinin belirlenmesi ve (iv) risklerin azaltılması yönünde düzeltici-önleyici-iyileştirici faaliyetlerin planlanması ve uygulanması basamaklarından oluşmaktadır (SKS-Hastane V6.1, 2020).

Sağlıkta hizmet kalitesi; 1. Söz verilen zamanda (timely), 2. Çalışan-hasta-hasta yakını merkezli (patient centered), 3. Adil (ırk, dil, din, kültür,.. farklılıklarına bakılmadan) (equal), 4. Etkin (güncel bilimsel esaslar üzerinde) (efficient), 5. Verimli (en az atıkla) (effective), 6. Emniyetli ve güvenli (safe and secure), ölçütleri üzerinden tanımlanır.

SKS-Hastane Seti Versiyon 6.1'de ilaç yönetiminin hastanede ilacın (ilaç, bir tanımıyla da dozu ayarlanmış zehirdir) dahil olduğu tüm süreçlerde etkin yönetimini sağlamak, hasta ve çalışana yönelik riskleri en aza indirmek olduğu vurgulanmış, ilaç yönetimi ile ilgili sorumlular ve sorumlulukların tanımlanması gerekliliği belirtilmiştir. Bu nedenle, ramak kala olayların eksik beyanına katkıda bulunan faktörlerin belirlenip hastane eczane sistemlerinin nasıl geliştirileceğini anlamaya doğru önemli bir adım olacaktır.

SKS-Hastane Seti Versiyon 6.1'e göre farklı bölümlerde risk değerlendirme çalışmaları yapılmıştır; Gür ve arkadaşları ameliyathane biriminde riskleri belirledikleri araştırmalarında, hasta

ve çalışan güvenliğini esas alan bir bakış açısı ile güvenli cerrahi kontrol listesinin etkin kullanılması ile birçok riskin (ramak kala olayın) ortadan kaldırılabileceğini belirtirken (Gür Ö. ve ark., 2019), Hamarat ve arkadaşları araştırmalarında, günümüzde yeni medya ilişkileri ile çıkan mahremiyeti zedeleyen ve bilgi güvenliğini tehdit eden risklerin sağlıkta risk yönetimi çerçevesinde değerlendirebileceğini göstermişlerdir. (Hamarat T. ve ark., 2022). Yılmaz ve arkadaşları da Anestezi Yoğun Bakım Birimi'nde hasta güvenliğini etkileyen riskler üzerinde bir çalışma gerçekleştirmiştir (Yılmaz ÖA ve ark., 2022). Sağlık hizmetlerinde risk analizi konusundaki bu üç araştırma farklı çalışma alanlarında gerçekleştirilmiş olsa da, Dokuz Eylül Üniversitesi Risk Yönetimi Yönergesi temel alınmıştır. Her üç çalışmada da riskleri belirleyenler, bilgi eşitleme toplantılarından sonra yapılan toplantılarda, ilgili birimlerin çalışanları olmuştur.

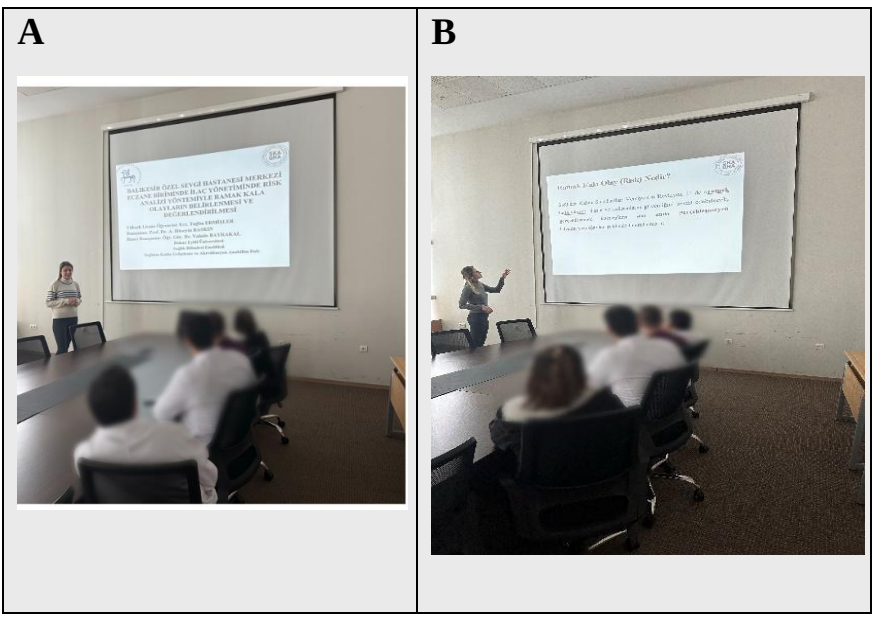
Yüksek lisans tez araştırması olan bu araştırmada da, Sağlıkta Kalite Standartları (SKS) Hastane Seti Versiyon 6.1 ve Dokuz Eylül Üniversitesi Risk Yönetimi Yönergesi esas alınarak, bu kez Balıkesir Özel Sevgi Hastanesi Merkezi Eczane Birimi'nde hasta güvenliği, ilaç güvenliği, ramak kala olay, istenmeyen olay, risk, risk yönetimi, risk analizi, kök neden analizi, risk haritası, etki ve olasılık gibi konulara ilişkin bilgi eşitlemesi yapılması (birinci toplantı), farkındalık oluşturulması, hasta ve ilaç güvenliğini etkileyebilecek ramak kala olayların (risklerin) belirlenmesi ve değerlendirilmesi amaçlanmıştır.

2.GEREÇ VE YÖNTEM

Yüksek lisans tez araştırmasının bir bölümü olan bu araştırma Balıkesir Özel Sevgi Hastanesi Merkezi Eczane Biriminde

gerçekleştirilmiştir. Dokuz Eylül Üniversitesi Girişimsel Olmayan Araştırmalar Etik Kurulundan 14.02.2024 tarih ve 7892- GOA protokol numaralı 2024/06-06 karar numarası ile etik kurul onayı ve 05.02.2024 tarihinde Balıkesir Özel Sevgi Hastanesi Başhekimliğinden kurum izni alınmıştır.

Balıkesir Özel Sevgi Hastanesi Eczane Biriminde, 15 Mart ve 22 Mart 2024 tarihlerinde, eczane birim çalışanlarıyla iki ayrı toplantı yapılmıştır. Bu toplantılara bir eczacı, bir eczacı teknisyeni, 1 tıbbi sekreter, bir destek personeli olmak üzere toplamda dört personel katılmıştır. İsimler anonimleştirilerek toplantı tutanağı tutulmuş ve kayıt altına alınmıştır. Kişisel Verilerin Korunması Kanunu çerçevesinde katılımcılardan izin alınarak, fotoğraf çekilmiştir (Şekil. 1) Toplantıda yöneltilen sorular bir sonraki toplantıda detaylı bilgi paylaşımı için not alınmıştır.



Şekil 1. A. 15 Mart 2024 tarihinde yapışan Balıkesir Özel Sevgi Hastanesi Eczane birim çalışanları bilgi eşitlemesi toplantısı **B.** 22 Mart 2024 tarihinde yapılan Balıkesir Özel Sevgi Hastanesi Eczane birim çalışanları ile birim içi ramak kala olayların (risklerin) belirlenmesi toplantısı

Balıkesir Özel Sevgi Hastanesi Merkezi Eczane birimi personeliyle;

1. 15 Mart 2024 tarihli toplantıda sağlık çalışanlarının ortak dilde buluşabilmeleri amacıyla ilaç güvenliği ve hasta güvenliği temel alınarak ramak kala olay, istenmeyen olay, risk, risk analizi, risk yönetimi gibi konulara ilişkin bilgi eşitlemesi yapılmıştır.
2. 22 Mart 2024 tarihli ikinci toplantıda eczane biriminde birim çalışanları ile birim içi ramak kala olaylar belirlenmiş, toplantı tutanaklarında isimler anonimleştirilerek yalnızca meslek grupları şeklinde imza altına alınmıştır.

Balıkesir Özel Sevgi Hastanesi Merkezi Eczane Biriminde, 25 Aralık 2014 tarihli Dokuz Eylül Üniversitesi Risk Yönetimi Yönergesi'ne göre risk analizleri yapılmış ve ramak kala olaylar (riskler) belirlenmiştir. Adı geçen yönergenin en önemli özelliği 5'li değil 10'lu değerlendirme ölçütleri ile daha ayrıntılandırılmış belirlemeler yapılmasıdır (cetvel olarak düşünürsek, metre ve santimetre cetvelleri arasındaki fark gibi).

3.BULGULAR ve TARTIŞMA

Balıkesir Özel Sevgi Hastanesi Merkezi Eczane Biriminde toplantı tutanakları ile kayıt altına alınan toplantılarda, eczane birimi

çalışanları, eczane iş akış süreçlerini gözden geçirerek ilaç yönetiminde 11 adet risk (ramak kala olay) belirlemiştir. Belirlenen 11 riskten yedisinin hasta güvenliği, üçünün ilaç güvenliği, bir tanesinin de çalışan güvenliği ile ilgili olduğu tespit edilmiştir (Tablo 1).

Tablo 1: Balıkesir Özel Sevgi Hastanesi Merkezi Eczane Birimi Çalışanları ile Birlikte Belirlenen İlaç Güvenliğini Etkileyebilecek Riskler (Ramak Kala Olaylar)

Ramak Kala Olay (Risk)	SKS Hedefi
1. Hastanede kullanılan bazı ilaçların çeşitli sebeplerden temin edilememesi dolayısıyla Eksik/Geç İlaç Uygulamaya Bağlı Tedavinin Gecikmesi Riski	Hasta Güvenliği
2. Eczane deposunda sıcaklık/nem değerlerinin limitler dışında kalma riski	İlaç Güvenliği
3. Eczacıya sözel order verilmesi sonucunda hastaya yanlış ilacın verilme riski	Hasta Güvenliği
4. İlaç hazırlama esnasında görünüşü/okunuşu benzer ilaçların karışma riski	Hasta Güvenliği
5. İlaç hazırlama esnasında order edilenden farklı uygulama yollu (tablet yerine i.v. gibi) ilacın verilme riski	Hasta Güvenliği
6. Soğuk zincire tabi ilaçların transferinde soğuk zincirin kırılma riski	İlaç Güvenliği
7. Yüksek riskli ilaçların diğer ilaçlarla karışma riski	Hasta Güvenliği
8. Pediatrik doz yerine yetişkin dozu gönderilme riski	Hasta Güvenliği
9. Orderların geç onaylanması ve tedavinin gecikmesi riski	Hasta Güvenliği

10. İlaçların tüketilemeden miatlarının geçme riski	İlaç Güvenliği
11. Depo personelinin iş kazası riski	Çalışan Güvenliği

Hastanede ilaç yönetimi ve eczacılık hizmeti, verilen sağlık hizmetinin vazgeçilmez bir parçasıdır. İlaç güvenliği; hasta güvenliği ve çalışan güvenliği faktörlerinin iyileştirilmesinde gerekli bir husus olduğundan sağlık hizmeti veren tüm birimlerde ilaç güvenliğini sağlayacak çalışmalar özenle planlanıp ve uygulanmalıdır. İlaç güvenliği ve ilaç yönetiminde yaşanacak sorunlar; hastaların hastanede kalış süresinin uzamasına ve tedavi giderlerinin artmasına da yol açabilmektedir. Tıpta ve teknolojide yaşanan gelişmelere ek olarak son yüzyılda ilaç endüstrisindeki gelişmeler ve çokça yeni etken madde keşfi olduğu düşünüldüğünde ilaca bağlı hataların önüne geçilebilmesi için sistematik çalışmalar planlanmalıdır (İlaç Güvenliği Rehberi- Versiyon 2.0, 2015).

Hastane eczanesi hasta bakımında hayati bir rol oynar ve istekte bulunulan ilacın, hedeflenen hastaya doğru bir şekilde ve tam zamanında verildiğinden emin olmaya odaklanır (Sağlık Hizmetlerinde Kalite Göstergelerinden: Emniyetli ve Güvenli). Hastane eczanesinin görevi ilaçları satın almak, uygun koşullarda saklamak ve dağıtmaktır. Bu faaliyetler uzman personellerin sorumluluğunda olan eczane lojistiği süreçleri olarak bilinir çünkü ilaçların belirli koşullar ve standartlar altında yönetilmesi gerekir (Romero A., 2013).

Hastane eczaneleri, sundukları hizmet sebebiyle yatan hastalar için ilaç güvenliği ve hasta güvenliği konularında, yüklü miktarda ilaç transferi ve depo faaliyeti gerçekleştirilmesi nedeniyle de çalışanlar bakımından riskli birimlerdir. Kurumun kalite yöneticileri; belirli zaman aralıklarında her birimin çalışanları ile birlikte çalıştıkları birimler için toplanarak riskleri belirlemeli, risk analizi çalışmalarını gerçekleştirip uygulamalara ve sonuçlarına sistemlerinde yer vermelidirler. Ek olarak birimlerde yapılan risk analizi çalışmalarının o birimin çalışanlarıyla birlikte yapılması, çalışanların fikirlerinin ve önerilerinin mutlaka dikkate alınması önemlidir. (Akın Yılmaz Ö ve ark, 2022).

4.SONUÇLAR

Araştırmada, Balıkesir Özel Sevgi Hastanesi Merkezi Eczane birimi personeliyle toplantı yapılarak istenmeyen olay, ramak kala olay, ilaç güvenliği gibi konulara ilişkin bilgi eşitlemesi yapıldıktan sonra, Eczane Biriminde gerçekleşebilecek ramak kala olaylar da birim çalışanlarının katılımı ve gönüllü destekleri ile örneklendirilmiştir.

1. Toplantılarda katılımcılar, yapılan bilgi eşitlemesi sunumlarından duydukları memnuniyeti dile getirmişlerdir,
2. Eczane biriminde çalışan personeller, verilen örneklerin bazılarına kendilerinin de tanık olduklarını, ilaç yönetiminde ramak kala olayların gerçekleşmesi durumunda doğuracağı sonuçların tahmin edilemez boyutlara ulaşabileceğini ifade etmişlerdir,

3. Hastanede ilaç yönetiminde ramak kala olayların eczane ayağında ne kadar kritik öneme sahip olduğunun fark edildiği belirtilmiş, süreçlerde mevzuata dayalı iyileştirme tavsiyeleri talep edilmiştir.
4. Öte yandan risk değerlendirme çalışmalarının yaygınlaştırılması ve standardizasyonu (normalleştirilmesi) için ilgili diğer kurum çalışanlarına da ulaşmasının sağlanması önerisinde bulunulmuştur,
5. Araştırmada, birim çalışanlarıyla birlikte saptanan ramak kala olayların ilaç güvenliği, hasta güvenliği ve çalışan güvenliği konularında yapılacak uygulamaların düzenlenmesinde önemli olduğu ve yine birimde aynı kurum kültürünü almış sağlık çalışanlarının uyumu ile de sunulan hizmetin kalitesinin artabileceği ortak sonucuna varılmıştır.

