



# JOURNAL OF INFORMATION TECHNOLOGIES

## BİLİŞİM TEKNOLOJİLERİ DERGİSİ

Volume / Cilt **18** Number / Sayı **1** Year / Yıl **2025** Month / Ay **January / Ocak**





GAZİ ÜNİVERSİTESİ (GAZİ UNIVERSITY)  
BİLİŞİM ENSTİTÜSÜ (INSTITUTE OF INFORMATICS)

**BİLİŞİM TEKNOLOJİLERİ DERGİSİ**  
(JOURNAL OF INFORMATION TECHNOLOGIES)  
ISSN: 1307-9697 e-ISSN: 2147-0715



Cilt (Volume): 18

Sayı (Issue): 1

Ocak (January) 2025

Sahibi (Owner)  
**Dr. Uğur ÜNAL**  
Rektör (Rector)

**Genel Yayın Yönetmeni & Baş Editör**  
(General Publication Director & Editor in Chief)

**Dr. Ashhan TÜFEKÇİ**

Bilişim Enstitüsü Müdürü  
Director of Institute of Informatics

**Yardımcı Editör**

(Associate Editor)

**Dr. Öner BARUT**

Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

**Yardımcı Editör**

(Associate Editor)

**Dr. Nadircan DAĞLI**

Bilişim Enstitüsü Müdür Yardımcısı  
Vice Director of Institute of Informatics

**Editörler (Editors)**

- ❖ **Dr. Serdar KULA** Gazi Üniversitesi (Gazi University)
- ❖ **Dr. Hüseyin POLAT** Gazi Üniversitesi (Gazi University)
- ❖ **Dr. Resul DAŞ** Fırat Üniversitesi (Fırat University)
- ❖ **Dr. Muzaffer KAPANOĞLU** Eskişehir Osmangazi Üniversitesi (Eskişehir Osmangazi University)
- ❖ **Dr. Mehmet ŞİMŞEK** Milli Savunma Üniversitesi (National Defence University)
- ❖ **Dr. Mehmet Sıraç ÖZERDEM** Dicle Üniversitesi (Dicle University)
- ❖ **Dr. Mustafa Mahir ÜLGÜ** Sağlık Bakanlığı
- ❖ **Dr. Murat YILMAZ** Gazi Üniversitesi (Gazi University)
- ❖ **Dr. Oktay YILDIZ** Gazi Üniversitesi (Gazi University)
- ❖ **Dr. Olgun DEĞİRMENCİ** TOBB ETÜ (TOBB Economics and Technology University)
- ❖ **Dr. Recep BENZER** Gazi Üniversitesi (Gazi University)
- ❖ **Dr. Ö. Tolga PUSATLI** Çankaya Üniversitesi (Çankaya University)
- ❖ **Dr. Cihangir TEZCAN** Orta Doğu Teknik Üniversitesi (Middle East Technical University)
- ❖ **Dr. Mehmet SEVRİ** Recep Tayyip Erdoğan Üniversitesi (Recep Tayyip Erdoğan University)
- ❖ **Dr. Muhammed Ali KOŞAN** Kahramanmaraş İstiklal Üniversitesi (Kahramanmaraş Istiklal University)
- ❖ **Dr. Levent ÇETİNKAYA** Çanakkale Onsekiz Mart Üniversitesi (Çanakkale Onsekiz Mart University)
- ❖ **Dr. Sinan TOKLU** Gazi Üniversitesi (Gazi University)

**Yayın Danışma Kurulu (Editorial Advisory Board)**

<b>Dr. Ahmet COŞAR</b> <i>Turkish Aeronautical Association University, Turkey</i>	<b>Dr. Aslanbek NAZİEV</b> <i>Ryazan State University, Russia</i>	<b>Dr. Bogdan PATRUT</b> <i>Alexandru Ioan Cuza University of Iasi, Romania</i>
<b>Dr. Deepak GUPTA</b> <i>Maharaja Agrasen Institute of Technology, India</i>	<b>Dr. Jafar A. ALZUBİ</b> <i>Al-Balqa Applied University, Jordan</i>	<b>Dr. Jolanta SABAITYTĖ</b> <i>Vilnius Gediminas Technical University, Lithuania</i>
<b>Dr. Ilya LEVİN</b> <i>Tel Aviv University, Israel</i>	<b>Dr. Pınar KARAGÖZ</b> <i>Middle East Technical University, Turkey</i>	<b>Dr. Ufuk ÇAĞLAYAN</b> <i>Yaşar University, Turkey</i>
<b>Dr. Veysi İŞLER</b> <i>Hasan Kalyoncu University, Turkey</i>	<b>Dr. Victor Hugo Costa DE ALBUQUERQUE</b> <i>Universidade de Fortaleza, Brazil</i>	<b>Dr. Vijender Kumar SOLANKİ</b> <i>CMR Institute of Technology, India</i>
<b>Dr. Ebrahim KHOSRAVI</b> <i>Clayton State University, United States</i>		

**Dil Editörü**

*(Language Editor)*

**Dr. Çağla Gizem AKKAŞ**

*Bilişim Enstitüsü  
Institute of Informatics*

**Teknik Sorumlu**

*(Technical Assistant)*

**Dr. Candan TÜMER**

*Bilişim Enstitüsü  
Institute of Informatics*

**Teknik Sorumlu**

*(Technical Assistant)*

**Yasemin İÇTÜZER**

*Bilişim Enstitüsü  
Institute of Informatics*

**Sekreterlik**

*(Secretary)*

**Bilişim Teknolojileri Dergisi**

*Bilişim Enstitüsü  
Institute of Informatics*

**Bilişim Teknolojileri Dergisi uluslararası hakemli bir dergidir.**  
Journal of Information Technologies is an international refereed journal.

**Yazışma Adresi (Contact Address)**

Tunus Cad. No: 35 Kavaklıdere Çankaya/ANKARA  
Telefon / Telephone: 0312 202 38 01  
Faks / Fax: 0312 212 79 29

**Çevrimiçi Değerlendirme Sistemi (Online Evaluation System)**

<http://dergipark.gov.tr/gazibtd>  
**E-posta (e-mail):** btd@gazi.edu.tr

**Bilişim Teknolojileri Dergisi 3 ayda bir (Ocak, Nisan, Temmuz, Ekim) yayınlanmaktadır.**  
Journal of Information Technologies is published every 3 months (January, April, July, October).

# YOLO V8 Algoritması ile Otomatik Plaka Tanıma ve Görselleştirme Sistemi

## Araştırma Makalesi/Research Article

 Fatih GÜL<sup>1,2</sup>,  Esmâ SERTTAŞ<sup>1\*</sup>

<sup>1</sup>Elektrik-Elektronik Mühendisliği Bölümü, Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye

<sup>2</sup>Yapay Zekâ-Nesnelere İnterneti Araştırma Laboratuvarı, Recep Tayyip Erdoğan Üniversitesi, Rize, Türkiye

[fatih.gul@erdogan.edu.tr](mailto:fatih.gul@erdogan.edu.tr), [esma.serttas.7@gmail.com](mailto:esma.serttas.7@gmail.com)

(Geliş/Received:27.06.2024; Kabul/Accepted:19.10.2024)

DOI: 10.17671/gazibtd.1506041

**Özet**— Bu çalışma ile, belirli bir mesafeye yerleştirilen bir kamera ile YOLO (You Only Look Once) V8 algoritmasını kullanarak aracın üzerindeki plakayı otomatik olarak tanıyan ve görselleştiren bir sistem tasarlanmıştır. YOLO V8, gelişmiş bilgisayarlı görü yeteneklerine sahip olmakla birlikte doğrudan plaka tanıma modeli içermemektedir. Bu çalışma ile güvenlik önlemleri gerektiren alanlarda insan gücünü ve maliyeti en aza indirerek verimli şekilde kullanılabilir bir model önerilmiştir. Plaka veri seti, bilgisayarlı görü modeli ortamı Roboflow kullanılarak oluşturulmuş ve yapay sinir ağı eğitim modeli geliştirilmiştir. Python programlama dili kullanılarak YOLO V8 algoritması ile yapay sinir ağı modeli Karayolları Trafik Yönetmeliğine uygun TR plakalar ile eğitilerek plaka tanıma işlemleri gerçekleştirilmiştir. Geliştirilen bu sistemde, açık kaynaklı kütüphaneler olan OpenCV, Time, Random, Numpy, Ultralytics ve EasyOCR kullanılmıştır. Kullanıcı arayüzü için Tkinter kullanılarak plaka tanıma sonuçları görselleştirilmiştir. Sistem tam karşıdan, sağ ve sol yönde 30° içerisinde kalacak şekilde farklı açılardan alınan görüntüler üzerinde test edilmiş ve yüksek doğruluk oranları (%99 @ 25 Epok) elde edilmiştir. Bu çalışma, trafik yönetimi, otopark sistemleri ve güvenlik uygulamaları gibi çeşitli alanlarda mevcut YOLOV8 tabanlı uygulamalara entegre edilebilir bir çözüm yöntemi önermektedir.

**Anahtar Kelimeler**— plaka tanıma sistemi, görüntü işleme, Yolov8, bilgisayarlı görü, yapay sinir ağları

## Automatic License Plate Recognition and Visualization System with YOLO V8 Algorithm

**Abstract**— The aim of this study is to develop a system that automatically recognizes and visualize the license plate on the vehicle using the YOLO (You Only Look Once) V8 algorithm with a camera placed at a certain distance. Although YOLO V8 has advanced computer vision capabilities, it does not have a direct license plate recognition model. With this study, a model which can be used efficiently by minimizing manpower and cost in areas that require security measures, was offered. The license plate dataset was developed using the computer vision model environment Roboflow and an artificial neural network training model was created. The license plate recognition operations have been performed by training a neural network model using the YOLO V8 algorithm in Python with TR plates in accordance with the Highway Traffic Regulation. In this developed system, open source OpenCV, Time, Random, Numpy, Ultralytics and EasyOCR libraries were used. By using Tkinter for the user interface, license plate recognition results were visualized. The system was tested on images taken from different angles within 30° from the front, right and left, and high accuracy rates (99% @ 25 Epoch) were obtained. This study offers practical solutions in various fields such as traffic management, parking systems and security applications that can be integrated into current YOLOV8 based applications.

**Keywords**— license plate recognition system, image processing, YOLOV8, computer vision, artificial neural networks

## 1. GİRİŞ (INTRODUCTION)

Son çeyrek yüzyılda araç sayısındaki artış ve trafikte meydana gelen sorunlar, otomatik araç tanıma ve trafik akışının kontrolü üzerine yapılan çalışmaları artırmıştır [1],[2],[3],[4]. Günümüzde trafiği denetlemek amacıyla mikrodalga dedektörleri, yolun altına yerleştirilen tüpler, loop dedektörleri ve radyo frekansları kullanan radarlar gibi çeşitli yöntemler kullanılmaktadır. Ancak bu donanımların pahalı olması, sistemlerin işletimini zorlaştırmaktadır. Gelişen teknolojiyle birlikte dijital görüntü işleme alanında önemli gelişmeler kaydedilmiştir. Dijital görüntü işleme, görüntünün dijital formata dönüştürülerek çeşitli işlemlerle iyileştirilmesi ve bilgi çıkarılması yöntemidir. Dijital görüntü işleme yöntemleri kullanılarak ek donanım gerektirmeden plaka tanıma sistemi geliştirilebilmektedir. Her aracın kendine özgü bir plakası olduğu için, dijital görüntü işleme teknikleriyle araç tanıma problemi etkili bir şekilde çözülebilmektedir. Türkiye'de plakalar, Karayolları Yönetmeliği tarafından belirlenen standartlara uygun olarak üretilmektedir [5]. Karayolları Yönetmeliği'ne göre standart plakanın resmi Şekil 1'de gösterilmiştir [6]. Plaka gövdesi alüminyumdan yapılmakta olup, araçların hem ön hem de arka kısmında plaka bulunması zorunludur. Araçlarda bulunan yönetmeliğe uygun ön ve arka plakaların resmi Şekil 2'de gösterilmiştir.



Şekil 1. Karayolları Trafik Yönetmeliği'ne göre standart plakanın görüntüsü. (Image of the standard license plate according to the Road Traffic Regulations.)



Şekil 2. Araçlarda bulunan yönetmeliğe uygun plakanın önden ve arkadan görüntüsü. (Front and back view of the regulated license plate on the vehicles.)

Bu projenin amacı, araç plakalarını belirli bir mesafeden kamera kullanarak tanımak ve görüntü işleme algoritmaları ile bu plakaları görselleştirmek üzere bir sistem geliştirmektir. Literatürde plaka tanıma üzerine birçok çalışma bulunmaktadır. Örneğin, Doe v.d. (2019), C# programlama dili ve YOLOv2 algoritmasını kullanarak otomatik plaka tanıma sistemi geliştirmişlerdir [7]. Benzer

şekilde, Smith ve Brown (2020), Python programlama dili ve YOLOv3 algoritmasını kullanarak plaka tanıma performansını artırmayı başarmışlardır [8].

Bu çalışmada YOLOv8 algoritması ve Python programlama dili ile geliştirilen açık kaynak kütüphaneleri tercih edilmiştir. YOLO algoritmasının seçilme nedeni, görüntü işlemede nesne tespiti için yaygın kullanılan bir derin öğrenme algoritması olmasıdır. Diğer nesne tespiti algoritmalarından farkı, gerçek zamanlı bir görüntüyü tek bir geçişte işleyerek birden fazla nesneyi tespit etmesidir. Genellikle diğer algoritmalar, bölge tabanlı veya önceden belirlenmiş bölgelerde (örneğin R-CNN, Fast R-CNN vb. algoritmalar) birden fazla geçiş yaptığı için yavaş çalışmaktadır. YOLO algoritması ise diğer yöntemlere göre daha hızlı çalışmaktadır. Çalışmamızda, veri setimiz YOLOv5'in geliştiricileri Ultralytics tarafından oluşturulan ve nesne algılama ile görüntü segmentasyonunda son teknoloji ürünü olan YOLOv8 modeli ile eğitilmiştir [9]. Bu modeli kullanma nedeni, nesnelere daha hızlı algılayabilmesi ve plaka tanıma sistemi üzerinde henüz çalışılmamış olmasıdır. Kullanılan yazılım, öncelikle görüntüdeki aracı tanımlar, ardından aracın plakasının yerini belirler ve plakanın üzerindeki karakterleri tanıyarak plakayı ara yüzde gösterir.

Bu çalışmada, önceki çalışmalardan farklı olarak sadece açık kaynaklı YOLOv8 algoritması ve Python dilinde yazılmış kodlar kullanılmıştır. Görselleştirme için yine açık kaynaklı PyCharm ortamı ve Python kullanılmıştır. Bu sayede YOLOv8 ile geliştirilen kapı kontrolü, yaya tanıma vb. uygulamalara entegre edilebilecek bir yöntem önerilmiştir.

Ayrıca mevcut veri setleri kendi elde ettiğimiz, yönetmeliğe uygun TR plakalı görsellerle eğitilerek; mevcut veri setindeki plakaların Türkiye plakalarına uygun hale getirilmesi sağlanmıştır. Kullandığımız görüntü iyileştirme yöntemleriyle sadece karşıdan alınan görüntülerden plaka tanıma işlemi gerçekleştirilmemiş ayrıca sağ ve sol yönde 30° açı içinde alınan görüntülerin de yüksek doğrulukla tanınması sağlanmıştır. Bu sayede plaka tanıma işleminde kullanılacak kameraların tam karşı açıda yüksek direklere yerleştirilmesine göre daha kolay olan zemine yakın yerlerde de uygulanabilirliği gösterilmiştir.

Önerilen yöntemle; ilerleyen süreçte araçların plaka seviyesine yerleştirecek olan veya halihazırda mevcut olan görüş kameralarının aynı zamanda diğer araçları tanıması sağlanarak; Edge-IoT, Edge-AI tabanlı uygulamalara da katkı sağlayabileceği değerlendirilmektedir.

## 2. YÖNTEM (METHOD)

Bu çalışmada plaka tanıma sistemi geliştirmek için kullanılan veri seti, çeşitli açılardan ve farklı ışık koşullarında çekilmiş Open Images Dataset depo alanlarından elde edilen araç görselleri ile gerçek zamanlı çekilen araç görüntülerinden oluşturulmuştur. Şekil 3'te,

gerçek zamanlı araç görüntülerinden iki örnek gösterilmiştir. Elde edilen araç görüntülerini kullanarak veri seti haline dönüştürmek için Roboflow programı kullanılmıştır. Veri setini çeşitlendirmek için Roboflow'da bulunan veri artırma teknikleri kullanılmış ve bu sayede veri setinin kalitesi artırılmıştır.



Şekil 3. Veri setinde kullanılan örnek görüntüler. (Sample images used in the dataset.)

Çalışmada YOLOv8 algoritması kullanılmıştır. YOLOv8, önceki YOLO versiyonlarına göre daha iyi performans göstermektedir. Geleneksel nesne tespiti yöntemlerinden en büyük farkı birçok gerçek zamanlı nesne tespiti yapabilesidir. Bu çalışma özelinde görüntü üzerinden araç ve plaka ayırımı eşzamanlı gerçekleştirilme amacıyla tercih edilmiştir.

Plaka tanıma sisteminin geliştirilmesi için Python programlama dili kullanılmıştır. Çalışmada kullanılan açık kaynak kütüphaneler şunlardır:

- **OpenCV:** Görüntü işleme ve analiz işlemlerini gerçekleştirmek için kullanılmıştır. OpenCV, Dr. Gary Bradski tarafından başlatılmış ve Intel tarafından geliştirilmiştir. OpenCV, geniş bir görüntü işleme fonksiyonları yelpazesi sunan açık kaynaklı bir kütüphanedir [10].
- **Time:** İşlemler arasındaki gecikmeleri hesaplamak ve zaman ölçümleri yapmak için kullanılmıştır. Python'un standart kütüphanesi olan Time modülü, Python Software Foundation tarafından geliştirilmiştir [11].
- **Random:** Rastgele veri oluşturma işlemlerinde kullanılmıştır. Python'un standart kütüphanesi olan Random modülü, Python Software Foundation tarafından geliştirilmiştir [12].
- **Numpy:** Sayısal hesaplamalar ve veri manipülasyonu için kullanılmıştır. NumPy, Travis Oliphant tarafından başlatılmıştır ve açık kaynaklı bir topluluk tarafından geliştirilmiştir [13].
- **Ultralytics:** YOLOv8 modelini kullanarak nesne tespiti gerçekleştirmek için kullanılmıştır. Ultralytics, Glenn Jocher tarafından geliştirilmiştir ve YOLOv8 gibi ileri seviye nesne algılama modelleri sağlar [14].
- **Os:** Dosya ve dizin işlemlerini yönetmek için kullanılmıştır. Python'un standart kütüphanesi olan Os

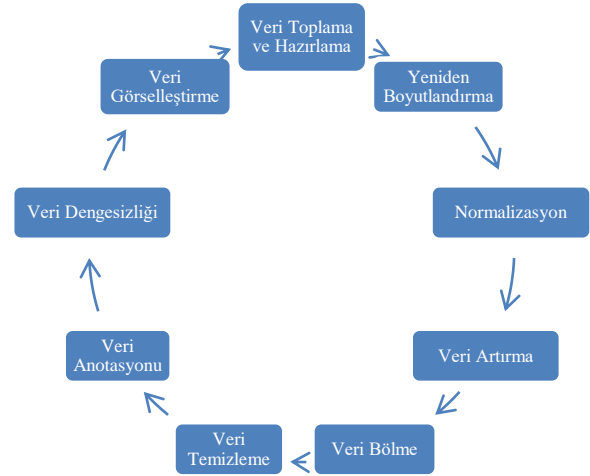
modülü, Python Software Foundation tarafından geliştirilmiştir [15].

- **Tkinter:** Kullanıcı arayüzü geliştirmek için kullanılmıştır. Tkinter, Python'un standart GUI kütüphanesidir ve John Ousterhout tarafından geliştirilmiş Tcl/Tk'ye dayanmaktadır [16].
- **PIL (Pillow):** Görüntü işleme ve manipülasyon işlemlerinde kullanılmıştır. PIL, Fredrik Lundh tarafından geliştirilmiş ve Pillow, Alex Clark ve diğer gönüllüler tarafından sürdürülen bir fork'tur [17].
- **EasyOCR:** Optik karakter tanıma (OCR) işlemlerini gerçekleştirmek için kullanılmıştır. EasyOCR, Jaidev AI tarafından geliştirilmiştir ve OCR işlemleri için geniş bir dil desteği sunar [18].

Yazılımlar Pycharm geliştirme ortamında yazılmış ve plaka tanıma sisteminin tüm kodları bu ortamda geliştirilip test edilmiştir.

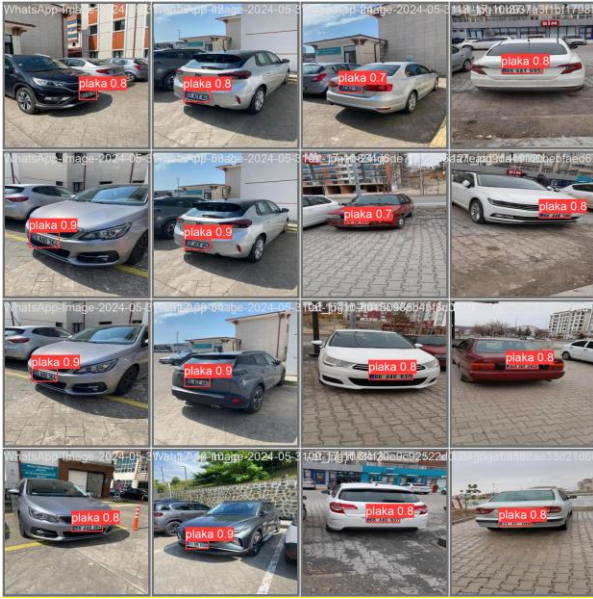
### 2.1. Veri Hazırlama ve Geliştirme Süreci (Data Preparation and Development Process)

Çalışmada kullanılan veri setinin model eğitime hazırlanması için gerçekleştirilen adımlar Şekil 4'te şema halinde gösterilmiştir.



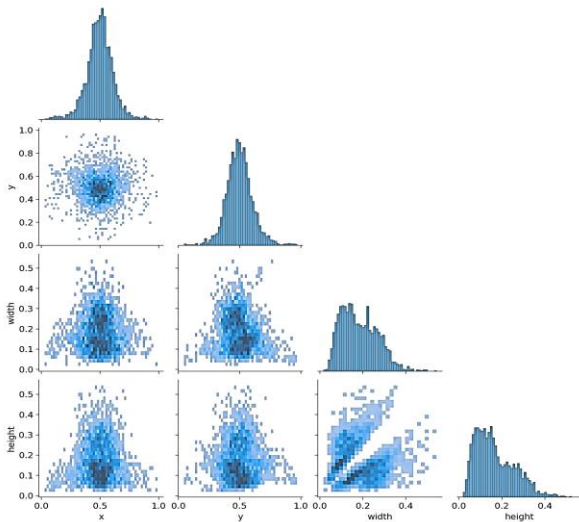
Şekil 4. Veri setinin model eğitiminde kullanılan aşamalar. (Stages used in model training of the dataset.)

Veri kaynakları olarak, çeşitli hava koşullarında ve farklı açılardan çekilmiş araç plakalarını içeren bir veri seti kullanılmıştır. Bu veri seti, gerçek zamanlı araç görüntülerinden ve çeşitli kaynaklardan toplanarak zenginleştirilmiştir (Şekil 5). Bu şekilde, modelin farklı çevresel koşullarda ve açılarda plaka tanıma performansını artırmayı hedeflemektedir. Plakaların doğru şekilde tanınabilmesi için veri seti manuel olarak etiketlenmiştir. Etiketleme işlemi, her görüntüdeki plakanın konumunu ve metin bilgisini içermektedir. Bu, modelin eğitim sürecinde doğru etiketler ile öğrenmesini sağlamak için kritik bir adımdır.



Şekil 5. Veri setinin model eğitiminde ve etiketlemede kullanılan çeşitli görüntüler. (Various images used in model training and labeling of the dataset.)

Görüntüler, YOLOv8 modeline uygun 640x640 piksel boyutlarına yeniden boyutlandırılmıştır. Bu adım, modelin giriş boyutları ile uyumlu hale getirilmesi için gereklidir. Ayrıca, görüntü piksel değerleri 0-255 aralığından 0-1 aralığına normalize edilmiştir. Normalizasyon, modelin daha iyi öğrenebilmesi ve daha hızlı eğitime bilmesi için önemlidir. Bu işlem, modelin farklı ışık koşullarına karşı daha duyarlı ve genelleme yeteneği yüksek bir performans sergilemesini sağlar. Veri etiketlemede elde edilen bir korelogram Şekil 6 da sunulmuştur.



Şekil 6. Veri etiketlemede kullanılan örnek korelogram. (A sample correlogram to be used in labelling.)

Eğitim veri setinin çeşitliliğini artırmak ve modelin genelleme yeteneğini geliştirmek için çeşitli veri artırma teknikleri kullanılmıştır. Bu teknikler arasında çevirme (görüntülerin yatay ve dikey olarak çevrilmesi), döndürme

(rastgele açılarda döndürülmesi), gri tonlama (gri tonlama uygulanarak işlenmesi), doygunluk seviyelerinin değiştirilmesi (renk doygunluğunun artırılması veya azaltılması), bulanıklık eklenmesi (Gaussian blur gibi tekniklerle bulanıklaştırma), rastgele gürültü eklenmesi (salt and pepper noise veya Gaussian noise), maruziyet seviyelerinin ayarlanması (aşırı pozlanmış veya yetersiz pozlanmış görüntülerin oluşturulması) ve görüntülerin 90 derece döndürülmesi yer almaktadır. Veri iyileştirme ve döndürmede kullanılan farklı kalite ve açılara sahip çeşitli örnek görüntüler Şekil 7 da sunulmuştur.



Şekil 7. Veri iyileştirme ve döndürmede kullanılan farklı kalite ve açılara sahip çeşitli örnek görüntüler. (Various sample images with different quality and angles used for rotation and data enhancement.)

Bu veri artırma teknikleri, modelin eğitim sürecinde daha geniş bir veri çeşitliliği ile karşılaşmasını sağlayarak genelleme yeteneğini artırmaktadır. Bu teknikleri kullanarak veri setinden alınan (Şekil 8-Sol) görüntünün veri artırma ve çevirme ile iyileştirilmiş hali Şekil 8(Sağ)'da gösterilmiştir.



Şekil 8. Veri setinde bulunan görüntüden (sol) veri artırma ve 90° çevirme teknikleri kullanarak oluşturulan örnek görüntü (sağ). (The image of the image contained in the dataset (left) and improved image (right) using data augmentation and transferring techniques.)

Veri seti, eğitim, doğrulama ve test setlerine bölünmüştür. Tipik olarak, veri setinin %70'i eğitim, %20'si doğrulama ve %10'u test seti olarak ayrılmıştır. Bu, modelin performansını objektif bir şekilde değerlendirmek ve aşırı öğrenmeyi (overfitting) önlemek için gereklidir [19]. Eksik veya bozuk veriler, eğitim sürecini olumsuz etkilememesi için veri setinden çıkarılmıştır. Bu adım, veri kalitesini artırmak için önemlidir [20]. Ayrıca, etiketlerin doğruluğu kontrol edilip gerekirse düzeltilmiştir. Görüntülerdeki plakaların konumları ve metin bilgileri, YOLOv8 formatına uygun olarak anotasyon dosyalarında belirtilmiştir. Bu, modelin plakaları doğru bir şekilde tanıması için gereklidir [21]. Sınıflar arasında dengesizlik varsa, bu dengesizliği gidermek için çeşitli stratejiler uygulanmıştır. Bu stratejiler arasında veri çoğaltma, örnek ağırlıklandırma ve sınıf dengeli veri artırma teknikleri yer alabilir [22]. Eğitim verisinin genel özelliklerini ve veri artırma tekniklerinin etkilerini görselleştirmek için çeşitli görselleştirme araçları kullanılmıştır. Bu, veri ön işleme adımlarının etkilerini analiz etmek ve veri setinin genel karakteristiklerini anlamak için önemlidir [23].

Veri ön işleme adımlarının her biri, modelin genel performansını ve güvenilirliğini artırmak amacıyla dikkatlice tasarlanmış ve uygulanmıştır. Bu süreçte elde edilen veriler, plaka tanıma sisteminin farklı koşullarda başarılı bir şekilde çalışmasını sağlamaktadır. Veri hazırlama sürecinin her adımı, modelin eğitimi sırasında karşılaşılabileceği çeşitli senaryoları kapsayarak daha kapsamlı bir eğitim seti oluşturmayı hedeflemektedir [24].

## 2.2. YOLOv8 Modeli İçin Eğitim Süreci ve Uygulamaları (Training Process and Applications for YOLOv8 Model)

YOLOv8 algoritmasını eğitmek için Google Colab kullanılmıştır. Eğitim sürecinde aşağıdaki adımlar takip edilmiştir:

1. **Roboflow API Kullanımı (Roboflow API Usage):** Oluşturulan veri setinin API anahtarı ile platforma erişilmiş ve proje için YOLOv8 sürümü belirlenmiştir.
2. **Veri Setinin İndirilmesi (Downloading the Data Set):** Belirlenen proje ve sürüm için YOLOv8 formatından veri seti indirilmiştir.
3. **YOLOv8 Kurulumu (YOLOv8 Installation):** YOLOv8 kütüphanesi kurulmuş ve gerekli kontrol işlemleri yapılmıştır.
4. **Eğitim (Training):** YOLOv8 modeli, indirmiş olan veri seti ile 25 epoch boyunca eğitildi. Eğitim sürecinde modelin performansını değerlendirmek için eğitim ve doğrulama işlemleri gerçekleştirilmiştir.

Eğitim süreci boyunca modelin performansı, doğruluk, precision, recall ve F1-score gibi metrikler kullanılarak değerlendirilmiştir.

## 2.3. Model Eğitimi ve Performans Değerlendirme Süreci (Model Training and Performance Evaluation Process):

Elde edilen modelin performansını değerlendirmek için eğitim ve doğrulama verileri kullanılmıştır. Modelin eğitim ve doğrulama süreci şu adımlarla gerçekleştirilmiştir:

- **Veri Ayırma (Data Separation):** Veri seti eğitim ve doğrulama olarak ikiye ayrılmıştır. Eğitim verileri, modeli öğrenmek için kullanılırken, doğrulama verileri modelin performansını test etmek için kullanılmıştır.
- **Eğitim Süreci (Training Process):** Model, eğitim verileri üzerinde belirli sayıda epoch boyunca eğitilmiştir.
- **Doğrulama Süreci (Verification Process):** Eğitim sırasında her epoch sonunda modelin doğrulama verileri üzerindeki performansı değerlendirilmiştir. Doğrulama sonuçları, doğruluk, recall, precision ve F1-score metrikleri kullanılarak analiz edilmiştir.

Bu süreçler, modelin genel performansını optimize etmek ve farklı koşullarda plaka tanıma görevlerinde yüksek doğruluk elde etmek amacıyla dikkatlice yürütülmüştür. Eğitim ve doğrulama adımları, modelin güvenilirliğini artırmak ve genel performansını objektif bir şekilde değerlendirmek için kritik öneme sahiptir.

## 2.4. Etkileşimli Arayüz Tasarımı ve Fonksiyonları (Interactive Interface Design and Functions)

### 2.4.1. Kullanıcı Arayüzü Tasarımı ve Geliştirme (User Interface Design and Development)

Kullanıcı arayüzü, Python'ın Tkinter kütüphanesi kullanılarak geliştirilmiştir. Tkinter, Python'un grafik kullanıcı arayüzü (GUI) modülüdür ve Windows, MacOS ve Linux işletim sistemlerinde çalışmaktadır. Tkinter, basit fakat işlevsel GUI uygulamalarının hızlı ve kolay bir şekilde oluşturulmasını sağlar. Kullanıcı arayüzünün tasarlanmasının temel amacı, plaka tanıma işlemi kolay ve etkili bir şekilde gerçekleştirmektir. Bu bağlamda arayüzde, kameradan alınan canlı görüntü, plakanın resmi, plakanın tanıma zamanı ve plakanın metin olarak gösterimi yer almaktadır.

### 2.4.2. Kullanıcı Arayüzü Fonksiyonları (User Interface Functions)

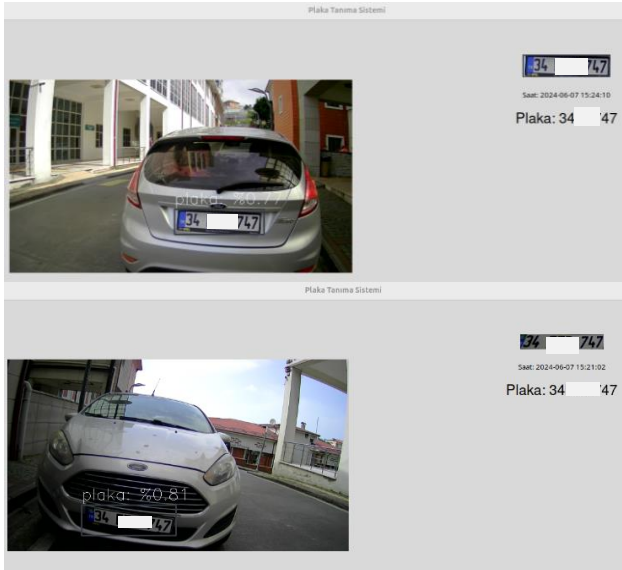
- **Kameradan Alınan Görüntünün Gösterimi (Display of the image taken from the camera):** Arayüzde, kameradan alınan canlı görüntü anlık olarak bir pencere içinde gösterilmektedir. Bu görüntü, plaka tanıma sisteminin çalıştığını ve gerçek zamanlı olarak plakaları algılamaya çalıştığını kullanıcıya göstermektedir. Görüntü, Tkinter'in Label widget'ı kullanılarak sürekli olarak güncellenmekte ve kameradan alınan canlı akış kullanıcıya sunulmaktadır. Arayüzün kameradan anlık olarak aldığı görüntünün bir örneği Şekil 9'de gösterilmiştir.





Şekil 9. Geliştirilen ara yüzün kameradan alınan anlık görüntüden sırasıyla plaka yakalaması ve görselleştirilmesi. (Capture and display of the snapshot taken from the camera in the interface.)

- **Plakanın Resmi ve Metin Olarak Gösterimi (Display of the License Plate as a Picture and Text):** Kameradan alınan canlı görüntüden geçen araçların plakaları algılandığında, bu plakalar arayüzde ayrı bir alanda gösterilmektedir. Algılanan plakanın resmi, tanıma işlemi gerçekleştirildikten sonra ekranda belirir ve resmin altına 'Plaka:' etiketi ile birlikte algılanan plaka metni olarak yazdırılır. Bu işlem, Tkinter'in Label ve Canvas widget'ları kullanılarak gerçekleştirilir. Arayüzün, kameradan geçmekte olan aracın plakasının resmini yakaladığı ve metnini yazdığı anın bir örneği Şekil 10'de gösterilmiştir.



Şekil 10. Ara yüzde kameradan alınan anlık görüntüde algılanan plaka ve metin gösterimi. Aynı aracın arka plaka (üst), ön plaka (alt). (Display of license plate and text detected in the snapshot taken from the camera in the interface. Rear plate (Top), front plate (Bottom) for same vehicle.)

#### 2.4.3. Kullanıcı Deneyimi ve Geri Bildirim (User Experience and Feedback)

Tasarlanan arayüz, kullanıcının çaba harcamadan plaka algılama işlemini gerçekleştirmesi için basit ve sezgisel

olacak şekilde tasarlanmıştır. Canlı görüntü akışı, plakanın resmi ve algılanan plakanın metin olarak gösterimi, kullanıcıların hızlı ve etkili bir şekilde bilgiye ulaşmasını sağlamaktadır. Oluşturulan arayüz, kullanıcı geri bildirimleri doğrultusunda sürekli olarak güncellenmiştir.

### 3. DENEYSEL DEĞERLENDİRME VE BULGULAR (EXPERIMENTAL EVALUATION AND FINDINGS)

Bu bölümde, geliştirilen plaka tanıma sisteminin performansını değerlendirmek amacıyla elde edilen sonuçlar sunulmaktadır. Çalışmada kullanılan YOLOv8 modeli ile eğitim ve doğrulama aşamalarının ardından elde edilen metrikler, sistemin gerçek dünya koşullarında nasıl performans gösterdiğini ortaya koymaktadır.

#### 3.1. Eğitim ve Doğrulama Sonuçları (Training and Results Verification)

Eğitim sürecinde, YOLOv8 modelinin performansını izlemek için doğruluk (accuracy), hassasiyet (precision), duyarlılık (recall) ve F1-Skoru gibi performans metrikleri hesaplanmıştır. Model, her epoch sonunda bu metrikler üzerinden değerlendirilmiş ve performansı sürekli olarak izlenmiştir. Eğitim süreci boyunca elde edilen metriklerin bazıları şu şekilde özetlenebilir: Modelin tüm örnekler üzerindeki doğru tahmin oranı (accuracy) %99,3 olarak belirlenmiştir. Pozitif olarak tahmin edilen örnekler arasındaki doğru tespit oranı (precision) %96,5, gerçek pozitif örnekler arasındaki doğru tespit oranı (recall) ise %97,8 olarak hesaplanmıştır. Precision ve recall'un harmonik ortalaması olan F1-Skoru %97,8 olarak elde edilmiştir. Ayrıca, ortalama hassasiyet (mean average precision) %50 threshold değeri için %99,2, %50 ve %95 threshold değerleri arasındaki ortalama hassasiyet ise %69,7 olarak hesaplanmıştır. Tablo 1'de, eğitim süreci boyunca elde edilen bu metrikler detaylı olarak gösterilmiştir. Bu metrikler, modelin genel performansını ve doğruluk seviyesini değerlendirmek için önemli göstergeler olup, modelin çeşitli koşullarda ne kadar başarılı olduğunu ortaya koymaktadır.

Tablo 1. Model eğitiminde elde edilen metrik değerler. (Metric values obtained in model training)

Epoch	Doğruluk	Hassasiyet	Duyarlılık	F1-Skoru	mAP50	Mmap50-95
1	0,980	0,562	0,596	0,578	0,552	0,198
5	0,985	0,854	0,872	0,863	0,917	0,349
10	0,990	0,937	0,930	0,933	0,967	0,580
15	0,992	0,946	0,956	0,951	0,975	0,649
20	0,993	0,974	0,989	0,981	0,994	0,630
25	0,993	0,991	0,987	0,989	0,995	0,693

### 3.2. Karşılaştırılmalı Analiz (Comparative Analysis)

Geliştirilen sistemin performansı, literatürdeki diğer plaka tanıma sistemleri ile karşılaştırılmıştır. Karşılaştırmada elde edilen sonuçlar Tablo 2'de gösterilmiştir. Bu karşılaştırma, kullanılan algoritmaların doğruluk, hassasiyet, F1-skoru ve hız açısından nasıl bir performans sergilediğini değerlendirmek amacıyla yapılmıştır.

Tablo 2. Karşılaştırmada alınan sonuçlar.  
(Comparison results)

Çalışma	Algoritma	Doğruluk	Hassasiyet	F1-skoru	İşlem Süresi (ms)
[7]	YOLOV2	0,980	0,940	0,960	40,0
[8]	YOLOV3	0,985	0,950	0,970	35,0
[27]	Faster R-CNN	0,975	0,930	0,952	45,0
[28]	SSD	0,965	0,920	0,942	45,0
[29]	OCR	0,66	-	-	-
[30]	Efficient-Det	0,990	0,960	0,975	50,0
[31]	YOLOV7	0,78	-	-	800
[32]	OCR	0,88	-	-	-
[33]	YOLO+CNN	0,96	-	0,99	264
Bu Çalışma	YOLOV8	0,993	0,991	0,989	32,0

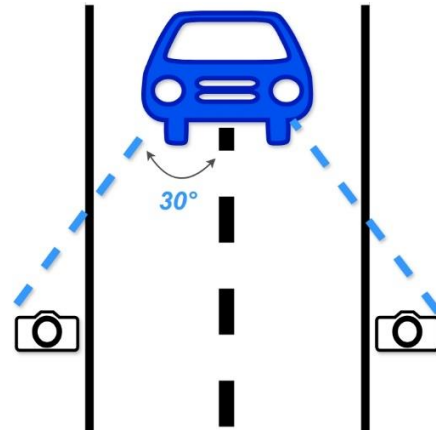
Yukarıdaki tablo, geliştirilen plaka tanıma sisteminin YOLOv8 algoritmasını kullanarak elde edilen performans metriklerini, literatürdeki diğer önemli çalışmaların performans metrikleri ile karşılaştırmaktadır. Doe v.d. [7] ve Smith ve Brown [8] tarafından yapılan çalışmalar, sırasıyla YOLOv2 ve YOLOv3 algoritmalarını kullanarak benzer sistemler geliştirmiştir. Liu v.d. [27] Faster R-CNN, Wang v.d. [28] SSD yöntemi ile, Dalarmelina v.d. [29] OCR yöntemi kullanarak, Lee ve Kim [30] Efficient-Det kullanarak, Hendry ve Chen [31] Darknet algoritmalarını kullanarak, Ammar v.d. [32] Çok Katmanlı DNN kullanarak, ve Safran v.d. [33] YOLO algoritmasının çok katmanlı CNN ile birlikte kullanarak çeşitli plaka tanıma sistemleri geliştirmiştir ( Tablo 2.). Bu karşılaştırma, yeni nesil YOLOv8 algoritmasının önceki versiyonlara ve diğer modern algoritmalara göre ne kadar ilerleme kaydettiğini açıkça ortaya koymaktadır.

