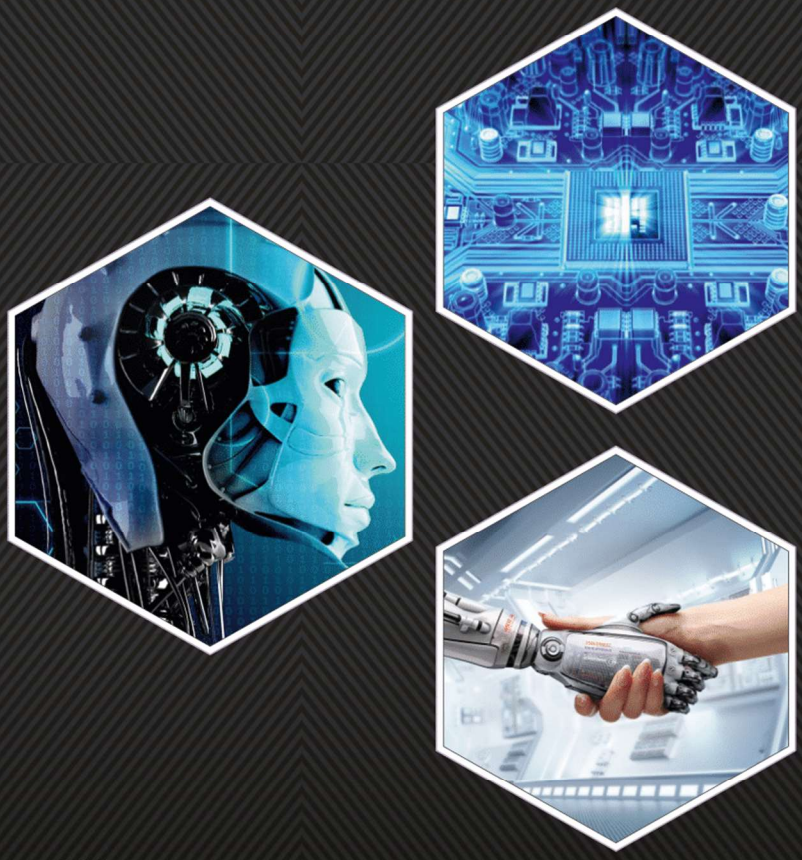
# BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ

## JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES



EDİTÖR : PROF. DR ERDİNÇ AVAROĞLU
ISSN 2717-8579


*Bilgisayar Bilimleri ve Teknolojileri Dergisi*

# BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ

**CİLT 5, SAYI 2**

**ISSN: 2717-8579**

**ARALIK 2024**


*Bilgisayar Bilimleri ve Teknolojileri Dergisi*

**Dergi Hakkında**

Bilgisayar Bilimleri ve Teknolojileri Dergisi bilim ve teknolojideki gelişmelere paralel olarak bilgisayar bilimleri ve teknolojileri alanında yeni gelişmelerle ilgili yapılan çalışmaları yayınlayan bir dergidir.

**Amaç & Kapsam**

BIBTED Dergisi,

- Bilgisayar Bilimleri ve Teknolojileri Dergisinin amacı bilgisayar alanında yapılan özgün çalışmaları yayınlamaktır. Yazım kurallarına uygun olarak hazırlanan eser, dergi editörlüğünce değerlendirme için hakemlere gönderilir. Bilgisayar Bilimleri ve Teknolojileri Dergisinde **KÖR HAKEMLİK** uygulaması mevcuttur. Yayımlanmasına, hakemlerin görüşü doğrultusunda Dergi Editör ve Yayın Kurulu karar verir. Gönderilen makaleler yayınlansın veya yayınlanmasın iade edilmez. Dergimizde yayınlanan yazıların her türlü sorumluluğu (bilimsel, mesleki, hukuki, etik vb.) yazarlara aittir. Yayınlanan yazıların telif hakkı dergiye aittir ve referans gösterilmeden aktarılamaz. Araştırmacılar arasındaki bilimsel iletişimi oluşturmak amacıyla aşağıda nitelikleri açıklanan, başka bir yerde yayımlanmamış makaleler Türkçe ve İngilizce olarak kabul edilmekte ancak Türkçe Kabul edilen makalenin özetinin ingilizce de basılması zorunluluğu vardır.

Aşağıdaki türlerdeki makaleler dergide yayına kabul edilmektedir:

- **Araştırma makalesi**: Özgün bir araştırmayı sonuçlarıyla birlikte sunan makale,
- **Derleme makale:** Bilgisayar Mühendisliği alanında belli bir konuda yeterli sayıda bilimsel makaleyi tarayıp, özetleyen, değerlendirme yapan ve bulguları yorumlayan makale,
- **Endüstriyel makale:** Bu alanda endüstride yapılan araştırma ve geliştirilen yeni ürün veya teknolojilerin açıklandığı makale,
- **Tez çalışması:** Lisansüstü düzeyde yapılan özgün bir tez çalışmasının genişletilmiş özetini içeren yazı,
- **Kitap yorumu:** Bilgisayar mühendisliği alanında yayınlanmış yeni bir kitabın tanıtılması ve değerlendirilmesi.
- Kısa Bildiri: Yapılan bir araştırmanın önemli bulgularını açıklayan yeni bir yöntem veya teknik tanımlayan yazılar.

Bütün yazıların Telif Hakkı Devri, yazarlarına bir form gönderilmek suretiyle alınır. Telif Hakkı Devir Formu göndermeyen yazarların yayınları işleme konmaz. Yayınlanmasına karar verilen yazılar üzerine yazarlarınca hiçbir eklenti yapılamaz.

Her yazı konusu ile ilgili en az iki hakeme gönderilerek şekil ve içerik bakımından incelettirilir. Dergide yayınlanabilecek nitelikteki yazılar dizgisi yapıldıktan sonra, yazarlarına gönderilerek baskı öncesi gözden istenir. Makale içinde, dergide basıldığı haliyle gözüken hataların sorumluluğu yazarlarına aittir. Hata, editörlük ofisinden kaynaklandığı takdirde düzeltme yayınlanabilir.

**Derginin Kapsamı;**

Bilgisayar Bilimleri ve Teknolojileri Dergisinin kapsamı, akıllı sistemler, algoritmalar, benzetim, bilgisayar ağları, bilgisayar grafiği, bilgisayarla görme, bilgisayar mimarisi, bilgiye erişim, bilimsel hesaplama, bilişim güvenliği, biyoenformatik, kriptografi, paralel işleme, doğal dil işleme donanım, görüntü işleme, hesaplama kuramı, işaret işleme, işletim sistemleri, makine öğrenmesi, mobil sistemler, modelleme, tıbbi bilişim, veri madenciliği, veri tabanı sistemleri, yazılım mühendisliği, siber güvenlik, yapay zeka dahil olmak üzere bilgisayar bilimleri ve teknolojilerin tüm alanları içerir.

**Yayınlanma Sıklığı** Yılda 2 sayı

**ISSN** 2717-8579

**WEB** https://dergipark.org.tr/tr/pub/bibted

**İletişim** eavaroglu@mersin.edu.tr / ttuncer@firat.edu.tr / kemaladem@gmail.com


*Bilgisayar Bilimleri ve Teknolojileri Dergisi*

*Bilgisayar Bilimleri ve Teknolojileri Dergisi*

# İçindekiler
## *Contents*

**ARAŞTIRMA MAKALELERİ;**
*RESEARCH ARTICLES;*

**DERLEME MAKALELERİ;**
*REVIEW ARTICLES;*

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Derleme Makalesi

# ISO/IEC 27001:2022 Bilgi Güvenliği Yönetim Sistemi Kapsamında Bilgi Güvenliği Risk Yönetimi ve Risk Analizi

**Melis BÖKE YAZICIOĞLU***

İskenderun Teknik Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği, Hatay, Türkiye

**Anahtar Kelimeler:**
Bilgi Güvenliği
Bilgi Güvenliği Yönetim Sistemi
Bilgi Güvenliği Risk Yönetimi
ISO/IEC 27001
ISO/IEC 27005

**ÖZ**

Siber saldırılardaki artış ile bilgi güvenliği ve Bilgi Güvenliği Yönetim Sistemi (BGYS) büyük önem kazanmıştır. BGYS'yi kurmak, kurumların bilgi varlıklarını belirleyerek, önemine göre risklerini tanımlayıp yönetmesine ve iş sürekliliğini sağlamasına destek olmaktadır. BGYS'nin oluşturulması, uygulanması ve yönetilmesi sürecinde Risk Yönetim sürecini de kapsayan birçok husus bulunmaktadır. Kurumlar, farkındalık eksikliği, maliyet ve süreç yönteminin zor olması sebepleri ile bu süreci devreye almaktan çekinmektedir. Bu makale, ISO/IEC 27001:2022 standardı çerçevesinde bilgi güvenliği risk yönetimi ve analizi süreçlerini ele almakta ve kurumların bu süreçleri etkin bir şekilde nasıl uygulayabilecekleri konusuna ışık tutmaktadır. Makalede, risk analizi için kullanılan çeşitli teknikler ve metodolojiler ele alınmıştır. İnceleme sonucunda, risk yönetim metodolojilerinin değinmiş olduğu temel noktaların ve temelde işletilmesi beklenen süreçlerin aynı ya da benzer olduğu tespit edilmiştir. Risk yönetim süreci, risklerin belirlenmesi, analiz edilmesi, değerlendirilmesi, yönetilmesi ve izlenmesi adımlarından oluşur. Her bir adımın önemi ve nasıl uygulanabileceği detaylı olarak incelenmiş olup uygulama aşamasında karşılaşılabilecek zorluklar ve çözüm önerileri analiz edilerek detaylandırılmıştır. Sürecin daha net ele alınabilmesi amacıyla edinilmiş deneyimler ve literatürde yapılan araştırmalar sentezlenerek örnek senaryolara yer verilmiştir.

# Information Security Risk Management and Risk Analysis within the Scope of ISO/IEC 27001:2022 Information Security Management System

**Keywords:**
Information Security
Information Security Management System
Information Security Risk Management
ISO/IEC 27001
ISO/IEC 27005

**ABSTRACT**

With the rise in cyber attacks, information security and the Information Security Management System (ISMS) have become crucial. Establishing an ISMS helps organizations identify their information assets, manage risks based on their significance, and ensure business continuity. The process involves various aspects, including Risk Management. Organizations often hesitate to start due to lack of awareness, cost concerns, and perceived complexity. This article explores information security risk management and analysis within the ISO/IEC 27001:2022 standard and provides guidance on effective implementation. It discusses techniques and methodologies for risk analysis. Fundamental points of risk management methodologies and expected processes are found to be similar or identical. The risk management process includes identifying, analyzing, evaluating, managing, and monitoring risks. Each step's importance and implementation are thoroughly examined, including challenges during implementation and proposed solutions. To clarify the process, example scenarios are provided based on research and practical experiences. This approach helps organizations understand and navigate the complexities of establishing and maintaining an effective ISMS.

* **Sorumlu Yazar**

*(bokemelis@gmail.com) ORCID ID 0009-0007-9055-1576

**e-ISSN: 2717-8579**

**Geliş Tarihi: 05/03/2024; Kabul Tarihi: 18/07/2024**

## 1. GİRİŞ

Bilgi, kâğıt veya başka ortamlar üzerine kaydedilmiş, anlaşılabilen ve iletilebilen veriler topluluğudur veya zihinde herhangi bir biçimde resmi veya gayri resmi olarak iletilen, kaydedilen, yayınlanan fikirlerin gerçek ve hayali ürünleridir (Sağsan, 2010).

Bilgi güvenliği, bilginin bir varlık olarak tehdit veya tehlikelerden korunması için doğru teknolojinin, doğru amaçla ve doğru şekilde kullanılarak, bilginin varlığının her türlü ortam üzerinde istenmeyen kişiler tarafınca elde edilmesini önleme girişimi olarak tanımlanmaktadır "(URL-7)".Bilgi güvenliği, en değerli kaynaklarından biri olan veri ve bilginin gizliliği, mahremiyeti, bütünlüğü ve kullanılabilirliği ile ilgilendiğinden, kuruluşların yönetiminde kilit bir rol oynar (Antunes vd., 2021). Bilgi güvenliğinde bu kavramlardan en az birinin tehlikede olma ihtimali ve gerçekleşmesi durumunda oluşturabileceği etki riski oluşturmaktadır. Gelişen teknoloji beraberinde birçok fayda getirse de bilinen ya da bilinmeyen pek çok tehdidi de beraberinde getirmektedir. Bu tehditlerin yönetilebilmesi ve azaltılabilmesi Risk Yönetim sürecini oluşturmaktadır.

Dünyaca kabul gören ve 2022 yılında güncellenmiş olan ISO/IEC 27001:2022 Bilgi Güvenliği Yönetim Sistemi Standardı kapsamında da Risk Değerlendirme sürecine önem verilmektedir. Her bir kurumun bu süreci işletmesinin bilginin güvenliğinin sağlanmasında kritik bir nokta olduğu görülmektedir. Risk Yönetim sürecini oluşturmak isteyen kurumların, temelde bunu nasıl uygulamaları ve yapmaları gerektiğini ve bu süreç içerisinde ne gibi metodolojilerin yer aldığını anlaması gerekir. Bu noktada kurumların veya kişilerin toplanmış, incelenmiş bir bilgiye ve benzer çalışmalara ihtiyaçları ortaya çıkmaktadır. Kurumlarda BGYS özelinde ve teknik kapsamda risklerin nasıl yönetilebileceği ile ilgili literatürdeki kaynaklar araştırılıp incelendiğinde, kapsamlı bir çalışma olmadığı, sunulan çalışmaların yeterli olmadığı ve kısa, özet bilgilere yer verildiği tespit edilmiştir.

Bu çalışmada, risk metodolojileri ve standartları araştırılarak, kurumların ya da kişilerin ihtiyaç duyabileceği risk metodolojilerinin, güncellenmiş olan IEC/ISO 27005:2022 Standardının gereklilikleri ile ortak paydada buluşan noktaları ve bunların risk analizi üzerindeki etkileri detaylı olarak incelenecektir. İnceleme ile kurumda risk yönetim sürecinin nasıl planlanabileceğine, yürütülebileceğine ilişkin önemli noktalar, karşılaşılabilecek zorluklar, çözüm önerileri, ISO 27001:2022'de Risk Yöntemindeki ana değişiklikler ve örnek risk değerlendirme tablosu paylaşılacaktır.

## 2. YÖNTEM

Araştırmada betimleme yöntemi kullanılmıştır. Betimleme Yöntemi, olayların, grupların, kurumların vb. çeşitli alanların ne olduğu ve bu sırada gerçekleşen eylemleri daha iyi anlayabilme, aktarabilme adına aralarındaki ilişkinin açıklandığı bir unsurdur (Kaptan, 1995).

Araştırma soruları aşağıdaki şekilde belirlenmiştir:

- BGYS ile Risk Yönetimi arasındaki ilişki nedir?
- Bilgi Güvenliği Risk Yönetimi nedir?
- Kurumlar için Bilgi Güvenliği Risk Yönetiminin önemi nedir?
- Kurumlar Risk Yönetimi yapmak için hangi metodolojileri kullanabilir?
- Metodolojilere göre kurumda Risk Yönetim süreci nasıl uygulanabilir?
- Risk Yönetim sürecinde hangi zorluklar ile karşılaşılabilir ve bu zorluklar nasıl çözümlenebilir?
- Araştırma kapsamında elde edilen bilgiler doğrultusunda hangi sonuç ve öneriler paylaşılabilir?

## 3. BİLGİ GÜVENLİĞİ YÖNETİM SİSTEMİNE GENEL BAKIŞ

BGYS, kurumun hassas bilgilerini yönetebilmek amacıyla benimsenen sistematik bir yaklaşımdır. BGYS'nin temel amacı hassas bilginin korunmasıdır (Marttin ve Pehlivan, 2010). BGYS, kurumun süreçlerini, çalışanları gibi tüm bilgi sistemlerini kapsamakta olup tüm üst yönetim tarafından desteklenmelidir.

Bir bilgi:

- Dijital sistemler üzerinde
- Bir dokümanda
- Depolama cihazlarında
- Bilgisayarlarda veya telefonlarda
- E-postalarda olmak üzere hemen hemen her yerde bulunabilmektedir.

Bilgi güvenliği, kurumlardaki çalışanlar ile hareket edilerek, iş süreçlerinin desteklenmesinde, işin sürekliliğinin güvenli bir şekilde sağlanmasında, kurumun dış veya iç tehditlerden korunmasında önemli bir role sahip olmakla birlikte bunları gerçekleştirirken 3 ana temel kavramı ele almaktadır.

**Gizlilik**, bilginin yetkisiz kişilerin erişmesini veya eline geçmesini engellemeyi amaçlamaktadır.

**Bütünlük**, bir bilginin yetkisiz kişiler tarafından değiştirilmesinin, silinmesinin veya yok edilmesinin engellenmesini ve bilginin bütünlüğünün korunması amaçlamaktadır.

**Erişilebilirlik**, bilginin veya bilgi bulanan bir sistemin yetkili kişiler tarafından ihtiyaç duyulması durumunda her zaman ulaşılabilir durumda olmasını amaçlamaktadır.

Bilgi varlıklarının korunması, taraflarda güven oluşturulabilmesi ve güvenlik kontrollerinin yeterli ve etkili seviyede uygulanmasını sağlamak için tasarlanmıştır. ISO/IEC 27001:2022 Bilgi Güvenliği Yönetim Sistemi, kurumsal yapıyı, politikaları,

planlama faaliyetlerini, sorumlulukları, uygulamaları, prosedürleri, prosesleri ve kaynakları içerir "(URL-3)".

ISO/IEC 27001:2022 Bilgi Güvenliği Yönetim Sistemi'nin kurum için birçok faydası ve avantajı bulunmaktadır. Bu faydalar "(URL-6)":

- Doğru, güvenilir ve geçerli bilgiler sağlamaktadır.
- Riskin minimize edilmesini sağlamaktadır.
- İş sürekliliğini veya kuruluşun faaliyet sürekliliğini sağlamaktadır.
- Bilginin gizliliğinin, bütünlüğünün ve erişilebilirliğinin korunmasını sağlamaktadır.
- Yasal zorunlulukların zorunlu kıldığı bazı gerekliliklerin sağlanmasına olanak tanımaktadır.
- Kurumsal saygınlık korunmasını sağlamaktadır.
- Bilgiye erişimin korunmasını sağlamaktadır.

ISO/IEC 27001:2022 kapsamında bir BGYS kurulurken Şekil 1'de yer alan ve "PUKÖ Döngüsü" olarak adlandırılan bir döngü kullanılmaktadır.


**Şekil 1.** Pukö Döngüsü

**Planla:** BGYS politikasının, amaçların, hedeflerin, süreçlerin ve prosedürlerin geliştirilmesidir (Marttin ve Pehlivan, 2010).

**Uygula:** BGYS politikasının, amaçların, hedeflerin, süreçlerin ve prosedürlerin uygulanmasıdır (Marttin ve Pehlivan, 2010).

**Kontrol Et:** BGYS politikasının, amaçların, hedeflerin, süreçlerin ve prosedürlerin performansının uygulanmasının değerlendirilmesi, ölçülmesi ve yazılı bir şekilde raporlanmasıdır (Marttin ve Pehlivan, 2010).

**Önlem Al:** Bir önceki adımda yer alan yönetimin gözden geçirme sonuçlarına bağlı olarak, düzeltici ve önleyici faaliyetlerin gerçekleştirilmesidir (Marttin ve Pehlivan, 2010).

Döngüde yer alan her bir adım tüm süreçler için uygulanarak BGYS kurulabilmektedir.

## 4. BİLGİ GÜVENLİĞİ YÖNETİM SİSTEMİNDE RİSK YÖNETİMİ

Risk yönetimi, potansiyel faydalara göre risk analizini, alternatiflerin değerlendirilmesini ve son olarak yönetimin en iyi eylem planı olarak belirlediği yöntemin uygulanmasını gerektirir (Marianne ve Barbara, 1996).

ISO 27001 Standardı ile Risk Yönetiminin yakın bir ilişkisi bulunmaktadır. ISO 27001 Risk Yönetim sürecini BGYS'nin merkezine yerleştirir ve organizasyonların bilgi güvenliği risklerini etkin bir şekilde yönetmelerini sağlar. Ayrıca, organizasyonların BGYS'yi diğer yönetim sistemleriyle bütünleştirmelerini sağlar. Bu, organizasyonların tüm risk yönetimi süreçlerini birleştirerek daha kapsamlı bir yönetim yaklaşımı benimsemelerine yardımcı olur.

Risk Yönetim süreci genel olarak Şekil 2.'de yer alan riskin tanımlanması, riskin analizi, azaltılması, riskin takibi ve riskin gözden düzenli aralıklarla yeniden gözden geçirilmesi gibi farklı süreçlerden oluşur. Risk yönetim süreçleri metodolojilere göre değişkenlik gösterse de uygulanacak adımlar birbirleriyle benzerlik göstermektedir.


**Şekil 2.** Risk Yönetim Süreci Yaşam Döngüsü

Seçilen metodoloji doğrultusunda her bir süreç uygulanarak risk yönetimi gerçekleştirilmekte ve alınacak aksiyona karar verilmektedir.

### 4.1. RİSK DEĞERLENDİRME SÜRECİNE GENEL BAKIŞ VE RİSK DEĞERLENDİRME METODOLOJİLERİ

Risk Değerlendirme, temelde gerçekleşmesi muhtemel risklerin tanımlanmasına ve bu risklerin analizine dayanmaktadır. Kurumlar, belirli bir varlığa yönelik hangi tehditlerin bulunduğunu ve bu tehditlerin risk düzeyini belirleyebilmek amacıyla risk değerlendirme sürecini kullanırlar (Peltier, 2005). Risk düzeyini sıfıra indirgemek olumsuz bir etki oluşturabileceğinden, kurumlar kendi risklerini ve uygun risk seviyelerini belirlemeli ve bunlara uygun aksiyon planı oluşturmalıdır. Risklerin değerlendirilmesinde ve analizi sürecinde Risk Yönetim sürecini kapsayan 3 ana unsur bulunmaktadır (Marianne ve Barbara, 1996). Bu unsurlar:

- Risk Değerlendirme kapsamının ve metodolojisinin belirlenmesi
- Verilerin toplanması ve analizi

- Risk Değerlendirme sonuçlarının yorumlanması

### 4.1.1. Risk Değerlendirme Sürecinde Kapsamın ve Metodolojinin Belirlenmesi

Riski değerlendirmenin ilk adımı, söz konusu olan BGYS sistemlerinin tanımlanması ve bu sistemler için gerçekleştirilecek olan risk değerlendirme metodolojisinin belirlenmesi veya oluşturulmasıdır. Kullanılabilecek birçok risk metodolojisi bulunmaktadır.

### 4.1.2. Risk Metodolojileri

**EBIOS Risk Metodolojisi:** EBIOS metodolojisi, Fransız Ulusal Güvenlik Ajansı SGDN (Secrétariat Générale de la Défense Nationale) bünyesinde Fransa Başbakanına bağlı bir hükümet kuruluşu olan DCSSI (Direction Centrale de la Sécurité des Systèmes d'information) tarafından 1995 yılında oluşturulmuştur. Yöntem, 5 adımı içerir (Mohamed vd., 2014).

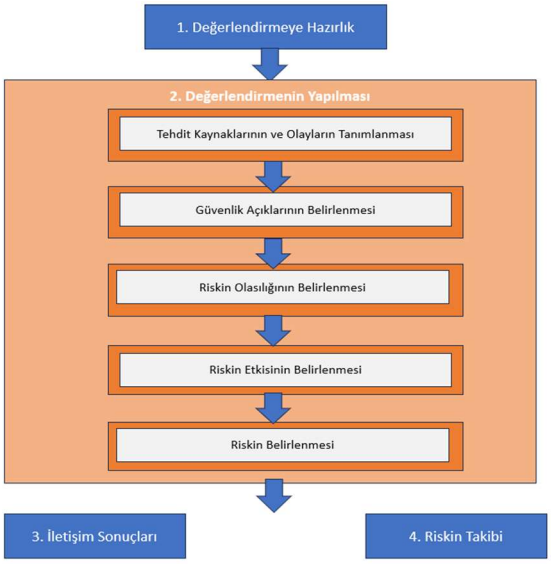
- Bağlam
- Güvenlik ihtiyaçları
- Tehdit analizi
- Güvenlik hedeflerinin belirlenmesi
- Güvenlik gereksinimlerinin belirlenmesi

EBIOS Risk Yönetim Metodolojisinde, risklerin değerlendirilmesi ve analizinde nitel bir yaklaşım benimsenmektedir.

**MEHARI Risk Metodolojisi:** MEHARI, CLUSIF tarafından geliştirilen ve Risicare (http://www.risicare.fr) şirketi tarafından yönetilen bir yazılım tarafından desteklenen risk analizi ve yöntemidir. İlk olarak 1996 yılında geliştirilen MEHARI, yöneticilere (operasyon yöneticileri, CISO, CIO, risk yöneticisi, denetçi) Bilgi ve BT kaynaklarının güvenliğini yönetme ve ilgili riskleri azaltma çabalarında yardımcı olmayı amaçlamaktadır. MEHARI, ISO/IEC 27001:2022 tarafından tanımlanan BGYS sürecine uygundur. Sürekli iyileştirme döngüsü elde etmek için güvenlik açığı kontrol noktaları listesine ve doğru bir izleme sürecine dayalı güvenlik planları geliştirmesine olanak tanır (Mohamed vd., 2014).

- Risk durumlarına ilişkin bilgileri belirleyin
- Optimum eylem planlarının oluşturulmasıyla sonuçlanan risk analizinin konsolidasyonuna ilişkin kurallar belirleyin
- Önemli riskleri analiz edin
- Güvenlik açıklarını analiz edin
- Riskleri azaltın ve yönetin
- Bilginin güvenliğini izleyin

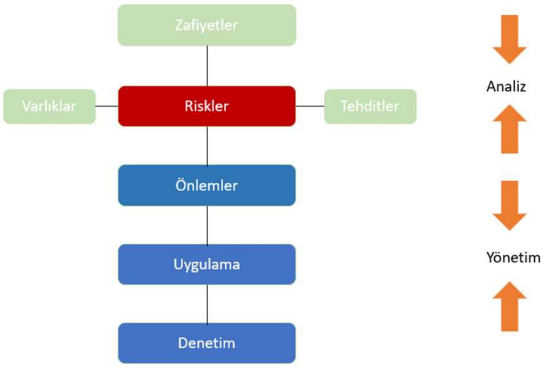
**NIST SP 800-30 Risk Metodolojisi:** NIST SP 800-30, Ulusal Standartlar ve Teknoloji Enstitüsü tarafından geliştirilen bir standarttır. Bilgi güvenliği risk değerlendirmesi için formüle edilmiş özel bir belge olarak yayınlanmakta ve özellikle BT sistemlerine yöneliktir. NIST SP 800-30'un kurumda nasıl uygulanabileceğini gösteren bir döngü (Şekil 3.) bulunmaktadır.




**Şekil 3.** NIST SP 800-30

**CRAMM Risk Metodolojisi:** CRAMM, yazılım tabanlı (Windows tabanlı) bir güvenlik riski değerlendirmesi ve risk yönetimi metodolojisidir. CRAMM, niceliksel bir metodolojiden ziyade niteliksel bir metodolojidir. CRAMM Şekil 4.'te yer alan üç temel aşamaya dayanmaktadır:

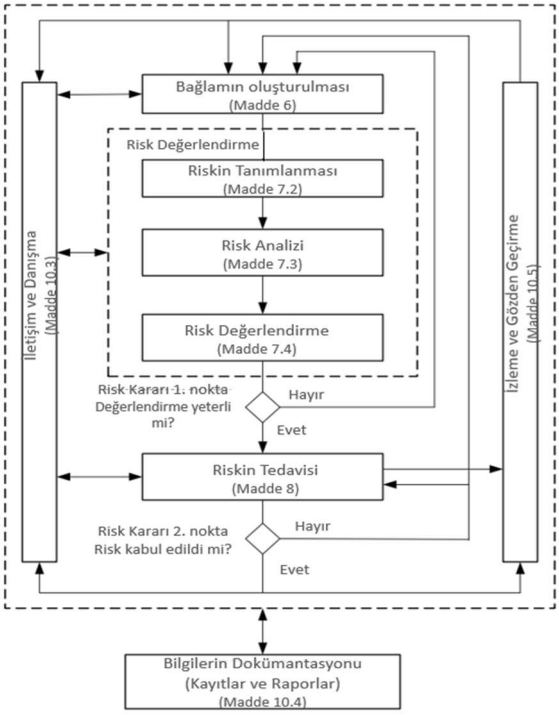
- Bilginin değerinin değerlendirilmesi ve iş sürecini destekleyen varlıkların belirlenmesi
- Hangi tehditlerin sistemi etkileyebileceğinin ve sistemin bu tehditlere karşı ne kadar savunmasız olduğunun belirlenmesi; riskler hakkında bir sonuca varmak
- Bir sonraki adım risk ölçümlerinin türetilmesidir ve bunlar tehdit, güvenlik açığı ve varlık değerinin birleşiminden elde edilir. Risk ölçümleri, oluşturulacak güvenlik gereksinimlerinin risk derecesine uygun olmasını sağlayacak şekilde ölçeklendirilir
- Mevcut kontrol önlemlerinde hangi iyileştirmelerin gerekli olduğu da dahil olmak üzere, risklerle nasıl mücadele edilebileceği belirlenir.




**Şekil 4.** CRAMM Risk Metodolojisi

**ISO/IEC 27005:2022 Bilgi Güvenliği Risk Yönetim Standardı:** ISO 27005'in amacı bilgi güvenliği risk yönetimi için yönergeler sağlamaktır. ISO/IEC 27001:2022'de belirtilen genel kavramları desteklemek ve risk yönetimi yaklaşımına dayalı olarak bilgi güvenliğinin etkin bir şekilde uygulanmasına yardımcı olmak için tasarlanmıştır. Risk analizinden risk tedavi planının oluşturulmasına kadar yapılandırılmış, sistematik ve titiz bir süreci belirtmesine rağmen, herhangi bir spesifik risk analizi yöntemini belirtmez. Birçok bilgi güvenliği kontrol hedefini ve genel kabul görmüş güvenlik kontrollerini açıklayan, bilgi güvenliği yönetimine yönelik uygulama kurallarını kapsayan bir standarttır (Mohamed vd., 2014).

ISO/IEC 27005:2022 Standardı kapsamında uygulanan ve Şekil 5'te belirtilen 8 adımdan oluşan bir risk yönetim süreci bulunmaktadır.




**Şekil 5.** ISO 27005 Bilgi Güvenliği Risk Yönetim Süreci (ISO/IEC 27005:2022 International Standard, 2022)

Süreç adımları:

- Bağlamın oluşturulması
- Riskin tanımlanması
- Riskin analizi
- Risk değerlendirilme
- Riskin tedavisi
- Bilgilerin dokümantasyonu
- İzleme ve gözden geçirme
- İletişim ve danışma

Risk Yönetim sürecinde risklerin değerlendirilmesine ve tedavi sürecindeki değişikliklere bağlı olarak Risk Yönetim süreci "Stratejik" ve "Operasyonel" olarak iki sürece ayrılabilir.

Stratejik Döngü; iş varlıkları, tehditler, hedefler, risk kaynakları, kuruluşun bağlamında yer alan değişiklikler olarak tanımlanmaktadır. Bunlar risk tedavi planının oluşturulmasında veya riskin değerlendirilmesi için girdi niteliğindedir.

Operasyonel Döngü ise, stratejik döngüde yer alanları kapsayan ve yapılan risk değerlendirmenin veya tedavinin gözden geçirilmesi gereken durumları kapsamaktadır.

Stratejik döngü daha uzun zaman bazında veya büyük değişiklikler meydana geldiğinde ve kuruluşun hedeflerine ulaşmaya çalıştığı ortam için uygulanırken, operasyonel döngü ise belirlenen ve değerlendirilen ayrıntılı risklere ve ilgili risk tedavisine bağlı olmanın yanı sıra risk yönetim süreci bağlamı dikkate alınarak yapılan tüm risk değerlendirmelerini içermekte ve daha kısa süreyi kapsayacak şekilde işletilmektedir.

## 6. RİSK DEĞERLENDİRME VE RİSK ANALİZİ

Risk değerlendirme sürecine başlayacak olan kurumlar kullanacağı risk metodolojisini belirlemiş olmalıdır. Risk metodolojisini belirleyen kurumların, sonraki süreçleri planlaması ve uygulaması gerekmektedir. Bu süreçlerin uygulanması sırasında zorluklar ile karşı karşıya kalınabilmektedir.

### 6.1. Risk Değerlendirme ve Analiz Sürecinde Karşılaşılabilecek Zorluklar ve Öneriler

Risk değerlendirme analiz sürecinde karşılaşılan zorluklar, genellikle bilgi yetersizliği, kapsam belirsizliği, tahmin hataları, önyargı ve ön kabuller, karmaşıklık, değerlendirme yöntemleri ve araçlarının karmaşıklığı, süreç yönetimi zorlukları gibi faktörlerden kaynaklanır.

- **Bilgi yetersizliği,** doğru risk analizi yapmak için gerekli olan bilgilere eksik erişim sağlanmasına veya belirli bilgilerin eksik olmasına neden olabilir.
- **Kapsam belirsizliği,** analizin doğru bir şekilde yönetilmesini zorlaştırabilir ve hangi varlıkların veya süreçlerin risk analizine dahil edileceği konusunda belirsizlik yaratabilir.

- **Tahmin hataları,** risklerin olasılığının veya etkisinin yanlış tahmin edilmesine yol açabilir ve yanlış kararlar alınmasına neden olabilir.
- **Önyargı ve ön kabuller,** analizin nesnel olmasını engelleyebilir ve doğru sonuçlara ulaşmayı zorlaştırabilir.
- **Karmaşıklık,** büyük ve karmaşık organizasyonlarda risk analizi sürecini yönetmeyi zorlaştırabilir ve analizin doğruluğunu etkileyebilir.
- **Değerlendirme yöntemleri ve araçlarının karmaşıklığı,** sürecin yürütülmesini zorlaştırabilir ve gereksiz karmaşıklığa yol açabilir.
- **Süreç yönetimi zorlukları**, paydaşlar arasında iş birliğini sağlamak ve iletişimi sürdürmek için ek bir engel oluşturabilir.

Bu zorlukların üstesinden gelmek için:

- İyi bir hazırlık ve planlama yapılmalı
- Kapsam net bir şekilde belirlenmeli
- Risk analizi için ihtiyaç duyulan tüm bilgilere erişim sağlanmalı
- Farklı görüş ve deneyimlere önem verilmeli
- Analiz yapılırken objektif olunmalı
- Basit ve uygulanabilir yöntemler tercih edilmeli
- İletişim ve iş birliği sağlanmalı
- Sürekli iyileştirme benimsenmeli

Zorluklar ile karşılaşılmaması adına sürecin başlangıcında yukarıda yer alan önerilerin uygulanması sonraki süreçlerin daha yönetilebilir olmasını sağlayacaktır. Risk metodolojisi ve kapsam net bir şekilde belirlendikten sonra bir sonraki süreçler uygulanmalıdır.

### 6.1.1. Verilerin Toplanması ve Analizi

Ebios veya ISO 27005 Risk Metodolojileri tercih edilecekse verilerin toplanması ve analiz aşamasından önce kurumların bağlamını oluşturması beklenmektedir. Bağlam, faaliyetleri etkileyen ya da faaliyetler sonucunda etkilenen yönetim sistemini iç ve dış hususlar olmak üzere belirlenmesini sağlayan bir kavramdır. ISO 27005 Standardına göre kuruluşun bağlamı tanımlanırken ya da belirlenirken, kurum bağlamını aşağıda yer alan gereksinimler kapsamında oluşturmalıdır.

- Organizasyonel hususlar
- İlgili tarafların temel gereksinimlerinin belirlenmesi
- Risk değerlendirmenin uygulanması
- Bilgi güvenliği risk kriterlerinin oluşturulması ve sürdürülmesi
- Uygun bir yöntemin seçilmesi

Diğer metodolojilerden biri ile devam edilmesi durumunda ise, verilerin toplanabilmesi için bakılabilecek çok sayıda alan bulunmaktadır. Riskin birçok bileşeni bulunmakta ve bir bilgi birçok alanda bulunabilmektedir. Verileri toplamak ve bunların analizini yapmak riskin tespit edilmesini ve tanımlanmasına destek olmaktadır.

### 6.1.2. Varlık Envanteri ve Varlıkların Değerlendirilmesi

Varlıklar; bilgi, yazılım, donanım, süreç, insan gibi içerisinde bilginin bulunduğu alanlardır. BGYS kapsamında bir kurum varlıklarını belirlemiş ise bu varlıklar üzerinden risklerin belirlenmesi için veriler toplayabilmekte ve analiz ederek riskini belirleyebilmektedir. Eğer kurum varlıklarını henüz belirlememiş ve bir varlık envanterine sahip değilse, kurumda yer alan departmanlar ile görüşmeler yapılarak varlık envanterleri çıkarılabilir. Varlık envanteri oluşturulurken aşağıda yer alan noktalar belirlenmelidir:

- Varlıkların gizlilik, bütünlük ve erişilebilirlik dereceleri
- Varlığın sahibi
- Varlıkların türü

### 6.1.3. Tehditlerin Tanımlanması

Tehdit, sisteme zarar verme potansiyeli olan varlıklar veya olaylardır. Tehditler, 3 kategoride değerlendirilmektedir.

**Doğal Tehditler:** Sel, deprem, kasırga, hortum gibi doğal afetler gibi olaylardır.

**İnsan Kaynaklı Tehditler:** İnsan kaynaklı tehditler kasıtlı ve kasıtsız olarak değerlendirilmektedir. Kasıtsız eylemler, hatalar veya ihmaller sebebiyle, kasıtlı eylemler ise insanlar tarafından bilinçli olarak yapılan dolandırma, kötü amaçlı yazılım gibi büyük kayba neden olabilecek tehditlerdir.

**Çevresel Tehditler:** Uzun süreli elektrik kesintileri, kirlilik, kimyasal atıklar gibi tehditlerdir.

Tehditler, ortaya çıkma olasılıklarını ve varlıklara zarar verme potansiyellerini belirlemek için tanımlanmalı ve analiz edilmelidir (Marianne ve Barbara, 1996).

ISO 27005:2022 kapsamında riskler tanımlanırken Olaya Dayalı Yaklaşım ve Varlığa Dayalı Yaklaşım olmak üzere 2 farklı yaklaşım türü uygulanmaktadır.

**Olaya Dayalı Yaklaşım:** Risk kaynaklarını, riskleri ve bu risklerin hedefe ulaşmayı ne ölçüde etkilediğine bağlı olarak stratejik senaryolar belirlenmektedir. Bu yaklaşımda olaylar ve sonuçlar genellikle üst yönetimin endişelerinin, risk sahiplerinin ve kuruluşun bağlamını belirlerken ortaya çıkan gereksinimler ile belirlenebilmektedir.

**Varlığa Dayalı Yaklaşım:** Varlıklar, tehditler ve güvenlik açıklarına bağlı olarak ayrıntılı bir operasyonel senaryolar belirlenmektedir. Bu yaklaşım, varlığa özgü tehditleri ve güvenlik açıklarını belirleyebilir ve kurumun bazı riskleri ayrıntılı şekilde incelemesine imkân

tanıyabilmektedir (ISO/IEC 27005:2022 International Standard, 2022).

Bu tehditlere ek olarak, kurumlar sektörü tanımalı ve sektörel tehditlerini ve var ise özel tehditlerini ve zafiyetlerini de belirlemelidir.

### 6.1.4. Güvenlik Açığı Analizi

Güvenlik açığı; güvenlik prosedürleri, teknik kontroller, fiziksel kontroller veya bir tehdit tarafından istismar edilebilecek diğer kontrollerdeki (veya bunların bulunmaması) bir durum veya zayıflıktır (Marianne ve Barbara, 1996).
Bu zayıflıklar sonucunda varlıkların ve bilgi güvenliğinin zarar görme ihtimali bulunmaktadır. Bu potansiyel açıklar belirlenmeli ve analiz edilmelidir. Bu noktada sızma testleri, tarama araçları, iç ve dış denetimlerde tespit edilen bulgular güvenlik açığı hususunda hem bir referans hem de risk için girdi oluşturmaktadır. Güvenlik açığı olarak tespit edilen zafiyet ya da bulgular risk olarak tanımlanabilmektedir.

### 6.1.5. Olasılıkların Belirlenmesi ve Değerlendirilmesi

Olasılık, bir tehdidin gerçekleşme sıklığının ya da ihtimalinin tahminidir. Kurumların kullandığı metodolojiye göre genellikle 3 ya da 5 dereceye sahiptir. 3 (Tablo 2) ve 5 (Tablo 1) dereceye ilişkin çizelgeler aşağıda detaylandırılmıştır.

**Tablo 1.** 5 Dereceli Olasılık Skalası

| Olasılık Düzeyi | Açıklama |
|---|---|
| 5- Neredeyse Kesin | Neredeyse her zaman gerçekleşebilir. Mevcut tedbirlerin yeterli olmaması nedeniyle sıklıkla tekrarlanabilir. |
| 4- Muhtemel | Sıklıkla olabilir. Mevcut tedbirlerin yeterli olmaması nedeniyle yılda birkaç kez tekrarlanabilir. |
| 3- Orta | Bazen olabilir. Mevcut tedbirlerin yeterli olmaması nedeniyle yılda bir kez tekrarlanabilir. |
| 2- Muhtemel değil | Meydana gelmesi çok mümkün olmasa bile olabilir. Mevcut kontroller ve tedbirler kısmen yeterli seviyededir. |
| 1- Nadir | Meydana gelmesi oldukça nadir. Mevcut kontroller yeterli seviyededir. Bu sayede tehdidin oluşması önlenebilmektedir. |

**Tablo 2.** 3 Dereceli Olasılık Skalası

| Düzey | Açıklama |
|---|---|
| 3- Yüksek | Sıklıkla meydana gelebilir |
| 2 -Orta | Ara sıra meydana gelmesi muhtemel |
| 1- Düşük | Çok mümkün olmasa da nadiren gerçekleşebilir. |

### 6.1.6. Tehditlere Göre Etkinin Belirlenmesi ve Değerlendirilmesi

Bir tehdidin meydana gelme olasılığı değerlendirildikten sonra, tehdidin kurum üzerinde oluşturabileceği etkisi belirlenmektedir. Etki değerinin belirlenmesinde de olasılıkta olduğu gibi genellikle 3 (Çizelge 4) ve 5 (Çizelge 3) dereceye sahip skala kullanılmaktadır. Riskin etki değeri belirlenirken kurumun genel misyonu, değer ve hedeflerinin nasıl etkileneceğine ek olarak Bilgi Güvenliğinin 3 temel prensibi olarak bilinen Gizlilik, Bütünlük ve Erişilebilirlik üzerindeki etkileri de göz önünde bulundurulmalıdır. Kurum, riskin gerçekleşmesi durumunda itibar kaybı, finansal kayıp, operasyonel süreçlerde aksama gibi zararlar görebilir.

**Tablo 3.** 5 Dereceli Etki Skalası

| Düzey | Açıklama |
|---|---|
| 5- Kritik | Çok ciddi kayıplara sebep olabilir. |
| 4- Yüksek | Ciddi kayıplara olabilir. |
| 3- Orta | Önemli kayıplara sebep olabilir. |
| 2- Düşük | Kurum için minör kayıplara sebep olabilir. |
| 1- Çok Düşük | Kurum için neredeyse önemsiz kayıplardır. |

**Tablo 4.** 3 Dereceli Etki Skalası

| Düzey | Açıklama |
|---|---|
| 3-Yüksek/ Kritik | Çok ciddi veya ciddi kayıplara sebep olabilir. |
| 2 – Orta | Önemli kayıplara sebep olabilir. |
| 1- Düşük | Kurum için minör kayıplara sebep olabilir. |

### 6.2. Olasılık ve Etki Analizine Göre Risklerin Değerlendirilmesi

Olasılık ve etki analizi kurumların riski doğru şekilde değerlendirmesini, ölçülebilmesini ve bu

değerlendirme neticesinde kontrollerin ve önlemlerin tanımlanarak etkin bir şekilde uygulanmasına destek sağlamaktadır. Riskler belirlendikten sonra ve sonuçların hem olasılık hem de etki değerleri belirlendikten sonra kurumlar risklerin kabul edilip edilmeyeceğini belirlemek için risk kabul kriterlerini uygulamalıdır (ISO/IEC 27005:2022 International Standard, 2022).
Riskin değeri,

$$Risk\ (R) = OxE \qquad (1)$$

Formülü ile hesaplanır. Hesaplamada yer alan O değeri Olasılık (Likelihood), E değeri ise Etki (Impact) olarak değerlendirilmektedir.

Tanımlanan risk için uygulanan mevcut bir kontrol var ise, kontrol metni içerisinde kontrolün ne olduğu ve nasıl çalıştığının detaylandırılması riskin yönetiminin kolaylaşmasında önemli bir rol oynamaktadır. Detaylandırma ISO 27005'te belirtilen bilgilerden yararlanılarak yapılabilir.

- Belgelenmiş bir kontrol olup olmadığı
- Kontrolün sahibinin kim veya hangi departman olduğu
- Nasıl izlendiği
- Nasıl kanıtlanabileceği
- İstisnalar

Bir risk, olasılık ve etkisine göre değerlendirilirken sanki herhangi bir kontrol uygulanmıyormuş gibi değerlendirilmelidir. Bu değerlendirme, doğal risk olarak tanımlanan riskin gerçek değerini göstermektedir.

Mevcutta uygulanan kontroller risk değerlendirme aşamasında yapılan skorlamaya dahil edilmemelidir. Bunun sebebi (ISO/IEC 27005:2022 International Standard, 2022):

- Bir veya daha fazla bilgi güvenliği riskini yönetmek için gerekli bir kontrol olmayabilir.
- BGYS tarafından yönetilmek için yeterince etkili olmayan bir kontrol olabilir.
- Halihazırda bilgi güvenliği ile ilgili olmayan başka bir konuda uygulanıyor olabilir.
- Bilgi Güvenliği için uygun olmayabilir.

Bu tespit yapıldıktan sonra mevcutta uygulanan kontroller var ise bu kontroller ile yeniden risk değerlendirme yapılır ve riskin mevcut skoru belirlenir. Uygulanması gereken ek kontrol veya önlemlerin gerekliliği tespit edilmelidir.

ISO 27005:2022 Standardı kapsamında uygulanan kontroller Önleyici, Tespit Edici ve Düzeltici olmak üzere 3 sınıfa ayrılmaktadır.

**Önleyici Kontrol,** bir veya daha fazla sonucun ortaya çıkmasına yol açabilecek bir bilgi güvenliği olayının meydana gelmesini engellemeyi amaçlamaktadır (ISO/IEC 27005:2022 International Standard, 2022).

**Tespit Edici Kontrol,** bir bilgi güvenliği olayının meydana geldiğini tespit etmeyi amaçlayan bir kontroldür (ISO/IEC 27005:2022 International Standard, 2022).

**Düzeltici Kontrol,** bir bilgi güvenliği olayının sonuçlarını sınırlamayı amaçlayan kontroldür (ISO/IEC 27005:2022 International Standard, 2022).

- Tespit edici kontroller, önleyici kontrollerin başarısız olması durumunda riski azaltmalıdır.
- Düzeltici kontroller, eğer tespit edici kontrollerin başarısız olması durumunda riski azaltmalıdır.
- Önleyici kontroller ise, düzeltici kontrollerin kullanılma olasılığını azaltmalıdır.