Kaynakça:

Akın Yılmaz Ö, Bayrakal V, Baskın AH. (2022, Aralık), "Hasta Güvenliğini Etkileyen Risklerin Yönetimi:Dokuz Eylül Üniversitesi Hastanesi Anestezi Yoğun Bakım Birimi Örneği", VIII. Uluslararası Sağlıkta Performans ve Kalite Kongresi, Antalya, ss.130-142

Dokuz Eylül Üniversitesi Risk Yönetimi Yönergesi, 2014

Gür Ö, Bayrakal V, Baskın AH. (2019), "Fatsa Devlet Hastanesi Ameliyathane Biriminde Risk Analizi Yöntemiyle Ramak Kala Olayların belirlenmesi ve Değerlendirilmesi", *Türkiye Klinikleri Adli Tıp ve Adli Bilimler Dergisi*, 16, No 3., sayfa 174-82

Hamarat T, Bayrakal V, Baskın AH. (2022, Aralık), "Sağlıkta Risk Yönetimi Çerçevesinde Sağlık Hizmetlerinde Mahremiyet ve Yeni Medya İlişkilerinde Bilgi Güvenliği: Öncül Bir Araştırma", VIII. Uluslararası Sağlıkta Performans ve Kalite Kongresi, Antalya, ss.201-218

İlaç Güvenliği Rehberi, Versiyon 2.0, Sağlık Bakanlığı Sağlıkta Kalite ve Akreditasyon Daire Başkanlığı, 2015

Romero A., Managing Medicines in the Hospital Pharmacy: Logistics Inefficiencies, Proceedings of the World Congress on Engineering and Computer Science,Vol II WCECS, 2013

Sağlıkta Kalite Standardı – Versiyon 6.1, Sağlık Bakanlığı Sağlıkta Kalite ve Akreditasyon Daire Başkanlığı, 2020

BÖLÜM VI

"Clarke Error Grid" Analizinin Sürekli Glukoz İzlem Cihazlarının Doğruluğunda Kullanımı

Memiş BOLACALI[1]

Giriş

Teknolojinin ilerlemesi ile birlikte diyabet hastaları arasında son yıllarda Sürekli Glukoz İzlem (SGİ) cihazları kullanımı giderek yaygınlaşmaktadır. SGİ cihazları son 2000'li yıllarda keşfedilmesine rağmen bu cihazların kullanımında ve sonuçlarında bazı problemler olduğu bilinmektedir. Bu problemlerden en önemlisi ise; SGİ cihazlarının sonuçları ile parmaktan kan glukozu sonuçları arasında farklıkların olmasıdır. Özellikle SGİ cihazlardan elde edilen sonuçlar, diyabet hastalarını etkilemektedir. SGİ cihaz sonuçlarındaki hatalara ait analizinin yapılması, bu cihazların klinik doğruluğu bakımından büyük önem arz etmektedir. SGİ cihazlarının

[1] Doç. Dr., Kırşehir Ahi Evran Üniversitesi, Tıp Fakültesi, Temel Tıp Bilimleri Bölümü, Biyoistatistik ve Tıp Bilişimi Anabilim Dalı, Kırşehir, Türkiye. ORCID ID: 0000-0002-4196-2359. bolacali@gmail.com

değerlendirilmesinde doğruluk/kesinlik terimi kullanılmaktadır. SGİ cihazlarının doğruluğu klinik ve analitik olarak iki şekilde yapılabilmektedir. Klinik doğruluk; ortalama performans doğruluğuna veya uç değerlerin sayısına dayanır. Analitik doğruluk ise (mean absolute relative difference= MARD), belli sayıda ve değişik glukoz düzeylerinde parmak kan glukoz değerleri ile SGİ cihaz ölçüm değerleri arasındaki ortalama farkı göstermektedir (Jernelv ve ark., 2019; Anand ve ark., 2020; Alsunaidi ve ark., 2021; Besen & Dervişoğlu, 2021).

SGİ cihaz sonuçlarındaki herhangi bir hata; sadece sapma yüzdesi değil, hem referans hem de ölçülen değerlerin mutlak değerine bağlıdır. Üstelik bu bağımlılık herhangi bir basit matematiksel ilişkiyle kolayca tanımlanamamaktadır. Bu nedenle, doğrusal regresyon analizi veya korelasyon katsayıları gibi standart basit istatistiksel yöntemler, hastaların SGİ cihaz sonuçlarındaki hataları izleme veya değerlendirme açısından yetersiz kalmaktadır. Bu bağlamda, 1987 yılında geliştirilen Clarke Error Grid Analizi (CEA)'nde; korelasyon katsayısı, linear regresyon, yüzde sapma ve ortalama farklılıklar dikkate alınarak 70-180 mg/dl glukoz değerine ilişkin standart referans aralığı kullanılarak geliştirilmiş bir yöntemdir. CEA, SGİ cihazının klinik doğruluğunu, kullanılan referans standart ölçüm cihazından elde edilen kan glukoz değerine göre karşılaştırarak sonuçların klinik olarak kabul edilebilir olup olmadığını gösteren bir yöntem olarak karşımıza çıkmaktadır. SGİ sonuçları ile referans ölçüm cihazın sonuçlarının karşılaştırıldığı biyoistatistik analiz yöntemlerinden biri olarak kabul edilen CEA temel prensibi; referans glikoz değerlerini χ eksenine ve etkinliği/doğruluğu test edilecek kan glukoz ölçüm cihazından elde

edilen sonuçları ise γ ekseninde işaretleyerek, verilerin dağılım grafiği oluşturmaya dayanmaktadır (Clarke ve ark., 1987; Poirier ve ark., 1998; Hackett & McCue, 2010).

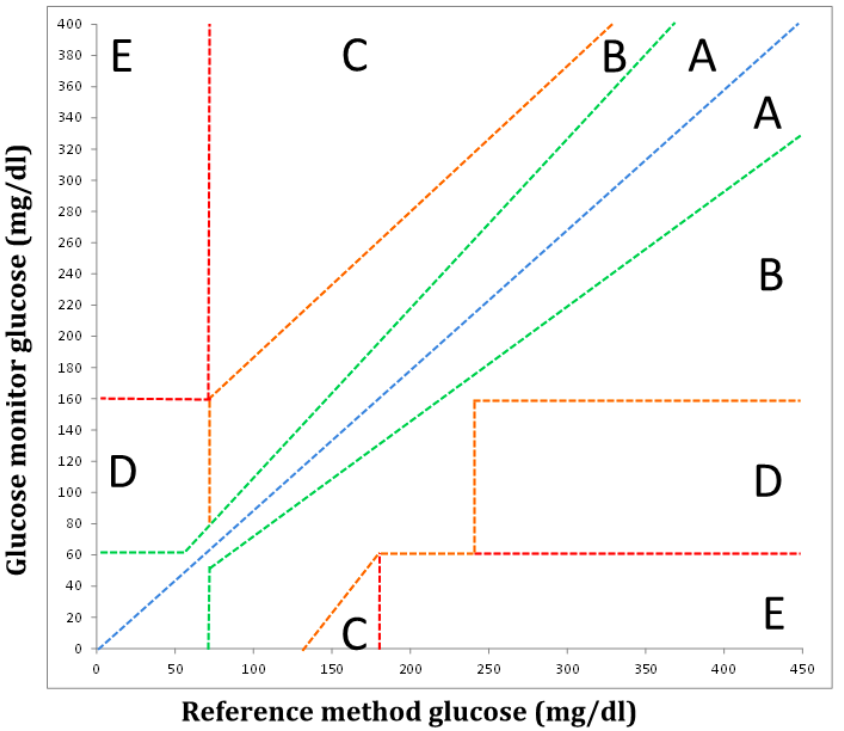
CEA'nde verilerin dağılım grafiğini klinik öneme sahip A'dan E'ye kadar değişen 5 farklı bölgeye ayrılmıştır (Clarke ve ark., 1987; Jernelv ve ark., 2019; Alsunaidi ve ark., 2021; Bachache ve ark., 2022). Grafik üzerindeki bölgelerden;

A bölgesindeki değerler; SGİ cihaz değerinin referans kan glukoz değerinden %20 veya daha az farklı olduğunu diğer bir ifadeyle hatanın olmadığını göstermektedir. Bu aralığa düşen değerlerin "Klinik Açıdan Doğru" olduğu ve doğru tedavi kararları alınmasını sağlayacak verileri ifade edilmektedir.

B bölgesindeki değerler; SGİ cihaz değerinin ölçüm değerinin referans kan glukoz seviyesi değerinden %20'den daha fazla sapan ve hiç tedavi gerektirmeyen veya klinik karar hatalarına yol açmayan kısaca "iyimser - iyi huylu" değerleri temsil etmektedir. Bu bölgelerdeki hatalar büyük olasılıkla önemli hipoglisemiye veya hiperglisemiye neden olmamaktadır. Kısaca; A ve B bölgelerinde gösterilen kan glukoz sonuçları klinik olarak kabul edilebilir sonuçları göstermektedir.

Şekil 1: Microsoft Excel'de Dağılım Grafiği ve Eklenen Çizgilerle Oluşturulmuş Clarke Error Grid Analiz Örneği

Kaynak: Mondal & Mondal, 2020



A ve B bölgeleri arasında kalan sonuçların yüzde dağılım oranı ne kadar yüksek olursa, değerlendirilen SGİ cihaz sonuçlarının klinik açıdan doğruluğunun o kadar yüksek olduğu sonucunu göstermektedir. Örneğin; piyasaya sürülmek için geliştirilen SGİ glukoz ölçüm cihazının sonuçları; elde edilen dağılım grafiğini üzerindeki A ve B bölgelerinde gösterilen verilerin toplam verilerin yüzdesi %95'den daha fazla olması durumunda bu sonuçların klinik olarak kabul edilebilir olduğunun bir kanıtını ifade etmektedir

(Pfützner ve ark., 2013; Bailey ve ark., 2015; Chekima ve ark., 2022; Alam ve ark., 2023; Hammour ve Mandic, 2023).

C, D ve E bölgelerinde işaretlenen sonuçlar, SGİ cihaz sonucu ile referans değerler arasında giderek artan ve potansiyel olarak olumsuz etkileri olan farklılıkları temsil etmektedir. Grafik üzerindeki bölgelerden;

C bölgesindeki değerler, "aşırı düzeltme" hatalarını içeren varileri temsil eder. Bu bölgedeki değerler, kan glukoz düzeyinin 70 mg/dl'nin altına düşmesine veya 180 mg/dl'nin üzerine çıkmasına neden olabilir ve kan şekeri seviyelerinin hedef aralığın dışına çıkmasına neden olabilecek yanlış tedavi kararları alınmasına neden olabilmektedir.

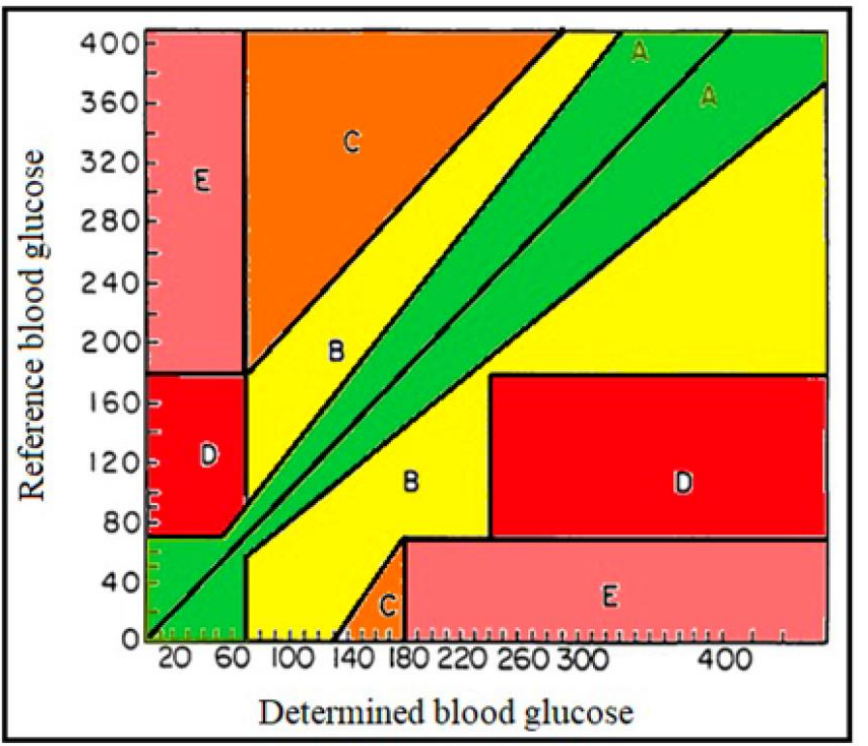
D bölgesindeki değerler, hataların "tedavi edilemeyen" verileri temsil etmektedir. Başka bir deyişle, referans değerleri hedefin dışındayken doğruluğu test edilen SGİ cihazın sonuçları hedef aralık içinde olduğunu göstermektedir. Bu nedenle klinik karar alırken gerçekte yüksek veya düşük kan şekeri düzeyleri dikkatten kaçabilir.

E bölgesindeki değerler, "Hatalı Sonuçlar"a ait verileri temsil etmektedir. Bu bölge içindeki değerler doğruluğu test edilen SGİ ölçüm cihazın sonuçları, referans değerlerinin tersini göstermektedir. Bu nedenle cihazın sonuçlarına göre verilen klinik kararlar, olması gerekenin tam tersi kararların verilmesine neden olmaktadır. Örneğin; gerçekte referans kan glukoz değeri yüksek olduğunda ek insülin verilmesi gerekirken, doğruluğu test edilen SGİ cihazın düşük kan glukoz sonucu vermesi hastanın ilave glukoz almasına veya gerçekte referans kan glukoz değeri düşük olması

gerekirken doğruluğu test edilen SGİ cihazın yüksek kan glukoz sonucuna bağlı olarak hastaya ilave insülin verilmesine neden olmaktadır.

Şekil 2. Bölgeleri renklendirilmiş Clarke Error Grid Grafik Örneği

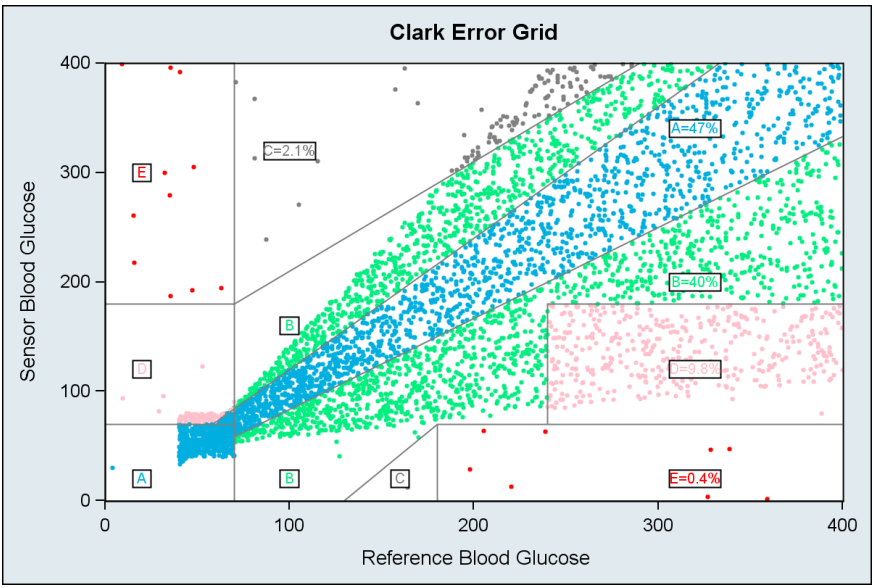
Kaynak: Bachache ve ark., 2022



Kısaca, C, D ve E bölgelerinde işaretlenen değerleri, değerlendirilen sonuçlar ile referans değerler arasında giderek artan ve kuvvetle muhtemel olumsuz etkileri olan farklılıkları temsil etmektedir (Clarke ve ark., 1987; Guillot ve ark., 2020; Scott ve ark., 2018; Costa ve ark., 2020; Bowman & Nichols, 2020; Xue ve ark., 2022). Sonuçları test edilen SGİ cihaz verilerinin, CEA analiz grafiği

üzerinde en az %95'i A ve B bölgelerinde olması; cihazın doğru ölçüm yaptığı ve klinik açıdan kullanılabilir olduğu sonucuna varılmasını sağlamaktadır. (Clarke ve ark., 1987; Beardsall ve ark., 2013)

Şekil 3. Verilerin İşaretlendiği Clarke Error Grid Analiz Örneği
Kaynak: SAS, 2024



Sonuç

Doğruluğu test edilen SGİ cihaz verilerinin analize ait grafik üzerinde en az %95'i A ve B bölgelerine düşmesi; cihazın doğru ölçüm yaptığı ve klinik açıdan kullanılabilir olduğu sonucuna varılmasını sağlamaktadır.