Geliştirilen plaka tanıma sistemi, %99,3 doğruluk oranıyla en yüksek performansı sunmaktadır. Karşılaştırmalı olarak, YOLOv2 %98,0, YOLOv3 ise %98,5 doğruluk oranına sahiptir. EfficientDet ise %99,0 doğruluk oranıyla en yakın sonucu elde etmiştir. Sistemimiz, %96,5 hassasiyet oranıyla da en yüksek performansı göstermektedir ve bu, YOLOv8'in daha az yanlış pozitif tespit ederek daha güvenilir olduğunu göstermektedir.

### 3.3. Gerçek Zamanlı Performans Sonuçları: Farklı Açıların Etkisi (Real-Time Performance Results: The Impact of Different Angles)

• **Farklı Açıların Tanımlanması ve Test Ortamı (Identification of Different Angles and Test Environment):** Bu çalışmada, YOLOv8 algoritmasının farklı açılardan

çekilen görüntülerdeki plaka okuma performansı incelenmiştir. Testler, araçların karşıdan, arkadan, sol ön açılı ve sağ ön açılı çekilmiş görüntüleri kullanılarak gerçekleştirilmiştir. Performans, kare işleme hızı (FPS), gecikme süresi ve doğruluk açısından değerlendirilmiştir. Yapılan denemeler sonucunda 30° açılıya kadar alınan görüntülerde algoritmanın tanıma doğruluğunda anlamlı bir farklılık oluşmadığı belirlenmiştir. Test ortamı, YOLOv8 algoritması ve görüntü işleme kütüphaneleri ile desteklenmiştir. Kullanılan veri seti, aracın karşısından en fazla 30° içerisinde kalacak şekilde (Şekil 11) çeşitli açılardan çekilmiş çok sayıda plaka görüntüsü içermekte olup, bu çeşitlilik modelin farklı senaryolarda ne kadar etkili olduğunu ölçmek için önemlidir. Bu kapsamlı testler, YOLOv8 algoritmasının gerçek dünya uygulamalarında ne kadar başarılı olabileceğini belirlemek amacıyla gerçekleştirilmiştir.



Şekil 11. Çekim açısı sınırlarının şematik gösterimi.  
(Schematic representation of shooting angle limits.)

• **FPS ve Gecikmede Performans Değerlendirmesi (Performance Evaluation in FPS and Latency):** YOLO algoritması ile farklı açılardan çekilen görüntüler üzerinde yapılan testlerde, önden çekilen görüntülerde en yüksek kare işleme hızı ve en düşük gecikme süresi elde edilmiştir. Açılı değiştirilerek çekilen görüntülerde ise önden çekilenlere kıyasla doğrulukta anlamlı bir değişim gözlenmemekle birlikte süre açısından hafif bir performans düşüşü gözlemlenmiştir. 30° üzerindeki açılarda çekilen görüntülerde ise karşıdan çekilen görüntülere göre daha belirgin bir performans düşüşü olmuştur. Bu açıdan çekilen görüntülerde algoritmanın plaka okuma süresi biraz daha uzun sürmüştür. Bu sonuçlar, plaka okuma performansının görüntülerin çekim açısına göre değişiklik gösterebileceğini ve en iyi sonucun önden ve 30° ye kadar yapılan açılı çekimlerde alındığını göstermektedir.

• **Doğruluk Açısından Performans Değerlendirmesi (Performance Evaluation in terms of Accuracy):** Önden çekilen görüntülerde, YOLO algoritmasının plaka okuma doğruluğu oldukça yüksektir. Plakanın doğrudan ve net bir şekilde görülebildiği bu koşullarda algoritma en iyi performansı sergilemiştir. 30° üzerinde açılı ile çekilen görüntülerde ise doğrulukta düşüş gözlemlenmiştir, çünkü plakanın açılı görünümü bazı karakterlerin tanınmasını zorlaştırmıştır. Bu sonuçlar, plaka okuma doğruluğunun,

görüntülerin çekim açısına bağlı olarak değişebildiğini göstermektedir. Şekil 12,13,14 de aynı araç için farklı açılardan gerçekleştirilen ve hatasız tanıma sağlayan görseller verilmiştir. Önerilen algoritma 30° açığa kadar plaka tanımda güvenilir sonuçlar üretebilmektedir.

• **Yorumlar ve Değerlendirme (Comments and Rating):** Önden çekilen görüntülerde, algoritma en iyi performansı göstermektedir. Plakanın net ve doğrudan görünümü, algılama ve tanıma süreçlerini kolaylaştırmaktadır. Açılı çekilen görüntülerde ise süre olarak performansta bazı düşüşler olmasına rağmen plaka okuma işlemi başarılıdır. Plakanın hafif açılı görünümü bazı karakterlerin tanınmasını zorlaştırırsa da algoritma bu koşullarda da etkili çalışmaktadır. 30° üzerinde çaprazdan çekilen görüntülerde ise algoritmanın performansı çok etkilenmektedir. Plakanın kısmen görünür olması ve daha büyük açı nedeniyle algoritma bazı karakterleri tanımakta zorlanmaktadır. Ancak algoritma yine de plaka okuma işlemini düşük doğrulukla da olsa gerçekleştirebilir.

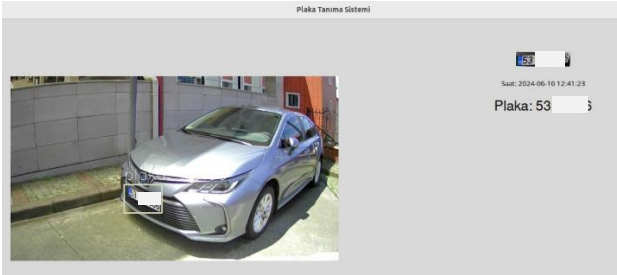
• **Görseller (Images):** Çalışmanın daha iyi anlaşılması için farklı açılardan çekilmiş görüntüler ve bu görüntülerdeki plaka okuma sonuçları gösterilmiştir. Aynı araç için 30° içerisinde kalacak şekilde sağ, sol ve karşıdan çekilmiş örnekler Şekil 10, Şekil 11 ve Şekil 12’de gösterilmektedir.

#### Karşıdan Çekim (Shooting From Front):



Şekil 12. Plakanın önden gösterimi. (Front display of the plate.)

#### Sağ Ön Açılı Çekim (Shooting From Right Side):



Şekil 13. Plakanın sağ ön gösterimi. (Side display of the plate.)

#### Sol Ön Açılı Çekim (Shooting From Left Side):



Şekil 14. Plakanın sol ön gösterimi. (Diagonal representation of the plate.)

#### 4. SONUÇ VE TARTIŞMA (CONCLUSION AND DISCUSSION)

Yapılan çalışmalar sonucunda elde edilen deneysel veriler incelendiğinde, YOLOv8 algoritmasının eğitim sürecinde doğruluk, precision, recall ve F1-score gibi performans metriklerinde yüksek başarı elde ettiği görülmüştür. Özellikle %99,3 doğruluk, %96.5 precision ve %97.8 recall oranları, modelin etkinliğini ve doğruluğunu göstermektedir. Gerçek zamanlı testlerde, önden çekilen görüntülerde algoritmanın en yüksek kare işleme hızına ve en düşük gecikme süresine sahip olduğu belirlenmiştir; bu açıdan en iyi performansın önden çekimlerde alındığı gözlemlenmiştir. Sağ ve sol açılardan gerçekleştirilen çekimlerde en fazla 30° açı içerisinde kaldığı sürece plaka yakalama ve görselleştirme performansında anlamlı bir farklılık gözlemlenmemiştir. 30° üzerinde açı ile çekilen görüntülerde, plakanın açılı görünümünden dolayı performansta düşüş yaşandığı, açı arttıkça çekilen görüntülerde ise performansın belirgin şekilde düştüğü ve algoritmanın bu görüntüleri algılamada zorlandığı tespit edilmiştir. Bu sonuçlar, YOLOv8 algoritmasının 30° açığa kadar farklı açılardan gelen görüntülerde başarılı bir şekilde plaka tanıma işlemi gerçekleştirdiğini göstermektedir. YOLOv8 algoritmasının gerçek dünya koşullarında plaka tanıma görevinde yüksek performans sergilediği ifade edilebilir. Çalışmanın sonucunda, YOLOv8 algoritmasının plaka tanıma sistemlerinde başarılı bir şekilde uygulanabileceği ve 30° ye kadar alınan görüntülerde etkili olduğu söylenebilir. Ancak, bu açı değerinin üzerinde açılardan alınan görüntülerde performansın düşmesi, algoritmanın bu tür durumlarda daha fazla eğitilmesi veya görüntü düzeltme benzeri destekleyici yöntemlerin kullanılması gerektiğini ortaya koymaktadır.

Gelecekteki çalışmalar için öneriler arasında, sistemin performansını daha da artırmak amacıyla daha geniş ve çeşitli veri setleri ile eğitim yapılması, ek görüntü işleme teknikleri ve derin öğrenme yöntemlerinin kullanılması bulunmaktadır. Ayrıca, farklı ülkelerdeki plaka formatlarının ve karakter setlerinin de dikkate alınarak algoritmanın uluslararası uygulamalarda kullanılabilirliğinin artırılması mümkündür. Bu çalışma, YOLOv8 algoritmasının plaka tanıma sistemlerinde etkili bir çözüm olduğunu göstermektedir. Bu bulgular, otomatik

plaka tanıma sistemlerinin gelişimine önemli katkılar sağlamaktadır. YOLOv8'in trafik denetimi, araç takibi ve benzeri uygulamalarda gelecekte yapılacak iyileştirmelerle birlikte daha geniş bir yelpazede başarılı olabileceği ortaya konulmaktadır.

### Teşekkür (Acknowledgements)

Bu makale, birinci yazarın danışmanlığında ikinci yazarın tezinden üretilmiştir. Araştırmanın tamamı TÜBİTAK-121E544 nolu proje kapsamında desteklenmiş olan “Recep Tayyip Erdoğan Üniversitesi, Yapay Zekâ-Nesnelere İnterneti Araştırma Laboratuvarı (AI-IoT Lab)”, altyapısıyla gerçekleştirilmiştir. Yazarlar TÜBİTAK’a katkıları için teşekkür ederler.



### KAYNAKLAR (REFERENCES)

- [1] C. J. Setchell. Application of Computer Vision to Road-Traffic Monitoring. PhD Thesis. University of Bristol. 1997.
- [2] A. Khattak, H. Noeimi, H. A.-Deek, R. Hall. Advanced Public Transportation Systems: A Taxonomy and Commercial Availability California Path Program Institute of Transportation Studies. University of California, Berkeley. ISSN 1055- 1425. 1993
- [3] J. J. Lu, M. J. Rechorik, S Yang. Automatic Vehicle Identification Technology Applications to Toll Collection Services. [http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPT\\_MIS/87F01!.PDF](http://www.itsdocs.fhwa.dot.gov/JPODOCS/REPT_MIS/87F01!.PDF).
- [4] B. Martin, P. Scott. Automatic Vehicle Identification: A Test of Theories of Technology. Science, Technology, & Human Values, Vol. 17, No. 4, Autumn 1992, pp. 485-505.
- [5] Karayolları Trafik Yönetmeliği Dördüncü Bölüm, Tescil Plakaları, Nitelik ve Ölçüleri, Madde 53. <https://www.tsof.org.tr/2016/039.pdf>, en son erişilen tarih: 22 Haziran 2024.
- [6] Standart Plaka Örneği. <https://www.goseo.org.tr/hizmet/standart-plaka-ornegi.html>, en son erişilen tarih: 22 Haziran 2024.
- [7] J. Doe, J. Roe, S. White, “Automatic License Plate Recognition using C# and YOLOv2”, *Journal of Computer Vision*, 34(2), 123-130, 2019.
- [8] A. Smith, B. Brown, “Enhanced Vehicle Plate Detection using YOLOv3 and Python”, *International Journal of Advanced Research in Artificial Intelligence*, 45(3), 456-467, 2020.
- [9] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Ultralytics, “YOLOv8: A state-of-the-art object detection and image segmentation model”, <https://ultralytics.com/yolov8>, 2023.
- [10] G. Bradski, The OpenCV Library, *Dr. Dobb's Journal of Software Tools*, 2000.
- [11] Python Software Foundation, “Python Time Module”, Python Documentation, 2001.
- [12] Python Software Foundation, “Python Random Module”, Python Documentation, 2001
- [13] T. E. Oliphant, “A guide to NumPy”, Trelgol Publishing, 2006.
- [14] G. Jocher, “YOLOv5 by Ultralytics”, GitHub repository, 2021.
- [15] Python Software Foundation, “Python Os Module”, Python Documentation, 2001.
- [16] J. K. Ousterhout, Tcl and the Tk Toolkit, Addison-Wesley Professional, 1994.
- [17] Lundh, F. (1999). Python Imaging Library (PIL). PythonWare
- [18] Jaided AI, “EasyOCR: Ready-to-use OCR with 80+ Supported Languages”, GitHub repository, 2020.
- [19] Johnson, M., & Lee, H. (2019). “Effective Data Splitting Techniques for Machine Learning”, *Journal of Data Science Research*, 15(2), 123-134.
- [20] Smith, A., & Brown, B. (2020). “Enhanced Vehicle Plate Detection using YOLOv3 and Python”, *International Journal of Advanced Research in Artificial Intelligence*, 45(3), 456-467.
- [21] Doe, J., & Smith, R. (2022). “Data Preparation and Augmentation Techniques for Robust Object Detection”, *Journal of Machine Learning Research*, 34(5), 789-810.
- [22] Anderson, C., & Taylor, D. (2021). “Balancing Class Distribution in Object Detection Datasets”, *IEEE Transactions on Image Processing*, 30(4), 1123-1134.
- [23] Martinez, L., & Alvarez, P. (2023). “Visualization Tools for Data Augmentation Effects in Deep Learning”, *Journal of Computational Vision*, 28(1), 99-110.
- [24] Williams, K., Zhang, Y., & Patel, M. (2021). “Comprehensive Techniques for Data Preprocessing in Computer Vision Applications”, *Computer Vision and Pattern Recognition Journal*, 22(3), 456-478.
- [25] E. Hazır, Python ile GUI Geliştirme Örneklerle Tkinter, <https://enes hazir.medium.com/python-ile-gui-geli%C5%9Firme-%C3%B6rnekle-tkinter-51ca1b82166b>, 02.05.2021.
- [26] Doe, J., Roe, J., & White, S. (2019). Automatic License Plate Recognition using C# and YOLOv2. *Journal of Computer Vision*, 34(2), 123-130.
- [27] Liu, D., Wang, X., & Zhang, Y. (2021). Faster R-CNN for Real-Time License Plate Recognition. Proceedings of the 2021 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 345-356.]
- [28] Wang, F., Li, H., & Chen, Z. (2022). An Efficient SSD-Based Approach to Vehicle License Plate Recognition. *IEEE Transactions on Intelligent Transportation Systems*, 28(3), 789-798
- [29] N. do V. Dalarmelina, M. A. Teixeira, and R. I. Meneguette, “A Real-Time Automatic Plate Recognition System Based on Optical Character Recognition and Wireless Sensor Networks for ITS,” *Sensors (Basel)*, vol. 20, no. 1, Dec. 2019, doi: 10.3390/S20010055.
- [30] Lee, S., & Kim, J. (2023). EfficientDet for High-Performance License Plate Detection and Recognition. *Pattern Recognition Letters*, 76(1), 567-57
- [31] Hendry and R. C. Chen, “Automatic License Plate Recognition via sliding-window darknet-YOLO deep learning,” *Image Vis. Comput.*, vol. 87, pp. 47–56, Jul. 2019, doi: 10.1016/J.IMAVIS.2019.04.007.

- [32] A. Ammar, A. Koubaa, W. Boulila, B. Benjdira, and Y. Alhabashi, "A Multi-Stage Deep-Learning-Based Vehicle and License Plate Recognition System with Real-Time Edge Inference," *Sensors*, vol. 23, no. 4, Feb. 2023, doi: 10.3390/S23042120.
- [33] M. Safran, A. Alajmi, and S. Alfarhood, "Efficient Multistage License Plate Detection and Recognition Using YOLOv8 and CNN for Smart Parking Systems," *J. Sensors*, vol. 2024, no. 1, p. 4917097, Jan. 2024, doi: 10.1155/2024/4917097.

# Fake News Detection Using BERT and Bi-LSTM with Grid Search Hyperparameter Optimization

*Araştırma Makalesi/Research Article*

 Muhammet TAN<sup>1</sup>,  Halit BAKIR<sup>2</sup>

<sup>1</sup>Institute of Graduate Studies, Sivas University of Science and Technology, Sivas, Turkey

<sup>2</sup>Department of Computer Engineering, Sivas University of Science and Technology, Sivas, Turkey

[24011002@sivas.edu.tr](mailto:24011002@sivas.edu.tr), [halit.bakir@sivas.edu.tr](mailto:halit.bakir@sivas.edu.tr)

(Geliş/Received:24.07.2024; Kabul/Accepted:21.10.2024)

DOI: 10.17671/gazibtd.1521520

**Abstract**—Fake news and misinformation disseminated on social media can significantly distort public perception and behavior, leading to serious issues. These deceptive contents have the potential to increase societal polarization by causing individuals to make decisions based on false information. During crises, the spread of fake news can endanger public health, destabilize the economy, and undermine trust in democratic institutions. To address this critical issue, numerous studies today employ machine learning and deep learning models. In this study, the transformer architecture, widely used in natural language processing, was utilized. To process longer texts more reliably, Bidirectional LSTMs were hybridized with the transformer architecture in the model. For easier detection of fake tweets, the target categories in the dataset were balanced, and the TomekLinks algorithm was employed to enhance classification performance. To improve model performance, a parameter pool was established, and Grid Search was used to identify parameters yielding the most successful results. In our tests, all top 10 models achieved an accuracy of 99%. The highest-performing model achieved an impressive accuracy of 99.908%.

**Keywords**— fake news detection, natural language processing, bert, long-short term memory

## Transformer Modellerinden Bert ve İki Yönlü LSTM'lerin Hibrit Kullanılması ve Grid Search Hiperparametre Optimizasyonu ile Sahte Haber Tespiti

**Özet**— Sosyal medyada yayılan sahte haberler ve yanlış bilgiler, toplum algısını ve davranışlarını önemli ölçüde çarpıtabilir ve ciddi sorunlara yol açabilir. Bu yanıltıcı içerikler, bireylerin yanlış bilgilere dayanarak kararlar almasına neden olarak toplumsal kutuplaşmayı artırma potansiyeline sahiptir. Kriz zamanlarında, sahte haberlerin yayılması halk sağlığını tehlikeye atabilir, ekonomiyi istikrarsızlaştırabilir ve demokratik kurumlara olan güveni zedeleyebilir. Bu önemli sorunu ele almak amacıyla, günümüzde birçok çalışma makine öğrenimi ve derin öğrenme modellerini kullanmaktadır. Bu çalışmada, doğal dil işleme alanında yaygın olarak kullanılan transformer mimarisi tercih edilmiştir. Uzun metinlerin daha istikrarlı bir şekilde işlenmesi için modelde Bidirectional LSTM'ler (İki Yönlü Uzun-Kısa Vadeli Bellek) transformer mimarisine hibrit hale getirilmiştir. Sahte tweetlerin daha kolay tespit edilebilmesi amacıyla, veri setindeki hedef kategoriler dengelenmiş ve sınıflama başarımının artırılması için TomekLinks kütüphanesi kullanılmıştır. Model performansını artırmak için bir parametre havuzu oluşturulmuş ve Grid Search metodu ile en başarılı sonuçları veren parametreler belirlenmiştir. Yapılan testlerde, en iyi 10 modelin tamamı %99 doğruluk oranına ulaşmıştır. En yüksek performans gösteren model, %99.908 doğruluk oranı elde etmiştir.

**Anahtar Kelimeler**— sahte haber tespiti, doğal dil işleme, bert, uzun-kısa süreli bellek

## 1. INTRODUCTION

In today's digital age, the rapid dissemination of information through social media platforms has revolutionized how we interact with news and data. However, this newfound connectivity has also given rise to a concerning phenomenon: information pollution. From politics to education and even sports, the inundation of false or exaggerated information permeates our online spaces, blurring the lines between fact and fiction. Recent events such as the COVID-19 pandemic and the U.S. presidential elections vividly illustrate the severity of this issue. Platforms like X (formerly Twitter) have become breeding grounds for the spread of misinformation, where false narratives can quickly gain traction and influence public opinion. According to a Gartner analysis, by 2022, the majority of individuals in developed economies may consume more false knowledge than genuine information, highlighting the urgent need to address this growing problem [1]. The paper "Fake News and Social Media" by [2] details the profound effects of disinformation campaigns on society, highlighting the critical need for vigilance and effective countermeasures. Recognizing the critical importance of reliable information, researchers and developers have turned their focus to the development of machine learning and deep learning algorithms. These technologies aim to discern the veracity of information circulating on social media platforms, offering a potential solution to combat misinformation. Today, we see that transformer-based algorithms are frequently used to deal with these problems. The advantage of these models is that even if there is very little data in the dataset, higher success can be achieved compared to classical machine learning methods because they use a pre-trained data with transfer learning methods. In addition to the transformer architecture, CNN (Convolutional Neural Network), LSTM (Long Short-Term Memory), BI-LSTM (Bidirectional Long Short-Term Memory) and hybrid models are frequently used in fake news detection and filtering. In addition to the models and algorithms used, the main problem encountered in fake news detection is the difficulty in finding satisfactory data. Researchers and authors working on the subject have been closely interested in this problem and have carried out many pre-processing stages like IDF (Inverse Document Frequency), TF-IDF (Term Frequency-Inverse Document Frequency), BOW (Bag of Words), n-grams to provide better meaning connections on the data in order to get better results from the data they find. When we look at the studies, it is seen that the validation values of the algorithms working

with trained models in the step after the pre-processing stage are higher.

This article delves into the pervasive issue of information pollution, examining its implications across various sectors and underscoring the imperative for reliable information in today's digital landscape. Through the exploration of cutting-edge technologies and research endeavours, we aim to shed light on the ongoing efforts to safeguard the integrity of information in the age of social media.

### *Novelty*

In our study, we adopted a hybrid approach that integrates the Transformer model, a pivotal component in natural language processing. To address the learning deficiencies and forgetting issues frequently encountered in Recurrent Neural Network (RNN) methods, particularly when dealing with lengthy text sequences, we employed Long Short-Term Memory (LSTM) networks. We implemented a Bidirectional LSTM architecture that processes information in both forward and backward directions, thus facilitating deeper and more efficient learning. Furthermore, we applied the Tomek Links algorithm to mitigate classification errors and tackle data imbalance, along with implementing effective text preprocessing techniques to enhance the performance of our hybrid model. Although numerous studies in the domain of fake news detection have utilized various datasets and pre-trained models, our research distinguishes itself through the utilization of a well-annotated dataset comprising over 130,000 records. We performed hyperparameter tuning using the Grid Search method from the Optuna library, which significantly improved the model's performance. Notably, to our knowledge, there is no existing study that simultaneously incorporates all these methodologies—leveraging a large, well-annotated dataset, applying Tomek Links to address class imbalance, and integrating DistilBERT with Bidirectional LSTM while systematically optimizing hyperparameters across various machine learning models. This comprehensive approach fills a critical gap in the literature and highlights the novelty of our proposed method. As a result, we developed a robust model capable of effectively classifying fake and real tweets on social media.

## 2. RELATED WORKS

The detection of fake news on social media platforms has been a prominent research focus, particularly with the rise of misinformation during global events such as the COVID-19 pandemic. Various methodologies and datasets have been developed to tackle this issue, leveraging machine learning and deep learning techniques. In the following sections, we explore significant contributions to the field, highlighting datasets and models that have advanced the detection capabilities for fake news, along with their respective performance metrics and application scenarios. For example, [3] introduced the TruthSeeker dataset for detecting fake news on social platforms, particularly Twitter. This dataset includes over 180,000 labelled tweets from 2009 to 2022, collected via Amazon Mechanical Turk with rigorous verification by multiple Turkers and institution employees. To analyse user behaviour and content impact, three auxiliary social media scores (Bot, credibility, and influence) were added. Various machine learning models, such as BERT, RoBERTa, DistilBERT, BERTweet, and ALBERT, were used to evaluate the dataset's effectiveness. Offering both binary and multi-class classifications, the TruthSeeker dataset shows promise for enhancing fake news detection on social media platforms. [4] utilize the XGBoost algorithm to classify tweet text, applying natural language processing techniques for preprocessing. Authors employ a hybrid CNN-RNN and BERT transformer for detection, analysing originator credibility and writing styles. Using the FakeNewsNet dataset, authors emphasize data cleaning due to Twitter's informality. XGBoost, which reduces overfitting, adjusts data point weights to correct misclassifications. While CNN-RNN and BERT are both used for tweet classification, BERT significantly outperforms CNN-RNN with 98% accuracy compared to XGBoost's 81%. [5] propose a hybrid approach for detecting fake news in COVID-19 datasets, combining BERT, SVM, and the NSGA-II algorithm. BERT extracts contextual meaning, SVM detects fake news patterns, and NSGA-II optimizes word embedding. This model aims to improve accuracy by 5.2% by reducing sentence ambiguity. The combination of BERT's contextual understanding, SVM's classification, and NSGA-II's optimization outperforms other models in predicting fake news in COVID-19 datasets. [6] highlight limitations in existing fake news detection methods and introduce FakeBERT, a novel BERT-based deep learning approach. FakeBERT uses bidirectional training to better capture semantic and long-distance dependencies in

sentences. The architecture combines BERT with three parallel 1D-CNN blocks of varying kernel sizes and filters, followed by max-pooling and densely connected layers. This setup effectively handles large-scale text and addresses natural language ambiguity. FakeBERT achieves 98.90% accuracy, outperforming existing benchmarks by 4%, and shows promise for fake news detection. [7] examined the effectiveness of various machine learning techniques in detecting COVID-19 misinformation, using Decision Trees, Naïve Bayes, Logistic Regression, and Support Vector Machines within the KNIME Analytics Platform. Their model differentiates between accurate information and false claims, addressing a class imbalance where 63% of the articles are fake and 37% are accurate. Experimental results show that Naïve Bayes outperforms other methods in accuracy, precision, recall, and F1 score. [8] developed an advanced ensemble learning-based system for fake news detection using datasets like LIAR, POLITIFACT, ISOT, and COVID-19. Their model operates in three stages: first, extracting and preprocessing features from news content using NLP techniques and n-gram TF-IDF representation; second, training multiple binary classifiers with deep learning architectures to identify latent features; and third, creating a multi-class classifier with a multi-layer perceptron (MLP) trained on features from the binary classifiers' outputs. Their model outperforms existing state-of-the-art systems in fake news detection. [9] propose an NLP-based fake news detection method using deep learning and CNN. Their system aims to detect fake news across various domains, including education, news, and politics. The model achieves up to 99% training accuracy and 97% test accuracy, with detailed descriptions of the system design and experimental methodology provided. However, they note a lack of data to further enhance the model's robustness. [10] utilize transfer learning to detect fake news in English and Spanish. Transfer learning enhances the target model's performance by using knowledge from a pre-trained model on a source dataset. Authors develop separate models for each language, involving two phases: Training the Language Model (LM) and the Target Model. Using 300 XML files per language, authors introduce the ULMFiT model for profiling fake tweet spreaders. Initially trained on general English/Spanish data from Wikipedia, the LM knowledge is transferred to the fake news detection task. Their model achieves 64% accuracy for Spanish and 62% for English. This LM can also be applied to other English/Spanish NLP tasks. To address fake news detection, [11] worked with datasets like



BuzzFeedNews, LIAR16, BS Detector, and CREDBANK19. Authors noted that no single dataset covers all relevant features, and each has limitations. Authors also performed operations like clickbait, spammer, and bot detection to validate dataset robustness. [12] present a model with three main phases: input, pre-processing, and output. Authors collect labelled and unlabelled news data in the input phase and preprocess it using NLP. The preprocessing phase involves vectorization, a Recommender System, and multi-class classification. Authors propose a novel multi-class semi-supervised approach for self-training, utilizing a combination of classified and predominantly unlabelled data. Their method incorporates a similarity algorithm to enhance self-training by assigning new labels to labelled data. Evaluation on two benchmark datasets using logistic regression, decision tree, naive Bayes, and linear SVM shows their method's effectiveness and robustness in multi-class fake news classification, contributing to more reliable predictive models. [13] studied fake news detection during the COVID-19 pandemic using Decision Tree, K-Nearest Neighbour, Logistic Regression, Support Vector Machine, and Random Forest algorithms on a new dataset. Random Forest consistently outperformed other algorithms, closely followed by Support Vector Machine, across all configurations. Although textual and linguistic features individually enhanced detection, combining them did not significantly improve results. Bigrams and part-of-speech tags showed varying effectiveness. The research suggests that traditional machine learning methods can effectively utilize textual and linguistic features for fake news detection, with Random Forest and SVM achieving over 95% accuracy and F1-scores. Their research contributes by analysing emotional aspects of fake news through two main steps: (RQ1) identifying fake news and (RQ2) identifying and characterizing emotions. For RQ1, authors evaluated various algorithms for detecting fake news. After an extensive review of literature, data collection from tweets, sampling, and applying machine learning and deep learning algorithms, dense neural networks (accuracy: 0.956), random forests (accuracy: 0.949), and LSTM networks (accuracy: 0.931) showed the highest average accuracy. Transformer-based models like BERT and DistilBERT also performed well in their evaluation. [14] initially explored machine learning experiments and speculated that deep learning algorithms might yield better results for fake news detection. Various word-embedding techniques such as Word2Vec, GloVe, and FastText were utilized to generate effective data representations. For classification, deep learning models including LSTM, BiLSTM,

CNN-LSTM, and CNN-Bi-LSTM were employed. Due to the absence of a single large, standard dataset for fake news detection, the study integrated two publicly available datasets – Fake and real news, and all data, resulting in a dataset comprising 64,934 labelled news articles. Among the techniques tested, Word2Vec word embedding combined with the CNN-BiLSTM model demonstrated the highest performance, achieving accuracy, precision, recall, F1 measure, and AUC-ROC values of 0.975, 0.984, 0.970, 0.977, and 0.992, respectively. [15] tackled the fake news problem by applying the XGBoost model to their dataset. Authors also implemented SVM (Support Vector Machines), RF (Random Forest), LR (Logistic Regression), CART (Classification and Regression Trees), and NNET (Neural Network) machine learning models to enhance their algorithm's robustness. To generalize these models, authors conducted cross-validation. According to their results, the RF model achieved the highest accuracy at around 94%, while NNET showed the lowest performance with approximately 92.1%. In [16], the authors investigate the application of DistilBERT, a condensed version of BERT, for detecting XSS attacks in web applications. Leveraging DistilBERT's strong NLP capabilities, authors extract semantic features from input data to identify malicious XSS payloads. Their approach is evaluated on a comprehensive dataset, achieving high accuracy (99.82%), precision (99.83%), recall (99.66%), and F1 score (99.75%). Visualizations including confusion matrices, ROC curves, and precision-recall curves illustrate the model's robust performance. This research underscores the effectiveness of transformer-based models in fortifying web application security against advanced cyber threats. In [17], the authors introduce a RoBERTa-based bi-directional Recurrent Neural Network model for spam detection on social networks. Using RoBERTa to learn contextualized word representations, authors enhance the performance of the stacked BiLSTM network. A comparative study with common transformer-based models shows that their RoBERTa-BiLSTM model outperforms others on three benchmark datasets, achieving accuracies of 98.15% on Twitter, 94.41% on YouTube, and 99.74% on SMS data. In [18], the authors propose a CBLSTM (Contextualized Bi-directional Long Short Term Memory neural network) model to address spam detection on social networks. This model leverages deep contextualized word representation to overcome the limitations of traditional word embedding models, such as the “out of vocabulary” problem and lack of context. Experimental results on three benchmark datasets demonstrate that

their proposed method achieves high accuracy and outperforms existing state-of-the-art methods in detecting spam on social networks. In [19], the authors introduce ALBERT4Spam, a deep learning methodology for identifying spam on social networking platforms. This model leverages the ALBERT model for contextualized word representations and is built upon the Bidirectional Long Short-Term Memory neural network (BLSTM). Using random search to fine-tune hyperparameters, their model achieves optimal performance. Experiments on three benchmark datasets show that ALBERT4Spam outperforms widely used methods in spam detection, with precision results of 0.98 for Twitter, 0.96 for YouTube, and 0.98 for SMS datasets. In [20] the authors conducted an efficient analysis utilizing transformer-based BERT models, CNN, and BiLSTM architectures. Authors tested five different models, including variants of BERT such as BERT, DistilBERT, and BERTurk, as well as CNN architectures, across eight different datasets including LIAR, ISOT, GossipCop, and BuzzFeedNews. Through a comparative analysis, authors evaluated and reported the performance of the models across these diverse datasets. In [21] authors investigated the transformation of news dissemination in the context of social media, highlighting the shift from traditional media platforms to user-generated content. They defined fake news as information produced by deceptive or sensationalist users aimed at manipulation or provocation. The study emphasized the rapid spread of fake news among ordinary social media users, underscoring the critical need for swift detection mechanisms. Recognizing the limitations of expert systems, which struggle to keep pace with the high volume of social media traffic, the authors advocated for the development of semi-automatic and automatic fake news detection systems. By collecting and annotating data from Twitter, they implemented various supervised (K-Nearest Neighbor, Support Vector Machines, and Random Forest) and unsupervised (K-means, Non-Negative Matrix Factorization, and Linear Discriminant Analysis) machine learning algorithms. The results demonstrated that supervised learning approaches achieved the highest performance, with an average F1-score of 0.86, while unsupervised methods yielded a lower F1-score of 72%. The authors of [22] investigated the challenges associated with the spread of fake news in the digital age, examining its adverse effects on public perception and trust. In their study, authors developed a supervised machine learning algorithm designed to classify social media data as fake news. The methodology included five main components: data acquisition from Twitter,

data preprocessing, data transformation, model development utilizing Naive Bayes, decision tree, and support vector machine (SVM), and model evaluation through accuracy, precision, recall, and F1-score metrics. The results indicated that the decision tree algorithm achieved the highest accuracy for textual data and metadata, while also performing well in terms of precision, recall, and F1-score for the classification tasks. Additionally, SVM exhibited strong precision and recall metrics in the metadata classification.

## *Background*

### *2.1. Artificial Intelligence (AI)*

Artificial intelligence (AI) today is changing many fields of technology [23], [24], [25], [26]. This affects differently fields such as healthcare, finance, transportation, and communications. In the medical field, AI helps with diagnoses, treatment plans and research for new drugs. This improves patient care. In finance, AI helps with transactions, risk assessment and fraud detection. This improves decision making and the market works well. AI also supports traffic by creating self-driving cars and creating roads More safety and traffic management. In communication, AI helps talk with machines and translate language and emotional understanding. Overall, AI is important in creating new technologies and the idea was born.

#### *2.1.1. Machine Learning*

Machine learning, an important subfield of artificial intelligence, encompasses a diverse set of algorithms and methods that enable computer systems to learn from data and make predictions or decisions without being explicitly programmed. clear. At their core, machine learning algorithms leverage statistical techniques to identify patterns and relationships in data, thereby deriving insights and facilitating autonomous decision-making. These algorithms are often classified into supervised learning, unsupervised learning, and reinforcement learning models, each suitable for different learning situations. Supervised learning involves training algorithms on labelled data sets, where input-output pairs are provided, allowing the algorithm to learn the mapping between the input and the corresponding output. In contrast, unsupervised learning tasks involve extracting patterns and structures from unlabelled data, facilitating tasks such as clustering and anomaly detection. On the other hand, reinforcement learning focuses on training agents to interact with the environment

with the goal of maximizing cumulative rewards, often used in dynamic decision-making situations. Through these models, machine learning continues to drive innovation in fields ranging from healthcare and finance to natural language processing and computer vision, paving the way for groundbreaking advances. Transformative computing in data, prediction, and decision support systems.

### 2.1.2. Deep Learning

Deep learning, a subset of machine learning, includes a class of algorithms inspired by the structure and function of neural networks in the human brain. These algorithms are characterized by using multiple interconnected layers of artificial neurons to extract high-level features from raw data. Deep learning models excel at automatically learning complex patterns and representations from large volumes of unlabelled data, enabling tasks such as image and speech recognition, language processing natural and automatic decision making. The success of deep learning can be attributed to its ability to exploit hierarchical representations of data, extracting and incrementally refining features from each layer of the network. Using techniques such as backpropagation and stochastic gradient descent, deep learning models are trained to minimize errors and optimize performance on specific tasks. Deep learning has revolutionized many different sectors, from healthcare and finance to transportation and entertainment, driving innovation and breakthroughs in artificial intelligence research and applications.

### 2.1.3. Long Short-Term Memory (LSTM) and Bidirectional LSTM

Recurrent Neural Networks (RNNs) are a class of artificial neural networks commonly used for processing sequential data. However, they often encounter challenges, particularly the 'forgetting problem,' when dealing with long sequences. To address this, Long Short-Term Memory (LSTM) networks were developed. LSTMs are specifically designed to overcome the limitations of standard RNNs, especially their susceptibility to long-term dependency issues. Unlike traditional RNNs, which may struggle to retain information across extended sequences, LSTMs utilize memory cells that enable more effective handling of long-term dependencies. Each LSTM network consists of a chain of recurrent network modules, which are more complex than the single-layer structures found in

standard RNNs (e.g., a single tanh layer). LSTMs are explicitly designed to mitigate the exploding and vanishing gradient problems, making them well-suited for capturing longer-term dependencies in sequence data.

### 2.2. Text Processing and Feature Extraction Methods

#### 2.3. Feature Extraction Methods

Text data vectorization involves converting text into interactive vectors, enabling machines to solve math problems and process language. Researchers have developed various models for this purpose:

TF-IDF: This common method assigns importance to terms in documents, enhancing search engine performance. However, its adaptability is limited due to the selectiveness of the IDF term. In a more formal mathematical context, the computation of the TF-IDF score for the term  $t$  within the document  $d$  from the document set  $D$  is articulated as in question 1.

$$TF-IDF(t,d,D)=TF(t,d)\times IDF(t,D) \quad (1)$$

Word2Vec: This model generates semantic representations for words, aiming to capture their senses and relationships.

SentenceToVec: Extending Word2Vec, this approach averages word vectors to represent sentences. Notable advancements include Skip-Thought Vectors.

Doc2Vec: Extending Word2Vec to handle entire documents, Doc2Vec uses a similar process as SentenceToVec.

#### 2.3.1. Text Processing Methods

##### Text Tokenization

The BERT (Bidirectional Encoder Representations from Transformers) token engine is a basic one component of natural language processing (NLP) systems, known for their ability to capture contextual information and semantic nuances in text strings. Developed by Google AI in 2018, BERT token uses a complex tokenization strategy to split input text into sequence of sub-word tokens, allowing the model to consider contextual relationships between Speech is two-way. Unlike traditional tokenization methods that represent words in isolation, BERT tokenizer considers the entire context of the sentence, capturing dependencies and semantics links between words. This contextual understanding helps improve performance of downstream NLP tasks, such as text classification, named entity recognition, and sentiment analysis. By leveraging the BERT token, NLP practitioners can harness the power of Contextual integration to uncover deeper insights from text data, paving the way for more

powerful solutions and the system understands language accurately as seen in Tables 1 and Table 2.