Riskin değerlendirilmesinde genellikle 3x3 (Şekil 6) veya 5x5'lik (Şekil 7) risk matrisleri kullanılmaktadır. Bu matrisler her ne kadar yaygın kullanılıyor olsa da kurumsal kendi risk matrislerini de belirleyebilmektedir.

| OLASILIK \ ETKİ | YÜKSEK | ORTA | DÜŞÜK |
|---|---|---|---|
| YÜKSEK | YÜKSEK | YÜKSEK | ORTA |
| ORTA | YÜKSEK | ORTA | DÜŞÜK |
| DÜŞÜK | ORTA | DÜŞÜK | DÜŞÜK |

**Şekil 6.** 3x3 Risk Değerlendirme Matrisi

| OLASILIK x ETKİ | | | Çok Yüksek | Yüksek | Orta | Düşük | Çok Düşük |
|---|---|---|---|---|---|---|---|
| | | | 5 | 4 | 3 | 2 | 1 |
| OLASILIK | Çok Yüksek | 5 | 25 | 20 | 15 | 10 | 5 |
| | Yüksek | 4 | 20 | 16 | 12 | 8 | 4 |
| | Orta | 3 | 15 | 12 | 9 | 6 | 3 |
| | Düşük | 2 | 10 | 8 | 6 | 4 | 2 |
| | Çok Düşük | 1 | 5 | 4 | 3 | 2 | 1 |

**Şekil 7.** 5x5 Risk Değerlendirme Matris ("URL-4")

Peltier 'in Bilgi Güvenliği Risk Analiz kitabında yer alan "Aksiyon Gereklilik Matrisi" risklerin olasılık ve etkisine göre aksiyon gerekliliği olup olmadığının tespitini desteklemek amacıyla kullanılmaktadır.

| OLASILIK \ ETKİ | YÜKSEK | ORTA | DÜŞÜK |
|---|---|---|---|
| YÜKSEK | A | B | C |
| ORTA | B | B | C |
| DÜŞÜK | C | C | D |

**Şekil 8.** Aksiyon Gereklilik Matrisi(Peltier, 2005)

Şekil 8'de yer alan Aksiyon Gereklilik Matrisi 'ne göre A, B, C ve D değerlerinin açılımları:

A – Mutlaka düzeltici faaliyet uygulanmalıdır.
B – Düzeltici faaliyet uygulanmalıdır.
C – Takip edilmeli ve izlenmelidir.
D – Şu anda herhangi bir aksiyon alınmasına gerek bulunmamaktadır.

### 6.3. Risk Değerlendirme Sonuçlarının Yorumlanması

Risk değerlendirme, kuruluş için gerçekten neyin önemli olduğunu yansıtan anlamlı bir çıktı üretmelidir. Risk değerlendirme birbiriyle ilişkili iki işlev olan riskin kabulü ve uygun maliyetli kontrollerin seçimini desteklemek için kullanılır (Marianne ve Barbara, 1996). Riskin değerlendirme sonuçları, risk skoruna (R) bağlı olarak mevcut kontroller, aksiyon planları, mevcut kontrollerin efektiflik derecesine göre yorumlanmalı ve riskin kabul edilebilir seviyede olması sağlanmalıdır. Kabul edilebilir seviye, kurum tarafından belirlenmiş olan ve kurumun kabul edebileceği seviyeye indirilmiş risk olarak tanımlanmaktadır. ISO 27001:2022 Standardını uygulayan her kurum ya da kuruluşta, üst yönetim tarafından karar verilmiş bir kabul edilebilir risk seviyesi bulunmaktadır (Durankaya vd., 2018). Kabul edilebilir seviye belirlenirken kurum tarafından belirlenmiş olan risk iştahı göz önünde bulundurulmalıdır.

Risk iştahı, üst yönetim tarafından belirlenen ve kurumun kabul edebileceği ya da tahammül edebileceği en yüksek risk seviyesidir. Aynı zamanda bu düzeyin üzerinde kalan risklerin onaylanamayacağını ve bu konuda önlem almanın gerekli olduğunu göstermektedir ("URL-2)".

Risk değerlendirme adımı tamamlandıktan sonra riskin ne yapılacağına karar verilmesi beklenmektedir. Bu karar sonucunda aşağıda yer alan 4 yöntemden biri uygulanmaktadır.

- Risk Kabulü
- Riskin Azaltılması
- Riskten Kaçınma
- Riskin Transferi

#### 6.3.1. Riskin Kabulü

Riskin kabulü, kurum tarafından güvenlik riskinin varlığının kabul edilmesini fakat ilgili risk için belirli gerekçeler sebebi ile herhangi bir iyileştirici aksiyon alınmayacağını ifade etmektedir. Risk iştahının üzerinde kalan fakat önlem ya da aksiyon alınması mümkün olmayan durumlarda risk kabulü yapmak uygulanabilecek yöntemlerden biridir. Bir kurumun risk kabul kriterleri, risk yöntem sürecinde genel bir yaklaşım olarak tanımlanır ve bilgi güvenliği politikası içerisinde yer alır. Sadece risk kabulü yapmak ve herhangi bir iyileştirici aksiyon almamak bazı durumlarda etkisiz olabilmekte ve potansiyel sonuçlar doğurabilmektedir.

Bir kuruluşun riskleri etkin bir şekilde azaltmadan kabul etmesi, zaman içinde kalan risklerin yüksek oranda birikmesine neden olabilir ve bu da güvenlik ihlalleri ve veri ihlalleri olasılığını artırabilir ("URL-1").

Risk kabul kararı, risklerin kabul edilebilir olduğu durumlarda ya da azaltma maliyetinin riskin potansiyel etkisinden daha ağır bastığı durumlarda verilebilir.

#### 6.3.2. Riskten Kaçınma

Riski tespit edilen bir varlığın kullanımından vazgeçmek riskten kaçınma olarak tanımlanmaktadır. Riskten kaçınma, güvenlik risklerini azaltmak için cazip bir yöntem gibi görünse de bazı dezavantajları bulunmaktadır.

- Daha güvenli bir teknolojiye geçiş gibi çeşitli ve daha yüksek maliyet gerektiren alternatifler doğurabilir.
- Güvenlik ihtiyaçlarını tam olarak karşılayabilecek güvenilir bir tedarikçi bulma konusunda zorluklarla karşılaşabilir. Bu durum, riskten kaçınma stratejilerinin uygulanmasında gecikmelere veya aksaklıklara yol açarak kurumu potansiyel tehditlere karşı savunmasız bırakabilir ("URL-1").
- Kurumun operasyonları üzerinde, işlerin aksamasına neden olabileceğinden olumsuz potansiyel etkileri olabilir.

Riskten kaçınmak her zaman uygulanabilir ve sürdürülebilir olamayabileceğinden uzun vadeli bir çözüm olarak kullanılması önerilmemektedir. Teknolojinin gelişmesiyle birlikte çok sayıda yeni güvenlik açıkları ve tehditleri de ortaya çıkmaktadır. Bu nedenle sadece riskten kaçınma stratejisini kullanmak uzun vadede olası tehditlere karşı kurum için tam bir koruma sağlama noktasında yetersiz kalabilmektedir ("URL-1").

#### 6.3.3. Riskin Transferi

Riskin transferi, potansiyel risklerin sorumluluğunun ya da yükünün üçüncü bir tarafa aktarılmasını içeren bir yöntemdir. Riskin transferi, hizmet sağlayıcı veya harici satıcılar ile yapılan sözleşmeler yolu ile gerçekleştirilebilir. Kurumlar riski transfer ederek, olası bir güvenlik ihlalinden doğacak olan maddi veya repütasyonel kayıpları indirgeyebilmekte ve aynı zamanda bu sayede hizmet satın alarak maliyeti de azaltabilmektedir.
Riskin transferinin getirmiş olduğu bazı dezavantajlar bulunmaktadır. Bu dezavantajlardan bir tanesi tüm risklerin aktarılamaması veya eksik aktarılmasıdır. Ayrıca, maliyet ve performans etkileri de dahil olmak üzere güvenlik risklerinin sonuçlarının aktarılması her zaman mümkün veya etkili olmayabilir ("URL-1").

**Tablo 5.** Risk Değerlendirme Rapor Örneği

| **Risk ID** | **Tespit Yılı** | **Riskin Tanımı** | **Riskin Sahibi** | **Riskin Tipi** | **Riskin Gizliliğe Etkisi** | **Riskin Bütünlüğe Etkisi** | **Riskin Erişilebilirliğe Etkisi** | **Riskin Gerçekleşme Olasılığı** | **Riskin Etkisi** | **Doğal Risk Seviyesi** | **Risk Değerlendirme Sonucu** | **Mevcut Aksiyon(lar)** | **Aksiyon Efektiflik Derecesi** | **Mevcut Kontrol Sonrası Riskin Değeri** | **Planlanan Aksiyon(lar)** | **Aksiyon Sonrası Riskin Gerçekleşme Olasılığı** | **Etkisi** | **Riskin Yeni Değeri** | **Riskin Durumu** |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 001 | 2024 | Riskin tanımı | Riskin Sahibi | İtibari (Repütasyonel), Finansal, Bilgi Kaybı Riski | Düşük, Orta Yüksek | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Riskin Azaltılması, Kabulü vb. | | Efektif, Az efektif, Efektif değil | Düşük, Orta, Yüksek | | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Düşük, Orta, Yüksek | Açık, Kapalı |

### 6.4. Riskin İyileştirilmesi ve İyileştirme Planı

#### 6.4.1. Riskin Azaltılması

Risk azaltma, riski yönetim tarafından kabul edilebilir bir düzeye indirmek için güvenlik kontrollerinin seçilmesini ve uygulanmasını içerir. Risk azaltma sürecinde aşağıda yer alan kontroller ve aksiyonlar uygulanabilmektedir.

**Koruma sağlayacak kontrollerin belirlenmesi:** Uygun kontrolleri seçerken aşağıdakiler dikkate alınmalıdır (Marianne ve Barbara, 1996).

- Kurumsal politika, mevzuat veya regülasyonlar
- Güvenlik, güvenilirlik ve kalite gereksinimleri
- Sistem performans gereksinimleri
- Güncellik, doğruluk ve bütünlük gereklilikleri
- Güvenlik önlemlerinin yaşam döngüsü maliyetleri
- Teknik gereksinimler
- Kültürel kısıtlamalar

Risk tedavi planı oluşturulurken dikkat edilmesi gereken bazı hususlar bulunmaktadır. Bu hususlar (ISO/IEC 27005:2022 International Standard, 2022):

- Risk düzeyi ve iyileştirmenin aciliyeti ile ilgili öncelikler
- Farklı kontrol türlerinin ve bu kontrollerin birleşiminin uygun olup olmadığını,
- Kontrolün uygulamaya konduğu an ile tamamen etkili ve çalışır duruma geldiği an arasında bir gecikme olup olmadığı.

**Artık riskin kabul edilmesi:** Riskin azaltılmasına yönelik gerekli aksiyonlar alındıktan sonra geriye kalan risk "Artık Risk" olarak adlandırılmaktadır. Artık riskler, kalan olasılık ve etkilerine göre yeniden değerlendirilmelidir. Yönetimin veya ekiplerin, kalan risklerin türünü ve ciddiyetini göz önünde bulundurarak, BT sisteminin işleyişinin kabul edilebilir olup olmadığına karar vermesi gerekir (Marianne ve Barbara, 1996). Karar doğrultusunda gerekiyorsa yeni bir aksiyon planı oluşturulmalıdır.

### 6.5. Riskin Dokümante Edilmesi

Tespit edilen ve değerlendirilen tüm risklerin dokümante edilmesi risk yönetim sürecinde kritik bir öneme sahiptir. Tespit edilen risklerin sayısı artmaya başladıkça, risklerin takibi ve hatırlanması güçleşmeye başlamaktadır. Bu nedenle risklerin dokümantasyonu bu noktada ve riskin detaylarını inceleme ya da yeniden hatırlama konusunda kaynak sağlayacaktır. Bir riski dokümante ederken aşağıda yer alan başlıklar veya Tablo 5'te yer alan örnek rapor taslağı kullanılabilir.

- Riskin tanımı/detayı
- Riskin sahibi
- Riskin kritiklik derecesi
- Alınmış veya alınması gereken aksiyon detayları

### 6.6. Riskin Takibi ve Gözden Geçirilmesi

Risk takibi ve gözden geçirilmesi Risk Yönetim Süreç Döngüsünün son halkasıdır. Bu bölümde risk değerlendirme süreci her ne kadar sona ermiş gibi görünse de aslında hala yaşayan bir süreç bulunmaktadır. İlgili risk için gerekli kontroller sağlanmış ve etkin önemler alınmış olsa bile risk

unsuru tamamen ortadan kalkmayabilir. Bazı riskler gerekli önlemler etkin bir şekilde alındıktan sonra kapanabilirken bazı riskler her zaman açık kalabilmektedir. Açık kalacak olan bu riskler kabul edilebilir seviyede ve her zaman gerçekleşme ihtimali devam eden risklerdir. Bu risklere:

- Deprem, yangın gibi doğal afet riskleri
- İnsan tarafından oluşabilecek riskler örnek verilebilir.

## 7. RİSK DEĞERLENDİRME VE ANALİZİ UYGULAMA ÖRNEKLERİ

Bir yazılım şirketinin risklerinin değerlendirme aşamalarının aşağıdaki gibi olduğu bir senaryo incelenmektedir.

**Hazırlık ve Planlama Aşaması:**

Bilgi güvenliği risklerini belirlemek amacıyla, uzmanlık alanları ve deneyimleri dikkate alınarak, bilgi güvenliği uzmanları, sistem yöneticileri ve ilgili departman temsilcilerinden oluşan bir risk yönetim ekibi kuruldu. Ekip, proje planını hazırlayarak sürecin adımlarını belirleyecektir. Proje planı, risk analiz sürecinin başlangıcından sonuna kadar olan zaman çizelgesini ve ekip üyelerinin sorumluluklarını belirtmektedir. Zaman çizelgesi içerisinde her adım için belirlenen süreler ve süre sonunda gerçekleştirilmesi gereken faaliyetler belirlenmelidir.

**Risklerin Tanımlanma Aşaması:**

Şirketin varlık envanteri oluşturuldu ve kritik bilgi varlıkları belirlendi. Varlıklar yazılım kodları, müşteri veri tabanları, proje belgeleri gibi kritik bilgi varlıklarını içermektedir. Bu varlıkların yanı sıra, şirketin sahip olduğu diğer önemli bilgi kaynakları (patent başvuruları vb.) da belirlenerek envantere eklenmiştir.

Potansiyel tehditler göz önüne alınarak, risk kaynakları belirlendi ve riskler dokümante edildi. Risklerin kataloglanması sürecinde, her bir tehdidin potansiyel etkisi ve olasılığı detaylı bir şekilde değerlendirildi. Örneğin, yazılım kodlarının yetkisiz erişime açık olması durumunda şirketin itibarı ve müşteri güveni ciddi şekilde zarar görebilirken, müşteri veri tabanının sızdırılması durumunda ise veri ihlali ve yasal yaptırımlar söz konusu olabilir. Bu adımla birlikte riskler kapsamlı şekilde değerlendirilmiştir.

**Risklerin Analiz Edilme Aşaması:**

Bu aşamada belirlenen risklerin olasılığı ve etkisi dikkate alınarak detaylı bir değerlendirme yapılmıştır. Her riskin potansiyel etkisi ve gerçekleşme olasılığı titizlikle incelenerek belirlenmiştir. Önceliklendirme yapılırken, risklerin kritiklik düzeyi ve etkileri göz önünde bulunduruldu. Özellikle, şirketin faaliyetleri üzerindeki potansiyel etkileri ve olası zararları değerlendirilerek risklerin önem sırası belirlendi. Örneğin, yazılım kodlarının sızdırılması gibi yüksek etkili ve yüksek olasılıklı riskler öncelikli olarak ele alındı. Bu tür bir olayın gerçekleşmesi durumunda, şirketin itibarı ciddi şekilde zarar görebilir, müşteri güveni sarsılabilir ve yasal yaptırımlarla karşılaşabilir. Bu nedenle, bu tür risklerin etkilerini azaltmak için öncelikli olarak çözüm stratejileri belirlendi ve uygulanması için gerekli kaynaklar tahsis edildi.

Diğer yandan, düşük etkili veya düşük olasılıklı riskler daha düşük öncelikle ele alındı. Bu tür riskler, şirketin operasyonları üzerindeki etkileri daha sınırlı olduğu için, daha az acil bir şekilde çözülmesi gereken riskler olarak değerlendirildi. Ancak, bu risklerin de göz ardı edilmemesi ve uygun önlemler alınarak yönetilmesi önemlidir,

**Risk Değerlendirme Aşaması:**

Riskleri değerlendirme aşamasında, belirlenen risklerin kabul edilebilirlik düzeyi belirlendi. Bu, risklerin şirketin tolerans seviyesine uygun olup olmadığının değerlendirilmesini içeriyordu. Kabul edilebilirlik düzeyi aşılan riskler için risk azaltma stratejileri geliştirildi ve uygulanması için gerekli önlemler alındı.

Örneğin, yazılım kod güvenliğinin artırılması, veri tabanı erişim kontrollerinin sıkılaştırılması gibi çeşitli önlemler bu stratejilere örnek olarak belirlendi. Yazılım kod güvenliğinin artırılması için, kod inceleme süreçleri güçlendirildi, güvenlik açıklarını tespit etmek için otomatik test araçları kullanıldı ve geliştiricilere güvenli kodlama eğitimleri verildi. Ayrıca, veri tabanı erişim kontrolleri sıkılaştırılarak, yetkisiz erişim girişimlerinin önlenmesi amaçlandı ve bu yönde teknik ve politika bazlı önlemler alındı.

Risklerin yönetilmesi için gereken kaynaklar ve bütçe belirlenirken, risk azaltma stratejilerinin uygulanması için gereken kaynakların ve maliyetlerin hesaplanması yapıldı. Bu, güvenlik yazılımlarının satın alınması, güvenlik eğitimlerinin düzenlenmesi gibi kaynak ve bütçe gereksinimlerini içeriyordu. Belirlenen kaynak ve bütçe gereksinimleri, risk azaltma stratejilerinin etkin bir şekilde uygulanmasını ve risklerin yönetilmesini sağlamak amacıyla kullanıldı. Bu sayede, şirketin bilgi güvenliğini artırmak için gerekli kaynakların ve bütçenin sağlanması ve yönetilmesi sağlandı.

**Risklerin Yönetilme Aşaması:**

Kabul edilebilirlik düzeyini aşan risklerin kontrol ve koruma önlemleri titizlikle uygulanarak, risklerin azaltılması ve etkilerinin minimize edilmesi sağlandı. Bu süreç, şirketin bilgi varlıklarını korumak ve olası güvenlik açıklarını kapatmak için kritik öneme sahipti.

Örneğin, yazılım kod güvenliğinin artırılması için otomatik test araçlarının kullanılması gibi önlemler alındı. Bu sayede, yazılım geliştirme sürecinde olası güvenlik açıkları daha erken tespit edilebilir hale geldi ve hızla çözüme kavuşturulabildi. Ayrıca, veri tabanı erişim yetkilerinin revize edilmesi gibi kontroller de

uygulandı. Bu sayede, yetkisiz erişim girişimlerini önlemek ve veri güvenliğini sağlamak amaçlandı.

Her bir kontrol ve koruma önlemi, riskin özelliğine ve potansiyel etkisine uygun olarak belirlendi ve sistematik bir şekilde hayata geçirildi. Bu önlemler sayesinde, şirketin bilgi güvenliği riskleri etkin bir şekilde yönetildi ve olası güvenlik tehditlerine karşı daha güçlü bir savunma mekanizması oluşturuldu.

**Risklerin İzlenmesi ve Gözden Geçirilme Aşaması:**

Uygulanan önlemlerin etkinliği düzenli aralıklarla izlendi ve değerlendirildi. Bu süreçte, değişen tehditler ve organizasyonel değişiklikler göz önünde bulunduruldu ve risk analizi sürekli olarak gözden geçirildi. Bu sayede, güvenlik önlemlerinin güncel kalması ve şirketin bilgi güvenliğinin sürekli olarak sağlanması amaçlandı.

Bu izleme ve değerlendirme süreci, aylık olarak düzenlenen toplantılar ve güvenlik raporlarının incelenmesi yoluyla gerçekleştirildi. Bu toplantılarda, risklerin güncel durumu değerlendirilerek, yeni tehditler ve güvenlik zafiyetleri hakkında bilgi paylaşımı yapıldı. Ayrıca, mevcut güvenlik önlemlerinin etkinliği tartışıldı ve alınması gereken ek önlemler belirlendi.

Bu süreç, şirketin bilgi güvenliğini sürekli olarak güçlendirmek ve olası risklere karşı hazırlıklı olmak için kritik bir rol oynadı. Değerlendirme sonuçlarına dayanarak, mevcut önlemlerin etkinliği artırıldı ve yeni güvenlik stratejileri belirlendi. Bu sayede, şirketin bilgi varlıklarını koruma kapasitesi sürekli olarak iyileştirildi ve güncel tehditlere karşı daha etkili bir şekilde savunma sağlandı.

Sonuç olarak, yazılım şirketi bilgi güvenliği risk yönetimi sürecinde başarılı bir ilerleme kaydederek kapsamlı bir varlık ve risk tanımlaması yapmış ve önemli tehditleri belirlemiştir. Kabul edilebilirlik düzeyini aşan riskler için etkili kontrol ve koruma önlemleri uygulanmıştır. Süreç sürekli izlenmiş ve değerlendirilmiş, böylece güvenlik önlemleri sürekli olarak güncel tutulmuştur. Bu sürecin şirketin bilgi güvenliğini güçlendirdiği ve sürdürülebilir bir güvenlik çerçevesi oluşturduğu belirlenmiştir.

## 8. SONUÇ ve ÖNERİLER

ISO 27001 Standardı çerçevesinde bilgi güvenliği risk yönetimi ve analizi süreçleri detaylı olarak incelenmiş ve bu süreçler kapsamında en sık kullanılan risk metodolojileri ele alınarak kurumların bu süreçleri etkin bir şekilde uygulayabilmesi noktasına ışık tutulmaktadır. Ekiplerin risk yönetimi konusunda sade ve yalın bir bilgiye ulaşabilmesi sağlanmakla birlikte, risk yönetimi konusunda ihtiyaç duyabileceği birçok konu ve soru detaylı olarak incelenmiştir. Bu inceleme akademik bir makale olmasının yanı sıra edinilmiş deneyimlerin ve uygulamaların da süzgeçten geçirilerek paylaşılmasını sağlamaktadır.

Teknolojinin son derece hızlı bir şekilde gelişmesi ile ortaya çıkabilecek pek çok bilgi güvenliği riski bulunmaktadır. Bilgi güvenliği risk yönetimi süreçlerinin önemi ve etkin uygulanmasının kurumların bilgi varlıklarını korumak ve güvenliklerini sağlamak açısından kritik olduğu ve bunun yanı sıra yasal uyumun sağlanmasında da gereklilik haline geldiği sonucu tespit edilmektedir. Bu gerekliliklerin sağlanması hususunda, kurumların hangi metodolojiyi seçmesi gerektiği ve nasıl bir süreç işletmesi gerektiği bilgisi, genellikle kurumlar için gerek sürecin tasarlanması gerek ise işletilmesi ve yönetilmesi konusunda zor olmaktadır. Bazı kurumlar düşük maliyet ve yönetim kolaylığı nedeniyle hizmet olarak satın almayı tercih etmektedir. Bu konuda kurumların çekimser davranmasında çeşitli sebepler bulunmakta, bu sebeplerden biri de sürecin tam anlaşılamaması ve anlaşılabilmesi için yeterli kaynağın bulunmamasıdır. Literatürde yapılan araştırmalar sonucunda da bilgi güvenliği risklerinin yönetimi konusunda yeterli kaynağın olmaması veya var olan kaynaklardaki bilgilerin yetersiz kalması bunu destekler niteliktedir.

Kurumsal bilgi güvenliği risklerinin tespit edilmesi, analizi ve doğru aksiyon planlanlarının belirlenmesi kurumlarda oluşabilecek olası bir tehditin ortaya çıkmasını azaltmaya ya da engellemeye olanak tanıması noktasında büyük önem oluşturmaktadır. Bu sayede kurumlar risklerini azaltarak kabul edilebilir seviyelere indirgeyebilmektedir.

Makalede incelenen risk yönetimi metodolojileri, temelde benzer adımları içermekte ve kurumların bu süreçleri etkin bir şekilde uygulayabilmelerine olanak tanımaktadır.

Hedef ve ihtiyaçlar belirlendikten sonra kurumlar kendi yöntemlerini de geliştirip uygulayabilirler ya da mevcut yöntemleri kullanmayı tercih edebilirler. Bu tamamen kurumun beklentilerine ve ihtiyaçlarına bağlı şekilde yapılanmaktadır.

Sonuçlar kapsamında öneriler ve gelecek çalışmalara yön verebilecek husular:

- Risk Yönetim Sürecinin uygulanabilmesi için BGYS kurma zorunluluğu bulunmasa da kurumların BGYS'yi kurmaları risk yönetimini kolaylaştırmaktadır.
- Risk Yönetim Sürecinin öneminin anlaşılması amacıyla hem üst yönetim hem de kurum çalışanları için farkındalık eğitimi verilebilir.
- 
- Risk analizi adımlarının ayrıntılı bir şekilde incelenmesi ve örnek senaryolar üzerinde değerlendirilmesi, kurumların riskleri daha iyi anlamalarına ve uygun önlemleri alabilmelerine yardımcı olmaktadır.

- Risk yönetim sürecinde karşılaşılabilecek zorlukların farkında olunması ve uygun çözümlerin geliştirilmesi önemlidir.

Son olarak, BGYS'de sürekli izleme ve değerlendirme sürecinin önemi vurgulanmaktadır. Değişen tehditler ve organizasyonel değişiklikler göz önünde bulundurularak, risk analizi sürekli olarak gözden geçirilmeli ve güncellenmelidir. Bu şekilde, kurumlar bilgi güvenliğini sürekli olarak güçlendirebilir ve olası risklere karşı hazırlıklı olabilirler.

**KAYNAKÇA**


Antunes, M., Maximiano, M., Gomes, R., ve Pinto, D. (2021). Information Security and Cybersecurity Management: A Case Study with SMEs in Portugal. Journal of Cybersecurity and Privacy, 1(2), 219-238.

Durankaya, İ., Gökşen, Y., Eminağaoğlu, M. (2018, Ekim) ISO/IEC 27001 ISO27001 Bilgi Güvenliği Yönetim Sisteminde Risk Analizi. IMISC 2018 Conference Proceedings, 29-33.

ISO/IEC 27005:2022 Information Security Risk Management International Standard. (2022).

ISO/IEC 27001:2022 Information Security Management System International Standard. (2022).

Kaptan, S., (1995), Bilimsel Araştırma ve İstatistik Teknikleri, Tekışık Web Ofset Yayınları. Ankara

Marianne S. ve Barbara G. (1996). NIST- Generally Accepted Principles and Practices, Special Publication (NIST SP)- 800-14.

Mohamed G., Sophia F., Hicham M., Adil S. (2014). Information Security Risk Assessment — A Practical Approach with a Mathematical Formulation of Risk, International Journal of Computer Applications (0975 – 8887).

Peltier, T. R. (2005). Information Security Risk Analysis (15-42).

Sağsan, M. (2010). Gelişmişliğin Vazgeçilmez Unsuru: Ulusal Bilgi Politikası.

Marttin, V., Pehlivan, İ. (2010). ISO 27001:2005 Bilgi Güvenliği Yönetimi Standardı ve Türkiye'deki Bazı Kamu Kuruluşu Uygulamaları Üzerine Bir İnceleme (49-56).

URL-1: https://www.6clicks.com/resources/answers/what-are-the-four-4-cybersecurity-risk-treatment-mitigation-methods
[Erişim Tarihi: 30.11.2023]

URL-2: https://finans.mynet.com/haber/detay/r/risk-istahi-nedir-risk-istahi-nasil-belirlenir/453728/
[Erişim Tarihi: 26.02.2024]

URL-3:https://it.bilgi.edu.tr/tr/guvenlik/iso-27001/
[Erişim Tarihi: 18.11.2023]

URL-4: https://ishayativedenetim.com/2021/03/26/risk-matrisi-nedir/
[Erişim Tarihi: 10.11.2023]

URL-5: https://www.beyaz.net/tr/guvenlik/makaleler/bilgi_guvenligi.html
[Erişim Tarihi: 17.05.2024]

URL-6: https://belgelendirme.ctr.com.tr/iso-27001.html
[Erişim Tarihi: 24.12.2023]

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

# Yüz Tanımada Derin Öğrenme Mimarilerinin ve Yüz Bulma Yöntemlerinin Karşılaştırılması

**Ayşe Merve Büyükbaş*[1], Ali Öztürk[2,3]**

*[1]KTO Karatay Üniversitesi, Ticaret ve Sanayi Meslek Yüksekokulu, Bilgisayar Teknolojileri Bölümü, Bilişim Güvenliği Teknolojisi, Konya, Türkiye*
*[2]KTO Karatay Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Bilgisayar Mühendisliği, Konya, Türkiye*
*[3]Havelsan A.Ş., Konya, Türkiye*

**Anahtar Kelimeler:**
Derin öğrenme
Yüz bulma
Yüz tanıma
Eigenfaces
Fisherfaces

**ÖZ**

Bu çalışmada, literatürde yaygın olarak kullanılan görüntü işleme tabanlı yöntemler ve AlexNet, ResNet-18, GoogleNet ve SqueezeNet gibi mimariler kullanılarak performans karşılaştırılması yapılmıştır. Yüzün resim üzerinde belirlenebilmesi için Viola-Jones algoritması kullanılmıştır. Bu algoritmada kaskad obje dedektörü, yüzü algılayıp kare içine alır. Viola-Jones algoritmasının doğruluk oranı %85,71 olarak bulunmuştur. FEI yüz veri tabanındaki sağ, sol ve orta pozlarla veri kümesi oluşturulmuştur. Eigenfaces ve Fisherfaces görüntü işleme yöntemlerinin analizi için Temel Bileşen Analizi (TBA) ve Doğrusal Ayrım Analizi (DAA) kullanılmıştır. Bu yöntemler oluşturulan veri kümesi üzerinde uygulanarak doğruluk oranları elde edilmiştir. Eigenfaces yöntemi veri kümesindeki bazı poz varyasyonları için fisherfaces yönteminden daha iyi sonuç vermiştir. Derin öğrenme metotlarından AlexNet, ResNet-18, GoogleNet ve SquuezeNet kullanılmıştır. Yüz tanıma yöntemlerinden Eigenfaces yönteminin en yüksek doğruluk oranı %76,66 ve derin öğrenmede ResNet-18'in en yüksek doğruluk oranı %100 olmuştur.

# Comparison of Deep Learning Architectures and Face Detection Methods in Face Recognition

**Keywords:**
Deep learning
Face detection
Face recognition
Eigenfaces
Fisherfaces

**ABSTRACT**

In this study, performance comparisons were made using image processing-based methods widely used in the literature and architectures such as AlexNet, ResNet-18, GoogleNet and SqueezeNet. Viola-Jones algorithm was used to determine the face on the image. In this algorithm, cascade object detector detects the face and frames it. The accuracy rate of Viola-Jones algorithm was found to be 85.71%. A dataset was created with right, left and middle poses in the FEI face database. Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) were used for the analysis of Eigenfaces and Fisherfaces image processing methods. These methods were applied on the created dataset and accuracy rates were obtained. Eigenfaces method gave better results than Fisherfaces method for some pose variations in the dataset. Deep learning methods AlexNet, ResNEt-18, GoogleNet and SqueezeNet were used. Among the face recognition methods, the highest accuracy rate of the Eigenfaces method was 76,66% and the highest accuracy rate of ResNet-18 in deep learning was 100%.

*Sorumlu Yazar

*(ayse.merve.buyukbas@karatay.edu.tr) ORCID ID 0000–0002–6534–7764
(ali.ozturk@karatay.edu.tr) ORCID ID 0000–0002–1797–2039

**e-ISSN: 2717-8579**

**Geliş Tarihi: 27/08/2024; Kabul Tarihi: 04/12/2024**

## 1. GİRİŞ

Yüz tanıma sistemleri, dijital resimler şeklinde kaydedilen insanlara ait yüz resimleri ile yapılan eğitimden sonra, yeni bir yüz resmi verildiğinde, kişinin kimliğinin belirlenmesi için geliştirilen sistemlerdir (Kekül vd., 2018).

Veri toplama, ön işleme, özellik çıkarma, model eğitimi, değerlendirme ve test etme bir makine öğrenimi yaklaşımı kullanılarak geliştirilebilir. Yüz resimlerinden oluşan veri kümesi ile, TBA (Temel Bileşen Analizi), Yerel İkili Örüntü (Local Binary Pattern: LBP), Evrişimsel Sinir Ağları (Convolution Neural Networks: CNN), Destek Vektör Makineleri (Support Vector Machine: SVM), Rastgele Orman (Random Forest) veya Yapay Sinir Ağları (Neural Networks) gibi algoritmalar eğitilebilir. Günümüzde yüz tanıma için kullanılan popüler yöntemler; Eigenface, Fisherface ve Yerel İkili Örüntü Histogramı (Local Binary Pattern Histogram: LBPH) algoritmalarıdır (Balanageshwara vd., 2023).

Yapay görme, makine görme, hesaplamalı görme veya görüntü analizi olarak da adlandırılan bilgisayarlı görme, bir bilgisayarı araç olarak kullanarak görüntülerden bilgi çıkarma işlemidir. Yüz tespiti, görüntüdeki bir veya daha fazla kişinin yüzünün, görüntü içindeki arka planı veya içinde bulunan diğer nesneleri göz ardı ederek bulunmasını sağlayan tekniktir. Başlangıçta bir sınıflandırıcıyı eğitmek için çok sayıda görüntüye ihtiyaç vardır. Örneğin, bir yüz dedektörü geliştirmek için yüz resmi içeren ve içermeyen görüntüler gereklidir (Cadena vd., 2023).

Gerçek zamanlı yüz tanıma, uygulamalarda oldukça popülerdir. Arya ve Tiwari, Eigenface, Fisherface ve LBPH algoritmalarını kullanan gerçek zamanlı otomatik bir yüz tanıma ve tespit sistemi yapmışlardır. Haar Cascade ile Eigenface, Fisherface ve LBPH algortiması kullandıklarında kişiyi $-30^0$ ila $+30^0$, $-60^0$ ila $+60^0$ ve $-60^0$ ila $+75^0$ açılardan başarıyla tespit edip tanıyabildiği sonucuna varmışlardır. Ayrıca kişinin ön yüzünü yukarıdan aşağıya veya tersi yönde döndürdüğü zaman kişiyi tespit edip, ön yüzü eğik olan kişiyi de tespit ederek tanıyabilmişlerdir. Ön yüzü eğik olan kişiyi tespit etmek ve tanımak için LBPH ve Fisherface kullanan sistem ±%10 eğim açısıyla en iyi çalışma performansını vermiştir. Ön yüz açısı tamamen yukarı ve aşağı olduğunda, Eigenface ve LBPH algoritmasını kullanan sistem hem normal ışık koşullarında (gündüz) hem de düşük ışık koşullarında (gece) en iyi sonucu vermiştir (Arya ve Tiwari, 2020).

Viola-Jones algoritması, bir insan yüzünün özelliklerini arayan pencere ile bir görüntüyü tarar. Özellikleri bulursa ve bir yüz olarak belirli bir değere sahipse, görüntünün belirli penceresinin yüz olduğu tahmin edilir. Farklı boyutlarda yüzlerin olduğu bir durumu çözmek için, pencere her görüntü için tekrarlanan işlemle ölçeklendirilir (Rahmad vd., 2020).

Holat ve Kulaç (2014) yaptığı çalışmada, yüz tanıma sistemi için kameradan alınan anlık görüntüyü ve bilgisayarda kayıtlı olan görüntüyü kullanarak yüz tespiti yapmışlardır. Görüntüyü edinme aşaması; test görüntüsüyle ile yüzü bulma aşamasından, yüz konum bilgisi üzerinden ön işlemenin yapılması, tanımanın yapılabilmesi için özellik çıkarıcı ve sınıflama işleminden oluşmaktadır. Yüz tanımada kullanılan TBA ve DAA (Doğrusal Ayrım Analizi) ve HE (Histogram Eşitleme) yöntemleri için Medyan, Gauss ve Laplace filtreleri kullanmışlardır. Yale veri tabanı ile yapılan denemelerde en iyi sonuçlar merkezden aydınlatılmış görüntüler eklendiğinde elde edilmiştir. En iyi yöntem ise LBP+HE+Medyan yönteminde %85 başarım oranına sahiptir. ORL veri tabanı ile yapılan denemelerde de en iyi sonuçlar LBP+HE+Medyan yönteminde %90 başarım oranı olarak bulunmuştur (Holat ve Kulaç, 2014).

Yapay sinir ağları tabanlı derin öğrenme mimarilerinden biri de evrişimsel sinir ağlarıdır. Günümüzde literatürde resim üzerinde sınıflama yöntemlerinde yaygın olarak kullanılmaktadır. Transfer öğrenimi için ön eğitim, özellik çıkarıcı ve kısmi özellik çıkarıcı yöntemler kullanılmıştır (Doğan ve Türkoğlu, 2019). Yapılan çalışmada, veri setindeki sınıflandırmaların AlexNet eğitim setiyle olan benzerliği transfer öğrenimi üzerindeki başarısını arttırmıştır. En yüksek doğruluk oranı %99,02 olarak Mnist veri kümesindedir (Fırıldak ve Talu, 2019).

Derin öğrenmenin gelişmesiyle CNN tabanlı yüz tanıma teknolojisi bu alanda temel yöntem haline gelmiştir. Algoritmalar açısından CNN konvolüsyon katmanı ile diğer konvolüsyon katmanları arasında paylaşım parametreleri vardır. Eğitilecek parametre sayısının bellek gereksinimlerine bağlı olarak azalması da bir avantajdır (Wang ve Li, 2018).

CNN fikri yapay sinir ağlarından geliştirilmiştir. Biyolojik sinir ağı, doğrusal olmama, eşzamanlılık, sağlamlık ve yüksek hata toleransıyla karakterize edilir. Bu özelliklerinden dolayı yapay sinir ağları görüntü işleme gibi alanlarda oldukça yaygın kullanmaktadır. Wang ve arkadaşları, araştırmasında 8 katmanlı geleneksel kedi türlerinin tanımlanması, 4 evrişimsel katmanı, 2 havuzlama katmanı, tam bağlı katman ve çıkış katmanından oluşan CNN modeli kullanmışlardır.

Konvolüsyon katmanı birden çok katmandan oluşur. Buradaki amaç, giriş verileri üzerinde işlem yaparak ve farklı özellikler oluşturarak konvolüsyonu gerçekleştirmektir. Aynı zamanda görüntü çözünürlüğünü azaltmak ve hesaplamayı kolaylaştırmak için, konvolüsyon katmanı görüntü kolerasyon ilkesine göre bir araya getirilmiştir. Geliştirilen modelde doğruluk testi %68,85'ten %79,41'e yükselmiştir (Wang vd., 2020).

Topal ve arkadaşları (2023) ImageNet üzerinde eğitilmiş CNN görüntülerine karşı evrişimsel algoritma (EA) tabanlı düşmanca saldırı önermişlerdir. EA tabanlı saldırıları, sınıflandırılma

konusunda güven olasılığı en az %75 olarak bulmuşlardır. Önerdikleri EA tabanlı saldırının, başarı oranı ve üretilen rakip görüntülerinin görsel kalitesi açısından rakiplerine göre üstün veya eşit olduğunu ortaya koymuştur (Topal vd., 2023).

## 2. YÖNTEM

Çalışma kapsamında yürütülen deneyler işlemci olarak Intel Core i7-10750H CPU, ekran kartı özelliği nVDIA GeForce RTX2060 Intel UHD Graphics, @2.60 GHz temel frekansı ve 16 GB RAM'e sahip 64 bit Windows 10 Home SL işletim sistemine sahip dizüstü bilgisayar üzerinde yapılmıştır. Gerçekleştirilen deneylerde MATLAB R2018b yazılımı kullanılmıştır.

Viola-Jones yöntemi görüntü düzleminin dışında döndürülen ve görüntü düzlemi etrafında döndürülen görüntülerdeki yüz varyasyonlarını inceler. Viola-Jones kaskad obje dedektörü, görüntü üzerinde oluşan pencereyi kaydırarak nesneleri algılamaya yarar. (Jones ve Viola, 2003).

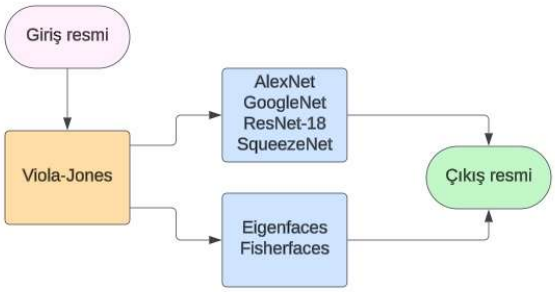
Yüz tanımanın yapılabilmesi için veri setindeki resimlerin ilk olarak Viola-Jones algoritmasından başarılı bir şekilde geçmesi gerekmektedir. Yapılan bu çalışma için toplam 686 yüz fotoğrafı bulunmaktadır. Görüntüler siyah beyaz renk uzayına indirgenmiştir. Yeni oluşturulan veri kümesindeki görüntü örneği ve boyutu Şekil 1'de gösterilmiştir.


**Şekil 1.** Viola-Jones algoritması ile tanınan yüz resmi

Giriş görüntüsündeki yüz resmini algılamak amacıyla kaskad obje dedektörü oluşturulur. Dedektör, fotoğraf üzerindeki yüzü arayarak belirlemektedir. Belirlenen yüz nesnesi pencere içine alınmaktadır.

Yüz bulma yöntemleri bilgiye dayalı yöntemler, değişmez özellikli yaklaşım yöntemleri, şablon eşleştirme yöntemleri ve görüntü işleme tabanlı yöntemler olmak üzere dört kategoride incelenmiştir. Bilgiye dayalı yöntemlerde, test görüntüsündeki insan yüzüne ait özellikleri çıkartmak ve bu özelliklere göre aday yüzleri bulmak kolaydır. Ancak arka planda oldukça başarılı sonuçlar üretmesine rağmen, farklı pozlardaki resimler için bu yöntemin uygulanması oldukça zorlayıcıdır. İnsana ait özellikleri kurallara veya kodlara dönüştürmek her zaman mümkün olmayabilir. Değişmez özellikli yaklaşım yöntemlerinde, aday resimlerdeki insan yüzlerini bulmak için renk yoğunluğu ve çeşitliliği, kenar ve desen gibi özellikler kullanılır. Ancak aydınlatma ve resimlerdeki diğer gürültüler yüzdeki özelliklerin bulunmasını zorlaştırmaktadır. Şablon eşleştirme yöntemleri, yüz resmindeki baskın türde olan özellikleri kullanarak verilen test görüntüsünün üzerindeki yüzleri bulmayı dener. Hesaplaması oldukça kolaydır, ancak yüze yakın yerlerde tarama yapılmazsa maliyetli olabilir. Görüntü işleme tabanlı yöntemlerde öncelikle giriş resmi üstünde ön işlem uygulanır. Test ve eğitim için oluşturulan resimler standart hale getirilir. Sınıflandırmanın yapılabilmesi için pozitif ve negatif algoritmalar sayesinde giriş verilerinin eğitilmesi gerekir. Görüntü işleme tabanlı yöntemler başarı oranı yüksek makine öğrenmesi algoritması kullanırlar. Hızlı ve etkin çalışmalarının yanı sıra başarısı ispatlanmış sonuçlar üretirler (Sütçüler, 2006).

Viola-Jones algoritması ile çerçeve içine alınan yüz resimleri derin öğrenme mimarileri ve yüz bulma algoritmaları olmak üzere iki farklı işlemden geçmiştir. Bu işlemler Şekil 2'de gösterildiği gibidir.




**Şekil 2.** Viola-Jones ile derin öğrenme mimarileri ve yüz tanıma algoritması

Giriş resmi, Viola-Jones algoritmasında algılandıktan sonra derin öğrenme mimarileri için AlexNet, GoogleNet, ResNet-18 ve SqueezeNet mimarileri kullanılmıştır. Yüz tanıma yöntemleri için Eigenfaces ve Fisherfaces yöntemleri kullanılmıştır.

### 2.1. Veri Kümesi

Yapılan bu çalışmada FEI yüz veri setindeki bazı imgeler kullanılmıştır (Thomaz, 2012). FEI yapay zekâ laboratuvarında, 19-40 yaş aralığındaki çalışanların ve öğrencilerin farklı açılardan çekilen pozları veri setini oluşturmuştur. Toplamda 200 farklı kişinin 14 farklı pozlarından oluşmuştur. Her görüntünün boyutu 640x480 pikseldir.

Test ve eğitim klasörlerinden oluşan bu veri kümesinde 140 yüz resmi derin öğrenme mimarileri, Eigenfaces ve Fisherfaces algoritmaları için ayrılmıştır. 546 yüz resmi derin öğrenme mimarileri için ayrılmıştır. Derin öğrenme mimarilerinin özelliklerine göre 224x224 veya 227x227 piksel boyutlarına indirgenerek yeni klasörde toplanmıştır. Toplam 686 adet görüntü yer almaktadır. Görüntü veri setlerinin %25'i test ve %75'i eğitim (train) klasörü olarak ikiye ayrılmıştır.

Veri kümesine ait görüntüler 14 farklı yüz pozlarından oluşmaktadır ve bu görüntüler Şekil 3'te gösterilmiştir.

**Şekil 3.** FEI bir kişiye ait 14 farklı poz (Thomaz, 2012)

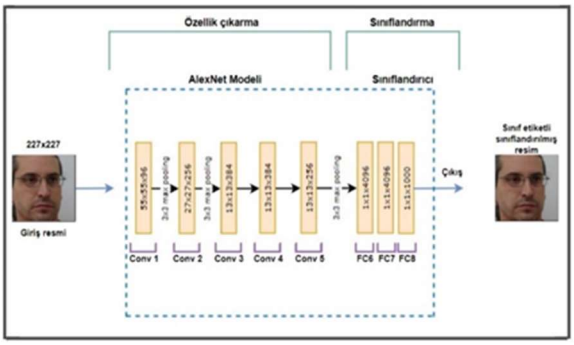
### 2.2. Derin Öğrenme Mimarileri

Derin öğrenme mimarileri konvolüsyon katmanları, havuzlama katmanları, tam bağlı ve sınıflandırma katmanından oluşmaktadır. Mimarideki ağlar, giriş verisi üzerindeki katmanlar arasında işlemlerden geçerek eğitilmektedir (Sert, 2020). Bu makalede yapılan çalışma için kullanılan mimariler AlexNet, ResNet, GoogleNet ve SqueezeNet olmak üzere dört farklı grupta incelenmiştir.

#### 2.2.1. AlexNet Mimarisi

AlexNet mimarisi ağ ağırlıkları olan sekiz katman içerir; ilk beşi evrişimli ve kalan üçü tamamen bağlı katmanlardır (Krizhevsky vd, 2017).

Giriş katmanı 227x227x3 boyutunda olup AlexNet mimarisi özellikleri Şekil 4'teki gibidir.



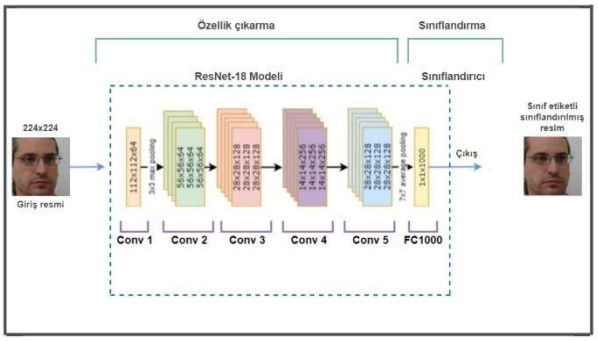

**Şekil 4**. AlexNet mimarisi özellikleri (Almabdy ve Elrefaei, 2019)

Giriş resmi ilk olarak Viola-Jones algoritmasından geçerek yüzü bulma işlemi tamamlanmıştır. Yüzü bulunan görüntüler, eğitim aşamasından geçerek örnek resimleri ve bunların sınıflandırmaları sağlanmıştır. Epoch değeri, eğitim aşaması boyunca geçecek olan her veri için ağ üzerinde gösterilmeyi sağlar. Mimarideki ağlar eğitilirken, doğruluğu arttırabilmek ve daha iyi sonuçlar elde edebilmek amacıyla epoch sayıları değiştirilerek denemeler yapılmıştır.

#### 2.2.2 ResNet Mimarisi

Derin kalıntı ağı veya ResNet mimarisi, He ve arkadaşları tarafından geliştirilmiş bir modeldir (He vd., 2016). Derin öğrenme eğitimindeki ikilemleri yenmek için oluşturulmuştur. Çünkü derin öğrenme eğitimi oldukça zaman alır ve belirli sayıda katmanla sınırlıdır. ResNet modelinin diğer mimari modellere kıyasla avantajı, mimari derinleşse bile bu modelin performansının düşmemesidir (Sarwinda vd., 2021).

ResNet18 mimarisi 72 katman ve 18 ağ derinliği, ResNet50 mimarisi 177 katman ve 50 ağ derinliği, ResNet101 mimarisi ise 347 katman ve 101 ağ derinliğine sahip önceden eğitilmiş ağ mimarileridir. Giriş görüntü katmanı 224x224x3 boyutundadır (Raghu vd., 2020). ResNet-18 mimarisinin özellikleri Şekil 5'te gösterilmiştir.