Kaynaklar

Alam, D. S., Rokhana, R., & Susetyoko, R. (2023, October). Non-Invasive Measurement of Blood Glucose Levels Using Linear Regression and Clark Error Grid Analysis Methods. In 2023 Sixth International Conference on Vocational Education and Electrical Engineering (ICVEE) (pp. 184190). IEEE.

Alsunaidi, B., Althobaiti, M., Tamal, M., Albaker, W., & AlNaib, I. (2021). A review of noninvasive optical systems for continuous blood glucose monitoring. Sensors, 21(20), 6820.

Anand, P. K., Shin, D. R., & Memon, M. L. (2020). Adaptive boosting based personalized glucose monitoring system (PGMS) for non-invasive blood glucose prediction with improved accuracy. Diagnostics, 10(5), 285.

Bachache, L. N., AlNeami, A. Q., & Hasan, J. A. (2022). Error grid analysis evaluation of noninvasive blood glucose monitoring system of diabetic Covid19 patients. International Journal of Nonlinear Analysis and Applications, 13(1), 36973706.

Bailey, T., Bode, B. W., Christiansen, M. P., Klaff, L. J., & Alva, S. (2015). The performance and usability of a factory-calibrated flash glucose monitoring system. Diabetes technology & therapeutics, 17(11), 787-794.

Besen, D. B., & Dervişoğlu, M. (2021). Diyabet Yönetiminde Teknoloji Kullanımı. Turkiye Klinikleri Journal of Internal Medicine, 6(2), 80-85.

Bowman, C. F., & Nichols, J. H. (2020). Comparison of accuracy guidelines for hospital glucose meters. Journal of Diabetes Science and Technology, 14(3), 546552.

Chekima, K., Wong, B. T. Z., Noor, M. I., Ooi, Y. B. H., Yan, S. W., & Chekima, B. (2022). Use of a continuous glucose monitor to determine the glycaemic index of ricebased mixed meals, their effect on a 24 h glucose profile and its influence on overweight and obese young adults' meal preferences. Foods, 11(7), 983.

Clarke, W. L., Cox, D., Gonder-Frederick, L. A., Carter, W., & Pohl, S. L. (1987). Evaluating clinical accuracy of systems for self-monitoring of blood glucose. Diabetes care, 10(5), 622-628.

Costa, D., Lourenço, J., Monteiro, A. M., Castro, B., Oliveira, P., Tinoco, M. C., ... & Rolanda, C. (2020). Clinical performance of flash glucose monitoring system in patients with liver cirrhosis and diabetes mellitus. Scientific Reports, 10(1), 7460.

Guillot, F. H., Jacobs, P. G., Wilson, L. M., Youssef, J. E., Gabo, V. B., Branigan, D. L., ... & Castle, J. R. (2020). Accuracy of the Dexcom G6 glucose sensor during aerobic, resistance, and interval exercise in adults with type 1 diabetes. Biosensors, 10(10), 138.

Hackett, E. S., & McCue, P. M. (2010). Evaluation of a veterinary glucometer for use in horses. Journal of Veterinary Internal Medicine, 24(3), 617621.

Hammour, G., & Mandic, D. P. (2023). An linear PPG-based blood glucose monitor: A proof of concept study. Sensors, 23(6), 3319.

Jernelv, I. L., Milenko, K., Fuglerud, S. S., Hjelme, D. R., Ellingsen, R., & Aksnes, A. (2019). A review of optical methods for continuous glucose monitoring. Applied Spectroscopy Reviews, 54(7), 543-572.

Mondal, H., & Mondal, S. (2020). Clarke error grid analysis on graph paper and microsoft excel. Journal of diabetes science and technology, 14(2), 499-499.

Pfützner, A., Klonoff, D. C., Pardo, S., & Parkes, J. L. (2013). Technical aspects of the Parkes error grid. Journal of Diabetes Science and Technology, 7(5), 1275-1281.

Poirier JY, Prieur NL, Campion L, et al. Clinical and statistical evaluation of self-monitoring blood glucose meters. Diabetes Care 1998;21: 1919–1924.

SAS, (2024). SAS istatistik paket programında Clark Error Grid Grafiği Çizimi. https://blogs.sas.com/content/graphicallyspeaking/2013/05/09/clark-error-grid-graph/ Erişim Tarihi: 2024.04.19.

Scott, E. M., Bilous, R. W., & Kautzky-Willer, A. (2018). Accuracy, user acceptability, and safety evaluation for the FreeStyle Libre flash glucose monitoring system when used by pregnant women with diabetes. Diabetes technology & therapeutics, 20(3), 180-188.

Xue, Y., Thalmayer, A. S., Zeising, S., Fischer, G., & Lübke, M. (2022). Commercial and scientific solutions for blood glucose monitoring—a review. Sensors, 22(2), 425.v

BÖLÜM VII

Investigating Health System Performance of OECD Countries by Using CRITIC and ARAS-G methods

Pakize YİĞİT[1]

1. Introduction

Decision-making may be defined as making the most optimal choice among existing alternatives according to known criteria. Accurate and rapid decision-making has become essential for all organizations in today's highly competitive global environment to identify opportunities, set goals, save limited resources, and enhance flexibility and productivity (Chakraborty, Raut, Rofin, & Chakraborty, 2023). For this reason, Multi-Criteria Decision-Making (MCDM) techniques provide decision-makers the ability to assist decision-makers in finding objective or appropriate solutions. MCDM techniques face the challenge of selecting the optimal option among several potential solutions based on several criteria. Since

[1] Dr.Öğr.Üye.., İstanbul Medipol Üniversitesi, Tıp Fakültesi, Biyoistatistik ve Tıp Bilişimi Bölümü, İstanbul/Türkiye, Orcid: 0000-0002-5919-1986,pyigit@medipol.edu.tr

there is frequently no option that dominates the others on all criteria, decision-makers typically search for a satisfying answer (Baležentis & Baležentis, 2014; Hafezalkotob, Hafezalkotob, Liao, & Herrera, 2019).

Healthcare and health policy problems need to have an optimal solution. There are several MCDA approaches providing healthcare decision-makers with transparent and consistent solutions (Marsh, Lanitis, Neasham, Orfanos, & Caro, 2014; Mühlbacher & Kaczynski, 2016). Chakraborty et al.(Chakraborty et al., 2023) categorize MCDM methods in healthcare decision making 11 headings: supply chain management, health information system, medical device and material selection, risk management, logistics, operations management in healthcare, miscellaneous, disease identification and treatment, quality in healthcare, waste management, COVID-19.

Health is crucial for long-term economic growth and the sustainability of nations (Green, 2007). There are three essential elements that cover healthcare, enhance health quality, and are valid for all individuals: health services, the concept of sustainable development (SD), and the health system, which varies according to countries (Pınarbaşı & Piyal, 2021). The issue of health has gained a place of constant and increasing importance in sustainable development, and it has been proven by SD indicators that there is no development without health (Tezcan, 2020).

Comparative Health Systems aims to study and compare different healthcare systems across countries or regions. MCDM methods use the comparison of health systems advancing healthcare

quality, accessibility, and equity globally. In this context, the study investigates and compares the healthcare system performance of OCED countries between 2011 and 2021 using hybrid MCDM models.

Grey additive ratio assessment (ARAS-G) applies grey values to evaluate and rank alternatives and to compare alternative scores with the best feasible alternative (Turskis & Zavadskas, 2010). By adapting the gray numbers to the ARAS method, an aggregate scoring is obtained for the period examined. With the ARAS-G method, instead of obtaining separate rankings for the 11-year period in the evaluation of the health systems of the countries, a total ranking is obtained. The CRITIC method is one of the weighting methods that uses correlations and standard deviations of criteria to have objective weights. CRITIC considers the conflicting relationships between the indicators and the contrast intensity of the attributes that the entropy method includes (Diakoulaki, Mavrotas, & Papayannakis, 1995). The technique is used to find objective weights for indicators to assess a country's health system.

The rest of this paper is organized as follows: Section 2 introduces the data. Section 3 describes the methods. Results are presented in section 4. The last section describes the conclusion and discussion.

2. Data

The data contains 38 OECD countries, creating a set of 11 indicators for a period of 11 years (2011-2021). It uses open databases: OECD stat and World Bank (OECD, 2024; The World Bank, 2024b). The latest available data for most indicators is from

2021. On the other hand, the latest available data for deaths from chronic diseases is 2019(The World Bank, 2024a).

The most commonly used indicators in the literature to examine the health systems of countries are examined (Atun et al., 2013, p. 65; Hadad, Hadad, & Simon-Tuval, 2013, p. 253; Popescu, Militaru, Cristescu, Vasilescu, & Matei, 2018, p. 1; Proksch, Busch-Casler, Haberstroh, & Pinkwart, 2019, p. 169). On a country basis, variables with a limited number of years reported in 11 years are excluded. For example, the number of physicians and nurses, which are common variables in the literature, are excluded from the study because they are reported for only the last four years for the USA. In addition, the country average is used for missing variables on a year-by-period basis. The indicators are:

- Infant mortality rate (IMR)
- Maternal mortality rate (MMR)
- Under the age of 5 mortality rate (U5R)
- Tuberculosis incidence (per 1000) (TI)
- Mortality from CVD, cancer, diabetes or CRD (ages 30-70) (CMR)
- Life expectancy (LE)
- GDP per capita (GDP)
- Current health expenditure per capita (all functions), PPP (CHE)
- Health Expenditures as % of GDP (HGDP)

- Government expenditure on health (GHE)
- Out-of-pocket health expenditures (OHE)

3. Methodology

The study aims to investigate OECD countries' health system performances and compare them using MCDM methods. Firstly, the Critic Method is applied to find the relative importance of the indicators used in the study. Secondly, ARAS-G method is used to rank the countries.

3.1. Criteria Importance Through Intercriteria Correlation (CRITIC)

CRITIC is a kind of MCDM objective weighting method introduced by Diakoulaki et al.(Diakoulaki et al., 1995, p. 763). This approach suggests calculating objective weights by accounting for the correlation between the criteria. It also uses the criteria's standard deviations after normalizing them between [0,1]. The method uses the following stages(Diakoulaki et al., 1995, pp. 764–765; Esra Aytaç Adalı & Ayşegül Tuş Işık, 2017, p. 90):

Step 1: The decision matrix obtains.

$$X = [x_{ij}]_{mxn} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{m1} \end{bmatrix} (i = 1,2..,m \; and \; j = 1,2,\dots n) \quad (1)$$

Step 2: The decision matrix is normalized according to equation 2:

$$X_{ij}^{*} = \frac{X_{ij} - \min(X_{ij})}{\max(X_{ij}) - \min(X_{ij})}; (i = 1,2 \dots, m \; alternatives, j = 1,2, \dots, n \; criteria). \quad (2)$$

Step 3: Both the criterion's standard deviation and its correlation with other criteria are considered when calculating the weights of the criteria (w_j).

$$w_j = \frac{C_j}{\sum_{j=1}^{n} C_j} \quad (3)$$

where Cj is the amount of data in the jth criterion (information measure), which is defined as

$$C_j = \sigma_j \sum_{j'=1}^{n} (1 - r_{jj'}) \quad (4)$$

σ_j is the standard deviation of jth criteria and $r_{jj'}$ is a correlation.

The correlation coefficient between the two criteria is denoted by $r_{jj'}$, and σ_j represents the standard deviation of the jth criterion. The higher value of C_j means the higher relative importance of the criteria.

3.2. Grey additive ratio assessment (ARAS-G)

ARAS method states that the utility function value, which determines the complex relative efficiency of a feasible alternative, is directly proportional to the relative effect of values and weights of the primary criteria examined in a project. The ARAS method was developed by Zavadskas and Turkskis in 2010 (Edmundas Kazimieras Zavadskas & Turskis, 2010, p. 159). Turskis and Zavadskas developed the ARAS-G method by integrating gray numbers into the ARAS method (Turskis & Zavadskas, 2010, pp. 597–600). Grey set theory is a field of study that examines the associations between different features (criteria) in an MCDM problem. It has proven useful for analyzing systems with imperfect information and incomplete samples(Chatterjee, Zavadskas, Roy, & Kar, 2018, p. 348). The steps of the ARAS method is given

below(Turskis & Zavadskas, 2010, pp. 600–602; Yildirim & Adiguzel Mercangoz, 2020, pp. 34–35):

Step 1: The grey decision-making matrix is formed. In this context, "m" refers to the number of options and "n" refers to the number of criteria. The symbol ⊗xij represents the gray number that indicates the performance value of the ith alternative in relation to the jth criterion. Lastly, ⊗x0j represents the optimal value of the jth criterion.

$$\otimes X = \left[\left[\otimes x_{ij}\right]\right]_{mxn} = \begin{bmatrix} [\otimes x_{01}] & [\otimes x_{02}] & \dots & [\otimes x_{0n}] \\ [\otimes x_{11}] & [\otimes x_{12}] & \dots & [\otimes x_{1n}] \\ [\otimes x_{21}] & [\otimes x_{22}] & \dots & [\otimes x_{2n}] \\ \vdots & \vdots & \dots & \vdots \\ [\otimes x_{m1}] & [\otimes x_{m2}] & \dots & [\otimes x_{mm}] \end{bmatrix} \quad (5)$$

$$i = 0,1, \dots m. \quad j = 1,2, \dots ., n$$

If optimal value of j criterion is unknown,it is j calculated as:

$$\otimes x_{0j} = \max_i \otimes x_{ij} \quad \text{if } \max_i \otimes x_{ij} \text{ is preferable} \quad (6)$$

$$\otimes x_{0j} = \min_i \otimes x_{ij} \quad \text{if } \min_i \otimes x_{ij} \text{ is preferable} \quad (7)$$