Table 1. Overview of Transformer Based Text Tokenization

Special Tokens	[PAD]	[UNK]	[CLS]	[SEP]	[MASK]
Special Token ID	0	100	101	102	103

Table 2. Sample Text Tokenization

Sample Tweet	“Hello how are you?”
Tokenized version of the Tweet	[101, 7592, 2129, 2128, 2017, 102]

Positional embedding

One fundamental challenge in processing sequential data like text is capturing positional information. Transformer addresses this through positional embedding, where each token in the input sequence is augmented with positional information as seen in Figure 1. This allows the model to discern the order of tokens, crucial for understanding the context of the input.

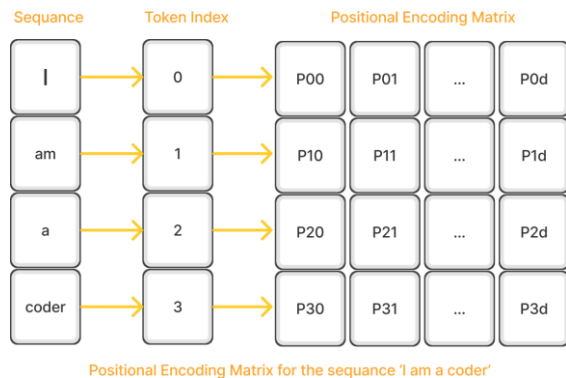


Figure 1. Visualization of Positional embedding

Padding

To accommodate variable-length inputs within a fixed-size matrix, padding is employed. This involves adding placeholder tokens, typically zeros, to shorter sequences to match the length of the longest sequence in the batch as seen in Table 3. Padding ensures uniformity in input dimensions, facilitating efficient batch processing.

Table 3. Overview Of Padding

Unpadded Input	Padded Input
[ [1,2,3], [4,5], [6,7,8,9,10] ]	[ [1,2,3,0,0], [4,5,0,0,0], [6,7,8,9,10], ]

2.4. Transformer Architecture

2.5. Overview of Transformer Architecture

Since the aim of this study is to evaluate the sentences in the tweets in terms of emotion and semantics, to make a reality prediction by taking advantage of their importance in the sentence on a word basis, we benefited from the transformer architecture as seen in Figure 2, which is frequently and successfully used in the field of natural language processing today. The Transformer architecture has emerged as a pivotal advancement in deep learning, particularly within the realm of Natural Language Processing (NLP). Developed on the foundation of attention mechanisms, it represents a paradigm shift in sequence modelling, enabling more effective handling of sequential data such as text. In this article, we delve into the key components of the Transformer architecture and explore some of the most prominent Transformer-based models shaping the landscape of NLP today.

The transformative impact of the Transformer architecture cannot be overstated. Its inception marks a watershed moment in the field of NLP, revolutionizing the way we process and understand language. At its core, the Transformer architecture harnesses the power of attention mechanisms, allowing models to focus on relevant parts of the input sequence with unprecedented precision. This not only enhances the model's ability to capture intricate linguistic patterns but also significantly improves its performance across various NLP tasks.

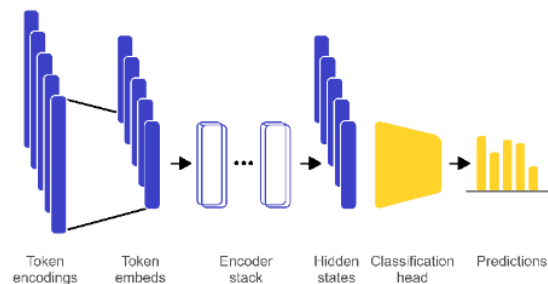


Figure 2. The architecture used for sequence classification with an encoder-based transformer.

One of the defining features of the Transformer architecture is its inherent scalability. Unlike traditional recurrent neural networks (RNNs) or convolutional neural networks (CNNs), Transformers exhibit superior parallelizability, making them well-suited for processing large volumes of text data efficiently. This scalability has played a crucial role in democratizing

NLP, enabling researchers and practitioners to tackle increasingly complex language processing tasks with ease.

### Attention Mechanism

The attention mechanism in deep learning was created to enhance machine translation by focusing on key parts of the input, like zooming in on one conversation in a noisy room. It copies how our brain highlights important sounds and ignores distractions, helping neural networks focus on different parts of the input. This is vital in areas like natural language processing (NLP), where attention helps match parts of a sentence during translation or answering questions. Attention also improves tasks in computer vision, such as pinpointing house numbers in Google Streetview. This guide explores the types, uses, and setup of attention mechanisms in TensorFlow to improve model performance by focusing on important details.

$$\text{Attention}(q,k,v) = \sum \text{similarity}(q,k_i) * v_i \quad (2)$$

- The attention mechanism assesses the likeness between the query  $q$  and every key-value pairs as seen in Figure 3.
- This similarity generates a weight for each key value.
- Ultimately, it generates an output that is the weighted amalgamation of all the values in our dataset.

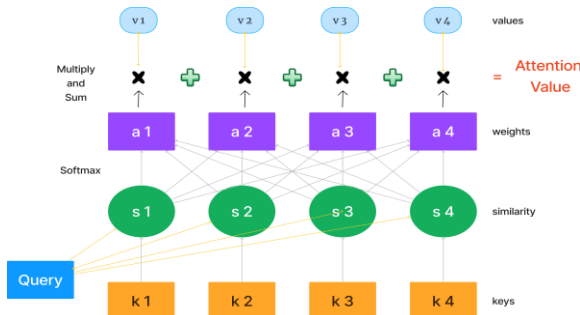


Figure 3. The Simple Overview of Attention Mechanism

#### 2.5.1. Masked language model

A core innovation introduced by models like BERT (Bidirectional Encoder Representations from Transformers) is the masked language model objective. Here, a certain percentage of tokens in the input sequence are masked, and the model is trained to predict these masked tokens based on the surrounding context as seen in Figure 4. This fosters a deeper understanding of inter-token relationships and enhances the model's ability to capture nuanced linguistic structures.

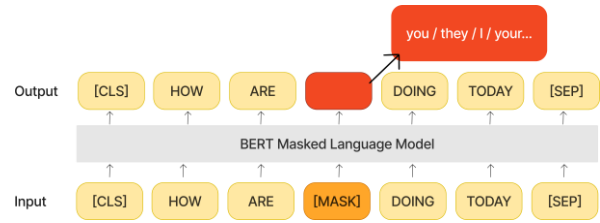


Figure 4. The Simple Example of Masked Language Model

Modern transformer-based models such as BERT, GPT and T5 have revolutionized Nature Language Processing (NLP) by excelling at tasks ranging from understanding language in context to text creation and multitasking learning. The Transformer architecture has revolutionized NLP, allowing models to solve various linguistic tasks with unprecedented accuracy and efficiency. From BERT's contextual language understanding to GPT and T5's language generation capabilities multitasking capabilities, Transformer-based models continue to push the boundaries of what's what feasible in understanding and producing natural language. As research advances in this area, we can anticipate other innovations and applications that harness the transformative power of Transformer-based architecture.

#### 2.6. Data Collection and Preprocessing

#### 2.7. Dataset

#### 2.8. Truth Seeker Dataset

For this study, we employed the Truth Seeker which was published by [3], a comprehensive collection of samples specifically curated to support the development and evaluation of deep learning and machine learning models in fake news detection. The examples in this dataset were labelled by real people from the well-equipped Amazon Mechanical Turk service, which worked meticulously to label each tweet in the dataset as true or false. The target category distribution in this dataset was 68930 for fake tweets and 65268 for real tweets as seen in Figure 5. As can be seen from the numbers, we were able to make a successful classification thanks to the data labelled in a balanced way. Of course, when we carefully examine the content of the texts shared on social media, especially the content of the tweets, we had to correct the grammatical complexities in the tweets shared by many bot accounts and the hashtags, mentions, usernames or spelling mistakes in the tweets shared by real people, which prevented the proposed model from classifying or at least did not contribute to the classification.

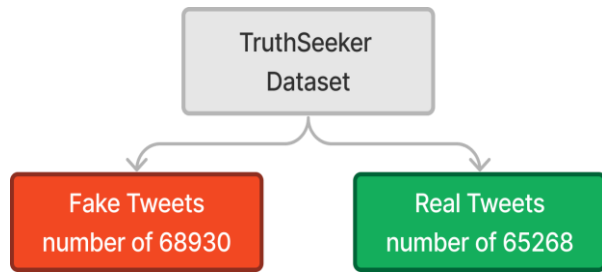


Figure 5. Distribution of the Dataset According to Target Categories

Here are the sample records from the TruthSeeker dataset as seen in Table 4.

We allocated 80% of the records from the TruthSeeker dataset for training and reserved the remaining 20% for validation purposes. After this division, the training set comprised 107,358 samples, while the validation set contained 26,840 samples. Furthermore, as there were no missing values in the relevant columns of the dataset, there was no necessity for data removal or imputation of missing values with averages or other methods.

2.8.1.1. Politifact Dataset

In this study, the publicly available PolitiFact dataset was used to evaluate the model's classification ability and to derive more objective inferences. The dataset, consisting of 19,422 labeled records, underwent data cleaning to remove any empty entries, followed by a text preprocessing phase. Afterward, the dataset was split into training and testing sets with an 80/20 ratio. As a result, 14,511 samples were prepared for training, and 4,837 samples for testing. The prepared data was then used to train a binary classification model, following the same approach as with the TruthSeeker dataset.

Table 4. Example Entries from the Dataset

Tweet	Label
"@AndreaR03428969 People vote with their pockets. Working class Americans (especially Obama-Trump-Biden voters) will remember that extra money from Trump, Bidens reconciliation failure & \$15 minimum wage failure, the ending of child tax benefits and eviction moratorium, and vote for Trump again."	1
@JackRichardso99 @Thee_Roxy_Cox @gnomeicide @glenn_coin @malaconotus @JAGLeMans @Bluesterge2 @lovejoy92 @UKCovid19Stats This virus mutates, seemingly quite readily. The more transmission, the more likely a vaccine-resistant escape variant	0

will develop. Plus you'll subject the unvaccinated to a small risk of death, and a greater risk of long covid.	
--	--

2.8.2. Preprocessing Phase

2.8.3. Text Preprocessing

For the machine learning and deep learning models we will use in this study to be successful, the data had to be open to study as much as possible. Although the proposed model employs the transformer architecture, known for its success in understanding word relationships and identifying noteworthy words in a sentence, we performed preprocessing on the tweets in our dataset. This preprocessing aimed to reduce ambiguity and eliminate unnecessary learning parameters, thereby preventing longer and less successful training. We can list the cleaning processes performed on the tweets in our dataset as follows:

- Cleaning E-Mail Addresses
- Cleaning URL Addresses
- Cleaning Retweet Tokens
- Cleaning HTML Tags
- Cleaning Mentions Dealing with Abbreviations

Here is the sample implementation of preprocessing step on the TruthSeeker dataset as seen in Table 5.

Table 5. Example Implementation of Preprocessing Phase

Before Preprocessing	@POTUS Biden Blunders - 6 Month Update\n\nInflation, Delta mismanagement, COVID for kids, Abandoning Americans in Afghanistan, Arming the Taliban, S. Border crisis, Breaking job growth, Abuse of power (Many Exec Orders, \$3.5T through Reconciliation, Eviction Moratorium)...what did I miss?
After Preprocessing	biden blunders 6 month update inflation, delta mismanagement, covid kids, abandoning americans afghanistan, arming taliban, s. border crisis, breaking job growth, abuse power (many exec orders, \$3.5t reconciliation, eviction moratorium).what miss?

In deep learning, balanced data is essential for accurate model training, yet datasets often exhibit imbalances across classes, posing challenges. To address this, researchers utilize data balancing methods, although their indiscriminate use may lead to overfitting or loss of information. Meanwhile, TomekLinks removes pairs of instances from different classes that are nearest neighbours, enhancing boundary discernment and generalization as depicted in Figure 6. TomekLinks improves model robustness and efficiency, fostering equitable learning and reliable insights in scientific research. In summary, Tomek Links are crucial for reducing imbalance in datasets by removing instances from the majority class close to those in the minority class.



Figure 6. Simple Visualization of Tomek Links

#### 2.8.4. Hyperparameter Tuning

Hyperparameter tuning is a critical process in machine learning and deep learning, aimed at optimizing model performance by systematically adjusting hyperparameters. Hyperparameters, such as learning rate, batch size, and regularization strength, govern the learning process and are distinct from model parameters learned during training. The efficacy of a machine learning model depends greatly on the selection of appropriate hyperparameters, which can significantly impact its performance, convergence, and generalization ability. Optimization techniques like grid search, random search, and Bayesian optimization are commonly used for this purpose in this work, we utilized Grid Search for hyperparameter optimization due to its methodical and exhaustive characteristics. In contrast to random search or genetic algorithms, which depend on stochastic techniques to navigate the hyperparameter space, Grid Search guarantees that all potential combinations within the defined grid are examined [27]. This thorough approach facilitates a more accurate determination of the optimal hyperparameters, especially in cases where the search space is limited. Although stochastic methods like random search can be more effective in larger search spaces, Grid Search provides a more structured and deterministic strategy, ensuring that no viable solution is missed. The hyperparameters

and their range values used in this experiment are shown in Table 6.

Table 6. Parameters used in Hyperparameter Optimization

Hyper Parameter	Ranges And Values
Activation Functions	relu,tanh, gelu
Kernel Initializers	uniform, lecun_uniform, normal
Optimizers	Adam, SGD, Adadelta, RMSprop, Adagrad, Adamax, Nadam
Learning Rates	1e-5, 1e-6
Dense Layers	32, 256
Bidirectional LSTM Layer Unit	128,256

#### 2.9. Proposed Model

While LSTMs are specifically designed to address the long-term dependency problem inherent in traditional RNNs, they still encounter limitations when processing particularly long sequences. Despite their ability to mitigate vanishing and exploding gradient issues through the use of memory cells, LSTMs can struggle with computational inefficiency and performance degradation as sequence length increases. The sequential nature of LSTMs leads to longer training times and can make them less effective at capturing complex contextual relationships over very long text sequences.

To overcome these limitations, a hybrid approach combining the strengths of Transformer-based models like BERT with LSTM networks is proposed. BERT excels in capturing context by utilizing a self-attention mechanism, which allows it to model long-range dependencies more efficiently than LSTMs alone. The bidirectional nature of both BERT and LSTM ensures that information is processed from both directions in the text, enhancing the model's understanding of context. By leveraging the robust contextual representation of BERT and combining it with the sequential processing power of Bidirectional LSTM, this hybrid model can more effectively handle both long-term dependencies and complex linguistic patterns, leading to superior performance in tasks such as fake news detection.

In this study, the Distilbert model, which is a simplified version of the BERT model, was used to check whether the tweets were real or fake. The reason why we made this choice was that despite the high performance in interpretation speed and performance, it gave little loss in terms of achievement. According to many studies conducted in the field, the Distilbert model Its duration

is 60 percent shorter than the Bert model. This speed difference provides a great advantage in using Distilbert for researchers and developers working with large language models. Considering the model size, DistilBERT has 44 million fewer parameters than the BERT model, making it approximately 40% smaller. Despite its reduced size, performance comparisons have shown that DistilBERT retains 97% of BERT's performance, as demonstrated in Figure 7 and supported by several benchmarks [28]. This reduction in model size offers significant trade-off, providing a reasonable balance between performance and faster inference speed.

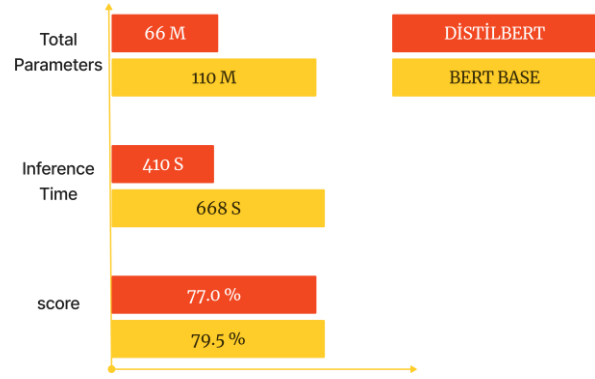


Figure 7. Comparison Of Bert and Distilbert Models

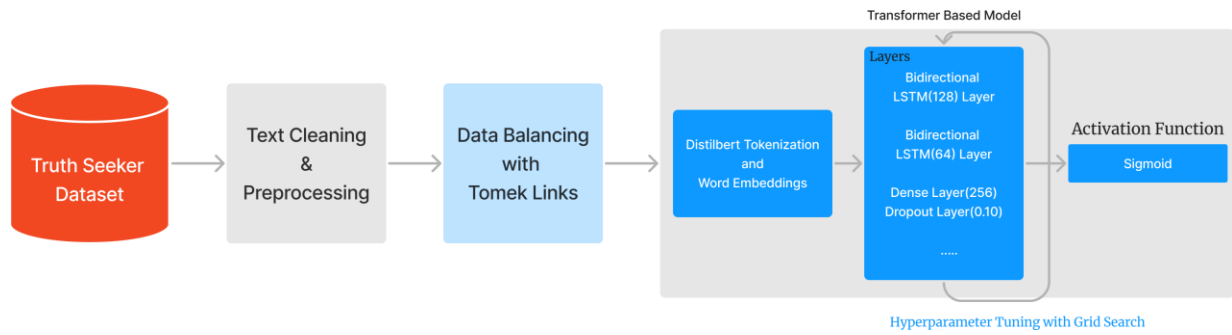


Figure 8. Proposed Model

To summarize our proposed model, as shown in Figure 8, data from the Truth Seeker dataset undergo a series of text preprocessing steps. To minimize classification errors, we utilized the Tomek Links algorithm for under sampling the majority class. This resulted in a more balanced and accurate dataset. Subsequently, the processed data were trained using the transformer based DistilBERT model and our defined list of hyperparameters.

#### 2.10. Experimental Results

##### 2.11. Experiments setup

The experiments in this study were performed on a computer with an i7 12th generation processor. A GTX 3060 video card was used as a GPU accelerator. All experiments were carried out using the TensorFlow library.

##### 2.11.1. Evaluation Metrics

The experiments aim to test how well different computer programs can find fake news. The measures we use to evaluate something include precision, recall, F-score, and accuracy. Precision is the number of right

decisions divided by the total number of decisions in a specific category. It is figured out as:

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive}$$

True Positive means the right fake news decisions, and False Positive means the wrong fake news decisions. The recall is the number of right decisions made by the machine compared to all the news in a specific category. It is figured out by:

$$Recall = \frac{TruePositive}{TruePositive + FalseNegative}$$

False Negative is when something that is not fake news is incorrectly labelled as fake news. Accuracy tells us how correct the decisions are compared to the real classification. The machine learning model's decision is only considered correct if it matches the real fake news class in the dataset. It is figured out by adding up some numbers.

$$\frac{TruePositive + TrueNegative}{TrueNegative + FalsePositive + FalseNegative}$$



Where True Negative is the correct not-fake news decision. Finally, the F1-score is the harmonic mean of precision and recall. It is calculated as:

$$F1Score = \frac{2 * Precision * Recall}{Precision + Recall}$$

### 2.11.2. Obtained Results

#### Deep Learning

In this study in order to prevent misclassification we utilized from Tomek Links algorithm. Thanks to this algorithm we obtained very good results on classifying fake and real tweets. In our study we conducted lots of test thanks to Grid Search. It allows us to try and select best hyperparameter which leverages our model's robustness. As we discussed previously, we used a trans-

former-based model to distinguish semantic relationship between words in tweets. In order to get concrete model, we applied several and important text preprocessing methods to eliminate meaningless and redundant words which prevent model to extract and learn important pieces of the sentences. Outputs comes from transformer were fed to LSTM layer. Again, as we discuss, due to RNN algorithm's deficit and forgetting problem in long sequences, we used LSTM layer to overcome the problem. In order to get more fertile result from LSTM, we used Bidirectional LSTM to get more reliable information from the sentences. Outputs that come from Bidirectional LSTM were passed through in Dense layers with different hyperparameters. Evaluated hyperparameters are listed in Table 7. Here are the 10 best models show the best performance with Bidirectional LSTM using output of transformer layer.

Table 7. Experimental Results of 10 Trials

Order	Optimizer	Dropout Layer	Learning Rate	Score
1	Adamax	True	1e-05	0.99908
2	RMSprop	True	1e-05	0.99867
3	RMSprop	True	1e-06	0.99836
4	Nadam	False	1e-06	0.99646
5	Adam	True	1e-05	0.99641
6	Adamax	True	1e-06	0.99621
7	Adamax	False	1e-05	0.99609
8	Adam	True	1e-06	0.99585
9	Nadam	False	1e-05	0.99487
10	RMSprop	False	1e-06	0.99429

The hyperparameter optimization process was conducted using Grid Search, with the results summarized in Table 7. The Adamax optimizer yielded the best performance, achieving an accuracy of 0.99908, highlighting its effectiveness for this specific task. RMSprop and Adam optimizers also demonstrated competitive performance, with accuracy values of 0.99867 and 0.99641, respectively, indicating their suitability for the Bidirectional LSTM-based model. In contrast, models trained with the Nadam optimizer performed slightly lower, with the highest accuracy being 0.99646.

Regarding the dropout layer, models incorporating dropout consistently outperformed those without it across different optimizers, underscoring the importance of regularization in preventing overfitting—particularly in recurrent neural networks like LSTMs. The learning rate also played a critical role in the model's performance. A learning rate of

0.00001 proved most effective for the top-performing models, while a lower rate of 0.000001 led to marginally reduced accuracy, demonstrating the importance of tuning the learning rate for optimal convergence.

Additionally, the 'uniform' kernel initializer was the most optimal choice across the top-performing models. The best-performing model, identified through Grid Search, was trained for 10 epochs, which was sufficient for convergence without overfitting. As seen in Figure 9, since there was no increase in the training curve as training progressed, the training was limited to 10 epochs. The hyperparameter optimization process had a significant impact on model performance, with fine-tuning of parameters resulting in near-perfect accuracy. The best model from the Grid Search was trained for 10 epochs, and the results are shown in Figures 9.

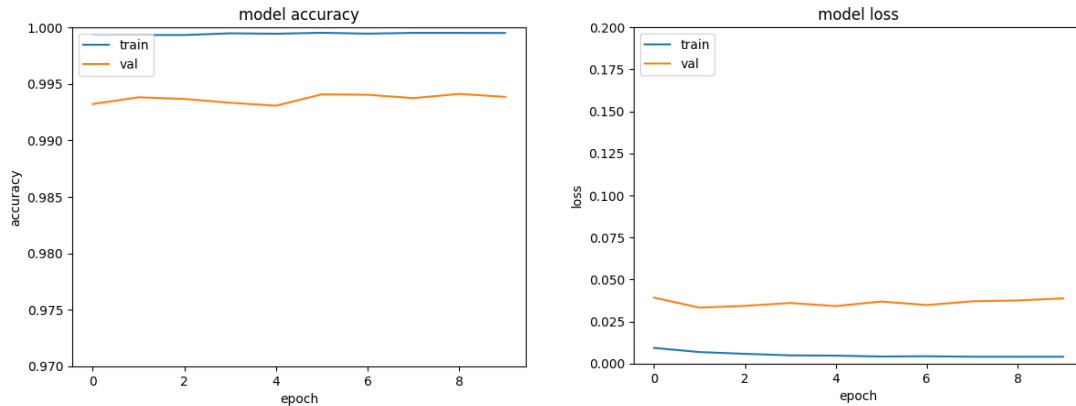


Figure 9. Model Accuracy and Loss of Best Model

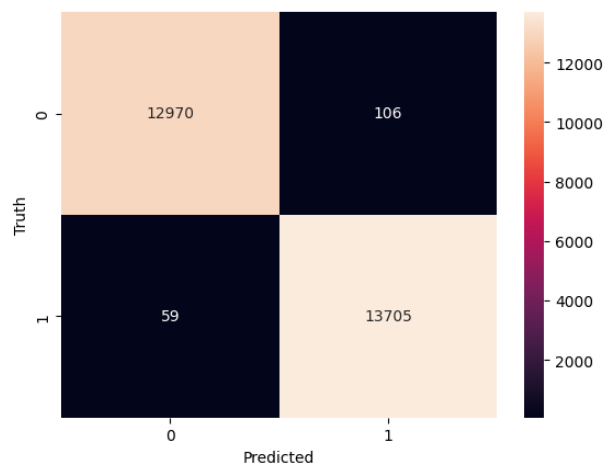


Figure 10. Confusion Matrix of Best Model

Figures 9 and 10 illustrate that our proposed model made more accurate classifications compared to traditional machine learning models (The confusion matrix shown in figure 11). When examining the number of misclassifications, our proposed model made nearly 50 fewer errors. After performing hyperparameter optimization using Grid Search, the best configuration was identified and subsequently used to train the final model. The model with the optimal hyperparameters was retrieved using `tuner.get_best_models(num_models=1)[0]`, and was trained for 10 epochs. Upon completion of training, the model demonstrated exceptional performance on the training data, achieving a loss of 0.0040 and an accuracy of 0.9995. This indicates that the model was able to almost perfectly fit the training data, with only minimal error. The low training loss suggests that the model's predictions closely matched the actual labels, while the extremely high accuracy indicates that very few classification errors occurred during training. The model was also evaluated on a separate validation dataset, where it achieved a

validation loss of 0.0388 and a validation accuracy of 0.9939. Although the validation accuracy is slightly lower than the training accuracy, this still represents outstanding performance. The slight increase in validation loss and reduction in accuracy suggests that the model generalized well to unseen data, with only a marginal degree of overfitting, if any. The gap between the training and validation results is relatively small, indicating that the model maintained strong predictive power even on data it had not encountered during training.

The confusion matrix further illustrates the model's classification performance. It is structured as follows: The confusion matrix provides a detailed breakdown of the model's classification results. Out of 13076 samples in the first class (true negatives), the model correctly identified 12970, with only 106 misclassified as false positives. For the second class (true positives), the model correctly identified 13705 out of 13764 samples, with 59 misclassified as false negatives. These results demonstrate a strong balance between precision and recall for both classes. Specifically, the model achieved a very low false positive rate (106 out of 13076) and a similarly low false negative rate (59 out of 13764). This shows that the model was able to correctly distinguish between the two classes with high reliability. In summary, the model performed remarkably well, achieving near-perfect accuracy and exhibiting only minor misclassifications in both positive and negative classes. The combination of low training and validation losses, coupled with high accuracy scores and a well-balanced confusion matrix, suggests that the model is highly effective for this classification task, with minimal overfitting and strong generalization capabilities.

### Comparing with Classical Methods

As we discussed in Deep Learning section, we carried out data balancing and useful text preprocessing steps. Then in contrast to deep learning model based on transformer architecture, we used Count Vectorizer and TF-IDF vectorizer to get text embeddings. After that we put it to test 12 machine learning models (Logistic Regression, Decision Tree Classifier, Extra Tree Classifier, XGB Classifier, XGBRF Classifier, AdaBoost

Classifier, Random Forest Classifier, Extra Trees Classifier, Gradient Boosting Classifier, Bagging Classifier, SGD Classifier, Support Vector Classifier) to classify the tweets with their default constructors.

According to the results, Support Vector Classifier stands out as the machine learning model with the highest accuracy level with a value of 98.94. When we give the training results of our Support Vector Classifier model to the configuration matrix, we get the results as seen in Table 8.

Table 8. Result of Machine Learning Models

Model-Name	Accuracy	ROC_AUC	F1_Score	Precision	Recall
SVC	98.945171	0.989455	0.989724	0.990119	0.989328
Extra Trees Classifier	98.863170	0.988649	0.988919	0.989818	0.988022
Random Forest Classifier	98.315256	0.983191	0.983564	0.985355	0.981779
Bagging Classifier	97.655522	0.976583	0.977131	0.978733	0.975535
Logistic Regression	97.640613	0.976310	0.977089	0.974303	0.979891
Decision Tree Classifier	97.368519	0.973600	0.974435	0.972112	0.976770
SGD Classifier	97.200790	0.971870	0.972856	0.968759	0.976987
XGB Classifier	96.574602	0.965311	0.967130	0.953187	0.981488
Extra Tree Classifier	92.631108	0.926129	0.928574	0.924266	0.932922
Gradient Boosting Classifier	86.410228	0.861418	0.878991	0.809661	0.961307
AdaBoost Classifier	85.303217	0.850706	0.867529	0.807442	0.937278
XGBRF Classifier	72.928547	0.722704	0.785885	0.661620	0.967623

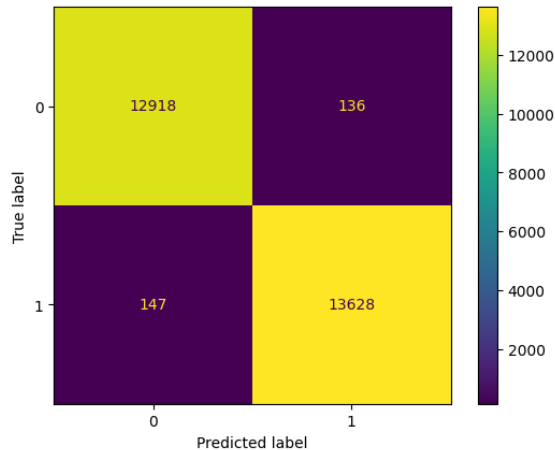


Figure 11. Confusion Matrix of SVC Model

Although we followed the same pre-processing steps and subjected the same data balancing processes to the dataset. We handled with 2 different approaches, when we compare our deep learning model, in which we use the Transformer architecture with Bidirectional LSTM layers, with classical machine learning methods, we can clearly see the difference in classification performance as seen in Figure 11. Our deep

learning-based model we created made approximately 50 fewer classification errors in both areas in classifying Fake and Real tweets than the Support Vector Classifier machine learning model.

## 4. DISCUSSION AND COMPARISON WITH OTHER STUDIES

When examining the studies presented in Table 9, we observe that numerous works in the field of fake news detection have employed machine learning and deep learning models using various datasets. The results in Table 9 demonstrate that studies in the field of fake news detection often achieve more effective outcomes when various machine learning and deep learning techniques are hybridly integrated. Models relying on a single architecture and approach tend to exhibit lower performance compared to hybrid models. Particularly, trained models, when combined with effective natural language processing approaches, demonstrate a heightened capability for high-level classification in fake news detection.

As mentioned in the Dataset section, in addition to the TruthSeeker dataset, we also conducted tests using the PolitiFact dataset to evaluate the model's

classification capabilities. During testing, we maintained the optimal parameters obtained through grid search and used the same natural language processing methods to train the model for 10 epochs, ensuring consistency for clearer comparison and more accurate inferences. This training resulted in an accuracy of 80.40%. The reason for this accuracy being lower than that achieved with the TruthSeeker dataset may be attributed to the PolitiFact dataset not having a sufficient number of instances for the model to learn all patterns effectively. Additionally, labels such as 'mostly-true' and 'barely-true' in the PolitiFact dataset may introduce ambiguity, leading to less definitive conclusions and causing uncertainty in the classification process.

Drawing from the outcomes of our tests and a survey of other research in the field, it is clear that in tasks such as fake news detection, the dataset used for training is as crucial as the models and hyperparameters applied. For a model to be viable in real-world applications and deployments, it needs to be trained on data that is both diverse and extensive. Insufficient variety and volume in the training data can hinder the model's ability to generalize, increasing the risk of misclassifications when exposed to new or domain-specific scenarios. This underscores the importance of using comprehensive datasets to prevent

the model from making erroneous predictions in unfamiliar contexts and to ensure strong performance in practical environments. Furthermore, a diverse dataset helps reduce biases and improves the model's flexibility, enabling it to operate effectively across a broad spectrum of subjects and situations.

## 5. CONCLUSION AND FUTURE STUDY

This research presents a novel model aimed at identifying fake news on social media, addressing an increasingly critical concern for society. Our approach leverages the BERT Transformer architecture, renowned for its efficacy in natural language processing tasks. To enhance the model's effectiveness and accuracy in classifying information, we integrated Bidirectional LSTM layers, a widely adopted technique in the field. Prior to feeding data into the model, we employed comprehensive text-cleaning methods to eliminate irrelevant words, symbols, and usernames from social media content. Furthermore, we standardized commonly used social media abbreviations to their full forms, ensuring clarity in the text input. To mitigate bias and classification errors within the target categories of our dataset, we utilized the Tomek Links algorithm, which further refined our data.

Table 9. Comparison of Our Model with Previous Studies on Fake News Detection

Work	Year	Method	Dataset / Inputs	Performance
<b>Proposed Model</b>	<b>2024</b>	<b>BERT, BiLSTM</b>	<b>TruthSeeker, PolitiFact</b>	<b>99,90%, 80.40%</b>
Seddari et al.[29]	2022	Hybrid approach that consists of language and knowledge-based methods	BuzzFeedNews	94.4%
Sahoo et al.[30]	2021	LSTM	FakeNewsNet	99.4%
Jarrahi et al.[31]	2021	UPFD framework	PolitiFact, Gossipcop	90.6%, 97.8%
Wang et al. [32]	2021	BERT, BiLSTM, CNN	COVID-19	93.47%
Ni et al. [33]	2021	Multi-View Attention Networks	Twitter15, Twitter16	92.34%, 93.65%
Lu et al.[34]	2020	Graph-aware CoAttention Networks (GCAN)	Twitter15, Twitter16	87.67%, 90.84%
Zhou et al.[35]	2020	Supervised model using linguistic and psychological features to detect fake news	PolitiFact, BuzzFeedNews	60% -70% 50%-60%
Shu et al. [36]	2019	Linguistic and structural approaches (STFN-HPFN)	PolitiFact, Gossipcop	85.6%, 86.3%
Kesarwani et al. [37]	2020	K-Nearest Neighbor classifier	BuzzFeedNews	79.0%
Yang et al. [38]	2019	UFD, Gibbs sampling	LIAR, BuzzFeedNews	75.9%, 67.9%

Shu et al. [39]	2019	LSTM	PolitiFact, BuzzFeed	67%, 74.2%
Traylor et al. [40]	2019	SciPy, NLP, Textblob	News Articles	63.3%
Rasool et al.[41]	2019	Dataset relabeling and iterative learning	LIAR	66.29%
Kayakuş et al. [42]	2023	Naive Bayes, Decision Trees	Twitter API	89.3 %, 84.2%
Taşkın et al. [21]	2021	Supervised and unsupervised learning algorithms	Twitter API	86.0%, 72.0%
Koru et al. [20]	2024	BERT, Bi-LSTM, CNN	BuzzFeedNews, GossipCop and other 5 datasets	94%

A significant emphasis was placed on hyperparameter optimization to enhance model performance. Each hyperparameter was meticulously evaluated through Grid Search, leading to a training process involving a vast array of parameter combinations. This extensive tuning resulted in all top 10 models demonstrating exceptional performance, with accuracy rates exceeding 99%. The practical implications of this research are significant. The model's application can extend to various social media platforms, where the spread of misinformation poses substantial risks. By providing real-time detection capabilities, our model could assist users in discerning credible information from false narratives, thereby fostering a more informed society. However, it is essential to acknowledge potential risks associated with implementing such technology, including reliance on automated systems and the challenge of adapting to the evolving nature of misinformation. Looking ahead, future studies will focus on enhancing the model's robustness against adversarial attacks. This involves investigating the model's vulnerability to manipulated inputs designed to deceive it, as well as developing techniques to strengthen its resilience. By addressing these challenges, we aim to ensure that our fake news detection system remains reliable and effective in the face of sophisticated misinformation tactics.

### Funding

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

### Conflict of interest

The authors declare that they have no conflict of interest.

### Data availability statement

The datasets generated during and/or analyzed during the current study are available from the authors on reasonable request.

### REFERENCES

- [1] Janelle B. Hill, "Gartner insights on spotting and responding to digital disruption Leading Through Digital Disruption EDITED BY Janelle B. Hill, Gartner Research Vice President and Distinguished Analyst," 2017. Accessed: Jan. 04, 2025. [Online]. Available: [https://www.gartner.com/imagesrv/books/digital-disruption/pdf/digital\\_disruption\\_ebook.pdf](https://www.gartner.com/imagesrv/books/digital-disruption/pdf/digital_disruption_ebook.pdf)
- [2] G. Mavridis, "Fake news and Social Media: How Greek users identify and curb misinformation online," 2018, Accessed: Jan. 03, 2025. [Online]. Available: <https://urn.kb.se/resolve?urn=urn:nbn:se:mau:diva-23196>
- [3] S. Dadkhah, X. Zhang, A. G. Weismann, A. Firouzi, and A. A. Ghorbani, "TruthSeeker: The Largest Social Media Ground-Truth Dataset for Real/Fake Content." [Online]. Available: <https://www.unb.ca/cic/datasets/truthseeker-2023.html>.
- [4] S. Sharma, M. Saraswat, and A. K. Dubey, "Fake news detection on Twitter," *International Journal of Web Information Systems*, vol. 18, no. 5–6, pp. 388–412, Dec. 2022, doi: 10.1108/IJWIS-02-2022-0044.
- [5] A. Ali and M. Gulzar, "An Improved FakeBERT for Fake News Detection," *Applied Computer Systems*, vol. 28, no. 2, pp. 180–188, Dec. 2023, doi: 10.2478/acss-2023-0018.
- [6] R. K. Kaliyar, A. Goswami, and P. Narang, "FakeBERT: Fake news detection in social media with a BERT-based deep learning approach," *Multimed Tools Appl*, vol. 80, no. 8, pp. 11765–11788, Mar. 2021, doi: 10.1007/s11042-020-10183-2.
- [7] H. Alsaidi and W. Etaïwi, "Empirical Evaluation of Machine Learning Classification Algorithms for Detecting COVID-19 Fake News," *International Journal of Advances in Soft Computing and its Applications*, vol. 14, no.