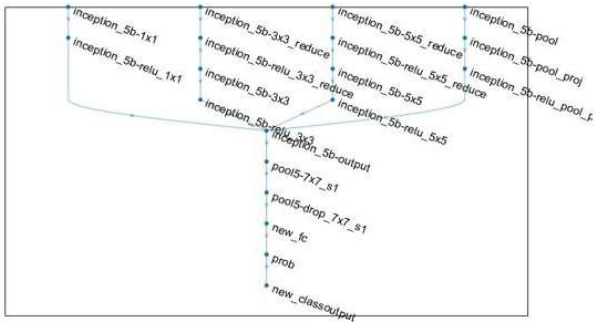

**Şekil 5.** ResNet-18 mimarisi özellikleri (Almabdy ve Elrefaei, 2019)

224x224x3 boyutundaki giriş resmi Viola-Jones algoritmasından geçtikten sonra, eğitime girmiştir. Eğitim aşaması; özellik çıkarma ve sınıflandırma adımlarından oluşmaktadır. Özellik çıkarma ve sınıflandırma adımından sonra sınıf etiketi oluşmuş resim elde edilmiştir.

#### 2.2.3 GoogleNet Mimarisi

Szegedy ve arkadaşları tarafından 2014'te ImageNet Büyük Ölçekli Görsel Tanıma Yarışması (ILSVRC) için kurulan GoogleNet adlı takım, iki farklı kategoride birinci olmuştur. İlk kategoride, ek eğitim verileriyle nesne tespitinde, algılama modelleri topluluğu için doğrulama oranı %44,5'tir. Sınıflandırma ve yerelleştirme olarak adlandırılan ikinci kategoride toplam 5 değer skorunda %6,66 hata vermiştir. Hebbian ilkesinden elde edilen çok ölçekli fikrini, evrişimli sinir ağı mimarisiyle birleştiren çalışmadır. Fikirleri birleştirmek toplam parametre sayısını azaltırken, evrişimsel katmanlardaki parametre sayısını önemli derecede attırmayı sağlamıştır (ImageNet Large Scale Visual Recognition Challenge, 2014).

Transfer öğrenimi, yeni problemin temeli olarak onu baştan eğitmek yerine önceden eğitilmiş ağı kullanmayı tercih eder (Şeker A, 2018). Derin öğrenme ile yapılan çalışmaların çoğunda başarı oranını arttırdığı gözlenmiştir. GoogleNet mimarisinde, konvolüsyon ve havuzlama katmanlarından geçen ağlar, transfer öğrenimi için eğitilmiştir. Eğitilmiş ağların son aşaması Şekil 6'da gösterildiği gibidir.




**Şekil 6.** Eğitilen ağın transfer öğrenimi

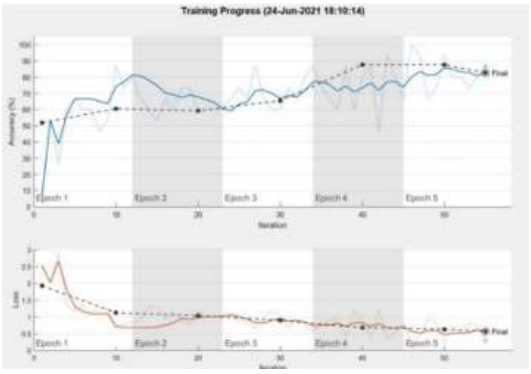
### 2.2.4 SqueezeNet Mimarisi

Iandola ve arkadaşları yapmış oldukları çalışmada, evrişimsel sinir ağlarında doğruluğu arttırmak için SqueezeNet mimarisini tasarlamışlardır. Tasarım için üç ana strateji kullanmışlardır: ilki 3x3 filtreleri 1x1 boyutlu filtrelerle değiştirmektir. Böylece dokuz kat daha az parametreye sahip filtre kullanmışlardır. İkinci olarak, filtrelere giriş kanalı sayısını 3x3 filtrelere düşürmektir. Mimari bir konvolüsyon katmanı, 8 ateşleme modülü ve son konvolüsyon katmanından oluşmaktadır (Iandola vd., 2016).

Çalışma için oluşturulan veri setindeki giriş yüz görüntüleri, 227x277x3 boyutuna çevrilerek ayrı klasöre eklenmiştir. Orijinal boyutlu görüntüler farklı klasörde tutulmuş olup, bu görüntülerin sınıflandırma örnekleri Şekil 7'de gösterilmiştir.

**Şekil 7.** SqueezeNet sınıflandırma örnekleri

Örnek sınıflandırma işleminden sonra, ağı tekrar eğiterek transfer öğrenimi gerçekleştirilmiştir. Eğitim aşamasındaki başarım oranı eğrisi Şekil 8'deki gibidir.




**Şekil 8.** ImageNet ağının eğitimi

### 2.3 Yüz Tanıma Yöntemleri

#### 2.3.1 Eigenfaces ile Yüz Tanıma

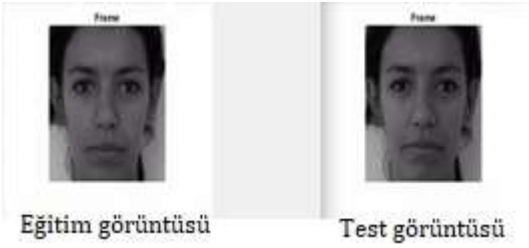
Eigenfaces ile yüz tanıma metodunda, veri setindeki tüm görüntüler siyah beyaz renk uzayına çevrilmiştir. Görüntünün piksel boyutları 180x200 olarak indirgenmiştir. Görüntü veri tabanı matrisini ortalama görüntüden farkı alınmış görüntü matrisi oluşturur. Görüntüler $N^2$xM boyutlu veri matrisi haline getirilmiştir. Eigenfaces ile elde edilebilir olması için, görüntü veri tabanı matrisiyle kovaryans matrisi işlemi gerçekleştirilir.

Eğitim setindeki 140 görüntüden 120 tanesi Viola-Jones algoritmasından başarıyla geçmiştir. Yüz tabanındaki kişilerin 14 farklı görüntüsü yerine, algoritmada algılanan 12 farklı yüz görüntüleri kullanılmıştır. Yeni oluşan eğitim setindeki bir kişinin 12 farklı görüntüsü Şekil 9'daki gibidir.

**Şekil 9.** Eğitim setindeki örnek görüntü kümesi

Kaskad obje dedektörü aracılığıyla kare içine alınan resimler kırpılarak, 431x360 örnek piksel boyutlu resim 180x200 piksel boyutuna düşürülmüştür.

10 ayrı kişiden oluşturulan 120 farklı fotoğrafın 30'u test, 90 tanesi eğitim klasörü olmak üzere iki klasöre ayrılmıştır. Eğitim ve test görüntüsünün karşılaştırması Şekil 10'da gösterilmiştir.




**Şekil 10.** Veri setindeki eğitim ve test görüntüsü

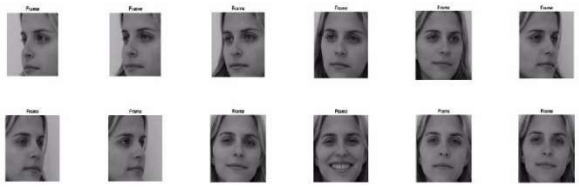
Veri setindeki iki görüntü birbiriyle karşılaştırıldığında, aynı kişiye ait olduğu görülmektedir.

#### 2.3.2 Fisherfaces ile Yüz Tanıma

Fisherfaces yaklaşımı DAA'ya dayanmaktadır. DAA sınıf içi değişkenliği en aza indirirken, sınıflar arası değişkenliği en üst düzeye çıkaran verilerin doğrusal kombinasyonlarını bulmayı amaçlayan denetimli bir boyutluluk azaltma yöntemidir (Çakmakoğlu, 2018). Giriş verilerindeki farklı sınıflar arasında en iyi ayrımı sağlayan yeni

indirgenmiş alt uzayı bulmaya çalışır. Bu temel fikir, TBA uygulamasına benzer şekilde yüz tanımaya uygulanır. Her yüz görüntüsü daha yüksek boyutlu bir uzayda nokta olarak kabul edilir. Ardından, fisherfaces adı verilen temel vektörleri elde etmek için verilere DAA uygulanır. Yüz görüntüleri daha sonra eşleştirmenin gerçekleştirildiği bu temelde yansıtılır (Aly M., 2006).

Eğitim setindeki 140 görüntünün 120 tanesi yüz bulma dedektöründen geçmiştir. Görüntüler siyah beyaz renk uzayında 180x200 piksel boyutuna indirgenerek oluşturulmuştur. Eğitim setindeki bir kişiye ait yüz görüntü kümesi örneği Şekil 11'de gösterilmiştir.



**Şekil 11.** Eğitim setindeki bir kişinin görüntüleri

Yüz görüntülerinde farklı yüz ifadeleri ve farklı açılardan çekimler olduğu görülmektedir.

## 3. BULGULAR

Bu bölümde ilk olarak Viola-Jones algılama dedektörü ile yüz görüntülerinin ne kadarının algılanıp algılanmadığı, buna bağlı olarak doğruluk oranları gösterilmiştir. Viola-Jones algoritmasından algılanarak geçen yüz resimleri yüz tanıma yöntemlerinde kullanılan veri setini oluşturmaktadır. Veri setindeki görüntüler eğitim ve test olmak üzere iki kategoriye ayrılmıştır. İkinci olarak, yüz tanımada Eigenfaces ve Fisherfaces yöntemleri kullanılmıştır. Eigenfaces yönteminde, giriş görüntü resmi için ortalaması ve matris boyutlarıyla özdeğer ve özvektörler yardımıyla ağırlık matrisi bulunmuştur. Öklid mesafesine göre karşılaştırması yapılarak görüntünün eş değeri test görüntüsünde elde edilmiştir. Fisherfaces yönteminde, giriş resmi üzerinden resmin ortalaması ve sınıf içi ortalaması aracılığıyla öz vektör matrisi bulunmuştur. Test ve eğitim görüntülerinin öklid mesafesi aracılığıyla eşleştirmeleri yapılmıştır. Çalışmanın sonraki adımlarında, derin öğrenme modellerinden evrişimsel sinir ağları olan AlexNet, ResNet-18, GoogleNet ve SqueezeNet mimarileri kullanılmıştır. Her bir mimari için sınıflandırma örnekleri ve eğitim aşaması uygulanmıştır. Yüz tanıma doğruluk oranları gösterilmiştir. Son olarak derin öğrenme modellerinde kullanılan hiper parametreler ve bu parametrelere bağlı olarak elde edilen bulgular gösterilmiştir.

Viola-Jones algoritması ile elde edilen yüz tespitindeki doğruluk oranları Tablo 1'de gösterilmiştir.

**Tablo 1.** Viola-Jones algoritmasının performansı

| | Toplam yüz resmi | Tespit edilen yüz resmi | Tespit edilemeyen yüz resmi | Doğruluk oranı (%) |
|---|---|---|---|---|
| Viola-Jones | 140 | 120 | 20 | 85,71 |
| Viola-Jones | 546 | 462 | 84 | 84,61 |

Viola-Jones algoritmasında toplam 686 görüntü içinde tespit edilen yüz resimleri 582, tespit edilemeyen yüz resimleri 104 tanedir. Veri setindeki en yüksek doğruluk oranı %85,71 olarak bulunmuştur.

### 3.1 Yüz Tanıma Yöntemleriyle İlgili Bulgular

Yüz tanıma yöntemlerinde kullanılan Eigenfaces ve Fisherfaces algoritmaları için sağ, sol ve orta poz olmak üzere üç farklı görüntü incelenmiştir. Veri setindeki görüntülerin hepsi test ve eğitim klasörü olarak ikiye ayrılmıştır. Eigenfaces algoritmasındaki doğruluk oranları Tablo 2'de gösterilmiştir.

**Tablo 2.** Eigenfaces ile yüz tanıma doğruluk oranı

| Eigenfaces | Sağ poz | Sol poz | Orta poz |
|---|---|---|---|
| Toplam yüz görüntüsü | 120 | 120 | 120 |
| Toplam test görüntüsü | 30 | 30 | 30 |
| Toplam eğitim görüntüsü | 90 | 90 | 90 |
| Doğru yüz görüntüsü | 13 | 6 | 23 |
| Yanlış yüz görüntüsü | 17 | 24 | 7 |
| Doğruluk oranı (%) | 43,33 | 20 | 76,66 |

Tablo 2'de elde edilen bulgularda en fazla yanlış bulunan yüz görüntüsü sol pozda, en fazla doğru bulunan yüz görüntüsü orta pozda olmuştur. Doğruluk oranları kıyaslandığında, en yüksek değer %76,66 bulunarak orta pozda gerçekleşmiştir.

Fisherfaces algoritmasında görüntüler sağ, sol ve ortadan olmak üzere üç ayrı kategoride incelenmiştir. Elde edilen yüz tanıma performansları Tablo 3'te gösterilmiştir.

**Tablo 3.** Fisherfaces ile yüz tanıma doğruluk oranı

| Fisherfaces | Sağ poz | Sol poz | Orta poz |
|---|---|---|---|
| Toplam yüz görüntüsü | 120 | 120 | 120 |
| Toplam test görüntüsü | 30 | 30 | 30 |
| Toplam eğitim görüntüsü | 90 | 90 | 90 |
| Doğru yüz görüntüsü | 18 | 6 | 22 |
| Yanlış yüz görüntüsü | 12 | 24 | 8 |
| Doğruluk oranı (%) | 60 | 20 | 73,33 |

Orta pozda en yüksek doğru yüz görüntüsü elde edilmiştir. Sol poz için en düşük doğruluk oranı %20 olduğu gözlenmiştir. Doğruluk oranlarında en yüksek başarım %73,33 olarak orta pozda bulunmuştur.

### 3.2 Derin Öğrenme Mimarileriyle İlgili Bulgular

Derin öğrenme mimarileriyle yüz tanımada doğruluk oranlarını attırabilmek için epoch sayı aralıkları ile çalışmalar yapılmıştır. AlexNet mimarisiyle yüz tanıma metodunda 3-10 epoch sayı aralığında denemeler yapılmıştır. Yapılan çalışmada elde edilen bulgular Tablo 4'te gösterilmiştir.

**Tablo 4.** AlexNet ile yüz tanıma için epoch sayıları ve doğruluk oranları

| AlexNet | Görüntü Sayısı | | | | |
|---|---|---|---|---|---|
| Epoch Değeri | 120 | 462 | 30 sağ | 30 sol | 30 orta |
| 3 | 77,78 | 76,45 | 87,27 | 88,46 | 82,69 |
| 4 | 83,95 | 81,61 | 90,91 | 88,46 | 88,46 |
| 5 | 91,36 | 87,10 | 90,91 | 92,31 | 86,54 |
| 6 | 83,95 | 88,06 | 94,55 | 90,38 | 92,31 |
| 7 | 86,42 | 89,68 | 92,73 | 90,38 | 90,38 |
| 8 | 79,01 | 92,26 | 90,91 | 92,31 | 88,46 |
| 9 | 90,12 | 92,90 | 90,91 | 88,46 | 92,31 |
| 10 | 90,12 | 92,58 | 96,36 | 94,23 | 96,15 |

AlexNet mimarisi doğruluk oranlarına göre; 120 görüntü için epoch değeri 5, 462 görüntü için 8, 30 sağ görüntü için 10, 30 sol görüntü için 10 ve 30 orta görüntü için 10 olarak belirlenmiştir. En yüksek doğruluk oranı %96,36 olarak sağ görüntüde bulunmuştur.

ResNet-18 mimarisiyle yüz tanımada daha iyi bulgular elde edebilmek için epoch sayı aralıkları ile denemeler yapılmıştır. Bulunan doğruluk oranları Tablo 5'te gösterilmiştir.

**Tablo 5.** ResNet-18 ile yüz tanıma için epoch sayıları ve doğruluk oranları

| ResNet-18 | Görüntü Sayısı | | | | |
|---|---|---|---|---|---|
| Epoch Değeri | 120 | 462 | 30 sağ | 30 sol | 30 orta |
| 3 | 86,42 | 91,61 | 84,62 | 84,62 | 86,54 |
| 4 | 93,83 | 72,26 | 86,54 | 86,54 | 96,15 |
| 5 | 88,89 | 81,29 | 92,31 | 88,46 | 86,54 |
| 6 | 96,30 | 88,39 | 90,38 | 92,31 | 92,31 |
| 7 | 98,77 | 92,26 | 98,08 | 92,31 | 96,15 |
| 8 | 96,30 | 92,90 | 94,23 | 94,23 | 100 |
| 9 | 97,53 | 98,71 | 94,23 | 94,23 | 98,08 |
| 10 | 97,53 | 95,81 | 92,31 | 92,31 | 96,15 |

Derin öğrenme mimarileri için oluşturulan 462 görüntü ve yüz tanıma yöntemleri için oluşturulan 120 görüntünün tamamı ve bu görüntülerin pozlara göre üç farklı görüntü kümeleri kullanılmıştır. Epoch değer aralığında 8 olarak belirlenen denemede en yüksek doğruluk oranı orta pozda %100 olarak gözlenmiştir.

GoogleNet mimarisiyle yüz tanımanın sağlanabilmesi için epoch değerleri ile denemeler yapıldığında bulunan doğruluk oranları için bulgular aşağıdaki Tablo 6'daki gibidir.

**Tablo 6.** GoogleNet ile yüz tanıma için epoch sayıları ve doğruluk oranları

| GoogleNet | Görüntü Sayısı | | | | |
|---|---|---|---|---|---|
| Epoch Değeri | 120 | 462 | 30 sağ | 30 sol | 30 orta |
| 3 | 71,60 | 69,03 | 80,77 | 84,62 | 84,62 |
| 4 | 71,60 | 76,77 | 88,46 | 88,46 | 86,54 |
| 5 | 76,54 | 75,81 | 88,46 | 88,46 | 82,69 |
| 6 | 79,01 | 86,15 | 88,46 | 88,46 | 84,62 |
| 7 | 86,42 | 87,74 | 88,54 | 88,46 | 86,54 |
| 8 | 86,42 | 86,45 | 88,46 | 92,31 | 90,38 |
| 9 | 90,12 | 90,97 | 94,23 | 84,62 | 92,31 |
| 10 | 93,83 | 89,35 | 94,23 | 94,23 | 90,38 |

Verilen değer aralıklarına göre en yüksek doğruluk %94,23 olarak sağ ve sol pozda, 9 ve 10 epoch değerlerinde bulunmuştur.

Derin öğrenme mimarilerinden SqueezeNet ile yüz tanıma için epoch değerlerinde elde edilen bulgular Tablo 7'de verilmiştir.

**Tablo 7.** SqueezeNet ile yüz tanıma için epoch sayıları ve doğruluk oranları

| SqueezeNet | Görüntü Sayısı | | | | |
|---|---|---|---|---|---|
| Epoch Değeri | 120 | 462 | 30 sağ | 30 sol | 30 orta |
| 3 | 61,73 | 63,87 | 83,64 | 84,62 | 82,69 |
| 4 | 69,14 | 75,48 | 83,64 | 84,62 | 84,62 |
| 5 | 75,31 | 79,35 | 83,64 | 86,54 | 82,69 |
| 6 | 80,25 | 82,58 | 85,45 | 86,54 | 86,54 |
| 7 | 81,48 | 88,39 | 87,27 | 92,31 | 90,38 |
| 8 | 77,78 | 88,06 | 89,09 | 92,31 | 88,46 |
| 9 | 91,36 | 90 | 87,27 | 82,69 | 84,62 |
| 10 | 97,53 | 85,16 | 94,55 | 84,62 | 84,62 |

SqueezeNet mimarisiyle yüz tanımada epoch sayılarının seçilmesinde 120 görüntüde 10, 462 görüntüde 9, sağda 10, solda 7 ve orta pozda 7 değerlerinde başarı oranlarının arttığı görülmüştür. En yüksek doğruluk oranı %97,53 olarak bulunmuştur.

Derin öğrenme mimarileri için Tablo 4, Tablo 5, Tablo 6 ve Tablo 7'de verilen doğruluk değerleri veri kümesinin %75 eğitim ve %25 test olarak ayrılması ile yapılan eğitim sonucunda, test veri kümesindeki performanslarını göstermektedir.

Tablo 4, Tablo 5, Tablo 6 ve Tablo 7'ye göre her bir derin öğrenme modeli için en yüksek doğruluk değerinin elde edildiği maksimum epoch değeri kullanılarak, derin öğrenme modelleri üzerinde 5 katlı çapraz doğrulama uygulanmıştır. Çapraz doğrulamada, derin öğrenme mimarileri için kullanılan yığın boyutu (mini batch size), maksimum epoch değeri (max epoch) ve başlangıç öğrenme katsayısı (initial learning rate) değerleri Tablo 8'de verilmiştir.

**Tablo 8.** Derin öğrenme mimarilerinde kullanılan hiper parametreler

| Mimari Modeli | Yığın Boyutu | Maksimum Epoch | Öğrenme Katsayısı |
|---|---|---|---|
| AlexNet | 15 | 10 | 0,0001 |
| ResNet-18 | 10 | 3 | 0,0001 |
| GoogleNet | 15 | 9 | 0,0003 |
| SqueezeNet | 15 | 5 | 0,0003 |

Tablo 8'de verilen yığın boyutu ve başlangıç öğrenme katsayısı değerleri, aynı zamanda veri kümesinin %75 eğitim ve %25 test şeklinde ayrımı ile yapılan performans değerlendirmesinde de kullanılmıştır.

**Tablo 9.** Derin öğrenme mimarileri için çapraz doğrulama sonuçları

| Mimari Model | Görüntü Sayısı | | | | |
|---|---|---|---|---|---|
| Çapraz Doğrulama | 120 | 462 | 30 sağ | 30 sol | 30 orta |
| AlexNet | 73,6 | 84,6 | 73,8 | 67,8 | 60,8 |
| ResNet-18 | 96,8 | 97 | 84,4 | 73,2 | 78,2 |
| GoogleNet | 84,4 | 86,6 | 62,8 | 60,8 | 64 |
| SqueezeNet | 88,8 | 93,6 | 76,6 | 58,8 | 60 |

Tablo 9'a göre en yüksek doğruluk değeri tüm mimariler için 462 resmin bulunduğu veri kümesi ile elde edilmiştir. AlexNet için en yüksek doğruluk değeri %84,6. ResNet için %97, GoogleNet için %86,6 ve SqueezeNet için %93,6 olarak bulunmuştur.

Bu çalışmada bulunan sonuçların, literatürde FEI veri kümesi kullanılarak yapılan çalışmalarla karşılaştırılması Tablo 10'da verilmiştir.

**Tablo 10.** FEI veri kümesiyle yapılan çalışmalar

| Referans | Mimari Model | Doğruluk Oranı |
|---|---|---|
| Almabdy ve Elrefaei (2019) | AlexNet+SVM | 97,50 |
| Curtidor vd., (2021) | RLD | 93,57 |
| Hassan vd., (2021) | LDA | 97,84 |
| Win vd., (2021) | MTCNN | 86,80 |
| Ayata ve Çavuş (2022) | ESA | 98,86 |
| Yapılan Çalışma | ResNet-18 | 97 |

Bu çalışmada kullanılan Resnet-18, Curtidor vd. ile Win vd.'nin çalışmalarından daha iyi sonuç vermiştir. Almabdy ve Elrefaei ile Hassan vd.'nin çalışmaları ile hemen hemen aynı sonucu vermiştir. Ancak Ayakta ve Çavuş'un çalışmasından biraz daha düşük performanstadır.

## 4. SONUÇLAR

Bu çalışmada, yüz tanıma algoritması görüntü tabanlı yöntemler ve derin öğrenme mimarileri kullanılarak eğitilmiş, doğruluk oranları kıyaslanmıştır. Yüz tanımanın ön adımında kullanılan yüz algılama için Viola-Jones algoritması tercih edilmiştir. Kare içinde bulunan yüz görüntüleri veri tabanı olarak kaydedilmiştir.

Yüz tanıma teknikleri içerisinden Temel Bileşen Analizi ile Eigenfaces yöntemi ve Doğrusal Ayrım Analizi ile Fisherfaces yöntemi kullanılmıştır. Öncelikle Eigenfaces algoritması kullanılarak sağ, sol ve ön olmak üzere üç farklı yüz görüntüsü

incelenmiştir. Test görüntüsünden seçilen resim bu metotlar aracılığıyla eğitime giren resimlerle karşılaştırılmıştır. Aynı kişiye ait olan görüntüler doğru, farklı kişiye ait olan görüntü eşleşmeleri yanlış olarak belirlenmiştir. Ön pozisyonlarda yüz tanıma sıklığının arttığı gözlemlenmiştir. Bununla birlikte, Fisherfaces algoritması incelendiğinde ön pozun yüz tanıma oranı en iyi sonucu vermiştir. İki algoritma birbiriyle kıyaslandığında, Eigenfaces algoritması %76,66 yüksek doğruluk oranına sahiptir.

Derin öğrenme yöntemlerinden AlexNet, ResNet-18, GoogleNet ve SqueezeNet mimarileri kullanılmıştır. Görüntüler sağ, sol, ön ve hepsi olarak dört farklı kategoride incelenmiştir. İlk olarak veri kümesi %75 eğitim ve %25 test olarak ayrılarak derin öğrenme yöntemlerinin performansları değerlendirilmiştir. En yüksek doğruluk oranını veren epoch sayısının belirlenebilmesi için 3-10 aralığında epoch değerleri için denemeler yapılmıştır. Elde edilen sonuçlarda AlexNet'te 10 epoch, ResNet-18'de 8 epoch, GoogleNet'te 9 ve 10 epoch değerlerinde, SqueezeNet'te 10 epoch değerlerinde en yüksek doğruluk oranları elde edilmiştir. Mimariler birbirleriyle karşılaştırıldığında en yüksek başarım ResNet-18 için %100 olarak bulunmuştur.

Ayrıca, derin öğrenme mimarileri üzerinde 5 katlı çapraz doğrulama uygulanmıştır. Elde edilen sonuçlara göre, en yüksek doğruluk oranı ResNet-18 için %97 olarak elde edilmiştir.

120, 462, 30 sağ, 30 sol ve 30 orta görüntü içerisinden en düşük performans Eigenfaces ve Fiherfaces için 30 sol pozda olmuştur. Görüntülerdeki sağ sol poz oranlarının doğruluk sayısını etkilediği gözlenmiştir. Derin öğrenme mimarilerinde ise AlexNet, ResNet-18 ve SqueezeNet'te 30 sol pozda en düşük sonuçlar elde edilmiştir. GoogleNet mimarisinde ise 30 sağ pozda en düşük sonucu vermiştir. Görüntü sayısı arttıkça ve epoch değeri düştükçe elde edilen doğruluk oranlarının azaldığı görülmüştür. Çapraz doğrulama için en düşük sonuç 30 sol pozda çıkmıştır. Doğruluğun arttırılması ve görüntülerdeki parametre değerlerine bağlı değişikliklerin en iyi performansı vermesi için daha farklı çalışmalar yapılabilir.

Sonuç olarak derin öğrenme mimarilerinin, görüntü işleme tabanlı yöntemlere kıyasla daha iyi sonuçlar verdiği gözlenmiştir. Görüntü işleme alanında çok popüler olan bu yöntemlerin gelecekte daha iyi sonuçlar vereceğine inanılmaktadır.

## KAYNAKÇA


Aly M., Face Recognition using SIFT Features, CNS 186 Term Project Winter, 2006.

Almabdy S., ve Elrefaei L., Deep Convolutional Neural Network-Based Approaches for Face Recognition, *Applied Sciences,* 9(20), 4397, 2019.

Arya Z., ve Twiari, V. (2020). Automatic Face Recognition and Detection Using OpenCV, Haar Cascade and Recognizer at Different Angle of Face, *International Journal of Engineering Research and Applications*, 10(6), 2020.

Ayata F., ve Çavuş H., (2022). Yüz Tanıma Sistemlerinde Kullanılan ESA, YGH-DVM ve DSA Algoritmalarının Performans Testleri, *Fırat Üniversitesi Fen Bilimleri Dergisi*, 34(1), 39-48.

Balanageshwara, S., Kareem, A., ve Kumara, V. (2023, June). Machine Learning Approach for a Novel Facial Recognition System. In *2023 8th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1178-1183). IEEE.

Cadena, J., Villa, M., Martínez, M., Acurio, J., ve Chacón, L. (2023). An Efficient Technique for Global Facial Recognition using Python and OpenCV in 2D Images. *WSEAS Transactions on Systems and Control*, *18*, 47-57.

Curtidor A., Baydyk T., ve Kussul E., (2021). Analysis of Random Local Descriptors in Face Recognition, *Electronics,* 10(11), 1358. https://doi.org/10.3390/electronics10111358

Çakmakoğlu A., (2018). *Finans Sektöründe Veri Madenciliği Teknikleri Kullanılarak Kampanya Modelleme,* Yüksek Lisans Tezi, İstanbul Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Doğan F., & Türkoğlu İ., Derin Öğrenme Modelleri ve Uygulama Alanlarına İlişkin Bir Derleme, *DÜMF Mühendislik Dergisi,* 10(2), 419-424, 2019.

Fırıldak K. ve Talu F. M., (2019). Evrişimsel Sinir Ağlarında Kullanılan Transfer Öğrenme Yaklaşımlarının İncelenmesi, *Computer Science*, 4(2), 88-95.

Hassan M. M., Hussein H. I., Eesa A. S., ve Mstafa R. J., (2021). Face redognition based on gabor feature extraction followed by fastica and LDA, *Computers, Materials & Continua,* 68(2), 1637-1659. https://doi.org/10.32604/cmc.2021.016467

Holat R., ve Kulaç S., (2014). *Yüz Bulma ve Tanıma Sistemleri Kullanarak Kimlik Tespitinin Yapılması,* Yüksek Lisans Tezi, Düzce Üniversitesi, Fen Bilimleri Enstitüsü, Düzce.

Iandola F. N., Han S., Moskewicz M. W., Ashraf K., Dally W. J., ve Keutzer K., SqueezeNet: AlexNet-Level Accuracy with 50x Fewer Parameters and <1MB Model Size, ArXiv, 1-5, 2016.

ImageNet Large Scale Visual Recognition Challenge 2014 (ILSVRC2014), 14 Ağustos 2024 tarihinde https://image-net.org/challenges/LSVRC/2014/results adresinden erişildi.

Jones, M., ve Viola, P. (2003). Fast Multi-view Face Detection, Mitsubishi Electric Research Laboratories, TR-20003-96.

Kekül, H., Bircan, H., ve Arslan, H. (2018). Yüz Tanıma Uygulamalarında Özyüzler ve Yapay Sinir Ağlarının Karşılaştırılması, *Uluslarası Yönetim Bilişim Sistemleri ve Bilgisayar Bilimleri Dergisi*, 2(1), 52.

Krizhevsky A., Sutskever I., ve Hinton G. E., (2012). ImageNet Classification with Deep Convolutional Neural Networks, *Advances in Neural Information Processing Systems*, 25(2),1097-1105.

Raghu S., Sriraam N., Temel Y., Rao S. V., ve Kubben P. L., EEG based multi-class seizure type classification using convolutional neural network and transfer learning*, Neural Networks*: the Official Journal *of the International Neural Network Society*, 206, 2020.

Rahmad C., Asmara R. A., Putra D. R. H., Dharma I, Darmono H., ve Muhiqqin I., (2020). Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection, *IOP Conference Series: Materials Science and Engineering*, 732(1), 012038.

Sarwinda D., Paradisa R. H., Bustamam A., ve Anggia P., (2021). Deep Learning in Image Classification using Residual Network (ResNet) Variants for Detection of Colorectal Cancer, *Procedia Computer Science,* 423-431(179), 1877-0509.

Sert Z., (2020, 27 Aralık). ESA (Evrişimsel Sinir Ağları) [Blog yazısı]. 13 Ağustos 2024 tarihinde erişim adresi: https://zeysert.medium.com/esa-evri%C5%9Fimsel-sinir-a%C4%9Flar%C4%B1-87d9bd986579

Sütçüler E., (2006). *Gerçek Zamanlı Video Görüntülerinden Yüz Bulma ve Tanıma Sistemi*, Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi, Fen Bilimleri Enstitüsü, İstanbul.

Şeker A., (2018). Evaluation of Fabric Defect Detection Based on Transfer Learning with Pre-trained AlexNet, *2018 International Conference on Artificial Intelligence and Data Processing (IDAP),* Malatya, Turkey, pp. 1-4.

Thomaz C. E., (2012). *FEI Face Database*, 13 Ağustos 2024 tarihinde erişim adresi: https://fei.edu.br/~cet/facedatabase.html

Topal A. O., Chitic R., ve Leprévost F. (2023). One evolutionary algorithm deceives humans and ten convolutional neural networks trained on ImageNet at image recognition, *Applied Soft Computing,* 143, 110397. doi: https://doi.org/10.1016/j.asoc.2023.110397

Wang J. ve Li Z., (2018). Research on Face Recognition Based on CNN*, IOP Conference Series: Earth and Environmental Science*, 170(3), 032110.

Wang D., Yu H., Wang D. ve Li G., (2020). Face Recognition System Based on CNN, *2020 International Conference on Computer Information and Big Data Applications (CIBDA),* 470-473.

Win H. P. P., Khine P. T. T., ve Tun K. N. N., (2021). Face Recognition System based on Convolution Neural Networks, *International Journal of Image Graphics and Signal Processing* (IJIGSP), 13(6), 23-29.

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Derleme Makalesi

# Siber Fiziksel Sistemler Alanında Türkiye'deki Akademik Eğilimler: Bir Bibliyometrik Analiz

**Ayşegül Yüksel[1], Tamer Eren[1*], Emel Güven[1]**

*[1]Kırıkkale Üniversitesi, Mühendislik ve Doğa Bilimleri Fakültesi, Endüstri Mühendisliği Bölümü, Kırıkkale, Türkiye*

**Anahtar Kelimeler:**
Siber Fiziksel Sistemler, Bibliyometrik Analiz, Lisansüstü Tez

**ÖZ**

Siber fiziksel sistemler (SFS), fiziksel dünyadaki süreçlerin ve bilgisayar tabanlı kontrol mekanizmalarının entegrasyonunu sağlayan sistemlerdir. Bu sistemler, endüstriyel otomasyondan sağlık hizmetlerine kadar geniş bir yelpazede uygulama alanına sahiptir. Bu çalışmanın amacı, Türkiye'de SFS üzerine yapılan lisansüstü tezlerin bibliyometrik analizini yaparak, bu alandaki araştırma eğilimlerini ve mevcut durumu anlamaktır. Çalışma kapsamında, Türkiye'de yayınlanmış 80 lisansüstü tez incelenmiştir. Yapılan analizler, Türkiye'de SFS alanına olan ilginin son yıllarda arttığını göstermiştir. Bu analizler en fazla tezin 2019 yılında yayınlandığını göstermektedir, tezlerin %85'i yüksek lisans, %15'i doktora tezidir, tezlerin %95'i devlet üniversitelerinde, en fazla ise İstanbul Teknik Üniversitesi'nde yapılmıştır. Sosyal Bilimler Enstitüsü en fazla tez yayınlanan enstitü olmuştur. Ana bilim dalları arasında İşletme ve Endüstri Mühendisliği öne çıkmaktadır. Tezler genellikle nicel yöntemler kullanılarak hazırlanmıştır. Anahtar kelime analizleri, SFS çalışmalarının "endüstri 4.0", "akıllı sistemler" ve "siber güvenlik" gibi temalarda yoğunlaştığını ortaya koymuştur. Elde edilen bulgular, Türkiye'deki SFS araştırmalarının genel durumu hakkında kapsamlı bir bilgi sunmakta ve gelecekte yapılacak çalışmalar için yol gösterici öneriler sunmaktadır. Çalışmanın sonuçları, Türkiye'de SFS alanındaki araştırmaların daha da gelişmesi için disiplinler arası iş birliklerinin artırılması, siber güvenlik ve etik konularının derinlemesine incelenmesi ve uluslararası iş birliklerinin güçlendirilmesi gerektiğini önermektedir.

# Academic Trends in Cyber-Physical Systems in Turkey: A Bibliometric Analysis

**Keywords:**
Cyber Physical Systems, Bibliometric Analysis, Postgraduate Thesis

**ABSTRACT**

Cyber-physical systems (CPS) are systems that enable the integration of processes in the physical world and computer-based control mechanisms. These systems have a wide range of applications from industrial automation to healthcare. The aim of this study is to understand the research trends and status in this field by conducting a bibliometric analysis of postgraduate theses on CPS in Turkey. Within the scope of the study, 80 postgraduate theses published in Turkey were examined. The analyses conducted showed that the interest in the field of CPS in Turkey has increased in recent years. These analyses show that the most theses were published in 2019, 85% of the theses were master's theses, 15% were doctoral theses, 95% of theses were made in state universities, and the most were made in Istanbul Technical University. The Institute of Social Sciences was the institute with the most theses published. Business and Industrial Engineering stand out among the main branches of science. Theses were generally prepared using quantitative methods. Keyword analyses revealed that SFS studies are concentrated on themes such as "industry 4.0", "smart systems" and "cyber security". The findings provide comprehensive information about the general status of SFS research in Turkey and provide guiding suggestions for future studies. The results of the study suggest that interdisciplinary collaborations should be increased, cybersecurity and ethical issues should be examined in depth, and international collaborations should be strengthened for the further development of SFS research in Türkiye.

[*]**Sorumlu Yazar**

(deringozaysegul@gmail.com), ORCID ID: 0000-0001-9999-0531
[*](tamereren@gmail.com) ORCID ID: 0000-0001-5282-3138
(emel-gvn@hotmail.com) ORCID ID: 0000-0001-6106-9720

**e-ISSN: 2717-8579**
**Geliş Tarihi: 09/08/2024; Kabul Tarihi: 02/12/2024**

## 1. GİRİŞ

Siber fiziksel sistemler (SFS), fiziksel süreçlerin ve bilgisayar tabanlı kontrol mekanizmalarının entegrasyonunu sağlayan, karmaşık ve çok katmanlı sistemlerdir. Bu sistemler, sensörler ve aktüatörler aracılığıyla fiziksel dünyadan veri toplar ve bu verileri analiz ederek belirli eylemleri gerçekleştirir. Endüstri 4.0'ın öncüsü olarak kabul edilen SFS'ler, akıllı üretim tesislerinden sağlık hizmetlerine, enerji yönetiminden akıllı şehir altyapılarına kadar geniş bir yelpazede uygulama alanına sahiptir (Lee, 2015; Baheti ve Gill, 2011). Bu sistemlerin, süreç optimizasyonu, enerji verimliliği ve güvenlik gibi kritik alanlarda sağladığı faydalar, onların modern endüstriyel ve sosyal altyapının ayrılmaz bir parçası haline gelmesine neden olmuştur (Karnouskos, 2011). Endüstri 5.0'ın, Endüstri 4.0 üzerine inşa edilen bir paradigma olarak, insan odaklı üretim süreçlerini teknolojiyle bütünleştirdiği vurgulanmaktadır. Ancak, Endüstri 5.0'ın getirdiği siber güvenlik riskleri de dikkate değerdir. Yeni sanayileşmiş ülkelerin siber güvenlik düzeyleri derinlemesine incelenmiş ve ülkelerin siber güvenliğe yönelik yaklaşımlarının bu geçiş sürecindeki kritik rolü ortaya konulmuştur (Duran, 2024).

SFS'lerin gelişimi, bilgi teknolojileri ve mühendislik disiplinlerinin bir araya gelmesiyle mümkün olmuştur. Bilgisayar bilimleri, kontrol teorisi, sensör teknolojisi ve ağ iletişimi gibi alanların bir arada kullanılması, bu sistemlerin karmaşıklığını ve işlevselliğini artırmaktadır (Rajkumar, Lee, ve Sha, 2010). Bu bağlamda, siber güvenlik de SFS'ler için kritik bir unsur haline gelmiştir. Sistemlerin kesintisiz ve güvenli bir şekilde çalışmasını sağlamak, potansiyel tehditlere karşı alınacak önlemlerle mümkündür (Öztürk, 2020). SFS alanında gerçekleştirilen bazı çalışmalar şu şekildedir; Bhadani (2024), akıllı şebekelerin SFS perspektifinden incelenmesini sağlamış ve bu entegrasyonun getirdiği teknik zorlukları, faydaları ve SFS'nin akıllı şebekelere katkı sağlama potansiyelini detaylandırmıştır. Harkat vd. (2024), SFS güvenliğine yönelik mevcut tehditler, saldırı türleri ve savunma mekanizmaları sistematik olarak incelenmiştir. Çalışma, SFS güvenlik açıklarının değerlendirilmesi ve güvenlik yapılarının güçlendirilmesinin yanı sıra etik ve toplumsal etkilerin de önemini vurgulamaktadır. Lou vd. (2024), insan-merkezli siber-fiziksel sistemlerin tasarım, üretim ve hizmet aşamalarındaki etkileri incelemiş, özellikle insan ve teknoloji entegrasyonunun, tasarım ve üretim süreçlerini optimize etmede oynadığı rolü vurgulamışlardır. Çalışma, insan-merkezli siber-fiziksel sistemler ile insan-robot etkileşimi, dijital ikiz, veri entegrasyonu ve toplumsal etkileşime yönelik teknolojiler gibi önemli konuları detaylandırmaktadır. Yu vd. (2023), SFS'nin güvenlik tehditlerini fiziksel, siber ve siber-fiziksel alanlarda ele alarak bu alanlardaki savunma mekanizmalarını kapsamlı bir şekilde incelemişlerdir. Çalışmada, SFS'nin artan güvenlik riskleri karşısında geliştirilmesi gereken savunma stratejileri ve gelecekteki araştırma yönelimleri de tartışılmaktadır. Canonico ve Sperlì (2023), sanayi üretim tesislerinde siber-fiziksel sistemlerin güvenliği kapsamlı bir şekilde ele alınmıştır. Çalışma, endüstriyel kontrol sistemlerini hedefleyen siber saldırı türlerini ve bu saldırılara karşı geliştirilen model tabanlı ve yapay zekâ tabanlı güvenlik önlemlerini sınıflandırarak analiz etmektedir. Bu yaklaşım, SFS güvenliğinde güncel tehditlerle başa çıkmak ve etkin savunma stratejileri oluşturmak için önemli bir metodolojik çerçeve sunmaktadır.

Türkiye'de SFS'ler konusundaki akademik çalışmaların artması, bu alandaki araştırmaların önemini vurgulamaktadır. Türk araştırmacılar, bu alanda hem teorik hem de uygulamalı çalışmalar yaparak literatüre önemli katkılarda bulunmuştur. Örneğin, Demirci (2019) SFS'lerin endüstriyel uygulamalarda kullanımına dair detaylı analizler sunmuştur. Ancak, bu çalışmaların kapsamlı bir bibliyometrik analizi yapılmamıştır. Bu eksiklik, Türkiye'deki SFS araştırmalarının genel durumunu ve bu alandaki eğilimleri tam anlamıyla anlamayı zorlaştırmaktadır.

Bu çalışmanın amacı, Türkiye'de SFS'ler üzerine yapılan 80 lisansüstü tezi bibliyometrik yöntemlerle analiz etmektir. Bu analizde, tezlerin yıllara göre dağılımı, kullanılan yöntemler, anahtar kelimeler, üniversiteler ve enstitüler gibi çeşitli kriterler değerlendirilecektir. Ayrıca, bu çalışmaların hangi konular etrafında yoğunlaştığı ve hangilerinin daha fazla ilgi gördüğü de incelenecektir. Böylece, Türkiye'deki SFS araştırmalarının mevcut durumu hakkında kapsamlı bir bilgi elde edilerek, gelecekteki araştırmalar için yönlendirici bilgiler sunulacaktır.

SFS'lerin hızlı gelişimi ve geniş uygulama alanı, bu alandaki bilimsel literatürün de hızla büyümesine yol açmıştır. Bu çalışmada önceki literatür taramaları ve mevcut çalışmalar ışığında SFS'lerin önemi ve bu alandaki araştırma boşlukları ele alınacaktır. Özellikle, Türkiye'de yapılan çalışmaların küresel literatürdeki yeri ve bu alandaki uluslararası iş birlikleri değerlendirilecektir. Çalışmanın sonunda, SFS'ler üzerine gelecekte yapılacak araştırmalar için öneriler sunulacaktır. Bu öneriler hem akademik araştırmaların derinleşmesini hem de uygulamalı projelerin geliştirilmesini teşvik edecektir.

## 2. BİBLİYOMETRİK ANALİZ

Bibliyometrik analizler, bilimsel literatürdeki dinamikleri ve araştırma alanlarındaki değişimleri anlamak için önemli bir araçtır. Bu yöntemle yapılan çalışmalar, belirli bir alanın zaman içindeki gelişimini ve bu alandaki araştırma eğilimlerini ortaya koyar (Karademir ve Akın, 2021). Örneğin, mühendislik alanındaki yayınların bibliyometrik analizini yapan Çakmak ve Öztürk (2019), Türkiye'deki mühendislik araştırmalarının son yıllarda nasıl bir gelişim gösterdiğini ve hangi

konuların öne çıktığını detaylı bir şekilde incelemişlerdir.

Bibliyometrik analizler, aynı zamanda araştırma iş birliklerinin ve atıf ağlarının haritalandırılmasına da olanak tanır. Şahin ve Uğurlu (2018), Türkiye'deki sağlık bilimleri alanındaki araştırma iş birliklerini analiz ederek, bu iş birliklerinin yayın kalitesine etkisini incelemişlerdir. Benzer şekilde, Demir ve Yılmaz (2020), eğitim bilimleri alanında yapılan tezlerin bibliyometrik analizini yaparak, bu alandaki araştırma trendlerini ve iş birliği ağlarını ortaya koymuşlardır.

Bibliyometrik analizler, araştırmacıların belirli bir alandaki bilgi birikimini ve mevcut araştırma yönelimlerini anlamalarına yardımcı olur. Ayrıca, bu analizler, gelecekteki araştırma yönelimlerini belirlemek ve stratejik araştırma planlamaları yapmak için de önemli bilgiler sunar (Özkan ve Erten, 2021). Özellikle, üniversitelerin akademik performanslarının değerlendirilmesinde bibliyometrik analizlerin kullanımı, kurumların güçlü ve zayıf yönlerini belirlemelerine ve bu doğrultuda stratejik planlamalar yapmalarına yardımcı olur (Turan ve Güner, 2022).