Step 2: The grey decision matrix is normalized:

$$\otimes \bar{X} = \left[\left[\otimes \bar{x}_{ij}\right]\right] = \begin{bmatrix} [\otimes \bar{x}_{01}] & [\otimes \bar{x}_{02}] & \dots & [\otimes \bar{x}_{0n}] \\ [\otimes \bar{x}_{11}] & [\otimes \bar{x}_{12}] & \dots & [\otimes \bar{x}_{1n}] \\ [\otimes \bar{x}_{21}] & [\otimes \bar{x}_{22}] & \dots & [\otimes \bar{x}_{2n}] \\ \vdots & \vdots & \dots & \vdots \\ [\otimes \bar{x}_{m1}] & [\otimes \bar{x}_{m2}] & \dots & [\otimes \bar{x}_{mn}] \end{bmatrix} \quad (8)$$

$$i = 0,1, \dots m. \quad j = 1,2, \dots ., n$$

If the optimal value is maximum, the normalization formula is as follows:

$$\otimes \bar{x}_{ij} = \frac{\otimes x_{ij}}{\sum_{i=0}^{m} x_{ij}}$$

(9)

$$\otimes x_{ij} = \frac{1}{\otimes x^{*}{}_{ij}};\ \otimes \bar{x}_{ij} = \frac{\otimes x_{ij}}{\sum_{i=0}^{m} \otimes x_{ij}} \qquad (10)$$

Step 3:Normalized weighted grey decision matrix obtained:

$$\otimes \hat{X} = \left[\left[\otimes \hat{x}_{ij}\right]\right]_{mxn} = \begin{bmatrix} [\otimes \hat{x}_{01}] & [\otimes \hat{x}_{02}] & \dots & [\otimes \hat{x}_{0n}] \\ [\otimes \hat{x}_{11}] & [\otimes \hat{x}_{12}] & \dots & [\otimes \hat{x}_{1n}] \\ [\otimes \hat{x}_{21}] & [\otimes \hat{x}_{22}] & \dots & [\otimes \hat{x}_{2n}] \\ \vdots & \vdots & \dots & \vdots \\ [\otimes \hat{x}_{m1}] & [\otimes \hat{x}_{m2}] & \dots & [\otimes \hat{x}_{mm}] \end{bmatrix}$$

(11)

$i = 0,1, \dots m.\quad j = 1,2, \dots ., n$

$$\otimes \hat{x}_{ij} = \otimes \bar{x}_{ij} . \otimes w_j \qquad i = 0,1, \dots m$$

(12)

Step 4: Optimality function is determined:

$$\otimes S_i = \sum_{j=1}^{n} \otimes \hat{x}_{ij} \qquad (13)$$

⊗Si represents is the grey value of the optimality function of ith alternative. The performance level of alternatives can be assessed based on the ⊗Si value.

The grey value of $\otimes S_i$ transforms crisp value as follows:

$$S_i = \frac{1}{2}(\otimes S_i(\underline{1}, \overline{u}), \quad i = 0,1,..m$$

(14)

Step 5: the utility degree K_i is calculated as given below:

4. $$K_i = \frac{S_i}{S_o}, \qquad i = 0,1,..m \qquad (15)$$

S_i and S_0 are obtained with Equation 14.

The estimated values K_i fall within the range of [0; 1] and can be arranged in ascending order, which represents the desired order of precedence. The relative efficiency of the reasonable alternative can be evaluated by evaluating the utility function values.

4. Application and Results

A hybrid method is used to investigate the health system performance of OECD countries using CRITIC and ARAS-G approaches. There are 11 indicators (criteria). The indicators of optimal values of IMR, MMR, U5R, TI, and CMR are minimum, meaning the lower values are better. On the other hand, the optimal values of LE, GDP, CHE, HGDP, and GHE are maximum, meaning the higher numbers are better.

For CRITIC method, firstly, the decision matrix is obtained and normalized (Equations 1 and 2). After that, the standard deviation for each feature is calculated, as shown in Table 1.

Table 1: Standard Deviation of Indicators

Indicators	IMR	MMR	U5R	TI	CMR	OHE	LE	HGDP	CHE	GHE	GDP
Standard Deviation	0.1613	0.1466	0.1955	0.1475	0.2191	0.1164	0.2349	0.1590	0.1664	0.1641	0.1918

The standard deviation shows the dispersion of the data. The higher values mean higher variability and the lower values mean lower variability. For this reason, higher standard deviation criteria tend to have higher importance, and small standard deviations tend to have lower importance. In this study, life expectancy and mortality from chronic diseases have the highest dispersion. Then,

the correlation matrix and Cj (information measure of indicators, Equation 4) are calculated. Lastly, the indicator's weights are calculated using Equation 3. The Cj values and weights of indicators are listed in Table 2.

Table 2: Weights of the Indicators

Indicators	CMR	LE	U5R	GDP	TI	IMR	MMR	GHE	CHE	HGDP	OHE
C_j	2.6761	2.4566	2.0166	1.7761	1.6451	1.6327	1.4637	1.3975	1.3794	1.3545	0.9969
Weights	14.24%	13.07%	10.73%	9.45%	8.75%	8.69%	7.79%	7.44%	7.34%	7.21%	5.30%

Mortality from chronic diseases has the highest weight (14%), followed by life expectancy (13%) and under-five mortality (11%). The weights for indicators derived from the CRITIC approach are next applied as the input in the ARAS-G method.

A gray decision matrix is first created to apply the ARAS-G method. The grey numbers are found from the data, which has 11 years (2011-2021). The gray decision matrix is defined using the maximum and minimum values for each indicator for each country. The grey decision matrix is given in Table 3. Secondly, the grey decision matrix is normalized using Equations 8-10. Then, the normalized matrix is weighted using the CRITIC weights (Formulas 11-12). Next, Equation 13 is applied to the weighted normalized grey matrix, and the grey numbers are transformed to crisp ones using Equation 14. Finally, the benefit degree (Ki) is obtained using Formula 15. the rankings of the countries are assigned.

The benefit degree values and rankings are reported in Table 4. The rankings show that Sweden is ranked first, with the utility score of 65.7 %, followed by Norway, with the utility score of 64%. Iceland, Switzerland, and the Netherlands are ranked 3rd,4th, and

5th with the utility score of 63.1%, 59.4%,55.9%, respectively. Türkiye ranked as 37th.

Table 3: Countries' health system ranking performed by ARAS-G

Country	K_i	Rank	Country	K_i	Rank
Sweden	0.65749972	1	Greece	0.45634212	20
Norway	0.63985471	2	New Zealand	0.45081377	21
Iceland	0.6317307	3	Canada	0.43857054	22
Switzerland	0.59413292	4	Slovenia	0.43856459	23
Netherlands	0.55916732	5	United Kingdom	0.42273484	24
Finland	0.55624636	6	France	0.42213901	25
Denmark	0.55207739	7	Poland	0.41505874	26
Israel	0.5439182	8	Slovak Republic	0.39902612	27
United States	0.53988164	9	Estonia	0.38297576	28
Ireland	0.53839478	10	Korea	0.37196512	29
Austria	0.53543854	11	Portugal	0.36603831	30
Italy	0.53433595	12	Hungary	0.32006494	31
Luxembourg	0.51939104	13	Lithuania	0.3146317	32
Japan	0.50737568	14	Latvia	0.2941343	33
Australia	0.50177859	15	Costa Rica	0.28943909	34
Germany	0.49548272	16	Chile	0.28341196	35
Czechia	0.47473288	17	Colombia	0.26381528	36
Spain	0.46125094	18	Türkiye	0.25853635	37
Belgium	0.45757744	19	Mexico	0.20299018	38

Table 4: Grey Decision Matrix

Country	IMR		MMR		U5R		TI		CMR		OHE		LE		HGDP		CHE		GHE		GDP	
Australia	3.1	3.8	1.9	5.2	3.8	4.5	6.1	7.2	8.6	9.8	1173	1638	81.9	83.3	8.54	10.68	3809	6226	2603	4588	49919	68198
Austria	2.7	3.6	1.3	8.6	3.3	4.2	4.9	9.2	10.4	12.2	1103	1449	81.1	82	10.03	12.10	4345	6690	3242	5241	44196	53518
Belgium	2.9	3.8	2.4	8.5	3.8	4.4	7.7	10	10.6	12.6	964	1348	80.5	82.1	10.36	11.20	4054	6022	3089	4674	41008	51850
Canada	4.4	5	4.8	8.6	5	5.7	4.8	5.9	9.6	10.8	1256	1700	81.6	82.3	10.26	13.04	4217	6278	2962	4578	42316	52669
Chile	5.6	7.7	9	21	6.6	8.5	14	18	10	12.7	550	998	79	81	6.80	9.73	1355	2677	806	1679	13174	16241
Colombia	16.5	21.2	45.3	83.2	12.8	17.8	31	41	9.7	10.5	190	331	74.5	76.8	6.746	9.018	792	1532	601	1201	5304	8264
Costa Rica	7.7	9.1	16.1	40.5	7.9	10.2	9.4	14	9.5	10.3	311	426	79.1	80.8	7.045	8.112	1176	1671	858	1244	9137	12669
Czechia	2.2	2.8	1.9	10.1	2.8	3.3	3.9	6.2	14.3	17.4	343	584	77.2	79.3	7.369	9.489	2228	4303	1885	3719	17830	26823
Denmark	2.1	3.6	1.6	5.3	3.6	4.1	3.8	7.4	10.8	12.9	687	943	79.9	81.6	10.098	10.824	4266	6372	3580	5429	53255	69269
Estonia	1.4	3.6	6.3	14.4	2	4.2	9.3	25	14.9	19.1	355	730	76.6	79	6.078	7.579	1466	3074	1111	2343	17402	27944
Finland	1.7	2.4	1.7	10.9	2.3	2.9	3.5	6.7	9.6	11.7	806	1062	80.6	82.1	9.043	10.25	3598	5252	2793	4190	42802	53505
France	3.5	3.9	7.6	8.9	4.1	4.3	7.6	9.1	10.6	11.8	829	1094	82.1	83	11.089	12.308	4162	6106	3163	5178	36653	43846
Germany	3	3.6	2.8	4.7	3.6	4.1	5	7.7	12.1	12.8	771	1094	80.6	81.3	10.778	12.934	4567	7518	3795	6424	41103	51427
Greece	2.9	4.2	1.8	11.3	4	4.3	2.1	5.4	12.5	13.9	724	1033	80.2	81.9	7.891	9.504	2011	2736	1158	1700	17617	25484
Hungary	3.3	5	5.5	25.8	4	5.9	3.7	15	22.1	24.5	588	757	74.3	76.5	6.282	7.481	1756	2749	1157	1992	12717	18753
Iceland	0.7	3.3	3.3	22.3	2.6	2.7	2.1	4.4	8.5	9.7	662	834	82.1	83.2	8.07	9.734	3312	5107	2650	4273	45996	74452
Ireland	2.8	3.6	1.5	6.2	3.2	4.1	4.8	9.8	9.7	11.9	1170	1325	80.9	82.8	6.709	10.606	4186	5861	2980	4536	48944	102002
Israel	2.4	3.6	1.2	8.1	3.4	4.4	2.1	7.7	8.8	10	751	1004	81.8	82.9	6.839	7.905	2082	3258	1303	2221	33156	52130
Italy	2.3	2.9	1.2	3.5	2.8	3.9	4.2	7.3	9	10.5	712	992	82.3	83.6	8.656	9.625	3037	4043	2291	3051	30242	38650
Japan	1.7	2.3	2.7	4.8	2.3	3.2	11	20	8.3	9.4	608	739	82.7	84.6	10.488	11.299	3741	4899	3133	4198	34961	49145

Korea	2.4	3	7.8	17.2	2.9	4	44	101	7.3	10	792	1579	80.6	83.6	5.915	9.331	1936	4189	1144	2610	25098	35142
Latvia	2.7	6.6	4.8	55.2	3.5	7.2	16	54	21.6	25.2	393	953	73.1	75.7	5.409	9.045	1078	3122	684	2169	13339	20930
Lithuania	2.8	4.8	3.3	14.2	3.5	5.6	26	65	19.3	23.7	429	1048	73.7	76.5	6.126	7.823	1485	3336	1054	2287	14264	23850
Luxembourg	2.5	4.7	16.4	32.4	2.7	2.8	5	9.8	9.5	11.3	691	814	81.1	82.7	5.071	5.952	4425	6274	3666	5397	105462	133712
Mexico	12.3	15.3	34.2	58.6	13.3	18.5	21	25	15.2	16	430	629	74.7	75.4	5.383	6.222	912	1262	482	649	8895	11490
Netherlands	3.3	3.8	1.2	5.3	4	4.3	4.1	6.7	10.3	12.2	792	1017	81.2	82.2	10.02	11.294	4567	6739	3776	5722	45193	58728
New Zealand	4.2	5.7	1.7	16.8	4.7	6.1	6.8	8	10.3	11.9	599	969	81	82.3	8.97	10.125	3132	4921	2533	3953	38388	49996
Norway	1.6	2.6	1.8	4.9	2.2	3.1	3	8.2	8.7	10.8	773	1017	81.4	83.3	8.722	11.249	4965	7043	4192	6026	68340	103554
Poland	3.6	4.7	1	2.5	4.4	5.7	9.4	24	17	20.1	436	695	75.5	78	6.265	6.583	1497	2522	1061	1828	12379	18050
Portugal	2.4	3.4	4.5	20.1	3.2	3.7	16	27	11	12.2	871	1409	80.6	81.9	9.306	11.137	2421	3830	1485	2421	19250	24661
Slovak Republic	4.5	5.8	1.8	9.9	5.9	6.7	2.8	8.5	15.5	19.7	396	584	74.6	77.8	6.666	7.751	1974	2522	1457	2011	16391	21768
Slovenia	1.6	2.9	4.8	9.2	2.4	3.1	4.1	10	11.4	13.3	632	1020	80.1	81.6	8.185	9.478	2377	3885	1745	2865	20890	29331
Spain	2.5	3.1	1.7	4.2	3.1	3.7	7.2	16	9.6	10.9	725	1160	82.4	84	8.945	10.746	2728	4087	1963	2926	25754	31678
Sweden	1.8	2.7	0.9	7	2.5	3	3.5	9.1	8.4	10.1	692	877	81.8	83.2	10.416	11.333	4460	6228	3768	5351	51545	61418
Switzerland	3.1	3.9	1.2	8.5	4	4.5	4.7	7.9	7.9	9.4	1840	2338	82.8	84	10.046	11.801	5229	7582	3329	5135	82153	93446
Türkiye	8.5	11.25	13	15.8	10.1	16.8	14	24	15.6	17.3	185	331	74.6	78.6	4.117	4.653	888	1560	703	1229	8639	12578
United Kingdom	3.8	4.2	4.5	6.7	4.2	5	7	15	10.3	12	606	928	80.4	81.4	9.596	12.365	3417	5467	2811	4539	40217	47440
United States	5.4	6.1	15.05	23.8	6.3	7.2	2.3	3.9	13.6	14.3	1557	4309	76.4	78.9	15.992	18.756	8081	12197	3943	10201	50066	70219

A comparative analysis is also performed to determine the reliability of the results. Country rankings are found for each year using the ARAS method. COPRAS-G ((Complex proportional assessment of alternatives with grey relations) method is also applied (E.K. Zavadskas & Kaklauskas, 1996; Edmundas Kazimieras Zavadskas, Kaklauskas, Turskis, & Tamošaitiene, 2008, p. 85). Turkey is also ranked 37th according to the COPRAS-G method. When evaluated year by year according to the ARAS method, Türkiye ranks 36th in the first six years of analysis (2011-2016) and 37th in the last five years of the examining period (2017-2021).