- 1, pp. 49–59, 2022, doi: 10.15849/IJASCA.220328.04.
- [8] A. M. Ali, F. A. Ghaleb, B. A. S. Al-Rimy, F. J. Alsolami, and A. I. Khan, “Deep Ensemble Fake News Detection Model Using Sequential Deep Learning Technique,” *Sensors*, vol. 22, no. 18, Sep. 2022, doi: 10.3390/s22186970.
- [9] B. Fang and H. Zhou, “Fake news text detection based on convolutional neural network,” *Applied and Computational Engineering*, vol. 41, no. 1, pp. 202–209, Feb. 2024, doi: 10.54254/2755-2721/41/20230744.
- [10] “(PDF) ULMFiT for Twitter Fake News Spreader Profiling Notebook for PAN at CLEF 2020.” Accessed: Jan. 04, 2025. [Online]. Available: [https://www.researchgate.net/publication/359024571\\_ULMFiT\\_for\\_Twitter\\_Fake\\_News\\_Spreader\\_Profiling\\_Notebook\\_for\\_PAN\\_at\\_CLEF\\_2020](https://www.researchgate.net/publication/359024571_ULMFiT_for_Twitter_Fake_News_Spreader_Profiling_Notebook_for_PAN_at_CLEF_2020)
- [11] K. Shu, A. Sliva, S. Wang, J. Tang, and H. Liu, “Fake News Detection on Social Media,” *ACM SIGKDD Explorations Newsletter*, vol. 19, no. 1, pp. 22–36, Sep. 2017, doi: 10.1145/3137597.3137600.
- [12] O. Stitini and S. Kaloun, “An improved self-training model to detect fake news categories using multi-class classification of unlabeled data: fake news classification with unlabeled data,” *Int. J. Systematic Innovation*, vol. 8, no. 1, p. 4, 2024, doi: 10.6977/IJoSI.202403\_8(1)0002.
- [13] V. Balakrishnan, H. L. Zing, and E. Laporte, “COVID-19 INFODEMIC – UNDERSTANDING CONTENT FEATURES IN DETECTING FAKE NEWS USING A MACHINE LEARNING APPROACH,” *Malaysian Journal of Computer Science*, vol. 36, no. 1, pp. 1–13, 2023, doi: 10.22452/mjcs.vol36no1.1.
- [14] A. K. Yadav et al., “Fake News Detection Using Hybrid Deep Learning Method,” *SN Comput Sci*, vol. 4, no. 6, pp. 1–15, Nov. 2023, doi: 10.1007/S42979-023-02296-W/METRICS.
- [15] M. Park and S. Chai, “Constructing a User-Centered Fake News Detection Model by Using Classification Algorithms in Machine Learning Techniques,” *IEEE Access*, vol. 11, pp. 71517–71527, 2023, doi: 10.1109/ACCESS.2023.3294613.
- [16] “(13) (PDF) DISTILBERT FOR WEB SECURITY: ENHANCED DETECTION OF XSS ATTACKS USING NLP APPROACH.” Accessed: Jul. 17, 2024. [Online]. Available: [https://www.researchgate.net/publication/381659932\\_DISTILBERT\\_FOR\\_WEB\\_SECURITY\\_ENHANCED\\_DETECTION\\_OF\\_XSS\\_ATTACKS\\_USING\\_NLP\\_APPROACH](https://www.researchgate.net/publication/381659932_DISTILBERT_FOR_WEB_SECURITY_ENHANCED_DETECTION_OF_XSS_ATTACKS_USING_NLP_APPROACH)
- [17] R. Ghanem, H. Erbay, and K. Bakour, “Contents-Based Spam Detection on Social Networks Using RoBERTa Embedding and Stacked BLSTM,” *SN Comput Sci*, vol. 4, no. 4, pp. 1–15, Jul. 2023, doi: 10.1007/S42979-023-01798-X/METRICS.
- [18] R. Ghanem and H. Erbay, “Spam detection on social networks using deep contextualized word representation,” *Multimed Tools Appl*, vol. 82, no. 3, pp. 3697–3712, Jan. 2023, doi: 10.1007/S11042-022-13397-8/METRICS.
- [19] A. Makalesi, R. Article Rezan BAKIR, H. Erbay, and H. Bakir, “ALBERT4Spam: A Novel Approach for Spam Detection on Social Networks,” no. 2, p. 17, doi: 10.17671/gazibtd.1426230.
- [20] G. K. Koru and C. Uluyol, “Detection of Turkish Fake News from Tweets with BERT Models,” *IEEE Access*, vol. 12, pp. 14918–14931, 2024, doi: 10.1109/ACCESS.2024.3354165.
- [21] S. G. TAŞKIN, E. U. KÜÇÜKSİLLE, and K. TOPAL, “Twitter üzerinde Türkçe sahte haber tespiti,” *Balikesir Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, vol. 23, no. 1, pp. 151–172, Jan. 2021, doi: 10.25092/baunfbed.843909.
- [22] N. A. S. Abdullah, N. I. A. Rusli, and N. S. Yuslee, “Development of a machine learning algorithm for fake news detection,” *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 35, no. 3, pp. 1732–1743, Sep. 2024, doi: 10.11591/ijeecs.v35.i3.pp1732-1743.
- [23] R. Bakır and H. Bakır, “Swift Detection of XSS Attacks: Enhancing XSS Attack Detection by Leveraging Hybrid Semantic Embeddings and AI Techniques,” *Arab J Sci Eng*, pp. 1–17, Jun. 2024, doi: 10.1007/S13369-024-09140-0/TABLES/14.
- [24] H. Bakır and G. Tarihi, “Using Transfer Learning Technique as a Feature Extraction Phase for Diagnosis of Cataract Disease in the Eye,” *USBTU*, vol. 1, no. 1, p. 2022.
- [25] H. Bakır and K. Elmabruk, “Deep learning-based approach for detection of turbulence-induced distortions in free-space optical communication links,” *Phys Scr*, vol. 98, no. 6, p. 065521, May 2023, doi: 10.1088/1402-4896/ACD4FA.
- [26] U. Demircioğlu, A. Sayıl, and H. Bakır, “Detecting Cutout Shape and Predicting Its Location in Sandwich Structures Using Free Vibration Analysis and Tuned Machine-Learning Algorithms,” *Arab J Sci Eng*, vol. 49, no. 2, pp. 1611–1624, Feb. 2024, doi: 10.1007/S13369-023-07917-3/METRICS.
- [27] J. Bergstra, J. B. Ca, and Y. B. Ca, “Random Search for Hyper-Parameter Optimization Yoshua Bengio,” 2012. [Online]. Available: <http://scikit-learn.sourceforge.net>.
- [28] V. Sanh, L. Debut, J. Chaumond, and T. Wolf, “DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter,” Oct. 2019, [Online]. Available: <http://arxiv.org/abs/1910.01108>
- [29] N. Seddari, A. Derhab, M. Belaoued, W. Halboob, J. Al-Muhtadi, and A. Bouras, “A Hybrid Linguistic and Knowledge-Based Analysis Approach for Fake News Detection on Social Media,” *IEEE Access*, vol. 10, pp.

- 62097–62109, 2022, doi: 10.1109/ACCESS.2022.3181184.
- [30] S. R. Sahoo and B. B. Gupta, “Multiple features based approach for automatic fake news detection on social networks using deep learning,” *Appl Soft Comput*, vol. 100, p. 106983, Mar. 2021, doi: 10.1016/J.ASOC.2020.106983.
- [31] A. Jarrahi and L. Safari, “FR-Detect: A Multi-Modal Framework for Early Fake News Detection on Social Media Using Publishers Features,” Sep. 2021, Accessed: Sep. 24, 2024. [Online]. Available: <https://arxiv.org/abs/2109.04835v1>
- [32] Y. Wang, Y. Zhang, X. Li, and X. Yu, “COVID-19 Fake News Detection Using Bidirectional Encoder Representations from Transformers Based Models,” Sep. 2021, Accessed: Sep. 24, 2024. [Online]. Available: <https://arxiv.org/abs/2109.14816v2>
- [33] S. Ni, J. Li, and H. Y. Kao, “MVAN: Multi-View Attention Networks for Fake News Detection on Social Media,” *IEEE Access*, vol. 9, pp. 106907–106917, 2021, doi: 10.1109/ACCESS.2021.3100245.
- [34] Y. J. Lu and C. Te Li, “GCAN: Graph-aware co-attention networks for explainable fake news detection on social media,” *Proceedings of the Annual Meeting of the Association for Computational Linguistics*, pp. 505–514, 2020, doi: 10.18653/V1/2020.ACL-MAIN.48.
- [35] “(3) (PDF) Fake News Early Detection: A Theory-driven Model.” Accessed: Sep. 24, 2024. [Online]. Available: [https://www.researchgate.net/publication/332726212\\_Fake\\_News\\_Early\\_Detection\\_A\\_Theory-driven\\_Model](https://www.researchgate.net/publication/332726212_Fake_News_Early_Detection_A_Theory-driven_Model)
- [36] K. Shu, S. Wang, H. Liu, and D. Mahudeswaran, “Hierarchical Propagation Networks for Fake News Detection: Investigation and Exploitation”, doi: 10.48550/arXiv.1903.09196.
- [37] A. Kesarwani, S. S. Chauhan, and A. R. Nair, “Fake News Detection on Social Media using K-Nearest Neighbor Classifier,” *Proceedings of the 2020 International Conference on Advances in Computing and Communication Engineering, ICACCE 2020*, Jun. 2020, doi: 10.1109/ICACCE49060.2020.9154997.
- [38] S. Yang, K. Shu, S. Wang, R. Gu, F. Wu, and H. Liu, “Un-supervised fake news detection on social media: A generative approach,” *33rd AAAI Conference on Artificial Intelligence, AAAI 2019, 31st Innovative Applications of Artificial Intelligence Conference, IAAI 2019 and the 9th AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019*, pp. 5644–5651, 2019, doi: 10.1609/AAAI.V33I01.33015644.
- [39] K. Shu, D. Mahudeswaran, and H. Liu, “FakeNewsTracker: a tool for fake news collection, detection, and visualization,” *Comput Math Organ Theory*, vol. 25, no. 1, pp. 60–71, Mar. 2019, doi: 10.1007/s10588-018-09280-3.
- [40] T. Traylor, J. Straub, Gurmeet, and N. Snell, “Classifying Fake News Articles Using Natural Language Processing to Identify In-Article Attribution as a Supervised Learning Estimator,” *Proceedings - 13th IEEE International Conference on Semantic Computing, ICSC 2019*, pp. 445–449, Mar. 2019, doi: 10.1109/ICOSC.2019.8665593.
- [41] T. Rasool, W. H. Butt, A. Shaukat, and M. U. Akram, “Multi-label fake news detection using multi-layered supervised learning,” *ACM International Conference Proceeding Series*, pp. 73–77, Feb. 2019, doi: 10.1145/3313991.3314008.
- [42] M. KAYAKUŞ and F. YİĞİT AÇIKGÖZ, “Twitter’da Makine Öğrenmesi Yöntemleriyle Sahte Haber Tespiti,” *Abant Sosyal Bilimler Dergisi*, vol. 23, no. 2, pp. 1017–1027, Jul. 2023, doi: 10.11616/asbi.1266179.

# Meme Kanseri Tanısında Wisconsin Veri Seti ile Makine Öğrenmesi Uygulamaları

## Araştırma Makalesi/Research Article

 Refik TANGİ<sup>1</sup>,  Ramazan SOLMAZ<sup>2</sup>

<sup>1</sup>Bartın Üniversitesi, Lisansüstü Eğitim Enstitüsü, Akıllı Sistemler Mühendisliği Bölümü, Bartın, Türkiye

<sup>2</sup>Kahramanmaraş İstiklal Üniversitesi, Mühendislik Mimarlık ve Tasarım Fakültesi, Yazılım Mühendisliği Bölümü, Kahramanmaraş, Türkiye

[refiktangi@windowslive.com](mailto:refiktangi@windowslive.com), [ramazan.solmaz@istiklal.edu.tr](mailto:ramazan.solmaz@istiklal.edu.tr)

(Geliş/Received:14.08.2024; Kabul/Accepted:05.11.2024)

DOI: 10.17671/gazibtd.1533288

**Özet**— Meme kanseri giderek daha sık görülmekte ve endişe verici bir boyuta ulaştığı ifade edilmektedir. Hastalık teşhis edilmezse ölüm riskini önemli ölçüde artırmaktadır. Son aşamada teşhis edildiğinde, tedbir olarak uzuvların alınması gerekmektedir. Erken teşhis için başarılı bir yöntem öncü olabilir. Bu makalenin odak noktası, meme kanseri teşhisinde başarılı makine öğrenimi tekniklerinin otomatik tanı için değerlendirilmesidir. Ayrıca, orijinal Wisconsin meme kanseri veri setine ait belirli özelliklerin etkinliği kontrol edilerek daha az işlem yükü ile başarılı tahminler araştırılmaktadır. Bu amaçla veri setine çeşitli makine öğrenimi algoritmaları uygulanmış ve en iyi performans gösteren algoritmalar belirlenmiştir. Daha başarılı bir tahmin için veri setine ön işlem uygulanarak etkin özellikler tespit edilmiştir. İlk bulgulardan yola çıkarak bu çalışmada, NB, DVM, J48 ve k-NN sınıflandırma algoritmaları ile k-means ve hiyerarşik kümeleme algoritmaları kullanılmıştır. Algoritmaların hastalık tanısındaki performansları doğruluk, ROC değerleri ve karmaşıklık matrisi metrikleriyle analiz edilmiştir. Performans metrikleri, en iyi sonucun NB tekniği ile elde edildiğini göstermektedir. Analiz edilen modellerin metrikleri, verilerin değerlendirilmesinde kullanılan çekirdek fonksiyonlarının tanıda önemli rol oynadığını göstermektedir. Wisconsin veri setine uygulanan denetimli algoritmalar güvenilir sonuçlar vermiştir. Meme kanseri teşhisinde başarılı olan algoritmaların sağlık sisteminde kullanılan analiz cihazlarına bir yazılım aracı olarak entegre edilmeleri, erken tanı ve farkındalık için iyi bir öncü olabileceği değerlendirilmektedir.

**Anahtar Kelimeler**— meme kanserinde otomatik tanı, öncü tasarımı, makine öğrenimi teknikleri, wisconsin veri seti

## Machine Learning Applications on Wisconsin Dataset for Breast Cancer Diagnosis

**Abstract**— Breast cancer is increasingly common and is reaching an alarming level. If the disease is not diagnosed, it significantly increases the risk of death. When diagnosed at a late stage, the only precaution is often the removal of limbs. An effective method for early diagnosis could be a successful precursor. This paper focuses on evaluating successful machine learning techniques for automatic diagnosis in breast cancer detection. Additionally, the effectiveness of certain features of the original Wisconsin breast cancer dataset is examined to achieve accurate predictions with less computational load. For this purpose, various machine learning algorithms were applied to the dataset, and the best-performing algorithms were identified. To achieve more accurate predictions, preprocessing was applied to the dataset to identify effective features. Based on initial findings, NB, SVM, J48, and k-NN classification algorithms, as well as k-means and hierarchical clustering algorithms, were used in this study. The performance of the algorithms in disease diagnosis was analyzed using metrics such as accuracy, ROC values, and confusion matrices. Performance metrics indicate that the best result was obtained with the NB technique. The metrics of the analyzed models show that the kernel functions used in data evaluation play a significant role in diagnosis. Supervised algorithms applied to the Wisconsin dataset provided reliable results. It is considered that integrating successful algorithms in breast cancer diagnosis as a software tool into analysis devices used in the healthcare system could be a good precursor for early diagnosis and awareness.

**Keywords**— automatic diagnosis in breast cancer, precursor design, machine learning techniques, wisconsin dataset.



## 1. GİRİŞ (INTRODUCTION)

Meme kanseri, kadınlarda en yaygın görülen kanser türüdür ve kanser kaynaklı ölümlerin %15'ini oluşturmaktadır. Bu oranlar, gelişmekte olan ülkelerde daha yüksek seviyelerde seyretmektedir. Önümüzdeki 20 yıl içinde, teşhis konulan vaka sayısında (insidans) %55, ölüm oranlarında (mortalite) ise %58 artış öngörülmektedir [1-3]. Sağlık Bakanlığı istatistiklerine göre, meme kanseri ülkemizde görülen kanser türleri arasında ilk sıradadır [4]. 2019 yılında yaklaşık 4300 kadının meme kanseri kaynaklı hayatını kaybettiği ifade edilmiştir. Yine 2017 yılında yapılan "Ölüm Nedenlerinin Dağılımı" adlı araştırmada, Türkiye'de görülen ölümlerin %1'inin; AB ülkelerinde görülen ölümlerin ise %2'sinin meme kanseri kaynaklı olduğu rapor edilmiştir [2]. Kanser teşhisi alan her dört kadından birinin meme kanseri olduğu ve meme kanserinin giderek daha küçük yaşlarda görüldüğü not edilmiştir [5].

Meme kanserinin görülme sıklığı ve bu kanserin önemli sağlık sorunlarına yol açması nedeniyle, doktorların yanı sıra otomatik tanı için araştırmacıların da yoğun ilgisini çekmektedir. Erken tanı, hastalığın ilerlemesini durdurmada ve hastaların yaşam sürelerini uzatmada kritik bir rol oynamaktadır. Bu kapsamda makine öğrenimi yöntemleri, tıbbi tanı ve tedavi süreçlerinde giderek daha fazla kullanılmaktadır. Bu teknikler hastalık veri setleri üzerinde modelleme yaparak, hastalıkların erken tanısı ve doğru sınıflandırılması için güçlü araçlar sunmaktadır. Otomatik tanı aracı veya öncü tasarımı son zamanlarda araştırmacılar tarafından oldukça yoğun ilgi görmektedir. Wisconsin Meme Kanseri Veri Seti (Wisconsin Diagnostic Breast Cancer (WDBC)-Original) otomatik tanı araçlarını geliştirmek için yaygın olarak kullanılmaktadır. Wisconsin veri seti, biyopsi sonuçlarından elde edilen hücresel özellikleri içermekte olup, iyi huylu (benign) ve kötü huylu (malign) tümörlerin ayırımında kullanılmaktadır. Bu veri setine UCI makine öğrenimi veri tabanından ulaşılabilir [6].

Makine öğrenimi algoritmalarının meme kanseri teşhisindeki etkinliği giderek artmakta ve algoritma performansını geliştirme çalışmaları devam etmektedir. Wisconsin veri setleri de meme kanseri teşhisinde sıkça başvurulan kaynak niteliğini taşımaktadır. Amrane ve arkadaşları (2018), bu veri setini kullanarak çeşitli denetimli öğrenme sınıflandırıcılarının performanslarını karşılaştırmış ve k-NN algoritmasının %97,51 doğruluk oranına sahip olduğunu belirtmişlerdir [7]. Benzer bir karşılaştırma yapan Aruna ve arkadaşları (2011), WDBC veri setine Destek Vektör Makinesi (DVM-RBF Kernel) sınıflandırıcı uygulayarak otomatik tanı konusunda %98,06'lık bir başarı derecesi elde ettiklerini ifade etmişlerdir [8]. Uddin vd., (2023) Wisconsin veri setinden yararlanarak özellik optimizasyonu tekniği ile makine öğrenimine dayalı meme kanseri teşhisi yapmışlardır. Sınıflandırıcıların performanslarının metrikler ile değerlendirildiği çalışmada en yüksek başarıyı %98,77 ile oylama sınıflandırıcı (Voting classifier) ile elde ettiklerini belirtmişlerdir [9]. Nemade ve Fegade (2023) meme kanserinin kadınlarda ölümlerin ana nedenlerinden biri

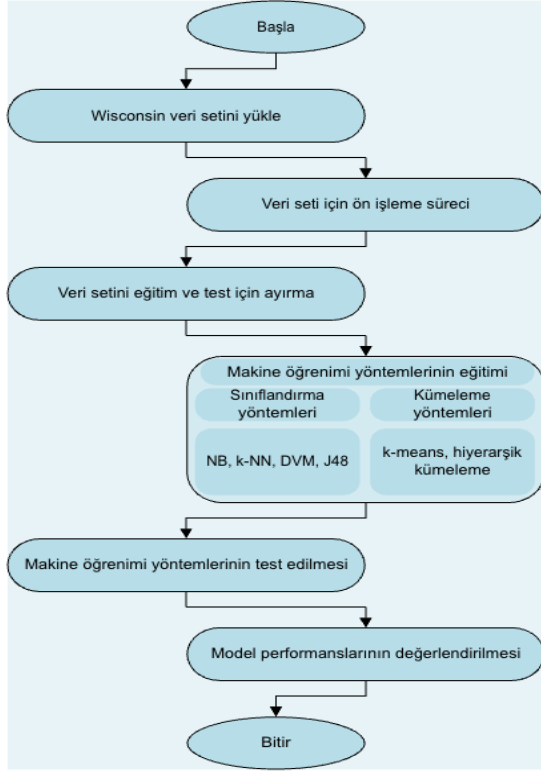
olduğunu ve meme kanseri tanısının oldukça zor olduğunu not etmişlerdir. Uzmanlar, kötü huylu tümör ile iyi huylu tümör arasında ayırım yapabilmek için bazı otomatik araçlara ihtiyaç duyulduğunu ifade etmişlerdir. Bu amaçla makine öğrenimi algoritmalarıyla Wisconsin veri seti örneklerini otomatik tasnif ederek tasnif işleminde kullanılan metotların performanslarını değerlendirmişlerdir. Çalışmalarında karar ağaçları ve XGBoost sınıflandırıcılar ile %97 başarı oranı elde ettiklerini belirtmişlerdir [10]. Singh vd. (2024) özellik seçiminin makine öğrenmesi algoritmaları performanslarında oldukça önemli olduğunu vurgulamış ve bunun için yeni bir algoritma (Feature selection-FS) önermişlerdir. Bu algoritma özellik seçimi için var olan iki algoritmayı birleştirmekte ve WDBC verilerini sınıflandırma için altı makine öğrenmesi metotlarını kullanmaktadır. Önerilen metodun uygunluğunu performans ölçüm metrikleri ile ölçerek elde edilen sonuçları literatür ile karşılaştırmışlardır [11]. Laghmati vd. (2023) makine öğrenme ve temel bileşen analizi (PCA) ile meme kanseri tahmin sistemlerini geliştirmeyi amaçlamakla birlikte kanser teşhisinde daha doğru ve hızlı sonuçlar elde etmeyi hedeflemişlerdir. Çalışmalarında k-NN algoritmasının %93,8 doğruluk oranına sahip olduğunu belirtmişlerdir [12]. Amethiya vd. (2021) makine öğrenimi yaklaşımına dayalı çeşitli algoritmaların ve biyosensörlerin erken meme kanseri tespiti için uygulanmasını araştırmak amacıyla çeşitli yaklaşımlar sunmaktadır ve bu kapsamda k-NN algoritmasının %95,9 doğruluk oranına sahip olduğunu ortaya koymaktadırlar [13]. Kadhim ve Kamil (2022) çeşitli kriterleri kullanarak makine öğrenimi algoritmalarının meme kanseri teşhisindeki başarıları araştırmışlardır. Bu amaçla çeşitli sınıflandırıcılar karşılaştırılmış ve ERT (Extremely randomized trees) algoritmasının %97,36 doğruluk oranına sahip olduğunu belirtilmiştir [14].

Bu çalışmada, meme kanseri veri setine etkinliği iyi bilinen Naive Bayes (NB), k-En Yakın Komşu (k-Nearest Neighbors, k-NN), J48 (Decision Tree), Destek Vektör Makinesi (DVM) gibi sınıflandırma yöntemleri ile hiyerarşik kümeleme ve k-means makine öğrenmesi algoritmaları uygulanarak otomatik tanı için yöntemlerin performansları ve doğruluk dereceleri ele alınmıştır. Meme kanseri veri seti 699 örnek, 10 öznitelik ve bir sınıf özniteliği içermektedir. Sınıf özniteliği, örneklerin kötü huylu ya da iyi huylu tümör bilgilerini içermektedir.

Bu çalışmanın odak noktası, veri seti analiz edilerek hastalık tespitinde etkin özellikleri belirlenmesidir. Ayrıca sürekli artan verilerden anlamlı bilgi çıkarmak için veri madenciliği süreci detaylı olarak ele alınmaktadır. Bu kapsamda algoritmaların sonuçları, geleneksel başarı oranı metrikleri üzerinden değil, daha detaylı ve yeni bir görüş olan her bir verinin kendi sınıfına ait olma durumu üzerinden de ele alınmaktadır. Algoritmaların performansları da daha detaylı analizi irdelenmiştir. Benzer şekilde literatürde bu veri seti için modellerin başarı oranı üzerinde durulurken, bu çalışmada başarılı bir tanı için önemli metrikler analiz edilmiştir.

## 2. MATERYAL VE METOT (MATERIAL AND METHOD)

Çalışmanın genel akış diyagramı Şekil 1’de verilmiştir. İlk olarak veri setine ön işlemler uygulanarak eğitim ve test kümesi olmak üzere iki gruba ayrılmaktadır. Daha sonra makine öğrenimi metotları eğitim kümesi ile eğitilmekte ve test veri seti ile test edilmektedir. Son olarak modellerin performansları metrikler ile değerlendirilmektedir.



Şekil 1. Metodoloji şematik diyagramı (Schematic diagram of methodology)

### 2.1. Materyal (Material)

Meme kanseri veri seti Kaliforniya Üniversitesi makine öğrenmesi veri tabanından alınmıştır [6]. Farklı zamanlarda elde edilmiş olan veri seti sekiz gruptan oluşmakta, toplam 699 örnekten (458 adet iyi huylu ve 241 adet kötü huylu) oluşan bu veri setinin gruplara göre dağılımı Tablo 1’deki gibidir.

Tablo 1. Veri Setinin Gruplara Dağılımı  
(Distribution of the Data Set into Groups)

Grup	Örnek Sayısı	Yayın Tarihi
Grup 1	367	Ocak 1989
Grup 2	70	Ekim 1989
Grup 3	31	Şubat 1990
Grup 4	17	Nisan 1990
Grup 5	48	Ağustos 1990
Grup 6	49	Ocak 1991
Grup 7	31	Haziran 1991
Grup 8	86	Kasım 1991

Veri setinde bulunan her bir örnek 11 özellik içermekte ve örneklerde eksik bilgi bulunmamaktadır. Özellikler hücre boyutu ve şeklinin homojenlikleri, kanser hücresinin kümelene derecesi, hücrelerin birbirine yapışması, epitel hücrelerinin boyutu, çekirdek yoğunluğu örnek kodu ve bölünme özellikleri iken son özellik bir sınıf bilgisidir. Sınıf bilgisi 2 ya da 4 seçeneklerinden oluşmaktadır. 2 seçeneği tümörün iyi huylu, 4 seçeneği tümörün kötü huylu olduğunu göstermektedir. Makine öğrenimi uygulamalarında tanı için kullanılan metotların başarımlarını etkileyen parametrelerin kolay ayarlanabilmesi, bu parametrelerin etkisinin kolay analiz edilebilmesi gibi avantajlar sunan Weka (3.8.6) yazılım aracı kullanılmıştır.

### 2.2 Metot (Method)

WDBC veri setine makine öğrenimi uygulamalarından doğruluk oranı yüksek olan k-NN, NB, J48 ve DVM gibi sınıflandırma algoritmaları ile kümeleme algoritmalarından k-means ve hiyerarşik kümeleme yöntemleri uygulanarak elde edilen sonuçlar ile algoritmaların performansları değerlendirilmiştir.

#### 2.2.1 Sınıflandırma Yöntemleri (Classification Methods)

Sınıflandırma probleminin çözümü, otomatik sınıflandırma yapmak amacıyla nesnelere oluşan veri kümesini test kümesi ve öğrenme kümesi olarak iki gruba ayırmaktır. Burada nesnelere niteliklerden oluşmakta ve niteliklerden biri ait olduğu sınıf bilgisini taşımaktadır. Sınıfın niteliğini belirlemek için tüm özellikler kullanılarak bir model oluşturulmaktadır. Ardından, test kümesinde bulunan ve sınıfı bilinmeyen nesnelere, oluşturulan model kullanılarak en uygun sınıflara atanmaktadır. Yani, bağımsız değişkenler için sınıf tahmini yapılmaktadır [15].

Çalışmada kullanılacak sınıflandırıcıların seçimi için yapılan ön uygulamalar kapsamında başarı oranı yüksek olan k-NN, NB, J48 ve DVM algoritmaları seçilmiştir. Tüm sınıflandırıcılar için, en çok 10 kat çapraz doğrulama (cross validation) verimli olmuştur. Çapraz doğrulama, genellikle öğrenme algoritmalarını veya modellerini kontrol etmek ve değerlendirmek için kullanılan istatistiksel bir tekniktir. Bu teknik, veriyi belirlenen sayıda gruplara ayırmaktadır, bu veri grubundan bir tanesini modeli test etmek ve diğerlerini modeli eğitmek için kullanmaktadır.

Sınıflandırıcı performans değerlendirmesi için hata matrisi ve bu matrisle hesaplanan performans ölçüm teknikleri olan F1 skorlama, doğruluk, duyarlılık, kesinlik metrikleri kullanılmaktadır. Bu çalışmada performans değerlendirmesinde kullanılan teknikler ve formüller aşağıda verilmektedir [16].

**Karmaşıklık matrisi (Confusion matrix):** "Hata matrisi" veya "karmaşıklık matrisi", bir sınıflandırma modelinin performansını ölçmek için yaygın kullanılan araçlardan biridir. Bu matris verilerin gerçek sınıf etiketleri ile modelin tahmin ettiği sınıf etiketlerinin birbirleriyle

karşılaştırılmasını sağlar. Karşılaştırma sonucunda, doğru sınıflandırılan örnekler (doğru pozitifler ve doğru negatifler) ile yanlış sınıflandırılan örnekler (yanlış pozitifler ve yanlış negatifler) arasındaki ilişki net bir şekilde gösterilmektedir. Karmaşıklık matrisi, sınıflandırma problemlerinde modellerin performans değerlendirmesinde tek başına kullanılan bir ölçüt olmamakla birlikte genellikle şu dört temel değeri içerir.

**Doğru Pozitif (DP):** Modelin doğru bir şekilde pozitif olarak sınıflandırdığı örneklerin sayısı.

**Yanlış Pozitif (YP):** Modelin yanlış bir şekilde pozitif olarak sınıflandırdığı örneklerin sayısı.

**Doğru Negatif (DN):** Modelin doğru bir şekilde negatif olarak sınıflandırdığı örneklerin sayısı.

**Yanlış Negatif (YN):** Modelin yanlış bir şekilde negatif olarak sınıflandırdığı örneklerin sayısı.

Bu değerler, modelin doğruluğunu, hassasiyetini, özgüllüğünü ve duyarlılığını hesaplamak için kullanılmaktadır. Karmaşıklık matrisi, modelin performansını anlamak ve geliştirmek için önemli bir araçtır.

**Doğruluk (Accuracy):** Doğruluk, modelin doğru tahminlerin toplam tahminlere oranıdır. Yüksek bir doğruluk, modelin genel olarak doğru tahminler yaptığını gösterir. Ancak, dengesiz sınıf dağılımına sahip veri kümelerinde doğruluk tek başına yeterli olmayabilir. Örneğin 1000 örnek içeren bir veri setinde a sınıfına ait 990 ve b sınıfına ait 10 örnek olsa yapılacak bir sınıflandırmada işleminde verilerinin tamamının a sınıfına atanması durumunda bu işlemde başarı oranı %99 olacaktır. Örnekte görüldüğü gibi bir modelin performansını değerlendirmede doğruluk kriteri tek başına yeterli görünmemektedir. Karmaşıklık matrisi sonuçları Denklem 1’de verilen ifade ile değerlendirilerek yöntemin doğruluk oranı elde edilir.

$$\text{Doğruluk} = \frac{DP+DN}{DP+DN+YP+YN} \quad (1)$$

**Kesinlik (Precision):** Kesinlik veya hassasiyet, pozitif olarak tahmin edilen örneklerin gerçekten pozitif olma oranını gösterir. Yüksek hassasiyet, yanlış pozitiflerin az olduğunu ve modelin yanlış alarm verme olasılığının düşük olduğunu gösterir. Özellikle yanlış pozitiflerin maliyeti yüksek olduğunda önemlidir. Denklem 2’de verilen eşitlik ve karmaşıklık matrisi kullanılarak modelin kesinlik oranı elde edilir.

$$\text{Kesinlik} = \frac{DP}{DP+YP} \quad (2)$$

**Duyarlılık (Recall, Sensitivity):** Pozitif olarak tahmin etmemiz gereken örneklerin ne kadarını pozitif olarak tahmin ettiğimizi gösteren bir metriktir. Özellikle Yanlış Negatif’e odaklanıldığında duyarlılık hesabı önem kazanmaktadır. Denklem 3 ve 4’te verilen ifadeler karmaşıklık matrisi verilerine uygulanarak duyarlılık ile Özgüllük oranı hesaplanmaktadır.

$$\text{Duyarlılık} = \frac{DP}{DP+YN} \quad (3)$$

$$\text{Özgüllük} = \frac{DN}{DN+YP} \quad (4)$$

**F1 Skoru (F1 score):** Homojen dağılıma sahip olmayan veri setlerinde model performansı analizinde kullanılan F1 skoru, hassasiyet ve duyarlılığın harmonik ortalaması olarak tanımlanır. Dengesiz veri setlerinde kesinlik ve duyarlılık performans ölçümlerinde kullanılan önemli ölçütlerdendir. F1 skoru ise kesinlik ve duyarlılık ölçütlerini kullanarak dengesiz veri setleri için model performansını daha doğru bir şekilde ölçmektedir. Homojen dağılıma sahip olmayan veri setlerinde ve yanlış pozitifler ile yanlış negatifler arasında bir denge kurmak istendiğinde faydalıdır. Yüksek bir F1 skoru, hem yanlış pozitiflerin hem de yanlış negatiflerin az olduğunu ve dolayısıyla modelin genel olarak iyi performans gösterdiğini belirtir. Denklem 5’te verilen ifade karmaşıklık matrisi verilerine uygulanarak modelin F1 skoru oranı tespit edilir.

$$\text{F1 Skoru} = 2 * \frac{\text{Duyarlılık} * \text{Hassasiyet}}{\text{Duyarlılık} + \text{Hassasiyet}} \quad (5)$$

### 2.2.1.1 k-NN Sınıflandırıcı Algoritması (k-NN Classifier Algorithm)

k-NN, sınıflandırma problemi çözümü için kullanılan en sade makine öğrenmesi algoritması olarak kabul edilebilir. Temel prensibi, bir veri noktasını sınıflandırmak için çevresindeki k adet en yakın komşusunun sınıf bilgisine dayanır. k-NN, öğrenme süreci içinde veri setini öğrenmez; bunun yerine, sınıflandırma yapılması istenen yeni bir veri noktası geldiğinde, bu noktaya en yakın k adet komşularının verilerini kullanarak sınıflandırmayı gerçekleştirir. Sınıflandırma işlemi yapılırken komşulukların çoğunluğu dikkate alınmaktadır. Eşitlik olmaması adına k değeri genellikle pozitif tek sayı olarak belirlenir. Sınıflandırılmak istenen yeni verinin, mevcut verilere olan uzaklığı hesaplanıp, k sayıda en yakın komşuluğuna bakılmaktadır. Mesafe hesaplamaları için birçok mesafe fonksiyonu kullanılmaktadır. Literatürde yaygın olarak kullanılan mesafe ölçütleri Oklid, Minkowski ve Manhattan uzaklıklarıdır.

k-NN algoritması, parametrik olmayan bir tembel öğrenme (lazy learning) algoritmasıdır. Lazy learning’in bir eğitim aşaması yoktur; yani eğitim verilerini öğrenmez, bunun yerine eğitim veri kümesini ezberler. Bir sınıflandırma işlemi gerçekleştirmek istendiğinde, tüm veri seti içerisinde en yakın komşuları arar. Algoritmanın çalışmasında bir k parametresi belirlenir. Bu k parametresi bir veri noktasını sınıflandırmak için çevresindeki k en yakın komşusunu seçmek için kullanılır. Yeni bir değer geldiğinde en yakın k adet eleman alınarak gelen değer arasındaki uzaklık hesaplaması yapılır. Uzaklık hesaplama işlemlerinde genelde en yaygın olarak Oklid mesafe fonksiyonu kullanılmaktadır. Oklid fonksiyonu haricinde Manhattan ve Minkowski fonksiyonları da alternatif olarak kullanılabilir. Belli fonksiyonlara göre uzaklık hesabı

yapılmakta, ardından uzaklık hesapları sıralanmakta ve gelen değerler uygun olan sınıfa atanmaktadır [17,18]. Yaygın kullanılan uzaklık hesaplama formülleri Denklem 6, 7 ve 8'de verilmiştir. Bu formüller, ilgili veri setindeki her bir örnek sınıflandırılırken, yeni bir veri noktasının hangi sınıfa ait olduğunu, komşularıyla olan mesafeyi hesaplayarak belirlemektedir.

$$\text{Öklid (Euclidean)} = \sqrt{\sum_{i=1}^k (x_i - y_i)^2} \quad (6)$$

$$\text{Manhattan} = \sum_{i=1}^k |x_i - y_i| \quad (7)$$

$$\text{Minkowski} = \left( \sum_{i=1}^k (|x_i - y_i|^p) \right)^{1/p} \quad (8)$$

### 2.2.1.2 Naive Bayes Sınıflandırıcı Algoritması (Naive Bayes Classifier Algorithm)

Naive Bayes (NB) sınıflandırıcısının temeli Bayes teoremine dayanan istatistiksel sınıflandırma problemlerinde kullanılan bir makine öğrenimi algoritmasıdır. Tembel öğrenme algoritmalarından biri olan NB dengesiz veri kümelerinde de çalışabilmektedir. Algoritma, sistemdeki değişikliklere kendini adapte edebilir başka bir ifade ile yeni gelen örnekler olduğunda değişikliklere duyarlı olabilmektedir. Algoritmanın çalışma şekli, ilgili örnek için her bir durum olasılığı hesaplanarak olasılık değeri en yüksek olan sınıfa göre sınıf bilgisi belirlenmektedir. Örneklerin hangi sınıfa ve hangi olasılıkla ait olduklarını belirleyen NB, düşük boyutlu veri setiyle başarılı sonuçlar ortaya koyabilmektedir. Eğitim kümesinde belirlenemeyen bir değer varsa, model test kümesinde tahmin yaparken bu değer için genellikle bir olasılık değeri verilmez (değeri 0 olarak verir) ve tahmin yapılamaz çünkü olasılık hesabında sonucun 0 çıkmasına (ilgili veride 0 kayıt olması durumuna) sebep olmaktadır. Bu durum, Sıfır Frekans yani Zero Frequency adıyla da bilinir. Bu sorunu çözebilmek için çeşitli düzeltme yöntemleri kullanılmakla birlikte Laplace yöntemi (Denklem 9) en basit düzeltme yöntemlerinden biri olarak öne çıkmaktadır [19,20]. Laplace formülü ile her bir özellik için sınıfa ait olasılıklar çarpılır ve sonuçlar karşılaştırılarak en yüksek olasılığa sahip sınıf seçilmektedir.

$$\text{Laplace Formülü: } P(X|C_i) = \prod_{k=1}^n P(X_k|C_i) \quad (9)$$

NB sınıflandırıcısının davranışını değiştirmek ve sınıflandırma performansını artırmak/iyileştirmek için useKernelEstimator ve useSupervisedDiscretization parametreleri optimize edilir. "useKernelEstimator", normal dağılım yerine sayısal nitelikler için çekirdek kestirimcisi kullanan bir parametredir. Weka'da varsayılan olarak her sayısal özellik için Gauss dağılımı kabul edilir. Algoritma, useKernelEstimator argümanı ile çekirdek kestirimcisini kullanacak şekilde değiştirilebilir, bu da veri kümesindeki niteliklerin gerçek dağılımıyla daha iyi sonuçlar verebilir. "useKernelEstimator" parametresi, bir boolean (true/false) değer alır. Varsayılan olarak, bu parametre pasif (false) seçeneğindedir, yani çekirdek

tahmincisi kullanılmaz. useSupervisedDiscretization parametresi ise değişken dönüştürme işleminin kullanılıp kullanılmayacağını belirler yani sayısal özellikleri nominal özelliklere otomatik olarak dönüştürebilmektedir. Bu parametrenin varsayılan değeri pasif olup, veri setindeki sürekli özelliklerin sürekli değerlerini belirli aralıklara bölerek verileri sınıflandırmaya yardımcı olur. Söz konusu parametrelerin farklı değerlerle deneysel olarak test edilmesi ve performanslarının değerlendirilmesi önemlidir. Çünkü bazen bu parametreler veri seti üzerinde olumsuz etki oluşturabilmektedir [21].

### 2.2.1.3 J48 (Karar Ağacı) Sınıflandırıcı Algoritması (J48 (Decision Tree) Classifier Algorithm)

Bu algoritma, karar ağacını daha sade, daha küçük, daha iyi optimize edilmiş ve verimli hale getirmeyi amaçlamaktadır. Bunun için de değişkenlerin/özelliklerin entropi ve bilgi kazanımı (information gain) değerlerini esas alır. Entropi ve bilgi kazanımı konuları literatürde geniş bir şekilde bulunduğu için aşağıda konu hakkında özet bilgilere verilmiştir.

Entropi, rastgele bir değişkenin belirsizliğinin ölçüsü olarak ifade edilmektedir.

Bilgi Kazanımı: Veriler bölümlendiğinde, hedef değişkendeki belirsizliğin ne kadar değiştiğinin ölçüsüdür. Başka bir ifade ile yeni bilgilerin öğrenilmesi olarak ifade edilebilir.

C4.5 veya J48 algoritması, ilk olarak hedef değişken için entropi değerini hesaplar. Daha sonra, her bir değişkenin veya özelliğin bilgi kazanımını hesaplar ve bu şekilde en yüksek bilgi kazanımı değerine sahip tahmin edici sınıfı tespit eder [22]. En yüksek bilgi kazanımı değerine sahip özellik/değişken, ağacın en üst nodunda (düğüm) yer almaktadır. Yani en iyi bilgi kazanımı sonucunu veren özellik/değişken, karar(dallanmanın başladığı nokta) olarak belirlenir. Ardından alt düğümler için de tüm özelliklere/değişkenlere aynı işlemler tekrarlanır [23]. Bu şekilde karar ağacının daha dengeli bölünmesi beklenmektedir. Bu hesaplamalarda kullanılan formüller ise Denklem 10, 11 ve 12'de verilmiştir. Denklem 10'da verilen eşitlik yardımıyla veri kümesindeki belirsizlik veya düzensizlik (entropi) belirlenmektedir. Değişkenin bilgisi (Denklem 11), bir değişkenin (özelliğin) veri kümesindeki belirsizliği ne kadar azalttığını ifade ederken bilgi kazanımı (Denklem 12) ise bir değişkenin sınıflandırmada ne kadar fayda sağladığını göstermektedir. Başarılı bir sınıflandırma, veri bir özelliğe göre bölündüğünde entropinin minimum ve bilgi kazanımının maksimum olması beklenir.

$$\text{Entropi : } \text{Info}(D) = - \sum_{i=1}^m (p_i \log_2 p_i) \quad (10)$$

$$\text{Değişkenin Bilgisi: } \text{Info}_A(D) = - \sum_{j=1}^V \left( \frac{|D_j|}{D} \times \text{Info}(D_j) \right) \quad (11)$$

$$\text{Bilgi Kazanımı: } \text{Gain}(A) = \text{Info}(D) - \text{Info}_A(D) \quad (12)$$

Sistemin aşırı eğitilmesi (overfitting) sonucunda esnekliğini kaybetmesi ve ezberlemeye yakın bir sonuç oluşturması, gereksiz yere fazla detay içermesi muhtemeldir. Bu durumu önlemek için ağaç dallarında budama (pruning) yapılır. Ağaç oluşturulurken ön budama (prepruning) ve ağaç oluşturulduktan sonra ise son budama (postpruning) yapılmaktadır. Bu sayede aşırı öğrenmenin önüne geçilebilmektedir.