Bibliyometrik analizlerin bir diğer önemli katkısı da belirli bir araştırma alanında hangi konuların yoğun olarak çalışıldığını ve hangi konuların daha fazla araştırma gerektirdiğini belirlemektir. Örneğin, Karademir ve Akın (2021), sosyal bilimler alanında yapılan bibliyometrik analizlerle, bu alandaki araştırma boşluklarını ve gelecekteki araştırma ihtiyaçlarını tespit etmişlerdir. Bu tür analizler, akademik literatürün genişletilmesine ve araştırma stratejilerinin geliştirilmesine önemli katkılar sağlar. Tüm bu bilgiler ışığında;

Kaya ve Zeren (2020), yaptığı çalışmada dijital pazarlama alanında ulusal yayınların bibliyometrik analizini yapmıştır. Savrun ve Mutlu (2019), yaptıkları çalışmada "şehir lojistiği" anahtar kelimesini içeren kitap, makale ve konferans bildirisinin bibliyometrik analizini yapmıştır. Eren vd. (2024), insansız hava aracı alanında yapılan yüksek Öğretim Kurumu Ulusal Tez Merkezinde yayınlanmış 341 adet çalışmasının bibliyometrik analizini yapmıştır. Yapılan araştırma sonucunda son yıllarda bu alanda yapılan çalışmaların elektrik ve elektronik mühendisliğinde gerçekleştirildiği tespit edilmiştir. Tekin vd. (2021), yaptıkları çalışmada tersine lojistiği içeren 2016-2020 yılları arasında yayınlanan araştırmaların bibliyometrik analizini gerçekleştirmiştir. Bu konuda en fazla çalışma yapan ülkenin Çin Halk Cumhuriyeti olduğu tespit edilmiştir. Öztürk ve Kurutkan (2020), kalite yönetimi üzerine bilim haritalama tekniği uygulamıştır. 1372 makalenin bibliyometrik analizini gerçekleştirmiştir. Gündüz ve Eren (2024), kent dirençliliği konusunda yayınlanan ulusal tezlerin bibliyometrik analizini gerçekleştirmiştir. Bu konuda en fazla çalışılan konuların afet, kent tasarım, kentsel direnç, sürdürülebilirlik, planlama, iklim değişikliği, akıllı kent, direnç olduğu görülmüştür. Sanlı vd. (2024), "siber güvenlik" anahtar kelimesi içeren lisansüstü tezlerin bibliyometrik analizi yapılmıştır. Bu alanda yayınlanan tezlerin en fazla devlet üniversitelerinde yapıldığı görülmüştür. Uzan (2024), akıllı şehirler ve yönetişim kavramlarını bir arada içeren makalelerin bibliyometrik analizi yapmıştır. Şahin vd. (2024), sürdürülebilirlik değeri kavramı üzerine nicel verilerin bibliyometrik analizi gerçekleştirilmiştir. Elmas vd. (2024), dijital finans alanında bilimsel araştırmaların bibliyometrik analizi gerçekleştirilmiştir. Özsaatci (2022), yaptığı çalışmada "Sosyal medya pazarlaması" ifadesini içeren çalışmalar üzerine bibliyometrik analiz yapmıştır. Bu alanda en çok çalışma yapan ülkenin ABD olduğu tespit edilmiştir. Küpçüoğlu vd. (2024) çalışması, Türkiye'de blok zincir (blockchain) teknolojisi üzerine yazılan lisansüstü tezlerin bibliyometrik analizini yaparak, bu alandaki akademik eğilimleri detaylandırmaktadır. 476 yüksek lisans ve doktora tezi üzerinde yapılan analizde, tezlerin danışman, üniversite, tez türü, yazıldığı yıl, enstitü ve ana bilim dalına göre dağılımları incelenmiştir. Çalışma, blok zincir teknolojisinin Türkiye'deki akademik ilgisinin özellikle fen bilimleri enstitülerinde yoğunlaştığını ve en çok bilgisayar mühendisliği anabilim dalında tezler yazıldığını göstermektedir. Gaferoğlu vd. (2024), Türkiye'deki tsunami konulu lisansüstü tezlerin bibliyometrik analizini yaparak, bu alanda yapılan araştırmaların dağılımını ortaya koymuştur. Çalışmada, 124 tez; türü, yılı, üniversitesi, üniversite türü, enstitüsü, ana bilim dalı, konusu ve yöntemi gibi çeşitli parametreler çerçevesinde incelenmiştir. Analiz sonucunda en fazla çalışmanın devlet üniversitelerinde ve özellikle fen bilimleri enstitüsünde hazırlandığı belirlenmiştir. Öztürk vd. (2024), Türkiye'de sağlık turizmi teması üzerine yazılmış 220 lisansüstü tezin bibliyometrik analizini sunmaktadır. Çalışmada, tezlerin üniversite türü, enstitü, ana bilim dalı, konu ve yöntem gibi çeşitli parametrelere göre dağılımı incelenmiştir. Bulgular, sağlık turizmi alanındaki tezlerin büyük ölçüde sosyal bilimler enstitüsüne bağlı olarak yazıldığını ve devlet üniversitelerinde yoğunlaştığını göstermektedir. Bu analiz, Türkiye'de sağlık turizmi alanındaki akademik araştırmaların mevcut durumunu anlamak ve gelecekteki çalışmalara rehberlik etmek amacıyla önemli veriler sunmaktadır. Pınarcı vd. (2024), Türkiye'de ekip çizelgeleme konusundaki lisansüstü tezlerin bibliyometrik analizini yaparak, bu alandaki akademik eğilimleri incelemektedir. 1991-2024 yılları arasında yapılan 30 lisansüstü tez üzerinden gerçekleştirilen analizde, tezlerin tür, üniversite, bilim dalı, kullanılan yöntem ve konu gibi çeşitli parametrelere göre dağılımları incelenmiştir. Çalışmanın bulguları, ekip çizelgeleme çalışmalarının özellikle 2008 yılında yoğunlaştığını ve en fazla yüksek lisans tezi olarak hazırlandığını göstermektedir. Ayrıca, tezlerin büyük çoğunluğunun Endüstri Mühendisliği Ana Bilim

Dalı'nda yazıldığı belirlenmiştir. Güven ve Eren (2024), YÖKTEZ veri tabanında yer alan endüstriyel kazalar konulu lisansüstü tezlerin bibliyometrik analizini gerçekleştirmişlerdir. Analizde ele alınan 72 çalışmalar tez türü, tez yılı, tezin yazıldığı üniversite, üniversite türü, enstitü, ana bilim dalı, konu ve yöntemlerine göre incelenmiştir.

Bu çalışmada ise SFS üzerine yayınlanmış lisansüstü tezler incelenmiş olup, yapılan literatür araştırmasında bu konu ile ilgili bibliyometrik analize rastlanmamıştır. Çalışmanın bu yönüyle literatüre katkı sağlayacağı düşünülmektedir.

## 3. YÖNTEM

Bu çalışmanın amacı, Türkiye'de SFS üzerine yapılan lisansüstü tezlerin bibliyometrik analizini gerçekleştirmektir. Bu analiz, SFS alanındaki araştırmaların genel eğilimlerini ve önemli konuları ortaya koymayı amaçlamaktadır. Çalışma kapsamında, Türkiye'de yayınlanan 80 lisansüstü tez incelenmiştir. Tezler, Yükseköğretim Kurumu Ulusal Tez Merkezi (YÖKTEZ) üzerinden temin edilmiştir. Arama kısmında "siber fiziksel sistemler" yazılarak; tez adı, yazar, danışman, konu, dizin, özet ve tez no kısımlarında aramalar yapılmıştır. Araştırma, 5 Temmuz 2024 tarihine kadar yayınlanan tezleri içermektedir.

Bu tezlerle ilgili olarak; yıl, tez türü, konu, üniversite, ana bilim dalı, üniversite türü, enstitü, yöntem (nitel/nicel), yöntem ayrıntısı, anahtar kelimeler ve konu başlıklarını içeren bir veri seti oluşturulmuştur. Elde edilen bu veri seti ile SFS alanında yayınlanan lisansüstü tezlerin içerik analizi gerçekleştirilmiştir.

Bu çalışma çerçevesinde aşağıdaki sorulara yanıt aranmıştır:

- Yıllara göre tez sayılarında nasıl bir eğilim gözlemlenmektedir?
- Tez türü dağılımı nasıldır?
- Tezlerin yayınlandığı üniversitelerin türleri nelerdir?
- Tezlerin üniversitelere göre dağılımı nasıldır?
- Tezlerin enstitülere göre dağılımı nasıldır?
- Ana bilim dallarına göre tez dağılımı nasıldır?
- Hangi konular SFS alanında daha fazla araştırılmıştır?
- Tezlerde kullanılan yöntemler nelerdir?
- Tezlerde en sık kullanılan anahtar kelimeler nelerdir?

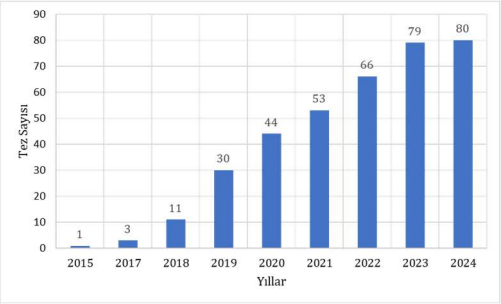
Bu sorulara verilen yanıtlar, Türkiye'de SFS alanında yapılan lisansüstü tezlerin genel eğilimlerini, önemli konuları ve araştırma yöntemlerini ortaya koyarak, alandaki durumu ve gelecekteki araştırma potansiyelini değerlendirmeyi amaçlamaktadır.

## 4. BULGULAR

Bu bölümde, Türkiye'de SFS alanında yapılan akademik çalışmaların bibliyometrik analizinden elde edilen bulgular sunulacaktır. Analiz, 80 lisansüstü tezi kapsamakta olup, çalışmaların yıllara göre dağılımı, kullanılan anahtar kelimeler, ana bilim dalları, konu, yöntem çeşitleri gibi çeşitli boyutlarını ele almaktadır.

### 4.1.Lisansüstü Tezlerin Yıllara Göre Kümülatif Dağılım Grafiği

Lisansüstü tezlerin yıllara göre kümülatif dağılım grafiği Şekil 1'de verilmiştir.

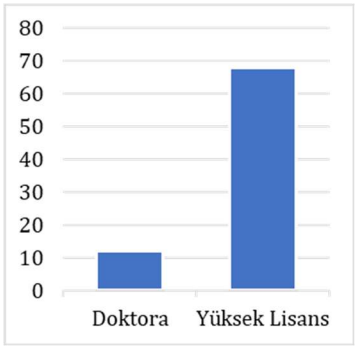



**Şekil 1:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Kümülatif Dağılımı

Verilen şekil incelendiğinde, en fazla çalışma %23,75 ile 2019 yılında yapılmıştır. Bunu %17,50 ile 2020 yılı, %16,25 ile 2022 ve 2023 yılları takip etmektedir. Bu yıllardaki yüksek tez sayısının nedenleri arasında SFS alanına olan ilginin artması ve teknolojik gelişmelerin hız kazanması sayılabilir. 2019'dan sonra tez sayılarında bir düşüş gözlemlenmiştir. Bu düşüşün sebepleri arasında küresel pandemi koşulları nedeniyle araştırma faaliyetlerinin yavaşlaması, laboratuvar ve saha çalışmalarının kısıtlanması ve ekonomik belirsizlikler nedeniyle araştırma bütçelerinde yaşanan azalmalar sayılabilir. Ayrıca bazı araştırmacıların tez çalışmalarını tamamlamakta zorlanmaları veya projelerini ertelemeleri de bu duruma neden olmuş olabilir

### 4.2.Lisansüstü Tezlerin Tez Türüne Göre Dağılımı

Şekil 2'de yayınlanan tezlerin türüne göre dağılımı verilmiştir.

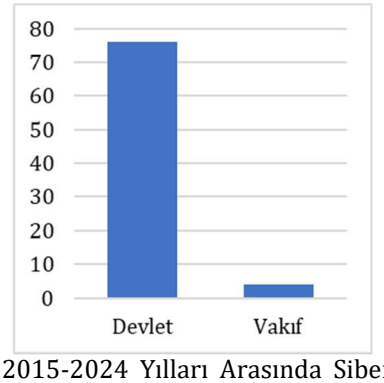



**Şekil 2:** Siber Fiziksel Sistemler Alanında 2015-2024 Yılları Arasında Tamamlanan Yüksek Lisans ve Doktora Tezlerinin Dağılımı

SFS alanında yazılan lisansüstü tezlerin türüne göre yapılan araştırma sonucunda, %15'inin doktora tezi, %85'inin ise yüksek lisans tezi olarak yayınlandığı görülmüştür. Bu sonuç, SFS alanında yüksek lisans tezlerinin daha fazla tercih edildiğini göstermektedir. Yüksek lisans tezlerinin daha fazla olmasının sebepleri arasında, bu alanda yapılan çalışmaların geniş bir uygulama alanına sahip olması ve yüksek lisans programlarının doktora programlarına göre daha kısa sürede tamamlanabilmesi yer almaktadır. Ayrıca, birçok öğrenci kariyerlerine hızlı bir başlangıç yapmak için yüksek lisans programlarını tercih etmektedir.

### 4.3.Lisansüstü Tezlerin Üniversite Türüne Göre Dağılımı

Şekil 3'te lisansüstü tezlerin üniversite türüne göre dağılımı grafiği verilmiştir.

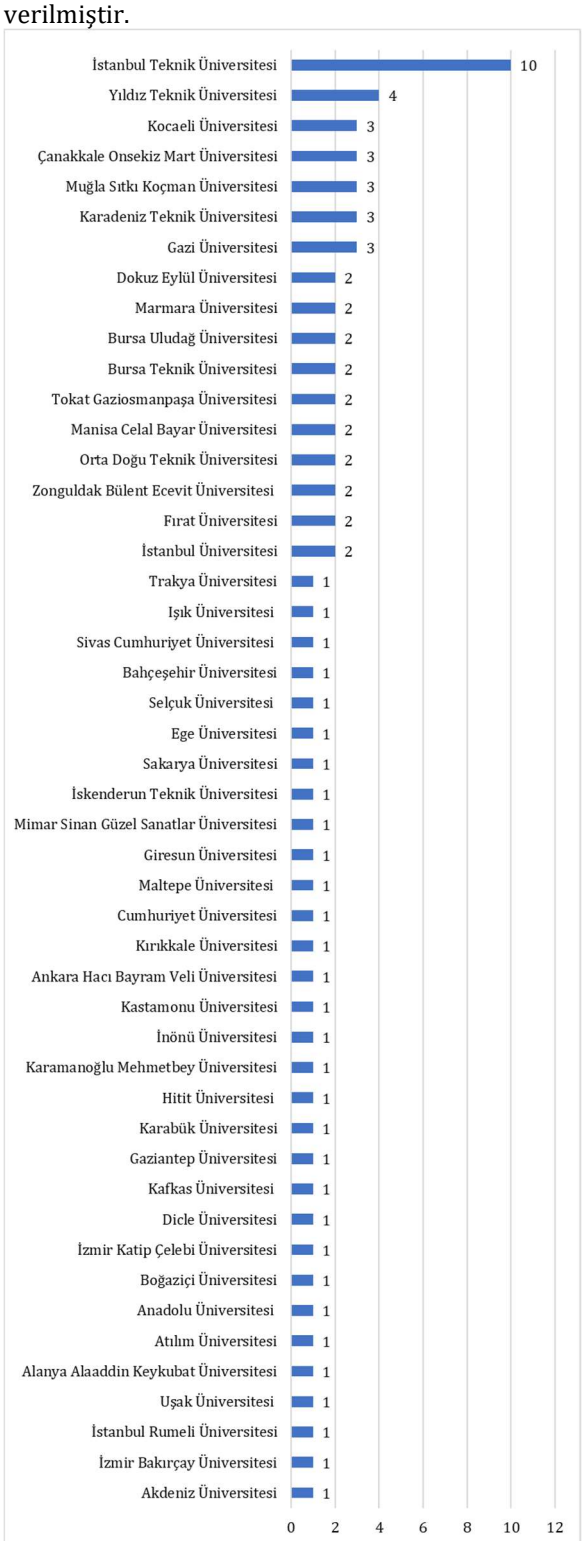



**Şekil 3:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Devlet ve Vakıf Üniversitelerine Göre Dağılımı

Oluşturulan veri setinin analizi sonucunda, SFS alanında yayınlanmış lisansüstü tezlerin %95'inin devlet üniversitelerinde, %5'inin ise vakıf üniversitelerinde yapıldığı görülmüştür. Bu sonuç, devlet üniversitelerinin SFS alanında daha fazla lisansüstü tez çalışmasına ev sahipliği yaptığını göstermektedir. Bunun sebepleri arasında, devlet üniversitelerinin daha geniş araştırma imkanlarına sahip olması, daha fazla kaynak ve laboratuvar imkânı sunması ve bu alandaki akademik kadronun daha geniş ve uzman olması sayılabilir. Ayrıca, devlet üniversitelerinin öğrenci sayısının daha fazla olması da bu duruma katkıda bulunmaktadır.

### 4.4.Lisansüstü Tezlerin Üniversitelere Göre Dağılımı

Bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin üniversitelere göre dağılımı incelenmiştir. Şekil 4'te lisansüstü tezlerin üniversitelere göre dağılımı grafiği verilmiştir.

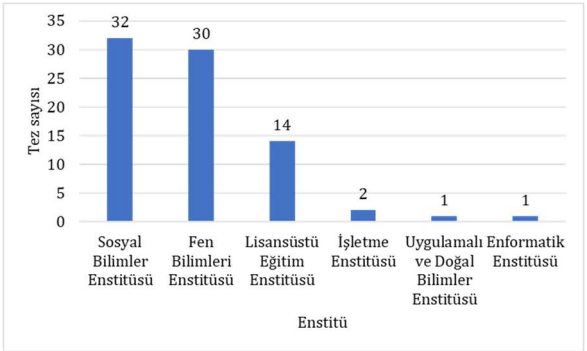



**Şekil 4:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Üniversitelere Göre Dağılımı

Analiz edilen 80 tezin hangi üniversitelerde yazıldığını gösteren tabloya göre, İstanbul Teknik Üniversitesi %12,5 ile en çok tez yayınlanan üniversite olmuştur. Bunu %5 ile Yıldız Teknik Üniversitesi takip etmektedir. Gazi Üniversitesi, Karadeniz Teknik Üniversitesi, Muğla Sıtkı Koçman Üniversitesi, Çanakkale Onsekiz Mart Üniversitesi ve Kocaeli Üniversitesi ise her biri %3,75 ile önemli bir yer tutmaktadır. İstanbul Üniversitesi, Fırat Üniversitesi, Zonguldak Bülent Ecevit Üniversitesi, Orta Doğu Teknik Üniversitesi, Manisa Celal Bayar Üniversitesi, Tokat Gaziosmanpaşa Üniversitesi, Bursa Teknik Üniversitesi, Bursa Uludağ Üniversitesi, Marmara Üniversitesi ve Dokuz Eylül Üniversitesi ise her biri %2,5 ile katkıda bulunmuşlardır. Diğer üniversiteler ise %1,25 oranında tez ile bu alana katkıda bulunmuşlardır. Bu dağılım, büyük ve teknik üniversitelerin, özellikle İstanbul Teknik Üniversitesi ve Yıldız Teknik Üniversitesi gibi kurumların, SFS alanında daha fazla araştırma ve tez çalışması yapıldığını göstermektedir. Bu durum, bu üniversitelerin geniş araştırma olanaklarına, ilgili laboratuvar ve altyapıya sahip olmaları ile açıklanabilir. Ayrıca, bu üniversitelerde SFS alanında çalışan uzman akademisyenlerin sayısının fazla olması da tez çalışmalarının yoğunluğunu artırabilir. Diğer yandan, birçok üniversitede sadece birer tez çalışması yapılmış olması, SFS alanındaki araştırma faaliyetlerinin Türkiye genelinde yaygınlaştığını ancak belirli üniversitelerde daha yoğunlaştığını göstermektedir. Bu durum, diğer üniversitelerde de bu alanda yapılan çalışmaların artırılması ve desteklenmesi gerektiğini ortaya koymaktadır.

### 4.5.Lisansüstü Tezlerin Enstitülere Göre Dağılımı

Bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin enstitülere göre dağılımı incelenmiştir. Şekil 5'te lisansüstü tezlerin enstitülere göre dağılımı grafiği verilmiştir.



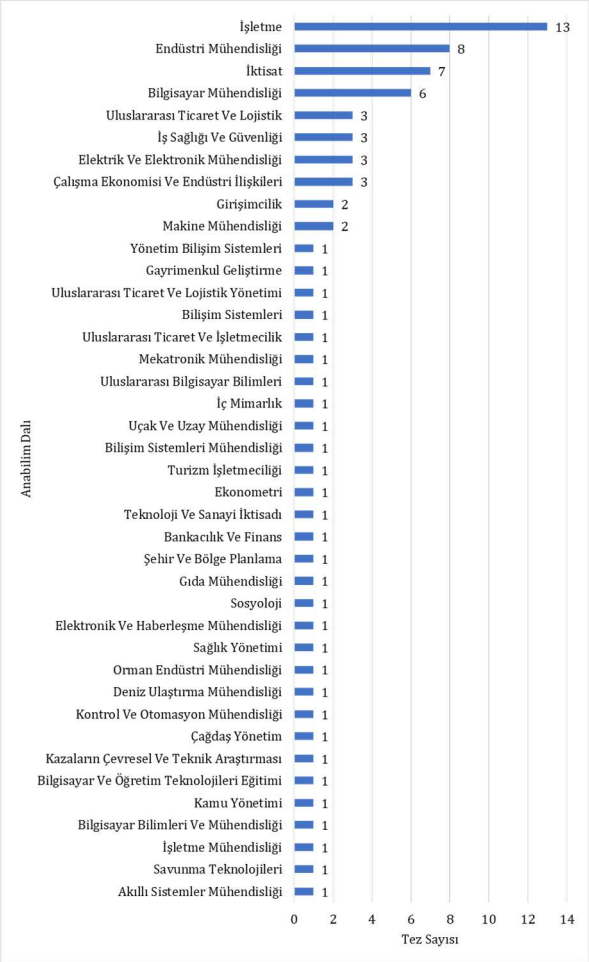

**Şekil 5:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Enstitülere Göre Dağılımı

Analiz edilen 80 tezin hangi enstitülerde yazıldığını gösteren tabloya göre, Sosyal Bilimler Enstitüsü %40 ile en fazla tez yayınlanan enstitü olmuştur. Fen Bilimleri Enstitüsü ise %37,5 ile onu takip etmektedir. Lisansüstü Eğitim Enstitüsü %17,5 ile üçüncü sırada yer almaktadır. İşletme Enstitüsü %2,5 ile daha az sayıda tez yayınlamıştır. Uygulamalı ve Doğal Bilimler Enstitüsü ile Enformatik Enstitüsü ise %1,25'er tez ile katkıda bulunmuşlardır.

Bu dağılım, Sosyal Bilimler Enstitüsü ve Fen Bilimleri Enstitüsü'nün, SFS alanında daha fazla araştırma ve tez çalışması yapılmasına ev sahipliği yaptığını göstermektedir. Bu durum, bu enstitülerin daha geniş araştırma imkanlarına sahip olmaları ve ilgili alanlarda uzman akademisyenlerin bulunması ile açıklanabilir. Sosyal Bilimler Enstitüsü'ndeki yüksek tez sayısı, SFS alanının sosyal bilimlerle olan etkileşiminin ve bu alanda yapılan çalışmalara olan ilginin yüksek olduğunu göstermektedir. Fen Bilimleri Enstitüsü'ndeki yüksek tez sayısı ise, bu alandaki teknolojik ve mühendislik odaklı çalışmaların yoğunluğunu işaret etmektedir. Diğer enstitülerdeki daha az sayıda tez, bu alanın geniş bir disiplinler arası yapıya sahip olduğunu ancak belirli enstitülerde daha fazla yoğunlaştığını göstermektedir. Bu durum, SFS alanındaki araştırma faaliyetlerinin daha fazla enstitüde teşvik edilmesi ve desteklenmesi gerektiğini ortaya koymaktadır.

### 4.6.Lisansüstü Tezlerin Anabilim Dallarına Göre Dağılımı

Bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin ana bilim dallarına göre dağılımı incelenmiştir. Şekil 6'da lisansüstü tezlerin anabilim dallarına göre dağılımı grafiği verilmiştir.



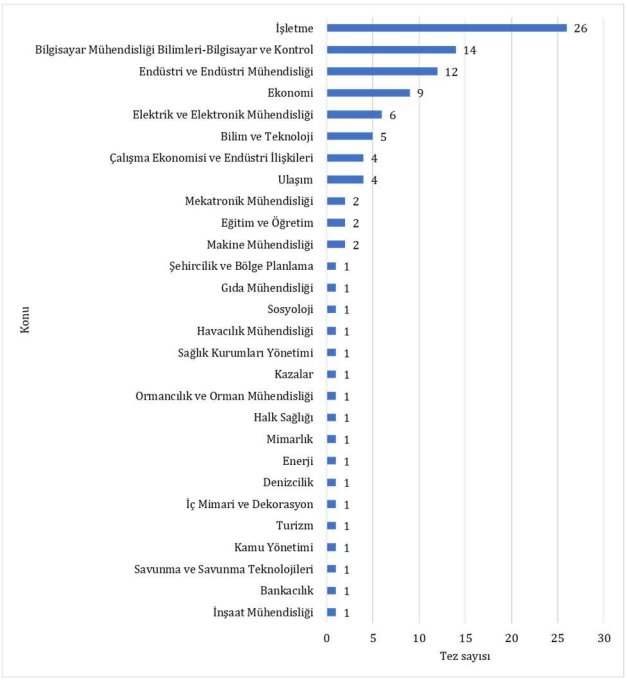

**Şekil 6:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Anabilim Dallarına Göre Dağılımı

Analiz edilen 80 tezin hangi ana bilim dallarında yazıldığını gösteren tabloya göre: İşletme ana bilim dalı, %16,25 ile en fazla tez yayınlanan alandır. Endüstri Mühendisliği %10 ile ikinci sırada yer almaktadır. İktisat %8,75 ve Bilgisayar Mühendisliği %7,5 ile önemli bir yer tutmaktadır. Elektrik ve Elektronik Mühendisliği, İş Sağlığı ve Güvenliği, Çalışma Ekonomisi ve Endüstri İlişkileri ile Uluslararası Ticaret ve Lojistik ana bilim dallarında ise her biri %3,75 oranında tez ile katkıda bulunulmuştur. Makine Mühendisliği ve Girişimcilik ana bilim dallarında her biri %2,5 oranında tez yayınlanmıştır. Diğer ana bilim dallarında ise her biri %1,25 oranında tez ile katkıda bulunulmuştur.

Bu dağılım, İşletme, Endüstri Mühendisliği, İktisat ve Bilgisayar Mühendisliği gibi bölümlerin, SFS konusundaki araştırmalar ve tez çalışmalarında daha fazla yoğunlaştığını ortaya koymaktadır. İşletme ve Endüstri Mühendisliği alanlarında yapılan tez çalışmalarının fazlalığı, bu disiplinlerin SFS ile ilgili operasyonel, yönetsel ve süreç optimizasyonu konularına yoğun ilgi gösterdiğini ortaya koymaktadır. İktisat ve Bilgisayar Mühendisliği alanlarında yapılan çalışmaların fazlalığı ise, bu disiplinlerin teknolojik gelişmelerin ekonomik etkilerini ve teknik altyapılarını incelemeye yönelik araştırmalara ağırlık verdiğini göstermektedir. Diğer ana bilim dallarında daha az sayıda tez bulunması, SFS araştırmalarının bazı disiplinlerde yoğunlaştığını, ancak bu çalışmaların çok çeşitli akademik alanlarla da bağlantılı olduğunu göstermektedir. Bu durum, SFS alanındaki araştırma faaliyetlerinin daha fazla ana bilim dalında teşvik edilmesi ve desteklenmesi gerektiğini ortaya koymaktadır.

### 4.7.Lisansüstü Tezlerin Konularına Göre Dağılımı

Bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin konu alanlarına göre dağılımı incelenmiştir. Şekil 7'de lisansüstü tezlerin konularına göre dağılımı grafiği verilmiştir.




**Şekil 7:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Konularına Göre Dağılımı

Analiz edilen 80 tezin konularına göre dağılımına bakıldığında, %32,5 ile İşletme alanında en fazla çalışma yapılmıştır. Bu durum, işletme disiplininin SFS konularına olan yoğun ilgisini ve bu alandaki süreçlerin yönetimi ve optimizasyonuna yönelik araştırmaların önemini vurgulamaktadır. SFS'nin iş süreçlerine entegrasyonu, verimlilik artırma, maliyet azaltma ve rekabet avantajı sağlama gibi konular, işletme alanında büyük ilgi görmektedir.

Bilgisayar Mühendisliği Bilimleri-Bilgisayar ve Kontrol alanında %17,5 oranında tez yazılmış olup, bu, bilgisayar mühendisliğinin SFS teknolojilerinin gelişimine ve kontrol sistemlerinin iyileştirilmesine katkı sağladığını göstermektedir. SFS genellikle bilgisayar sistemleri ve kontrol mühendisliği ile yakından ilişkilidir, bu nedenle bu alanda yapılan araştırmaların sayısı oldukça yüksektir.

Endüstri ve Endüstri Mühendisliği %15 oranıyla üçüncü sırada yer almaktadır. Bu durum, endüstri mühendisliğinin üretim süreçlerinin iyileştirilmesi, verimlilik artırma ve süreç optimizasyonuna olan ilgisini ortaya koymaktadır. SFS'nin endüstriyel uygulamalarda kullanımı, üretim hatlarının daha verimli ve esnek hale getirilmesini sağlamaktadır.

Ekonomi %11,25 ile önemli bir yer tutmaktadır. SFS'nin ekonomik etkileri, maliyet analizleri ve ekonomik modellemeler gibi konular, ekonomi disiplininde yoğun olarak araştırılmaktadır. İşletme ve ekonomi, SFS'nin iş dünyasına ve makroekonomik etkilerine dair önemli veriler sağlamaktadır.

Elektrik ve Elektronik Mühendisliği %7,5 oranında katkıda bulunmuş olup, bu alandaki teknik gelişmelerin SFS ile entegrasyonunun araştırıldığını göstermektedir. Elektrik ve elektronik mühendisliği, SFS'nin donanım ve yazılım bileşenlerinin geliştirilmesi ve iyileştirilmesinde kritik bir rol oynamaktadır.

Bilim ve Teknoloji %6,25 oranında temsil edilmiştir. Bu disiplinler, SFS'nin sosyal ve teknolojik boyutlarını ele almakta ve bu alandaki yeniliklerin iş gücü dinamikleri üzerindeki etkilerini incelemektedir.

Çalışma Ekonomisi ve Endüstri İlişkileri ile Ulaşım konularında her biri %5 oranında tez yazılmıştır. Bu alanlar, SFS'nin iş gücü dinamikleri üzerindeki etkilerini ve ulaşım sistemlerindeki uygulamalarını araştırmaktadır.

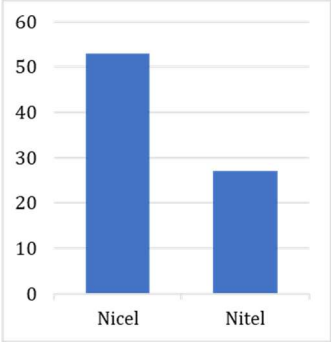
Makine Mühendisliği, Eğitim ve Öğretim ile Mekatronik Mühendisliği konularında her biri %2,5 oranında tez yazılmıştır. Bu, SFS'nin bu mühendislik disiplinlerindeki uygulamalarını ve eğitim süreçlerindeki etkilerini göstermektedir.

Diğer konular ise her biri %1,25 oranında tez ile temsil edilmiştir. Bu, SFS'nin geniş bir yelpazede disiplinler arası bir çalışma alanı olduğunu ve birçok farklı konu ile ilişkili olduğunu göstermektedir. Örneğin, iç mimarlık, halk sağlığı ve savunma teknolojileri gibi alanlar, SFS'nin bu sektörlerdeki özel uygulamalarını ve etkilerini araştırmaktadır.

Bu dağılım, belirli disiplinlerin SFS alanında daha fazla yoğunlaştığını ve bu konuların araştırma faaliyetlerinde öne çıktığını göstermektedir. Bu durum, SFS alanındaki araştırma faaliyetlerinin daha fazla disiplinler arası iş birliği ile genişletilmesi ve desteklenmesi gerektiğini ortaya koymaktadır. Aynı zamanda, SFS'nin farklı sektörlerdeki potansiyelini ve uygulama alanlarını genişletmek için daha fazla araştırma yapılmasının önemini vurgulamaktadır.

### 4.8.Lisansüstü Tezlerin Kullandığı Yöntemlere Göre Dağılımı

Bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin araştırma yöntemlerine göre dağılımı incelenmiştir. Şekil 8'de lisansüstü tezlerin kullandığı yöntemlere göre dağılımı grafiği verilmiştir.




**Şekil 8:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Kullandığı Yöntemlere Göre Dağılımı

Analiz edilen 80 tezin hangi araştırma yöntemleri ile yazıldığını gösteren tabloya göre: Nicel yöntemler 53 tezde kullanılmıştır. Bu, tezlerin %66'sının nicel araştırma yöntemlerini tercih ettiğini göstermektedir. Nicel yöntemler, genellikle veri toplama, istatistiksel analiz ve ölçülebilir sonuçlar elde etme amacıyla kullanılır. SFS alanında yapılan çalışmaların büyük bir kısmı, sistem performansını ölçme, modelleme ve simülasyon yapma gibi konularda yoğunlaştığından, nicel yöntemlerin yaygın olarak tercih edilmesi anlaşılabilir bir durumdur. Nitel yöntemler ise 27 tezde kullanılmıştır. Bu, tezlerin %34'ünün nitel araştırma yöntemlerini tercih ettiğini göstermektedir. Nitel yöntemler, genellikle derinlemesine analiz, teori geliştirme ve karmaşık fenomenlerin anlaşılması amacıyla kullanılır. SFS alanındaki bazı çalışmalar, sosyal etkileri, kullanıcı deneyimlerini ve organizasyonel değişimleri anlamak için nitel yöntemleri tercih etmektedir.

Bu dağılım, SFS alanında yapılan araştırmaların büyük bir kısmının ölçülebilir ve veriye dayalı sonuçlar elde etmeyi hedeflediğini göstermektedir. Nicel yöntemlerin ağırlıklı olarak kullanılması, bu alandaki araştırmaların teknik ve mühendislik odaklı olduğunu ve sistemlerin performansını optimize etmeyi amaçladığını ortaya koymaktadır. Ancak nitel yöntemlerin de önemli bir yer tuttuğu görülmektedir. Bu, SFS'nin sadece teknik boyutlarıyla değil, aynı zamanda sosyal, ekonomik ve organizasyonel boyutlarıyla da ilgilenen araştırmaların yapıldığını göstermektedir. Nitel araştırmalar, kullanıcı deneyimlerinin ve insan-makine etkileşimlerinin derinlemesine incelenmesini sağlayarak, SFS'nin daha geniş bir perspektiften anlaşılmasına katkı sağlamaktadır. Bu sonuçlar, SFS alanında araştırma yaparken hem nicel hem de nitel yöntemlerin dengeli bir şekilde kullanılması gerektiğini ve her iki yöntem türünün de alana değerli katkılar sunduğunu göstermektedir.

Aynı zamanda bu çalışmanın bir parçası olarak Türkiye'de SFS alanında yazılan lisansüstü tezlerin kullandıkları yöntem ayrıntılarına göre dağılımı incelenmiştir. Şekil 9'da lisansüstü tezlerin

kullandığı yöntem ayrıntılarına göre dağılımı grafiği verilmiştir.


**Şekil 9:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerin Kullandığı Yöntem Ayrıntılarına Göre Dağılım

Analiz edilen 80 tezin hangi yöntemlerle yazıldığını gösteren tabloya göre: Nitel araştırma yöntemleri, 22 tez ile en sık kullanılan yöntemdir. Nitel araştırmalar, derinlemesine analiz, teori geliştirme ve karmaşık fenomenlerin anlaşılması amacıyla kullanılır. Bu durum, SFS alanında sosyal etkiler, kullanıcı deneyimleri ve organizasyonel değişimlerin önemini göstermektedir. Anket ve SPSS analizi yöntemi 6 tezde kullanılmıştır. Bu yöntem, veri toplama ve istatistiksel analiz için yaygın olarak tercih edilmektedir. SPSS, araştırmacılara verilerini detaylı bir şekilde analiz etme imkanı sunar. Derin öğrenme yöntemleri 2 tezde yer almaktadır. Derin öğrenme, SFS'nin büyük veri ve yapay zeka uygulamalarında önemli bir rol oynamaktadır. Diğer yöntemler ise her biri 1 tezde kullanılmış olup, geniş bir yelpazede çeşitlilik göstermektedir.

Bu dağılım, SFS alanındaki tezlerin geniş bir metodolojik çeşitliliğe sahip olduğunu göstermektedir. Araştırmacılar, tezlerinde hem nitel hem de nicel yöntemleri kullanarak, farklı perspektiflerden kapsamlı analizler yapmaktadırlar. Nicel yöntemlerin yaygın kullanımı, SFS'nin teknik ve mühendislik odaklı doğasını yansıtırken, nitel yöntemlerin de önemli bir yer tutması, bu alandaki sosyal ve organizasyonel boyutların da derinlemesine incelendiğini ortaya koymaktadır.

### 4.10.Lisansüstü Tezlerin Anahtar Kelimelere Göre Dağılımı

Anahtar kelimeler, akademik makalelerde kullanılan temel kavramlar ve konular hakkında bilgi sağlar. Bu analizde, SFS lisansüstü tezlerinde en sık kullanılan anahtar kelimeler ve bu kelimelerin yoğunluğu incelenmiştir. Anahtar kelime analizi, SFS alanında hangi konuların daha fazla ilgi gördüğünü ve araştırmaların hangi temalar etrafında yoğunlaştığını belirlemek amacıyla yapılmıştır. Şekil 10'da SFS alanında yazılan lisansüstü tezlerin anahtar kelimelerinden oluşan kelime bulutu verilmiştir.


**Şekil 10:** 2015-2024 Yılları Arasında Siber Fiziksel Sistemler Alanında Tamamlanan Lisansüstü Tezlerde En Sık Geçen Anahtar Kelimeler İle Oluşturulan Kelime Bulutu

## 5. SONUÇLAR

Bu çalışmada, Türkiye'de SFS üzerine yapılan 80 lisansüstü tez bibliyometrik analiz yöntemleri kullanılarak incelenmiştir. Elde edilen bulgular, bu alandaki araştırma faaliyetlerinin genel durumunu, eğilimlerini ve gelecekteki araştırma yönelimlerini anlamaya yardımcı olmuştur.

Singh vd. (2024), siber-fiziksel sistemlerin (SFS) üretim ve tedarik zinciri yönetimindeki küresel eğilimlerini analiz ederek, Endüstri 4.0, IoT ve bulut bilişim gibi teknolojilerin SFS'in gelişimindeki kritik rolünü ortaya koymuşlardır. Türkiye özelinde değerlendirildiğinde, bu teknolojilerin üretim sistemlerine entegrasyonu öncelikli hedeflerden biri olmasına rağmen, uygulama sürecinde karşılaşılan zorluklar ve entegrasyonun olgunluk düzeyi gelişmiş sanayi ülkelerine kıyasla daha geride kalmaktadır. Türkiye'de SFS uygulamalarının genellikle büyük

ölçekli ve devlet destekli projelerde yoğunlaşması, endüstriyel altyapının ve yerel teknoloji ekosisteminin sınırlı olması nedeniyle teknolojinin yaygın adaptasyonunu yavaşlatmaktadır. Bu bağlamda, Singh vd. (2024), tarafından küresel eğilimlere dair sunulan bulgular, Türkiye için potansiyel bir yol gösterici niteliğindedir. Çalışmada belirtilen IoT ve Endüstri 4.0 entegrasyon modelleri, Türkiye'nin dijital dönüşüm çabalarına örnek teşkil edebilir ve Türkiye'deki araştırmacılara daha etkin stratejiler geliştirme konusunda ışık tutabilir. Bu çalışmanın farkı, Türkiye özelinde gerçekleştirilen lisansüstü tezleri inceleyerek, Türkiye'nin SFS alanındaki mevcut akademik eğilimlerini ortaya koymasıdır. Yaacoub vd. (2020), SFS güvenliğine küresel bağlamda yaklaşarak güvenlik açıklarını ve çözüm yöntemlerini kapsamlı bir şekilde analiz etmişlerdir. Ancak Türkiye'deki SFS uygulamalarında güvenlik, genellikle temel önlemlerle sınırlı kalmış ve daha ileri düzey güvenlik stratejileri geliştirilmemiştir. Bu çalışma, Türkiye'de SFS güvenliği konusundaki akademik çalışmaların mevcut düzeyini ortaya koyarken, güvenliğin özellikle enerji ve sanayi sektörlerinde gelişim gerektirdiğini vurgulamaktadır. Ayrıca, Yıldız ve Gejam (2022) tarafından gerçekleştirilen bibliyometrik analiz çalışması, SFS ve siber güvenlik alanındaki uluslararası iş birliği düzeylerinin ve araştırma sayılarının Türkiye'deki mevcut durumdan daha ileri olduğunu ortaya koymuştur.

Bu çalışma ise Türkiye'deki SFS alanında yapılan tezleri inceleyerek, yerel eğilimler ve ihtiyaçlar doğrultusunda gelecekteki araştırmalara yönelik öneriler sunmaktadır. Türkiye'nin uluslararası iş birliğini güçlendirmesi, SFS ve siber güvenlik üzerine daha fazla akademik yayın yapması gerektiği bu çalışmanın bir diğer önemli sonucudur.

Türkiye'deki SFS araştırmalarının daha da gelişmesi için bu çalışma, diğer araştırmalardan farklı olarak, disiplinler arası iş birliğinin teşvik edilmesi, SFS güvenliği ve etik boyutlarının özellikle kritik altyapılar ve hassas alanlarda araştırılmasına dikkat çekmektedir. Bu yönüyle çalışma, Türkiye'nin SFS alanında uluslararası düzeydeki bilimsel eğilimleri yakalamasına ve kendi stratejik gereksinimlerine uygun çözümler üretmesine yönelik bir yol gösterici niteliğindedir.

Analiz sonuçları, Türkiye'de SFS alanına olan ilginin son yıllarda arttığını göstermektedir. Özellikle 2015 yılından itibaren makale sayısında belirgin bir artış gözlenmiş, bu da SFS'nin akademik ve endüstriyel alanlarda artan önemini yansıtmaktadır Analiz sonuçları, Türkiye'de SFS alanına olan ilginin son yıllarda arttığını göstermektedir. Özellikle 2015 yılından itibaren tez sayısında belirgin bir artış gözlenmiş, bu da SFS'nin akademik ve endüstriyel alanlarda artan önemini yansıtmaktadır. En fazla tez 2019 yılında yayınlanmış olup, bu dönemdeki teknolojik ilerlemeler ve Endüstri 4.0'ın yaygınlaşması bu artışa katkıda bulunmuştur.

Tez türlerine göre dağılımda, yüksek lisans tezlerinin %85 oranında daha fazla olduğu, doktora tezlerinin ise %15 oranında olduğu görülmüştür. Bu durum, lisansüstü düzeyde SFS alanına yönelik yoğun bir ilgi olduğunu göstermektedir. Üniversite türlerine göre analizde, tezlerin %95'inin devlet üniversitelerinde, %5'inin ise vakıf üniversitelerinde yapıldığı belirlenmiştir. İstanbul Teknik Üniversitesi, en fazla tez yayınlanan üniversite olarak öne çıkmaktadır. Enstitülere göre dağılımda, Sosyal Bilimler Enstitüsü %40 ile en fazla tez yayınlayan enstitü olmuştur. Bu, sosyal bilimlerin SFS alanındaki önemini ve bu alanda yapılan çalışmaların çeşitliliğini göstermektedir. Ana bilim dallarına göre analizde, İşletme (%21.25), Bilgisayar Mühendisliği (%17.5), ve Endüstri Mühendisliği (%12.5) gibi disiplinler öne çıkmaktadır. Bu, SFS'nin işletme süreçleri, bilgisayar teknolojileri ve endüstriyel uygulamalar üzerindeki etkisini vurgulamaktadır. Tezlerde kullanılan yöntemler incelendiğinde, %66'sının nicel, %34'ünün ise nitel yöntemler kullandığı görülmüştür. Bu durum, SFS araştırmalarının büyük ölçüde veriye dayalı ve ölçülebilir sonuçlar elde etmeyi amaçladığını ortaya koymaktadır. Anahtar kelime analizleri, SFS konusundaki çalışmaların genellikle "endüstri 4.0", "akıllı sistemler" ve "siber güvenlik" gibi temalar etrafında yoğunlaştığını ortaya koymuştur. Bu, SFS'nin çok disiplinli doğasını ve çeşitli uygulama alanlarını yansıtmaktadır. Elde edilen bulgular ışığında, Türkiye'de SFS alanındaki araştırmaların daha da gelişmesi için bazı önerilerde bulunulabilir:

SFS alanında daha geniş bir disiplinler arası iş birliği teşvik edilmeli, mühendislik, bilgisayar bilimleri, sosyal bilimler ve ekonomi gibi farklı alanlar bir araya getirilerek yenilikçi çözümler üretilmelidir.

SFS'lerin güvenliği ve etik boyutları, özellikle kritik altyapılar ve sağlık gibi hassas alanlarda daha fazla araştırılmalıdır. SFS teknolojilerinin pratik uygulamaları üzerine odaklanılmalı ve endüstriyel iş birlikleri artırılmalıdır. Bu, akademik bilginin pratiğe aktarılmasını hızlandırabilir ve teknolojik gelişmeleri destekleyebilir. Türkiye'deki araştırmacılar, uluslararası projelere katılımı artırmalı ve küresel ağlarda daha aktif rol almalıdır. Bu, bilgi ve teknoloji transferini kolaylaştıracak ve yerel araştırma kapasitesini güçlendirecektir.

Türkiye'de SFS alanında yapılan çalışmalar, bu alandaki akademik literatüre önemli katkılarda bulunmuş ve çeşitli uygulama alanlarında yenilikçi çözümler sunmuştur. Ancak, bu alanda daha fazla araştırma yapılması ve mevcut bilgi birikiminin genişletilmesi gerekmektedir. Bu çalışmanın bulguları, SFS araştırmalarının gelecekteki yönelimlerini belirlemek için önemli bir kaynak oluşturmakta ve bu alanda yapılacak çalışmalar için yol gösterici bilgiler sunmaktadır.

**KAYNAKÇA**


Aksungur, B. N, Sever, H., Güven, E., ve Eren T. (2024). İnsansız Hava Araçları Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Türkiye İnsansız Hava Araçları Dergisi, 6(1), 21-29.

Baheti, R., ve Gill, H. (2011). Cyber-physical systems. In T. Samad ve A. M. Annaswamy (Eds.), The Impact of Control Technology (pp. 161-166). IEEE Control Systems Society.

Bhadani, U. (2024). Smart Grids: A Cyber–Physical Systems Perspective. International Research Journal of Engineering and Technology (IRJET), 11(06), 801.

Canonico, R., ve Sperlì, G. (2023). Industrial cyber-physical systems protection: A methodological review. Computers ve Security, 135, 103531.

Çakmak, İ., ve Öztürk, A. (2019). Türkiye'de Mühendislik Alanındaki Bilimsel Yayınların Bibliyometrik Analizi. *Bilim ve Teknoloji Dergisi*, 25(3), 451-467.

Demir, K., ve Yılmaz, E. (2020). Eğitim Bilimleri Alanında Yapılan Tezlerin Bibliyometrik Analizi. *Eğitim ve Bilim Dergisi*, 45(2), 123-138.

Demirci, M. (2019). Siber fiziksel sistemler üzerine Türkiye'deki akademik çalışmalar: Bir bibliyometrik analiz (Yüksek lisans tezi). Yüksek Öğretim Kurulu Ulusal Tez Merkezi.

Duran, Z. (2024). Endüstri 5.0'a Geçişte Siber Güvenlik: Yeni Sanayileşen Ülkeler Üzerine Bir İnceleme. Eskişehir Osmangazi Üniversitesi Sosyal Bilimler Dergisi, 25(3), 745-760.

Elmas, B., Çelik, M., ve Korkmaz, E. (2024). Dijital Finans Araştırmalarının Bilim Haritalama Teknikleri ile Bibliyometrik Analizi. Muhasebe ve Finansman Dergisi, (103), 113-134.

Gaferoğlu, İ., Kaya, S., Kalemler, Y. B., Güven, E. ve Eren, T. (2024). Tsunami Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Urban 21 Journal, 2(2), 153-164.

Gündüz, H., ve Eren, T. (2024). Kent Dirençliliği Konusunda Yapılan Lisansüstü Tezlerin Bibliyometrik Analizi. Resilience, 8(1), 73-82. https://doi.org/10.32569/resilience.1440796

Güven, E. ve Eren, T. (2024). Endüstriyel Kaza Araştırmalarına Yönelik Bibliyometrik İnceleme: Tezler Üzerine Bir Çalışma, kisgd, vol. 8, no. 2, pp. 85–93, https://doi.org/10.33720/kisgd.1426403

Harkat, H., Camarinha-Matos, L. M., Goes, J., ve Ahmed, H. F. (2024). Cyber-physical systems security: A systematic review. Computers ve Industrial Engineering, 109891.

Karademir, A., ve Akın, H. (2021). Sosyal Bilimlerde Bibliyometrik Analiz: Araştırma Eğilimleri ve Gelecek Perspektifler. *Sosyal Bilimler Araştırma Dergisi*, 18(2), 231-249.

Karnouskos, S. (2011). Cyber-Physical Systems in the SmartGrid. 19th International Conference on Software, Telecommunications and Computer Networks (SoftCOM 2011), Split, Croatia, 1-6. https://doi.org/10.1109/softcom.2011.6069429

Küpçüoğlu, E., Alakaş, H. M., ve Eren, T. (2024). Blok Zincir Üzerine Yazılan Yüksek Lisans ve Doktora Tezlerinin Bibliyometrik Analizi. Bilişim Teknolojileri Dergisi, 17(4), 281-293. https://doi.org/10.17671/gazibtd.1453335

Lee, E. A. (2015). The past, present and future of cyber-physical systems: A focus on models. Sensors, 15(3), 4837-4869. https://doi.org/10.3390/s150304837

Lou, S., Hu, Z., Zhang, Y., Feng, Y., Zhou, M., ve Lv, C. (2024). Human-cyber-physical system for Industry 5.0: A review from a human-centric perspective. IEEE Trans. Autom. Sci. Eng, 1-18.

Ozsaatci, F. G. B. (2022). Sosyal medya pazarlaması alanındaki yayınların bibliyometrik analizi. İşletme Araştırmaları Dergisi, 14(4), 3177-3192.

Özkan, M., ve Erten, E. (2021). Üniversitelerin Akademik Performanslarının Bibliyometrik Analizi: Stratejik Araştırma Planlamasına Katkılar. *Yükseköğretim ve Bilim Dergisi*, 31(1), 89-105.

Öztürk, H. (2020). Türkiye'de siber fiziksel sistemler üzerine yapılan akademik çalışmalar: Bibliyometrik bir değerlendirme (Yüksek lisans tezi). Yüksek Öğretim Kurulu Ulusal Tez Merkezi.