Spearman's correlation coefficient analysis assesses the reliability of ranking in the unknown domain by comparing the ranking results obtained from alternatives for every possible pair of MCDM methods (Antucheviciene, Zakarevicius, & Zavadskas, 2011; Chatterjee et al., 2018). They present at Table 6. The Spearman rank correlation between ARAS-G and COPRAS-G ranking results is very high (0.945). In addition, the relationships between yearly rankings with ARAS and ARAS-G results are all very high (r>0.90). It shows the reliability of the performing results.

Table 5: Spearman Rank Correlations of the Ranking Methods

Methods	**COPRAS-G**	**ARAS 2011**	**ARAS 2012**	**ARAS 2013**	**ARAS 2014**	**ARAS 2015**	**ARAS 2016**	**ARAS 2017**	**ARAS 2018**	**ARAS 2019**	**ARAS 2020**	**ARAS 2021**
ARAS-G	.945**	.952**	.902**	.930**	.935**	.934**	.935**	.949**	.948**	.942**	.954**	.959**

**p<0.001

4.Discussion and Conclusion

The study evaluates the health system performance of OECD countries according to selected health indicators. The data contains 38 OECD countries and 11 indicators between 2011 and 2021. The study uses CRITIC and ARAS-G methods to compare and rank the countries.

To begin with, the indicator's weights are assigned according to the CRITIC method. The CRITIC method determines the relative importance of the criteria using their standard deviations and correlations. For this reason, the method provides an objective and unbiased way to find the significance of the indicators. The most important indicators are mortality from chronic diseases, life expectancy, and the under-five mortality rate.

After that, the ARAS-G method is used to rank the countries. The indicator's weights obtained from the CRITIC technique are used in the ARAS-G method to evaluate countries. In this stage, the time dimension is converted to grey numbers. Thus, instead of examining the performance of countries separately over 11 years, a single score is obtained for the study period. Sweden, Norway, Iceland, Switzerland, and the Netherlands are the best-performing countries, respectively, while Mexico, Turkey, Colombia, Chile, and Costa Rica are the worst.

The proposed methodology is also compared with COPRAS-G and ARAS (each period separately) methods. When pairwise comparisons are examined, it is determined that there are strong relationships between the methods. Thus, it can be concluded that

the methodology is reliable for investigating the studied MCDM problem.

Although the study presents an original perspective to the literature, it has several limitations. Firstly, health system performance is evaluated with restricted indicators for only OECD countries. More advanced studies can be performed with more countries and indicators. In addition to this, using different MCDM approaches may provide different results. Future studies can be conducted by increasing the number of countries and variables and using newly developed MCDM methods.

References

Antucheviciene, J., Zakarevicius, A., & Zavadskas, E. K. (2011). Measuring congruence of ranking results applying particular MCDM methods. *Informatica*, *22*(3), 319–338. https://doi.org/10.15388/informatica.2011.329

Atun, R., Aydin, S., Chakraborty, S., Sümer, S., Aran, M., Gürol, I., … Akdağ, R. (2013). Universal health coverage in Turkey: Enhancement of equity. *The Lancet*, *382*(9886), 65–99. https://doi.org/10.1016/S0140-6736(13)61051-X

Baležentis, T., & Baležentis, A. (2014). A Survey on Development and Applications of the Multi-criteria Decision Making Method MULTIMOORA. *Journal of Multi-Criteria Decision Analysis*, *21*(3–4), 209–222. https://doi.org/10.1002/mcda.1501

Chakraborty, S., Raut, R. D., Rofin, T. M., & Chakraborty, S. (2023). A comprehensive and systematic review of multi-criteria decision-making methods and applications in healthcare. *Healthcare Analytics*, *4*(May), 100232. https://doi.org/10.1016/j.health.2023.100232

Chatterjee, K., Zavadskas, E. K., Roy, J., & Kar, S. (2018). Performance evaluation of green supply chain management using the grey DEMATEL–ARAS model. In *Springer Proceedings in Mathematics and Statistics* (Vol. 225). Springer Singapore. https://doi.org/10.1007/978-981-10-7814-9_24

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). *D E T E R M I N I N G OBJECTIVE WEIGHTS IN M U L T I P L E*

CRITERIA PROBLEMS : THE CRITIC M E T H O D. *22*(7), 763–770.

Esra Aytaç Adalı, & Ayşegül Tuş Işık. (2017). Critic and Maut Methods for the Contract Manufacturer Selection Problem. *European Journal of Multidisciplinary Studies* , *2*(5), 88–96.

Green, A. (2007). *An introduction to health planning for developing health systems (3rd ed.)* (3rd ed.). London: Oxford University Press.

Hadad, S., Hadad, Y., & Simon-Tuval, T. (2013). Determinants of healthcare system's efficiency in OECD countries. *European Journal of Health Economics*, *14*(2), 253–265. https://doi.org/10.1007/s10198-011-0366-3

Hafezalkotob, A., Hafezalkotob, A., Liao, H., & Herrera, F. (2019). An overview of MULTIMOORA for multi-criteria decision-making: Theory, developments, applications, and challenges. *Information Fusion*, *51*(November 2018), 145–177. https://doi.org/10.1016/j.inffus.2018.12.002

Marsh, K., Lanitis, T., Neasham, D., Orfanos, P., & Caro, J. (2014). Assessing the value of healthcare interventions using multi-criteria decision analysis: A review of the literature. *PharmacoEconomics*, *32*(4), 345–365. https://doi.org/10.1007/s40273-014-0135-0

Mühlbacher, A. C., & Kaczynski, A. (2016). Making Good Decisions in Healthcare with Multi-Criteria Decision Analysis: The Use, Current Research and Future Development of MCDA. *Applied*

Health Economics and Health Policy, *14*(1), 29–40. https://doi.org/10.1007/s40258-015-0203-4

OECD. (2024). OECD Stat.

Pınarbaşı, Ş., & Piyal, B. (2021). Sürdürülebilir Kalkınma Hedefi Üç'ün Sağlık Kapsayıcılığı İşlevi. *ESTÜDAM Halk Sağlığı Dergisi*, *7*(2), 0–3.

Popescu, M. E., Militaru, E., Cristescu, A., Vasilescu, M. D., & Matei, M. M. M. (2018). Investigating health systems in the European Union: Outcomes and fiscal sustainability. *Sustainability (Switzerland)*, *10*(9), 1–28. https://doi.org/10.3390/su10093186

Proksch, D., Busch-Casler, J., Haberstroh, M. M., & Pinkwart, A. (2019). National health innovation systems: Clustering the OECD countries by innovative output in healthcare using a multi indicator approach. *Research Policy*, *48*(1), 169–179. https://doi.org/10.1016/j.respol.2018.08.004

Tezcan, N. (2020). Sürdürülebilir Kalkınma Amaçları Kapsamında Türkiye'de Sağlık Göstergelerinin Analizi. *İstanbul Ticaret Üniversitesi Sosyal Bilimler Dergisi*, *19*(Prof. Dr. Sabri ORMAN Özel Sayısı), 202–217. Retrieved from https://dergipark.org.tr/en/pub/iticusbe/issue/56181/768478

The World Bank. (2024a). Mortality from CVD, cancer, diabetes or CRD between exact ages 30 and 70 (%). Retrieved from https://genderdata.worldbank.org/en/indicator/sh-dyn-ncom-zs?gender=total#data-table-section

The World Bank. (2024b). World Bank Group Data. Retrieved June 19, 2023, from https://www.worldbank.org/en/home

Turskis, Z., & Zavadskas, E. K. (2010). A novel method for multiple criteria analysis: Grey Additive Ratio Assessment (ARAS-G) method. *Informatica*, *21*(4), 597–610. https://doi.org/10.15388/informatica.2010.307

Yildirim, B. F., & Adiguzel Mercangoz, B. (2020). Evaluating the logistics performance of OECD countries by using fuzzy AHP and ARAS-G. *Eurasian Economic Review*, *10*(1), 27–45. https://doi.org/10.1007/s40822-019-00131-3

Zavadskas, E.K., & Kaklauskas, A. (1996). Multicriteria Evaluation of Building (Pastatų sistemotechninis įvertinimas). *Vilnius: Technika.*

Zavadskas, Edmundas Kazimieras, Kaklauskas, A., Turskis, Z., & Tamošaitiene, J. (2008). Selection of the effective dwelling house walls by applying attributes values determined at intervals. *Journal of Civil Engineering and Management*, *14*(2), 85–93. https://doi.org/10.3846/1392-3730.2008.14.3

Zavadskas, Edmundas Kazimieras, & Turskis, Z. (2010). A new additive ratio assessment (ARAS) method in multicriteria decision-making. *Technological and Economic Development of Economy*, *16*(2), 159–172. https://doi.org/10.3846/tede.2010.10

BÖLÜM VIII

Application of Cox Regression and Decision Tree Models to Heart Disease Data.

Zeynep YILMAZ[1]
Özlem GÜRÜNLÜ ALMA[2]

GİRİŞ

Doku ve hücrelerin ihtiyaç duydukları oksijen ve besin maddelerinin vücuda yeterli miktarda pompalanması kalbin en önemli görevidir. Kalbin yeterli miktarda kanı pompalayamaması durumunda kalp yetmezliği ortaya çıkar. Kalp yetmezliği durumunda kalp çalışır ancak yeterli ve ihtiyaç duyulan miktarda kanı pompalayamaz. Kan dolaşımının yetersizliğinden dolayı bazı doku ve damarlarda kan birikir. Buna "konjestif" denir. Kalp

[1] Öğr.*Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey,* Orcıd:-

[2] Prof.Dr., *Muğla Sıtkı Koçman University, Department of Statistics, Muğla, Turkey*,, Orcıd: 0000-0001-6978-9810

yetmezliği hastalarının çoğunluğunda meydana geldiğinden dolayı bu rahatsızlığa genellikle "KKY" denir *(Narin, A., İşler, Y., & Özer, M., 2014).*İlk kez Thomas Lewis tarafından "kalbin muhteviyatının yeterince boşalamaması durumu" şeklinde tanımlanan kalp yetmezliğinin birçok farklı tanımı mevcut olmakla birlikte Paul Wood da kalp yetmezliğini kalbin vücudun ihtiyacı için gerekli olan dolaşımı sağlayamaması şeklinde tanımlamıştır *(Köse, G., Kurutkan, M. N., & Orhan, F. 2020).*

Kalp yetmezliği, hastaların yaşam kalitesini olumsuz etkilemektedir. Kalp yetmezliği olan hastaların günlük yaşam aktivitelerini yerine getirmede güçlük yaşadıkları, ekonomik, cinsel ve psikososyal sorunlarının olduğu, özellikle iş yaşamında, aile, arkadaş ilişkilerinde sorunlarla karşı karşıya kaldıkları saptanmıştır *(Oksel, E., Akbıyık, A., & Koçak, G., 2016).* Bunun yanı sıra, bireyin yaşam kalitesini etkileyen en önemli değişkenlerden biri olan uyku kalitesi; ilerleyen yaş, kronik hastalıklar ve buna bağlı gelişen solunum bozuklukları neticesinde azalmaktadır. Kalp yetmezliğinde fonksiyonel kapasitenin sınırlanmış olması ve çok sayıda semptomun görülmesi gibi önemli stresörler uykuyu olumsuz yönde etkilemektedir *(GÖKÇE, S., & MERT, H.,2015).*

Günümüzde sağlık sektörü alanında hastalıkların tespiti için makine öğrenmesi yöntemleri etkin bir şekilde kullanılmakta olup, literatürde kalp hastalıkları için yapılmış çalışmalar bulunmaktadır. Bu çalışmalardan bazıları: Aktaş Potur E. ve Enginel N.,2021 yılında kalp yetmezliği hastalarının sağ kalımlarının sınıflandırma algoritmaları ile tahmin edilmesi üzerinde çalışmıştır. Çoşar M. Ve Deniz E. (2021) makine öğrenimi algoritmaları kullanarak kalp hastalıklarının tespit edilmesi üzerinde çalışmışlardır. Kalp ve damar

hastalığı teşhisinde Frank ve Asuncion, 2010 yılında Yapay Sinir Ağları ve Bayes algoritmaları kullanmıştır. 270 hastadan alınan 13 özellikle kurulan modelde veri setinde kalp hastası olan ve kalp hastası olmayan şeklinde iki sınıfa ayrılmış ve veri setine uygulanan algoritmaların sonuçları performans açısından karşılaştırılmıştır.

Bu çalışmada, Nisan-Aralık (2015) döneminde Kardiyoloji Enstitüsü ve Faisalabad-Pakistan Müttefik Hastanesi'ne başvuran kalp yetmezliği hastalarının hayatta kalma verileri ele alınmıştır (Ahmad T.,Munir A., Haider Bhatti S. , Aftab M., Raza M.A., 2017). Veri seti kalp yetmezliğine yakalanmış 203'ü Ex 96'sı sansürlü olan 299 Pakistanlı bireyin hasta kayıtlarından oluşmaktadır. Hasta bireylerin 105'ini kadınlar, 194'ünü erkekler oluşturmaktadır. Yaş aralıkları ise minimum 40 yaş, maximum 95 yaş olmak üzere ortalama 61 yaş aralığındadır. Bireyler minimum 4 gün, maximum 285 gün ile ortalama 130 gün boyunca izlenmişlerdir. Tablo 1'de değişkenlerin tanımlayıcı istatistikleri görülmektedir. Kalp yetmezliğinin sebep olduğu ölümlerin, aşağıda verilen değişkenler ile anlamlı bir biçimde açıklanıp açıklanmadığını test etmek için Kaplan-Meier hayatta kalma analizi ve Cox Regresyon Analizi kullanılmıştır.

- Anaemia : Kansızlık
- Creatinine_phosphokinase : Kreatinin fosfokinaz, kaslarda bulunan enzim türüdür
- Diabetes : Şeker hastalığı
- Ejection_fraction (EF) : Kalp döngüsünde kanın ne kadarı vücuda pompalandığını gösterir.