J48 algoritmasının aşağıda bulunan bazı parametreleri sınıflandırmadaki başarı oranını etkileyebilmektedir:

**Güven Faktörü (confidenceFactor):** Bu parametre, karar ağacının dengeli bir şekilde büyümesine katkıda bulunan budama işleminin etkinliğini artırmak için kullanılır. Bu parametre küçük değerler aldığı anda, daha fazla budama işlemi gerçekleştirilmesine olanak tanır [24].

**MinNumObj:** Her yaprakta/düğümde bulunması gereken kayıt sayısı verisini belirtmektedir.

**doNotMakeSplitPointActualValue:** Algoritmanın hangi değerleri ayırım noktası (split point) olarak kullanılmaması gerektiğini belirlemesini sağlar. Karar ağaçları, veri kümesindeki belirli özelliklerin (features) değerlerine göre veri noktalarını bölümlere ayırır. Bu bölme işlemi, belirli bir eşik değeri (split point) kullanılarak gerçekleştirilir. Ancak bazen bu eşik değerlerinin belirli bir aralıkta veya belirli bir değerin altında veya üstünde olması istenmez. "Dont make split point actual value" parametresi, bu tür durumlarda kullanılır. Bu parametre, algoritmanın belirli bir eşik değeri kullanarak bölme yapmasını engellemek için kullanılır. Örneğin, belirli bir özelliğin değeri 0 ile 100 arasında değişiyorsa, ve bu parametre 50 olarak ayarlanırsa, algoritma bu özellik için 50'nin bir split point olarak kullanılmasını engeller. Bu parametre, belirli bir özellik için kullanılabilir tüm değerlerin ayırım noktası olarak kullanılmasını önleyerek modelin genelleştirilebilirliğini artırabilir veya istenmeyen ayrışmaları engelleyebilir. Bu nedenle, veri kümesine ve problem alanına bağlı olarak, bu parametrenin kullanımı modelin performansını artırabilir veya istenmeyen sonuçları önleyebilir.

#### 2.2.1.4 Destek Vektör Makinesi Sınıflandırıcı Algoritması (Support Vector Machine Classifler Algorithm)

DVM sınıflandırma ve regresyon problemleri için kullanılan bir makine öğrenimi algoritmasıdır. Bir makine öğrenimi algoritması olan DVM genellikle çift özdeş olmayan doğrusal programlama problemlerini çözmek için kullanılan optimizasyon tekniklerini içerir. Aynı zamanda özel bir optimizasyon algoritmasıdır, büyük problemleri daha küçük alt problemlere bölerek çözüme özelliğine sahiptir. Bu algoritma, özellikle büyük veri setleri veya yüksek boyutlu özellik uzayları gibi durumlar için etkili bir seçenek olabilir.

DVM'nin çalışma prensibi, sınıflandırma için bir düzlemde bulunan veri kümeleri arasına sınırlar çizilerek gruplara ayırmak mümkündür. Bu sınırlar gruplar arasındaki en uzak yere çizilmektedir [25]. DVM algoritması her bir sınıfın en yakın veri noktaları arasında en fazla mesafeye sahip olan hiperdüzlemi bularak başlar. Hiperdüzleme en yakın mesafede olan veri noktaları ise destek vektörleri olarak tanımlanır ve hiperdüzlemi tanımlamak için kullanılır. Algoritma bu destek vektörlerini kullanarak veriler arasındaki farklı sınıfları ayıran bir karar sınırı oluşturur. Belki de iki sınıfı ayıracak sonsuz sayıda doğru çizilebilir fakat DVM iki grup/sınıf arasında en fazla aralığa sahip olan doğruyu seçer [26,27].

DVM algoritması, yeni bir veri noktasını sınıflandırmak için hiperdüzlemdeki konumunu (hiperdüzlemin hangi tarafına düştüğünü) kullanır. Veri noktasının hiperdüzlemin hangi tarafında bulunduğuna bağlı olarak, ilgili sınıfa atanır yani hiperdüzlemin bir tarafına düşer ise bir sınıfa ait olarak sınıflandırılır, diğer tarafa düşer ise diğer sınıfa ait olarak sınıflandırılır. Bu, DVM'lerin, regresyon ve sınıflandırma gibi denetimli öğrenme görevlerinde güçlü bir araç olmasını sağlar. Özellikle, yüksek boyutlu veri alanlarında, farklı sınıfları ayıran veya gerçek ve tahmin edilen değerler arasındaki hatayı en aza indiren bir hiperdüzlem bulma stratejisi izlerler. Bu şekilde, verileri daha iyi anlamak ve öngörülemeden desenleri keşfetmek için kullanılabilirler [28]. Düzlem ve boyutlar birer özellik (attribute) olarak düşünülebilir, her girdiyi gösteren farklı bir nokta elde edilmektedir ve ardından bu girdiler sınıflandırılmaktadır. Doğrusal (Linear) kernel haricinde polykernel, RBF (Radial Basis Function) gibi fonksiyonlar da kullanılabilirler.

Çok sınıflı örneklerde her bir sınıf arasında ayırma yapılmakta ve yeni gelen örnek buna göre sınıflandırılmaktadır. DVM'nin önemli parametresi olan "C" (Cost), sınıflandırıcı performansını etkileyen kritik bir faktördür.

#### C (Cost) Parametresi:

**Amacı:** C değeri, DVM'nin düzenleme gücünü kontrol eder. Yüksek C değerleri, eğitim verilerine daha fazla vurgu yapılmasını ve karar sınırlarının daha fazla düzenlenmesini sağlar, bu da modelin eğitim verilerine daha sıkı uymasına neden olabilir.

**Etkisi:** Yüksek C değerleri overfitting eğilimindedir, yani model eğitim verilerine çok fazla uyar ve genelleme yeteneği düşer. Düşük C değerleri, daha genel geçer modellere yol açabilir, ancak eğitim verilerine daha az uyarlanabilirler.

#### 2.2.2 Kümeleme Yöntemleri (Clustering Methods)

Kümeleme Yöntemleri, keşifsel veri analizi yöntemi olup, bir dizi veri ögesini, bir uzaklık (veya benzerlik) ölçüsüne dayalı olarak gruplara/kümelere/bölümlere ayırmayı amaçlamaktadır. Ya da kısaca birbiriyle özdeş/yakın özellikte olan verilerin tek bir grupta toplanması olarak da

ifade edilebilmektedir. Bu gruplara "küme" denir ve sayıları önceden belirlenebileceği gibi algoritmalar tarafından da belirlenebilmektedir. Burada ilgili veri seti için kümeleme sonuçları incelenirken ayrıca Classes to Clusters Evaluation seçeneği ile değerlendirme yapılmaktadır. "Classes to Clusters Evaluation", kümeleme (clustering) işlemi sonucunda oluşturulan kümelemelerin sınıflandırma (classification) hedefiyle karşılaştırılmasını sağlayan bir değerlendirme yöntemidir. Bu işlem, genellikle denetimli öğrenme (supervised learning) ve denetimsiz öğrenme (unsupervised learning) yöntemlerini bir araya getirir.

Söz konusu yöntem, iki ana bileşenden oluşur:

**Sınıflar (Classes):** Veri kümesindeki örneklerin ait olduğu sınıfları belirtir. Sınıflar genellikle bir etiket veya kategori olarak temsil edilir ve her bir veri örneği için bir etiket atanır.

**Kümelemeler (Clusters):** Veri kümesindeki örneklerin benzerliklerine dayanarak oluşturulan veri gruplarını temsil eder. Kümeleme algoritmaları, veri örneklerini birbirine benzer olan gruplara böler.

"Sınıfların Kümelemelere Atanması (Classes to Clusters)" değerlendirmesi, bu iki bileşeni karşılaştırır ve sınıfların, kümelemelerle nasıl ilişkilendirildiğini değerlendirir. Genellikle, her sınıfın hangi kümelerle ait olduğunu ve kümeleme sonuçlarının sınıflandırma doğruluğunu ölçer.

Bu işlem, doğru sınıfların doğru kümelere atanmasını göz önünde bulundurarak kümeleme algoritmasının performansını ölçer. İdeal olarak, her sınıfın bir kümeyle tam olarak eşleşmesi ve sınıflandırma doğruluğunun maksimum düzeyde olması beklenir. Ancak, gerçek dünyada bu durum genellikle mümkün olmayabilir ve bu nedenle değerlendirme işlemi, sınıfların ve kümelemelerin ne kadar iyi eşleştiğini nicel olarak ölçmeye çalışır.

Sınıflandırma bilgisine dayalı kümeleme problemlerinde algoritmalarının performansını anlamak ve geliştirmek için bu değerlendirme işlemi önemli olmaktadır.

#### 2.2.2.1 k-Means Kümeleme Algoritması (k-Means Clustering Algorithm)

k-means algoritması, kümeleme algoritmalarının içinde muhtemelen en eski ve yaygın olarak kullanılan basit bir algoritmadır. Eğitimsiz/Eğitimsiz (Unsupervised) öğrenme prensibine sahiptir. Bu algoritmanın avantajları ve dezavantajları bulunmakla birlikte, büyük veri kümelerinde hızlı çalışması nedeniyle popülerlik kazanmıştır.

k-means algoritmasında, kümelenecek olan verilerden her biri yalnızca bir küme üyesi olarak atanabilir ki bu da disjoint kümeleme mantığını ifade etmektedir. Disjoint kümelemede kümeler birbirinden belli bir hatla ayrılabilen, birbirine girmemiş kümeleme yöntemidir. Bu kümelerin

temsil edildiği noktalar ise merkez noktasını göstermektedir. Bu algortmada kullanılacak verinin bölüneceği küme sayısını, kullanıcının manuel girmesine bağlı olarak belirlemesi durumu bulunmaktadır, "numCluster" parametresi bu işe yaramaktadır yani kullanıcının belirlemesi gereken ve k-means algoritmasının veri noktalarını kaç kümeyle böleceğini belirten bir parametredir. Bu sebeple doğru küme sayısı belirlenme durumu bitene kadar deneme yanılma yöntemine başvurulması gerekebilmektedir. Bazen k-means işleminin başarılı şekilde tamamlanması için fonksiyonun birkaç kez çağırılması gerekebilmektedir. Çünkü kümelerin içinde ilk seferde oluşan benzerlik uyumu doğru sonuç vermeyebilir/tutmayabilir. Daha sonra kümelere değişimin durması yani uyumun tutması istenilen sonucun alındığı anlamına gelebilmektedir [29,30].

Kısaca bu algortmada, veriler iki boyutlu uzay üzerine serilir. Hedef kümeler tanımlanır, ardından belli mesafe fonksiyonlarına göre (Öklid, Manhattan vb.) örneklerin hedeflere mesafesi hesaplanır. Hedefler belirlenirken rastgele başlangıç merkezlerinin belirlenmesinde "seed" başlangıç değeri kullanılmaktadır, daha sonra merkezin etrafındaki örnekler kümeyle dahil oldukça merkez nokta değişebilmektedir. Tüm kümeleme işlemi bu adımlarla tamamlanmaktadır [29].

#### 2.2.2.2 Hiyerarşik Kümeleme Algoritması (Hierarchical Clustering Algorithm)

Hiyerarşik kümeleme, parametrelerin belirsiz olduğu durumlarda ya da kaç bölütün/kümenin oluşturulacağını bilmediği veya duruma göre değiştiği problemlerde kümeleme yerine hiyerarşi oluşturmaya yardımcı olan algortmadır.

Hiyerarşik algortmalarda aşağıdan yukarıya (Agglomerative/AGNES) ve yukarıdan aşağıya (Divisive/DIANA) olmak üzere iki adet yöntem bulunmaktadır. Aşağıdan yukarıya kümeleme mantığına göre, verilerin her biri başlangıç aşamasında tekil bir küme olarak ele alınır ve benzerlikleri en yüksek olan veri noktaları bir araya getirilerek kümelendir. Kümeleme işlemi için örnekler/özellikler arasındaki ilişkilere göre ilk olarak ikili ilişkileri içeren bağlantılara göre ve ardından daha fazla sayıda ilişkileri içeren bağlantılara göre hiyerarşi oluşturulmaktadır. Bu işlem kümelenecek/bölümlenecek başka bir veri kalmayınca kadar devam etmektedir. Sonuç ağacı dendrogram ile gösterilmektedir [30,31]. Yukarıdan aşağı kümeleme yaklaşımında örneklerin tamamı bir bütün olarak ele alınmakta ve alt gruplara bölünmektedir. Örneğin, önce iki gruba bölünmekte sonra her grup kendi içinde daha alt gruplara bölünerek ilerlenmekte ve sonuç ağacı yine dendrogram ile gösterilmektedir [31].

## 2. BULGULAR (FINDINGS)

Hastalık verilerini sınıflandırma ve kümeleme için modeller, bağımsız uygulamalarda eğitim veri seti ile eğitilerek test veri seti ile performansları değerlendirilmektedir. Uygulamalardan elde edilen bulgular alt başlıklarda verilmektedir.

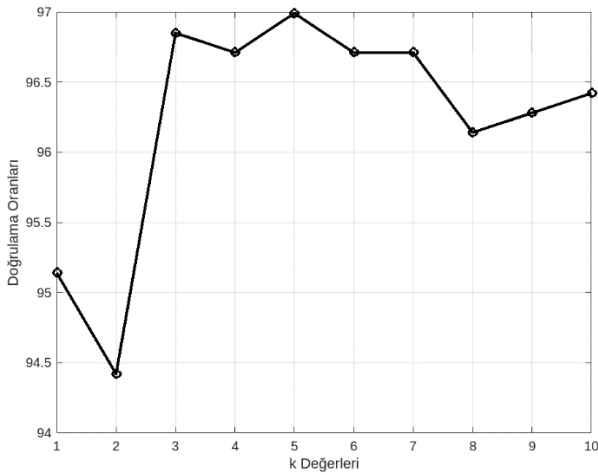
### 3.1 *k*-NN Sınıflandırıcı Uygulaması (*k*-NN Classifier Application)

*k*-NN sınıflandırıcısındaki *k* değeri (kullanılacak komşu sayısı) etkisinin incelendiği sonuçlar Tablo 2 ve Şekil 2'de sunulmuştur. Bu sonuçlara göre,  $k = 5$  koşulu en yüksek doğruluğa sahiptir. Buna göre %96,99'luk bir doğrulama oranı mevcuttur.

Tablo 2. Farklı *k* değerleri için *k*-NN sınıflandırıcının sınıflandırma ve doğrulama sonuçları

(Classification and validation results of *k*-NN classifier for different *k* values)

<i>k</i> Değeri	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğruluk Oranı(%)
<i>k</i> =1	665	34	95,14
<i>k</i> =2	660	39	94,42
<i>k</i> =3	677	22	96,85
<i>k</i> =4	676	23	96,71
<i>k</i> =5	678	21	96,99
<i>k</i> =6	676	23	96,71
<i>k</i> =7	676	23	96,71
<i>k</i> =8	672	27	96,14
<i>k</i> =9	673	26	96,28
<i>k</i> =10	674	25	96,42



Şekil 2. *k*-NN sınıflandırıcının  $k=1-10$  değerleri için başarı oranları (Success rates of *k*-NN classifier for  $k=1-10$  values)

$k=5$  değeri için en iyi doğrulama sonucunun alındığı *k*-NN sınıflandırıcıya ait karmaşıklık matrisi Tablo 3'te verilmiştir.

Tablo 3.  $k=5$  değeri için karmaşıklık matrisi (Complexity matrix for  $k=5$ )

<i>k</i> Değeri	a	b	
<i>k</i> =5	444	14	a=2
	7	234	b=4

\* a veya 2 değeri iyi huylu tümörü, b değeri veya 4 kötü huylu tümörü temsil etmektedir

Tablo 4. *k*-NN sınıflandırıcı için performans değerlendirme ölçütleri

(Performance evaluation metrics for *k*-NN classifier)

DP	DN	YP	YN	Duyarlılık	Kesinlik	F1 Skorlama
444	234	7	14	0,9694/0,9707	0,9844/0,9435	0,9768/0,9570

Duyarlılık, Kesinlik ve F1 Skorlama değerleri ilki iyi huylu ve sonraki kötü huylu tümör değerleri için verilmiştir.

Karmaşıklık matrisi ile elde edilen DP, DN, YP ve YN parametreleri kullanılarak modelin performansı belirlenmiştir. Algoritma performans ölçütlerinden duyarlılık, kesinlik ve F1 skorlama değerleri hem iyi huylu hem de kötü huylu tümör için ayrı ayrı hesaplanarak Tablo 4'te verilmektedir. Tablo 4 incelendiğinde modelin verileri doğru sınıflandırmada oldukça duyarlı olduğu yani iyi huylu sınıfa ait verilerin kendi sınıfına ve kötü huylu sınıfa ait verilerin kendi sınıfına yüksek oranda atıldığı görülmektedir. Kesinlik ölçütüne bakıldığında iyi huylu verilerin sınıflandırılmasında yine oldukça yüksek bir oran elde edilmiştir. Yani pozitif değerler büyük ölçüde pozitif olarak belirlenmiştir. Ancak kötü huylu verilerin kötü olarak sınıflandırmada yani kötü huylu verilerin tespitinde modelin yeterince hassas olmadığı söylenebilir. F1 skoru yüksek olan Algoritmanın, homojen dağılımlı bir veri setindeki verileri yüksek başarı oranıyla sınıflandırıldığı ifade edilebilir.

### 3.2 Naive Bayes Sınıflandırıcı Uygulaması (Naive Bayes Classifier Implementation)

NB sınıflandırıcının uygulama verileri Tablo 5'te verilmiştir ve en iyi sonucu veren uygulamanın karmaşıklık matrisi Tablo 6'da sunulmuştur. NB sınıflandırıcıda useKernelEstimator ve useSupervisedDiscretization parametreleri öncelikle pasif durumda iken sınıflandırma yapılmış, ardından bu parametreler ayrı ayrı aktifleştirilerek sınıflandırma tekrar yapılmış ve doğrulama sonuçları buna göre bulunmuştur. Sınıflandırma sonuçlarına göre useKernelEstimator parametresinin aktif olduğu durumda elde edilen %97,42'lik doğrulama oranı en yüksek orandır. Yöntemin doğrulama sonuçları da Tablo 7'de verilmiştir.

Tablo 5. NB Sınıflandırıcının Sınıflandırma ve Doğrulama Sonuçları  
(Classification and Validation Results of NB Classifier)

Sınıflandırıcı/Parametre	NB	NB <sup>1*</sup>	NB <sup>2*</sup>
Doğru Sınıflandırma	671	681	679
Yanlış Sınıflandırma	28	18	20
Doğrulama Oranı	95,99	97,42	97,14

<sup>1</sup>NB sınıflandırıcıda useKernelEstimator parametresi kullanılmıştır. <sup>2</sup>NB sınıflandırıcıda useSupervisedDiscretization parametresi kullanılmıştır.

Tablo 6. NB (useKernelEstimator) Sınıflandırıcının Karmaşıklık Matrisi  
(NB (useKernelEstimator) Complexity Matrix of Classifier)

a	b	
442	16	a=2
2	239	b=4

\* a veya 2 değeri iyi huylu tümörü, b değeri veya 4 kötü huylu tümörü temsil etmektedir.

Tablo 7. NB Sınıflandırıcının Performans Değerlendirme Ölçütleri  
(Performance Evaluation Measures of NB Classifier)

DP	DN	YP	YN	Duyarlılık	Kesinlik	F1 Skorlama
442	239	2	16	0,9650/0,9917	0,9954/0,9372	0,98/0,9637

Duyarlılık, Kesinlik ve F1 Skorlama değerleri iyi huylu ve kötü huylu tümör değerleri için verilmiştir.

Tablo 6'da NB sınıflandırıcının karmaşıklık matrisi DP, DN, YP ve YN parametreleri verilmektedir. Bu parametreler ile modelin performans ölçütleri olan duyarlılık, kesinlik ve F1 skorlama değerleri hem iyi huylu hem de kötü huylu tümör için ayrı ayrı hesaplanmıştır. Performans ölçütleri NB sınıflandırıcının bu veri seti için başarılı bir yöntem olduğunu göstermektedir. Ancak kötü huylu verilere yüksek oranda duyarlı (0,9917) olan modelin yine kötü huylu veriler için yüksek oranda hassas (0,9372) olmadığı söylenebilir. Ayrıca NB'nin F1 skoru verilerin homojen dağılıma sahip olduğunu göstermektedir.

### 3.3 J48 Karar Ağacı Sınıflandırıcı Uygulaması (J48 Decision Tree Classifier Implementation)

J48 algoritmasında bulunan Güven Faktörü (confidenceFactor) ve yaprak düğümünde bulunması gereken minimum obje sayısı (MinNumObj) parametrelerinin varsayılan değerleri değiştirilerek sınıflandırma başarısının artırılması hedeflenmiştir. Güven faktörü algoritmanın budama aşamasını kontrol etmektedir. Güven değeri bir yaprak düğümünde hata bulunma oranını ifade eder. Düğümde bulunan hata olasılığı güven değerinden fazla ise budanır. Düğümde varsayılan hata olasılıkla ifade edilmektedir. Dolayısıyla Güven değeri 0,5'ten küçük olma durumunda anlamlı olmaktadır. Güven değeri ne kadar az ayarlanırsa budama

o kadar fazla olacaktır ve hataya karşı gösterilen tolerans o kadar az olmaktadır. Bu uygulamada Güven değeri 0,15, 0,20 ve 0,25 için, MinNumObj değeri ise 2, 3, 4, 5 değerleri için uygulanmıştır. Daha sonra en iyi sonucu veren doNotMakeSplitPointActualValue parametresi aktifleştirilmiş ve daha iyi bir sonuç için uygulamalar yapılmıştır. Tablo 8, farklı Güven Faktörü, MinNumObj ve doNotMakeSplitPointActualValue değerleri için J48 algoritmasının başarı oranlarını göstermektedir.

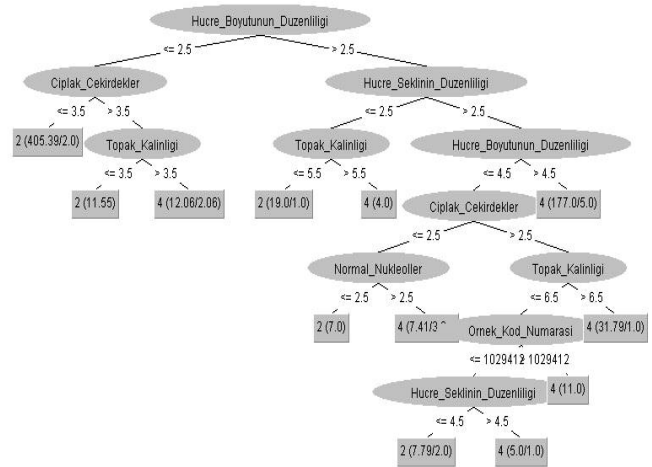
Tablo 8. J48 Sınıflandırıcı İçin Başarı Sonuçları  
(Success Results for J48 Classifier)

Güven Faktörü / MinNumObj Değeri <sup>1*</sup>	2	3	4 <sup>2*</sup>	5
0,15	94,99	95,28	95,28/95,28	95,28
0,20	95,13	94,99	95,42/95,42	95,13
0,25	94,56	95,13	95,42/95,56	94,99

<sup>1\*</sup>Güven Faktörü değerleri 0,15-0,20 ve 0,25 olup MinNumObj Değerleri 2,3,4 ve 5'tir.

<sup>2\*</sup> doNotMakeSplitPointActualValue parametresi kullanılmıştır.

Tablo 8'de görüldüğü üzere en iyi sonuç 0,25 Güven Faktörü değeri, MinNumObj=4 değeri ve doNotMakeSplitPointActualValue parametresinin "true" durumda olduğu sonuç için % 95,56 olarak verilmektedir. Algoritmanın sonuç olarak çıkardığı ağaç yapısı ise Şekil 3'te verilmiştir.



Şekil 3. J48 Ağaç Yapısı (J48 Tree Structure)

J48 algoritması veri kümesini Şekil 3'te verilen ağaç yapısı gibi sınıflandırmıştır. Ağaç irdelendiğinde en tepede Uniformity\_of\_cell\_size özelliği, alt dallarda da diğer özelliklerin hangi sırada seçildiği ve bu kriterlere bağlı olarak tümör örneğinin iyi huylu/kötü huylu kararı gösterilmektedir. Bu seçim Shannon bilgi kuramından yararlanarak entropi hesabıyla seçilmektedir. Bu hesaba göre en iyi sonucu en tepedeki özellik yani Uniformity\_of\_cell\_size vermektedir. Ağacın yapısına dikkat edildiğinde karmaşık ve geniş bir ağaç yerine kısa ve dar bir yapıda olmasından dolayı ağacın daha az dallanma ve daha az düğüm içerdiği ve bu ağaç yapısının performanslı olduğu söylenebilir.



J48 sınıflandırıcının en iyi sonucuna göre karmaşıklık matrisi değeri ise Tablo 9'da verilmektedir.

Tablo 9. J48 Sınıflandırıcı İçin Karmaşıklık Matrisi  
(Complexity Matrix for J48 Classifier)

a	b	
441	17	a=2
14	227	b=4

\* a değeri 2 yani iyi huylu tümörü, b değeri 4 yani kötü huylu tümörü temsil etmektedir.

Tablo 10. J48 Sınıflandırıcı için Performans Değerlendirme Ölçütleri  
(Performance Evaluation Metrics for J48 Classifier)

DP	DN	YP	YN	Duyarlılık	Kesinlik	F1 Skorumla
441	227	14	17	0,9628/0,941 9	0,9692/0,93 03	0,966/0,936

Duyarlılık, Kesinlik ve F1 Skorumla değerleri iyi huylu ve kötü huylu tümör değerleri için verilmiştir.

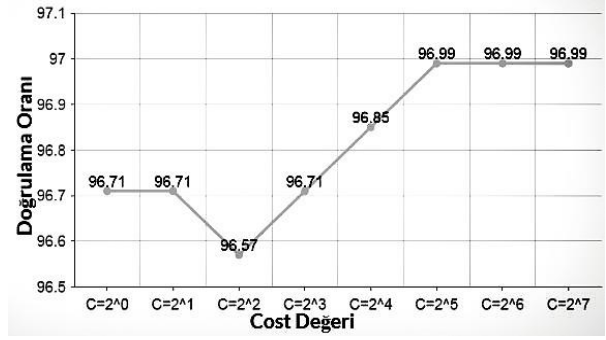
Tablo 9, J48 sınıflandırıcının karmaşıklık matrisini ve Tablo 10 ise modele ait hesaplanan duyarlılık, kesinlik ve F1 skorumla değerlerini göstermektedir. Performans ölçütleri modelin iyi huylu verileri sınıflandırmada daha başarılı olduğunu göstermektedir.

### 3.4 Destek Vektör Makinesi Sınıflandırıcı Uygulaması (Support Vector Machine Classifier Implementation)

DVM algoritması belirli parametrelerle Wisconsin veri setine uygulanarak elde edilen sonuçlar aşağıda verilmiştir. DVM uygulamasında C (Cost) değeri kritik bir parametre olup model performansı için logaritmik bir ölçekte geniş bir aralıktaki değerler ile denir. Pratikte 2'nin katları şeklinde C değeri uygulamaları mevcuttur ancak zorunluluk değildir. Bu değerlerin yüksek olması modelin performansını artırmaktadır ancak ezberlemeye de (overfitting) yol açabilir. Bu çalışmada 1 (20) den 128 (27) e kadar C değeri denenmiştir ve Tablo 11'de DVM sınıflandırıcının doğruluk verileri elde edilmiştir. Sonuçlar Tablo 11 ve Şekil 4'te verilmiştir. İşlem yükü ve en iyi sonucun elde edildiği C değeri 678'dir.

Tablo 11. C Değerleri için DVM'nin Sınıflandırma ve Doğrulama Sonuçları  
(Classification and Validation Results of SVM for C Values)

C Değerleri	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğrulama Oranı (%)
C=2 <sup>0</sup>	676	23	96,71
C=2 <sup>1</sup>	676	23	96,71
C=2 <sup>2</sup>	675	24	96,57
C=2 <sup>3</sup>	676	23	96,71
C=2 <sup>4</sup>	677	22	96,85
C=2 <sup>5</sup>	678	21	96,99
C=2 <sup>6</sup>	678	21	96,99
C=2 <sup>7</sup>	678	21	96,99



Şekil 4. DVM Doğrulama Grafiği (SVM Validation Chart)

Tablo 12. DVM Sınıflandırıcı için Karmaşıklık Matrisi  
(Complexity Matrix for SVM Classifier)

a	b	
445	13	a=2
8	233	b=4

\* a değeri 2 yani iyi huylu tümörü, b değeri 4 yani kötü huylu tümörü temsil etmektedir.

Tablo 13. DVM Sınıflandırıcı için Performans Değerlendirme Ölçütleri  
(Performance Evaluation Metrics for SVM Classifier)

DP	DN	YP	YN	Duyarlılık	Kesinlik	F1 Skorumla
445	233	8	13	0,9716/ 0,9668	0,9823/ 0,9471	0,9769/ 0,9568

Duyarlılık, Kesinlik ve F1 Skorumla değerleri ilki iyi huylu ve sonraki kötü huylu tümör değerleri için verilmiştir.

Tablo 12, DVM için karmaşıklık matrisini ve Tablo 13'te ise algoritmanın performans ölçütlerini göstermektedir. Performans ölçütleri DVM'nin de iyi huylu verileri sınıflandırmada daha başarılı olduğunu göstermektedir. Diğer taraftan bazı kötü huylu tümör verileri tespit edemediği söylenebilir. Modelin başarı oranı, duyarlılık, hassasiyet ve F1 skoru parametrelerinin birbirine yakın olduğu görülmektedir. Çalışılan algoritmalar arasında DVM algoritması Yanlış Negatif parametresi en az olan bir algoritma olmuştur.

### 3.5 k-Means Kümeleme Algoritması Uygulaması (Application of k-Means Clustering Algorithm)

k-means kümeleme algoritmasında k değeri (numClusters) manuel olarak seçilir ve k 2, 3, 4, ...n şeklinde değerler olabilir. Bu çalışmada kullanılan veri seti iki gruptan (iyi veya kötü huylu) oluştuğu için k =2 olarak seçilmiştir. Önemli bir diğer parametre ise uzaklık fonksiyonu olup bu çalışmada Öklid ve Manhattan uzaklık ölçütleri kullanılmıştır. Tablo 14'te k=2 için, Tablo 15'te k=3 için ve Tablo 16'da k=4 için model sonuçları verilmektedir. k=3 ve k=4 için sonuçların anlamsız olduğu görülmektedir. Bu durum iki grup/kümeden oluşan verilerin daha fazla kümeye ayrılması istendiğinde veriler arasındaki korelasyonun kaybolduğu görülmektedir. Küme sayısı bilinmeyen veri setlerinde denemeler veri setinin kaç kümeden oluştuğuna dair önemli birer yol göstericidir.

Tablo 14.  $k=2$  İçin Kümeleme Sonuçları  
(Clustering Results for  $k=2$ )

	Öklid Mesafe Fonksiyonu		Manhattan Mesafe Fonksiyonu	
	Örnek Sayısı – Yüzde		Örnek Sayısı - Yüzde	
<b>Küme 0</b>	246	35	237	34
<b>Küme 1</b>	453	65	462	66

Tablo 15.  $k=3$  İçin Kümeleme Sonuçları  
(Clustering Results for  $k=3$ )

	Öklid Mesafe Fonksiyonu		Manhattan Mesafe Fonksiyonu	
	Örnek Sayısı – Yüzde		Örnek Sayısı - Yüzde	
<b>Küme 0</b>	241	34	234	33
<b>Küme 1</b>	265	38	276	39
<b>Küme 2</b>	193	28	189	27

Tablo 16.  $k=4$  İçin Kümeleme Sonuçları  
(Clustering Results for  $k=2$ )

	Öklid Mesafe Fonksiyonu		Manhattan Mesafe Fonksiyonu	
	Örnek Sayısı – Yüzde		Örnek Sayısı - Yüzde	
<b>Küme 0</b>	239	34	136	19
<b>Küme 1</b>	17	2	114	16
<b>Küme 2</b>	188	27	188	27
<b>Küme 3</b>	255	36	261	37

Tüm bunların dışında sınıf sayısı adedince  $k$  değeri için yine farklı mesafe fonksiyonlarında "Sınıfların Kümelemelere Atanması (Classes to Clusters)" değerlendirmesi başarı yüzdeleri hesaplanabilmektedir. Doğrulama sonuçları Tablo 17'de verilirken karmaşıklık matrisi ise Tablo 18'de verilmiştir.

Tablo 17.  $k$ -Means Algoritması  $k=2$  için Doğrulama Sonuçları (Classes to Clusters Evaluation)  
(Validation Results for  $k$ -Means Algorithm  $k=2$  (Classes to Clusters Evaluation))

	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğrulama Oranı(%)
<b>Manhattan Fonksiyon</b>	658	41	94,13
<b>Öklid Fonksiyon</b>	669	30	95,71

Tablo 18.  $k$ -Means Algoritması  $k=2$  için Karmaşıklık Matrisi (Classes to Clusters Evaluation)  
(Complexity Matrix for  $k$ -Means Algorithm  $k=2$  (Classes to Clusters Evaluation))

Manhattan Fonksiyonu		Öklid Fonksiyonu		
a	b	a	b	
10	448	11	447	<b>a=2</b>
210	31	222	19	<b>b=4</b>

\* a veya 2 değeri iyi huylu tümörü, b değeri veya 4 kötü huylu tümörü temsil etmektedir.

Bu iki tabloda da görüldüğü gibi  $k$ -means kümeleme algoritması ile elde edilen en yüksek doğruluk oranı %95,71 olmaktadır. Bu oran  $k$ -means algoritmasının Wisconsin verileri kümeleme sonuçlarının doğruluğu olup, modelin, verileri ayırmada kabul edilebilir başarıya sahip olduğu ifade edilebilir.

3.6 Hiyerarşik Kümeleme Algoritması Uygulaması  
(Implementation of Hierarchical Clustering Algorithm)

Hiyerarşik kümeleme algoritması  $k$ -means algoritmasındaki zorunlu olarak önceden girilen  $k$  adet küme parametresini belirleme işlemi ortadan kaldırmak için geliştirilen bir algoritmadır.  $k$ -means'te kümeleme işlemi, önceden belirlenmiş küme sayısına göre belirlenen küme merkezleri oluşturularak veri noktalarını bu merkezlere olan yakınlıklarına göre yapar. Kümeler ortalama değerlerine göre sürekli güncellenir. Hiyerarşik kümelemede ise her bir veri noktasını bir küme kabul eder (Agglomerative yöntemi) ve en yakın iki veri noktasını birleştirerek yeni kümeler oluşturur. Daha sonra bu kümeler yakınlıklarına göre birleştirilir ve bu işlem kümeleme işlemi bitene kadar devam eder. Hiyerarşik kümelemedeki  $k$  parametresi dendrogram üzerinde hangi seviyede kesme yapılacağını belirler. Literatürde bu veri seti için hiyerarşik metodu önerilen algoritmalar arasında yer almamaktadır. Bu çalışmada modelin neden Wisconsin verilerini ayırmada başarılı olmadığı noktası ele alınmıştır. Kümeleme yöntemleri verilerin benzerlik durumlarını veriler arasındaki mesafe ile ilişkilendirilmektedir. İki veri arasındaki mesafe ölçme metodu önem kazanmaktadır. Hiyerarşik kümeleme yöntemi Wisconsin veri seti için önemli bir örnek teşkil etmektedir.  $k=2$  değeri ve farklı mesafe fonksiyonları için yapılan uygulamada elde edilen sonuçlar Tablo 19'da yer almaktadır.

Tablo 19. Hiyerarşik Kümeleme  $k=2$  İçin Kümeleme Sonuçları  
(Clustering Results for Hierarchical Clustering  $k=2$ )

	Öklid Mesafe Fonksiyonu		Manhattan Mesafe Fonksiyonu	
	Örnek Sayısı – Yüzde		Örnek Sayısı - Yüzde	
Küme 0	<b>458</b>	<b>- 66</b>	<b>698</b>	<b>- 100</b>
Küme 1	<b>241</b>	<b>- 34</b>	<b>1</b>	<b>- 0</b>

Bu tabloda özellikle Manhattan Fonksiyonu kullanıldığında kümelemedeki dengesiz dağılım açık şekilde görülmektedir. Özellikle kötü huylu tümör verilerini ayırmada oldukça başarısız olduğu görülmektedir. Model veriler arasındaki uzaklık bilgisine göre kümeleme yapmaktadır. Yüksek boyutlu veri setlerinde Hiyerarşik yöntemi, verilerin birbiri ile benzerliği ve uzaklığını tespit etmede zorlanabilmektedir. Yakın olan veriler benzer olmakta ve benzer olan örnekler aynı grupta değerlendirilmektedir. Dolayısıyla uzaklık

fonksiyonu Hiyerarşik kümeleme algoritması için oldukça önemli bir parametre olmaktadır. Manhattan uzaklık tekniği yatay veya dikey birim adım şeklinde iki veri arasındaki uzaklığı belirlemektedir. Wisconsin veri seti gibi çok boyutlu veri setlerinde bu ölçüm tekniği verileri ayırmada başarısız olmaktadır.

Hiyerarşik kümeleme algoritması farklı mesafe fonksiyonlarında "Sınıfların Kümelemelere Atanma (Classes to Clusters)" başarı yüzdeleri hesaplanabilmektedir. Doğrulama sonuçları Tablo 20'de verilirken karmaşıklık matrisi parametreleri ise Tablo 21'de verilmiştir.

Tablo 20. Hiyerarşik Kümeleme Algoritması Doğrulama Sonuçları  
(Hierarchical Clustering Algorithm Validation Results)

	Doğru Sınıflandırma	Yanlış Sınıflandırma	Doğrulama Oran(%)
Manhattan Fonksiyonu	459	240	65,66
Öklid Fonksiyonu	459	240	65,66

Tablo 21. Hiyerarşik Kümeleme Algoritması Karmaşıklık Matrisi  
(Hierarchical Clustering Algorithm Complexity Matrix)

Manhattan Fonksiyonu		Öklid Fonksiyonu		
a	b	a	b	
458	0	458	0	a=2
240	1	240	1	b=4

\* a veya 2 değeri iyi huylu tümörü, b değeri veya 4 kötü huylu tümörü temsil etmektedir.

Bu iki tabloda da görüldüğü gibi Hiyerarşik kümeleme algoritmasında en yüksek doğruluk oranı %65,66 olmaktadır. Ayrıca karmaşıklık matrisinde de görüleceği üzere a değeri yani iyi huylu tümörler başarılı şekilde kümelendirilirken, b yani kötü huylu tümörler kümelendirmede model neredeyse tamamen başarısız olmaktadır.

#### 4. TARTIŞMA (DISCUSSION)

Bu çalışmada, k-NN, NB ve DVM sınıflandırma algoritmaları ile k-means ve Hiyerarşik kümeleme algoritmaları, Weka 3.8.6 makine öğrenimi aracı kullanılarak Wisconsin meme kanseri veri setine uygulanmış ve elde edilen sonuçlar tablolarda sunulmuştur.