Öztürk, N., ve Kurutkan, M. N. (2020). Kalite yönetiminin bibliyometrik analiz yöntemi ile incelenmesi. *Journal of Innovative Healthcare Practices*, *1*(1), 1-13.

Öztürk, S., Taş, B. S., Kaplan, D., Keskin, S., et al. (2024). Sağlık Turizmi Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Sağlıkta Performans Ve Kalite Dergisi, 21(3), 184-204.

Pınarcı, E. Ş., Vuruşkan, C. T., Güven, E., Eren, T. (2024). Türkiye'de Ekip Çizelgeleme Konulu Lisansüstü Tezlerin Bibliyometrik Analizi. Harran Üniversitesi Mühendislik Dergisi, 9(2), 118-130. https://doi.org/10.46578/humder.1509219

Rajkumar, R., Lee, I., Sha, L., ve Stankovic, J. (2010). Cyber-Physical Systems: The Next Computing Revolution. Design Automation Conference, Anaheim, CA, 731-736. https://doi.org/10.1145/1837274.1837461

Sanlı, Y. B., Baltacı, F., Güven, E., ve Eren, T. (2024). Siber Güvenlik Çalışmaları Üzerine Bibliyometrik Analiz. *Bilişim Teknolojileri Dergisi*, *17*(3), 223-229.

Savrun, B., ve Mutlu, H. M. (2019). Kent Lojistiği Üzerine Bibliyometrik Analiz. Kent Akademisi, 12(2), 364-386.

Singh, N., Panigrahi, P. K., Zhang, Z., ve Jasimuddin, S. M. (2024). Cyber-physical systems: a bibliometric analysis of literature. Journal of Intelligent Manufacturing, 1-37.

Şahin, M. D. Sürdürülebilirlik Değeri Kavramına Yönelik Bibliyometrik Analiz. Dumlupınar

Üniversitesi Sosyal Bilimler Dergisi, (81), 321-338.

Şahin, S., ve Uğurlu, B. (2018). Türkiye'deki Sağlık Bilimleri Alanındaki Araştırma İş Birliklerinin Bibliyometrik Analizi. *Sağlık Bilimleri Dergisi*, 22(4), 321-337.

Tekin, M., Öztürk, D., ve Bahar, İ. (2021). Tersine lojistiğin bibliyometrik analizi. Aksaray Üniversitesi İktisadi ve İdari Bilimler Fakültesi Dergisi, 13(3), 87-100.

Turan, F., ve Güner, M. (2022). Bibliyometrik Analiz Yöntemi ile Üniversitelerin Akademik Performans Değerlendirmesi. *Akademik Araştırmalar ve Çalışmalar Dergisi*, 34(1), 56-72.

Uzan, H. K. Akıllı Şehir ve Yönetişim Temalı Makaleler Üzerine Bibliyometrik Analiz. Yönetim Bilimleri Dergisi, 22(53), 938-960.

Yaacoub, J. P. A., Salman, O., Noura, H. N., Kaaniche, N., Chehab, A., ve Malli, M. (2020). Cyber-physical systems security: Limitations, issues and future trends. Microprocessors and microsystems, 77, 103201.

Yıldız, B., ve Gejam, E. H. Y. (2022). Cyber-Physical Systems and Cyber Security: A Bibliometric Analysis. OPUS Journal of Society Research, 19(45), 35-49.

Yu, Z., Gao, H., Cong, X., Wu, N., ve Song, H. H. (2023). A Survey on Cyber–Physical Systems Security. IEEE Internet of Things Journal, 10(24), 21670-21686.

Zeren, D., ve Kaya, N. (2020). Dijital pazarlama: Ulusal yazının bibliyometrik analizi. Çağ Üniversitesi Sosyal Bilimler Dergisi, 17(1), 35-52.

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

# Comparative Analysis of Machine Learning Algorithms in Stock Price Prediction

**Hakan Murat Karaca[1], Umut Dökmen[1*]**

*[1]Manisa Celal Bayar Üniversitesi, Bilgisayar Mühendisliği, Manisa, Türkiye*

**Keywords:**
Machine Learning
Algorithms
Stock
Regression
Supervised Learning

**ABSTRACT**

Stock is part of a company's principal. A person who buys stock of a company shares the profit or loss of this company. Large volume transactions are made on stock exchanges where stocks are traded. Stock prices are difficult to predict because they are affected by many variables, but when they can be predicted, great benefits are provided. Prediction of stock prices is possible with today's computers using machine learning algorithms. Machine learning provides more successful results than fundamental and technical analysis in stock price prediction. In our study, daily closing price predictions were made by collecting approximately 5-years data of the top 5 stocks with the highest market value traded in BIST 100 between 2016 and 2020. Multiple linear regression, Bayesian regression, random forest, decision trees, support vector machines, artificial neural networks algorithms were applied to include maximum 22 features and the results were compared. The most successful result was obtained in the artificial neural networks algorithm. To achieve the highest success, data pre-processing, normalization, cross-validation, parameter optimization and feature selection were applied. It has been observed that using these methods together increases the success.

# Hisse Senedi Fiyat Tahmininde Makine Öğrenimi Algoritmalarının Karşılaştırmalı Analizi

**Anahtar Kelimeler:**
Makine Öğrenmesi
Algoritmalar
Hisse Senedi
Regresyon
Gözetimli Öğrenme

**ÖZ**

Hisse senedi bir şirketin anaparasının bir parçasıdır. Bir şirketin hisselerini satın alan kişi, bu şirketin kar veya zararına ortak olur. Hisse senetlerinin işlem gördüğü borsalarda büyük hacimli işlemler yapılmaktadır. Hisse senedi fiyatları birçok değişkenden etkilendiğinden tahmin edilmesi zordur ancak tahmin edilebildiğinde büyük faydalar sağlanır. Hisse senedi fiyatlarının tahmini, makine öğrenmesi algoritmalarını kullanan günümüz bilgisayarları ile mümkün olmaktadır. Makine öğrenimi, hisse senedi fiyat tahmininde temel ve teknik analize göre daha başarılı sonuçlar sağlamaktadır. Çalışmamızda 2016-2020 yılları arasında BIST 100'de işlem gören en yüksek piyasa değerine sahip 5 hisse senedinin yaklaşık 5 yıllık verileri toplanarak günlük kapanış fiyatı tahminleri yapılmıştır. Çoklu doğrusal regresyon, bayesian regresyon, rastgele orman, karar ağaçları, destek vektör makineleri, yapay sinir ağları maksimum 22 özelliği dahil edecek şekilde uygulanmış ve sonuçlar karşılaştırılmıştır. En başarılı sonuç yapay sinir ağları algoritmasında elde edilmiştir. En yüksek başarıyı elde etmek için veri ön işleme, normalleştirme, çapraz doğrulama, parametre optimizasyonu ve özellik seçimi uygulanmıştır. Bu yöntemlerin bir arada kullanılmasının başarıyı arttırdığı gözlemlenmiştir.

*Sorumlu Yazar

(hakanmkaraca@gmail.com) ORCID ID 0000-0003-2144-2994
*(umut.dokmen@cbu.edu.tr) ORCID ID 0000-0001-6919-4278

**e-ISSN: 2717-8579**

**Geliş Tarihi: 19/12/2023; Kabul Tarihi: 16/12/2024**

## 1. INTRODUCTION

A stock is a part of the principal of a company. People who buy a company's stock share in the profit and loss of that company. The stock signifies the special relationship between the company and the person who buys the stock (Summers, 2007). Companies open their stocks to investors through the stock market in order to increase their financial capacity and their capital. The expectation that the value of the stock will increase creates demand for that stock. This demand increases the value of the stock. On the contrary, the expectation that the value of the stock will decrease requires selling the stock and the price will decrease. Investors aim to make a profit by buying stocks that will rise in the future. For this reason, it is very important for investors to be able to predict the stock price.

Machine learning is widely used in the field of finance, as it is in many fields. Many companies use machine learning in stock trading. It is able to make very wise investment decisions and reduce financial risks for people. Many studies have shown that machine learning-based applications are more successful than traditional stock trading strategies. These results increase the applications of artificial intelligence and machine learning in the field of finance day by day (URL-1, URL-2).

Stock prices are volatile. There are many internal and external factors that affect the stock price. Internal factors, profit distribution policy, capital increase, financial structure, management, field of activity of the enterprise (Hürer, 1995). In this study, Opening Price, High Price, Low Price, Volume, Net Profit for the Period, Resource and Dividend Income Factors, which are internal variables affecting the stock price, are included in the calculations. External factors included in the calculation in this study: BIST Stars Tradded Value, BIST Stars Tradded Volume, BIST 100 Index, BIST 100 Volume (TL), BIST 100 Difference, Dollar-TL, Euro-TL, XAU-USD, Brent Oil, S&P 500 Index, Euro Stoxx 50 Index, Interest and Inflation. These factors affect the stock price differently. The effect of these factors on the stock price can be calculated by statistical methods. But there are other factors that affect the stock price. These are the political situation in the country and the world, financial expectations, sectoral expectations, unexpected events. The political situation in the country and the world or natural disasters cannot be predicted by numerical methods. However, since the policies of the country and the world will affect the external factors used in this study, the indirect effect on the stock can be calculated.

In the study of Ghana, Awaisa and Muzammula, they tried to predict the stock prices of Apple, Amazon and Google with time series forecasting algorithms and observed that the exponential smoothing results gave greater accuracy (Ghani, 2019; Muzammul 2019). Sarode, Tolani, Kak and Lifna used real-time data along with news analysis. With LSTM (Long Short-Term Memory), an artificial neural network architecture, they presented a system that decides whether to buy the stocks of different companies (Sarode, 2019; Tolani, 2019; Kak 2019; Lifna 2019). In their study, Usmani, Adil, Raza and Ali tried to predict the end-of-day closing performance of the Karachi Stock Exchange (KSE) using machine learning. In the study, it was found that the multi-layer perceptron showed the best performance and the feature that affected the index the most was the oil price (Usmani, 2019; Adil, 2019; Raza, 2019; Ali, 2019). Tipirisetty made stock price prediction using deep learning. In addition to quantitative analysis, his study also analyzed textual public news from online news sources and concluded that "accuracy increases when textual information is used in stock price prediction" (Tipirisetty, 2018). For stock price prediction, Sinngh collected 10 years of data from yahoo finance and used LSTM and linear regression for prediction. RMSE is used as the evaluation metric. RMSE was found 2.04 for linear regression and 0.43 for LSTM (Singh, 2021). In his study, Guo tried to predict the S&P 500 index. LSTM, arima and garch models are used and found that the model in which these 3 models were used together gave more successful results than the model in which LSTM was used alone (Guo, 2022).

In this study, the daily closing price of the stocks of the top 5 companies as seen table1 with the highest market value in BIST(Borsa Istanbul) 100 is predicted. For the training of machine learning models, approximately 5 years of historical data between 2016 and 2020 were collected and combined from borsaistanbul.com, investing.com and isyatırım.com. In the study, 80% of the data set was used for training and 20% for testing. In order for machine learning algorithms to give more successful results, data preprocessing has been implemented. The effect of the normalization methods used in the data set on the model success was investigated. The parameter changes in the models were examined and their effects on the results were investigated. Optimum parameters are selected to get the highest success. By using feature selection methods, the success of the model has been increased. The success of machine learning algorithms used in the study has been compared.

One of the aims of the study is to predict the stock price and give direction to the investors. Another purpose is to analyze the factors affecting stock prices with machine learning methods and to give an idea to financial analysts. Another purpose is to optimize prediction success by using internal and external factors that affect the price of stocks together as features.

In similar studies in literature, values close to the closing price such as opening, high, low or global features such as gold, dollar, euro, interest, inflation etc. were included in the model. Differently in this

study, the performance of the machine learning algorithms aimed to increase by including the period net profit, dividend income, resource features obtained from the company's internal balance sheet.

## 2. MACHINE LEARNING

Machine learning is an artificial intelligence field that aims to give the machine the ability to learn without programming it directly. There are broadly three types of machine learning: supervised learning, unsupervised learning, and reinforcement learning. In stock price prediction, the supervised learning technique, which covers all prediction problems, is used because the future price is predicted from the past, known data set. Since the prices which is the output, we get in the stock price prediction is numerical, the task is called prediction. To predict stock prices, the computer learns patterns from past stock prices. The difference between the predicted price and the actual price is called the loss function. The machine improves its performance a little more with each experience. In practice, experience means training data. Therefore, we cannot easily distinguish between machine learning and statistical approaches. The goal of supervised learning is minimizing the loss function. In the stock price prediction machine tries to minimize the difference between the actual stock price with predicted stock price. In supervised learning the machine learns a predictive model that maps the features of the data to an output. Machine aims to learn a model predicting parameters (Molnar, 2019; Goodman, 2019; Kaminsky, 2019; Lessler, 2019).

### 2.1 Machine Learning Algorithms Using in the Project

Multiple linear regression is a linear regression with multiple independent variables. The equation form is also similar to simple linear regression. Both types of regression are ultimately linear.

$$y_i = \beta_i + \beta_1 x_{1i} + \beta_2 x_{2i} + \cdots + \beta_n x_{ni} + e_i \quad (1)$$

The dependent variable y in this study is the “stock price” we aim to find. The independent variables, represented by $x$ in formula(1) are the features in the model for “qnbfb model” such as Opening, High, Low, Difference, Star Market Transaction Volume, Star Market Transaction Amount, Bist 100 Index, Bist100 Volume, Bist100 Difference %, Dollar-TL, XAU-USD, Euro- TL, SP 500, Brent Oil, Euro Stoxx 50, Interest %, Inflation %, Net Profit for period, Resources, Dividend income.

Support vector regression (SVR) is a supervised machine learning method used in prediction problems. Regression analysis is performed to analyze the relationship between a dependent variable and one or more independent variables. SVR formulates an optimization problem by learning a regression function to map input prediction variables to observed output values. SVR is another version of support vector machines which is classification algorithm. However SVM produces a class label i.e. a binary output. SVR is the solution to the regression problem consisting of a real-valued function prediction. The aim in SVR is to find the optimal width hyperplane containing the most appropriate line, that is, the maximum data point as seen figure 1. SVR does not try to minimize the difference between the actual value and the predicted value as in other regression models. It tries to best fit the data within a certain threshold value. The distance between the boundary line and the hyperplane is called the threshold value (Zhang and Donnel 2020).

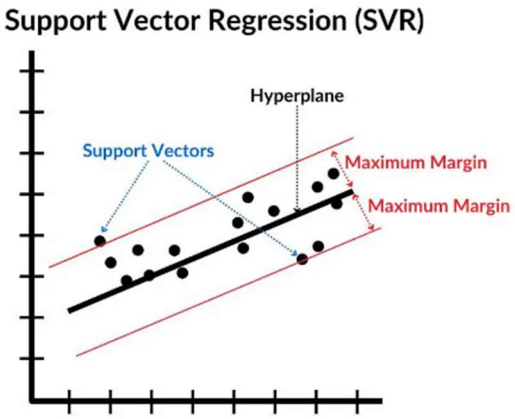



**Figure 1.** Support Vector Regression

Linear regression and logistic regression models cannot be successful when the relationship between the features and the dependent variable is not linear or when the features interact with each other. In such cases, tree-based models can be used. Tree models split data multiple times according to certain cutoff values in the features. Many subsets are created by splits from all data. The final, terminal subsets form leaves, and the other inner subsets form nodes. The average of the training data from this subdivided subset is used to estimate the outcome at each leaf node (Skinea. 2017).

The random forest algorithm is based on drawing more than one decision tree for the same dataset and using these decision trees together. Random forest algorithm can be used for classification and regression. While getting the regression result, the average of more than one separated decision tree is taken (Liu., Wang, Zhang 2012).

Bayesian regression is a type of linear regression based on Bayes’ theory. In the Bayesian approach, the uncertainty in the w vector is characterized by a probability distribution p(w). Bayes’ theorem applies this distribution through

observations of data points and the probability function of the data.

$$P(\beta|y, X) = \frac{P(y|\beta,X)*P(\beta|X)}{P(y|X)} \quad (2)$$

P (β|y, X) is the posterior probability distribution of the parameters when the inputs and outputs are known in formula 2. This is found by multiplying the data probability P(y|β, X) by the prior probability of the parameters and dividing by a normalization constant (URL-10).

A neural network is an oriented structure that connects a simple input layer called a neuron to the output layer with weighted connections to larger structures. A neuron is connected with n input channels, each expressed in synaptic weight $w_i$. Each input from the neuron is multiplied by its weights and they are summed. An optional bias might be added to this sum. The summed result is then put into an activation(threshold) function. This function can be sigmoid, hyperbolic tangent, ReLU or any other function. The input produces an output after filtering it with the activation function (Bonaccorso, 2017).

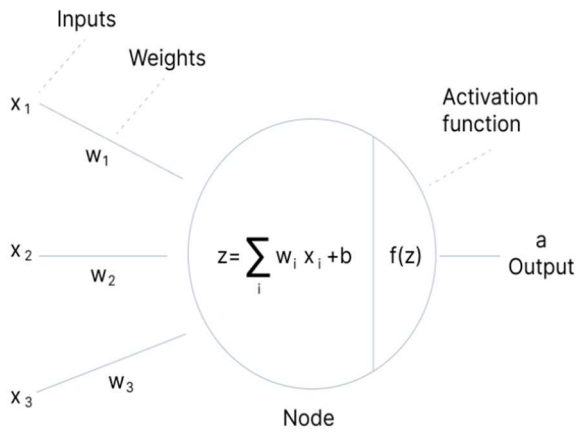



**Figure 2** The structure of a simple neural network.

### 2.2. Data Set

Data has been collected from Borsa Istanbul formal website and investing.com. The daily index, opening, high, low, volume, difference, exchange rates, XAU-USD (golden ounce price), brent oil, S&P 500, data of each stock are taken from the 'investing.com' website. Market transaction volume, Bist 100 transaction amount, Bist100 index, Bist 100 volume, Bist 100 difference data are obtained from borsaistanbul.com. Net profit for the period, resource, dividend income are collected from www.isyatirim.com.tr [URL-3]. Data from different sources are combined in a single table. Since financial data is not open on holidays, only working days are added to the data set. While adding internationally valid features such as brent oil, S&P 500, lines corresponding to holidays in Turkey were not included in the data set.

**Table 1**. Names of stocks examined

| Stock Code | Company Name |
|---|---|
| QNBFB | Qnb Finansbank |
| ENKAI | Enka İnşaat Ve Sanayi A.Ş |
| FROTO | Ford Otomotiv Sanayi A.Ş |
| EREGL | Ereğli Demir Ve Çelik Fabrikalari T.A.Ş |
| KCHOL | Koç Holding A.Ş |

**Table 2.** Explanation of features

| Feature | Explanation |
|---|---|
| Date | Indicates the date on which the relevant feature was obtained. |
| Close price | Indicates the closing value of the stock on the specified date(Aslan, 2020). |
| Opening price | Indicates the opening price of the relevant stock (Aslan 2020). |
| High | Refers to the highest value of the related feature (column) during the day (Aslan, 2020). |
| Low | Refers to the lowest value of the related attribute (column) during the day(Aslan, 2020). |
| Volume | Indicates the trading volume of the relevant stock during the day. |
| Difference % | Indicates the change in the day-to-day price of the relevant stock as a percentage (Aslan, 2020). |
| BIST Stars Traded Value (TL) | Number of transactions in the star market. |
| BIST Stars Traded Volume | It is the trading volume of the market in which the shares with a market value of 300 million TL and above of the portion offered to the public in the first listing to the stock exchange are traded (URL-4). |
| BIST 100 index | It consists of the 100 stocks traded in Borsa Istanbul with the highest market value and trading volume and is the main index of the Equity Market(URL-5). |

| | |
|---|---|
| BIST 100 volume (TL) | It is the total value of daily trading transactions in the BIST 100 (URL-6). |
| BIST 100 difference | It is the change of the BIST100 Index Value information announced by Borsa Istanbul at the end of the trading day according to the value of the next day (Karagöz, 2020). |
| Dollar-TL | TL equivalent of the dollar on the relevant date. |
| Euro-TL | TL equivalent of the euro on the relevant date. |
| XAU-USD | The dollar price of an ounce of gold on the relevant date. |
| Brent Oil | Dollar price of brent oil on the relevant date. |
| S&P 500 | Stock market index of 500 major US stocks by Standard and Poor's(URL-7). |
| Euro Stoxx 50 | Stock index of 50 stocks from 11 Eurozone countries designed by Stoxx (URL-8). |
| Interest | The policy interest rate used by the Turkey Central Bank is the interest rate applied in one-week repo transactions. Decisions on policy rates are taken by the Monetary Policy Committee (MPC) (URL-9). |
| Net profit for the period | It indicates the net profit of the company in that period. |

### 2.3 Data Preprocessing

There are many factors that affect the success of machine learning algorithms. The most important of these is the representation and quality of the data set. Data must be preprocessed to improve quality. Machine learning suffers when there is too much irrelevant and redundant data. In machine learning studies, a significant amount of time is spent in data preprocessing. Data preprocessing is unavoidable as it is impossible to have a preprocessing algorithm that works on all datasets, providing reliable and effective performance. In data preprocessing, operations such as data cleaning, normalization, conversion, feature selection are performed (Kotsiantis, Kanellopoulos, Pintelas 2006; Alexandropoulos, Kotsiantis, Vrahatis 2016)

The normalization technique is used to transform the features to the same scale. In this way, it reduces the difference between the predicted value and the real value. Different feature normalization methods can be used when the actual distribution of features is not known beforehand.

Min max normalization is one of the most used methods to standardize data. For each component, it converts the element's base estimate to zero, the extreme value to 1, and the other values to a decimal between 0 and 1.

$$\hat{X}_i = \frac{x_i - x_{min}}{x_{max} - x_{min}} \quad (3)$$

Where $x_{min}$: minimum value in X feature $x_{max}$: maximum value in X feature $\hat{X}_i$: scaled X

(Raju, Lakshmi, Jain, Kalidindi, Padma 2020)

The standard scaler(ss) is a method in which the distribution approaches normal by averaging each feature and scaling its variance to 1. In the formula, the mean is subtracted from the true value and divided by the variance.

$$\hat{X}_i = \frac{x_i - \bar{X}}{\sigma} \quad (4)$$

Where $\hat{X}_i$ : normalization version of x

σ: standard derivation

$\bar{X}$: mean

$x_i$: each observation from a sample

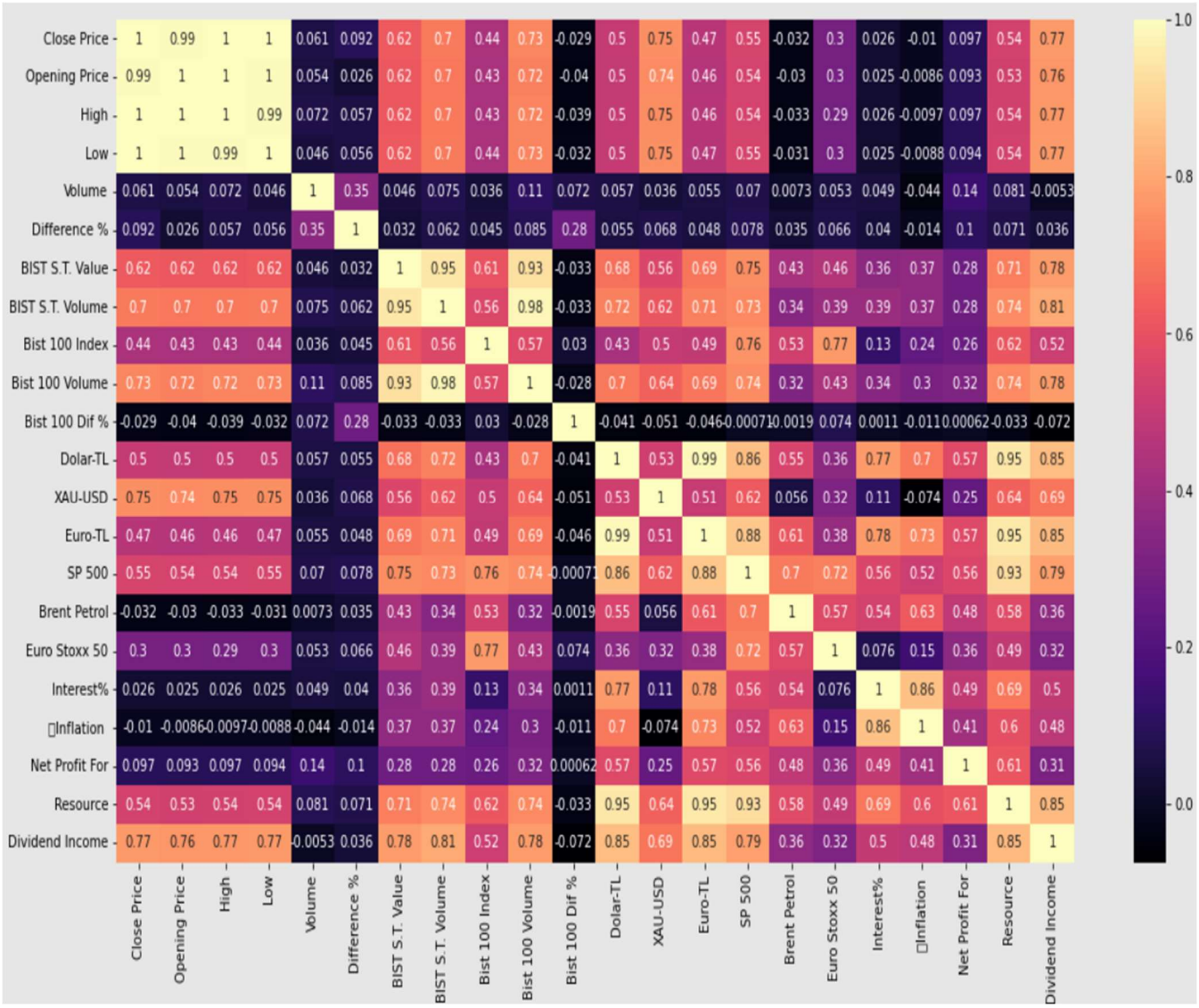
(Ferreira, Le 2019)

In order to the data to be of good quality, empty lines were removed during the data preprocessing stage, all the features were collected numerically, and dummy variables were not included in the model.

### 2.4 Analysis of Features

In this study, the target is the stock price as the dependent variable, since the desired value is the stock price. The remaining features are independent variables. It is important to analyze which features affect the stock price and how much.

**Table 3.** Heatmap of the data set




According to the heat map in table 3, the factors that affect the stock price the most are the features closest to 1 in the heat map. According to the analyzed data set, the current opening high and low values affect the stock price the most. The reason for this is that the stock price we are looking for is very close to the opening, high and low values of the same day. This was actually something we could see before the heatmap. Here there are other values close to 1. For example, with a value of 0.77 in dividend income, it is seen that it significantly affects the stock price. It is understood that the gold ounce price, which comes after that, with a value of 0.75, is also a feature that affects the stock price. Bist 100 volume and star market trading volume are among the features that significantly affect the dependent variable. These ratios can be used when selecting features to increase success.

### 2.5 Predicting Stock Price Using Machine Learning Algorithms

The stages applied in machine learning are:

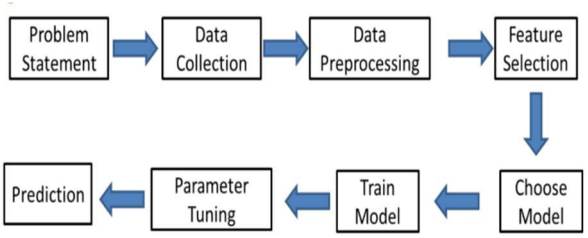



**Figure 3.** Machine learning steps

#### 2.5.1 Creating machine learning model

In the machine learning model, the data set is divided into 80% training and 20% testing. The aim was to predict the closing price of the stock for current day. Therefore, the y dependent variable is

the stock price, represented by x, the independent variables being the remaining features.

In the first stage, all the features were included in the system and the model was created. No feature selection and normalization has been done. The aim here is to determine which methods and which algorithms make the best predictions. Success found using no methods is compared to the success of predictions found using certain methods.

### 2.5.2 Cross validation

Cross validation is a resampling method to avoid memorization and generalize the model. In cross validation, the data set is divided into subsamples. Separate training and test samples are created for each sub-sample. The training and testing part of each sub-samples are different samples. The model learns from different parts of the data in each sub-sample. The model's estimate error is calculated for all sub-samples and their average is the model's error (King, 2021; Orhobor, 2021; Taylor, 2021). In this study, k-fold cross validation technique was used.

### 2.5.3. Feature selection

Through feature selection methods, the computation time of machine learning algorithms can be reduced, prediction success can be increased, and data can be better understood. There are many methods for feature selection in literature. These methods can be roughly classified as filter methods, wrapper methods, embedded and hybrid methods.

The purpose of feature selection is exclude unnecessary features that negatively affect the model that cause a decrease in success. Which feature combination will give the most successful result can be found by brute force method by trying one by one. However, this job is feasible only in models with very few features. It will be very expensive to calculate this in models with many features (Chandrashekar, 2014; Sahin, 2014; Jovic, 2015; Bogunovic, 2015).

## 3. RESULTS OF MACHINE LEARNING ALGORITHMS

This study and other studies in the literature give the result that there is a relationship between the stock market and macroeconomic variables. There are macroeconomic variables (such as interest rate, inflation, exchange rate, oil prices, gold prices) as well as intra-firm factors (such as firm performance, dividends, incomes, changes in the board of directors) that affect the stock price traded in the stock markets. In this study, internal factors and macroeconomic variables affecting stock prices were taken as features and it was investigated how much these features affect stock prices. Accordingly, machine learning models were created, and feature selections were made to increase success. Algorithm performances and used methods were compared.

The success of the test results in the studies created with the QNBFB data set was very high. The reason for this is that the target variable to be found is very close to 3 features. The QNB index values are very close to the “Opening, High, and Low” features. According to table 3, these 3 independent variables in the QNB index have a high correlation. Knowing the "Opening, High and Low" features of the model while training has greatly increased the success in the prediction. However, this situation is not effective and useful when predicting stock price in practice. Because in practice, the target is to predict the end-of-day closing index. The end-of-day features "low, high, open" may not be known at the beginning. Therefore, the model was created without using these 3 features for training while predicting the "Enkai" stock price in order to be more realistic in its application to daily life.

### 3.1. Effect of Machine Learning Algorithm to Result

In the study, when comparing algorithms, min-max normalization, which was the best normalization method before, was used, except for support vector regression and artificial neural network. In SVR and ANN, on the other hand, standard scaler(ss) normalization was used because it gave better results. Models were created with the parameters that gave the best results in parameter comparison before. No parameter optimization was done for MLR. For SVR, the highest success was achieved in the model where the kernel was determined linearly. The best results were obtained with default parameters for DTR and BRR. The model was created with various estimator numbers for RFR. The best result was obtained with the RFR created with 300 estimators. The effect of epoch number and learning rate on success for ANN was investigated and the optimum result was obtained with the parameters where learning rate was 0.0001 and epoch number was 1000.

Table 4 shows the parameters and normalization methods for which the algorithms were most successful, and the model was created with the same number of features and the performances of the algorithms were compared.

The success of the machine learning algorithms used in this study was mostly high. Reason for this may be enough data used in the models, using of optimum normalization, the appropriate parameter selection, the quality of the data set, and the appropriate selection of the features. In feature selection, the heatmap is basically used. Features that do not affect the closing price of the stock were removed from the model and more successful results were obtained with the remaining 18 features.

**Table 4**. Comparing the results of machine learning models.

| Test Name | Algorithm | R Square | MAE | MSE | Feature Number | Methods | Parameters |
|---|---|---|---|---|---|---|---|
| Eregli_test1 | MLR | 0.99320 | 0.01010 | 0.00035 | 18 | Min-max, cv | default |
| Eregli_test2 | SVR | 0.98759 | 0.07178 | 0.00865 | 18 | ss, cv | kernel=linear |
| Eregli_test3 | DTR | 0.99538 | 0.00679 | 0.00022 | 18 | Min-max, cv | default |
| Eregli_test4 | RFR | 0.99537 | 0.00762 | 0.00023 | 18 | Min-Max, cv | n_estimators=300 |
| Eregli_test5 | BRR | 0.99303 | 0.00993 | 0.00035 | 18 | Min-Max, cv | default |
| Eregli_test6 | ANN | 0.99875 | 0.00477 | 0.00006 | 18 | Ss, cv | epoch=1000, lr=0.0001 |

**Figure 4**. R square results for test data in machine learning model

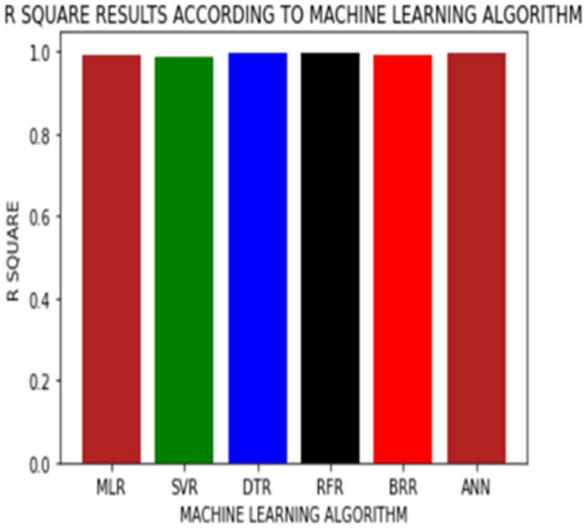



**Figure 5.** R square results for 104 unseen data in machine learning model

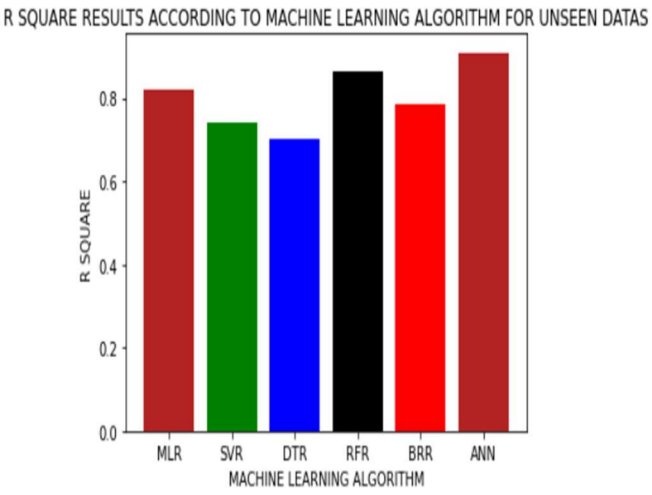



**Table 5.** Comparison of the actual values with the predicted values for QNBFB stock

| Real Endex vs Prediction | | |
|---|---|---|
| | 0 | 0 |
| 0 | 4.959 | 4.902183 |
| 1 | 4.189 | 4.174437 |
| 2 | 5.790 | 5.643054 |
| 3 | 4.278 | 4.247966 |
| 4 | 4.730 | 4.600707 |
| .. | ... | ... |
| 424 | 6.061 | 6.045184 |
| 425 | 4.430 | 4.345840 |
| 426 | 4.394 | 4.342442 |
| 427 | 4.305 | 4.230426 |
| 428 | 39.000 | 39.637070 |

[429 rows x 2 columns]

**Table 6.** Random Forest Regression Actual vs Predicted Values for Unseen Data

| | Eregli Endeks | 0 |
|---|---|---|
| 0 | 25.654 | 24.239450 |
| 1 | 26.513 | 25.376003 |
| 2 | 27.075 | 24.821627 |
| 3 | 26.496 | 25.458620 |
| 4 | 26.864 | 25.919260 |
| .. | ... | ... |
| 99 | 32.420 | 34.205427 |
| 100 | 32.880 | 34.027070 |
| 101 | 34.740 | 35.350970 |
| 102 | 34.200 | 35.613210 |
| 103 | 34.240 | 35.328943 |

[104 rows x 2 columns]

**Table 7.** Neural Network (MLP) Actual vs Predicted Values for Unseen Data

| | Eregli Endeks | 0 |
|---|---|---|
| 0 | 25.654 | 27.431393 |
| 1 | 26.513 | 26.643353 |
| 2 | 27.075 | 27.583457 |
| 3 | 26.496 | 26.989741 |
| 4 | 26.864 | 26.284076 |
| .. | ... | ... |
| 99 | 32.420 | 34.401145 |
| 100 | 32.880 | 35.831437 |
| 101 | 34.740 | 35.985915 |
| 102 | 34.200 | 36.425303 |
| 103 | 34.240 | 35.251362 |

[104 rows x 2 columns]

### 3.2 Effect of Feature Selection Methods to Result

In Table 8, the performances of the algorithms are compared before and after the feature selection method is applied. As it can be understood from Table 8, feature selection methods increased the success in all algorithms. The most significant increase in success has been in SVR, DTR and ANN algorithms. For all algorithms, backward elimination (bwe) and forward feature selection(ffs) methods have been tested. The success for MLR, SVR and DTR algorithms has increased with the forward feature selection method. The success for the RFR, BRR and ANN algorithms has increased with the backward elimination method.

**Table 8.** Comparing machine learning methods with and without feature selection method for unseen data.

| Test Name | Algorithm | R Square | MAE | MSE | Feature Number | Methods | Parameters |
|---|---|---|---|---|---|---|---|
| Eregli_test7 | MLR | 0.8208 | 1.4129 | 2.7662 | 18 | min-max, cv | default |
| Eregli_test8 | MLR | 0.8336 | 1.3587 | 2.5685 | 17 | ffs,min-max, cv | default |
| Eregli_test9 | SVR | 0.7402 | 1.6551 | 4.0114 | 18 | ss, cv | kernel=linear |
| Eregli_test10 | SVR | 0.8492 | 1.2793 | 2.3286 | 9 | ffs, ss, cv | kernel=linear |
| Eregli_test11 | DTR | 0.7025 | 1.7568 | 4.5945 | 18 | min-max, cv | default |
| Eregli_test12 | DTR | 0.7659 | 1.4455 | 3.6146 | 12 | Ffs, min-max, cv | default |
| Eregli_test13 | RFR | 0.8629 | 1.2241 | 2.1161 | 18 | min-max, cv | n_estimators=300 |
| Eregli_test14 | RFR | 0.8656 | 1.2255 | 2.0748 | 17 | bwe, min-max, cv | n_estimators=300 |
| Eregli_test15 | BRR | 0.7867 | 1.5828 | 3.2934 | 18 | min-max, cv | default |
| Eregli_test16 | BRR | 0.7868 | 1.5468 | 3.2913 | 16 | bwe, min-max, cv | dafault |
| Eregli_test17 | ANN(MLP) | 0.9089 | 0.9544 | 1.4056 | 18 | ss, cv | act=logistic, hidden layer=4 max_iter=1 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Eregli_test18 | ANN(MLP) | 0.9424 | 0.7649 | 0.8885 | 7 | bwe, ss, cv | act=logistic, hidden layer=4 max_iter=1000 |

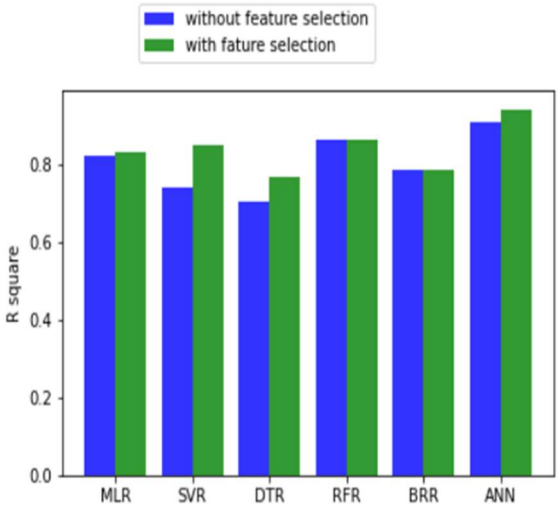



**Figure 6.** Effect of feature selection method in algorithm performances

## 4. CONCLUSION AND SUGGESTIONS

The success of the machine learning algorithms applied in this study was generally high. In order to the success to be high, 1070 data are used in machine learning models.

The effect of machine learning algorithm on success was investigated and the most successful results were obtained when ANN was applied according to table 4, figure 4 and 5. According to Table 5, the most successful algorithm, ANN, predicted the QNBFB stock dated 6 January 2022 as 39.63 with a real index value of 39.00. Since more successful results were obtained for unseen data, MLP, which is the ANN type, was applied. When the success was measured for the observed data, results were close to each other for all algorithms. While the SVR was given according to figure 4 for the data with the most unsuccessful result among the 6 algorithms applied, it gave DTR according to figure 5 for the data not seen.

The random forest algorithm has achieved better performance than the decision tree algorithm because the random forest consists of many decision trees trained from different subsets of the data. Taking averages over many decision trees reduces variance and over-fitting. Artificial neural networks can learn the non-linear relationship between dependent and independent variables. It does this by using a large number of neuron layers with non-linear activation functions. ANN can capture patterns that random forest algorithms and decision trees cannot capture (Robbach, 2018). Therefore, ANN have shown better performance than linear models and tree models.

In order to increase the success of machine learning algorithms in the study, forward selection and backward elimination methods, which are feature selection methods, were applied. According to Table 8, the feature selection methods increased the success of R square and decreased the error measures in the results obtained from testing the final stock closing price of 2021 for unseen data that is not in the training data set. For unseen data, the R square success of MLR algorithm increased from 82.0% to 82.3%, MAE and MSE decreased with forward feature selection method. Similarly, applying the forward feature selection method in the SVR algorithm increased the success of R square from 74.0% to 84.9% and greatly reduced the error measures. The forward feature selection method in DTR algorithm increased the success of R square from 70.2% to 76.5% and decreased the error measures. Since the forward feature selection method did not increase the success in the RFR algorithm, the backward elimination method was applied. Although the backward elimination method did not significantly increase the performance of the RFR algorithm, it increased the success of the R square from 86.2% to 86.5% and slightly decreased the error rates. The forward feature selection method for the BRR algorithm did not increase the success, but the success increased slightly with the backward elimination method. In the MLP algorithm, which is the ANN type, the backward elimination method increased the R square success of the model from 90.8% to 94.2% and was successful by reducing the error rates.

In the study, cross validation technique was applied to prevent overfitting. In order to show that machine learning models learn without overfitting, predictions are made for the 104-days index of 2021, which is not in the training dataset of the models. As seen in Table 6 and Table 7 prediction results close to the actual value were found in the prediction of these 104 unseen data. For this reason, it has been shown that the machine learning models applied in the study can be generalized.

In this study, machine learning models were created by taking 3 internal and 18 external features in the models with the highest number of features. The number of internal features can be increased in future work. The number of features can be increased by adding the features used in this study by performing sentiment analysis with daily data

compiled from container data or other financial news sites. In this study, 5 years of data were collected. The success of machine learning algorithms can be increased by adding more rows to the training data set.

**REFERENCES**


Summers, D. (2007) Longman Business English Dictionary, Pearson Longman, London, 594 p.

URL-1: https://dataconomy.com/2023/01/11/stock-prediction-machine-learning.
[Access date: 20.04.2023]

URL-2: https://builtin.com/machine-learning/machine-learning-stock-prediction.
[Access date: 20.04.2023]

Hürer, E. (1995) Hisse Senedi Fiyatını Etkileyen Faktörler ve İMKB Üzerine Bir Uygulama, İstanbul University, İstanbul, 208 s.(Master Thesis)

Ghani, M., Awais, M., Muzammul (2019), Stock Market Prediction Using Machine Learning(ML) Algorithms, *Advances in Distributed Computing and Artificial Intelligence Journal*, 4, pp. 97-116.

Sarode, S., Tolani, H., Kak, P., Lifna, C. (2019) Stock Price Prediction Using Machine Learning Techniques, *International Conference on Intelligent Sustainable Systems* (ICISS), Palladam, India.

Usmani, M., Adil, S., Raza, K., Ali, S. (2016) Stock Price Prediction Using Machine Learning Techniques. *3rd International Conference On Computer And Information Sciences* (ICCOINS), Kuala Lumpur, Malaysia.

Tipirisetty, A. (2018) Stock Price Prediction using Deep Learning. San Jose State University, Department of Computer Science, California, 54s. (Master Thesis)

Singh, S. Stock Prediction using Machine Learning, California State University, Computer Science, California, 2021, 16s. (Master Thesis).

Guo, Y. Stock Price Prediction Using Machine Learning, Sodertorn University, School of Social ScienceMaster, Economics, Stockolm, 2022, 34. (Master Thesis).

Molnar C. (2019) Interpretable Machine Learning, Lulu.com, 314 p.

Bi, Q., Goodman, K. E., Kaminsky, J., Lessler, J.(2019) What is machine learning? *A primer for the epidemiologist, American journal of epidemiology*, 188(12), 2222-2239.

URL-3 //www.isyatirim.com.tr/tr-tr/analiz/hisse/Sayfalar/Temel-Degerler-Ve-Oranlar.
[Access date: 12.12.2022]

Aslan, B. (2020), Derin Öğrenme ile Borsa Verileri Üzerinde Tahminleme Yapılması, Ege Üniversitesi, İzmir, 61. (Master Thesis)

URL-4: https://borsaistanbul.com/tr/sayfa/506/pazarlar
[Access date: 18.12.2022]

URL-5: https://www.alnusyatirim.com/bist-100
[Access date: 18.12.2022]

Karagöz, S. (2020), Payların Kapanış Fiyatlarının Makine Öğrenmesi Yöntemleri ile Tahmin Edilmesi,, İstanbul , 118. (master Thesis).

URL-6: https://bigpara.hurriyet.com.tr/
[Access date: 21.11.2022]

URL-7: https://en.wikipedia.org/wiki/S%26P_500
[Access date: 22.12.2022]

URL-8: https://en.wikipedia.org/wiki/EURO_STOXX
[Access date: 22.12.2022]

URL-9 https://www.tcmb.gov.tr/
[Access date: 23.12.2022]

Kotsiantis, S.B., Kanellopoulos, D., Pintelas P.E.(2006), Data Preprocessing for Supervised Leaning. *International Journal of Computer Science Volume* 1, pp. 111-117

Alexandropoulos, S.N., Kotsiantis S.B., Vrahatis M.N. (2019), Data Preprocessing in Predictive Data Mining, Cambridge University Press 34 E1.

King, R., Orhobor, O., Taylor, C (2019) Cross-Validation is Safe to Use, Nature Machine Intelligence. 2021, 3, pp. 276-276.

Daniel, B. (2021) Cross-Validation, Data Science Laboratory, 2, pp. 542-545.

Chandrashekar, G., Sahin, F., A Survey on Feature Selection Methods. Computers & Electrical Engineering, 2014, 40, pp. 16-28.

Jović, A. and Brkić, K., Bogunović, N. A Review of Feature Selection Methods with Applications. *38th International Convention on Information and Communication Technology*, 2015, Croatia.

Zhang, F., O'Donnel, L. Support Vector Regression, Machine Learning Methods and Applications to Brain Disorders. 2020, 7, pp. 123-140

Raju, G., Lakshmi, K., Jain, V., Kalidindi, A., Padma V., Study the Influence of Normalization/Transformation process on the Accuracy of Supervised Classification. *Third International Conference on Smart Systems and Inventive Technology (ICSSIT)*, 2020, India.

Skinea, S., Data Science Design Manual, New York, USA, 2017, 453 s.

Liu, Y., Wang, Y., Zhang, J. New Machine Learning Algorithm: Random Forest. *International*

*Conference on Information Computing and Applications*, 2012, pp 246-252)

URL-10
https://towardsdatascience.com/introduction-to-bayesian-linear-regression
[Access date: 20.04.2023]

Bonaccorso, G. Machine Learning Algorithms., *Packt Publishing, Birmingham*, UK, 2017, 337s.

Ferreira, P., Le. D., Zincir-Heywood N., Exploring Feature Normalization and Temporal Information for Machine Learning Based Insider Threat Detection. *15th International Conference on Network and Service Management (CNSM),* 21-25 October, 2019, Halifax, NS, Canada.

Robbach P. Neural Networks vs. Random Forests – Does it always have to be Deep Learning? *Computer Science*, 2018.

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

**Araştırma Makalesi**

# Real Random Number Generation by Chemical Reactions Based on Quantum Wave Equation

**Muharrem Tuncay Gençoğlu*[1], Tuncay Genç[2]**

[1] *Fırat Üniversitesi, Teknik Bilimler MYO, Elâzığ, Türkiye*
[2]*Emniyet Müdürlüğü, Elâzığ, Türkiye*

**Keywords:**
Chemical reaction,
Random number generator,
True random number generator

**ABSTRACT**

Random Number Generators are software or hardware components that allow the production of unpredictable number sequences without any pattern or relationship between them. Various studies have been conducted with different techniques regarding RNG. In these studies, the difficulties of random number generation and the high cost negatively affect the efficiency of the developed generators. Many different methods have been used in real random number generation, and even quantum random number generators have been developed to make predictability difficult. Quantum Random Number Generators; are a tepe of generator based on the laws of Quantum physics instead of classical physics. In photonic-based RNG, random numbers are generated after various software and hardware operations by utilizing the uncertainty of photons. This study, it is aimed to develop a true random number generator using chemical reactions that have not been studied before. Data was produced by using sensors and other hardware elements together, the values produced were taken as seed values and assigned as input to the algorithm used in generating random numbers, and true random numbers were produced and these numbers were tested in detail with known test methods.