- High_blood_pressure : Yüksek tansiyon
- Platelets (PLT) : Trombositler
- Serum_creatinine : Serum kreatinin, kan dolaşımındaki seviyeleri kontrol eder.
- Serum_sodium : Serum sodyum
- Smoking : Sigara kullanımı
- Death_Event : Ölüm olayı

Tablo 1. Değişkenlerin Tanımlayıcı İstatistikleri

	N	Minimum	Maximum	Mean	Std. Deviation
age	299	40	95	60,83	11,895
creatinine_phosphokinase	299	23	7861	581,84	970,288
ejection_fraction	299	14	80	38,08	11,835
platelets	299	25100	850000	263358,03	97804,237
serum_creatinine	299	,5	9,4	1,394	1,0345
serum_sodium	299	113	148	136,63	4,412
time	299	4	285	130,26	77,614
Valid N (listwise)	299				

Değeri	Ölüm Olayı	Anemi	Diabet	Yüksek Kan Basıncı	Cinsiyet	Sigara Kullanımı
0	203(ex)	170	174	194	105	203
1	96 (sansürlü)	129	125	105	194	96

1. KAPLAN – MEİER HAYATTA KALMA ANALİZİ ve SONUÇLARI

Kaplan-Meier Analizi, 1958'de Kaplan ve Meier tarafından yaşam fonksiyonunun tahminine yönelik geliştirilen Product-Limit (PL) bir yöntemidir *(Bardakçi, S., 2017).* Yaşam sürelerine ilişkin

verileri zaman aralıklarına bölmeden yaşam ve ölüm fonksiyonlarını hesaplamada kullanılmaktadır. Bu analiz yöntemiyle, yaşam süresinin tahmin edilmesinde kullanılan yaşam tablolarına benzerlik göstermektedir. Lakin Kaplan-Meier yöntemini yaşam tablolarından ayıran en belirgin fark, aralık tahminleri yerine, olayın gerçekleşmesine kadar geçen süreyi ele almasıdır *(KARAPINAR, D., ve ZORLUTUNA, Ş., 2022).*

Birçok tahmin probleminde tüm bireylerin ölçümlemelerini tamamlamak imkânsızdır. Örneğin tıbbi çalışmalarda, hastaların yaşam sürelerini hesaplarken ölmeden önce izi kaybedilen iletişim kurulamayan ve bazı bilinmeyen nedenlerden ölen hastalar çalışmadan çıkarılmak istenir. Doğru bilgiler elde edebilmek için bu gözlemlerin analiz edilmesi gerekir. Böyle sansürlenmiş gözlemlerin olduğu örneklemlerde, yaşam süresi t'den büyük gözlemlerin sayısı kesin olarak bilinemeyecektir. Bu durum, sansürlemenin olmadığı durumda elde edilen yaşam fonksiyonunun tahmininin sansürlenmenin olduğu duruma uyarlanmasını gerektirir. Bu uyarlama sonucu elde edilen yaşam fonksiyonunun tahmini "Çarpım Limit (Ç-L) Tahmini" ya da "Kaplan-Meier (K-M) Tahmini" olarak bilinir. Kaplan ve Meier 1958 yılında "Product Limit Tahmini" olarak da bilinen bu yöntemi bulmuşlardır. Bu yöntemle sansürlenmiş verilerin olmadığı durumda yaşam fonksiyonunun tahmini;

$$\hat{S}(t) = \frac{\text{t zamanında yaşayan birey sayısı}}{n} = \frac{Nt}{n}$$

Olarak tanımlanır.

Hem product limit hem actuarial tahminleri aşağıdaki genel prosedürleri temel alır:

- ✓ *Zaman ölçeği uygun olarak seçilen aralıklara ayrılır.* $(0, u_1)$, (u_1, u_2)…
- ✓ *Her bir aralık için* (u_{j-1}, u_j), *bir tahmin* $p_j = P_j / P_{j-1}$ (u_{j-1} *'den sonra hayatta kalan öğeler* u_j *'nin ötesinde hayatta kalır) ile tanımlanan koşullu olasılık tahmin edilir.*
- ✓ *t bölümleme noktasıysa, t'nin ötesinde hayatta kalan popülasyondaki P(t) oranı ; t'den önceki tüm aralıklar için tahmin edilen* p_j *'lerin çarpımı olarak tahmin edilir.*

Başlangıçta, hiçbir zaman aralığının aynı zamanda sansürleme ve ölüm durumunu aynı anda içermediği kabul edilmektedir.

$P = Pr(T > u_j) = S(u_j)$, $j=1, 2…k$

$j=0$ *için,* $S(u_0) = S(0) = P_0 = 1$ *olur.*

$p_j = \frac{P_j}{P_{j-1}}$ *Olduğundan,*

j=1 için, $p_1 = \frac{P_1}{P_0} = P_1 \rightarrow P_1 = p_1$

j=2 için, $p_2 = \frac{P_2}{P_1} \rightarrow P_2 = p_1 * p_2$

j=3 için, $p_3 = \frac{P_3}{P_2} \rightarrow P_3 = p_1 * p_2 * p_3$

$P_j = \prod_{j=!}^{k} p_j$ *Olarak bulunur.*

$1 * p_1 * p_2 * \ldots * p_j \rightarrow$ *Koşullu Sağkalım (Yaşam) Olasılıklarının Çarpımı*

$S(t) = \prod_{k=1}^{j} P_k \rightarrow$ *Sağkalım (Yaşam) Olasılığı*

$\hat{S}(t) = \prod_{k=1}^{j} \widehat{P_k} \rightarrow$ *Sağkalım (Yaşam) Olasılığının Tahmini*

Koşullu Ölme Olasılığı

$q_j = 1 - p_j$, *j=1,2,…*

$\hat{q}_j = \frac{d_j}{n_j} \rightarrow \widehat{p_j} = 1 - \hat{q}_j = 1 - \frac{d_j}{n_j} = \frac{n_j - d_j}{n_j}$, *j= 1,2,…*

$\hat{S}(t) = \prod_{k=1}^{j} \widehat{P_k} = \prod_{k=1}^{j} (\frac{n_k - d_k}{n_k})$

Bazı durumlarda aynı hastalığa yakalanan bireyler iki ya da daha fazla gruba ayrılarak her gruba farklı bir tedavi yöntemi uygulanabilir. (Örneğin farklı ilaç, farklı ameliyat gibi) Farklı yöntemlerle tedavi edilen hastalar için birden fazla yaşam fonksiyonu hesaplanabilir. Bu yaşam fonksiyonlarının birbiriyle olan farklılıkları test edilebilir.Kaplan – Meier analizinde farklı gruplar arası karşılaştırmada kullanılan bazı özel testler Tarone – Ware Testi, Log – Rank Testi, Breslow – Wilcoxon Testi, Mantel – Cox Testi'dir. Çalışmada yapılan analizler SPSS 20 ve SPSS Clementine 12 programları kullanılarak elde edilmiştir.

Veri seti için Cinsiyete göre Kaplan Meier analizinden elde edilen sonuçlar Tablo 2'de verilmiştir.

Tablo 2. Kaplan-Meier Analiz Sonuçları

sex	Total N	N of Events	Censored	
			N	Percent
0	105	71	34	32,4%
1	194	132	62	32,0%
Overall	299	203	96	32,1%

Means and Medians for Survival Time

sex	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
0	168,433	7,649	153,440	183,425	187,000	6,647	173,971	200,029
1	165,609	5,503	154,824	176,394	186,000	14,743	157,103	214,897
Overall	166,535	4,457	157,799	175,272	186,000	6,056	174,131	197,869

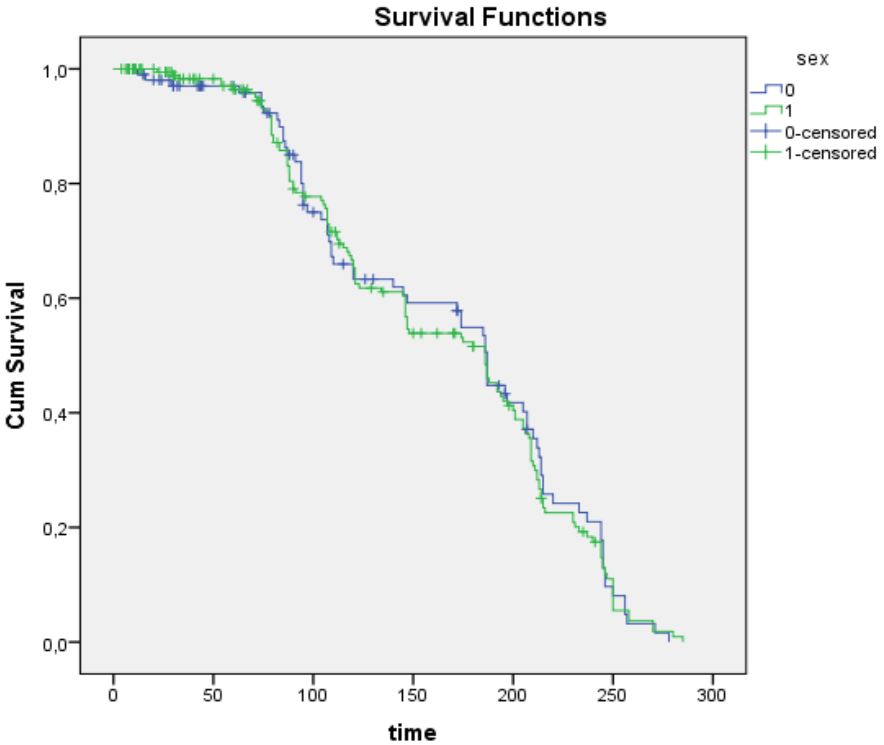
a. Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	,026	1	,872
Breslow (Generalized Wilcoxon)	,058	1	,809
Tarone-Ware	,080	1	,778

Test of equality of survival distributions for the different levels of sex.

Yukarıda verilen Kaplan-Meier çıktısında, cinsiyete (0=Kadın, 1= Erkek) göre analiz yapılmıştır. 105 hasta kadın bireyin 71'i Ex olup, kalan 34'ü ise sansürlü veridir. 194 Hasta erkek bireylerde ise 132'si Ex olmuş, 62'si sansürlü verdir. Her iki grup için gözlemlenemeyen bireylerin yüzdesi %32'dir. Ortalama sağkalım süresi; Kadınlar için 168 gün, erkekler için 166 gündür. %95 güven aralığına bakıldığında ise alt ve üst sınırlar birbirine oldukça yakın olduğundan dolayı, gruplar arasında ciddi farklar olmadığı söylenebilir.



Log-Rank Test (p=0,872) (p=0,05)'ten oldukça büyük olduğundan dolayı, istatistiksel açıdan kadın ve erkek hasta bireyler arasında anlamlı bir farklılık olmadığını gösterir. Yani İki hasta grubun sağkalım süreleri benzerlik gösterir. Sağkalım eğrileri de bu hipotezi destekler nitelikte olup, cinsiyet gruplarının sağkalım olasılıkları izlem süresince eşdeğer olduğunu ifade etmektedir.

Ejection Fraction değişkenine göre Kaplan Meier analiz sonuçları aşağıda verilmiştir.

Ejeksiyon fraksiyon (EF) değeri; %40'ın altında toplam 219 hastadan 142'si Ex, 77'si de sansürlüdür. %40'ın üstünde 80 hastadan 61'i Ex, 19'u da sansürlü hastalardır.

- <%40 : Kalp fonksiyonlarının bozulduğunu ve kalp yetmezliği riskinin arttığını gösterir.
- >%40 : Kalp fonksiyonlarının iyi durumda olduğunu ve kalp yetmezliği riskinin azaldığını gösterir.

EF	Total N	N of Events	Censored	
			N	Percent
40 ve Altı	219	142	77	35,2%
40 Üstü	80	61	19	23,8%
Overall	299	203	96	32,1%

Means and Medians for Survival Time

EF	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
40 ve Altı	175,042	5,211	164,829	185,254	194,000	7,049	180,184	207,816
40 Üstü	145,324	8,049	129,548	161,101	121,000	18,250	85,231	156,769
Overall	166,535	4,457	157,799	175,272	186,000	6,056	174,131	197,869

a. Estimation is limited to the largest survival time if it is censored.

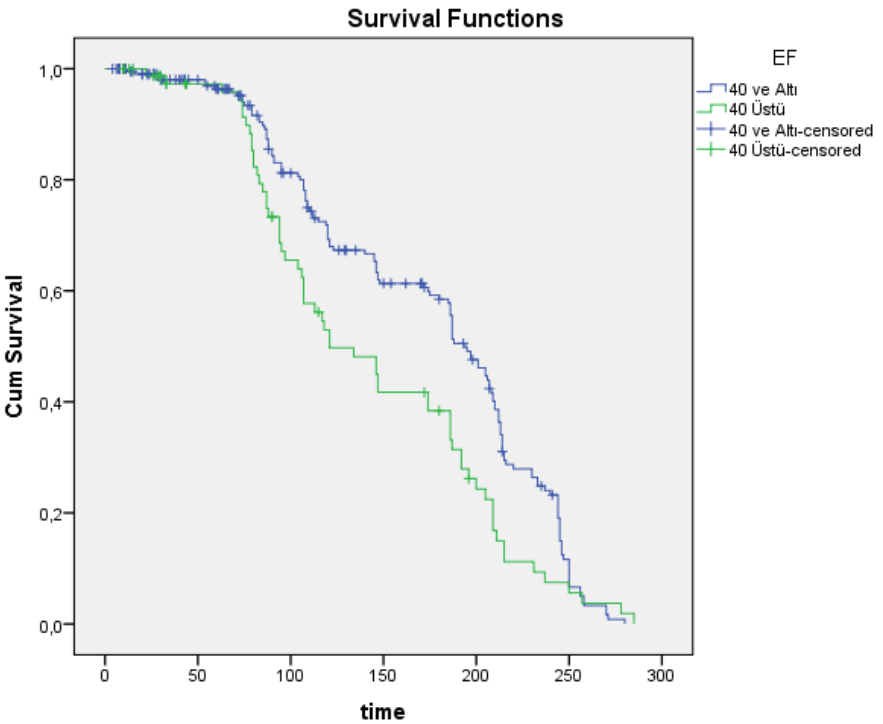
Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	6,777	1	,009
Breslow (Generalized Wilcoxon)	10,539	1	,001
Tarone-Ware	10,514	1	,001

Test of equality of survival distributions for the different levels of EF.

EF için ortalama ve medyan hayatta kalma süresi %95 güven düzeyi ile ölçülmüştür. EF değerleri ≤%40, >%40 olarak gruplandırılmıştır. Sırasıyla ortalama yaşam süreleri 175 gün ve 145 gündür. Bu bireylerin hayatta kalma olasılıkları %50'ye düştüğü süreler: ≤%40 için 194. gün, >%40 için de 121.gündür.

Log-Rank değeri (p=0,009) çıkmıştır. Sonuçlar, Ejeksiyon fraksiyonunun hayatta kalma süreleri üzerinde etkili olduğu istatistiksel olarak anlamlı bulunmuştur.



Anemi değişkenine göre Kaplan Meier analiz sonuçları aşağıda verilmiştir.

Case Processing Summary

anaemia	Total N	N of Events	Censored	
			N	Percent
Yok	170	120	50	29,4%
Var	129	83	46	35,7%
Overall	299	203	96	32,1%

Kansızlığı olmayan 170 hasta bireylerden 120'si Ex olmuş, 50'si ise gözlemlenmemiş sansürlü bireylerdir. Kansızlığı olan 129 bireyden ise 83'ü Ex olup, 46'sı da sansürlü bireyleri oluşturur.