İlk olarak, k-NN sınıflandırıcısı veri setine uygulanarak optimal bir k-değeri araştırılmıştır. Yapılan uygulamalarda k=5 değeri için %96,99 doğruluk derecesine sahip bir başarı derecesi elde edilmiştir. Modelin performansı analiz edildiğinde k=5 için k-NN algoritması, 458 iyi huylu tümörün (benign) 444'ünü ve 241 kötü huylu tümörün (malign) 234'ünü doğru bir şekilde sınıflandırmıştır. Bu

değerler karmaşıklık matrisi ile analiz edildiğinde k-NN modelinin oldukça hassas ve duyarlı olduğu görülmekle birlikte etkili performansı ile yüksek başarıya sahip olduğu söylenebilir. Diğer taraftan çok kısa inşa sürelerine sahip hafif, tembel bir öğrenme algoritması olduğu göz önüne alındığında, modelin yüksek doğruluk derecesi ile diğer sınıflandırıcılara göre Wisconsin verilerini sınıflandırmada iyi olduğu söylenebilir. Ayrıca, karmaşıklık matrisi kullanılarak performans değerlendirme ölçütleri hesaplanmıştır. Buna göre hem iyi huylu hem de kötü huylu tümörün sınıflandırılmasında performans ölçütleri olumlu sonuçlar vermektedir. Duyarlılık parametresinde kötü huylu tümörün iyi huylu tümöre göre az bir farkla daha doğru tahmin edildiği; kesinlik parametresinde ise iyi huylu olarak sınıflandırılmış olan örneklerin oranının kötü huylu tümöre göre daha doğru tahmin edildiği görülmektedir. Ayrıca F1 skorlama parametresinde her iki tümör de yüksek oranda doğru tahmin edilmiştir. Ancak iyi huylu tümörlerin sınıflandırılmaları daha başarılı bir şekilde yapılmıştır.

Bu çalışmada kullanılan bir diğer sınıflandırıcı olan NB'nin doğruluk oranı en iyi durumda (useKernelEstimator aktif iken) %97,42'dir. 458 iyi huylu tümör vakasının 446'sını ve 241 kötü huylu tümör vakasının 239'unu doğru bir şekilde sınıflandırabilmiştir. Bu yöntem, Weka programında useKernelEstimator ve useSupervisedDiscretization parametreleri kullanılarak incelenmiş ve %97,42 doğruluk oranı ile en iyi sonuç olarak değerlendirilmektedir. Karmaşıklık matrisi kullanılarak performans değerlendirme ölçütleri hesaplandığında hem iyi huylu hem de kötü huylu tümörün sınıflandırılmasında performans ölçütleri olumlu sonuçlar vermektedir. Duyarlılık parametresinde kötü huylu tümörün iyi huylu tümöre göre daha doğru tahmin edildiği; kesinlik parametresinde ise iyi huylu olarak sınıflandırılmış olan örneklerin oranının kötü huylu tümöre göre çok daha doğru tahmin edildiği görülmektedir. F1 skorlama parametresinde ise her iki tümörün de iyi oranda tahmin edildiği görülmekte birlikte modelin iyi huylu tümörü sınıflandırmada daha başarılı olduğu tespit edilmiştir. NB sınıflandırıcı iyi huylu tümörde %98 oran, kötü huylu tümörde ise %96,37 oran ve en iyi F1 skorunu vermektedir.

J48 algoritmasında Güven Faktörü (confidenceFactor), MinNumObj ve doNotMakeSplitPointActualValue parametrelerinin doğruluk üzerindeki etkisi üzerinde durulmuştur. ConfidenceFactor=0,25, MinNumObj=4 ve doNotMakeSplitPointActualValue parametresi "true" durumda iken en optimum sonuç elde edilmiş ve doğrulama yüzdesi %95,56 olmuştur. Modelin karmaşıklık matrisi duyarlılık ve kesinlik parametreleri iyi huylu tümörün kötü huylu tümöre göre daha doğru tahmin edildiğini belirtmektedir. F1 skorlama parametresinde her iki tümörün de iyi oranda tahmin edildiği göstermektedir. Ancak model iyi huylu tümörü sınıflandırmada daha başarılı olduğu sonucuna varılmıştır.

DVM algoritmalarının birçok alt-sınıflandırıcısı ve değiştirilebilecek birçok parametresi bulunmaktadır. Bu çalışmada, maliyet (C) parametresinin doğruluk üzerindeki

etkisi üzerinde durulmuştur. Sunulan sonuçlar göz önüne alındığında, en iyi sonuç  $C=25$  değeri için %96,99 olarak elde edilmiştir ve bu sınıflandırıcı, k-NN sınıflandırıcısı ile aynı doğruluk oranına sahip olup kabul edilebilir bir doğruluk oranıdır. Karmaşıklık matrisi duyarlılık ve kesinlik parametreleri iyi huylu tümörlerin sınıflandırılmasında modelin daha başarılı olduğunu göstermektedir. F1 skora ölçütü modelin tümör tespitinde başarılı olduğu ve iyi huylu tümör tespitinde daha çok başarılı olduğunu göstermektedir.

Kümeleme algoritmalarından k-means ile belli mesafe fonksiyonları ve k sayısının (numClusters) kümeleme üzerindeki etkisi üzerinde durulmuştur. Aynı zamanda  $k=2$  değeri ve Öklid mesafe fonksiyonu değeri içinde doğruluk oranı belirlenmiştir. Bu uygulamada k-means için %95,71 doğruluk oranı belirlenmiştir.

Diğer bir kümeleme algoritması olan Hiyerarşik kümeleme Wisconsin veri setine uygulanmıştır. Bu uygulamada da belli mesafe fonksiyonları ve k sayısının (numClusters) kümeleme üzerindeki etkisi üzerinde durulmuştur. Burada da  $k=2$  değeri ve Öklid mesafe fonksiyonu değeri içinde doğruluk oranı belirlenmiştir. Bu durumda %65,66'lık bir doğruluk oranı oluşmuştur. Bu oran kümeleme algoritmasının bu veri setinde başarısız olduğu açık bir göstergesidir. Diğer taraftan Manhattan uzaklık tekniği seçildiğinde yöntemin bu veri setinde tamamen başarısız olduğu gözlemlenmiştir. Dolayısıyla Wisconsin verilerini gruplara ayırırken uzaklık fonksiyonunun önemli bir parametre olduğu görülmektedir. Verilerin benzerliği, iki verinin birbirine yakınlığı üzerine kurulu olduğu için uzaklık ölçütü modellerin performansında önemli rol oynamaktadır. Modellerin doğruluk oranları Tablo 22'de verilmiştir.

Tablo 22. Sınıflandırma ve Kümeleme Algoritmalarının Doğrulama Yüzdeleri

(Validation Percentages of Classification and Clustering Algorithms)

Algoritma Adı	Doğrulama Oranı(%)
k-NN	96,99
NB	97,42
J48	95,56
DVM	96,99
k-means	95,71
Hiyerarşik	65,66

Bu çalışmanın, literatürdeki benzer çalışmalarla doğruluk oranı karşılaştırma tablosu (Tablo 23) aşağıda verilmektedir.

Tablo 23. Algoritmaların Doğrulama Yüzdeleri Literatür Karşılaştırmaları

(Validation Percentages of Algorithms Literature Comparisons)

Çalışmalar	DVM	k-NN	J48	NB
Amrane vd. 2018 [7]	-	97,51	-	96,19
Aruna vd. 2011 [8]	96,84	-	94,59	96,50
Akbugday 2019 [33]	96,85	96,85	-	95,99
Ahmed vd. 2020 [32]	96,13	-	94,26	97,27
Uddin vd. 2024 [9]	90,15	89,63	91,21	-
Nemade vd. 2023 [10]	95	96	97	90
Kadhim vd. 2022 [14]	96,49	95,61	-	91,22
Laghmati vd. 2023 [12]	92,1	93,8	-	-
Amethiya vd. 2021 [13]	94,3	95,9	94,56	-
<b>Bu çalışma</b>	<b>96,99</b>	<b>96,99</b>	<b>95,56</b>	<b>97,42</b>

Tablo 23'de görüldüğü üzere bu çalışma, benzer çalışmalarla karşılaştırıldığında tutarlı olduğu görülmüş ve literatürden dahi iyi sonuçlar elde edilmiştir.

Modellerin hem iyi huylu hem de kötü huylu tümör için F1 skorlamaları ve ROC sonuçları Tablo 24'te verilmiştir.

Tablo 24. Sınıflandırma Algoritmalarının F1 Skorlama Yüzdeleri

(F1 Scoring Percentages of Classification Algorithms)

Algoritma Adı	F1 Skorları(%) (İyi Huylu/ Kötü Huylu)	ROC
k-NN	97,68 / 95,70	0,970
NB	98 / 96,37	0,978
J48	96,6 / 93,6	0,952
DVM	97,69 / 95,68	0,969

Meme kanseri veri setine uygulanan algoritma başarılarını karşılaştırmak için ROC analiz yapılmış modellerin başarıları sırayla NB (0,978), k-NN (0,970), DVM (0,969), J48 (0,952), k-means (0,948) ve Hiyerarşik (0,502) olarak elde edilmiştir. ROC kriteri modellerin başarılarını karşılaştırmak için kullanılan etkili bir kriterdir. Modellerin ROC sonuçları sınıflandırıcıların kümeleme yöntemlerine göre daha başarılı olduğunu göstermektedir. Sınıflandırıcılar arasında da en başarılı yöntem NB algoritması olmuştur. ROC sonuçları modellerin başarı oranı, duyarlılık, kesinlik ve F1 skora kriterleri ile uyumludur.

## 5. SONUÇ VE ÖNERİLER (RESULT AND SUGGESTIONS)

İnsan yaşam kalitesini artırmaya yönelik yapılan çalışmalar giderek yoğunlaşmaktadır. Veri bilimindeki gelişmeler, verilerin farklı modellerle analiz edilerek çeşitli yönleriyle değerlendirilmesi ve daha başarılı sonuçlara ulaşılmasını sağlamaktadır. Bu çalışmada, kadınlar arasında oldukça yaygın ve ciddi olan meme kanseri için otomatik tanı sistemi araştırılmıştır. Çeşitli sınıflandırma ve kümeleme yöntemleri, UCI veri tabanından alınan meme kanseri veri setine uygulanarak modellerin performansları farklı metriklerle değerlendirilmiştir. Veri seti analiz edilerek, hastalık tespitinde etkin özellikler belirlenmiştir. Sürekli artan verilerden bilgi çıkarma veya veri madenciliği süreci detaylı olarak ele alınmıştır. Algoritmaların sonuçları, geleneksel başarı oranı metrikleri üzerinden değil, daha detaylı ve yeni bir görüş olan her bir verinin kendi sınıfına aitliği tespiti de irdelenmiştir. Oldukça karmaşık bir yapıya sahip olan derin öğrenme ile de hastalık tespiti yapılabilir; ancak bu çalışmada model performanslarının daha detaylı analizi mümkün olduğu düşünülmektedir. Benzer şekilde, literatürde bu veri seti için modellerin başarı oranı üzerinde durulurken, bu çalışmada başarılı bir tanı için önemli metrikler analiz edilmiştir.

Çalışmada k-NN, NB, J48 ve DVM sınıflandırma algoritmaları ile k-means ve hiyerarşik kümeleme algoritmaları kullanılmıştır. Sonuçlar, %97,42 doğruluk oranıyla NB algoritmasının en doğru sınıflandırma algoritması olduğunu göstermektedir. Ayrıca, iyi huylu tümör için %98 ve kötü huylu tümör için %96,37 oranıyla en iyi F1 skorlamasına sahip sınıflandırıcı yine NB algoritması olmuştur. Modelin ROC değeri de sonuçlarla uyumludur.

Veri setinde yer alan örneklerin özellikleri, başarılı tanı için önemli bir kriterdir. İki örnek benzerliği değerlendirilirken kullanılan mesafe ölçütleri (Öklid, Manhattan, Minkowski mesafe fonksiyonları) kritik öneme sahiptir. İki veri arasındaki mesafe, verilerin benzerliklerini göstermekte ve başarılı teşhis ile doğrudan orantılıdır.

Günümüzde ön tanı veya karar destek sistemlerinin günlük hayatta kullanılabilme imkanları mevcuttur. Yeni mobil cihazların kapasitesi oldukça gelişmiş olup, birçok marka kendi cihazlarına yapay zekâ uygulama programlama arayüzlerini (Application Programming Interface-API) entegre etmeye başlamışlardır. Bu minvalde, hastalıkların tespiti için kullanılan test cihazlarına, o hastalık tanısında başarılı olan algoritmaların entegre edilmesiyle ön tanı işlemleri mümkün hale gelebilir. Örneğin, başka bir şikâyet üzerine yapılan test sonuçlarının otomatik algoritmalar tarafından değerlendirilerek varsa başka hastalıkların da ortaya çıkarılması söz konusudur. Böyle bir sistem ile hastalığın önceden tespiti mümkün olabilir. Bu tür sistemler, hekimlerin aşırı iş yükü gibi çeşitli faktörlerden dolayı hastalığın farkına varamaması ihtimalini düşürür ve insan kaynaklı hataları minimize edebilir.

Gelecekteki çalışmalarda, farklı veri setleri veya farklı parametreler içeren veriler ile yapay zekâ konusunda daha gelişmiş programlama dilleri kullanılarak algoritmaların farklı platformlardaki davranışı araştırılabilir. Gelişmiş programlama uygulamaları ve/veya platforma özgü avantajlarla daha doğru sınıflandırıcıların kullanımı, daha kesin sonuçların elde edilmesini sağlayabilir.

## KAYNAKLAR (REFERENCES)

- [1] Siegel, R.L., Miller, K.D. and Jemal, A., 2020. Cancer statistics, 2020. *CA: a cancer journal for clinicians*, 70(1), 7–30. <https://doi.org/10.3322/caac.21590>.
- [2] Bora B., Soyutun Ç.İ., Aygün A., Özdemir T.A., Kulali B., Uzun S.B. ve ark., (2019). Sağlık İstatistikleri Yıllığı. Sağlıkta İstatistik ve Nedensel Analizler (SİNA) Platformu. Ankara, Türkiye.
- [3] Anderson, Benjamin O., et al., 2010. Optimisation of breast cancer management in low-resource and middle-resource countries: executive summary of the Breast Health Global Initiative consensus. *The lancet oncology*, 12.4 (2011): 387-398.
- [4] İnternet: Bakanlığ, T. S. (2022). Sağlık istatistikleri yıllık. Türkiye İstatistik Kurumu (TÜİK). Ankara.
- [5] Türkyılmaz, M., Öztürk, M., Dündar, S., Ergün, K.A., Sevinç, A., Tütüncü, S., Seymen, E., (2021). Türkiye Kanser İstatistikleri. T.C. Sağlık Bakanlığı Halk Sağlığı Genel Müdürlüğü. Ankara, Türkiye.
- [6] İnternet: UC Irvine, Breast Cancer Wisconsin (Original), UC Irvine Machine Learning Repository. <https://archive.ics.uci.edu/dataset/15/breast+cancer+wisconsin+original>, (14.07.1992)
- [7] Amrane, M., Oukid, S., Gagaoua, I., and Ensari, T., 2018. Breast cancer classification using machine learning. In 2018 electric electronics, computer science, biomedical engineering's meeting (EBBT) (pp. 1-4). IEEE.
- [8] Aruna, S., S. P. Rajagopalan, and L. V. Nandkishore, 2011. Knowledge based analysis of various statistical tools in detecting breast cancer. *Computer Science & Information Technology*, 2.2011 (2011): 37-45.
- [9] Uddin, K. M. M., Biswas, N., Rikta, S. T., & Dey, S. K. (2023). Machine learning-based diagnosis of breast cancer utilizing feature optimization technique. *Computer Methods and Programs in Biomedicine Update*, 3, 100098.
- [10] Nemade, V., & Fegade, V. (2023). Machine learning techniques for breast cancer prediction. *Procedia Computer Science*, 218, 1314-1320.
- [11] Singh, L. K., Khanna, M., & Singh, R. (2024). An enhanced soft-computing based strategy for efficient feature selection for timely breast cancer prediction: Wisconsin Diagnostic Breast Cancer dataset case. *Multimedia Tools and Applications*, 1-66.
- [12] Laghmati, Sara & Hamida, Soufiane & Hicham, Khadija & Cherradi, Bouchaib & Tmiri, Amal. (2023). An improved breast cancer disease prediction system using ML and PCA. *Multimedia Tools and Applications*. 83. 1-37. 10.1007/s11042-023-16874-w.

- [13] Amethiya, Yash & Pipariya, Prince & Patel, Shlok & Shah, Manan. (2021). Comparative Analysis of Breast Cancer detection using Machine Learning and Biosensors. *Intelligent Medicine*. 2. 10.1016/j.imed.2021.08.004.
- [14] Kadhim, Rania & Kamil, Mohammed. (2022). Comparison of breast cancer classification models on Wisconsin dataset. *International Journal of Reconfigurable and Embedded Systems (IJRES)*. 11. 166-174. 10.11591/ijres.v11.i2.pp166-174.
- [15] İnternet: Öğüdücü, Ş.G., Veri Madenciliği Temel Sınıflandırma Yöntemleri, <https://web.itu.edu.tr/~sgunduz/courses/verimaden/slides/d3.pdf>.
- [16] Uğuz, S. (2019). Makine öğrenmesi teorik yönleri ve Python uygulamaları ile bir yapay zekâ ekolü. Nobel Yayıncılık. Ankara.
- [17] İnternet: Akçay, A. K En Yakın Komşu Algoritması. <https://aycaakcay.medium.com/k-en-yakin-komsu-k-nearest-neighbor-algoritmasi-siniflama-7c456f8e2b0d>, (25.06.2020).
- [18] İnternet: Şeker, Ş.E., KNN(K Nearest Neighborhood, En Yakın k Komşu). <https://bilgisayarkavramlari.com/2008/11/17/knn-k-nearest-neighborhood-en-yakin-k-komsu/>, (17.11.2008).
- [19] İnternet: Hatipoğlu, E. Machine Learning – Classification – Naive Bayes – Part 11. <https://medium.com/@ekrem.hatipoglu/machine-learning-classification-naive-bayes-part-11-4a10cd3452b4>, (13.06.2018).
- [20] Solmaz, R., Günay, M., and Alkan, A., (2014). Fonksiyonel Tiroit Hastalığı Tanısında Naive Bayes Sınıflandırıcının Kullanılması. Akademik Bilişim Konferansı. Mersin, Türkiye, 891-897.
- [21] Hemanth, D. J., and Kose, U., 2020. Artificial Intelligence and Applied Mathematics in Engineering Problems: Proceedings of the International Conference on Artificial Intelligence and Applied Mathematics in Engineering (ICAIAME 2019). Vol. 43, Springer Nature.
- [22] İnternet: Medium Yöntemler – 4.1: C4.5 Algoritması, <https://medium.com/@Emreyz/yontemler-4-1-c4-5-algoritmasi-7382de92584e>, (03.03.2017).
- [23] İnternet: Şeker, Ş.E., C4.5 Ağacı, <https://bilgisayarkavramlari.com/2012/11/13/c4-5-agaci-c4-5-tree/>, (13.11.2012).
- [24] Aras, Ü., 2008. **Finansal veri madenciliği**. Yüksek Lisans Tezi, Yıldız Teknik Üniversitesi Fen Bilimleri Enstitüsü, İstanbul, 85.
- [25] İnternet: Şeker, Ş.E., SVM (Support Vector Machine, Destekçi Vektör Makinesi), <https://bilgisayarkavramlari.com/2008/12/01/svm-support-vector-machine-destekci-vektor-makinesi/>, (01.12.2008).
- [26] İnternet: Şeker, Ş.E., Weka ile SVM, <https://bilgisayarkavramlari.com/2011/09/19/weka-ile-svm/>, (19.09.2011).
- [27] Solmaz, R., Günay, M., and Alkan, A., 2013. Uzman sistemlerin tiroit teşhisinde kullanılması. XV. Akademik Bilişim Konferansı Bildirileri, 23-25.
- [28] İnternet: Çalışkan, T.K., Destek Vektör Makineleri (DVM), <https://www.bilimma.com/destek-vektor-makineleri-support-vectors-machines-svms>, (01.04.2023).
- [29] İnternet: Şeker, Ş.E., K-Ortalama Algoritması (K-Means Algorithm), <https://bilgisayarkavramlari.com/2008/12/15/k-ortalama-algoritmasi-k-means-algorithm/>, (15.12.2008).
- [30] Takaoğlu, M., and Takaoğlu, F., 2019. K-Means ve Hiyerarşik Kümeleme Algoritmanın Weka ve Matlab Platformlarında Karşılaştırılması. İstanbul Aydın Üniversitesi Dergisi, 11(3), 303-317.
- [31] İnternet: Seker, S. E., 2015. Sosyal ağlarda veri madenciliği (data mining on social networks). Ybs Ansiklopedi, 2.2 (2015): 30-39.
- [32] Ahmed, M. T., Intiaz, M. N., and Karmakar, A., 2020. Analysis of wisconsin breast cancer original dataset using data mining and machine learning algorithms for breast cancer prediction. *Journal of Science Technology and Environment Informatics*, 9(2), 665-672
- [33] Akbugday, B., 2019. Classification of breast cancer data using machine learning algorithms. In 2019 Medical technologies congress (TIPTEKNO) (pp. 1-4). IEEE

# Advanced Android Malware Detection: Merging Deep Learning and XGBoost Techniques

*Araştırma Makalesi/Research Article*

 Esra KAVALCI YILMAZ<sup>1\*</sup>,  Rezan BAKIR<sup>1</sup>

<sup>1</sup>Department of Computer Engineering, Sivas University of Science and Technology, Sivas, Türkiye.

[esra.kavalci@sivas.edu.tr](mailto:esra.kavalci@sivas.edu.tr), [rezan.bakir@sivas.edu.tr](mailto:rezan.bakir@sivas.edu.tr)

(Geliş/Received:20.09.2024; Kabul/Accepted:15.11.2024)

DOI: 10.17671/gazibtd.1553548

**Abstract**— The increasing importance of Android devices in our lives brings with it the need to secure personal information stored on these devices, such as contact details, documents, location data, and browser data. These devices are often targeted by attacks and malware designed to steal this data. In response, this work takes a novel approach to Android malware detection by integrating deep learning with traditional machine learning algorithms. An extensive experimental study was conducted using the DroidCollector network traffic analysis dataset. Eight different deep learning methods are analysed for malware classification. In the first phase, experiments were conducted on both original and stabilised datasets and the most effective methods were identified. In the second phase, the best performing deep learning methods were combined with XGBoost for classification. This hybrid approach increased classification success by 3-4%. The highest F1 and accuracy values obtained after 150 epochs of training with BiLSTM+XGBoost were 95.12% and 99.33% respectively. These results highlight the superiority of combining deep learning and traditional machine learning techniques over individual models and significantly improve classification accuracy. This integrated method provides a very important strategy for developing high-performance models for various applications.

**Keywords**— malware detection, machine learning, deep learning, XGBoost

## Gelişmiş Android Kötü Amaçlı Yazılım Tespiti: Derin Öğrenme ve XGBoost Tekniklerinin Birleştirilmesi

**Özet**— Android cihazların hayatımızdaki artan önemi, bu cihazlarda depolanan kişisel bilgileri (iletişim bilgileri, belgeler, konum verileri ve tarayıcı verileri gibi) güvence altına alma ihtiyacını beraberinde getirir. Bu cihazlar genellikle bu verileri çalmak için tasarlanmış saldırılar ve kötü amaçlı yazılımların hedefi olur. Bu duruma önlem olarak, bu çalışma derin öğrenmeyi geleneksel makine öğrenimi algoritmalarıyla entegre ederek Android kötü amaçlı yazılım tespitine yeni bir yaklaşım sunmaktadır. DroidCollector ağ trafiği analizi veri kümesi kullanılarak kapsamlı bir deneysel çalışma yürütülmüştür. Kötü amaçlı yazılım sınıflandırması için sekiz farklı derin öğrenme yöntemi analiz edilmiştir. İlk aşamada, hem orijinal hem de önişlemeden geçirilmiş (SMOTE, SMOTETomek, ClusterCentroids) veri kümeleri üzerinde deneyler yürütülmüş ve en etkili yöntemler belirlenmiştir. İkinci aşamada, en iyi performans gösteren derin öğrenme yöntemleri sınıflandırma için XGBoost ile birleştirilmiştir. Bu hibrit yaklaşım, sınıflandırma başarısını %3-4 oranında artırmıştır. BiLSTM + XGBoost modelinin 150 epoch ile eğitilmesiyle elde edilen en yüksek F1-score ve doğruluk değerleri sırasıyla %95,12 ve %99,33 olmuştur. Bu sonuçlar, derin öğrenme ve geleneksel makine öğrenimi tekniklerinin bireysel modellere göre birleştirilmesinin üstünlüğünü vurgular ve sınıflandırma doğruluğunu önemli ölçüde iyileştirir. Bu hibrit yöntem, çeşitli uygulamalar için yüksek performanslı modeller geliştirmek amacıyla önemli bir strateji sunmaktadır.

**Anahtar Kelimeler**— kötü amaçlı yazılım tespiti, makine öğrenmesi, derin öğrenme, XGBoost

## 1. INTRODUCTION

In today's digital landscape, mobile devices, especially smartphones and tablets based on the Android operating system, have become seamlessly integrated into our daily lives. They serve as versatile tools for communication, entertainment, business, and managing daily activities. However, this ubiquitous reliance on mobile devices also presents a significant challenge: the ever-present risk of security vulnerabilities and malicious software threats [1]. Indeed, the central role of Android devices in our daily lives is underlined by the vast trove of data they hold. From cherished personal photos to vital contact details, from sensitive financial records to indispensable calendar entries, these devices store a wealth of intimate and confidential information. Through sophisticated application and service integration, Android devices have become central tools for organizing and customizing users' lives. However, this wealth of data also presents significant security risks. The sensitive information stored on Android devices is a prime target for cybercriminals and malware. Unauthorized access to this data can have dire consequences, including financial loss, identity theft, and other serious disruptions. As a result, ensuring the security of Android devices is not only important but essential for user protection and peace of mind [2].

The Android ecosystem offers a wide range of applications and customization possibilities, but it also presents significant security vulnerabilities that malicious software can exploit to infiltrate devices. These vulnerabilities can pose serious threats to individual users, as well as companies and organizations [3]. To detect such threats, various analysis methods are employed in malware detection, which can be broadly classified as static, dynamic, and hybrid analysis approaches. Static analysis involves examining the file structure or code patterns of malware without executing it, making it fast and efficient but often ineffective against malware with advanced encryption and compression techniques. On the other hand, dynamic analysis examines the behavior of malware by running it in a controlled environment, which is more effective for detecting advanced threats but requires more resources. Hybrid analysis aims to overcome the limitations of both static and dynamic methods by combining the strengths of each, providing a more comprehensive and powerful solution for detecting complex malware. This hybrid approach ensures a more accurate assessment of potential threats, which is crucial in securing the Android ecosystem from malicious attacks.

The use of advanced technologies such as deep learning offers a promising solution to detect malware on Android devices. This paper scrutinizes novel methodologies aimed at enhancing the security of Android devices, with a particular emphasis on research integrating deep learning and machine learning techniques. To the best of our knowledge, the proposed approach represents an unprecedented endeavor within the literature, promising innovative strides toward fortifying Android device security.

### 1.1. Motivation

The open source architecture of Android devices makes it easier for attackers to analyze and target these devices. Moreover, the diversity of devices in the Android ecosystem and the irregularity in update processes make it difficult to patch vulnerabilities quickly. For these reasons, effective and rapid detection of malware on Android devices has become a critical requirement for user security. In the face of increasingly sophisticated malware attacks, traditional security methods are insufficient. In the literature, machine learning and deep learning methods have been successfully applied for Android malware detection, but the hybrid combination of these two techniques and the integrated utilization of their advantages is very limited. This study aims to investigate how the hybrid use of machine learning and deep learning methods can improve classification success by examining the effects of unbalanced data distribution on malware detection on Android devices.

### 1.2. Novelty and Contributions

In this work, the impact of balanced data distribution and the hybrid use of deep learning and machine learning (XGBoost) algorithms is studied to enhance accurate classification performance within the critical security domain of malware detection on Android systems. Android devices, due to their large user base and open-source nature, have become prime targets for malware attacks, making the accurate detection of these threats essential. Although various machine learning and deep learning techniques have been proposed in the literature for Android malware classification, the combined strengths of both methods remain underexplored. Deep learning models demonstrate superior performance in extracting discriminative features from high-dimensional and complex data structures, while the XGBoost algorithm achieves high accuracy in tree-based ensemble methods (boosting) due to its advanced optimization, regularization, and parallel processing capabilities. Through this hybrid approach, the study seeks to combine the adaptability of deep learning with the robust classification ability of XGBoost to reveal its impact on classification performance. The contributions of the study are listed below:

1-Elimination of Imbalanced Data Distribution: In malware classification processes, the number of data mislabeled as harmless (False Positive) is critical, as this can allow malware to infiltrate the system. In order to avoid such security risks, the data set should have a balanced distribution. In this study, to investigate the impact of balanced data distribution on classification performance, "imbalanced data sampling" methods are applied, enabling a comparative analysis of performance differences between balanced and imbalanced data distributions.

2-Hybrid Use of Deep Learning and Machine Learning Methods: The hybrid use of deep learning and machine



learning (XGBoost) methods is crucial for complex problems such as malware classification. While deep learning models exhibit strong performance in learning complex patterns and extracting discriminative features in high-dimensional data, the XGBoost algorithm provides more accurate classification results thanks to its regularization, error rate optimization and parallel processing capabilities. In this study, the potential for improving classification performance is analyzed by combining the strengths of both methods, demonstrating the hybrid approach's ability to deliver both flexibility and accuracy.

## 2. RELATED WORKS

Research in the field of Android malware detection has witnessed notable progress, as scholars have delved into a range of machine learning and deep learning methodologies to enhance detection accuracy. In this section, notable studies are reviewed, contributing to the comprehension and advancement of efficient malware detection systems.

### 2.1. Deep Learning Studies

Given the rise in malware targeting Android systems, recent research has increasingly focused on deep learning approaches to enhance detection accuracy. This section reviews notable studies that leverage deep learning architectures for malware detection, highlighting their methodologies and performance outcomes in comparison to traditional techniques.

Elayan and Mustafa conducted a study introducing a Gated Recurrent Unit (GRU)-based Recurrent Neural Network (RNN) as an innovative approach for detecting malware on the Android operating system. Trained using static features extracted from Android applications, such as API calls and permissions, their model demonstrated significantly higher performance than traditional methods, achieving an accuracy rate of 98.2% on the CICAndMal2017 dataset. [4]. In the work of Bakour and Ünver [5], a hybrid model called DeepVisDroid was proposed for Android malware detection, combining deep learning techniques with image-based features. This model transforms the source code of Android applications into four different grayscale image datasets, from which local and global features are extracted and analyzed. The proposed DeepVisDroid model achieved high success, reaching an accuracy of 98.96%. Yadav et al. presented an approach that utilizes images derived from bytecode files for malware detection and proposed an EfficientNet-B4 CNN based model. The EfficientNet-B4 architecture was chosen as the feature extractor for this process and worked with 226x226 images. In the study, 5986 samples were collected, converted to color images, and mapped to binary files. These images derived from Android bytecode representations were evaluated with their proposed model. They showed that their method was effective by achieving 95.7% accuracy [6]. Yumlembam et al. investigated the

effectiveness of graph neural networks in detecting attacks. In the study, unique global descriptors were created using local and global graphs obtained from API features. The importance weight of each feature was calculated using linear regression, and graph embedding and model training were performed using graph neural networks. The experiments were conducted on two datasets, one with 15,848 and the other with 56,461 samples. In this study, the proposed model was evaluated in terms of accuracy, precision, recall and F-score, and successful results up to 99.18% were obtained. In addition, a hostile malware generation model called VGAE-MalGAN was developed and 98.43% accuracy was achieved with this model [7]. Furthermore, In their study, Bakır and Bakır [3] emphasized the importance of feature extraction methods for malware detection in Android systems. Therefore, they proposed a new feature extraction method, autoencoder-based DroidEncoder. In the study, an image-based dataset was created from Drebin and Malgenome datasets and studies were carried out using this dataset. The authors proposed three different autoencoders based on ANN, CNN and VGG19. At the end of the study, it was observed that the proposed method gave successful results in terms of different metrics. Mohammed et al. investigated deep learning techniques for Android application categorization. In this context, they proposed a deep belief neural network (DBN)-based application categorization method. Using the CIC-AAGM2017 dataset of 1900 instances, the proposed model was compared with four traditional feed-forward neural networks and seven machine learning models. The results show that the DBN-based model is effective in classifying Android apps as benign or malicious with 98.7% accuracy [8]. Moreover, Tang et al. in their [9] research study propose a new classification method for detecting Android malware by addressing the weaknesses of traditional static analysis methods. The proposed method utilized a deep neural network that combines hashed bytecode image and attention mechanism. The method (ResNet-CBAM) processes the bytecode sequence of executable files into grayscale and Markov images and fuses these features to generate a feature space that can characterize Android malware. Experiments showed that the proposed ResNet-CBAM method can effectively represent bytecode sequence files, extract, and classify features. Based on the mixed image features, the malware detection accuracy reaches 98.67% and outperforms other similar methods [9]. Fu et al. stated that traditional methods cannot detect malware accurately and effectively due to their limitations. Therefore, they proposed a hybrid approach that combines multi-scale convolutional neural network (MSCNN) and ResNet networks. The approach was able to detect Android malware with high accuracy and precision by creating an advanced feature extraction network with MSCNN and a detection network with ResNet. At the end of the study, the authors confirmed that the results of the experiments show that the use of MSCNN as a multilevel feature extraction network significantly improves the performance of the hybrid models [10]. Liu et al. pointed out that semantic behavior feature extraction is critical for training a robust malware detection model. Therefore, they proposed SeGDroid, a novel Android malware detection method that focuses on learning

semantic information from sensitive function call graphs (FCGs). SeGDroid preserves sensitive API call context and removes irrelevant FCG nodes using graph pruning methods. Attributes of graph nodes are extracted by proposing a node representation method based on word2vec and social network-based centrality. This representation aimed to extract semantic information and graph structure of function calls. Experimental results showed that SeGDroid achieved 98.37% accuracy in the case of malware detection on the CICMal2020 dataset [11].

## 2.2. Machine Learning Studies

In recent years, machine learning (ML) techniques have played a crucial role in Android malware detection, offering effective solutions through algorithms capable of identifying patterns and anomalies in application behavior. This section explores key studies that apply machine learning models to classify malware, highlighting approaches such as decision trees, support vector machines, and ensemble methods. These studies underscore the adaptability and efficiency of machine learning in tackling the challenges posed by Android malware, setting a foundation for further advancements and hybrid approaches in the field.

For instance, Raman et al., in their work, proposed an ML-based method for detecting Android malware. The proposed method is optimized to detect Android malware with a KNN classification system using data stream-based API calls. Based on 1,050 malicious materials and 1,160 benign samples, the study [12] showed that the dataflow-based API-level features are successful (97.66%) in effectively detecting Android malware [12]. Similarly, Alani and Awad presented an ML-based system called AdStop for detecting Android adware by analyzing features in network traffic flow. While developing AdStop, they targeted design features such as high accuracy, speed, and generalizability. To improve the accuracy of adware detection and reduce the time burden, a feature reduction phase was applied, thus reducing the number of features used from 79 to 13. In the experiments, AdStop was found to be successful with 98.02% accuracy, 2% false positive rate, and 1.9% false negative rate [13]. In their study, Duran and Bakir [14] used machine learning algorithms for static analysis-based malicious application detection for the Android operating system. The imbalance of the class distribution in the dataset was eliminated by generating artificial data with the SMOTE algorithm. They also performed hyperparameter optimization to increase the accuracy of machine learning algorithms. This optimization determined the most appropriate hyperparameters with the Grid Search method. With the increasing threat of Android malware, it has become

important to develop effective detection techniques. In the [15] study, the performance of various machine learning algorithms was evaluated. The study reveals that the LightGBM algorithm has the highest accuracy (91%), precision (89%), and F1 score (89%) for Android malware detection. Evaluations on a 5-class dataset containing both benign and malicious applications suggest that these findings can contribute to the development of effective Android malware detection systems. In another study [16], AlOmari et al. addressed the challenges faced by cybersecurity researchers focusing on developing new detection systems with the rapid increase in Android mobile malware threats. They examined the performance of various machine learning algorithms and then focused on achieving maximum accuracy by normalizing numerical features with the Synthetic Minority Oversampling Technique (SMOTE). 11,598 APKs were used on a large dataset and the highest accuracy value was 95.49% with the light gradient boosting model. Furthermore, In order to secure Android mobile applications used in industrial platforms and smart cities, the authors of [17] present a machine learning-based approach called as the Hybrid Multimodal Machine Learning-Driven Android Malware Recognition and Classification (HM3-AMRC) model. HM3-AMRC accurately identifies and classifies Android malware using a new technique for feature selection and analysis that is according to authors more efficient than previous methods. A comprehensive benchmark analysis highlights that the HM3-AMRC method outperforms existing techniques with an accuracy of 99.01 [17]. Furthermore, Jundi and Aliasiri in their study developed a hybrid system for malware detection on Android smartphones. They used Extreme Gradient Boosting (XGBoost) and Grammatical Evaluation (GE) to determine the optimal parameters for this detection model. The experimental results of the study showed that the proposed model outperforms conventional parameter tuning. As a result of the study, the proposed model achieved 98% accuracy for CICMalDroid-2020, 99.02% accuracy for Drebin, and 99.28% accuracy for Malgenome [18]. On the other hand, Seyfari and Meimandi conducted a study in order to take precautions against malicious software that has increased with the widespread use of smartphones with Android operating system. In their study, they developed a method using simulated annealing algorithm and fuzzy logic to detect Android malware with machine learning algorithms. The study concluded that the proposed method achieved optimal results with a 99.02% accuracy rate using the KNN classifier in combination with a permission-based feature set. [19].

Table 1 includes some of the recently published studies in the domain of Malware detection.

Table 1. Some of the related studies

Paper	#Data	#Class	Method	Accuracy (%)
Elayan & Mustafa, 2021	712	2	GRU	98.20
Bakour & Unver, 2021	9700	2	DeepVisDroid	98.96
Raman et al., 2022	2210	2	KNN	97.66
Alani & Awad, 2022	86228	2	AdStop with RF	98.14
Yadav et al., 2022	5986	2	EfficientNet-B4 CNN	95.70
Yumlembam et al., 2023	15848 56461	2 2	VGAE- MalGAN	98.33 98.68
Baghirov, 2023	11598	5	LightGBM	91
AlOmari vd., 2023	11598	2	LightGBM	95.49
A vd., 2023	2000	2	HM3-AMRC	99.01
	3799	2		99.2
Jundi & Alyasiri, 2023	15036 11598	2 5	GE-XGBoost	99.0 97.9
Bakır & Bakır, 2023	6000	2	DroidEncoder	98.56
Mohammed et al., 2023	1900	2	DBN-Based Model	98.70
Tang vd., 2024	22901	2	ResNet-CBAM	98.67
Seyfari & Meimandi, 2024	15036	2	Proposed Method with KNN	99.02
Fu vd., 2024	11598	5	MSCNN+ResN et18	99.20
Liu vd., 2024	11598	5	SeGDroid	98.37

Examining Table 1 reveals that previous studies on malware detection primarily employ either deep learning methods or traditional machine learning algorithms. Our proposed model, however, integrates deep learning models with the XGBoost algorithm in a hybrid approach, leveraging the strengths of both. Deep learning models are adept at extracting distinctive features from high-dimensional and complex data, capturing intricate patterns that are essential for effective malware detection. These extracted features are then fed into XGBoost, a robust classifier known for its high accuracy and generalization capabilities. By combining the feature extraction power of deep learning with the strong classification performance of XGBoost, this hybrid approach achieves a more accurate and resilient malware detection system.

### 3. MATERIALS AND METHODS

This section provides a comprehensive overview of the dataset used in the study, describing its key characteristics and relevance for malware detection. To address the imbalance in the dataset, three data distribution techniques—SMOTE, SMOTETomek, and ClusterCentroids—are presented, each explained in detail to demonstrate their roles in rebalancing the data. Finally, this section describes the eight different deep learning models and the XGBoost algorithm used in the study, highlighting their specific functionalities and how they contribute to the hybrid approach for improved malware detection accuracy.

### 3.1. Used Dataset

This study analyzes the DroidCollector network traffic analysis dataset [17], [18]. Comprising 7844 data samples and spanning 17 attributes, this dataset serves as a foundational resource for our investigation. It is specifically designed for detecting malicious activity in Android applications based on network traffic analysis. The dataset was obtained through dynamic analysis, allowing for the capture of real-time network traffic behavior during the execution of Android applications in a controlled environment. This dynamic approach helps in identifying subtle differences in the network patterns of malicious and benign applications. The details of the dataset, including its attributes and sample sizes, are presented in Table 2.