# Kuantum Dalga Denklemi Tabanlı Kimyasal Reaksiyonlarla Gerçek Rastgele Sayı Üretme

**Anahtar Kelimeler:**
Kimyasal reaksiyonlar,
Rasgele sayı üretimi,
Gerçek rasgele sayı üretimi

**ÖZ**

Rastgele Sayı Üreteçleri, aralarında herhangi bir örüntü veya ilişki olmayacak şekilde tahmin edilemeyecek sayı dizileri üretilmesini sağlayan yazılımsal veya donanımsal bileşenlerdir. RSÜ ile ilgili farklı tekniklerle çeşitli çalışmalar yapılmıştır. Bu çalışmalarda rastgele sayı üretiminin zorlukları ve maliyetin yüksek olması geliştirilen üreteçlerin verimliliğini olumsuz etkilemektedir. Gerçek rastgele sayı üretiminde çok farklı yöntemler kullanılmış hatta tahmin edilebilirliği zorlaştırmak için kuantum rastgele sayı üreteci dahi geliştirilmiştir. Kuantum Rastgele Sayı Üreteçleri; klasik fizik yerine Kuantum fiziği yasalarının temel alındığı bir üreteç çeşididir. Fotonik tabanlı KRSÜ'de fotonların belirsizliğinden faydalanılarak çeşitli yazılımsal ve donanımsal işlemlerden sonra rastgele sayılar üretilir. Üretilen bu sayılar, tahmin edilemeyecek seviyede güçlü rastgele sayılardır. Ancak bu yöntemin hem insan sağlığı hem de maliyet açısından olumsuzlukları mevcuttur. Bu çalışmada, özellikle radyoaktif rastgele sayı üreteçlerine alternatif olacak ve maliyeti düşürmek adına daha önce çalışılmamış olan kimyasal reaksiyonlar kullanılarak gerçek rastgele sayı üreteci geliştirilmesi amaçlanmıştır. Donanımsal kaynaklar ve kimyasal reaksiyonlar birlikte kullanılarak gerçek rastgele sayılar üretilmiştir. Sensörler ve diğer donanım elemanlarının ortak kullanımıyla veri üretilmiş, üretilen değerler tohum değeri olarak alınıp, rastgele sayı üretmede kullanılan algoritmaya girdi olarak atanarak gerçek rastgele sayılar üretilmiş ve bu sayılar bilinen test yöntemleriyle detaylı olarak test edilmiştir.

* **Muharrem Tuncay Gençoğlu**

*(mt.gencoglu@firat.edu.tr) ORCID ID 0000 – 0002 – 8784 – 9634
(tncygnc@gmail.com) ORCID ID 0000 – 0002 – 8325 – 3243

**e-ISSN: 2717-8579**

**Geliş Tarihi: 05/10/2024; Kabul Tarihi: 16/12/2024**

## 1. INTRODUCTION

The random number (RN) is the number we obtain by mathematically and evenly distributing the elements in a series whose members are known so that new choices cannot be predicted from previous choices (Chaitin,2001).

The history of random numbers goes back a long way. Dice, coins, and other devices have been used to generate random numbers in random elections and games of chance. In particular, dice were used to make important decisions such as inheritance sharing and presidential elections. In addition to dice, card games, coins, spinning wheels, etc. objects were also used as early random number generators.

Random numbers began to be used in later years, especially in cryptology. It has been used to generate keys in encryption.

Currently, images, patterns and 3D objects are created using random data through certain programs and computers. Random numbers are used in secure communication applications where only the receiver and transmitter know the content or in data-hiding applications where only the user needs to know the content (Daemen,2013). The fields where random numbers are used include sampling, entertainment, modeling, simulation and testing, decision-making, cryptography, computer games, computer programming and electronic design(Robinson,1998;Schoukens,1988; Schindler,2002).

Random numbers are very important for ensuring the confidentiality and reliability of the encryption process (Avaroğlu,2017; Tuncer and Genç,2019). The use of random numbers in cryptographic applications increases the encryption strength.

The situations that can be used as sources of randomness are listed as follows:

- Time spent during electrostatic release in radioactive decay
- Thermal noise caused by a resistor or diode element
- Parameter instability between independently operating oscillators
- The charging time of the semiconductor capacitor for a certain period
- Air turbulence on a hard disk
- An arbitrary amount of software-based audio from a microphone or image from a camera [8].

Although there are many different RNG structures for generating random numbers, it is generally possible to divide them into three classes. These are called pseudo random number generators (PRNG), true random number generators (RRNG) and hybrid random number generators (HRNG). The classification of random number generators is shown in Figure 1 (Koç,2009):

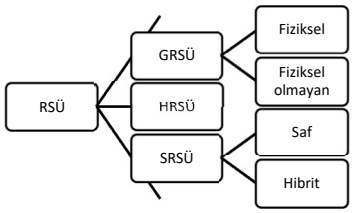



**Figure 1.** Classification of Random Number Generators

In general, random properties must be met in random number generation. RRNG is based on a physical state known to be random. Sources that generally produce noise or noise sources found in nature can be used as examples. In other words, even if the RRNG is run twice under exactly the same conditions, it produces two unrelated sequences of random numbers. PRNGs are number generators that are based on predictable equations, contain random data, and calculate random data generation in their processor in a limited situation.

The differences between RRNG and PRNG are shown in Table 1 (Von Neumann,1951):

**Table 1.** Differences between RRNG and PRNG

| RNG | Sufficiency | Determinism | Periodicity |
|---|---|---|---|
| **PRNG** | Perfect | Deterministic | Periodic |
| **GRNG** | Weak | Nondeterministic | Nonperiodic |

Figure 2 shows the general structure of real random number generation.

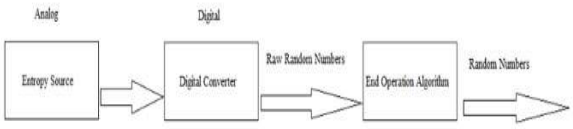



**Figure 2.** General structure of random number generation

Random numbers form the most important part of many systems and applications where security is at the forefront. In critical areas such as cryptographic applications, games of chance, and password generators, the security of the application is based on random numbers (Wold,2011; Yakut et al.,2019). Yakut proposed a random number generator that can be easily generated using any digital data source (Yakut,2021).

In such security applications, RRNG is generally preferred. The security of encrypted systems is based on the fact that confidential data or keys are known by authorized individuals and cannot be guessed by other individuals.

Random values are needed here to make it difficult for others to guess the secret information.

Malicious users can create security vulnerabilities by exploiting the weaknesses of random number generation methods. Therefore, the

use of random numbers in important areas such as security makes it important that these numbers are close to real random numbers and have the properties of real random numbers (Chaitin,2001; Sanguinetti,2014). Apart from security, random numbers also play an important role in simulating real events. The use of different sources as entropy sources in RRNG is available in the literature (Wold,2011). However, the generation of random numbers in uncontrolled environments outside the system poses a problem for the security of the system (Avaroğlu,2014).

RRNG generally consists of three blocks. These:

- Entropy (noise) source,
- Sampler (digitizer),
- End operation algorithms.

The concept of entropy constitutes the second law of thermodynamics theory. Entropy is defined as a measure of the qualitative disorder and randomness of a system (Kapur and Kesavan,2014). The sampler ensures the necessary sampling of the noise signal, and this structure can be expressed as the production mechanism for physical noise sources (Wold,2011). Thus, it is possible to obtain a digitized signal from an analog signal.

There are different approaches to sampling, and the sampler, along with the entropy source, has an important role in determining the quality of the numbers produced. Postprocessing is often used to increase the randomness present in the signal. This operation is applied to long-bit sequences and propagates in an autocorrelated manner. Here, the coefficient of two adjacent bits in the connected bit stream is greater than the coefficient of connection between the distant bits. Therefore, the relationships between bits that are close to each other are stronger than the relationships between bits that are far away from each other. In post processing algorithms, better results are obtained from statistical tests by rearranging the autocorrelation instead of a simple compression process. The postprocessed signal has a more uniform distribution and random appearance compared to its pure form.

Random numbers generated by post processing are more resistant to side-channel analysis attacks and less affected by environmental factors. Therefore, post processing algorithms make the generator more secure.

There are different post processing algorithms, such as XOR verification, Von Neumann verification, extractor function, cryptographic hash algorithms and resilient function (Dichtl,,2007). RRNG cryptological applications and chaos-based PRNG-RRNG applications have been developed on FPGAs (Yıldırım,2012; Özkaynak,2013; Özkaynak,2014; Özkaynak,2015). Cirauqui et al. performed correlation analysis and compared the effects that potentially hidden correlations in random or pseudo-random flows can have in some physical MC simulations (Cirauqui et al.,2024) Daojing et al. have created a Software Random Number Generator Applicable to the Internet of Things (Daojing et al. ,2024). Luis et al. used quantum mechanics for pseudo-random number generation based on simulated quantum processes as a source of entropy Luis et al.,2025).

### 1.1. RRNG Designs in Literature

In their study, Voris et al. suggested that the accelerometer and temperature on the Wireless Identity and Sensing Platform (WISP) are better sources of entropy than other sensors (Voris et al.,2011). However, using only these two sensors in motionless environments where the temperature does not change is insufficient for random number generation. In Mitra's study, a true random number generator suitable for generating seed values was proposed. RRNG was implemented with a dual-fed operational amplifier (Bedekar and Shee,2015). Hennebert et al. found that the best candidates as possible sources of entropy are accelerometers, magnetometers, vibration sensors, and internal clock sensors (Hennebert et al.,2013).

In their study, Bedekar and Shee presented a practical method for assessing the GRSS by using microelectromechanical system sensors (MEMS) (accelerometer, gyroscope and compass) (Bedekar and Shee,2015). In another study by Vivier et al., a pseudorandom number generator was designed based on an n-cube without a Hamilton cycle. Since this method, which has passed classical tests, is carried out only with integers, the security was evaluated as weak as a result of the NIST test (Vivier et al.,2017). In another study, Akgül et al. designed only an interface and did not introduce a generator (Akgül et al. 2019). In 2020, a study on pseudorandom number generation was conducted and tested by Rezk et al. (Rezk et al.,2020).

In one recent study, Avaroğlu and Tuncer designed a new true random number generator based on an S-box (Avaroğlu and Tuncer,2020). The disadvantage of this work is that there is a correlation with the generalized bit sequence coming from the entropy source.

Cryptography is a fundamental component of network security and therefore cybersecurity [28]. The most important problem in public key cryptography is finding a unique and nonrepeatable key. There are two methods for generating the key. The first is a rigorous and powerful mathematical algorithmic approach. The second is to imitate nature.

In 2020 and beyond, studies focused on quantum random number generators (Smith et al.,2020; Lin et al.,2020; Kavulich et al. 2021). In a study on random number generation with quantum technology, unpredictable random numbers were produced by using photons obtained from photo frames taken from phone cameras (Dutang and Wuertz,2009). Since generating random numbers

using quantum technology is costly and these techniques are not widely used today, there are thoughts that they will be widely used in the future with the development of quantum technology (Gençoğlu,2021).

It is important to ensure certain features when generating numbers in a random number generator. It should be as random as possible, randomness should be ensured over long periods, and the generated random numbers should be reproducible, calculable and reusable when necessary. Can a true random number generator be created that provides all these features and is efficient, low-cost, and easy to use? question became the source of motivation for this study.

### 1.2. Quantum Wave Equation

The wave equation is a partial differential equation that has a very important place in physics. Wave equations, which have a very wide usage area, have started to be used in cryptography in recent years (Gençoğlu and Agarwal,2021). When the wavefunction is used in the Schrödinger equation, it is also called a quantum wavefunction. This equation provides information about the future behavior of a dynamic system and predicts the distribution of outcomes by analytically and precisely predicting the likelihood of events. The combination of a physical system consisting of a particle and a wavefunction is one of the assumptions of quantum mechanics. The wavefunction can be complex (Gençoğlu,2013).

## 2. MATERIALS AND METHODS

In this study, a new hybrid approach is proposed for the use of quantum wave equationbased algorithms in cryptography by using seed values obtained through chemical reactions, combining mathematical calculations and natural phenomena. This work aims to generate a quantum wave equation-based, low-cost random number using chemical reactions for nonreproducible, unpredictable and efficient real random number generation (RRNG) that exhibits good statistical properties. For this purpose, the following hypotheses have been proposed:

1. The use of chemical reactions to generate real random numbers has a positive impact on the efficiency and cost of the generated random numbers.
2. The use of chemical reactions as a noise source in generating real random numbers is important because of their good statistical properties.
3. Chemical reactions that can be used as alternatives to radioactive random number generators positively affect the development of lowcost random number generators.

The greatest disadvantages of existing RRNGs and PRNGs are that they are costly and predictable. The disadvantages of QRNGs, which are the most reliable in terms of unpredictability, are cost, negativities caused by radioactivity and difficulty of use. In the model we propose, since the numbers obtained from different seed values in the random number generation process will be combined with a function f;

- Even if a part of the generated number sequence is obtained, it is impossible to obtain the other part.
- The number sequence does not contain periodic results.
- The produced sequences will not have any hidden correlations within themselves.

The first step in the theoretical approach and method followed in this study is seed data generation. For this, a seed was planted from a corn cob and data was obtained from the chemical reactions occurring in the plant and the environment during the germination and growing process. Then, these data were used as input in a mathematical algorithm based on quantum wave equations to produce true random numbers.

The first step in the theoretical approach and method followed while carrying out this study is seed data generation. For this purpose, a seed was planted from a corn cob, and data were obtained from the chemical reactions occurring in the plant and the environment during the germination and growing process. Then, these data were used as input in a quantum wave equation-based mathematical algorithm to generate true random numbers.

Afterwards, the obtained random numbers were tested.

Stage 1

Chemical reactions have been used as noise sources. The light source, soil, water, precision scale were used to calculate the weight gain, and humidity and thermometer were used to measure the ambient humidity and temperature. The weight values of the corn plants on the precision scale were recorded at regular intervals, as were the humidity and temperature changes. The data obtained in this direction are shown in Table 2.

**Table 2.** Weight, Humidity and Temperature Values According to Measurement Order

| Measurement Order | Weight | Humidity | Temperature(°C) |
|---|---|---|---|
| 1 | 277.33 | 51.0 | 26.8 |
| 2 | 271.81 | 45.0 | 28.1 |
| 3 | 272.24 | 52.0 | 23.6 |
| 4 | 271.41 | 55.0 | 24.1 |
| 5 | 270.68 | 54.0 | 24.3 |
| 6 | 268.67 | 44.0 | 24.1 |
| 7 | 268.35 | 48.0 | 24.3 |
| 8 | 263.73 | 54.0 | 23.0 |
| 9 | 262.33 | 47.0 | 23.0 |
| 10 | 261.8 | 42.0 | 22.6 |
| 11 | 261.28 | 53.0 | 23.0 |
| 12 | 260.0 | 52.0 | 22.6 |
| 13 | 258.50 | 53.0 | 22.9 |
| 14 | 258.31 | 48.0 | 22.3 |
| 15 | 258.07 | 55.0 | 23.1 |
| 16 | 257.75 | 44.0 | 27.2 |
| 17 | 257.61 | 45.0 | 27.1 |
| 18 | 257.37 | 45.0 | 27.3 |
| 19 | 257.27 | 46.0 | 27.4 |
| 20 | 257.1 | 48.0 | 27.5 |
| 21 | 255.12 | 48.0 | 27.3 |
| 22 | 299.15 | 47.0 | 28.8 |
| 23 | 298.91 | 46.0 | 28.8 |
| 24 | 298.48 | 48.0 | 28.9 |
| 25 | 296.28 | 45.0 | 28.9 |
| 26 | 295.88 | 45.0 | 29.0 |
| 27 | 295.41 | 44.0 | 29.7 |
| 28 | 288.49 | 45.0 | 29.5 |
| 29 | 287.42 | 46.0 | 29.6 |
| 30 | 286.92 | 43.0 | 29.3 |
| 31 | 286.74 | 43.0 | 29.4 |
| 32 | 286.45 | 42.0 | 29.6 |

| | | | |
|---|---|---|---|
| 33 | 286.02 | 40.0 | 29.8 |
| 34 | 283.39 | 41.0 | 28.3 |
| 35 | 282.82 | 41.0 | 29.3 |
| 36 | 281.79 | 45.0 | 28.3 |
| 37 | 281.63 | 47.0 | 28.4 |
| 38 | 281.42 | 52.0 | 23.7 |
| 39 | 281.34 | 51.0 | 23.8 |
| 40 | 281.14 | 50.0 | 23.5 |
| 41 | 280.98 | 56.0 | 24.1 |
| 42 | 280.85 | 53.0 | 24.2 |
| 43 | 279.71 | 51.0 | 24.4 |
| 44 | 278.92 | 51.0 | 22.9 |
| 45 | 278.77 | 51.0 | 23.7 |
| 46 | 277.39 | 51.0 | 24.2 |
| 47 | 276.92 | 57.0 | 24.5 |
| 48 | 276.82 | 56.0 | 24.5 |
| 49 | 276.67 | 55.0 | 24.7 |
| 50 | 276.39 | 52.0 | 24.0 |
| 51 | 276.32 | 56.0 | 24.3 |
| 52 | 276.23 | 63.0 | 24.4 |
| 53 | 276.06 | 59.0 | 24.2 |
| 54 | 276.0 | 62.0 | 24.4 |
| 55 | 275.84 | 59.0 | 24.3 |
| 56 | 275.77 | 60.0 | 24.3 |
| 57 | 275.38 | 57.0 | 24.0 |
| 58 | 274.7 | 57.0 | 23.6 |
| 59 | 287.25 | 53.0 | 23.1 |
| 60 | 287.18 | 52.0 | 23.1 |

Stage 2

A mathematical formulation was developed by taking into account existing applications in the literature. The quantum wave equation, is a quadric differential equation known as the Schrödinger equation;

$$-\frac{\hbar^2}{2m}\frac{\partial^2\Psi(x)}{\partial x^2} + V(x)\Psi(x) = E\Psi(x) \qquad (1)$$

The general solution of this equation is a linear combination as follows:

$$\Psi(x) = ACos(kx) + BSin(kx) \ (2)$$

Here, a new wavefunction is obtained from the solution of equation (2), where t: time, k: wave vector ($2\pi/\lambda$), λ: wavelength, x: position and w: frequency [35];

$$\psi(x,t) = \frac{1}{\sqrt{2}}[\cos(wt - kx) + sin(wt - kx)] \quad (3)$$

Each of the values obtained from moisture, heat and mass changes was taken as the seed value in equation (3), and different random numbers were generated. The algorithms used for RRNG, which are based on the quantum wave equation, were written using the Python program. Python codes are shown in Figure 3.

The generation of random numbers using the available data is shown in Figure 4.

The general view of the proposed architecture for the random number generator algorithm obtained using the Python program is given in Figure 5 The architecture proposed in Figure 5 is a mathematical function-based random number generator architecture recommended by the data analyzed using the determined parameters. Blocks other than dashed arrows represent analysis stages of statistical randomness processes, which is the basic condition that must be met for generated random numbers.

Stage 3

A run test was used to check the randomness of the results. The run randomness test is a statistical test used to check randomness in data. This nonparametric test uses datasets to determine whether the data presented are random or tend to follow a pattern (Bujang and Sapri, 2018).

The first step in run testing is to count the number of runs in the data array. A run is defined as a series of consecutive positive or negative values.

$$Z = \frac{R - \bar{R}}{S_R}$$

Here,

R= Observed number of runs

$R$=Expected number of runs

$$R = \frac{n_1 n_2}{n_1 + n_2}$$

$S_R$= Standard deviation of the number of runs

$$S_R^2 = \frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1+n_2+)^2\,(n_1+n_2-1)}$$

$n_1$, $n_2$= Number of positive and negative values in the series

Comparing the calculated Z-statistic with the Z critical value for a certain confidence level (Z critical =1.96 for 95% confidence level), if |Z|>Z critical, the numbers are not random Bujang and Sapri, 2018). The test Python codes are given in Figure 6.

## 3. FINDINGS AND DISCUSSION

In this study, in the real random number generator design prepared using corn plants, the plant's weight change and the humidity and temperature values of the environment when the plant's weight was measured were taken as seed values and these values were used to generate random numbers in precise functions using the Python language. The Run test was used to measure the reliability of the numbers.

Run test, which is one of the methods used to test the homogeneity of the data, is a test in which the data to be examined is assumed to come from the same mass and are independent of each other or the opposite assumption can be checked. According to the result of this test, if the data are from the same mass and independent of each other, these series are called simple random numbers. Therefore, it was evaluated that the most reliable analysis could be made with the run test according to the data we have.

In the run test, to say that the numbers are random, the Z value must be less than 1.96. In our study, the Z value was found to be 0.5242377083205431 as a result of the Run test.

```
import math
import xlrd
import xlwt

#excelin olduğu adres
loc = ("C:\\Users\\Tuncay\\Desktop\\proje\\veri.xlsx")
wb = xlrd.open_workbook(loc)
sheet = wb.sheet_by_index(0)
EXCEL_FILES_FOLDER = 'C:\\Users\\Tuncay\\Desktop\\proje\\'
workbook = xlwt.Workbook()
worksheet = workbook.add_sheet('data')
#excel_file_path = EXCEL_FILES_FOLDER+'result.xlsx'
#workbook.save(excel_file_path)
#k=277.33
#n=51
#s=26.8

#k=1. satır 1. sütun
#n=1. satır 2. sütun
#n=1. satır 3. sütun
for i in range(132):
    k=float(sheet.cell_value(i, 0))
    n=float(sheet.cell_value(i, 1))
    s=float(sheet.cell_value(i, 2))

    x1=1/math.sqrt(2)*(math.cos(0.5777*1.6-42.6630*k)+math.sin(0.5777*1.6-42.6630*k))
    x2=1/math.sqrt(2)*(math.cos(4.2817*1.6-5.7567*n)+math.sin(4.2817*1.6-5.7567*n))
    x3=1/math.sqrt(2)*(math.cos(0.2493*1.6-98.7421*s)+math.sin(0.2493*1.6-98.7421*s))


    h1=5*x1*(1-x1)+(3-0.9999)*math.sin(math.pi*x1)/3
    h2=5*x2*(1-x2)+(3-0.9999)*math.sin(math.pi*x2)/3
    h3=5*x3*(1-x3)+(3-0.9999)*math.sin(math.pi*x3)/3

    i1, d1 = divmod(h1, 1)
    o1=round(d1,4)

    i2, d2 = divmod(h2, 1)
    o2=round(d2,4)

    i3, d3 = divmod(h3, 1)
    o3=round(d3,4)

    top=o1+o2+o3
    sonuc=math.pow(math.e,math.sin(math.pi*top))
    worksheet.write(i, 0,sonuc)
    workbook.save('result.xls')
```

**Figure 3.** Random number generator

The formation of random numbers with the available data is shown in Figure 4.

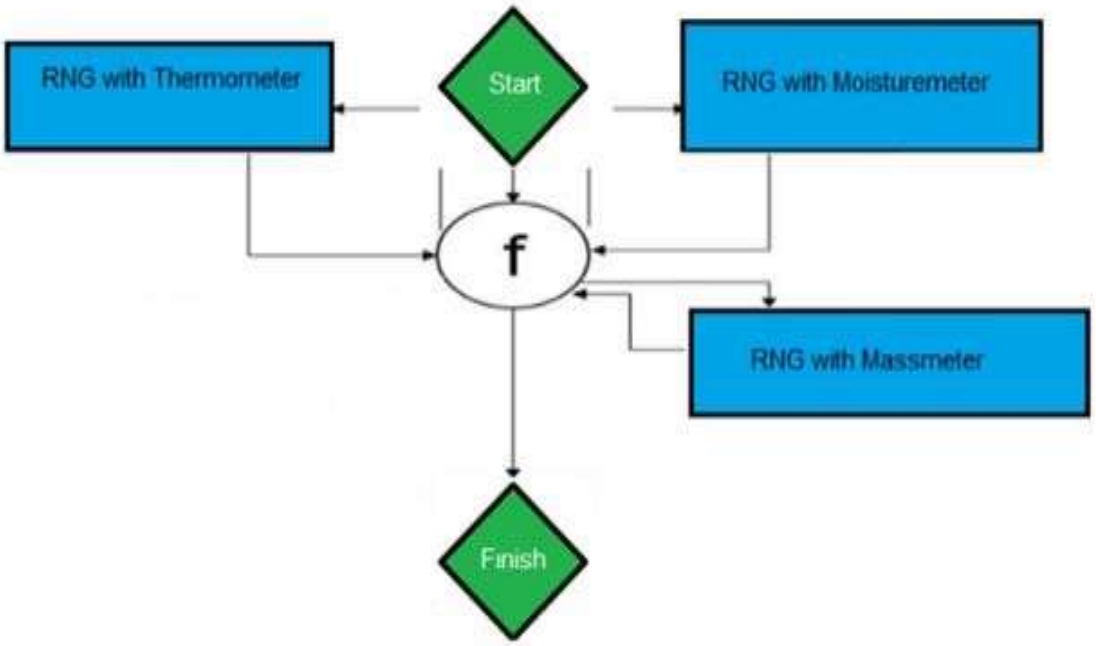



**Figure 4.** Combining Random Number Generators with the f Function

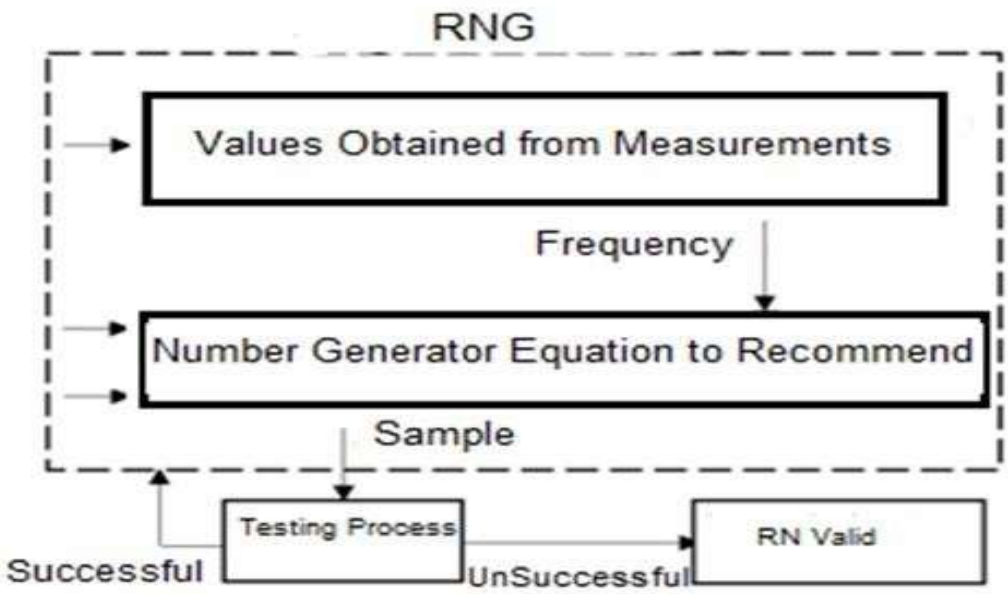



**Figure 5.** Overview of the Proposed Architecture

```
import random
import math
import statistics
import xlrd
import xlwt

def runsTest(l, l_median):

        runs, n1, n2 = 0, 0, 0

        # Checking for start of new run
        for i in range(len(l)):

                # no. of runs
                if (l[i] >= l_median and l[i-1] < l_median) or \
                                (l[i] < l_median and l[i-1] >= l_median):
                        runs += 1
                # no. of positive values
                if(l[i]) >= l_median:
                        n1 += 1

                # no. of negative values
                else:
                        n2 += 1

        runs_exp = ((2*n1*n2)/(n1+n2))+1
        stan_dev = math.sqrt((2*n1*n2*(2*n1*n2-n1-n2))/ \
                                        (((n1+n2)**2)*(n1+n2-1)))

        z = (runs-runs_exp)/stan_dev

        return z


loc = ("C:\\Users\\Tuncay\\Desktop\\test\\result.xls")
wb = xlrd.open_workbook(loc)
sheet = wb.sheet_by_index(0)
l = []
for i in range(132):
    k=float(sheet.cell_value(i, 0))
    print(k)
    l.append(k)

l_median= statistics.median(l)
Z = abs(runsTest(l, l_median))
print('Z-statistic= ', Z)
```

**Figure 6.** Run -Test Python Codes

## 4. CONCLUSIONS

In the proposed model, since the numbers obtained from different starting values are combined with an f function in the random number generation process; Even if a part of the number sequence is generated, it is impossible to obtain the other part. Since the number sequence does not contain periodic results, the generated sequences do not have hidden correlations within themselves. Therefore, it was concluded that the numbers found are not related to each other and have sufficient randomness.

The value obtained as a result of the run test is 0.524 and shows the reliability of the data. Some of the issues considered in the selection of a random number generator are cost, speed, installation, and performance values. It has been observed that the chemical reactions and real number generation presented in this study provide superiority over its competitors in terms of both cost and ease of use. The proposed method offers a different perspective that can be a source for future studies in this field. It is thought to be a guide for new research to be conducted in the future.

Various generators can be designed with data to be obtained from existing plants using appropriate

mechanisms. The technique used can be developed and placed in a cabin system, and a hybrid random number generator can be designed by obtaining more data in a shorter time.

**ACKNOWLEDGMENTS**

Muharrem Tuncay Gencoglu was supported by TUBİTAK (121E323).

**Author Contributions**

Writing – Original draft, conceptualization, and methodology were performed by the MTG. The software, experimental methods, results and outcomes were evaluated via TG.

**Funding**

This study was produced from the master's thesis titled "Quantum Wave Equation Based Real Random Number Generator with the Effects of Chemical Reactions", which is the output of the project supported by TÜBİTAK [121E323].

**REFERENCES**

Chaitin, GJ. (2001). Exploring Randomness, London, Springer.

Daemen, J., Rijmen V. (2013). The Design of Rijndael: AESThe Advanced Encryption Standard, New York, Springer Science & Business Media.

Robinson SO., Dessart, DJ. (1998). Teaching and Learning of Algorithms in School Mathematics, USA, National Council of Teachers of Mathematics.

Schoukens, J., Pintelon, R., van der Ouderaa, E., Renneboog. (1998) J. Survey of excitation signals for FFT based signal analyzers, IEEE Transactions on Instrumentation and Measurements, 37(3), 342-352.

Schindler, W., Killmann, W. (2002). Evaluation criteria for true (physical) random number generators used in cryptographic applications, Cryptographic Hardware and Embedded Systems.

Avaroğlu, E. (2017). LFSR soru girdisi ile puf tasarımının gerçeklenmesi, Fırat Üniversitesi Mühendislik Bilimleri Dergisi. 29(2), 15–21.

Tuncer, SA., Genç, Y. (2019). İnsan hareketleri tabanlı gerçek rastgele sayı üretimi. 8(1), 261–269.

Yalçın M., Suykens J., Vandewalle J. (2004). True Random Bit Generation from a Double Scroll Attractor, IEEE Trans. Circuits Syst.. 51(7), 1395-1404.

Koç, Ç. K. (2009). Cryptographic Engineering, SpringerVerlag.

Von Neumann, J. (1951). Various Techniques Used in Connection with Random Digits, National Bureau of Standards Applied Mathematics Series. 12, 36-38.

Wold, K. (2011). Security Properties of a Class of True Random Number Generators in Programmable Logic, Doctoral Degree, Gjøvik University College, Doctor of Philosophy in Information Security.

Sanguinetti, B., Martin, A., Zbinden, H., Gisin, N. (2014). Quantum random number generation on a mobile phone, Physical Review. 4(3), 031056.

Avaroğlu, E. (2014). Donanım Tabanlı Rastgele Sayı Üretecinin Gerçekleştirilmesi, Doktora Tezi, Fırat Üniversitesi, Fen Bilimleri Enstitüsü.

Kapur, JN., Kesavan, HK. (1992). Entropy Optimization Principles and Their Applications, Netherlands, Springer.

Dichtl, M. (2007). Bad and good ways of post processing biased physical random numbers, International Workshop on Fast Software Encryption.

Yıldırım, S. (2012). A True Random Number Generator in FPGA for Cryptographıc Applications, Master's degree, Middle East Technical University, Graduate School of Natural and Applied Sciences.

Özkaynak, F. (2013). Security problems for a pseudorandom sequence generator based on the Chen chaotic system, Computer Physics Communications.184(9), 2178-2181.

Özkaynak, F. (2020). Cryptographically secure random number generator with chaotic additional input, Nonlinear Dynamics. 78, 2015-2020.

Özkaynak, F. (2015). Kriptolojik Rasgele Sayı Üreteçleri, Türkiye Bilişim Vakfi Bilgisayar Bilimleri ve Mühendisliği Dergisi. 8(2), 37-45.

Voris, J., Saxena, N., Halevi, T. (2011). Accelerometers and randomness: perfect together, Proceedings of the fourth ACM conference on Wireless network security, Hamburg, Germany.

Mitra, M. (2012). A Low-Cost Lightweight Random Number Generator Implementation, International Journal of Engineering Research & Technology. 1(10), 1-9.

Hennebert, C., Hossayni, H., Lauradoux, C. (2013). Entropy harvesting from physical sensors, Proceedings of the sixth ACM conference on

Security and privacy in wireless and mobile networks, Budapest, Hungary.

Bedekar, N., Shee, C. (2015). A Novel Approach to True Random Number Generation in Wearable Computing Environments Using MEMS Sensors. Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics. 8957, 530-546.

Contassot-Vivier, S., Couchot, JF., Guyeux, C., Heam, P.C. (2017). Random Walk in a N-Cube Without Hamiltonian Cycle to Chaotic Pseudorandom Number Generation: Theoretical and Practical Considerations, International Journal of Bifurcation and Chaos. 27(1), 1750014.

Akgül, A., Arslan, C., Arıcıoğlu, B. (2019). Design of an Interface for Random Number Generators based on Integer and Fractional Order Chaotic Systems, Chaos Theory and Applications. 1(1), 1-18.

Rezk, A., Madian, A., Radwan, A., Soliman, A.M. (2019). Multiplierless Chaotic Pseudo Random Number Generators, AEU- International Journal of Electronics and Communications. 113, 152947.

Avaroğlu, E., Tuncer T. (2020). A novel S-box-based postprocessing method for true random number generation, Turk J Elec Eng & Comp Sci. 28, 288-301.

Khan, F. U., Bhatia, S. (2012). A Novel Approach to Genetıc Algorıthm Based Cryptography, International Journal of Research in Computer Science. 2(3), 7-10.

Hurley-Smith, D., Hernandez-Castro, J. (2020). Quantum Leap and Crash: Searching and Finding Bias in Quantum Random Number Generators, ACM Transactions on Privacy and Security. 23(3), 1-25.

Lin, X., Wang, S., Yin, Z.Q. (2020). Security analysis and improvement of source independent quantum random number generators with imperfect devices, Npj Quantum Information. 6(1), 100.

Kavulich, J., Van Deren, B., Schlosshauer, M. (2021). Searching for evidence of algorithmic randomness and incomputability in the output of quantum random number generators, Physics Letters; 2021. A (388), 127032.

Dutang, C., Wuertz. D. (2009). A note on random number generation, Overview of Random Generation Algoritms.

Gençoğlu, MT. (2021). Quantum cryptography, quantum communicatıon and quantum computing problems and solutions, Turkish Journal of Science and Technology. 16 (1), 97-101.

Gençoğlu, MT., Agarwal, P. (2021). Use of Quantum Differential Equations in Sonic Processes, Applied Mathematics and Nonlinear Science. 6(1), 21-8.

Gençoğlu, MT. (2013). Complex solutions for Burgers-Like equation, F.U. Türkish Journal of Science and Technology. 8(2), 121-123.

Bujang MA., Sapri, F. (2018). An Application of the Runs Test to Test for Randomness of Observations Obtained from a Clinical Survey in an Ordered Population, Malaysian Journal of Medical Sciences. 25, 146-151.

Yakut, S., Tuncer, T., Ozer, A. B. (2019). Secure and Efficient Hybrid Random Number Generator Based on Sponge Constructions for Cryptographic Applications. *Elektronika Ir Elektrotechnika*, *25*(4), 40-46. https://doi.org/10.5755/j01.eie.25.4.23969

Yakut, S., Tuncer, T., Ozer, A. B. (2020). A New Secure and Efficient Approach for TRNG and Its Post-Processing Algorithms, Journal of Circuits, Systems and Computers. 29:15.

Yakut, S. (2021). Random Number Generator Based on Discrete Cosine Transform Based Lossy Picture Compression. NATURENGS, 2(2), 76-85. https://doi.org/10.46572/naturengs.1009013

Yakut, S. (2022). Kayıplı Resim Sıkıştırma Algoritmalarını Temel Alan Rastgele Sayı Üreteci. Adıyaman Üniversitesi Mühendislik Bilimleri Dergisi, 9(18), 571-580. https://doi.org/10.54365/adyumbd.1145590

He, D., Huang, W., Chen, L., Chan, S. (2024). A Secure and Efficient Software Random Number Generator Applicable to the Internet of Things, IEEE *Internet of Things Journal*, 1-12. doi: 10.1109/JIOT.2024.3468451.

Santa Cruz, L.J.M., Faina, L.F., Souza Pereira, J.H. (2025). Exploring quantum systems for pseudo-random number generation. *Quantum Stud.: Math. Found.* **12**, 3. https://doi.org/10.1007/s40509-024-00348-1

Cirauqui, D., Ángel, M., Guillem, G.M., Corominas, G., Graß, T., Grzybowski, P.R., Muñoz-Gil, G., Saavedra,J.R.M., Lewenstein, M. (2024). Comparing pseudo- and quantum-random number generators with Monte Carlo simulations. APL Quantum, 1 (3): 036125. https://doi.org/10.1063/5.0199568

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

# An Efficient Steganography Method Based on Chaotic Functions and XOR Operation for Data Hiding

**Selman Yakut**[*1]

[1]*Inonu University, Faculty of Engineering, Software Engineering Department, Malatya, Türkiye*

**Keywords:**
Steganography
Chaotic functions
Logistic Map
Tent Map

**ABSTRACT**

The advancing technology and digitalizing world have increased the importance of secure data transmission. Steganography, a technique that ensures secure data communication, is a critical component of data security. Derived from the term meaning "hidden writing" in Turkish, steganography is based on the principle of embedding the data to be hidden into a carrier medium. While historically applied using primitive methods, steganography has transitioned to the use of modern techniques and methods in today's digitalized era. In this study, a steganography method based on chaotic functions and the XOR operation is proposed. The proposed method consists of two stages. In the first stage, data embedding, the data to be hidden is first converted into binary format. This binary data is then subjected to an XOR operation with a tent map sequence. The resulting final data is embedded into a grayscale image by determining its embedding positions using a logistic map. In the second stage, data extraction, the embedded message is retrieved using the logistic map, and the extracted message is XORed with the tent map to recover the original data. The effectiveness of the proposed method was evaluated using commonly employed metrics such as PSNR, MSE, and SSIM on images in the literature. The results demonstrate that the proposed method offers a robust structure against steganalysis techniques while ensuring critical security parameters.

# Veri Gizlemede Kaotik fonksiyonlar ve XOR İşlemi Tabanlı Etkili bir Steganografi Yöntemi

**Anahtar Kelimeler:**
Steganografi
Kaotik fonksiyonlar
Logistic harita
Tent harita

**ÖZ**

Gelişen teknoloji ve dijitalleşen dünya, güvenli veri iletiminin önemini artırmaktadır. Steganografi, verilerin güvenli bir şekilde iletilmesini ele alan ve veri güvenliğinin kritik bir parçasını oluşturan bir tekniktir. Türkçe'de "gizli yazı" anlamına gelen steganografi, gizlenmek istenen verinin taşıyıcı bir veri aracılığıyla aktarılması esasına dayanır. Tarihsel olarak ilkel yöntemlerle uygulanan steganografi, dijitalleşen dünya ile birlikte modern tekniklerin ve yöntemlerin kullanımına geçiş yapmıştır. Bu çalışmada, kaotik fonksiyonlar ve XOR işlemi tabanlı bir steganografi yöntemi önerilmektedir. Önerilen yöntem iki aşamadan oluşmaktadır. Birinci aşama olan veri gömme işleminde, gizlenecek veri önce ikilik formata dönüştürülür. Ardından bu veri, tent map dizisi ile XOR işlemine tabi tutulur. Bu işlem sonucunda elde edilen nihai veri, logistic map kullanılarak gri seviye bir görüntünün gömüleme pozisyonları belirlenerek yerleştirilir. İkinci aşama olan veri çıkarma işleminde, logistic map yardımıyla gömülü mesaj çıkarılır ve çıkarılan bu mesaj tent map ile XOR işlemine tabi tutularak orijinal veri elde edilir. Önerilen yöntemin etkinliği, literatürdeki görüntüler üzerinde gerçekleştirilen PSNR, MSE ve SSIM gibi metriklerle test edilmiştir. Sonuçlar, yöntemin steganaliz tekniklerine karşı dayanıklı bir yapı sunduğunu ve güvenlik parametrelerini sağladığını göstermiştir.

[*]**Corresponding Author**

[*](selman.yakut@inonu.edu.tr) ORCID ID 0000-0002-0649-1993

**e-ISSN: 2717-8579**

**Geliş Tarihi: 29/11/2024; Kabul Tarihi: 28/12/2024**

## 1. INTRODUCTION

Data security is a fundamental requirement in the advancing technological and digital world. Numerous algorithms, security systems, protocols, and similar approaches have been proposed to ensure data security. One of the critical components of data security is steganography, which translates to "hidden writing." Steganography involves embedding any message that needs to be transmitted into a carrier medium in a way that prevents it from being detected (Kipper, 2019). While historically performed using primitive methods, steganography now requires modern techniques and approaches due to technological advancements. The ever-growing volume of data and its transmission in today's digital era further emphasizes its importance.

In the literature, various steganography methods have been proposed depending on factors such as the type of carrier data and the method used (Kipper, 2019). The carrier medium and the type of transmitted message can include various formats such as video, images, text, or audio signals (Cheddad, Condell, Curran, & Mc Kevitt, 2010). Additionally, the techniques used may differ based on whether the data is manipulated in the spatial domain or frequency domain (Karakış, Gürkahraman, Çiğdem, Öztoprak, & Topaktaş, 2021). The Least Significant Bit (LSB) algorithm is widely used in various data types, particularly images and videos (Akyüz, 2021). To enhance the security of steganography algorithms and approaches, additional operations and functions are employed alongside LSB bits. One such function is chaotic functions, which exhibit chaotic properties based on specific parameters (Yakut, Tuncer, & Ozer, 2019). These functions are utilized in various fields to address complex problems (Özbay, 2023). In the literature, chaotic functions are applied in diverse ways to develop steganography methods. By employing chaotic functions, the selection of LSB bits for embedding data can lead to effective and secure steganography methods.

In this study, a new steganography method is proposed for secure data transmission. The proposed method combines chaotic functions and the XOR operation to provide a complex and secure steganographic structure. Initially, the data to be embedded is encrypted by XORing it with a chaotic sequence generated using the tent map. The encrypted data is then embedded into the LSB bits of pixels at specified positions in a grayscale image, with these positions determined using the logistic map. The proposed method ensures two levels of security by utilizing chaotic functions for both the encryption process and the determination of embedding positions in the grayscale image. Furthermore, tests conducted using analysis tools such as PSNR, SSIM, and MSE demonstrate the effectiveness of the proposed approach. Additionally, the random outputs and efficiency of the chaotic functions used further validate the effectiveness of the proposed method.

This study is organized as follows: Section 2 presents the related literature. Section 3 describes the proposed approach. Section 4 provides the experimental results of the proposed approach. Finally, Section 5 concludes the study.

## 2. LITERATURE REVIEW

In the literature, numerous approaches to steganography have been proposed based on parameters such as the type of carrier data, the data to be transmitted, and the methods employed. Among these, the use of images as carriers for transmitting secret messages is one of the most commonly utilized steganographic mediums. For such transmissions, one of the earliest and simplest approaches involves embedding data into the least significant bits (LSBs) of the carrier medium. Additionally, the use of frequency domain transformations for data embedding is also prevalent (Kipper, 2019). To determine the embedding positions and ensure the security of transmitted data, various approaches have been explored in the literature. Among these, chaotic functions are widely employed. Chaotic functions are frequently used in areas such as data hiding and encryption to enhance data security (Akyüz, 2021). Their properties of randomness and unpredictability make them highly effective in increasing security in steganographic applications. The use of chaotic functions in data hiding is well-documented in the literature.

Khalil et al. proposed a novel data hiding algorithm utilizing 1D and 2D chaotic maps combined with LSB techniques for concealing various types of data (images, text, audio) within cover images of different dimensions, achieving successful test results (Khalil, Sarhan, & Alshewimy, 2024). Pak et al. introduced an improved one-dimensional chaotic map, demonstrating superior performance compared to existing models, and aimed to enhance the robustness of the LSB steganography algorithm against attacks using this model (Pak et al., 2020). Tiwari et al. implemented data hiding in images by employing two chaotic systems: the first determined the pixel positions for data embedding, while the second set the initial conditions for the first chaotic system (Kumar Tiwari, Rajpoot, K. Shukla, & Karthikeyan, 2015). Nasr et al. combined chaotic Henon, Baker, and Arnold maps with audio steganography to propose a secure data hiding method for image encryption (Nasr et al., 2024).

Ghosh et al. combined a chaotic 2D classical map and a linear feedback shift register, demonstrating that this approach effectively enhances encryption security and plays a vital role in ensuring data privacy and security in medical image steganography within healthcare networks (Ghosh, Saha, Pal, & Jha, 2024). Alzubi et al.

introduced a novel image hiding formula based on a chaotic map system employing Tent map, Singer map, and Logistic map for pixel- and bit-level scrambling, which can be used in sensitive fields such as military and healthcare (Alzubi, Alzubi, Suseendran, & Akila, 2019). Kumar and Hussaini proposed an effective method combining an artificial neural network and a cyclic chaos algorithm to select the best cover image, enhancing visual quality, hiding capacity, and security (Kumar & Hussaini, 2021).

López Torres et al. proposed a crypto-steganography algorithm combining chaos, DNA coding, and edge-based techniques, achieving high similarity and low error between the original and stego images (López Torres, Alvarado-Nieto, Amaya-Barrera, & Parra, 2024). Durafe and Patidar presented a novel and effective color image steganography model, combining DNA-hyperchaotic encryption and DWT-SVD embedding techniques using unique fractal cover images, which can be applied across various fields (Durafe & Patidar, 2024). Karakış et al. demonstrated a method for medical image steganography that conceals data in non-tumor pixels using discrete wavelet transform and k-means clustering-based segmentation, preserving image fidelity while securely storing large patient data (Karakış et al., 2021).

Ranjithkumar et al. utilized chaotic maps to propose a video steganography method with three-layered security, embedding data into the spatial domain of cover video frames to ensure confidentiality (Ranjithkumar, Ganeshkumar, & Senthamilarasu, 2021). Nagarajegowda and Krishnan introduced an efficient video steganography method embedding audio or image-based secret data into a cover video, using a hybrid algorithm combining 2D-Henon and 3D-Logistic maps for encryption (Nagarajegowda & Krishnan, 2024).

Madhu et al. combined a dynamic 8-bit XOR algorithm with the AES encryption algorithm to securely store hidden data in images, achieving effective results (Madhu, Vasuhi, & Samydurai, 2024). Balkesen and Koçer proposed an approach that embeds AES-encrypted data into random bit positions of the cover image (Balkesen & Koçer, 2020).

The literature highlights the widespread use of XOR operations and chaotic functions in steganographic approaches. Their combined application facilitates the development of effective and secure data embedding techniques. Chaotic functions, known for their ability to generate unpredictable sequences, offer robust solutions in steganography. Additionally, XORing data with chaotic functions before embedding can provide dual-layer security.

## 3. PROPOSED METHOD

In this study, a steganography method utilizing Logistic Map and Tent Map chaotic functions in combination with the XOR operation is proposed for secure data transmission. The proposed method consists of two main components: embedding the secret message into the carrier data and extracting the embedded message from the carrier data. The chaotic functions, which form the core of the proposed method, are introduced first, followed by a detailed discussion of the data embedding and extraction processes.