Means and Medians for Survival Time

anaemia	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
Yok	173,366	5,933	161,736	184,995	192,000	5,515	181,190	202,810
Var	156,334	6,611	143,376	169,292	147,000	17,170	113,347	180,653
Overall	166,535	4,457	157,799	175,272	186,000	6,056	174,131	197,869

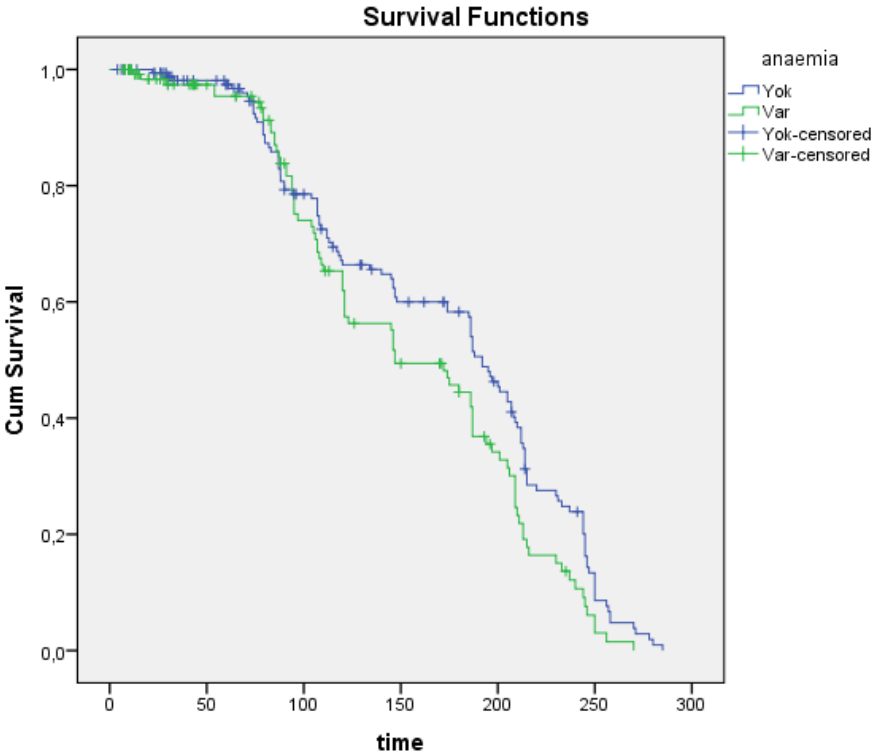
a. Estimation is limited to the largest survival time if it is censored.

Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	5,419	1	,020
Breslow (Generalized Wilcoxon)	2,577	1	,108
Tarone-Ware	3,832	1	,050

Test of equality of survival distributions for the different levels of anaemia.

Yukarıdaki çıktıda, ortalama ve medyan yaşam süreleri %95 Güven Düzeyi ile ölçülmüştür. Kansızlığı olmayan hasta bireylerin yaşam olasılığı %50'ye 192. Günde düşerken, kansızlığı olan hasta bireylerin ise 147. Günde düşmüştür. Ortalama yaşam süresi, kansızlığı olmayan hastalarda 173 gün, kansızlığı olan hastalarda 156 gün olduğu görülür. Verilen istatistiklerden de anlaşılacağı gibi kansızlık, hastaların hayatta kalma sürelerini olumsuz yönde etkilemektedir. Log-Rank değeri (p=0,020) İstatistiksel olarak anlamlı çıkmıştır. Bu da Kansızlığın hayatta kalma süreleri üzerinde önemli bir etkisinin olduğunu göstermektedir. Aşağıda alternatif hipotezin doğruluğu Kaplan-Meier hayatta kalma eğrisiyle görselleştirilmiştir.



Yüksek kan basıncı değişkenine göre Kaplan Meier analiz sonuçları aşağıda verilmiştir.

high_blood_pressure	Total N	N of Events	Censored	
			N	Percent
Yok	194	137	57	29,4%
Var	105	66	39	37,1%
Overall	299	203	96	32,1%

Yüksek tansiyonu olmayan 194 hasta bireyin 137'si Ex olup, 57'si gözlemlenememiş sansürlü veridir. Yüksek tansiyonu olan

hasta bireylerde ise 105 hasta bireyden 66'sı Ex olmuş, 39'u da sansürlü veridir.

Means and Medians for Survival Time

	Mean[a]				Median			
			95% Confidence Interval				95% Confidence Interval	
high_blood_pressure	Estimate	Std. Error	Lower Bound	Upper Bound	Estimate	Std. Error	Lower Bound	Upper Bound
Yok	174,931	5,492	164,167	185,696	192,000	8,567	175,209	208,791
Var	148,620	7,126	134,653	162,587	146,000	31,631	84,003	207,997
Overall	166,535	4,457	157,799	175,272	186,000	6,056	174,131	197,869

a. Estimation is limited to the largest survival time if it is censored.

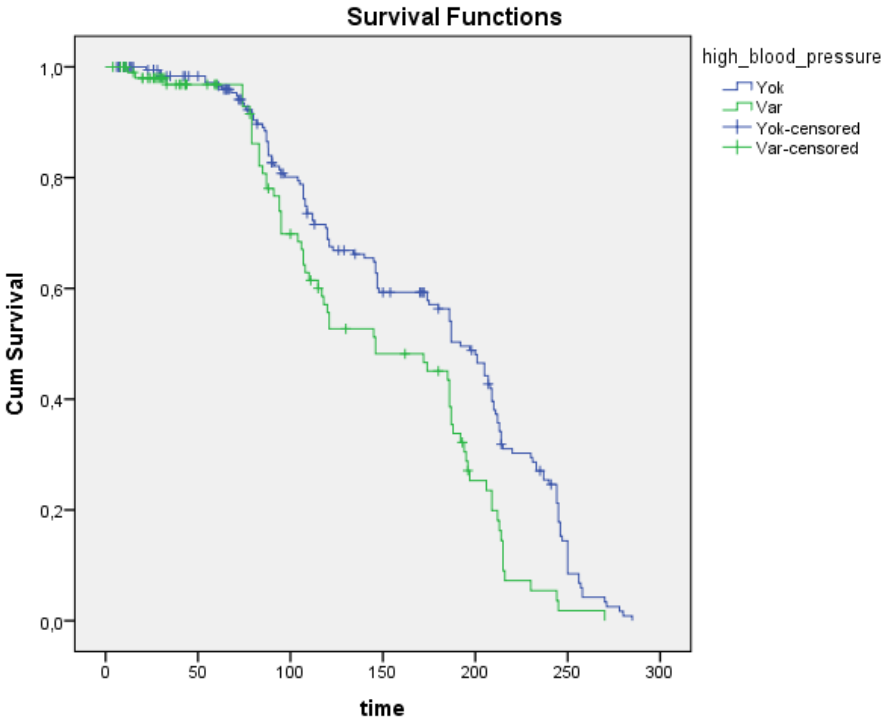
Overall Comparisons

	Chi-Square	df	Sig.
Log Rank (Mantel-Cox)	12,377	1	,000
Breslow (Generalized Wilcoxon)	7,446	1	,006
Tarone-Ware	9,790	1	,002

Test of equality of survival distributions for the different levels of high_blood_pressure.

%95 güven düzeyinde, yüksek tansiyonu olmayan hasta bireylerin ortalama yaşam süresi 175 gün iken yüksek tansiyonu olan bireylerde ortalama yaşam süresi 149 gündür. Bu iki grubun yaşam olasılıkları %50'ye düştüğü zaman noktaları yüksek tansiyonu olmayan hastalarda 192. Günde iken, yüksek tansiyonu olan hastalarda 146. Güne düşmüştür. Log-Rank değeri (p=0,000) istatistiksel olarak anlamlı çıkmıştır. Testlerin sonucunda, Yüksek tansiyonun varlığı, hayatta kalma süreleri üzerinde anlamlı ve etkili olduğu söylenebilir.

Aşağıda verilen Kaplan-Meier hayatta kalma eğrisi ile sonuçlar daha anlaşılır halde görselleştirilmiştir.



2. COX REGRESYON YÖNTEMİ ve SONUÇLARI

Cox regresyon yöntemi, bir hastanın yaşam süresi (bağımlı değişken) ile birden fazla açıklayıcı değişken arasındaki ilişkiyi ortaya çıkaran istatistiksel bir yöntemdir. Cox regresyon analizinde hedef, yaşam verilerinin genel durumunu yansıtacak bir model oluşturmaktır. Bu yolla, yaşam süresi üzerinde etkili olduğu düşünülen bağımsız değişkenlerin etkileri eş zamanlı olarak açıklanabilmekte, yani ölçülebilmektedir. Cox regresyon modeli, klinik bir çalışmada hastaların yaşamını analiz etmek için kullanıldığında, tedavinin etkisini diğer değişkenlerin etkisinden ayırmaya yardımcı olur *(Yay, M., Çoker, E., & Uysal, Ö., 2007).*

Parametrik modellerin gerektirdiği varsayımların (normallik, bağımsızlık vb.) sağlanmadığı durumlarda Cox regresyon analizi, parametrik analizlerden daha etkilidir. Cox regresyon modelinin temel varsayımları şu şekilde açıklanabilir: bağımsız değişkenlerin risk (hazard) fonksiyonu üzerindeki etkileri log-lineerdir ve bağımsız değişkenlerin log-lineer fonksiyonu ile risk fonksiyonu arasındaki ilişki çarpımsaldır *(Arı, A., & Onder, H., 2013).*

Açıklayıcı değişken X vektörüne sahip olan birey ya da birim için t anında hazard fonksiyonu;

$$h\,(t;\,X)\;=\;h_0(t)\;g\;(\,X;\,\beta\,)$$

Biçiminde ifade edilmektedir.

Burada $g\;(\,X;\,\beta\,)$ çeşitli biçimlerde gösterilebilmektedir ve genellikle şeklinde kullanılmaktadır. Buradaki $g(X)$, birey ya da birimlerin özelliklerini yansıtan X vektörünün tehlike fonksiyonu üzerindeki çarpımsal etkisini belirleyen bir fonksiyondur *(Demir, A., 2017).*

Veri seti için Cox regresyon analizi sonuçları Tablo 3 verilmiştir.

Tablo 3. Cox Regresyon Analizi Sonuçları

-2 Log Likelihood	Overall (score)			Change From Previous Step			Change From Previous Block		
	Chi-square	df	Sig.	Chi-square	df	Sig.	Chi-square	df	Sig.
1783,209	27,262	9	,001	25,970	9	,002	25,970	9	,002

a. Beginning Block Number 1. Method = Enter

Variables in the Equation

	B	SE	Wald	df	Sig.	Exp(B)
age	,006	,007	,778	1	,378	1,006
anaemia	-,302	,151	4,008	1	,045	,740
creatinine_phosphokinase	,000	,000	,668	1	,414	1,000
diabetes	,142	,150	,887	1	,346	1,152
ejection_fraction	,014	,007	3,906	1	,048	1,014
high_blood_pressure	-,516	,158	10,638	1	,001	,597
serum_creatinine	-,189	,116	2,644	1	,104	,828
serum_sodium	-,019	,019	1,033	1	,309	,981
sex	-,063	,154	,168	1	,682	,939

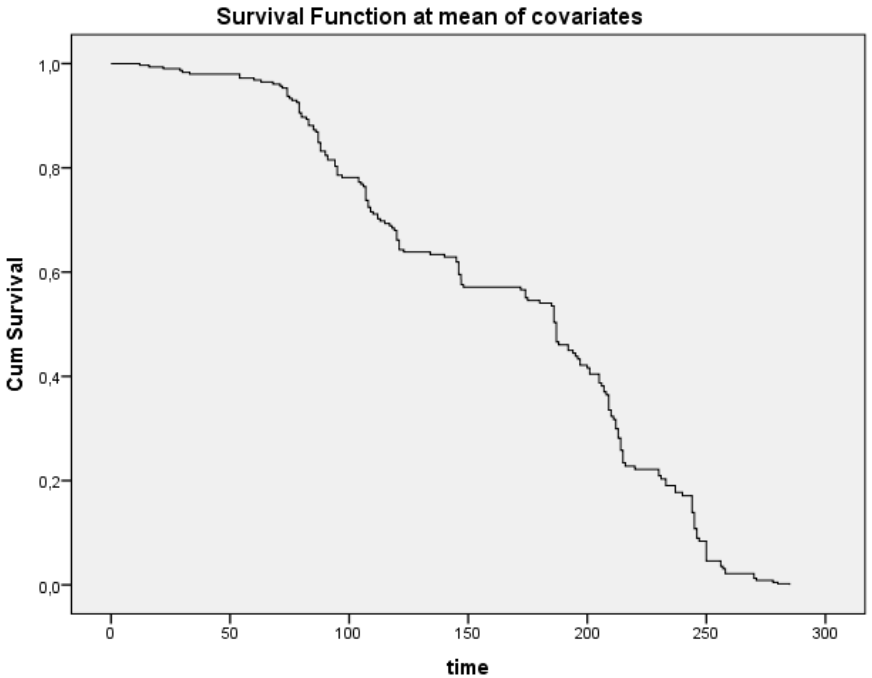
Covariate Means

	Mean
age	60,622
anaemia	,579
creatinine_phosphokinase	571,091
diabetes	,568
ejection_fraction	38,326
high_blood_pressure	,653
serum_creatinine	1,353
serum_sodium	136,796
sex	,361

Model uyumuna bakıldığında; Kalp yetmezliği olan hasta bireylerin Ex(ölüm) olayını Age, Anaemia, Creatinine phosphokinase, Diabetes, Ejection fraction, High blood pressure, Serum creatinine, Serum sodium ve sex değişkenleri tarafından %95 güven düzeyinde (p=0,002) anlamlı bir şekilde açıklamaktadır.

Denklemdeki değişkenler göz önüne alındığında; Anaemia ile Ex arasında negatif yönde β= -0,302 kadar değişim, high blood pressure ile Ex arasında negatif yönde β= -0,516 kadar değişim, Ejection fraction ile Ex arasında pozitif yönde β=0,014 kadar değişim olduğunu göstermektedir. Değişkenlerin anlamlılığı ise

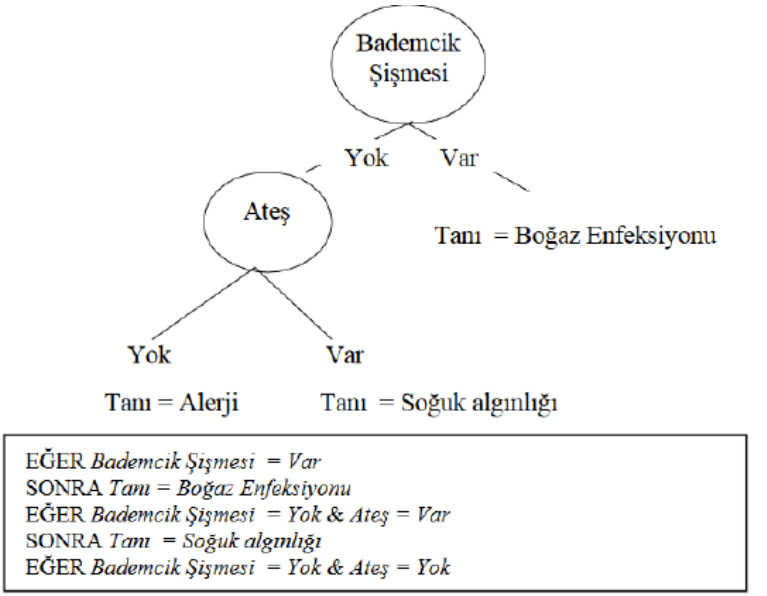
sırasıyla (p=0,045), (p=0,001) ve (p=0,048) bulunmuş olup, p-significance değerleri bu 3 değişken için istatistiksel açıdan anlamlı çıkmıştır. Anaemia, high blood pressure ve Ejection fraction değişkenlerinin kategorileri arasında oluşan farklılıklar aşağıda Hayatta kalma fonksiyonları ile görselleştirilmiştir.