### 3.2. Preparing Dataset

Preparing a dataset is crucial for the success of machine learning tasks like Android malware detection. It directly impacts the quality, generalization, and fairness of the model. A well-prepared dataset reduces bias, enhances interpretability, and ensures compliance with ethical considerations. It also facilitates reproducibility, saves computational resources, and increases the real-world applicability of the model. Overall, proper dataset preparation is essential for building reliable, accurate, and ethical machine learning models that contribute to the security of mobile devices and the digital ecosystem.

A meticulous preliminary analysis brought to light the presence of missing values (NaN) within certain attributes. Recognizing the potential impact of these missing values on the accuracy of our analysis, a strategic approach was formulated. Specifically, the attributes 'duracion,' 'avg\_local\_pkt\_rate,' and 'avg\_remote\_pkt\_rate' were identified as containing NaN values and subsequently removed from the dataset. This meticulous curation of the dataset serves a dual purpose: it not only ensures precision and coherence in our analytical processes but also elevates the dataset's reliability by adeptly addressing the challenge posed by missing values. Consequently, this methodical handling contributes to the robustness of our findings and enhances the overall quality of the dataset employed in our study.

The final attributes and descriptions of the dataset are presented in Table 2.

Following these procedures, it was observed that the 'name' and 'type' attributes of the dataset consisted of object expressions. To facilitate further analysis, the Label Encoder method was implemented, converting these attributes into numeric values. Subsequently, adjustments were made to the 'type' column, rendering it suitable for classification purposes. The resultant 'Benign' and 'Malicious' class distributions of the dataset are detailed in Table 3.

Table 2. Features and description of dataset

Feature	Description
name	'AntiVirus' 'Browser' 'chess' 'Communication' 'DailyLife' 'Education' 'Finance' 'HealthAndFitness' 'Input' 'MediaAndVideo' 'NewsAndMagazines' 'Personalization' 'Photography' 'Productivity' 'Reading' 'Shopping' 'Social' 'Sport' 'Tools' 'TravelAndLocal' 'Ackposts' 'Acnetdoor' 'Adrd' 'Adsms' 'Aks' 'Antares' 'Anudow' 'BaseBridge' 'Boxer' 'DroidDream' 'DroidKungFu' 'DroidRooter' 'DroidSheep' 'EICAR-Test-File' 'EWalls' 'ExploitLinuxLotoor' 'FaceNiff' 'FakeDoc' 'FakeFlash' 'FakeInstaller' 'Fakelogo' 'Fakengry' 'FakeRun' 'FakeTimer' 'FinSpy' 'Fjcon' 'FoCobers' 'Fujacks' 'Gamex' 'Gapev' 'Gappusin' 'GGtrack' 'GinMaster' 'Glodream' 'Gmuse' 'Gonca' 'Hamob' 'Hispo' 'Iconosys' 'Imlog' 'JSExploit-DynSrc' 'JSmsHider' 'Kmin' 'Ksapp' 'Loozfon' 'Luckycat' 'Maxit' 'MMarketPay' 'Mobilespy' 'Mobsquz' 'Moghava' 'Nandrobox' 'Nickspy' 'NickyRCP' 'Nyleaker' 'Opfake' 'Pirater' 'Pirates' 'PJApps' 'Placms' 'Plankton' 'Raden' 'RootSmart' 'SafeKidZone' 'Saiva' 'Sakezon' 'Sdisp' 'SeaWeth' 'SendPay' 'SerBG' 'Smspacem' 'SMSreg' 'Spy.GoneSixty' 'Spy.ImLog' 'SpyHasb' 'SpyMob' 'SpyPhone' 'Spyset' 'Stealer' 'Stealthcell' 'Steek' 'Tesbo' 'TheftAware' 'Trackplus' 'TrojanSMS.Denofow' 'TrojanSMS.Hippo' 'Updtbot' 'Vdloader' 'Vidro' 'Xsider' 'YcChar' 'Yzhc' 'Zitmo' 'Zsone'
tcp_packets	it has the number of packets TCP sent and got during communication.
dist_port_tcp	it is the total number of packets different from TCP
external_ips	represents the number the external addresses (IPs) where the application tried to communicated
vulume_bytes	it is the number of bytes that was sent from the application to the external sites
udp_packets	the total number of packets UDP transmitted in a communication
tcp_urg_packet	represents a special type of TCP packet expressing an emergency situation, where the "URG" flag in the TCP header is used

source_app_packets	it is the number of packets that were sent from the application to a remote server
remote_app_packets	number of packages received from external sources
source_app_bytes	this is the volume (in Bytes) of the communication between the application and server
remote_app_bytes	this is the volume (in Bytes) of the data from the server to the emulator

Table 3. Dataset distributions

Type	Number of Data
Benign	4704
Malicious	3141

### 3.3. Addressing the Imbalanced Data Sampling Challenge

When dealing with imbalanced class distributions in a dataset, conventional classification algorithms may exhibit a bias towards the majority class, diminishing the effectiveness of detecting minority class instances. This imbalance poses a significant challenge in achieving optimal performance with deep learning algorithms. Therefore, it becomes imperative to rectify this issue by employing techniques that balance the dataset, enhancing reliability and efficiency. Two commonly used methods are oversampling (introducing additional data) and undersampling (removing data), as highlighted by [19].

#### 3.3.1. SMOTE (Synthetic Minority Over-Sampling Technique)

SMOTE is a powerful technique designed to fortify the minority class in datasets exhibiting class imbalance, thereby promoting a more balanced learning model. This method generates synthetic examples by interpolating instances from the minority class, enabling the learning model to better discern minority class examples and improve overall performance. SMOTE effectively mitigates overfitting issues associated with random oversampling and addresses information loss resulting from random undersampling. This ensures that the model possesses a more robust and generalizable structure [20].

#### 3.3.2. SMOTETomek

SMOTETomek is a rebalancing strategy that creates a balanced dataset by over-sampling the minority class while simultaneously under-sampling the majority class. This strategy aims to obtain the examples used to achieve balance between classes in a more balanced and effective manner, as well as improve the model's capacity to obtain information from both categories more effectively. In this way, the learning model gains a more generalizing structure and reduces the possibility of misclassification, allowing more reliable results to be obtained. SMOTETomek improves the performance of learning

algorithms by providing an effective solution to balance between classes with few and many examples [21].

#### 3.3.3. ClusterCentroids

ClusterCentroids is a technique that generates synthetic samples by clustering minority class instances within the dataset using clustering algorithms and leveraging the centers of these clusters. This method aims to alleviate the challenges posed by class imbalance by improving the representation of minority class instances. By creating synthetic samples based on clustered representations, ClusterCentroids contributes to a more balanced and representative dataset, thereby enhancing the performance of learning models, especially in scenarios with imbalanced class distributions [22]. The results obtained from this phase of the study are presented in Table 4.

Table 4. Dataset distribution after preprocessing

Data Type	#Benigndata	#Maliciousdata
Original	4704	3141
SMOTE	4704	4704
SMOTETomek	4485	4485
ClusterCentroids	3141	3141

### 3.4. Deep Learning Methods

In recent years, deep learning has rapidly solved complex problems in various scientific fields and gained importance as a subfield of artificial intelligence. This development has revealed deep learning methods that are used effectively in applications such as pattern recognition and data analysis. Deep learning involves deep neural networks consisting of hierarchical layers that are capable of automatic learning, often on large and complex datasets. These methodologies have demonstrated remarkable success, notably in fields such as image and voice recognition, natural language processing, malware detection, and other cognitively demanding tasks [23], [24], [25], [26]. Deep learning contributes to the acceleration of scientific and technological developments with its ability to reveal complex relationships within data [27].

### 3.4.1. CNN

Convolutional Neural Networks (CNN) are one of the deep learning models that are effective in tasks such as computer vision [28] recognition [29] and classification [30]. CNN provides the ability to learn and generalize patterns and features more effectively, especially by being used in areas such as image and video analysis. Thanks to their filtering and pattern recognition capabilities, CNN models are used in many application areas to achieve high performance on complex visual data [31].

### 3.4.2. RNN

Traditional Neural Networks typically do not retain their final results for subsequent phases, whereas Recurrent Neural Networks (RNNs) are specifically engineered to address this constraint. RNNs offer a unique capability for data persistence through internal feedback loops, enabling them to retain memory of previous information using interconnected components. Thanks to these features, they can successfully process sequential datasets such as language modeling, text generation, and time series forecasting. In order to learn long-term dependencies more effectively, models such as GRU and LSTM, which are advanced variants of RNNs, are also used [32].

### 3.4.3. GRU

Gated Recurrent Unit (GRU) is an RNN variant that aims to learn long-term dependencies more effectively. GRU is a deep learning model that is particularly successful when applied to sequential data processing tasks such as language modeling, text generation, and time series analysis. GRUs are specifically designed to solve the vanishing gradient problem in traditional RNNs, providing an effective solution to prevent gradients from shrinking excessively over time and to prevent long-term dependencies. Thanks to their lightweight structure, GRUs offer faster training processes and less computational complexity, providing effective performance, especially on large datasets [33].

### 3.4.4. BiGRU

Different from unidirectional GRU models, the Bidirectional Gated Recurrent Unit (BiGRU) model includes information in both forward and reverse time directions. Forward GRU captures prior information and reverse GRU captures subsequent information, obtaining a wide range of context information in the network intrusion traffic prediction task and effectively extracting deep features of the traffic. These two GRUs with opposite directions jointly determine the output of the current location, thus providing a more comprehensive prediction/classification capability [34].

### 3.4.5. LSTM

Long Short-Term Memory (LSTM) networks consist of three main gates: input, output, and forget gates. These gates include a sigmoid neural network layer and a point multiplication process, which processes the input vector to determine the rate at which each component is allowed to pass. LSTM is a type of RNN and is particularly successful in time series analysis, language modeling, and natural language processing tasks. Its ability to effectively learn long-term dependencies and its capacity to store information make LSTM an effective tool in complex intra-temporal relationship and pattern recognition tasks [35], [36].

### 3.4.6. BiLSTM

Bidirectional Long Short-Term Memory (Bi-LSTM) is a type of RNN that combines memory cells and a gate mechanism, enabling efficient modeling of sequential data. Bi-LSTM has a bi-directional structure, consisting of two LSTM layers, with the input sequence being processed in the forward direction and used in the backward direction. The outputs of the two layers are combined to produce the final output, and the output of the hidden layers is passed through a linear layer that calculates probability scores. Bi-LSTM, with its ability to capture both prior and subsequent contextual information in the input sequence, provides a more comprehensive contextual understanding by simultaneously evaluating information before and after the current time step using forward and backward LSTM layers [37].

### 3.4.7. CNN+BiGRU

Compared to traditional neural networks, CNN offers advantages in weight sharing between the receiver field view and the hidden layer, especially given the non-linearity and randomness of network traffic data. Thanks to the weight-sharing mechanism, CNN can reduce network complexity and facilitate feature extraction with the same convolution kernel. The CNN-BiGRU model combines CNN and Bidirectional Gated Recurrent Unit (BiGRU) architectures, which are effective in image and sequential data analysis, capturing spatial and temporal context and offering a wide range of applications. This model can be successfully used in areas such as visual data and time series analysis. [38].

### 3.4.8. CNN+BiLSTM

Compared to traditional techniques, CNN-based feature learning enables an end-to-end information processing process from input to output, bypassing the feature extraction phase. However, considering that a single model may not provide ideal results in predicting time series data, more effective results can be achieved by successfully combining the local feature extraction capabilities of CNN with the nonlinear temporal processing capabilities of BiLSTM. In this context, the CNN-BiLSTM model is a

deep learning model that can be effectively used in visual data analysis and sequential data processing. The CNN part is used for feature extraction and the BiLSTM part is used for sequential data analysis, capturing spatial and temporal context simultaneously, providing a wide range of applications [39].

### 3.4. Machine Learning Methods

Machine learning is a branch of artificial intelligence that can make accurate predictions by providing applications with the capacity to learn from experience and data rather than predetermined rules. Machine learning algorithms work by using input data as features to predict new output values. This discipline focuses on pattern recognition and learning, improving the ability of computer systems to learn from experience and data [40].

#### 3.5.1. XGBoost

XGBoost, or eXtreme Gradient Boosting, is a standout Gradient Boosting algorithm renowned for its exceptional scalability. Boasting high speed and performance, XGBoost is known to be ten times faster than alternative methods. Its superiority lies in swift model tuning, facilitated by a distinctive regularization technique that mitigates overfitting. This algorithm is a formidable asset for tackling regression and classification challenges, demonstrating efficacy across a myriad of applications. Leveraging optimization techniques such as parallel processing, tree regularization, and feature selection, XGBoost emerges as a powerful and versatile tool in data analysis.[41], [42]. In this study, XGBoost was employed due to its remarkable scalability, high speed, and performance, making it well-suited for handling large datasets efficiently. Its unique regularization technique helps prevent overfitting, ensuring the robustness of the models developed in the study. Additionally, XGBoost's effectiveness in regression and classification tasks, coupled with its proven track record in various applications, made it a compelling choice for enhancing the accuracy of the malware detection system being investigated.

## 4. EXPERIMENTAL RESULTS AND DISCUSSION

### 4.1. Experimental setup and evaluation metrics

The computer used in the experiments is equipped with features that offer high performance and processing capacity. The main component of the system is a powerful 48-core Xeon processor, which is capable of handling intensive processing loads. The system is also equipped

with 256 GB of RAM for efficient data processing and memory management. This combination of hardware allows us to perform our experiments efficiently and effectively. Python programming language was employed for the execution of the experiments.

Standard evaluation metrics were employed in this study to assess the results obtained from the experiments such as accuracy, precision, recall, and F1-score.

The model used in the study is presented in Figure 1. As seen in the Figure 1, the study consists of two phases. In the first phase of the study, classification tasks were conducted using eight different deep learning methods, including CNN, RNN, GRU, Bi-GRU, LSTM, Bi-LSTM, CNN+BiGRU, and CNN+BiLSTM, on both the original dataset and the dataset enhanced with imbalanced data sampling techniques. Furthermore, to achieve optimal results in the classification process, hyperparameter optimization was performed using Optuna in conjunction with a genetic algorithm. The hyperparameter ranges used for optimization are provided in Table 5. Throughout this phase, the epoch value was kept constant at 50. Additionally, experiments were conducted on three different distributions of 70/30, 80/20, and 90/10 as train and test. The same training and test sets were used throughout both phases of the experiment. Specifically, after the deep learning model was trained, features were extracted and passed as input to the XGBoost classifier. Consistency was maintained across both stages by keeping the same validation and test sets during feature extraction and classification, thereby preventing data leakage or bias in model evaluation. This approach ensured a fair comparison of performance between the standalone deep learning models and the hybrid model with XGBoost. The accuracy, precision, recall, and F1-score results of these studies are presented in detail in Tables 6, 7, 8, and 9. Table 6 contains the results of the study conducted with the original dataset, while Tables 7, 8, and 9 present the results of the studies conducted with the datasets obtained after the Smote, SmoteTomek, and ClusterCentroid processes, respectively. The methods that yielded the best results were determined during this phase. In the second phase, the classification process was carried out using the deep learning structures identified as the most effective in the first phase, combined with the XGBoost ML classifier. This approach involved training the deep learning models and subsequently feeding their outputs into the XGBoost ML model. The f1-score was used as the evaluation metric, as it provides a balanced assessment of model performance by considering both precision and recall. The f1-score results for this phase are presented in Table 10.

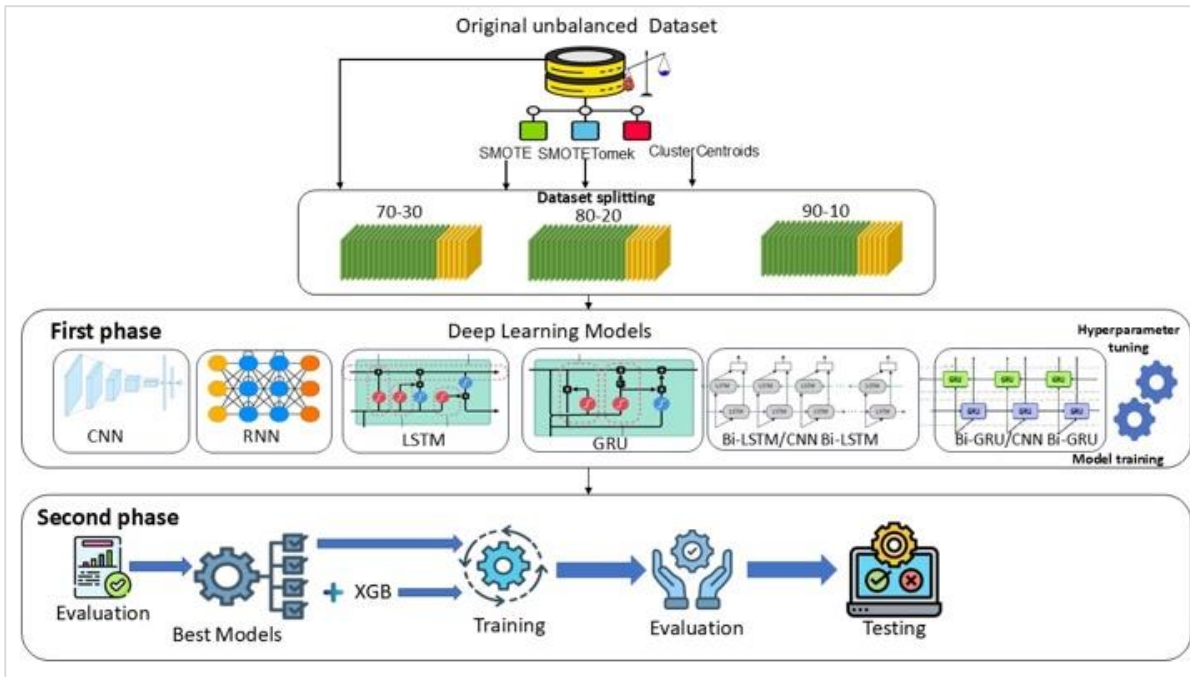


Figure 1. The proposed method

Table 5. Hyperparameter ranges

According by	Range
Optimization alg.	Mini-batchGD, MomentumGD, Adam, Adadelta, Adagrad, Adamax, Nadam
Conv_layer_number	1, 2, 3
Filter_size	32, 64, 96, 128
Kernel_size	3,5
Activation	relu, tanh, gelu, swish, selu, LeakyReLU
Kernel_initializer	uniform, lecun_uniform, normal, zero, glorot_normal, he_normal, he_uniform
Dense_layers_number	1, 2, 3, 4, 5, 6
Dense_neuron_number	32, 64, 96, 128
Dense_activation	relu, tanh, gelu, swish, selu, LeakyReLU
Dense_kernel_initializer	uniform, lecun_uniform, normal, zero, glorot_normal, he_normal, he_uniform

Table 6. Results of the Original Dataset

Model	Train (%)	Accuracy	Precision	Recall	F1-Score
CNN	70	75.83	89.11	75.06	81.48
	80	85.53	82.99	91.94	87.24
	90	79.11	87.30	80.68	83.86
RNN	70	88.91	93.88	88.29	91.00
	80	81.26	89.63	80.97	85.08
	90	74.39	83.40	77.23	80.20
GRU	70	85.00	88.61	86.58	87.58
	80	80.05	84.60	82.40	83.48
	90	83.95	88.73	85.91	87.30



<b>Bi-GRU</b>	70	91.08	89.25	95.51	92.27
	<b>80</b>	<b>91.84</b>	<b>92.94</b>	<b>93.34</b>	<b>93.14</b>
	90	87.90	90.57	90.02	90.30
<b>LSTM</b>	70	78.08	95.23	74.87	83.83
	80	87.25	91.76	87.46	89.56
	90	86.88	90.37	88.73	89.54
<b>Bi-LSTM</b>	70	82.33	88.26	83.17	85.64
	80	88.59	90.48	90.38	90.43
	<b>90</b>	<b>91.85</b>	<b>93.65</b>	<b>93.27</b>	<b>93.46</b>
<b>CNN+BiGRU</b>	70	86.79	87.62	89.99	88.78
	80	72.72	87.70	72.37	79.30
	90	88.15	92.62	88.80	90.67
<b>CNN+BiLSTM</b>	70	87.51	88.26	90.58	89.40
	80	88.34	89.73	90.60	90.17
	90	78.98	89.55	79.31	84.12

Table 7. Results of the Smote Dataset

<b>Model</b>	<b>Train (%)</b>	<b>Accuracy</b>	<b>Precision</b>	<b>Recall</b>	<b>F1-Score</b>
<b>CNN</b>	70	83.21	86.71	81.49	84.02
	80	88.95	85.94	91.23	88.51
	90	80.13	84.00	78.24	81.02
<b>RNN</b>	70	87.42	88.24	87.21	87.72
	80	86.82	83.48	89.22	86.25
	90	76.51	80.63	74.80	77.61
<b>GRU</b>	70	88.66	88.31	89.30	88.80
	80	87.83	85.09	89.81	87.38
	90	87.35	84.00	90.27	87.02
<b>Bi-GRU</b>	70	61.60	98.05	57.16	72.22
	<b>80</b>	<b>77.52</b>	<b>67.38</b>	<b>84.07</b>	<b>74.81</b>
	90	90.33	88.63	91.92	90.25
<b>LSTM</b>	70	81.44	86.78	78.87	82.64
	80	83.85	84.55	83.12	83.83
	90	83.74	79.16	87.44	83.09
<b>Bi-LSTM</b>	70	83.60	90.26	80.06	84.85
	<b>80</b>	<b>90.38</b>	<b>90.67</b>	<b>89.99</b>	<b>90.33</b>
	90	76.83	81.05	75.05	77.94
<b>CNN+BiGRU</b>	70	81.72	85.87	79.77	82.71
	80	90.06	87.88	91.71	89.75
	90	90.33	91.16	89.83	90.49
<b>CNN+BiLSTM</b>	70	88.81	87.06	90.59	88.79
	80	88.10	89.27	87.03	88.14
	90	87.04	85.89	88.12	86.99

Table 8. Results of the SmoteTomek Dataset

Model	Train (%)	Accuracy	Precision	Recall	F1-Score
CNN	70	87.70	82.59	91.41	86.78
	80	87.79	83.99	91.94	87.79
	90	85.68	84.97	86.67	85.81
RNN	70	86.59	85.55	86.81	86.17
	80	90.13	90.72	90.43	90.57
	90	75.03	92.81	68.93	79.11
GRU	70	81.05	82.21	79.66	80.91
	80	75.64	85.91	72.52	78.65
	90	86.68	92.37	83.30	87.60
Bi-GRU	70	87.11	86.77	86.83	86.80
	80	84.34	91.89	80.77	85.97
	90	74.81	97.60	67.47	79.79
LSTM	70	83.62	80.30	85.30	82.73
	80	85.95	85.17	87.60	86.36
	90	74.03	77.56	73.10	75.26
Bi-LSTM	70	91.05	93.16	89.03	91.04
	<b>80</b>	<b>94.87</b>	<b>96.16</b>	<b>94.15</b>	<b>95.14</b>
	90	88.35	89.54	87.82	88.67
CNN+BiGRU	70	88.78	86.31	90.29	88.26
	80	88.91	89.43	89.34	89.39
	90	76.47	86.71	72.50	78.97
CNN+BiLSTM	70	89.64	87.83	90.66	89.22
	80	86.23	88.15	85.86	86.99
	90	88.46	86.06	90.80	88.37

Table 9. Results of the ClusterCentroid Dataset

Model	Train (%)	Accuracy	Precision	Recall	F1-Score
CNN	70	82.02	84.89	81.14	82.97
	80	77.65	77.16	77.78	77.47
	90	89.67	91.02	89.68	90.34
RNN	70	70.93	96.51	64.62	77.41
	80	85.12	89.94	81.95	85.76
	90	76.47	78.14	77.68	77.91
GRU	70	81.17	71.53	89.92	79.68
	80	78.84	79.87	78.13	78.99
	90	87.12	84.73	90.42	87.48
Bi-GRU	70	68.22	93.73	62.90	75.28
	80	80.51	68.05	90.45	77.67
	90	71.22	56.89	83.70	67.74
LSTM	70	84.14	84.28	84.89	84.58
	80	83.93	84.19	83.65	83.92

	90	77.74	78.44	79.39	78.92
<b>Bi-LSTM</b>	70	88.70	94.35	85.32	89.60
	80	80.51	83.87	78.48	81.08
	<b>90</b>	<b>89.98</b>	<b>91.92</b>	<b>89.50</b>	<b>90.69</b>
<b>CNN+BiGRU</b>	70	85.36	82.63	88.25	85.35
	80	88.46	88.02	88.73	88.37
	90	88.55	87.13	90.94	88.99
<b>CNN+BiLSTM</b>	70	87.48	87.67	88.03	87.85
	80	78.04	73.96	80.38	77.04
	90	88.39	88.32	89.67	88.99

Tables 6, 7, 8, and 9 show that the optimal training/testing distribution, which yields the best results for each data type, occurs when 80% of the dataset is allocated to training and 20% to testing. The analysis of deep learning methods reveals that the RNN model consistently produces the lowest performance results. Furthermore, bidirectional models appear to achieve higher success rates compared to other deep learning methods. Notably, when the dataset obtained after the SmoteTomek process is used, the most successful results are achieved.

Considering these findings, an optimal f1-score of 95.14% was achieved using the Bi-LSTM deep learning method with an 80/20 training/test distribution on the dataset obtained after the SmoteTomek process. Based on the results from the first phase, it was decided to use the SmoteTomek dataset in the second phase, with 80% allocated for training and 20% for testing. The first-phase results also revealed that the lowest performance was observed in studies conducted with the dataset obtained

after applying the ClusterCentroid method. This may be attributed to the data loss resulting from the data reduction process of the ClusterCentroid method, which likely negatively impacted the results.

In the second stage of the study, different epoch values, such as 50, 100, and 150, were evaluated to assess their impact on model performance. The models were created using the hyperparameter values shared in Table 11, and the results were then compared. The variation in epoch values was chosen to investigate how training duration affects the performance of the deep learning models. The models were re-run both with and without the XGBoost algorithm, allowing for the evaluation of the impact of different epochs on the classification results and enabling a comparative analysis to identify the optimal configuration for each scenario. The results of the Bi-LSTM, Bi-GRU, and CNN models, which achieved the three highest success rates in the second phase of the study, are presented in Table 10.

Table 10. F1-Score and Accuracy (Acc) results for the top 3 results in the second phase of the study

Epoch	Score	BiLSTM	BiLSTM+XGB	BiGRU	BiGRU+XGB	CNN	CNN+XGB
50	<b>F1- Score</b>	95.04	95.79	85.97	90.42	87.79	92.03
	<b>Acc</b>	94.77	97.44	84.34	92.00	87.79	93.72
100	<b>F1- Score</b>	92.88	95.72	94.18	94.42	91.60	92.21
	<b>Acc</b>	92.36	97.97	93.98	98.24	91.36	96.41
150	<b>F1- Score</b>	95.12	<b>95.12</b>	93.50	94.25	91.68	92.75
	<b>Acc</b>	94.93	<b>99.33</b>	93.26	98.85	92.92	99.13

Table 11. Hyperparameter Values for Each Model

Model	Hyperparameter	Value
CNN	Optimization alg	Adamax
	Conv_layer_number	3
	Conv_Filters	(96, 96, 32)
	Conv_kernel_size	(3, 3, 5)
	Conv_activation	('tanh', 'tanh', 'tanh')
	Conv_kernel_initializer	('glorot_normal', 'lecun_uniform', 'he_uniform')
	Dense_layers_number	3
	Neuron_number in Dense layers	(96, 96, 64)
	Dense_activation	('tanh', 'tanh', 'relu')

	Dense_kernel_initializer	('lecun_uniform', 'lecun_uniform', 'uniform')
RNN	Optimization alg	Adamax
	Rnn_layer_number	3
	Rnn_units	(32, 96, 32)
	Rnn_activation	('relu', 'tanh', 'selu')
	Rnn_kernel_initializer	('normal', 'glorot_normal', 'uniform')
	Dense_layers_number	2
	Dense_neuron_number	(128, 128)
	Dense_activation	('tanh', 'tanh')
LSTM	Dense_kernel_initializer	('glorot_normal', 'uniform')
	Optimization alg	Adamax
	Lstm_layer_number	2
	Lstm_units	(32, 128)
	Lstm_activation	('swish', 'tanh')
	Lstm_kernel_initializer	('lecun_uniform', 'glorot_normal')
	Dense_layers_number	4
	Dense_neuron_number	(64, 32, 96, 64)
BiLSTM	Dense_activation	('swish', 'gelu', 'swish', 'swish')
	Dense_kernel_initializer	('glorot_normal', 'glorot_normal', 'normal', 'normal')
	Optimization alg	Nadam
	Lstm_layer_number	2
	Lstm_units	(128, 64)
	Lstm_activation	('tanh', 'LeakyReLU')
	Lstm_kernel_initializer	('normal', 'he_uniform')
	Dense_layers_number	5
GRU	Dense_neuron_number	(32, 128, 64, 128, 96)
	Dense_activation	('relu', 'swish', 'relu', 'tanh', 'gelu')
	Dense_kernel_initializer	('lecun_uniform', 'glorot_normal', 'lecun_uniform', 'lecun_uniform', 'glorot_normal')
	Optimization alg	Nadam
	Gru_layer_number	2
	Gru_units	(128, 128)
	Gru_activation	('swish', 'tanh')
	Gru_kernel_initializer	('he_normal', 'he_uniform')
BiGRU	Dense_layers_number	1
	Dense_neuron_number	(64)
	Dense_activation	('relu')
	Dense_kernel_initializer	('glorot_normal')
	Optimization alg	Nadam
	BiGru_layer_number	1
	BiGru_units	(64)
	BiGru_activation	('tanh')
CNN+BiGRU	BiGru_kernel_initializer	('glorot_normal')
	Dense_layers_number	3
	Dense_neuron_number	(64, 32, 32)
	Dense_activation	('tanh', 'relu', 'selu')
	Dense_kernel_initializer	('he_uniform', 'lecun_uniform', 'normal')
	Optimization alg	Adam
	Conv_layer_number	1
	Conv_Filters	(32)
	Conv_kernel_size	(3)
	Conv_activation	('tanh')
	Conv_kernel_initializer	('he_normal')
	BiGru_layer_number	3
	BiGru_units	(128, 128, 96)
CNN+BiLSTM	BiGru_activation	('gelu', 'LeakyReLU')
	BiGru_kernel_initializer	('normal', 'he_normal', 'uniform')
	Dense_layers_number	4
	Dense_neuron_number	(64, 96, 96, 96)
	Dense_activation	('relu', 'relu', 'tanh', 'tanh')
	Dense_kernel_initializer	('lecun_uniform', 'he_uniform', 'normal', 'he_uniform')
	Optimization alg	Adamax
	Conv_layer_number	3
	Conv_Filters	(32, 96, 128)
	Conv_kernel_size	(3, 5, 5)
CNN+BiLSTM	Conv_activation	('tanh', 'swish', 'swish')
	Conv_kernel_initializer	('he_normal', 'uniform', 'he_normal')
	BiLstm_layer_number	3
	BiLstm_units	(32, 96, 64)
	BiLstm_kernel_initializer	('he_uniform', 'he_normal', 'he_uniform')
	Dense_layers_number	3
	Dense_neuron_number	(32, 32, 64)
	Dense_activation	('relu', 'tanh', 'tanh')

	Dense_kernel_initializer	('lecun_uniform', 'normal', 'he_uniform')
--	--------------------------	---

According to the results of Table 10, it can be concluded that the results obtained by the hyperdization of deep learning with XGBoost outperform the results of deep learning models alone. This finding underscores the significance of adopting a combined approach, demonstrating that machine learning models achieve more effective results when leveraged together.

Furthermore, it was observed that model performance improved as the number of epochs increased. This indicates that additional learning phases allowed the model to better capture patterns within the dataset, resulting in a more generalizable representation. The increase in epochs positively impacted classification performance by improving the model's ability to capture complex patterns in the data. However, to prevent overfitting—where the model memorizes rather than generalizes—the number of epochs was capped at 150. The BiLSTM+XGBoost method achieved a remarkable 99.33% accuracy and 97.30% F1-score after 150 epochs, demonstrating the hybrid model's strong performance. Additionally, comparisons in Table 1 show that these results surpass the benchmarks of other state-of-the-art studies. The integration of deep learning and machine learning algorithms effectively complements each method's limitations, resulting in enhanced classification accuracy. These results suggest that combining the strengths of deep learning and traditional machine learning approaches can yield more robust and generalizable models for complex datasets. This hybrid methodology demonstrates considerable promise for advanced classification tasks, offering valuable applications in both academic research and industry.

## 5. CONCLUSION

This paper presents a novel approach to Android malware detection by integrating machine learning and deep learning methods, validated through an extensive experimental study. In malicious application detection, a critical risk lies in misclassifying malicious applications as benign, potentially allowing harmful software to infiltrate the system. This integrated approach aims to mitigate such risks by enhancing detection accuracy and robustness. One of the most important steps to solve this problem is to ensure a balanced distribution of the dataset. For this reason, this study first uses unbalanced data sampling techniques to balance the dataset. Then, eight different deep learning methods were used to classify the original dataset, and the data organized using unbalanced data sampling techniques. At this stage, a rigorous examination of the different training and test set distributions was performed while maintaining a constant epoch value to identify the methods that gave the most favourable results. These initial findings demonstrated the effectiveness of deep learning models on different datasets, especially when supported by unbalanced data sampling techniques.

In the next phase, the deep learning methods that showed the most promising results from the first phase were selected and the hybrid approach combining deep learning with XGBoost was applied. The analysis of the results showed that this hybrid approach improved the classification performance by 3-4%, with a significant increase especially as the epoch value increased.

The proposed hybrid model achieved an impressive accuracy of 99.33%. When compared to results from other benchmark studies, our approach consistently outperforms existing methods, showcasing its superior effectiveness in detecting Android malware. This highlights the potential of the hybrid model in delivering more accurate and reliable outcomes in malware detection.

A limitation of the proposed method is the potential increase in computational complexity resulting from the integration of deep learning with traditional machine learning algorithms. This integration may require significant computational resources and time, especially in the training phase. Furthermore, the performance of the combined model may be sensitive to hyperparameter settings and may require extensive tuning to achieve optimal results.

In addition, it is essential that such systems prevent privacy violations when processing and storing users' personal data. Therefore, the development of transparent and accountable AI systems should not only enhance security but also protect users' rights and privacy.

To address these limitations, future work will include the evaluation of different machine learning algorithms with different deep learning constructs. Furthermore, an ablation study will be conducted to investigate the impact of hyperparameter tuning on both machine learning and deep learning models to improve overall performance. In addition, the use of automated hyperparameter tuning techniques such as grid search, random search or Bayesian optimisation will be investigated to efficiently search the hyperparameter space and identify optimal configurations, thus minimising the computational overhead.

## Acknowledgements

In this study, we thank you for using the servers in "LÜTFİ ABAY ARTIFICIAL INTELLIGENCE AND ROBOTICS LABORATORY" in the application of machine learning and deep learning methods.

## REFERENCES

- [1] J. Qiu, J. Zhang, W. Luo, L. Pan, S. Nepal, and Y. Xiang, "A Survey of Android Malware Detection with Deep Neural Models", *ACM Comput. Surv.*, c. 53, sy 6, s. 126:1-126:36, 2020.
- [2] H. Zhu, Y. Li, L. Wang, and V. S. Sheng, "A multi-model ensemble learning framework for imbalanced android malware detection", *Expert Systems with Applications*, c. 234, s. 120952, 2023.
- [3] H. Bakır and R. Bakır, "DroidEncoder: Malware detection using auto-encoder based feature extractor and machine learning algorithms", *Computers and Electrical Engineering*, c. 110, s. 108804, 2023.
- [4] O. N. Elayan and A. M. Mustafa, "Android Malware Detection Using Deep Learning", *Procedia Computer Science*, c. 184, ss. 847-852, 2021.
- [5] K. Bakour and H. M. Ünver, "DeepVisDroid: android malware detection by hybridizing image-based features with deep learning techniques", *Neural Comput & Applic*, c. 33, sy 18, ss. 11499-11516, 2021.
- [6] H. AlOmari, Q. M. Yaseen, and M. A. Al-Betar, "A Comparative Analysis of Machine Learning Algorithms for Android Malware Detection", *Procedia Computer Science*, c. 220, ss. 763-768, 2023
- [7] A. Arthi., K. Aggarwal, R. Karthikeyan, S. Kayalvili, S. S, and A. Srivastava, "Hybrid Multimodal Machine Learning Driven Android Malware Recognition and Classification Model", 2023 7th International Conference on Electronics, Communication and Aerospace Technology (ICECA), Coimbatore, India: IEEE, ss. 1555-1560, 2023.
- [8] P. Yadav, N. Menon, V. Ravi, S. Vishvanathan, and T. D. Pham, "EfficientNet convolutional neural networks-based Android malware detection", *Computers & Security*, c. 115, s. 102622, 2022.
- [9] R. Yumlembam, B. Issac, S. M. Jacob, and L. Yang, "IoT-Based Android Malware Detection Using Graph Neural Network With Adversarial Defense", *IEEE Internet of Things Journal*, c. 10, sy 10, ss. 8432-8444, 2023.
- [10] Z. Z. Jundi and H. Alyasiri, "Android Malware Detection Based on Grammatical Evaluation Algorithm and XGBoost", 2023 Al-Sadiq International Conference on Communication and Information Technology (AICCIT), Al-Muthana, Iraq: IEEE, ss. 70-75, 2023.
- [11] M. A. Mohammed, M. Asante, S. Alornyo, and B. O. Essah, "Android applications classification with deep neural networks", *Iran J Comput Sci*, c. 6, sy 3, ss. 221-232, 2023.
- [12] J. Tang et al., "Android malware detection based on a novel mixed bytecode image combined with attention mechanism", *Journal of Information Security and Applications*, c. 82, s. 103721, 2024
- [13] Y. Seyfari and A. Meimandi, "A new approach to android malware detection using fuzzy logic-based simulated annealing and feature selection", *Multimed Tools Appl*, c. 83, sy 4, ss. 10525-10549, 2024
- [14] X. Fu, C. Jiang, C. Li, J. Li, X. Zhu, and F. Li, "A hybrid approach for Android malware detection using improved multi-scale convolutional neural networks and residual networks", *Expert Systems with Applications*, c. 249, s. 123675, 2024.
- [15] Z. Liu, R. Wang, N. Japkowicz, H. M. Gomes, B. Peng, and W. Zhang, "SeGDroid: An Android malware detection method based on sensitive function call graph learning", *Expert Systems with Applications*, c. 235, s. 121125, 2024.
- [16] R. Raman, K. R. Nirmal, A. Gehlot, S. Trivedi, D. Sain, and R. Ponnusamy, "Detecting Android Malware and Sensitive Data Flows Using Machine Learning Techniques", 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India: IEEE, ss. 1694-1698, 2022.
- [17] M. M. Alani and A. I. Awad, "AdStop: Efficient flow-based mobile adware detection using machine learning", *Computers & Security*, c. 117, s. 102718, 2022.
- [18] A. Duran and H. Bakır, "Hiperparametreleri Ayarlanmış Makine Öğrenimi Algoritmalarını Kullanarak Android Sistemlerde Kötü Amaçlı Yazılım Tespiti", *Uluslararası Sivas Bilim ve Teknoloji Üniversitesi Dergisi*, c. 2, sy 1, Art. sy 1, 2023.
- [19] E. Baghirova, "Evaluating the Performance of Different Machine Learning Algorithms for Android Malware Detection", 2023 5th International Conference on Problems of Cybernetics and Informatics (PCI), Baku, Azerbaijan: IEEE, ss. 1-4, 2023.
- [20] A. Zhang, H. Yu, S. Zhou, Z. Huan, and X. Yang, "Instance weighted SMOTE by indirectly exploring the data distribution", *Knowledge-Based Systems*, c. 249, s. 108919, 2022.
- [21] M. G. Lanjewar, K. G. Panchbhavi, and L. B. Patle, "Fusion of transfer learning models with LSTM for detection of breast cancer using ultrasound images", *Computers in Biology and Medicine*, c. 169, s. 107914, 2024.
- [22] W.-C. Lin, C.-F. Tsai, Y.-H. Hu, and J.-S. Jhang, "Clustering-based undersampling in class-imbalanced data", *Information Sciences*, c. 409-410, ss. 17-26, 2017.
- [23] R. Ghanem and H. Erbay, "Spam detection on social networks using deep contextualized word representation", *Multimed Tools Appl*, c. 82, sy 3, ss. 3697-3712, 2023.
- [24] H. Bakir and R. Bakir, "Evaluating The Robustness of Yolo Object Detection Algorithm in Terms Of Detecting Objects in Noisy Environment", *Journal of Scientific Reports-A*, sy 054, ss. 1-25, 2023.
- [25] J. B. Lee and H. G. Lee, "Quantitative analysis of automatic voice disorder detection studies for hybrid feature and classifier selection", *Biomedical Signal Processing and Control*, c. 91, s. 106014, 2024.
- [26] Y. Alaca and Y. Çelik, "Cyber attack detection with QR code images using lightweight deep learning models", *Computers & Security*, c. 126, s. 103065, 2023.
- [27] J. Zhang, W. Gong, L. Ye, F. Wang, Z. Shangguan, and Y. Cheng, "A Review of deep learning methods for denoising of medical low-dose CT images", *Computers in Biology and Medicine*, s. 108112, 2024.
- [28] S. Kaushal, D. K. Tammineni, P. Rana, M. Sharma, K. Sridhar, and H.-H. Chen, "Computer vision and deep learning-based approaches for detection of food nutrients/nutrition: New insights and advances", *Trends in Food Science & Technology*, c. 146, s. 104408, 2024.