### 3.1. Chaotic Functions

Chaotic functions, due to their sensitivity to initial conditions and their capacity to generate randomness, are widely used in security-critical applications such as encryption, steganography, and random number generation (Yakut et al., 2019; Yakut, Tuncer, & Özer, 2020). Although these functions have simple mathematical equations, they exhibit chaotic behavior under specific parameter values. For these parameter values, chaotic functions produce unpredictable and non-deterministic sequences.

These functions are highly sensitive to initial conditions in their chaotic parameter ranges. In other words, a small change in the initial value can lead to significant differences in the system's subsequent behavior. This sensitivity makes them highly suitable for applications such as random number generation and encryption, where unpredictability and randomness are essential.

#### 3.1.1. Logistic Map

The logistic map generates the next value x( n+1) based on the previous value $x(n)$ using the following equation:

$$x(n+1) = r * x(n) * \big(1 - x(n)\big) \qquad (1)$$

Here, $xn$: Represents the state at the n-th iteration, taking values between 0 and 1. $r$: A control parameter, typically chosen between 0 and 4, which determines the behavior of the logistic map.

The value of the control parameter $r$ dictates whether the logistic map exhibits chaotic behavior. For instance, values of $r \geq 3.57$ tend to trigger chaotic dynamics.

#### 3.1.2. Tent Map

The Tent Map is a mathematical example of a chaotic dynamical system commonly used in applications requiring randomness or chaos. It generates a chaotic sequence by iteratively updating a value within a specific range (e.g., [0,1]).

The Tent Map produces new values through iterations based on an initial value and a control parameter.

Mathematically, the Tent Map is defined by an $x$ value and a parameter $\mu$ as follows:

$$x(n+1) = \begin{cases} \mu * x(n), & x(n) < 0.5 \\ \mu * (1 - x(n)), & other \end{cases} \quad (2)$$

Here, $xn$: Represents the value at each iteration. $\mu$: A control parameter (typically set to 2.0) that influences the chaotic behavior of the system. Adjusting the value of $\mu$ directly affects the chaotic dynamics of the Tent Map.

### 3.2. Embedding Function

In the proposed method, the final state of the data to be embedded and its embedding positions are determined using chaotic functions and the XOR operation. The pseudocode for the embedding function is provided in Algorithm 1. Initially, the message to be hidden is converted into a byte array in UTF-8 format. It is then encrypted using the chaotic sequence generated by the Tent Map function through an XOR operation. This encryption enhances the security of the message and allows decryption during the extraction process using the same Tent Map sequence. Subsequently, the encrypted bytes are split into bits and embedded into random bit positions specified by the Logistic Map sequence. The Logistic Map determines the exact locations within the image where each bit will be embedded, enabling the message to be concealed by altering the specified bit positions in the image. This process ensures that the data is hidden within the image with no noticeable changes to its visual appearance.

**Algorithm 1.** Embedding function pseudocode

```
- Embeding Function
  Input: image_path, message (secret message)
  Output: Image with embedded message (stego_image)

  - Convert the message to a byte array
  - Open the image at image_path, and load into array
  - Generate chaotic sequences:
    - length = size of message in bytes * 8
    - logistic_seq = call logistic map function
    - tent_seq = call tent map function
  - Encrypt the message with XOR
  - Convert encrypted bytes to bits
  - Embed each bit of encrypted message in positions defined by Logistic map:
    - Create flat_img_data
    - Loop for i = 0 to length of encrypted_bits:
      - pos = i mod flat_img_data length
      - bit_pos = logistic_seq[i]
      - Clear bit at bit_pos in flat_img_data[pos] and replace it with encrypted_bits[i]
  - Reshape flat_img_data to original image shape and return stego_image
```

### 3.3. Extracting Function

The algorithm for extracting the hidden message in the proposed method is provided in Algorithm 2. During the extraction process, the values of the chaotic functions are recalculated. These values are then used to determine the bit positions. Using the same chaotic sequences and positions, the embedded bits are read from the image, and the original byte sequence is reconstructed. The extracted bit sequence is then decrypted by applying the XOR operation with the Tent Map sequence, recovering the original message. In this method, chaotic sequences not only ensure randomness during the embedding process but also act as keys for XOR encryption, enhancing security. By leveraging the randomness properties of chaotic systems, the message is securely encrypted and inconspicuously embedded within an image.

**Algorithm 2.** Extracting function pseudocode

```
- Extracting Function
  Input: Image with embedded message (stego_image)
  Output: Message (secret message)

- Load the stego image, and load into array
  - Generate chaotic sequences:
    - bit_length = message_length * 8
    - logistic_seq = call logistic map function
    - tent_seq = call tent map function
  - Extract encrypted bits from specified positions:
    - Create flat_img_data (flattened image data)
    - Initialize an empty extracted_bits array
    - Loop for i = 0 to bit_length:
      - pos = i mod flat_img_data length
      - bit_pos = logistic_seq[i]
      - Extract bit at bit_pos in flat_img_data[pos] and add it to extracted_bits[i]
  - Convert extracted_bits back to bytes
  - Decrypt the extracted data with XOR
  - Convert byte array to a UTF-8 string and return
return UTF-8 string (secret message)
```

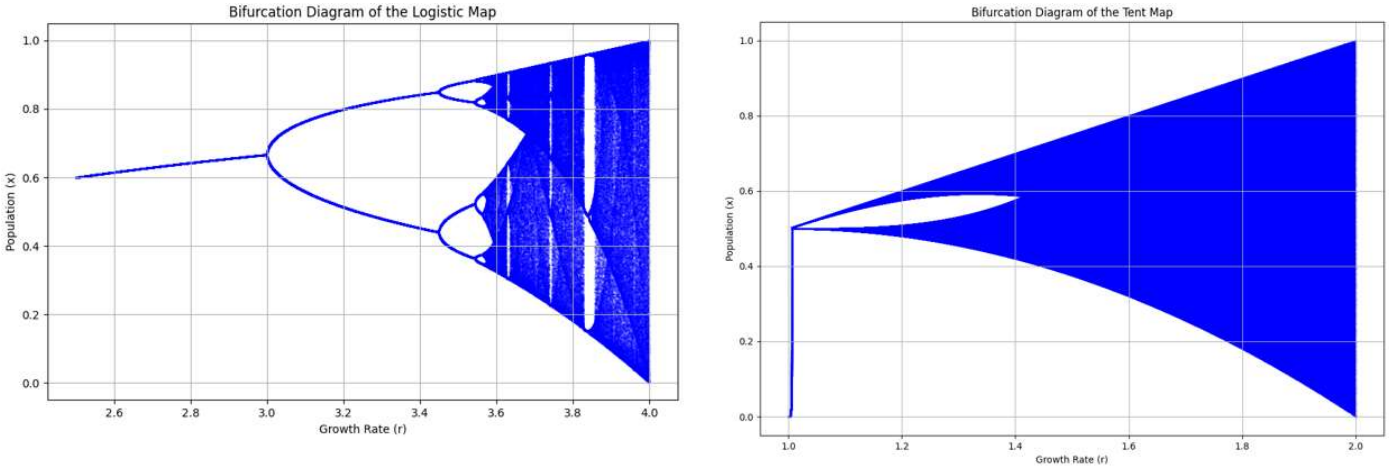
## 4. EXPERIMENTAL RESULS

Various parameters were used to assess the security and effectiveness of the proposed approach. These parameters primarily include robustness and imperceptibility. Robustness refers to the system's ability to retrieve the embedded data even when subjected to various attacks. Additionally, since chaotic functions are employed in the proposed approach, the randomness of these functions was also evaluated. Furthermore, the combined use of XOR encryption and chaotic bit selection provides a dual-layer security mechanism. XOR is a simple yet powerful operation commonly used in encryption and steganography due to its bit-level functionality, reversibility, and masking capabilities. Its ability to recover the original data when applied twice with the same key offers a significant advantage in security applications. Thus,

the proposed method ensures both data confidentiality and security. To demonstrate the effectiveness of the method, the message embedded in grayscale images was: "In this study, a steganography approach based on chaotic functions and the XOR method was proposed."

The chaotic functions used in the proposed approach, namely the Logistic Map and Tent Map, enhance the security of data embedding by leveraging their randomness properties. Chaotic functions such as the Logistic Map and Tent Map are widely utilized in data hiding and encryption to ensure data security. The unpredictability and randomness properties of chaotic maps are essential for enhancing security in steganographic applications. In the proposed approach, chaotic functions form the core of data masking and embedding processes. These functions were selected because of their simplicity and effectiveness. Parameter values were carefully chosen to ensure the chaotic behavior of the maps, as these maps exhibit extreme sensitivity to initial conditions within their chaotic parameter ranges.

In bifurcation diagrams, it can be observed that as the control parameter values of the Logistic and Tent Maps increase, their dynamic behaviors evolve. For both maps, the system remains stable at lower parameter values, with the population converging to a single fixed point. In the Logistic Map, as the parameter value increases, the system transitions to periodic oscillations, resulting in bifurcations that produce 2, 4, 8, and so on, periodic cycles. When the parameter exceeds 3.57, the system becomes entirely chaotic. Similarly, in the Tent Map, the system converges to a fixed point within the range of 0–1, exhibits periodic behavior within the range of 1–2, and demonstrates chaotic behavior for parameter values greater than 2.




(a) (b)

**Figure 1.** (a) Bifurcation diagram of the Logistic Map, (b) Bifurcation diagram of the Tent Map

### 4.1. Imperceptibility Anaysis

Imperceptibility refers to the inability to detect changes in the carrier image after data embedding. While these changes may sometimes be visually noticeable, they are often evaluated using various steganalysis methods. To measure the difference between the carrier image and the stego image containing the embedded data, the PSNR (Peak Signal-to-Noise Ratio) metric was used. PSNR quantifies pixel-level differences between two images and compares them against noise levels. A high PSNR value indicates that the embedded data is imperceptible. PSNR provides a numerical representation of the similarity between the original image and a distorted or reconstructed image. It is generally accepted that a PSNR value above 30 dB signifies imperceptibility of the embedded data. The relevant equation is provided in Equation (3). In the equation, $R$ denotes the maximum pixel value in the image.

$$PSNR = 10.\log_{10}(\frac{R^2}{MSE}) \quad (3)$$

MSE (Mean Squared Error) represents the average squared difference between corresponding pixels of two images. It is calculated as the mean of the squared differences between the pixel values of the original image and the reconstructed (or distorted) image. The formula for MSE is provided in Equation (4).

In the equation: $m$, $n$: Represent the dimensions of the image (height and width). $I(i, j)$: Denotes the pixel value at $(i, j)$ in the original image. $K(i, j)$: Denotes the pixel value at $(i, j)$ in the reconstructed (or distorted) image.

$$MSE = \frac{1}{m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}(I(i,j) - K(i,j))^2 \quad (4)$$

SSIM (Structural Similarity Index) is widely used to evaluate the performance of image processing techniques such as compression, image

restoration, noise reduction, and others. The operations for calculating SSIM are provided in Equation (5).

In the equation: $I(i, j)$: Represents the intensity of the pixel at $(i, j)$ in the original (carrier) image. $I'(i, j)$: Represents the intensity of the pixel at $(i, j)$ in the stego image (containing the hidden message). $M$, $N$: Denote the dimensions of the image, where $M$ is the number of rows and $N$ is the number of columns.

$$\text{SNR} = 10.\log_{10}\left(\frac{\sum_{i=1}^{N}\sum_{j=1}^{M} I(i,j)^2}{\sum_{i=1}^{N}\sum_{j=1}^{M}(I(i,j) - I(i,j))^2}\right) \quad (5)$$

The SPSNR, MSE, and SSIM test results for the proposed approach are presented in Table 1. The obtained results indicate that the differences between the two images are minimal, demonstrating that the steganography process has a negligible impact on both visual quality and structural similarity. These parameters serve as indicators of the effectiveness of the proposed method.

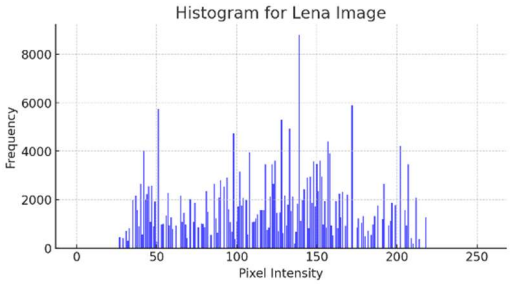
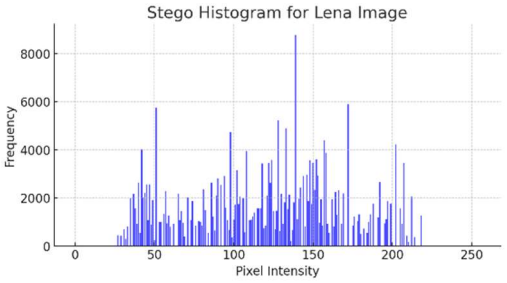
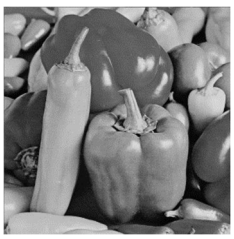
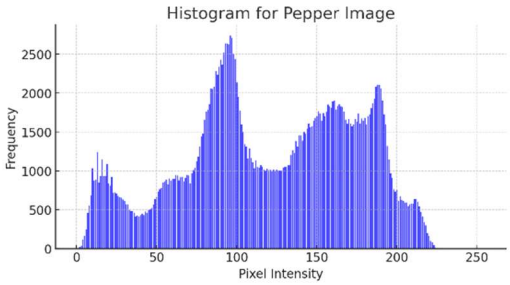
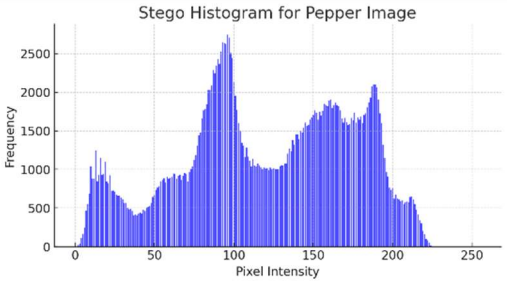
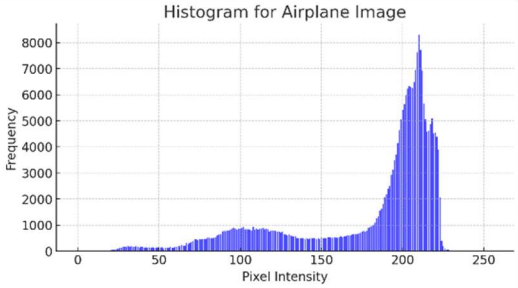
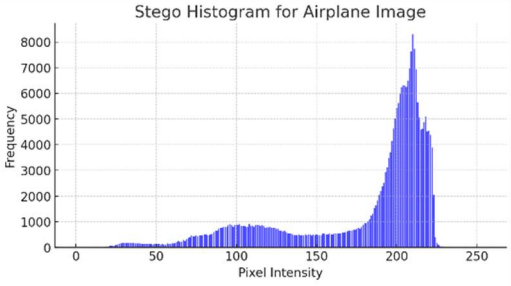
**Table 1.** SPSNR, SSIM, and MSE test results of the proposed method

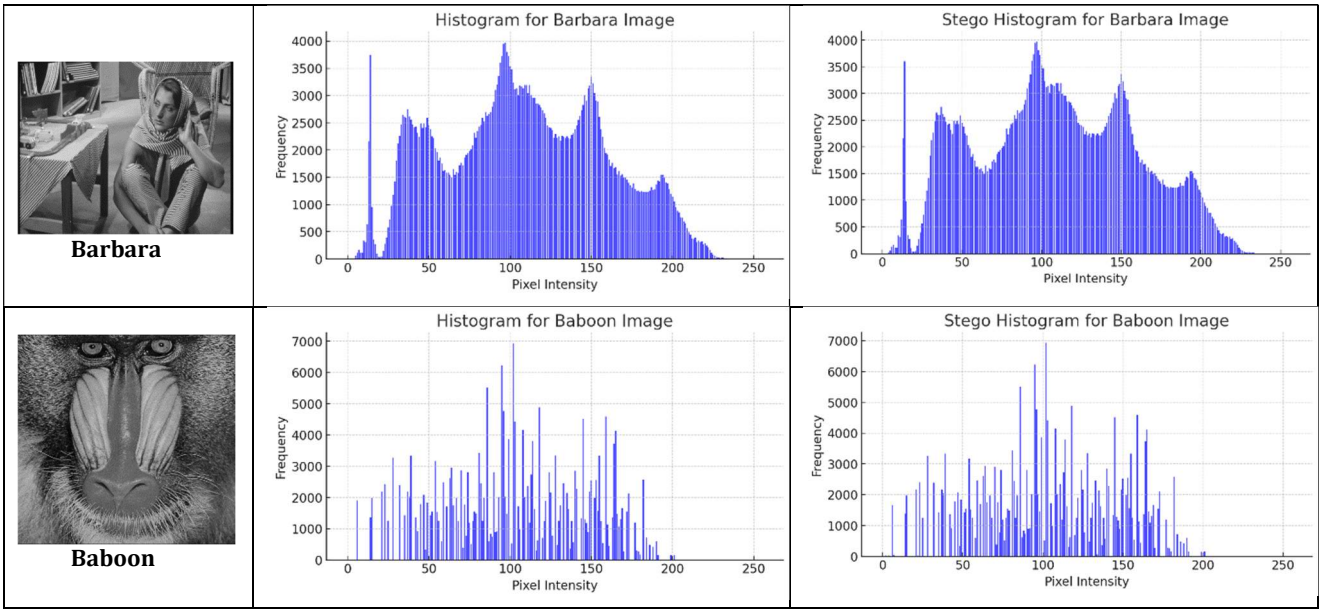
| | MSE | SSIM | PSNR |
|---|---|---|---|
| Lena | 0.0085 | 0.9979 | 68.86 dB |
| Paper | 0.010967 | 0.9990 | 67.73 dB |
| Airplane | 0.0084 | 0.9984 | 68.88 dB |
| Barbara | 0.0067 | 0.9994 | 69.88 dB |
| Baboon | 0.0096 | 0.9992 | 68.33 dB |

### 4.2. Histogram Analysis

Histogram analysis is a fundamental tool for evaluating the security and imperceptibility of steganography. Histograms assess the impact of the embedding process on visual quality and reveal differences between the original image and the stego image. The histogram results for the proposed method are presented in Table 2. In the proposed approach, histogram analysis examines changes in pixel intensity distributions when the message is embedded into the image (stego image). The similarity of the histograms indicates that the steganography process is imperceptible. Since there are no significant changes in the histogram, the embedding process remains undetectable.

**Table 2.** Histogram analysis results of the proposed method for commonly used ımages in the literature

| Orijinal resim | The histogram value of the original image | The histogram value of the stego image |
|---|---|---|
|  Lena |   |   |
|  Paper |   |   |
|  Airplane |   |   |




## 5. CONCLUSION

In this study, an effective data hiding algorithm was proposed by combining chaotic functions and XOR operations. In the proposed approach, the Logistic Map was used to determine the pixels where data would be embedded, while the Tent Map was employed to generate data for encrypting the original message. The original data was encrypted by XORing it with the output of the Tent Map chaotic function. By combining two chaotic functions with XOR operations, the algorithm provides a dual-layered security mechanism, preventing the extraction and decryption of the embedded data. The proposed algorithm was tested using commonly used images from the literature, with results evaluated through PSNR, MSE, and SSIM metrics. Additionally, the histogram analysis results for the same images were presented. The test results demonstrate the success of the proposed method on these images. Moreover, the randomness tests for the chaotic functions indicate that the algorithm prevents predictability. The tests conducted on grayscale images and the successful application results further validate the effectiveness of the proposed algorithm.

In future work, the proposed method is intended to be adapted for video steganography. The performance of the chaotic functions enables the algorithm to be applied to such data. Additionally, the application of the proposed method to data in the frequency domain could provide effective solutions for these approaches as well.

## REFERENCES

Akyüz, D. (2021). Yeni Kaotik Video Steganografi Metodu. İstanbul Ticaret Üniversitesi.

Alzubi, J. A., Alzubi, O. A., Suseendran, G., & Akila, D. (2019). +A Novel chaotic map encryption methodology for image cryptography and secret communication with steganography. International Journal of Recent Technology and Engineering, 8(1C2), 1122–1128.

Balkesen, C., & Koçer, H. E. (2020). Şifrelenmiş Verileri Rast Gele Piksel Yaklaşımı ile Bir Görüntüye Gömme. European Journal of Science and Technology, (September), 123–130. https://doi.org/10.31590/ejosat.802191

Cheddad, A., Condell, J., Curran, K., & Mc Kevitt, P. (2010). Digital image steganography: Survey and analysis of current methods. Signal Processing, 90(3), 727–752. https://doi.org/10.1016/j.sigpro.2009.08.010

Durafe, A., & Patidar, V. (2024). Image Steganography Using Fractal Cover and Combined Chaos-DNA Based Encryption. Annals of Data Science, 11(3), 855–885. https://doi.org/10.1007/s40745-022-00457-x

Ghosh, S., Saha, A., Pal, T., & Jha, A. K. (2024). A comparative analysis of chaos theory based medical image steganography to enhance data security. Procedia Computer Science, 235, 1024–1033. https://doi.org/10.1016/j.procs.2024.04.097

Karakış, R., Gürkahraman, K., Çiğdem, B., Öztoprak, I., & Topaktaş, A. S. (2021). Bölütlenen beyin bölgelerinin tıbbi görüntü steganografi için değerlendirilmesi. Gazi Üniversitesi Mühendislik-Mimarlık Fakültesi Dergisi, 36(4), 2301–2314. https://doi.org/10.17341/gazimmfd.753989

Khalil, N., Sarhan, A., & Alshewimy, M. A. M. (2024). A secure image steganography based on LSB technique and 2D chaotic maps. Computers and Electrical Engineering, 119(PB), 109566. https://doi.org/10.1016/j.compeleceng.2024.109566

Kipper, G. (2019). Investigator's Guide to Steganography. New York.

Kumar, M., & Hussaini, T. (2021). A Neural Network Based Image Steganography Method using Cyclic Chaos and Integer Wavelet Transform. 2021 Asian Conference on Innovation in Technology, ASIANCON 2021, 1–6. https://doi.org/10.1109/ASIANCON51346.2021.9544831

KumarTiwari, A., Rajpoot, A., K. Shukla, K., & Karthikeyan, S. (2015). A Robust Method for Image Steganography based on Chaos Theory. International Journal of Computer Applications, 113(4), 35–41. https://doi.org/10.5120/19817-1637

López Torres, E. A., Alvarado-Nieto, D., Amaya-Barrera, I., & Parra, C. A. S. (2024). Crypto-steganographic model using chaos and coding based in deoxyribonucleic acid. International Journal of Electrical and Computer Engineering, 14(4), 4239–4247. https://doi.org/10.11591/ijece.v14i4.pp4239-4247

Madhu, D., Vasuhi, S., & Samydurai, A. (2024). Dynamic 8-bit XOR algorithm with AES crypto algorithm for image steganography. Signal, Image and Video Processing, 18(Suppl 1), 429–445. https://doi.org/10.1007/s11760-024-03165-6

Nagarajegowda, S., & Krishnan, K. (2024). An adaptive approach for multi-media steganography using improved chaotic map and discrete cosine transform. Signal, Image and Video Processing, 18(10), 6695–6711. https://doi.org/10.1007/s11760-024-03345-4

Nasr, M. A., El-Shafai, W., El-Rabaie, E. S. M., El-Fishawy, A. S., El-Hoseny, H. M., Abd El-Samie, F. E., & Abdel-Salam, N. (2024). A robust audio steganography technique based on image encryption using different chaotic maps. Scientific Reports, 14(1), 22054. https://doi.org/10.1038/s41598-024-70940-3

Özbay, F. A. (2023). A modified seahorse optimization algorithm based on chaotic maps for solving global optimization and engineering problems. Engineering Science and Technology, an International Journal, 41, 101408. https://doi.org/10.1016/j.jestch.2023.101408

Pak, C., Kim, J., An, K., Kim, C., Kim, K., & Pak, C. (2020). A novel color image LSB steganography using improved 1D chaotic map. Multimedia Tools and Applications, 79(1–2), 1409–1425. https://doi.org/10.1007/s11042-019-08103-0

Ranjithkumar, R., Ganeshkumar, D., & Senthamilarasu, S. (2021). Efficient and secure data hiding in video sequence with three layer security: an approach using chaos. Multimedia Tools and Applications, 80(9), 13865–13878. https://doi.org/10.1007/s11042-020-10324-7

Yakut, S., Tuncer, T., & Ozer, A. B. (2019). Secure and efficient hybrid random number generator based on sponge constructionsfor cryptographic applications. Elektronika Ir Elektrotechnika, 25(4), 40–46. https://doi.org/10.5755/j01.eie.25.4.23969

Yakut, S., Tuncer, T., & Özer, A. B. (2020). A New Secure and Efficient Approach for TRNG and Its Post-Processing Algorithms. Journal of Circuits, Systems and Computers, 29(15). https://doi.org/10.1142/S0218126620502448

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

# Investigation of the Status of Artificial Intelligence Courses in Medical Education Curriculum in Turkey

**Kerem Gencer*[1], Gülcan Gencer[2]**

[1]*Afyon Kocatepe University, Faculty of Engineering, Department of Computer Engineering, Afyonkarahisar, Turkey*
[2]*Afyonkarahisar Health Sciences University, Faculty of Medicine, Department of Biostatistics and Medical Informatic, Afyonkarahisar, Turkey*

**Keywords:**
Artificial Intelligence
Medical education
Curriculum
Deep learning
Machine learning

**ABSTRACT**

Artificial Intelligence (AI) is a rapidly advancing technology with significant impacts across various sectors. Alongside advancements in healthcare, medical education is also evolving under the influence of AI. This transformation is driving major changes in the healthcare sector by improving clinical decision-making processes through increased data utilization and the support of drug-machine interactions. The aim of this study is to examine the current state of AI courses in medical education in Turkey, compare the curricula of private and public universities, and evaluate the integration of AI into medical education.The curricula of 112 universities providing medical education in Turkey were analyzed through their official websites, focusing on courses related to AI in healthcare, computer-assisted courses, and programming languages. It was observed that AI courses in healthcare have been recently incorporated into university curricula and have significant potential for further development. These courses are primarily theoretical, with practical components available only in a few universities. Additionally, AI courses are more prevalent in the curricula of public universities compared to private ones. The study concludes that AI courses should hold a more prominent place in medical education and include more practical applications. While public universities have taken greater strides in this area, there is still room for improvement. In conclusion, AI is becoming an integral part of medical education, and healthcare professionals' knowledge in this field will play a critical role in improving future healthcare services.

# Türkiye'de Tıp Eğitimi Müfredatlarında Yapay Zeka Derslerinin Durumunun Araştırılması

**Anahtar Kelimeler:**
Yapay Zeka
Tıp eğitimi
Müfredat
Derin öğrenme
Makine öğrenmesi

**ÖZ**

Yapay Zeka (YZ), çeşitli sektörlerde önemli etkileri olan, hızla ilerleyen bir teknolojidir. Sağlık hizmetlerindeki ilerlemelerle birlikte tıp eğitimi de yapay zekanın etkisi altında gelişiyor. Bu dönüşüm, artan veri kullanımı ve ilaç-makine etkileşimlerinin desteklenmesi yoluyla klinik karar alma sürecini geliştirerek sağlık sektöründe önemli değişikliklere yol açmaktadır. Bu çalışmanın amacı Türkiye'de tıp eğitiminde yapay zeka derslerinin mevcut durumunu incelemek, özel ve devlet üniversitelerinin müfredatlarını karşılaştırmak ve yapay zekanın tıp eğitimine entegrasyonunu değerlendirmektir. Türkiye'de tıp eğitimi veren 112 üniversitenin müfredatları resmi internet siteleri üzerinden incelenerek sağlıkta yapay zeka ile ilgili dersler, bilgisayar destekli dersler ve programlama dilleri ele alındı. Türkiye'de sağlık hizmetlerinde yapay zeka derslerinin yakın zamanda üniversite müfredatına dahil edildiği ve daha da geliştirilmeye açık olduğu gözlemlendi. Bu dersler öncelikle teoriktir ve uygulamalı dersler yalnızca birkaç üniversitede mevcuttur. Ayrıca devlet üniversitelerinin müfredatlarında yapay zeka dersleri özel üniversitelere göre daha yaygındır. Tıp eğitiminde yapay zeka derslerinin daha önemli bir yere sahip olması ve daha pratik uygulamalar içermesi gerektiği sonucuna varılmıştır. Devlet üniversiteleri bu konuda daha fazla adım atmış olsa da hâlâ geliştirilecek noktalar var. Sonuç olarak yapay zeka tıp eğitiminin ayrılmaz bir parçası haline geliyor ve sağlık profesyonellerinin bu alandaki bilgisi gelecekteki sağlık hizmetlerinin iyileştirilmesinde kritik bir rol oynayacak.

***Corresponding Author**

*(keremgencer09@hotmail.com) ORCID ID 0000-0002-2914-1056
(gencergulcan@gmail.com) ORCID ID 0000-0002-3543-041X

**e-ISSN: 2717-8579**

**Geliş Tarihi: 22/07/2024; Kabul Tarihi: 28/12/2024**

## 1. INTRODUCTION

Artificial intelligence is the ability of a machine to simulate cognitive processes like speech recognition, image identification, and captioning (Nilsson, 1998). To put it simply, AI models are employed to identify patterns in vast amounts of data in order to generate extremely precise predictions for a variety of activities. More digital and dynamic opportunities are provided to pupils. Old textbooks and the predetermined atmosphere of the classroom may not always provide these opportunities (Imran and Jawaid, 2020).

A rapidly expanding phenomena, artificial intelligence (AI) will soon have an impact on numerous industries, including medical education. AI has improved quickly due to the exponential growth in data and processing power (Lee, Wu, Li, and Kulasegaram, 2021). AI boosts learning capacity and offers a decision support system at scales that are revolutionizing the future of healthcare. According to (Noorbakhsh-Sabet, Zand, Zhang, and Abedi, 2019), this technology is employed in the diagnosis and prognosis of diseases, therapy optimization and outcome prediction, medication development, and public health. In the past ten years, the use of artificial intelligence has helped to partially address several problems in education, including language processing, reasoning, planning, and cognitive modeling.

(Kolachalama and Garg, 2018) emphasize the increasing recognition of the potential value that machine learning can bring to the medical community. If this trend continues, in the coming years, we may see numerous AI-focused products and technologies integrated into the healthcare ecosystem. In this scenario, it becomes essential to determine whether a medical professional is willing to adopt these tools as part of their repertoire and, if so, how they can receive training in the art and science of machine learning algorithms. Medical schools should start creating curriculum time for machine learning and begin recognizing the changes in healthcare by accepting machine learning.

The integration of artificial intelligence (AI) into medical education curricula is supported for several compelling reasons. AI is transforming the healthcare sector by enabling faster and more accurate medical diagnoses, improving patient treatment, and optimizing the management of healthcare services. This transformation underscores the need for medical students to understand and effectively apply these technologies. As highlighted by Ejaz et al. (2022), AI is driving significant changes in clinical decision-making processes, a shift that is increasingly reflected in medical education.

One critical aspect of AI in medical education is its role in data analysis. AI tools and techniques provide medical students with the ability to analyze large datasets and extract meaningful insights, which are essential for advancing medical research and understanding diseases (Alexandru, Radu, and Bizon, 2018). Tolentino et al. (2024) emphasize that training in AI equips students with the skills to draw meaningful inferences from complex data, a crucial competency for future healthcare professionals.

AI also has the potential to enhance communication with patients. By using AI-supported tools, medical students can improve their ability to communicate effectively, making interactions more efficient and empathetic. However, the integration of AI into healthcare also brings challenges such as security and ethical concerns. Medical students need training to address these issues and ensure patient privacy is protected, as noted by Gupta and Sao (2011).

Moreover, AI serves as the foundation for future medical applications, including telemedicine and personalized medicine. Zhang et al. (2024) stress that medical students must be equipped with AI knowledge to specialize in these areas and adapt to emerging healthcare technologies.

Incorporating AI into medical education is not only essential for preparing students for their future careers but also for ensuring that patients receive better care and the healthcare system operates more efficiently. Comparing AI-related courses in Turkish medical faculties, both private and public, highlights several benefits for students and healthcare professionals.

First, the quality of education varies between institutions, with differences in course content, resources, and teaching methods. Comparing these elements helps students identify universities that offer more comprehensive and up-to-date AI education. Second, some universities provide opportunities for students to develop AI projects or gain practical experience through internships, enabling them to apply theoretical knowledge in real-world settings. Third, collaborations and research initiatives between universities enhance the depth of knowledge students gain in AI. Access to research projects and laboratory facilities enriches their learning experience. Finally, knowledge of AI can significantly benefit students in their future careers. Proficiency in AI, regardless of their university, makes students more competitive and better prepared to excel as healthcare professionals in an increasingly technology-driven field. In conclusion, comparing AI courses in healthcare between private and state universities helps provide better educational opportunities and future career prospects. Therefore, it is essential to carefully assess education programs, opportunities, and costs when making decisions (Cleophas and Zwinderman, 2015; Topol, 2019).

### 1.1. Artificial intelligence and medical education

Due to its capacity to facilitate learning, artificial intelligence (AI) is mostly used in medical education to deliver individualized feedback. The assessment of

students' learning has received less attention, in part because of the lack of digitization and the delicate nature of exams. Data integrity must also be protected when manipulating massive data. The technical difficulties of developing AI applications and the need for new approaches to evaluating AI's efficacy must be addressed if methodological advancements are to promote the acceptance of AI (Chan and Zary, 2019). This is especially true given that AI has the ability to provide individualized learning. It should be remembered that producing capable and skilled doctors is the ultimate goal to comprehend AI's potential in medical education better. The dilemma of whether we should strive to modify teachers or target learning emerges when contemplating the learning capacity of AI, just as any improvements in medical education should be driven by this ultimate goal (Masters, 2019).

### 1.2. Curriculum and artificial intelligence

The evaluation of curricula is a complex and administrative procedure that strongly supports the need for automation. It's interesting that not many of the studies we evaluated discussed the use of AI in reviewing medical curricula. There isn't much usage of AI in curriculum evaluation, despite any possible advantages it might offer over conventional approaches (C.-K. Chen, 2010).

### 1.3. Artificial intelligence and clinical applications

The use of AI in medicine has been covered by (Noorbakhsh-Sabet et al., 2019) the importance of machine learning for clinical applications, translation, and public health has been the authors' primary attention. They have suggested that AI may be implemented in medical education at three levels: curriculum development and analysis, learning and assessment, and curriculum evaluation, as it is crucial for the future growth of medicine and healthcare services.

### 1.4. Artificial intelligence and learning support

As it is crucial for the education of future doctors, (Garg, 2020) has recommended using AI to control the effectiveness of the curriculum and the general happiness of medical students. With the use of AI, instructional materials can be tailored and adaptive for students, and they can be further developed with their input. It enables students to recognize their knowledge gaps and take appropriate action to fill them, making learning assessments more thorough, efficient, and objective.

### 1.5. Artificial intelligence and data privacy

On the other side, due to technical improvements, much data must be collected and shared, raising privacy issues. Healthcare machine-learning applications require a broad assessment of the significance of privacy in clinical, translational, and public health applications, focusing on data sharing and genetic information (Noorbakhsh-Sabet et al., 2019).

### 1.6. Artificial intelligence and telemedicine

In particular, the Virtual Inquiry System is an online virtual patient system that uses AI technology to educate hospitals, medical faculties, and interns. The system aggregates many actual patient records as well as expert and AI-compiled particular instances. Medical students make diagnoses through questions, simulated physical examinations, and additional studies of virtual patients. Medical remote education is a form of instruction that is conducted both online and offline in real-time and is not location- or time-bound. Web-based teaching techniques like Twitter enable learning, communication, and sharing. In particular, virtual hospital tours and mobile nursing in clinical application teaching play a key role in medical education (Zhao, Li, and Feng, 2018).

### 1.7. Artificial intelligence and curriculum analysis

Modern information technology, such as data centers, instructional resource libraries, cloud platforms, student recruitment, educational process administration, and evaluation, is the foundation for the influence of artificial intelligence technology. This increases the administration of continuous medical education's efficacy and service quality. Through computer systems, common elements in the processes of student recruiting, announcement, acceptance, instruction by instructors, and course setup can be exchanged. Through information technologies (web pages, mobile phones), essential hospitals, health administrative departments, clinical instructors, and departments can communicate the same information. Synchronizing lessons through computer information systems can facilitate data exchange, information sharing, and course collaboration (Imran and Jawaid, 2020). It is essential to develop medical education curricula to include artificial intelligence (Paranjape, Schinkel, Nannan Panday, Car, and Nanayakkara, 2019).

### 1.8. Artificial intelligence in medical applications

Numerous studies have been conducted on artificial intelligence in the field of medicine. These include the detection of lung nodules from lung X-rays (Deo, 2015), risk prediction models for anticoagulant therapy, automatic defibrillator implantation in cardiomyopathy (Lip, Nieuwlaat, Pisters, Lane, and Crijns, 2010), stroke and stroke mimicry (O'Mahony et al., 2013), modeling CD4+ T-cell heterogeneity, outcome prediction in infectious diseases (Lu et al., 2015), arrhythmia detection in electrocardiography (Y. Chen et al., 2018), and cancer

susceptibility detection, including histological or pathological assessments, in the development of high-throughput technologies such as genomics, proteomics, and imaging (Cruz, 2007; Kourou, Exarchos, Exarchos, Karamouzis, and Fotiadis, 2015).

In general, it can be observed that artificial intelligence is being more actively used in medical education and is becoming widely integrated into curricula. However, there is significant heterogeneity and weak consensus among studies regarding the content and presentation of artificial intelligence curricula. To address these inconsistencies and facilitate the broader adoption and implementation of standardized artificial intelligence curricula in medical education, further research is needed to establish a standard competency framework (Lee et al., 2021).

## 2. MATERIAL AND METHODS

This study aimed to evaluate the curricula of state and private universities providing medical education in Turkey, utilizing an observational descriptive research approach. The research was conducted by collecting and analyzing curriculum data through the official websites of universities. The data focused on the presence and distribution of courses related to artificial intelligence, computer-assisted courses, and programming languages within medical education curricula. Among the universities, 81 were state universities, and 31 were private universities. The examined data were collected from universities' course programs, websites, and information packages. The data includes the names of courses in medical education curricula, their distribution by semester/class, course hours, credits, and ECTS (European Credit Transfer and Accumulation System) information. Information on introductory information technology/computer courses offered as compulsory or elective in any university class was excluded. The data analysis process for this study was conducted using R version 4.0.2. Descriptive statistics were employed to calculate the frequencies and proportions of AI-related courses offered in state and private universities, providing a clear overview of the prevalence of these courses within the medical education curricula. A comparative analysis was also performed to identify differences in the availability and structure of AI-related courses between state and private universities, highlighting variations in their integration into the educational framework. Additionally, the distribution of AI-related courses across different semesters was examined to understand how these courses are positioned within the overall curriculum. The tidyverse package in R was utilized for efficient data cleaning and manipulation, ensuring the reliability and accuracy of the analysis.

## 3. RESULTS

This study provides a detailed examination of the curricula of private and state universities offering medical education in Turkey. The investigation focused on courses such as Artificial Intelligence in Healthcare, Computer-Aided Courses, and Programming Languages within the provided curricula. The curricula of a total of 112 universities in Turkey offering medical education were researched via their official websites and course programs. Among these, 81 were state universities, and 31 were private universities. Figure 1 illustrates the number of universities with a direct "Artificial Intelligence in Medicine/Healthcare" course in their medical education curriculum and the average course hours. There were 11 state universities, 9.8%, and six private universities, 5.3%. Notably, these courses were offered electives in state and private universities (Figure 1).

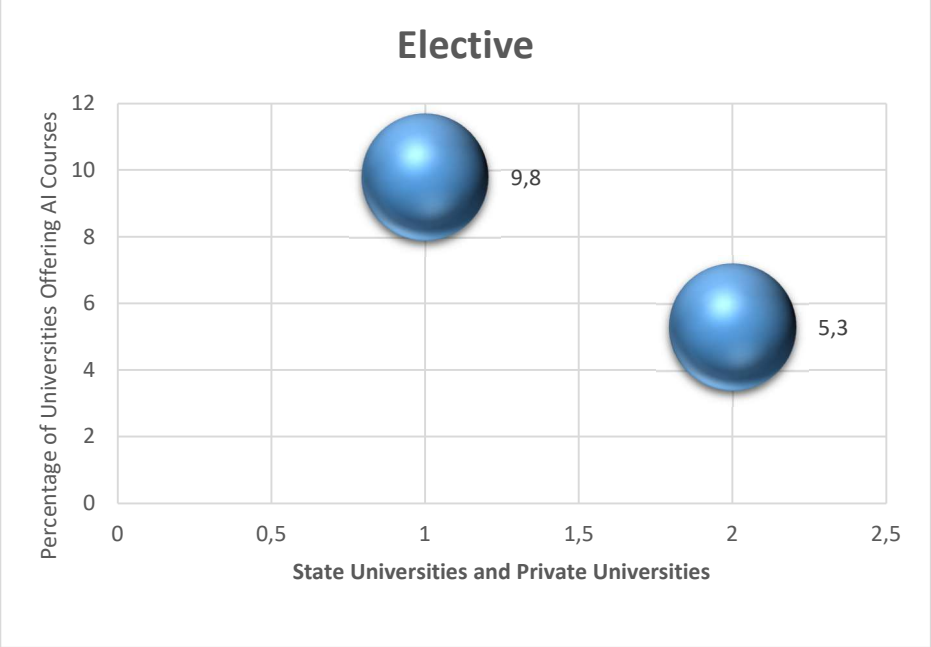



**Figure 1.** Number of Universities Offering a Course with the title "Artificial Intelligence in Medicine/Healthcare" in the Medical Education Curriculum.

Table 1 shows the number of universities offering courses related to computers and technology in the medical education curriculum, even though they do not have the title "Artificial Intelligence," All of these courses are elective. These courses are taught with various labels such as Image Processing Analysis, Computer-Aided Design, Information and Technology Management, Computer Applications, Web Design, Medical Informatics, Bioinformatics, Web Technologies, Robotic Medicine, Network Fundamentals, Medical Computing and Applications, and the Use of Computers in Medicine. It can be observed that State universities are leading in this aspect with 23 universities (20.5%), while private universities have two universities (0.02%).

**Table 1.** Distribution of Universities Offering Elective Computer and Technology Courses in Medical Education Curricula.

| Compulsory/ Elective | Some course names in the curriculum | Public University | | Private University | |
|---|---|---|---|---|---|
| | | n | Mean | n | Mean |
| Elective | Image processing analysis, Computer-aided design, Information and technology management, Computer applications, Web design, Medical informatics, Bioinformatics, Web technologies, Robotic medicine, Network fundamentals, Medical computers and applications, Computer use in medicine | 23 | % 20.5 | 2 | % 0.02 |

* Mean (%) values indicate the proportion of universities (state or private) offering the listed courses as part of their medical education curricula.

Table 2 reveals that when universities were examined, it was determined that there were courses covering different programming languages. In state universities, it was found that 6.2% of them offer programming language courses. However, in the case of private universities, no programming language courses were identified during the survey.

**Table 2.** Distribution of Universities Offering Elective Programming Language Courses in Medical Education Curricula.

| Compulsory/ **Elective** | **Course Names** | **Public University** | | **Private University** | |
|---|---|---|---|---|---|
| | | **n** | **Mean** | **n** | **Mean** |
| Elective | Introduction to Python Programming Language, Linux Operating System, Introduction to Python Programming/Data Science, Programming with Matlab, Robotic Coding, Python coding, Programming languages, software, Technology and Software | 7 | % 6.2 | - | - |

*Mean (%) values indicate the proportion of universities (state or private) providing elective programming-related courses in the curricula

Table 3 provides information on the availability of artificial intelligence courses in medical education curricula, specifically focusing on whether these courses are offered during the pre-clinical or clinical phases of education. The data reveals that artificial intelligence courses are not offered during the clinical phase in either state or private universities. Generally, these courses are provided in state universities during Semester 2, while private universities provide them during Semesters 1, 2, and 3.

**Table 3.** Semester-Wise Distribution of Artificial Intelligence Courses in Medical Education Curricula.

| Compulsory/ **Elective** | **Public University** | | | | **Private University** | | |
|---|---|---|---|---|---|---|---|
| | **Sem.1 (n-Mean)** | **Sem. 2 (n-Mean)** | **Sem. 1-2 (n-Mean)** | **Sem. 1-2-3 (n-Mean)** | **Sem. 1 (n-Mean)** | **Sem. 2-3 (n-Mean)** | **Sem. 1-2-3 (n-Mean)** |
| Artificial intelligence/artificial intelligence in medicine/health | (1-0.012) | (4-0,.49) | (3-0.037) | (2-0.024) | (1-0.032) | (2-0.064) | (3-0.096) |

*Mean (%) values indicate the proportion of universities (state or private) providing artificial intelligence-related courses in the specified semesters.

Table 4 examines artificial intelligence in healthcare courses in the medical education curriculum regarding course hours, credits, and ECTS (European Credit Transfer and Accumulation System) credits. In public universities, these courses generally have one theoretical hour, one credit, and 1 ECTS credit per week. In private universities, however, these courses typically have two academic hours, two credits, and 2 ECTS credits per week. Practical methods are available in public universities but not in private universities. Additionally, it is noted that in private universities, the theoretical hours, credits, and ECTS credits allocated to these courses are relatively higher (e.g., 3 ECTS, 5 ECTS, 8 ECTS...).

**Table 4.** Comparison of Course Hours, Credits, and ECTS for Artificial Intelligence Courses in Medical Education Curricula.

| Compulsory/ **Elective** | **Public University** |
|---|---|

| | Theoric course hour | | | Pr. Cour. Hour | Credit | | | ECTS | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | Oth | 2 | 1 Credit | 2 Credit | Oth | 1 | 2 | Oth |
| Artificial intelligence/artificial intelligence in medicine/health | 0.061 | 0.024 | 0.037 | 0.024 | 7 | 2 | 2 | 6 | - | 5 |
| | **Private University** | | | | | | | | | |
| | **Theoric course hour** | | | **Pr. Cour. Hour** | **Credit** | | | **ECTS** | | |
| | **1** | **2** | **3** | **2** | **1 Credit** | **2 Credit** | **Oth** | **1** | **2** | **Oth** |
| | 0.064 | 0.096 | 0.032 | - | - | 3 | 3 | - | 2 | 4 |

*Oth: Other, Mean values indicate the average course hours (theoretical and practical), credits, and ECTS credits for artificial intelligence-related courses in public and private universities.

The aims and learning outcomes of artificial intelligence in health courses at public and private universities in Turkey were investigated. It was observed that the curricula generally had similar dreams and learning objectives. Table 5 presents the aims and learning outcomes of general artificial intelligence in health courses taken from the medical education curriculum as an example.

**Table 5.** Examples of Aims and Learning Outcomes for Artificial Intelligence Courses in Medical Education Curricula.

| Section | Details |
|---|---|
| **Aim** | To recognize artificial intelligence (AI) technologies in healthcare for diagnosing diseases, improving treatment outcomes, and reducing costs. The course provides foundational knowledge on AI principles, applications in healthcare, privacy considerations, and programming techniques for developing AI applications with medical images and data. |
| **Objectives** | Students will learn to select and apply modern methods of AI in healthcare. |
| **Learning Outcomes** | |
| **Knowledge** | 1. Understand the fundamentals of programming.<br>2. Learn the basic principles of artificial intelligence.<br>3. Comprehend the impact of AI in medicine.<br>4. Familiarize with examples of AI applications in healthcare.<br>5. Use programming to create AI applications with medical images and data. |
| **Skills** | 1. Develop AI applications using medical images and data through programming. |
| **Attitude** | 1. Recognize the significance of artificial intelligence applications in medicine. |

## 4. LIMITATIONS

When collecting curriculum information through university websites, there is a risk of encountering incomplete or inaccurate information due to some university websites being outdated or lacking essential details. There may have been changes in curricula or course programs during your research process. Such changes could impact the results of your study. Limitation of Research Method: Observational studies aim to describe observed situations rather than establish cause-and-effect relationships. Therefore, this study seeks to define the current status of artificial intelligence courses in medical education rather than directly assessing their effects.

## 5. RECOMMENDATIONS

Based on the reasons discussed in the article and the current situation, we propose a series of recommendations aimed at enhancing the integration of artificial intelligence (AI) into medical education curricula. In the short term, AI courses should be made compulsory across public and private universities to raise awareness about its applications and benefits in healthcare. Additionally, Biostatistics and Medical Informatics courses should be updated to include topics like data analysis, AI, and machine learning, supplemented with sessions led by experts to provide practical knowledge. Competitions and hackathons can be organized to foster creativity and problem-solving skills, while educational resources such as books, videos, and software tools should be developed and distributed to support learning.