3.KARAR AĞAÇLARI YÖNTEMİ

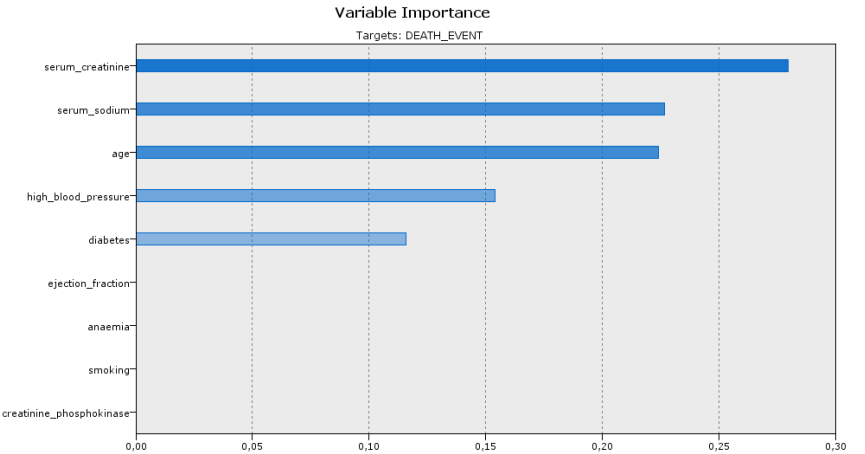
Karar ağaçları, ağaç yapısına benzetilen ve oldukça fazla kullanılan gözetimli öğrenme algoritmalarındandır. Karar ağaçları yapıları itibariyle hiyerarşik yapı olup dallanarak ilerleyen bir yol izlemektedir. Karar ağacı algoritması büyük veri setlerinde bazı karar kuralları uygulayarak daha küçük veri setlerine ayırmak için kullanılır. Karar ağaçları üç elemandan oluşmaktadır. İlk elamanı iç

karar düğümleridir ki giriş verileri bu kısımda test edilir ve bu iç karar düğümleri soruların sorulduğu hangi yöne yöneleceğine karar veren düğümlerdir. İkinci bir karar ağacı elamanı ise daldır. Dallar soruların cevaplarını temsil eder. Üçüncü eleman ise uç yapraklardır ve sınıf etiketleri bu kısımda bulunur. Şekil "de örnek bir karar ağacı ve kural yapısı verilmiştir (Bilekyiğit S., 2022).



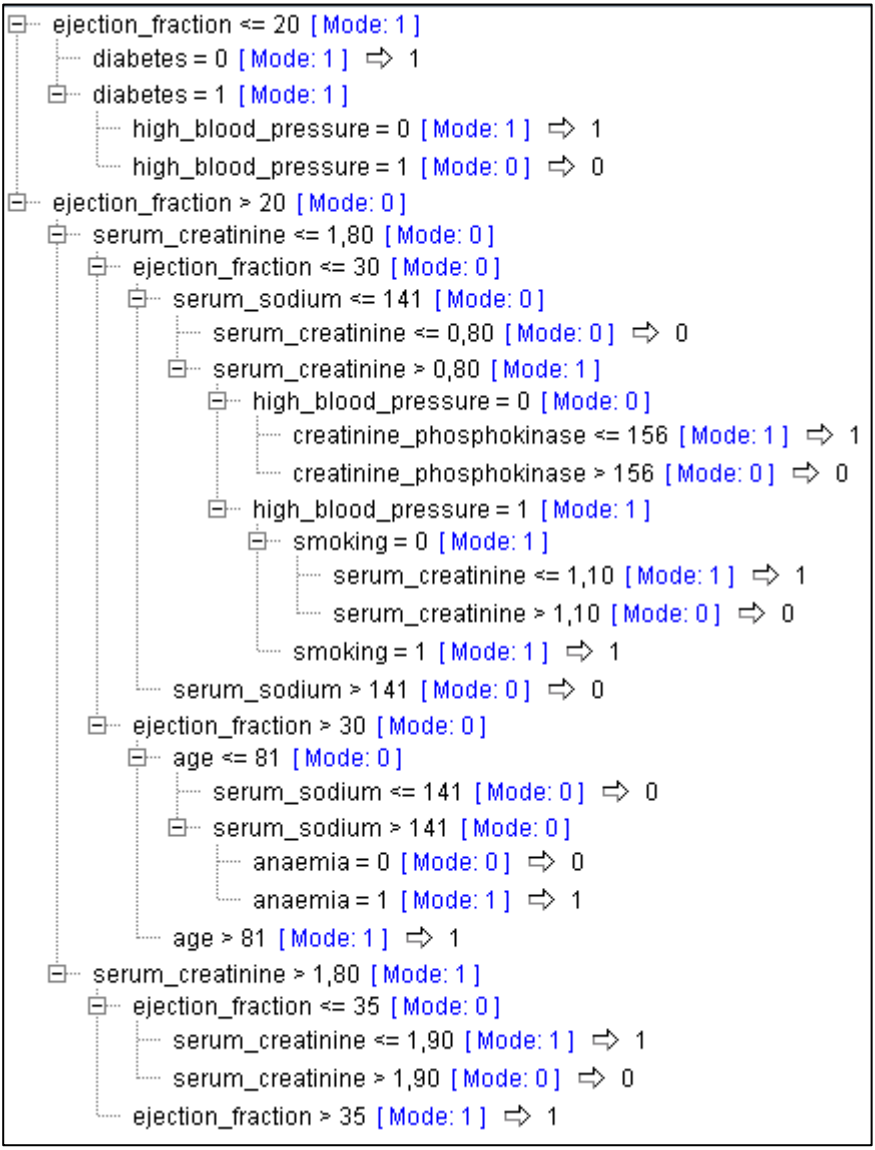
Şekil 1. *Karar Ağacı Yapısı*

Veri seti için uygulanan karar ağacı analizi sonucuna göre: en önemli değişken serum creatinin, serum sodium, yaş, yüksek kan basıncı ve diyabet olarak belirtilmiştir.



Şekil 2. *Karar Ağacı Analizi Sonucu Önemli Değişkenlerin Gösterimi*

Serum creatinine (Serum Kreatinin), ölüm olayı üzerinde büyük bir etkisi olduğundan en önemli değişkendir. Serum sodium (Serum Sodyum), age (Yaş) high_blood_pressure (Yüksek Tansiyon) ve diabetes (Şeker hastalığı) gibi önem derecelerine göre devam eder.

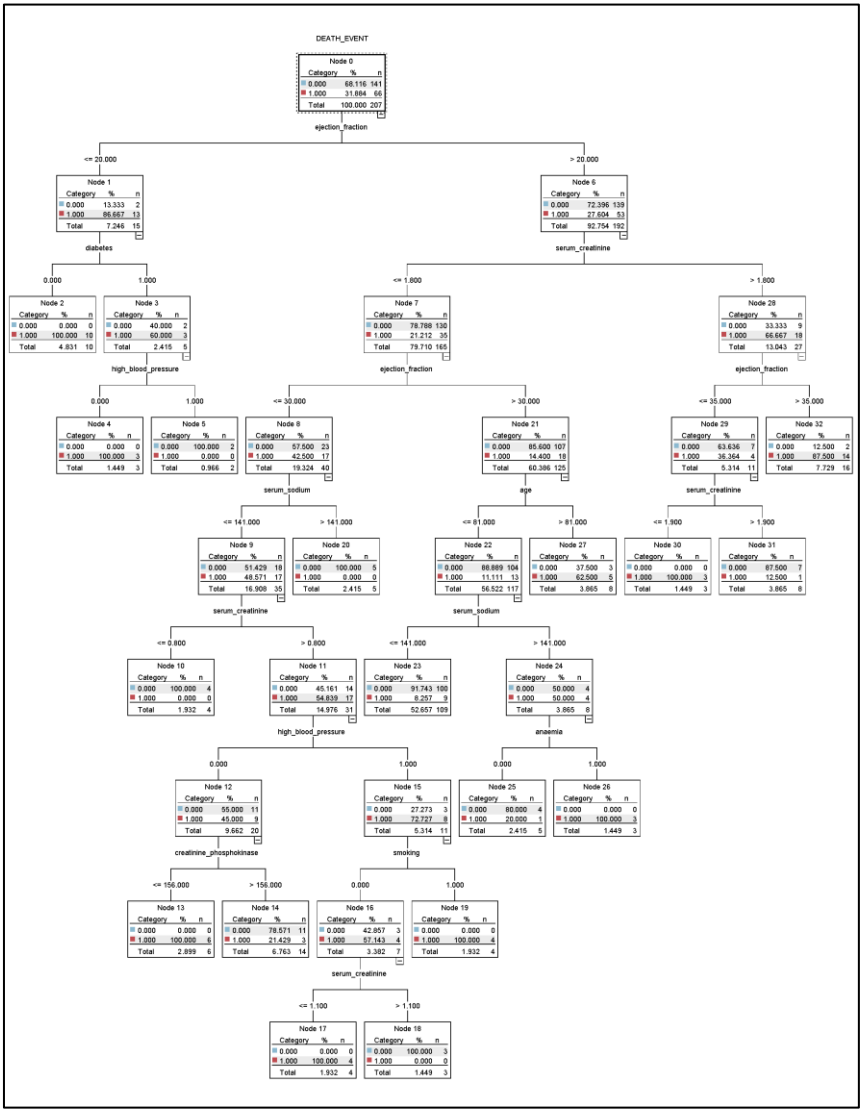


Şekil 3. Karar Ağacı Kural Yapısı

Bu ağaç yapısında; ejeksiyon fraksiyonu 20'den küçükse, ölüm riski %13. Dİyabet yoksa ölüm riski yoktur. Diyabet varsa ve

yüksek tansiyon yoksa, ölüm riski %0. Diyabet varsa ve yüksek tansiyon da varsa, ölüm riski %100. Ejeksiyon fraksiyonu 20'den büyükse, ölüm riski %72. Serum kreatinin ≤ 1.80 ise Ölüm riski %79, büyükse ölüm riski %33.Bu da bize EF değeri büyüdükçe, ölüm riskinin arttığını gösterir.

Bu ağacın ana belirleyicileri ejection fraction ve serum creatinine'dir



Şekil 3. Karar Ağacı Yapısı

- Ejection Fraction ≤ 20: Ölüm riski yüksek.
    - Diyabet varsa, ölüm riski %40.

- Diyabet yoksa, ölüm riski %0.

- Ejection Fraction > 20: Daha karmaşık bir yapı.
    - Serum creatinine ≤ 1.80: Ölüm Riski, %79.
        - Ejection Fraction ≤ 30: Ölüm Riski, %57.
            - Serum sodium ≤ 141: Ölüm Riski, %51.
                - Serum creatinine ≤ 0.80: Ölüm riski %100.
                - Serum creatinine > 0.80: Ölüm Riski, %45. Yüksek tansiyon, kreatinin fosfokinaz, sigara kullanımı ve serum kreatinin seviyelerine bağlı olarak ölüm riski değişir.
            - Serum sodium > 141: Ölüm riski %100.
        - Ejection Fraction > 30: Ölüm Riski, %86. Yaş ve serum sodyum seviyelerine bağlı olarak ölüm riski değişir.
    - Serum creatinine > 1.80: Ölüm riski %33. ejeksiyon fraksiyonu ve serum kreatinin seviyelerine bağlı olarak ölüm riski değişir.

Bu ağacın ana belirleyicileri arasında ejection fraction, serum creatinine ve serum sodium öne çıkıyor. Ayrıca, yaş, diyabet, yüksek tansiyon, anemi ve sigara kullanımı da dikkate alınıyor.

Results for output field DEATH_EVENT

Comparing $C-DEATH_EVENT with DEATH_EVENT

'Partition'	1_Training		2_Testing	
Correct	188	90,82%	65	70,65%
Wrong	19	9,18%	27	29,35%
Total	207		92	

Şekil 4. C5.0 algoritması ile elde edilen modelin doğruluk oranı

Duyarlılık %90, Seçicilik %29, Genel Doğruluk %72 olarak bulunmuştur.

KAYNAKLAR

1. Ahmad T.,Munir A., Haider Bhatti S. , Aftab M., Raza M.A., 2017. Survival analysis of heart failure patients: A case study, Plos One. https://doi.org/10.1371/journal.pone.0181001

2. Aktaş Potur E., Erginel N., 2021. Kalp Yetmezliği Hastalarının Sağ Kalımlarının Sınıflandırma Algoritmaları ile Tahmin Edilmesi. Avrupa Bilim ve Teknoloji Dergisi, Özel Sayı 24, S. 112-118.

3. Arı, A., & Onder, H. (2013). Farkli veri yapilarinda kullanilabilecek regresyon yöntemleri. *Anadolu Tarım Bilimleri Dergisi*, *28*(3), 168-174.

4. Bardakçi, S. (2017). Aktüeryal veri analizinde istatistiksel yöntemlerin kullanımı: Yangın hasarı ve trafik kazası verileriyle bir uygulama

5. Bilekyiğit S., 2022. Kalp Yetmezliği Riskinin Makine Öğrenmesi Yöntemleri ile Analiz Edilmesi. Karamanoğlu Mehmet Bey Üniversitesi, Fen Bilimleri Enstitüsü Yüksek Lisans Tezi.

6. Çoşar M. Ve Deniz E. (2021) Makine Öğrenimi Algoritmaları Kullanarak Kalp Hastalıklarının Tespit Edilmesi, Avrupa Bilim ve Teknoloji Dergisi, Özel Sayı 28, S. 1112-1116.

7. Demir, A. *Tütün kullanımını bıraktıktan sonra tütüne yeniden başlama süresini etkileyen faktörlerin cox regresyon ile analizi* (Master's thesis, Sosyal Bilimler Enstitüsü).

8. Frank, A., ve Asuncion, A. (2010). UCI Machine Learning Repository. Retrieved

9. from http://archive.ics.uci.edu/ml

10. GÖKÇE, S., & MERT, H. (2015). Kalp Yetmezliği Olan Hastaların Uyku Kalitesi ve İlişkili Etmenlerin İncelenmesi. *Journal of Education & Research in Nursing/Hemşirelikte Eğitim ve Araştırma Dergisi*, *12*(2).

11. KARAPINAR, D., & ZORLUTUNA, Ş. (2022). İşsizlik süresine etki eden faktörlerin yaşam analizi yöntemleri ile araştırılması. *Journal of Life Economics*, *9*(1), 1-19.

12. Köse, G., Kurutkan, M. N., & Orhan, F. (2020). Kalp yetmezliği konusunda en çok atıf alan ilk 100 makalenin bibliyometrik analizi. *Sağlık Akademisyenleri Dergisi*, *7*(2), 92-104.

13. Narin, A., İşler, Y., & Özer, M. (2014). KONJESTİF KALP YETMEZLİĞİ TEŞHİSİNDE KULLANILAN ÇAPRAZ DOĞRULAMA YÖNTEMLERİNİN SINIFLANDIRICI PERFORMANSLARININ BELİRLENMESİNE OLAN ETKİLERİNİN KARŞILAŞTIRILMASI. *Dokuz Eylül Üniversitesi Mühendislik Fakültesi Fen ve Mühendislik Dergisi*, *16*(48), 1-8.

14. Oksel, E., Akbıyık, A., & Koçak, G. (2016). Kronik kalp yetmezliği olan hastalarda öz-bakım davranışlarının incelenmesi. *İzmir Kâtip Çelebi Üniversitesi Sağlık Bilimleri Fakültesi Dergisi*, *1*(2), 1-8.

15. Yay, M., Çoker, E., & Uysal, Ö. (2007). Yaşam analizinde Cox regresyon modeli ve artıkların incelenmesi. *Cerrahpaşa Tıp Dergisi*, *38*(4), 139-145.