For the medium term, hands-on training programs and projects should be introduced into medical school curricula to help students apply theoretical AI knowledge in practical scenarios. Establishing bioinformatics departments with compulsory AI-focused courses during pre-clinical and clinical stages of education is essential. Workshops and conferences for interns and specialty students can be used to teach clinical applications of AI, such as diagnostic tools and patient management systems. Collaborating with healthcare companies to provide internship opportunities and real-world project experience, as well as developing AI-driven virtual patient simulations, will create immersive and practical learning environments.

In the long term, professional development programs should be created to update existing healthcare professionals on advancements in AI and machine learning. Universities should establish dedicated research centers to support AI research and innovation in healthcare. Faculty specializing in AI should be recruited and trained to ensure high-quality teaching in medical education. Postgraduate AI training programs can enhance the career prospects of healthcare professionals, and mechanisms should be developed to regularly revise AI-related curricula, ensuring alignment with emerging technologies and trends.

To address challenges in implementing these recommendations, practical solutions have been proposed. Regional AI hubs equipped with the necessary infrastructure and software can bridge technological gaps, supported by government grants and public-private partnerships. National AI educator training programs, coupled with scholarships and grants, can address the shortage of qualified educators. Streamlined academic approval processes, including fast-track committees and pilot programs, can accelerate curriculum changes. Resource shortages can be mitigated by creating centralized repositories of open-access AI educational materials, collaborating with global platforms like Coursera and edX. Flexible, online postgraduate training programs can accommodate healthcare professionals' schedules, while university liaison offices can expedite industry collaborations.

To enhance student participation, strategies such as gamification, offering certifications, and organizing AI-themed competitions and hackathons have been recommended. Reliable evaluation systems can be designed using AI-based adaptive testing tools, and advisory panels can be established to ensure continuous curriculum updates in line with rapid advancements in AI. Finally, cost-sharing initiatives among universities, along with government and private sector funding, can address the high implementation costs.

These recommendations are designed to strengthen medical education in Turkey, align it with global advancements in artificial intelligence, and prepare healthcare professionals for an AI-driven future. By addressing challenges with practical solutions, this framework provides a comprehensive roadmap for educators and policymakers.

## 6. CONCLUSION AND DISCUSSION

In recent years, health-related artificial intelligence courses have begun to be taught in Turkish classrooms. This rate is more significant in public universities than in private universities. In the future, new positions for medical students and doctors based on their proficiency with artificial intelligence will become available due to technological advancements in medical education. Medical education must engage with technology-enhanced learning and artificial intelligence in addition to its conventional biomedical and clinical sciences concentration. It is time for medical institutions to update their curricula to include material on artificial intelligence and machine learning, as well as a focus on empathy and honesty. This will guarantee that graduates use these AI technologies and are prepared to operate in the healthcare setting that AI has revolutionized. The purpose of education is to gather knowledge and leave a legacy by passing it on to succeeding generations and inspiring them to create using educational resources.When the literature is studied, it becomes clear that artificial intelligence is being employed more actively and extensively across the curriculum in medical education on a global scale. However, there is a considerable variety and a lack of agreement among research on the delivery and content of AI curricula. Although there is a lot of research on the subject, there is little agreement on what and how to teach AI in medical schools. More study is required to overcome these discrepancies and provide a consistent competency framework that can promote increased adoption and implementation of a standardized AI curriculum in medical education. To identify the values that medical educators should incorporate in curricula to adapt medical education to different healthcare contexts, including digitalized healthcare systems and the digital

student generation in a hyperconnected world, (Han et al., 2019) synthesized and introduced representative educational programs. Artificial intelligence was used by (Güner and Çomak, 2011) to predict student success not only in the field of medicine but also in the field of engineering.

(Zhao et al., 2018), an AI system may grasp clinical information, outperform the majority of human test takers, and make accurate clinical diagnoses based on electronic medical records, sometimes more precisely and consistently than humans, according to research conducted by a team of scientists in China. Doctors and other health professionals are usually defined by their knowledge and skill.

(Pucchio et. al, 2021) suggest including formal training on these topics in resident medical curricula and continuing medical education. Although teaching big data, digital technologies, artificial intelligence, and machine learning is urgent in health education, medical schools cannot teach students how to use, interpret, and apply these technologies in clinical settings and deliver health care. Therefore, the existence of artificial intelligence courses and computer-aided courses in private and public universities has been investigated, and the current situation has limitations when Turkey is considered within the scope of developing countries among OECD countries. Within the context of curriculum development studies, this research is anticipated to offer recommendations on how much technology should be incorporated into medical education.

Çalışkan et al. (2022) conducted an e-Delphi study focusing on the competencies required for the integration of AI in medical education. This study emphasizes that AI courses should go beyond providing only technical knowledge and provide students with the skills necessary to interpret AI outputs and apply them in clinical decision-making processes. Similarly, in our study, it was observed that AI courses in universities in Turkey were mostly theoretical and lacked practical components. The findings of Çalışkan et al. support the need to add practical applications to curricula.

Sridharan and Sequeira (2024) provide examples of how AI can be applied in classroom teaching and student assessment processes. Through a case study on pharmacology and therapeutics, they emphasize that AI can enhance personalized learning experiences and provide dynamic feedback to students. Our study similarly determined that AI has the potential to transform teaching methods in medical education, and this article further reinforces the importance of AI in course design to improve the quality of learning and support student success.

Zhang et al. (2024) examined the integration of AI into medical education from a global perspective and revealed the challenges encountered in this process. The study highlights issues such as lack of resources, lack of instructor training, and lack of standardized curricula. Our study also revealed similar obstacles in Turkey, and the findings of Zhang et al. support the need for Turkish universities to align their AI education initiatives with international standards. In addition, the role of AI in future medical practices, especially in personalized medicine and telemedicine, is highlighted. This further increases the importance of our recommendations for AI education.

**Author contribution statements**

Author1 and Author2 assisted in planning and implementing the study, conducted data analysis, wrote the manuscript, and contributed to critical discussions. Author1 supported study planning and implementation, data analysis, and manuscript drafting and offered constructive feedback. Author2 and Author1 played a role in study planning, execution, data analysis, and manuscript writing. Author2 conducted laboratory work and was responsible for manuscript writing. The final manuscript was reviewed and approved by Author2 and Author1.

## Ethics committee approval and conflict of interest statement

Afyonkarahisar University of Health Sciences Clinical Research Ethics Committee has unanimously decided that the ethics committee approval is not required for the study "Artificial Intelligence Course in Medical Education Curriculum: University Evaluation in Turkey," at the meeting numbered 2023/4 dated 07.04.2023. Relevant regulations and guidelines are conducted in all methods. Informed consent is not required for this study. The author declares that they have no competing interests with this article.

## REFERENCES

Alexandru, A. G., Radu, I. M., & Bizon, M.-L. (2018). Big Data in Healthcare-Opportunities and Challenges. *Informatica Economica*, 22(2).

Chan, K. S., & Zary, N. (2019). Applications and Challenges of Implementing Artificial Intelligence in Medical Education: Integrative Review. *JMIR Medical Education, 5*(1), e13930. doi:10.2196/13930.

Chen, C.-K. (2010). Curriculum Assessment Using Artificial Neural Network and Support Vector Machine Modeling Approaches: A Case Study. *IR Applications*, 29. Association for Institutional Research (NJ1).

Chen, Y., Wang, X., Jung, Y., Abedi, V., Zand, R., Bikak, M., & Adibuzzaman, M. (2018). Classification of Short Single-Lead Electrocardiograms (ECGs) for Atrial Fibrillation Detection Using Piecewise Linear Spline and XGBoost. *Physiological Measurement*, 39(10), 104006.

Cleophas, T. J., & Zwinderman, A. H. (2015). *Machine learning in medicine: A complete overview* (Vol. 21): Springer.

Cruz, J. (2007). Bryce. *Cuadernos Hispanoamericanos*, 689, 59. Retrieved from https://www.scopus.com/inward/record.uri?eid=2-s2.0-61149179761&partnerID=40&md5=c9b49bd722e5b041dd13cdd6b4447413.

Çalışkan, S. A., Demir, K., & Karaca, O. (2022). Artificial intelligence in medical education curriculum: An e-Delphi study for competencies. *PLoS One*, *17*(7), e0271872.

Deo, R. C. (2015). Machine learning in medicine. *Circulation, 132*(20), 1920-1930.

Ejaz, H., McGrath, H., Wong, B. L., Guise, A., Vercauteren, T., & Shapey, J. (2022). Artificial intelligence and medical education: A global mixed-methods study of medical students' perspectives. *Digital Health*, 8, 20552076221089099.

Garg, T. (2020). Artificial intelligence in medical education. *The American journal of medicine,* 133(2), e68.

Gupta, A., & Sao, D. (2011). The constitutionality of current legal barriers to telemedicine in the United States: Analysis and future directions of its relationship to national and international health care reform. *Health Matrix,* 21, 385.

Güner, N., & Çomak, E. (2011). Mühendislik Öğrencilerinin Matematik I Derslerindeki Başarısının Destek Vektör Makineleri Kullanılarak Tahmin Edilmesi. *Pamukkale University Journal of Engineering Sciences,* 17(2).

Han, E.-R., Yeo, S., Kim, M.-J., Lee, Y.-H., Park, K.-H., & Roh, H. (2019). Medical education trends for future physicians in the era of advanced technology and artificial intelligence: an integrative review. *BMC Medical Education,* 19(1), 460. doi:10.1186/s12909-019-1891-5.

Imran, N., & Jawaid, M. (2020). Artificial intelligence in Medical education: Are we ready for it. Pakistan Journal of Medical Sciences, 36(5), 857-859. doi:10.12669/pjms.36.5.3042.

Kolachalama, V. B., & Garg, P. S. (2018). Machine learning and medical education. *NPJ Digital Medicine,* 1(1), 54. doi:10.1038/s41746-018-0061-1.

Kourou, K., Exarchos, T. P., Exarchos, K. P., Karamouzis, M. V., & Fotiadis, D. I. (2015). Machine learning applications in cancer prognosis and prediction. *Computational and structural biotechnology journal,* 13, 8-17.

Lee, J., Wu, A. S., Li, D., & Kulasegaram, K. (2021). Artificial Intelligence in Undergraduate Medical Education: A Scoping Review. *Academic Medicine,* 96(11S). Retrieved from https://journals.lww.com/academicmedicine/Fulltext/2021/11001/Artificial_Intelligence_in_Undergraduate_Medical.14.aspx.

Lip, G. Y. H., Nieuwlaat, R., Pisters, R., Lane, D. A., & Crijns, H. J. G. M. (2010). Refining Clinical Risk Stratification for Predicting Stroke and Thromboembolism in Atrial Fibrillation Using a Novel Risk Factor-Based Approach: The Euro Heart Survey on Atrial Fibrillation. *Chest,* 137(2), 263-272. doi:https://doi.org/10.1378/chest.09-1584.

Lu, P., Abedi, V., Mei, Y., Hontecillas, R., Hoops, S., Carbo, A., & Bassaganya-Riera, J. (2015). Supervised learning methods in modeling of CD4+ T cell heterogeneity. *BioData mining,* 8, 1-21.

Masters, K. (2019). Artificial intelligence in medical education. *Medical Teacher,* 41(9), 976-980.

Nilsson, N. J. (1998). Artificial intelligence: a new synthesis. Morgan Kaufmann.

Noorbakhsh-Sabet, N., Zand, R., Zhang, Y., & Abedi, V. (2019). Artificial Intelligence Transforms the Future of Health Care. *The American Journal of Medicine,* 132(7), 795-801. doi:https://doi.org/10.1016/j.amjmed.2019.01.017

O'Mahony, C., Jichi, F., Pavlou, M., Monserrat, L., Anastasakis, A., Rapezzi, C., . . . Investigators, f. t. H. C. O. (2013). A novel clinical risk prediction model for sudden cardiac death in hypertrophic cardiomyopathy (HCM Risk-SCD). *European Heart Journal,* 35(30), 2010-2020. doi:10.1093/eurheartj/eht439.

Paranjape, K., Schinkel, M., Nannan Panday, R., Car, J., & Nanayakkara, P. (2019). Introducing Artificial Intelligence Training in Medical Education. *JMIR Medical Education,* 5(2), e16048. doi:10.2196/16048

Pucchio, A., Eisenhauer, E. A., & Moraes, F. Y. (2021). Medical students need artificial intelligence and machine learning training. *Nature Biotechnology,* 39(3), 388-389.

Sridharan, K., & Sequeira, R. P. (2024). Artificial intelligence and medical education: application in classroom instruction and student assessment using a pharmacology & therapeutics case study. BMC Medical Education, 24(1), 431.

Tolentino, R., Baradaran, A., Gore, G., Pluye, P., & Abbasgholizadeh-Rahimi, S. (2024). Curriculum frameworks and educational programs in AI for medical students, residents, and practicing physicians: scoping review. *JMIR Medical Education*, 10(1), e54793.

Topol, E. (2019). *Deep medicine: how artificial intelligence can make healthcare human again*: Hachette UK.

Zhang, W., Cai, M., Lee, H. J., Evans, R., Zhu, C., & Ming, C. (2024). AI in Medical Education: Global situation, effects and challenges. *Education and Information Technologies*, 29(4), 4611-4633.

Zhao, H., Li, G., & Feng, W. (2018, 10-11 Aug. 2018). *Research on Application of Artificial Intelligence in Medical Education.* Paper presented at the 2018 International Conference on Engineering Simulation and Intelligent Control (ESAIC).

BİLGİSAYAR BİLİMLERİ VE TEKNOLOJİLERİ DERGİSİ
JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGIES
http://dergipark.org.tr/tr/pub/bibted

Araştırma Makalesi

# Positioning Security Cameras in The Central Transportation Networks of Barcelona With Minimum Cost via The Malatya Minimum Vertex Cover Algorithm

**Cemalettin Sonakalan*[1], Furkan Öztemiz[1]**

*[1]Inonu University, Engineering Faculty, Software Engineering Department, Malatya, Türkiye*

**Keywords:**
Minimum Vertex Cover
Security Cameras
Graph Theory
Malatya Centrality

**ABSTRACT**

The Minimum Vertex Cover issue (MVCP) is a significant NP-complete optimization issue in graph theory. Its objective is to find a set of nodes that covers all edges of a given graph and contains the minimum number of nodes. Many different approaches and algorithms have been tried for this issue. Nevertheless, as the MVCP problem is an optimization problem, solutions are usually non-heuristic and only work under certain constraints Moreover, the proposed methods do not achieve the expected effect and the solution sets may change with each iteration. Having a minimum number of nodes in a network with a minimum coverage area improves network efficiency, reduces energy consumption, and allows for more efficient resource utilization. This study aims to control all streets in a popular neighborhood in Barcelona with a minimum number of security cameras. The Malatya Vertex Cover method is used to locate the optimal number of security cameras around the area. For modeling, the area is transformed into a graph using Google Earth. Each intersection represents a node. The graph was modeled using R programming language. Then, with the Malatya Vertex Cover algorithm, the Malatya centrality values of the nodes of the graph will be calculated. This centrality value is obtained from the sum of the ratio of the degree of each node to the degree of its neighbors. For the MVCP solution, the node of the graph with the highest Malatya centrality value is selected and added to the solution set. Then, this node and its edge links are removed from the graph. When the edges are completely covered, the process is terminated. As a result of this analysis, a low-cost solution is achieved by using the minimum number of security cameras to cover the entire region.

# Malatya Minimum Vertex Cover Algoritması ile Barselona'nın Merkezi Ulaşım Ağlarında Güvenlik Kameralarının Minimum Maliyetle Konumlandırılması

**Anahtar Kelimeler:**
Minimum Köşe Örtüsü
Güvenlik Kameraları
Çizge Teorisi
Malatya Merkezilik

**ÖZ**

Minimum Vertex Cover Problemi (MVCP), çizge teorisinde önemli bir NP-complete optimizasyon problemidir. Amacı, verilen bir grafın tüm kenarlarını kapsayan ve en az sayıda düğüm içeren bir düğüm kümesini bulmaktır. Bu problem için birçok farklı yaklaşım ve algoritma denenmiştir. Ancak MVCP problemi bir optimizasyon problemi olduğundan, çözümler genellikle sezgisel olmayıp belirli kısıtlamalar altında sonuç vermektedir. Bir ağda düğümlerin en az sayıda kapsanması ağın verimliliğini yükseltir, enerji tüketimini düşürür ve kaynakların daha verimli kullanılmasını sağlar. Bu çalışma Barcelona şehrinde popüler bir muhitteki cadde ve sokakların tümünü en az sayıda güvenlik kamerasıyla kontrol edilmesini hedefler. Bölge Google Earth kullanılarak çizgeye uygun modellenmiştir. Her bir kavşak bir düğümü temsil etmektedir. R programlama dili kullanılarak çizge oluşturulmuştur. Ardından Malatya Vertex Cover algoritmasıyla çizgenin düğümlerinin Malatya merkezilik değerleri hesaplanacaktır. Bu merkezilik değeri her bir düğümün derecesinin, komşularının derecesine oranının toplamından elde edilmektedir. MVCP çözümü için ise çizgenin en yüksek Malatya merkezilik değerine sahip olan düğümü seçilerek çözüm kümesine eklenir. Sonrasında bu düğüm ve kenar bağlantıları çizgeden çıkarılır. Kenarlar tamamen kapsandığında, işlem sonlandırılır. Bu analiz sonucunda tüm bölgeyi kapsayacak şekilde en az sayıda güvenlik kamerası kullanarak düşük maliyetli çözüm sağlanmıştır.

***Cemalettin Sonakalan**

*(yusufcs1234@gmail.com) ORCID ID 0009-0002-1391-3485
(furkan.oztemiz@inonu.edu.tr) ORCID ID 0000-0001-5425-3474

**e-ISSN: 2717-8579**
**Geliş Tarihi: 08/09/2024; Kabul Tarihi: 31/12/2024**

## 1. INTRODUCTION

A graph is a fundamental data structure consisting of nodes and edges(Thulasiraman et al., 2011). This form is commonly used in practical situations and the field of computer science(Hark and Karcı, 2022). The strategies and techniques applied to graphs facilitate the development of solutions and approaches in these domains. Nevertheless, graphs encompass numerous NP-complete problems(Thulasiraman et al, 2016). Among these, the node covering problem stands out as a key issue. This problem involves identifying which nodes should be selected to ensure that every edge in the graph is covered(Thulasiraman et al., 2011). When the objective is to cover all edges with the fewest possible nodes, this NP-complete problem is referred to as the minimum vertex covering problem (Khattab et al., 2022).

Minimum Vertex Cover Problem (MVCP) is frequently used in computer science to model today's problems (Angel, 2022) and (Wang, 2017). For this reason, the algorithms and solutions proposed for MVCP can be utilized in various applications (Dagdeviren, 2021). For example, the use of MVCP is important in various fields such as the detection of nodes for the protection of healthcare data (Angel, 2022), the generation of non-repetitive static genetic sequences (Hossain et al., 2020), the management of traffic, which is a common problem in today's world (Gusev, 2020), and the optimized use and positioning of hardware with limited facilities (Dagdeviren, 2021). In addition, security applications in wireless networks, routing and monitoring in wireless mobile networks (Yigit et al. 2022 and Yigit et al., 2021), link monitoring and formation in ad hoc networks (Dagdeviren, 2016), etc. are also areas where MVCP has been addressed. However, because the Minimum Vertex Cover is an NP-Complete optimization problem, it usually cannot be solved within an polinomial period. Therefore, exact solutions are presented for relatively simpler graphs or specified graphs. Various methods and approaches have been proposed to solve MVCP, but there is no method that produces optimal solutions to MVCP in polynomial time. Usually, after a certain number of iterations or processing steps, a set of possible solutions close to the optimal solution is identified. However, changing initial conditions and the characteristics of the method used may prevent the MVCP from determining the solution set. Real-world applications make solving the MVCP even more difficult due to the complexity and size.

In this work, Malatya Vertex Cover Algorithm(MVCA) is used to efficiently handle an NP-complete problem like MVCP. This algorithm calculates the Malatya centrality score for each node. When calculating this value, the node's own degree is taken into account as well as the degrees of its neighboring nodes. Then, Malatya centrality determines which nodes to select. The node with the highest MC value is selected and added to the solution set. Then, this node and its associated edges are removed from the graph. For the newly formed graph, the MVCA is applied again and new Malatya centralization scores are calculated. The operations on the graph continue until the vertex cover solution set is found. MVCA is a robust method with an efficient greedy approach. Thanks to the stable solution sets it produces, it can be used in real-time systems and large graphs.

### 1.1. Motivation To Work

Our contributions to the paper:

1. The main motivation of this study is to provide a high level of security in a city like Barcelona, which hosts millions of tourists every year, and to reduce the current cost and provide energy efficiency while providing this security.

2. As a result of this study, many touristic destinations in the world will be approached with a similar approach to this study and the concerns about the security problem will be eliminated. And the Barcelona region selected for analysis was prepared specifically for this study and a graph model was created. One of the most unique parts of the study is the Barcelona network developed specifically for this study.

3. In addition, Malatya Vertex Cover Algorithm was applied for the first time in the Barcelona region, which is a unique transportation network not previously available in the literature. Since the algorithm is an efficient method that provides optimum or near optimum solutions in polynomial time, successful results were obtained.

4. MVCA is applied to transportation networks for the first time in this study and the results are analyzed.

## 2. RELATED WORK

Since the MVCP is a popular graph theoretic problem, many methods with various approaches have been developed.

Khattab, Mahafzah, and Sharieh developed a hybrid algorithm using a combination of chemical reaction optimization and best-first search algorithms to solve the Minimum Vertex Cover problem, this approach provides remarkable performance improvements on large-scale problems (Khattab, 2022). Dinur and Safra analyzed the difficulty of approximate solutions to the Vertex Cover problem, showing that this problem is NP-hard and the theoretical limits of approximating the optimal solution, thus providing an important

perspective on the effectiveness of algorithms(Dinur, 2005). Wang et al.vdeveloped a Polynomial Time Approximation Scheme (PTAS) for the Minimum Weighted Connected Vertex Cover problem in three-dimensional wireless sensor networks(Wang, 2017). Öztemiz and Karcı provide concrete data on how the Vertex Cover problem can be applied on real-world transportation networks by analyzing the centrality of intersections in the transportation network of Malatya province(Öztemiz and Karci, 2021 ). (Angel, 2022) used graph theory approaches to protect medical information systems against cyber-attacks and showed how the Vertex Cover problem can be applied to the development of security strategies. Again, the research by (Yigit and Dagdeviren, 2022), and Challenger made significant contributions in improving network security and efficiency by investigating Vertex Cover algorithms with self-stabilizing capacity in IoT-based wireless sensor networks (Yigit and Dagdeviren, 2022).

Khattab, Sharieh, and Mahafzah improved the efficiency and accuracy of the algorithm by developing the Most Valuable Player (MVP) algorithm to solve the Minimum Vertex Cover problem (Khattab et al., 2019). Guo, Quan, and Chen achieved significant success in solving the Minimum Vertex Cover problem using the Membrane Evolutionary Algorithm (MEAMVC), which aims to achieve a solution by mimicking biological evolutionary processes(Guo, 2019). Xie et al. in their work called Test-cost-sensitive rough set based approach, addressed the Minimum Weight Vertex Cover problem and developed more effective solutions with this method, especially considering cost sensitivity (Xie, 2018). Jovanovic, Sanfilippo, and Voß used the Fixed Sets Search algorithm for the Multi-objective Minimum Weighted Vertex Cover problem and showed that this method is effective in multi-objective optimization problems (Jovanovic, 2022). Li et al. developed a new local search algorithm called NuMWVC to solve the Minimum Weighted Vertex Cover problem and proved that this method is especially effective for large graphs(Li et al., 2019).

These works provide innovative algorithms and theoretical analyses then developing solution for the Minimum Vertex Cover problem, providing significant advances in both academic research and practical applications. The analysis of various optimization techniques, security applications, and real-world networks makes it possible to address the Vertex Cover problem with a multifaceted approach. In particular, innovative methods such as chemical reaction optimization, evolutionary algorithms and local search techniques have made great progress in solving the Vertex Cover problem. In this context, the diversity and depth of research provides a broader perspective on the solution of the Vertex Cover problem.

It can be said that the methods developed for solving MVCP generally have exac, heuristic and greedy approaches. Exac methods aim for the optimum solution and usually all possibilities are tried. Therefore, it is not possible to produce solutions for large graphs. Heuristic methods run more than one iteration and can produce different results in each run. Greedy methods usually use a priority value for node selection and proceed by selecting the node that they think is the best in the current situation. Malatya Vertex Cover algorithm also produces deterministic results with an efficient greedy approach.

## 3. MATERIAL AND METHOD

There are two important aspects of the presented work. First, Malatya Vertex Cover algorithm is joined to transportation networks for the first time then analyzed. The second is the network selected for analysis. The region analyzed in the study is the center of the city of Barcelona. The reason why Barcelona was chosen is that it is a popular city known to many people and has a developed transportation network.In the following sections, detailed information about both the graph and the methodology is given. Figure 1 shows a visualization of the process of transforming the transportation network of the city of Barcelona into a graph. The selected region was specially designed for the study. The generated graph is unweighted and undirected

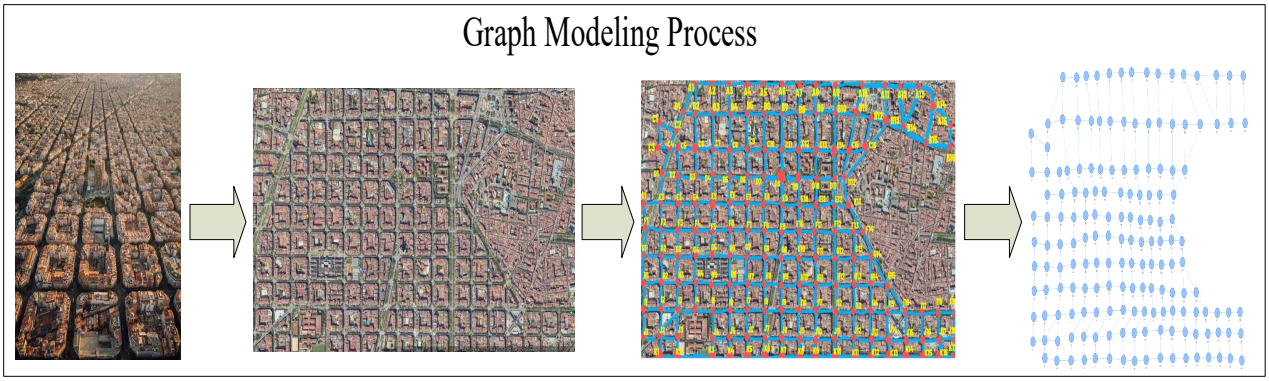



**Figure 1.** The Process of Barcelona Down Center Graph

## 4. METHOD

Barcelona is the region of interest in this study due to the fact that it hosts more than 10 million tourists per year on average and the increasing security concerns in such touristic areas. Thanks to the successful results of the study on a real transportation network model, it is thought that it can provide successful results on any transportation network. The region under study is shown in Figure2.

**Figure 2.** Barcelona Downtown

A bird's eye view of the area subject to analysis is shown in Figure 3

**Figure 3.** Birdview of Related Area

In the process of designing the network, the popular area of the city of Barcelona was first marked with Google Earth. During this marking, a bird's eye view was obtained to model the intersections and streets more accurately. Each intersection in the image represents a node of the graph and the line between each node represents the edges of the graph. After the whole region was marked, each node was given a name. At this stage, the visual model of the graph was complete. See Figure 4.

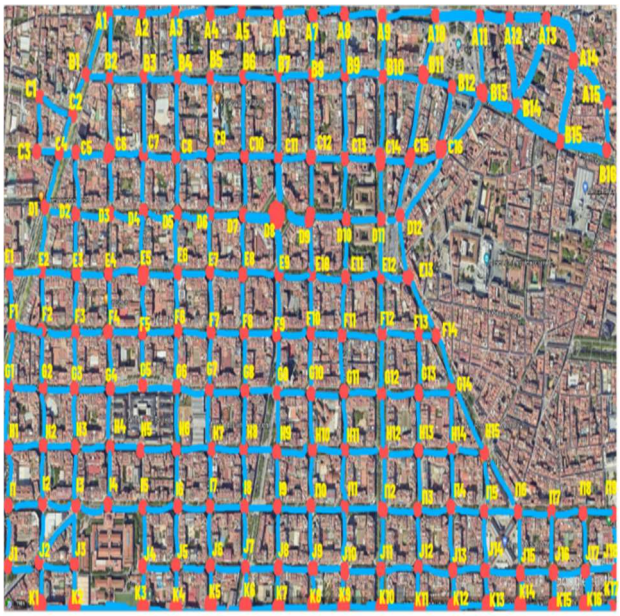


**Figure 4.** Identify nodes

After the visual model of the network was completed, the graph was modeled again in R programming language in order to apply the algorithm. Malatya Vertex Cover algorithm when applied to the graph, we get. The 169 nodes in the graph before applying the algorithm are shown in Figure 5.

The data set we obtained was simulated to the image on the real map using libraries such as igraph, visNetwork, shiny and an interactive image was provided.

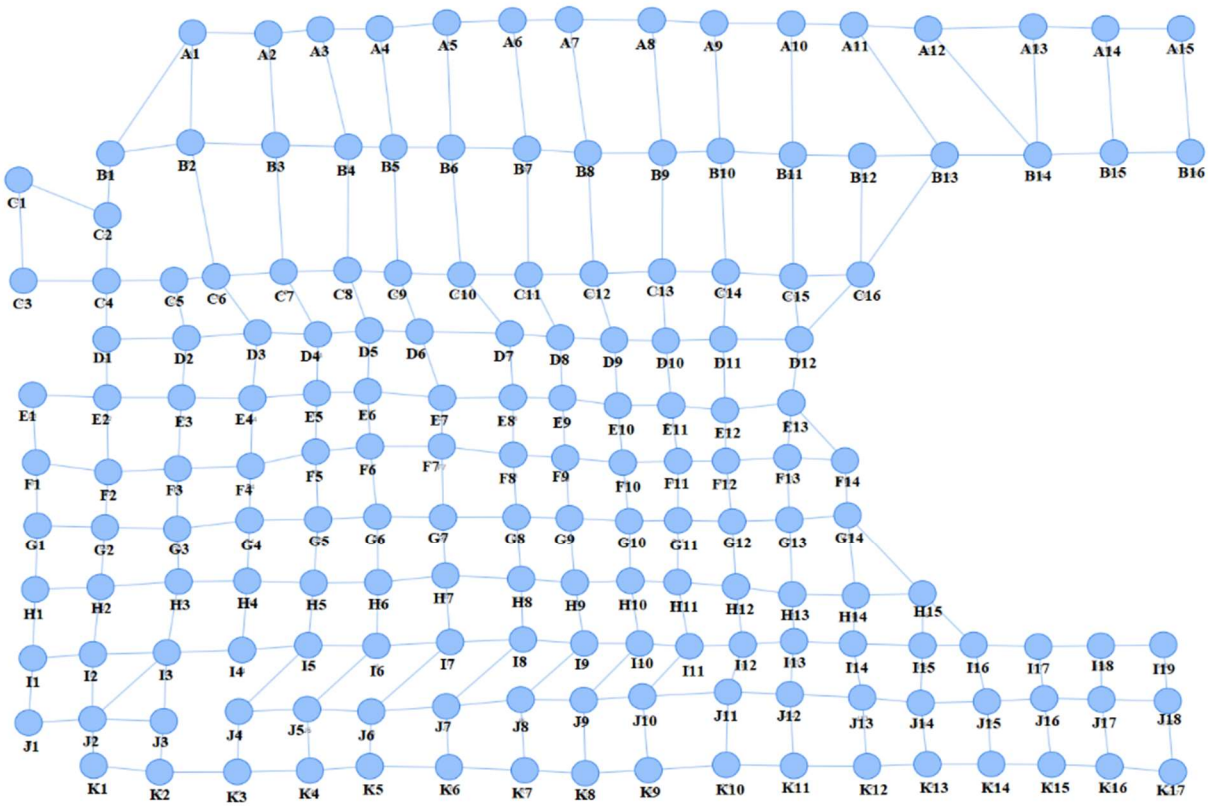



**Figure 5.** Graph model in R language

### 4.1. Malatya Vertex Cover Algorithm

The most critical parameter in solving the MVCP problem is the process of selecting the nodes. Malatya algorithm is used to guide this processThis algorithm computes a centrality value for each node and generates an efficient solution set for MVCP. Differently from previous optimization methods, the centrality value clearly defines the nodes. Operations start after Malatya centrality values are calculated. The nodes are ranked starting from the one with the highest centrality value and edge coverage is performed. A new graph is created by removing the selected node and the edges connected to it. These steps are repeated until all edges are covered and finally the minimum edge coverage set is obtained (Yakut et al., 2023). The extended code for this approach is shown in table 1.

**Table 1.** Proposed Algorithm Pseudo Code

Minimum Vertex-Cover (Karci et al, 2022)

```
calculateMalatyaCentrality <- function(graph) {
  centrality_values <- c()
  vertex_list <- V(graph)  # Get the list of vertices in the graph
 for (vertex in vertex_list) {
   vertex_degree <- degree(graph, v = vertex)

  neighbors_degrees <- degree(graph, v = neighbors(graph, v = vertex))

  # Centrality calculation

  centrality_value <- vertex_degree / neighbors_degrees

  centrality_values <- c(centrality_values, sum(centrality_value))
}
 return(centrality_values)  # Return the calculated centrality values}
# Function to find the vertex with the highest Malatya centrality
findMaximumCentralityVertex <- function(graph) {
  centrality_values <- calculateMalatyaCentrality(graph)
  centrality_data <- data.frame(centrality_values)
 return(order(centrality_data$centrality_values, decreasing = TRUE)[1]) }
# Function to find the vertex cover of the graph
findVertexCover <- function(graph) {
  vertex_cover <- c()  # Initialize an empty vertex cover
  while (ecount(graph) > 0) {  # Continue until all edges are covered
   max_vertex <- findMaximumCentralityVertex(graph)
   vertex_cover <- append(vertex_cover, V(graph)[max_vertex])
   # Remove the selected vertex and its edges from the graph
   graph <- delete_vertices(graph, max_vertex)}
 return(vertex_cover)  # Return the resulting vertex cover}
```

#### 4.1.1. Calculation of Malatya Centrality Value

Centrality plays a critical role in graph theory and network analysis because it allows nodes to be weighted and assigned values based on their location. It is often used to find the most influential nodes in the network and several algorithms have been developed for this purpose. The approach, known as degree centrality, is determined by the links on the node and is the basis of popular algorithms such as PageRank.

For solving the Minimum Vertex Cover Problem (MVCP), calculating Malatya centrality values is an important step. Malatya algorithm, which is a pragmatic and efficient method, is used to calculate these values. Malatya algorithm calculates Malatya centrality values for each node separately. In this calculation, the node's own degree and the degrees of its neighboring nodes are used.

The obtained centrality values are used to determine the nodes to be selected for vertex cover. This approach allows the MVCP to be solved in polynomial time, making it possible to complete the solution in finite steps and in a predictable time.

The Malatya algorithm uses the following formula to calculate the Malatya centrality value $\psi(v_i)$ of each node (Yakut et al., 2023)

$$\psi(v_i) = \sum_{\forall v_j \in N(v_i)} \frac{d(v_i)}{d(v_j)} \quad (1)$$

Here, the Malatya centrality value of each node is the sum of the node's own degree divided by the degree of each neighboring node. This calculation shows that the number and degree of neighboring nodes are as important as the degree of the node itself.

#### 4.1.2. Working and Spacetime Complexities

The Malatya centrality algorithm has been applied to calculate the centrality values for all nodes in the graph, exhibiting a time complexity of $O(n^2)$, which depends on the number of nodes and edges. Additionally, the algorithm requires a maximum of two adjacency matrices, resulting in a space complexity of $|V^2|$(Yakut et al., 2023).

## 5. EXPERIMENTAL RESULTS

Table 2 shows the nodes selected according to the centrality values taken as a criterion when applying the Malatya Vertex Cover algorithm. 89 nodes were obtained from the entire graph.

**Table 2.** Nodes selected by the algorithm in the example graph

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| A1 | A3 | A5 | A7 | A9 | A11 | A12 | A14 |
| B1 | B3 | B5 | B7 | B9 | B11 | B13 | B14 |
| C1 | C4 | C6 | C8 | C10 | C12 | C14 | C16 |
| D2 | D4 | D6 | D8 | D10 | D12 | E2 | E4 |
| E6 | E8 | E10 | E12 | E13 | F1 | F3 | F5 |
| F7 | F9 | F11 | F13 | G2 | G4 | G6 | G8 |
| G10 | G12 | G14 | H1 | H3 | H5 | H7 | H9 |
| H11 | H13 | H15 | I1 | I3 | I4 | I6 | I8 |
| I10 | I12 | I14 | I16 | I18 | J2 | J4 | J6 |
| J8 | J10 | J12 | J14 | J16 | J18 | K2 | K4 |
| K6 | K8 | K10 | K12 | K14 | K16 | | |

The initial 169 nodes were reduced to 88 nodes after the algorithm was applied. The selected nodes are colored in different yellow colors (Figure 6).

These 88 selected nodes cover all edges of the graph. Considering that security cameras are positioned only on the selected yellow nodes, an energy and cost saving of almost 50% is achieved.

The GMin algorithm with a similar greedy approach was applied on the same transportation network and the result was compared with MVCA. GMin, developed by M. Goldberg and colleagues, uses the degree centrality value to select the vertex with the lowest centrality at each iteration as a member of the independent set. The selected vertex and its neighbors are then removed from the graph (Goldberg et al. 2005). Using the mathematical relation between independent set and vertex cover, Maximum Independent Set Members + Minimum Vertex Cover Members = Total number of Vertexes (Alipour and Salari, 2022), the VC locations in the transportation network of Barcelona city were determined.

After applying Gmin to the transportation network, the Vertex Cover value was determined as 92. When the results obtained are analyzed, it is seen that MVCA produces better results than Gmin, which is popularly used in the literature with a value of 88. In other words, with MVCA, all roads can be monitored by installing security cameras at 88 locations detected in the transportation network, while this value is 92 with the Gmin algorithm. MVCA provides a significant cost advantage by using 4 fewer cameras. These results show that MVCA produces results that are successful enough to be included in the literature for solving real world problems.

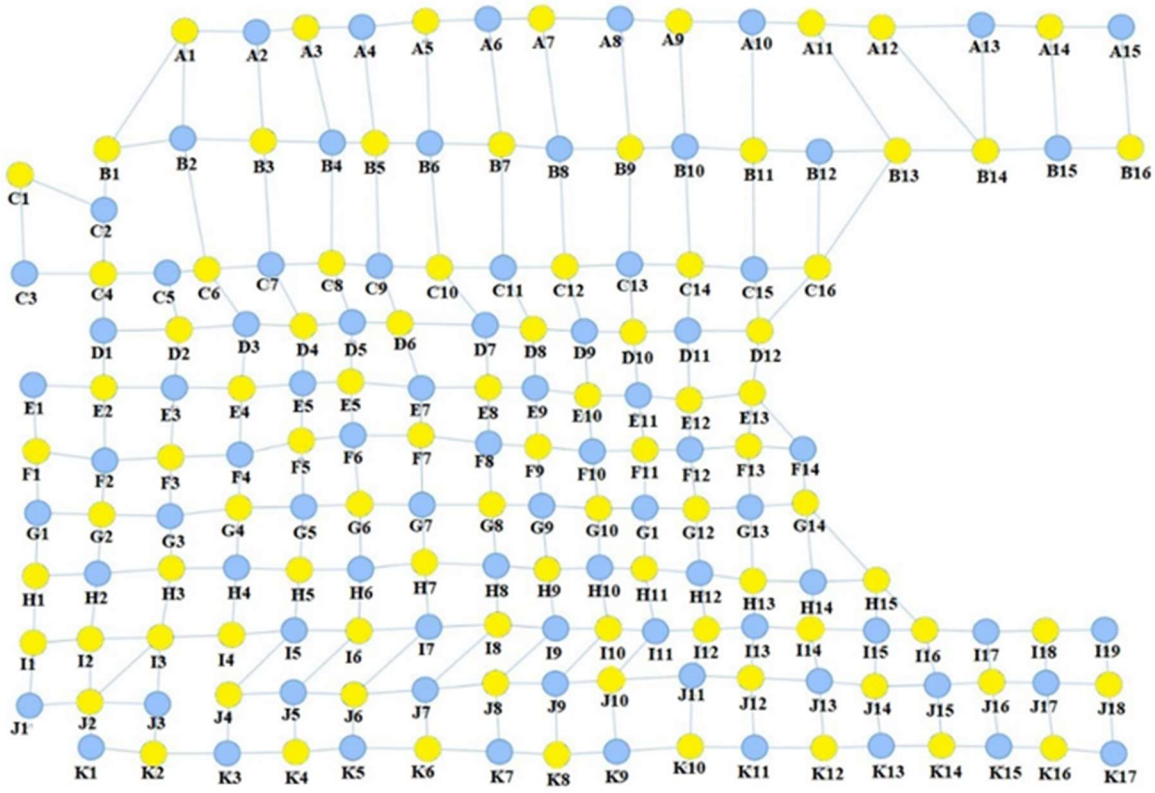



**Figure 6.** Nodes obtained as a result of the algorithm application those collored yellow

## 6. RESULTS

In popular touristic areas, it is necessary for each region to install security cameras for tourists to travel more safely and to monitor possible thefts, etc. It would be quite costly to install these cameras at all intersections in the entire region and monitor the area. However, this issue has been meticulously addressed in this study. Thanks to the Malatya Vertex Cover algorithm, 169 intersection points were initially identified for the region whose boundaries were determined, but this number was reduced to 88 by the application of the algorithm. In this way, the region can be monitored with far fewer cameras than usual. Both cost and energy savings were achieved. The transportation network used belongs to the city of Barcelona and has not been studied before in the literature. In addition, this is the first time such an analysis has been done in this region using the Malatya Vertex Cover algorithm. Considering that there are many such touristic regions in the world, this study has paved the way for other regions and the study can be extended for other purposes with different sectors and stakeholders.

## REFERENCES

Thulasiraman, K., Swamy, M, NS. (2011). Graphs: theory and algorithms. Montreal: John Wiley & Sons.

Hark, C., Karcı, A. (2022). A new multi-document summarisation approach using saplings growing-up optimisation algorithms : Simultaneously optimised coverage and diversity. https://doi.org/10.1177/0165551522110184 1.

Thulasiraman, K., Arumugam, S., Brandstädt, A., and Nishizeki, T. (2016). Handbook of Graph Theory, Combinatorial Optimization, and Algorithms, Boca Raton: Chapman & Hall/CRC.

Khattab, H., Mahafzah, B, A., and Sharieh, A. (2022). A hybrid algorithm based on modified chemical reaction optimization and best-first search algorithm for solving minimum vertex cover problem. Neural Comput. Appl., vol. 34(18), pp. 15513–15541, https://doi.org/10.1007/s00521-022-07262-w.

Dinur, I., and Safra, S. (2005). On the hardness of approximating vertex cover. Ann. Math. 162(1), 439–485, https://doi.org/10.4007/annals.2005.162.439.

Angel, D. (2022). Protection of Medical Information Systems Against Cyber Attacks: A Graph Theoretical Approach. Wirel. Pers. Commun., 126(4), 3455–3464, https://doi.org/10.1007/s11277-022-09873-x

Wang, L., Du, W., Zhang, Z., and Zhang, X. (2017). A PTAS for minimum weighted connected vertex cover $P_3$ problem in 3-dimensional wireless sensor networks. J. Comb. Optim., 33(1), 106–122. https://doi.org/10.1007/s10878-015-9937-z

Hossain, A. (2020). Automated design of thousands of nonrepetitive parts for engineering stable genetic systems. Nat. Biotechnol., 38(12), 1466–1475. https://doi.org/10.1038/s41587-020-0584-2

Gusev, V, V. (2020). The vertex cover game: Application to transport networks. Omega, 97, 102102. https://doi.org/10.1016/j.omega.2019.08.009

Dagdeviren, Z., A. (2021). Weighted Connected Vertex Cover Based Energy-Efficient Link Monitoring for Wireless Sensor Networks Towards Secure Internet of Things. IEEE Access, 9, 10107–10119. https://doi.org/10.1109/ACCESS.2021.3050930

Yigit, Y., Dagdeviren, O., and Challenger, M. (2022). Self-Stabilizing Capacitated Vertex Cover Algorithms for Internet-of-Things-Enabled Wireless Sensor Networks. Sensors, 22(10), 3774. https://doi.org/10.3390/s22103774

Yigit Y., Dagdeviren, Z. A., Dagdeviren, O., and Challenger, M. (2021). Performance Evaluation of Capacitated Vertex Cover Algorithms for Security Applications in Wireless Sensor Networks. in 7th International Conference on Electrical, Electronics and Information Engineering: Technological Breakthrough for Greater New Life, ICEEIE. https://doi.org/10.1109/ICEEIE52663.2021.9616719

Yigit, Y., Akram, V. K., and Dagdeviren, O. (2021) Breadth-first search tree integrated vertex cover algorithms for link monitoring and routing in wireless sensor networks. Comput. Networks, 194, 108144. https://doi.org/10.1016/j.comnet.2021.108144

Dagdeviren, Z. A. (2022) A Metaheuristic algorithm for vertex cover based link monitoring and backbone formation in wireless Ad hoc Networks. Expert Syst. Appl., 213(PA), 118919. https://doi.org/10.1016/j.eswa.2022.118919.

Öztemiz, F., Karci, A. (2021). Malatya İli ulaşım ağı kavşak noktalarının Merkezlilik Analizi. Gazi Üniversitesi Mühendislik Mimarlık Fakültesi Dergisi 37(1),511-528. https://doi.org/10.17341/gazimmfd.834255

Yigit, Y., Dagdeviren, O., and Challenger, M., (2022). Self-Stabilizing Capacitated Vertex Cover Algorithms for Internet-of-Things-Enabled Wireless Sensor Networks. Sensors, 22(10), 3774. https://doi.org/10.3390/s22103774

Hebatulla,h K., Sharieh, A. (2019). Most Valuable Player Algorithm for Solving Minimum Vertex Cover Problem. Basel A. January 2019International Journal of Advanced Computer Science and Applications 10(8) https://doi.org/10.14569/IJACSA.2019.0100821

Guo, Quan, and Chen (2019). achieved significant success in solving the Minimum Vertex Cover problem using the Membrane Evolutionary Algorithm (MEAMVC)

Xiaojun, X., Xiaolin, Q., Chunqiang, Y., Xingye, X. (2018). Test-cost-sensitive rough set based approach for minimum weight vertex cover problem. Applied Soft Computing, 64, 423-435. https://doi.org/10.1016/j.asoc.2017.12.023

Jovanovic. R., Sanfilippo, AP., & Voß, S. (2022). Fixed set search applied to the multi-objective minimum weighted vertex cover problem. Journal of Heuristics, 28(4), 481–508. https://doi.org/10.1007/s10732-022-09499-z

Li, R., Hu, S., Cai, S., Gao, J., Wang, Y., & Yin, M. (2019). NuMWVC: A novel local search for minimum weighted vertex cover problem. Journal of the Operational Research Society, 71(9), 1498–1509. https://doi.org/10.1080/01605682.2019.1621218

Yakut, S., Öztemiz, F., & Karci A. (2023). A new robust approach to solve minimum vertex cover problem: Malatya vertex-cover algorithm. J Supercomput 79, 19746–19769 https://doi.org/10.1007/s11227-023-05397-8

Karci, A., Yakut, S., & Öztemiz, F. (2022). A New Approach Based on Centrality Value in Solving the Minimum Vertex Cover Problem: Malatya Centrality Algorithm. Computer Science, 7(2),

81-88.
https://doi.org/10.53070/bbd.1195501

Yakut, S., Öztemiz, F., & Karci, A. (2023). A New Approach Based on Centrality Value in Solving the Maximum Independent Set Problem: Malatya Centrality Algorithm. Computer Science, 8(1), 16-23. https://doi.org/10.53070/bbd.122452

Goldberg, M., Hollinger, D., Magdon-Ismail, M. (2005). Experimental Evaluation of the Greedy and Random Algorithms for Finding Independent Sets in Random Graphs. In: Nikoletseas, S.E. (eds) Experimental and Efficient Algorithms. WEA 2005. Lecture Notes in Computer Science, 3503. Springer, Berlin, Heidelberg. https://doi.org/10.1007/11427186_44

Alipour, S., & Salari, M. (2022). Brief announcement: Distributed algorithms for minimum dominating set problem and beyond, a new approach. In 36th International Symposium on Distributed Computing (DISC 2022), 246, 40:1-40:3. https://doi.org/10.4230/LIPIcs.DISC.2022.40