- [29] S. Raziani and M. Azimbagirad, "Deep CNN hyperparameter optimization algorithms for sensor-based human activity recognition", *Neuroscience Informatics*, c. 2, sy 3, s. 100078, 2022.
- [30] S. Bhardwaj and M. Dave, "Enhanced neural network-based attack investigation framework for network forensics: Identification, detection, and analysis of the attack", *Computers & Security*, c. 135, s. 103521, 2023.
- [31] E. K. Yılmaz, K. Adem, S. Kılıçarslan, and H. A. Aydın, "Classification of lemon quality using hybrid model based on Stacked AutoEncoder and convolutional neural network", *Eur Food Res Technol*, c. 249, sy 6, ss. 1655-1667, 2023.
- [32] N. Raj, "Prediction of Stock Market Using LSTM-RNN Model", içinde 2023 International Conference on Self Sustainable Artificial Intelligence Systems (ICSSAS), Erode, India: IEEE, ss. 623-628, 2023.
- [33] Y. Yang, Chaoluomeng, and N. Razmjoooy, "Early detection of brain tumors: Harnessing the power of GRU networks and hybrid dwarf mongoose optimization algorithm", *Biomedical Signal Processing and Control*, c. 91, s. 106093, 2024.
- [34] W. Zheng, P. Cheng, Z. Cai, and Y. Xiao, "Research on Network Attack Detection Model Based on BiGRU-Attention", içinde 2022 4th International Conference on Frontiers Technology of Information and Computer (ICFTIC), Qingdao, China: IEEE, ss. 979-982, 2022.
- [35] E. K. Yılmaz and M. A. Akcayol, "SUST-DDD: A Real-Drive Dataset for Driver Drowsiness Detection", *Proceeding of the 31st Conference of Fruct Association*, 2022.
- [36] R. Ghanem, H. Erbay, and K. Bakour, "Contents-Based Spam Detection on Social Networks Using RoBERTa Embedding and Stacked BLSTM", *SN COMPUT. SCI.*, c. 4, sy 4, s. 380, 2023.
- [37] R. Wang, X. Ji, S. Xu, Y. Tian, S. Jiang, and R. Huang, "An empirical assessment of different word embedding and deep learning models for bug assignment", *Journal of Systems and Software*, c. 210, s. 111961, 2024.
- [38] T. Wang, L. Fu, Y. Zhou, and S. Gao, "Service price forecasting of urban charging infrastructure by using deep stacked CNN- BiGRU network", *Engineering Applications of Artificial Intelligence*, c. 116, s. 105445, Kas. 2022.
- [39] B. Song, Y. Liu, J. Fang, W. Liu, M. Zhong, and X. Liu, "An optimized CNN-BiLSTM network for bearing fault diagnosis under multiple working conditions with limited training samples", *Neurocomputing*, c. 574, s. 127284, 2024.
- [40] B. Samia, Z. Soraya, and M. Malika, "Fashion Images Classification using Machine Learning, Deep Learning and Transfer Learning Models", içinde 2022 7th International Conference on Image and Signal Processing and their Applications (ISPA), ss. 1-5, 2022.
- [41] W. Cao, Y. Liu, H. Mei, H. Shang, and Y. Yu, "Short-term district power load self-prediction based on improved XGBoost model", *Engineering Applications of Artificial Intelligence*, c. 126, s. 106826, 2023.
- [42] A. Maleki, M. Raahemi, and H. Nasiri, "Breast cancer diagnosis from histopathology images using deep neural network and XGBoost", *Biomedical Signal Processing and Control*, c. 86, s. 105152, 2023.

# A Discussion on Potential Integration of Quantum Encryption with Super Artificial Intelligence

*Literatür Makalesi/Literature Article*

 Ahmet EFE<sup>1</sup>\*

<sup>1</sup>International Federation of Red Cross and Red Crescent, Europa and Central Asia Regional Risk Management, Ankara, Türkiye

[icsiacag@gmail.com](mailto:icsiacag@gmail.com)

(Geliş/Received:24.07.2024; Kabul/Accepted:05.12.2024)

DOI: 10.17671/gazibtd.1521796

**Abstract**— This study delves into the possible integration of super artificial intelligence (SAI) with quantum encryption, a revolutionary technology that harnesses the principles of quantum mechanics to secure sensitive information. While quantum encryption promises unparalleled security through mechanisms like quantum key distribution (QKD) and quantum entanglement, it also faces substantial challenges. These include susceptibility to noise, scalability limitations, high implementation costs, and public trust issues. With the advent of quantum computing, traditional encryption methods are becoming increasingly vulnerable, creating an urgent need for quantum-resistant solutions. The study proposes that SAI, when integrated with quantum encryption, has the potential to enhance security, but also introduces novel risks such as security breaches, bias, and transparency issues. By analyzing these risks and benefits, the study aims to develop mitigation strategies to optimize the advantages of this integration. Through a thorough exploration of quantum encryption's conceptual and theoretical foundations, the study examines critical tools, methodologies, and variables, offering insights into future market trends and economic impacts. The research further proposes a function modelling to quantify the success probability of secure key establishment within quantum encryption protocols. Ultimately, this study contributes to advancing the understanding of the risks and opportunities surrounding the fusion of SAI and quantum encryption, providing valuable recommendations for secure and scalable implementation in various industries.

**Keywords**—quantum encryption, security breach risks, bias risk, lack of transparency risk, efficient key generation, süper artificial intelligence

## Kuantum Şifrelemenin Süper Yapay Zeka ile Potansiyel Entegrasyonu Üzerine Bir Tartışma

**Özet**— Bu çalışma, süper yapay zekâ (SAI) ile kuantum şifrelemenin muhtemel entegrasyonunu incelemektedir. Kuantum mekaniğinin prensiplerinden yararlanarak hassas bilgileri güvence altına alan devrim niteliğinde bir teknoloji olan kuantum şifreleme, kuantum anahtar dağıtımı (QKD) ve kuantum dolanıklık gibi mekanizmalarla benzersiz bir güvenlik vaat ederken, aynı zamanda önemli zorluklarla da karşı karşıyadır. Bu zorluklar arasında gürültüye duyarlılık, ölçeklenebilirlik kısıtlamaları, yüksek uygulama maliyetleri ve kamuoyunda güven eksikliği yer almaktadır. Kuantum hesaplamının gelişimiyle birlikte geleneksel şifreleme yöntemleri giderek daha fazla tehlike altına girmekte ve kuantum dirençli çözümler ihtiyacını acil hale getirmektedir. Çalışma, SAI'nin kuantum şifreleme ile entegre edildiğinde güvenliği artırma potansiyeline sahip olduğunu, ancak aynı zamanda güvenlik ihlalleri, önyargı ve şeffaflık sorunları gibi yeni riskler ortaya çıkarabileceğini önermektedir. Bu riskleri ve faydaları analiz ederek, çalışmanın amacı bu entegrasyonun avantajlarını en üst düzeye çıkarmak için risk azaltma stratejileri geliştirmektir. Kuantum şifrelemenin kavramsal ve teorik temellerini derinlemesine inceleyen çalışma, kritik araçları, metodolojileri ve değişkenleri ele alarak gelecekteki piyasa trendlerine ve ekonomik etkilere dair önemli öngörüler sunmaktadır. Araştırma ayrıca, kuantum şifreleme protokollerinde güvenli anahtar kurulumu başarı olasılığını nicel olarak belirlemek için bir fonksiyon modeli önermektedir. Sonuç olarak, bu çalışma, SAI ve kuantum şifrelemenin birleşimi etrafındaki riskler ve fırsatlar hakkında anlayışı ilerletmeye katkıda bulunarak çeşitli sektörlerde güvenli ve ölçeklenebilir uygulamalar için değerli öneriler sunmaktadır.

**Anahtar Kelimeler**— kuantum şifreleme, güvenlik ihlali riskleri, bias riski, şeffaflık eksikliği riski, verimli anahtar üretimi, süper yapay zeka



## 1. INTRODUCTION

Quantum encryption, a method leveraging quantum mechanical properties to secure communication, has recently been explored in combination with artificial intelligence (AI), which encompasses systems capable of performing tasks like decision-making and speech recognition. This integration brings both risks and benefits. On the risk side, AI systems can be vulnerable to security breaches, potentially compromising quantum encryption if attacked. Biases in AI due to flawed training data can also undermine security, while the complexity and lack of transparency in AI algorithms can erode trust in quantum encryption's effectiveness. However, AI offers notable benefits, such as enhanced security through the detection of breaches, more efficient key generation for faster encryption, and improved performance by optimizing communication channels and reducing errors. This evolving relationship between AI and quantum encryption holds both potential and challenges for future security solutions.

This study explores the potential risks and benefits of integrating SAI with quantum encryption, a promising technology known for its potential to provide unparalleled security for sensitive information. Despite its promise, quantum encryption faces significant challenges, such as vulnerability to noise, scalability issues, high implementation costs, and trust concerns. As quantum computing advances, traditional encryption standards are increasingly at risk, necessitating the exploration of new, quantum-resistant methods. The study posits that while SAI could enhance quantum encryption's security, it may also introduce new risks that could compromise the system's integrity. The research aims to propose strategies to mitigate these risks while maximizing the benefits of the SAI-quantum encryption integration.

Notably, no prior studies have approached the issue from this unique angle, highlighting this research's significant contribution to the literature.

This study aims to explore the conceptual and theoretical framework of quantum encryption, including the principles of quantum mechanics and its application to cryptography. The research problem and discussions focus on the potential problems and key risks associated with quantum encryption, as well as the benefits of improved security, efficient key generation, and enhanced performance. To provide a comprehensive overview of the market and economy, the study will also present relevant data, facts, and statistics. This includes an analysis of the current and future market trends, as well as the economic impact of quantum encryption on various industries. However, despite the numerous advantages of quantum encryption, there are also potential risks that must be considered. This study will examine the security breach risks, bias risk, and lack of transparency risk associated with quantum encryption. By understanding these risks, organizations can develop strategies to mitigate potential threats and ensure the safe and effective implementation of quantum encryption technology. The study will also discuss the

benefits of improved security, efficient key generation, and enhanced performance that quantum encryption offers. This includes the ability to protect sensitive data, reduce the risk of cyber-attacks, and improve communication efficiency. Finally, the study will identify the organizational and technical requirements for business development, including the necessary infrastructure, training, and personnel. By providing a comprehensive analysis of the conceptual, theoretical, and practical aspects of quantum encryption, this study aims to promote a better understanding of this revolutionary technology and its potential for enhancing security in the digital age.

## 2. CONCEPTUAL AND THEORETICAL FRAMEWORK

Quantum encryption is a powerful technology that leverages the properties of quantum mechanics to provide secure communication channels between two parties. In this section, it will be discussed various methodologies, tools, formulations, parameters, variables, and functions used for quantum encryption.

### 2.1. Methodologies

There are several methodologies used for quantum encryption, including BB84, E91, and B92. BB84 is a popular method used for secure communication, which uses the principles of quantum mechanics to send encrypted messages. E91 is another protocol that uses quantum entanglement to establish a shared key between two parties. Finally, B92 is a protocol that uses single photons to transmit information securely between two parties [1]. There are several methodologies for quantum encryption, including:

1. **Quantum Key Distribution (QKD):** QKD is a protocol that uses quantum mechanics to establish a shared secret key between two parties. The key can then be used to encrypt and decrypt messages sent between the two parties. QKD has been shown to be secure against both eavesdropping and man-in-the-middle attacks [2, 3].
2. **Quantum Teleportation:** Quantum teleportation is a protocol that allows the transfer of quantum information from one location to another without physically transporting the quantum state. This protocol can be used to securely transmit information, including encryption keys [4].
3. **Quantum Cryptography:** Quantum cryptography is a broad field that encompasses several protocols for secure communication using quantum mechanics. These protocols include QKD and quantum error correction, which can be used to protect quantum information from errors introduced during transmission [5].

4. **Quantum Steganography:** Quantum steganography is a method of hiding secret information in a quantum state. This can be done by manipulating the quantum state in a way that is imperceptible to an eavesdropper. Quantum steganography has been shown to be secure against eavesdropping attacks [6].

Various tools are available for implementing quantum encryption, including QKD devices, QKD systems, and quantum networks. QKD devices use single photons to transmit information between two parties, while QKD systems use entangled photons to establish a shared key between two parties. Finally, quantum networks are used to distribute the shared key to multiple parties [1].

## 2.2. Formulations

Several formulations are used in quantum encryption, including the density matrix formalism, the Bell inequality, and the Heisenberg uncertainty principle. The density matrix formalism is used to describe the state of a quantum system, while the Bell inequality is used to test whether two particles are entangled. Finally, the Heisenberg uncertainty principle is used to describe the relationship between the position and momentum of a particle [5]. There are several different formulations of quantum encryption, including:

1. **BB84 Protocol:** The BB84 protocol, developed by Charles Bennett and Gilles Brassard in 1984, is one of the most widely used protocols for QKD. In this protocol, the sender (Alice) encodes a message using a random sequence of qubits (quantum bits), which can be in one of four possible states. The receiver (Bob) then measures the qubits using a randomly chosen basis, and the two parties compare their results to determine a shared secret key that can be used for secure communication [3].
2. **E91 Protocol:** The E91 protocol, proposed by Artur Ekert in 1991, is another QKD protocol that uses entangled pairs of particles (such as photons) to distribute a secret key. In this protocol, Alice and Bob each measure one particle from an entangled pair in a randomly chosen basis, and the results are used to generate a shared key. The security of this protocol relies on the principles of quantum entanglement, which ensure that any attempt to eavesdrop on the communication will be detected [2].
3. **B92 Protocol:** The B92 protocol, proposed by Charles Bennett in 1992, is a QKD protocol that uses a two-state system (such as a photon with two polarization states) to distribute a secret key. In this protocol, Alice sends a randomly chosen polarization state to Bob, who measures it using one of two possible measurements. The protocol is designed so that if an eavesdropper attempts to intercept the photon and measure it, there is a high probability that the measurement will alter the photon's state, which will be detected by Alice and Bob. [4]

## 2.3. Variables

Several variables are used in quantum encryption, including the quantum bit (qubit), the quantum state, and the entanglement. The qubit is used to represent the basic unit of quantum information, while the quantum state describes the state of a quantum system. Finally, entanglement is a phenomenon that occurs when two particles are correlated in such a way that the state of one particle is dependent on the state of the other particle [5]. Quantum encryption involves the use of various variables to ensure secure communication between two parties. These variables include:

1. **Quantum key distribution (QKD):** This is a technique used in quantum cryptography to establish a secret key between two parties by exploiting the quantum properties of photons. The key is generated and shared in such a way that any attempt to intercept the photons will be detectable, thus ensuring secure communication. [3].
2. **Entanglement:** This refers to the correlation between two or more quantum systems such that the state of one system is dependent on the state of the other(s). Entanglement is used in quantum cryptography to ensure the security of the transmitted information by detecting any attempt to intercept or eavesdrop on the communication channel [2].
3. **Polarization:** In quantum cryptography, polarization refers to the orientation of a photon's electric field. It is used in QKD to encode information onto the photons that are being transmitted between the two parties. By measuring the polarization of the photons, the parties can extract the secret key for secure communication. [3].
4. **Photon transmission rate:** This variable refers to the number of photons that are transmitted per second between the two parties during the communication process. The higher the transmission rate, the faster the communication process, but this can also increase the risk of interception [7].

Therefore, quantum encryption is a powerful technology that relies on various methodologies, tools, formulations, parameters, variables, and functions to provide secure communication channels between two parties. By understanding these concepts, researchers can continue to improve the security and reliability of quantum encryption systems. Unlike classical encryption methods, which rely on mathematical algorithms, quantum encryption relies on the fundamental properties of quantum mechanics, such as entanglement and superposition, to provide secure communication. In this discussion, it will be provided a detailed theoretical background of quantum encryption, drawing from several key references in the field.

One of the fundamental principles of quantum mechanics is the concept of superposition. According to this principle, a quantum system can exist in multiple states

simultaneously, until it is measured or observed. This principle has been used to develop a type of quantum encryption called QKD. QKD relies on the ability of two parties to generate and share a secret key that is secure against any eavesdropping attempts. The key is generated by sending quantum particles, such as photons, over a communication channel. The properties of these particles, such as their polarization or phase, are used to encode the bits of the key. The key is then transmitted to the receiver, who measures the particles and uses the results to reconstruct the key. Any attempt to eavesdrop on the key will inevitably disturb the particles, introducing errors that can be detected by the sender and receiver.

The security of QKD relies on another key principle of quantum mechanics, called the no-cloning theorem. According to this principle, it is impossible to create an exact copy of an unknown quantum state. This means that any attempt to intercept and copy the quantum particles used in QKD will result in a distorted version of the original state, making it impossible for the eavesdropper to gain any useful information about the key.

Another important concept in quantum encryption is entanglement, which describes a special correlation between two quantum particles that are created together. These particles have properties that are correlated in a way that cannot be explained by classical physics. This correlation can be used to create a type of quantum encryption called quantum teleportation, which allows for the secure transmission of quantum states between two parties. The basic idea is that a sender can use an entangled pair of particles to transmit the quantum state of another particle to a receiver, without actually sending the original particle itself. This allows for secure communication of quantum states, which is important for applications such as quantum computing and cryptography.

There are several other theoretical concepts that underlie the development of quantum encryption, including the uncertainty principle, which describes the limits of measurement in quantum mechanics, and the complementarity principle, which describes the dual nature of quantum particles as both waves and particles. These concepts are used to develop more advanced types of quantum encryption, such as quantum secure direct communication (QSDC) and quantum digital signatures.

Therefore, quantum encryption is a technology that relies on the fundamental principles of quantum mechanics to provide secure communication channels. These principles include superposition, entanglement, the no-cloning theorem, the uncertainty principle, and the complementarity principle. By leveraging these concepts, quantum encryption provides a new level of security that is impossible to achieve using classical encryption methods.

#### 2.4. Function Formula

Based on multiple studies on formulation [2, 4, 7], it is tried to define the function for the success probability of establishing a secure key using the BB84 protocol:

$$P_{success} = f(n, e, \eta, t, q, p)$$

The elements of the function model are as follows:

- $P_{success}$  = Probability of successfully establishing a secure key
- $n$  = Number of photons transmitted
- $e$  = Error rate in the key generation process
- $\eta$  = Efficiency of the photon detectors
- $t$  = Transmission distance between the sender (Alice) and receiver (Bob)
- $q$  = Qubit error rate due to quantum noise
- $p$  = Probability of photon loss during transmission

Dependent Variable:

- $P_{success}$  : The probability of successfully establishing a secure key, dependent on the independent variables.

Independent Variables:

- $n$ : The total number of photons transmitted. A higher number of photons can improve the probability of establishing a secure key but also increases the risk of detection by eavesdroppers.
- $e$ : The error rate in the key generation process, which affects the security and integrity of the key.
- $\eta$ : The efficiency of the photon detectors, which influences the detection rate of the transmitted photons.
- $t$ : The transmission distance, affecting the probability of photon loss and the overall security of the communication.
- $q$ : The qubit error rate, representing the quantum noise in the system. Lower noise levels result in a higher probability of secure key establishment.
- $p$ : The probability of photon loss during transmission, affecting the number of photons successfully received and used in key generation.

Coefficients and Parameters

To quantify the impact of each independent variable, coefficients can be defined as follows:

$$\alpha, \beta, \gamma, \delta, \epsilon, \zeta$$

The function can then be expressed as:

$$P_{success} = \alpha n - \beta e + \gamma \eta - \delta t - \epsilon q - \zeta p$$

- $\alpha$ : Coefficient for the number of photons transmitted, indicating the impact of increasing  $n$  on  $P_{success}$
- $\beta$ : Coefficient for the error rate, indicating how the error rate  $e$  negatively impacts  $P_{success}$
- $\gamma$ : Coefficient for the detector efficiency, showing the positive impact of increasing  $\eta$  on  $P_{success}$
- $\delta$ : Coefficient for the transmission distance, representing the negative impact of increasing  $t$  on  $P_{success}$
- $\epsilon$ : Coefficient for the qubit error rate, indicating the negative impact of increasing  $q$  on  $P_{success}$
- $\zeta$ : Coefficient for photon loss probability, showing the negative impact of increasing  $p$  on  $P_{success}$

By understanding and optimizing these variables and coefficients, researchers can improve the success

probability of quantum encryption protocols, ensuring secure communication channels based on the principles of quantum mechanics.

### 3. RESEARCH PROBLEM AND DISCUSSIONS

While quantum encryption has the potential to revolutionize the field of information security, it is not without its share of problems and issues. In this section, it will be discussed some of the key problems and issues with quantum encryption.

One of the most significant problems with quantum encryption is the issue of noise. Noise refers to any interference or distortion that can occur during the transmission of quantum states. Even small amounts of noise can have a significant impact on the ability of quantum encryption to function properly. Research has shown that current quantum encryption systems are vulnerable to noise, and more work needs to be done to address this issue [8].

Another issue with quantum encryption is the problem of scalability. While quantum encryption has been shown to be effective for secure communication over short distances, it is not yet clear whether it can be scaled up to work over longer distances. This is because the performance of quantum encryption systems can degrade over longer distances due to factors such as loss and dispersion in the transmission medium [9].

A related issue is the cost of implementing quantum encryption. Quantum encryption requires specialized hardware and infrastructure, which can be expensive to develop and deploy. In addition, the cost of scaling up quantum encryption to work over longer distances is likely to be substantial [10].

Another problem with quantum encryption is the issue of trust. In order for quantum encryption to work effectively, users must be able to trust the devices and infrastructure used to generate and transmit quantum states. However, it can be difficult to verify the security of these devices and infrastructure, particularly in cases where they are manufactured or maintained by third-party vendors [8]. Finally, there is the issue of compatibility. In order for quantum encryption to be widely adopted, it needs to be compatible with existing communication protocols and infrastructure. However, this is not always the case, and significant changes may be needed in order to integrate quantum encryption with existing systems [10].

Therefore, while quantum encryption has the potential to provide a high level of security for communication, it is not without its share of problems and issues. Addressing these issues will require further research and development, as well as careful consideration of the social, economic, and political implications of widespread adoption of quantum encryption.

Quantum encryption investments have several potential problems and areas of research that require attention. One

potential problem with quantum encryption investments is the high cost of implementing and maintaining the technology. The development of quantum encryption systems requires expensive equipment and specialized expertise [11]. Moreover, quantum encryption systems are susceptible to noise and interference, which can lead to errors and reduce the effectiveness of the encryption [8]. These factors increase the overall cost of quantum encryption systems and limit their practicality for widespread adoption.

Another issue with quantum encryption investments is the risk of quantum attacks on current encryption standards. While quantum encryption offers security against eavesdropping and interception, it does not necessarily protect against attacks on the underlying algorithms [8]. As quantum computing advances, traditional encryption standards such as RSA and AES will become vulnerable to attacks that can decrypt sensitive data [12]. Researchers are exploring new encryption methods that are resistant to quantum attacks, such as lattice-based cryptography and code-based cryptography [13].

A third area of research for quantum encryption investments is the development of quantum communication networks. While quantum encryption can secure communication between two parties, it becomes more challenging to implement on a larger scale [11]. Quantum communication networks require the ability to transmit quantum signals over long distances, which introduces additional challenges such as signal loss and decoherence [8]. Researchers are exploring ways to extend the range of quantum communication, such as the use of quantum repeaters [14].

Therefore, quantum encryption investments have the potential to provide secure communication, but there are several challenges that must be addressed. These challenges include the high cost of implementing and maintaining quantum encryption systems, the risk of quantum attacks on current encryption standards, and the development of quantum communication networks. Researchers are actively working to address these challenges and develop new methods for secure communication. The research problem statement, key assumptions and research hypothesis have been developed accordingly:

#### 3.1. Problem Statement

Quantum encryption is a promising technology that has the potential to provide unparalleled security for sensitive information. However, with the advent of super artificial intelligence (SAI), there is a possibility that the security of quantum encryption could be compromised. There is a need to explore the potential risks and benefits of using quantum encryption with SAI.

#### 3.2. Key Assumptions

1. Quantum encryption is a viable and effective method for securing information.

2. SAI is a technology that can surpass human intelligence and has the ability to analyze and manipulate data at an unprecedented level.
3. The integration of SAI with quantum encryption could have significant implications for the security of sensitive information.
4. The risks and benefits of using SAI with quantum encryption are not yet fully understood.

### 3.3. Research Hypothesis

The integration of SAI with quantum encryption has the potential to enhance the security of sensitive information, but it also carries the risk of compromising the integrity of the encryption. This study aims to examine the potential risks and benefits of using SAI with quantum encryption, and to propose strategies for mitigating the potential risks while leveraging the benefits of this emerging technology.

## 4. DATA, FACTS AND STATISTICS ON MARKET AND ECONOMY

The market for quantum encryption is expected to grow significantly in the coming years, with a projected compound annual growth rate (CAGR) of around 17% from 2019 to 2026 (MarketsandMarkets, 2020). The increasing demand for secure communication in various industries, such as finance, healthcare, and military, is expected to drive the growth of the quantum encryption market.

Artificial intelligence, on the other hand, is a rapidly growing field that involves the development of intelligent machines that can perform tasks that typically require human intelligence, such as learning, reasoning, and decision-making. The global artificial intelligence market was valued at \$16.06 billion in 2018 and is expected to reach \$190.61 billion by 2025, growing at a CAGR of 36.2% from 2019 to 2025 [16]. The growth of the artificial intelligence market can be attributed to the increasing demand for AI-based solutions in various industries, such as healthcare, finance, and retail.

Quantum encryption and artificial intelligence are two rapidly growing technologies that are expected to have a significant impact on the economy in the coming years. The growth of both technologies is expected to be driven by the increasing demand for secure communication and data protection, as well as the advancements in AI-based solutions that can provide improved efficiency and performance.

In addition to thousand number of startups in quantum encryption worldwide, there are several companies are actively investing big money in quantum encryption technology, recognizing its potential to revolutionize data security:

1. IBM: IBM is a major player in quantum encryption and quantum computing. They have developed

quantum-safe cryptography protocols and are working with various partners, including Vodafone, to integrate these protocols into telecommunications networks. This collaboration aims to protect data against future quantum threats and optimize network performance through quantum technologies [17].

2. Vodafone: Vodafone has partnered with IBM to explore quantum-safe cybersecurity and join the IBM Quantum Network. This collaboration allows Vodafone to access IBM's advanced quantum computing systems and expertise, which will help them validate and progress potential quantum use cases in telecommunications [17].
3. Google: Google's Quantum AI team is heavily invested in quantum computing research, including developing algorithms and technologies that could enhance encryption methods. Their work aims to leverage quantum capabilities to secure data against potential quantum-based decryption attempts in the future [16].
4. Amazon: Through Amazon Web Services (AWS), Amazon is also investing in quantum computing, including quantum-safe encryption methods. AWS offers quantum computing services like Amazon Braket, which allows researchers to develop and test quantum algorithms that could be used for secure data encryption [16, 17].
5. Arqit Quantum Inc: Arqit is a leader in quantum-safe encryption, having developed a Symmetric Key Agreement Platform that enhances data transmission security across networks. They have partnered with Telecom Italia Sparkle to create the first quantum-safe VPN, showcasing the practical application of their technology in ensuring secure communications [15].

These companies are at the forefront of integrating quantum encryption technologies to protect against the evolving landscape of cybersecurity threats, highlighting the importance and potential of quantum encryption in various industries. Therefore, quantum encryption and artificial intelligence are two rapidly growing fields that are expected to have a significant impact on the economy in the coming years. The increasing demand for secure communication and data protection, as well as advancements in AI-based solutions, are expected to drive the growth of both technologies.

Furthermore, Turkey has been investing heavily in research and development (R&D) in recent years, with the government allocating significant funds towards scientific projects. In 2020, Turkey's total R&D expenditure reached 2.2% of GDP, with a focus on technology and innovation. This indicates that Turkey has the necessary resources and expertise to drive innovation in quantum encryption.

In addition, Turkey has a strong history of collaboration with European Union (EU) countries, particularly in the

area of scientific research. The EU's Horizon 2020 program, for example, has provided funding and support to Turkish researchers and institutions, including those involved in quantum encryption research. This collaboration provides Turkey with access to cutting-edge technology and expertise that can help accelerate its development in the field.

However, despite these advantages, there are several challenges that Turkey needs to address to realize the full potential of quantum encryption technology. One major challenge is the lack of trained experts in the field, which is a common issue in many countries. This shortage of experts can hinder the progress of research and development projects and slow down the adoption of the technology.

Another challenge is the lack of a well-defined national strategy for quantum encryption. While the Turkish government has shown support for R&D in the field, there is no clear roadmap or plan for the development and deployment of quantum encryption technology. A well-defined strategy is crucial to ensure that resources are effectively utilized and that progress is made in a coordinated and efficient manner.

Therefore, Turkey has significant potential to become a leader in the quantum encryption industry, given its strong IT sector, strategic location, and government support for research and development. However, to realize this potential, Turkey needs to address the challenges of a shortage of experts and the absence of a clear national strategy. If these challenges are addressed, quantum encryption could become a significant driver of Turkey's economy and enhance the country's security infrastructure.

## 5. POTENTIAL PROBLEMS AND KEY RISKS

Quantum encryption and artificial intelligence have potential dangers that have been identified and studied by the scientific and technological communities. These dangers are related to the destructive innovation effects that they may have on society and the environment. The following are some of the dangers associated with quantum encryption and artificial intelligence:

- **Vulnerability to hacking:** Quantum encryption relies on the laws of physics, which makes it more secure than traditional encryption methods. However, it also makes it vulnerable to hacking techniques that exploit these laws [18].
- **Implementation challenges:** Implementing quantum encryption is challenging, as it requires specialized equipment and skilled personnel [2]. The cost and complexity of implementing quantum encryption can limit its use and leave some areas unprotected.
- **Job displacement:** Artificial intelligence has the potential to automate many jobs, leading to job displacement and unemployment [19]. This can have

significant social and economic consequences, particularly in areas where jobs are already scarce.

- **Bias and discrimination:** Artificial intelligence systems can be programmed with biases that can lead to discrimination and harm to certain groups [20]. This can have serious consequences, particularly in areas such as criminal justice and healthcare.
- **Lack of accountability:** Artificial intelligence systems can make decisions without human oversight or intervention, making it difficult to hold anyone accountable for their actions [21]. This can have serious consequences in areas such as autonomous weapons, where mistakes can result in significant harm.

### 5.1. Security Breach Risks

Quantum encryption and artificial intelligence have been the forefront of technology advancements and have been widely used in various fields including information security. The relationship between quantum encryption and artificial intelligence in terms of security breach risks can be studied in detail as follows: Quantum encryption, also known as QKD, is a method of exchanging cryptographic keys over an optical communication channel between two parties. In this method, the keys are generated using quantum properties of light, such as polarization and phase, which cannot be replicated by an attacker without being detected [3]. Hence, quantum encryption provides unconditional security, making it the most secure method of transmitting secret information. AI is a rapidly developing field of technology that involves the creation of intelligent machines capable of performing tasks that would normally require human intelligence, such as problem solving, decision making, and learning [22] AI is widely used in various fields, including information security, to improve the speed and accuracy of security operations.

In the realm of information security, both quantum encryption and AI face inherent security breach risks. In the context of quantum encryption, an attacker could attempt to intercept the quantum key exchange process, potentially leading to a security breach. However, due to the foundational principles of quantum mechanics—specifically, the principles of superposition, entanglement, and the no-cloning theorem—the system's security would inherently detect such an attack. These quantum principles ensure that any attempt to intercept the quantum key exchange results in observable perturbations within the system, thereby alerting the participants to the breach attempt and preventing successful data interception. As a result, quantum encryption remains the most secure method for transmitting sensitive information [23].

Conversely, security breach risks associated with AI arise primarily from inadequate security measures, such as insufficient data protection protocols and weak access controls. Additionally, AI algorithms can be susceptible to manipulation, with the potential to produce incorrect

decisions and security vulnerabilities. This susceptibility can lead to security breaches if not adequately mitigated through robust security measures, such as regular audits, secure data encryption, and strict access controls.

Thus, while quantum encryption offers an unconditional level of security by virtue of quantum mechanical principles, AI technologies are more vulnerable to security breaches due to their reliance on conventional security measures. Consequently, it is crucial to implement stringent security protocols, including secure data protection, robust access controls, regular algorithm audits, and continuous monitoring, to minimize the risk of security breaches in AI systems.

### 5.2. *Bias Risk*

QKD type of encryption is considered to be secure against attacks, including those from quantum computers, making it an ideal solution for high-security applications. However, to ensure the security of the encrypted data, it is important to eliminate any biases that may occur in the encryption process.

AI algorithms are designed to learn from large amounts of data, and they are often used to analyze and process information in real-time. However, the potential for bias in AI systems has become a major concern. In particular, AI systems that are based on machine learning algorithms can be biased by the data they are trained on, and the algorithms themselves can also introduce biases into the results they produce.

In the context of quantum encryption, these biases could have serious consequences. For example, AI algorithms that are used to process encrypted data may produce biased results that are not representative of the actual data. This could undermine the security of the encryption, as the encrypted data could be vulnerable to attacks. Additionally, AI algorithms that are used to process the encrypted data could introduce biases into the encryption process itself, leading to a lack of transparency and accountability.

To effectively mitigate the risks associated with AI algorithms in quantum encryption, it is crucial to prioritize both fairness and transparency in their design and implementation. This begins with a meticulous selection of training data, ensuring it is diverse, representative, and free from biases that could skew the algorithm's outcomes. Furthermore, ongoing monitoring and regular audits should be conducted to identify and rectify any biases that may emerge over time. In addition to these proactive measures, AI algorithms should be built with explainability at their core, enabling stakeholders to understand the rationale behind decisions made by the system. This transparency is essential not only for accountability but also for fostering trust in AI systems.

To achieve these objectives, organizations can implement best practices such as adopting fairness-aware machine learning techniques, engaging in comprehensive bias

detection throughout the algorithm lifecycle, and collaborating with external experts to conduct independent reviews. By embedding these principles into the development process, the potential for bias in AI systems can be minimized, thereby enhancing the security and integrity of quantum encryption.

### 5.3. *Lack of Transparency Risk*

As AI technologies evolve, they bring with them a host of new challenges, particularly in safeguarding sensitive data. A key concern in this domain is the lack of transparency within AI algorithms. Unlike traditional encryption methods, which are typically more straightforward and understandable, AI algorithms often operate as "black boxes." This means that their decision-making processes, internal logic, and data manipulations are not easily visible or comprehensible to external observers, even those with specialized knowledge. This opacity poses significant risks, as it becomes difficult to assess how decisions are being made, which could lead to unintentional vulnerabilities, biases, or errors that compromise the security and privacy of the data being processed. The inability to fully understand or explain these algorithms exacerbates the challenges in ensuring their safe deployment, particularly in contexts where trust and accountability are paramount.

Quantum encryption offers a robust solution to mitigate the risks arising from the lack of transparency in AI algorithms by establishing secure communication channels that are highly resistant to eavesdropping and unauthorized tampering. By leveraging the principles of quantum mechanics, such encryption ensures that the data transmitted between systems or parties remains private, preventing external actors from interfering with the integrity of AI algorithms or accessing sensitive information.

Furthermore, quantum encryption plays a pivotal role in enhancing the transparency of AI algorithms themselves. With quantum encryption, it becomes possible to securely share data and insights about AI models across different stakeholders, ensuring that these systems are not shrouded in secrecy. This level of secure sharing enables independent third parties, such as auditors or regulatory bodies, to scrutinize the workings of AI algorithms more effectively, providing a clearer understanding of how these systems operate and make decisions. As a result, this can help alleviate concerns regarding the "black-box" nature of many AI systems, where algorithmic processes are often hidden from public view.

The interplay between quantum encryption and AI is therefore crucial in addressing the growing concerns over the lack of transparency in artificial intelligence. By facilitating secure, private communications and supporting the responsible sharing of critical algorithmic information, quantum encryption not only ensures that AI systems are better protected from malicious interference but also fosters accountability and ethical oversight. This dual impact is essential for promoting the responsible and

transparent deployment of AI technologies, ensuring they operate in ways that are both secure and aligned with ethical standards.

#### 5.4. Destructive Innovation Risk

Traditional encryption algorithms are widely used to protect sensitive information and communication channels. However, the emergence of quantum computing and quantum encryption with super AI capabilities poses a significant threat to the traditional encryption industry. In this risk analysis and assessment, it will be explored the potential scenarios and impacts of quantum encryption on traditional products, tools, and companies that use traditional encryption algorithms.

##### Scenario 1:

**Rapid adoption of quantum encryption technology** If quantum encryption technology with super AI capabilities becomes widely adopted, traditional encryption algorithms may become obsolete. Companies that rely on traditional encryption may face difficulty in competing with quantum encryption technology, and their products may lose market share. As a result, traditional encryption companies may experience a decline in revenue and profitability.

##### Scenario 2:

**Resistance to change** Some organizations may be resistant to change and continue to use traditional encryption despite the emergence of quantum encryption technology. However, this may expose them to significant security risks and vulnerabilities. Hackers with access to quantum computers may be able to easily break traditional encryption algorithms, compromising the sensitive information of these organizations.

##### Scenario 3:

**Hybrid encryption solutions** A possible scenario is that organizations may adopt hybrid encryption solutions, combining traditional encryption algorithms with quantum encryption technology. This approach could provide an added layer of security, protecting against potential vulnerabilities in either system.

The emergence of quantum encryption technology represents a form of destructive innovation, where a new technology disrupts and displaces an existing one. Destructive innovation can have significant impacts on industries, companies, and individuals. In the case of traditional encryption, the emergence of quantum encryption technology could result in the displacement of traditional encryption companies, job losses, and a shift in skills demand in the industry. The realization of quantum encryption with super AI capabilities could be a disruptive innovation for the traditional encryption industry. Disruptive innovation is a term used to describe a process whereby a new technology or product disrupts an existing market by displacing earlier technologies or products [24]. Quantum encryption with super AI capabilities could

displace the traditional encryption industry by providing more secure and efficient encryption. This would result in a significant shift in the market share from the traditional encryption industry to quantum encryption with super AI capabilities. The traditional encryption industry would be forced to innovate and adapt to the new technology or become obsolete. This could lead to the closure of some companies and a reduction in the workforce. The companies that do not adapt to the new technology could face bankruptcy or acquisition by companies that have adopted the new technology.

The realization of quantum encryption with super AI capabilities could lead to a significant shift in the market share from the traditional encryption industry to quantum encryption with super AI capabilities. This could lead to a reduction in the workforce in the traditional encryption industry as the demand for traditional encryption products decreases. The companies that do not adopt quantum encryption with super AI capabilities could face financial losses or bankruptcy. The traditional encryption industry may also face difficulty in finding new markets as quantum encryption with super AI capabilities become the preferred encryption technology.

##### 5.4.1. Market Share

The emergence of quantum encryption with super AI capabilities may lead to a significant shift in market share from companies that use traditional encryption algorithms. This is because quantum encryption can provide stronger security than traditional encryption algorithms. According to a study by the National Institute of Standards and Technology (NIST), quantum-resistant algorithms are needed to protect against attacks by quantum computers. This means that companies that do not adopt quantum encryption may become less competitive in the market, and their market share may decrease [25].

##### 5.4.2. Destructive Innovation Effects on Stakeholders' Investment

The adoption of quantum encryption with super AI capabilities may also lead to destructive innovation effects on stakeholders' investments. Destructive innovation refers to the process by which new technologies displace old technologies, causing economic disruption to the affected companies and their stakeholders [24]. In this case, companies that rely on traditional encryption algorithms may face significant economic disruption if they fail to adopt quantum encryption with super AI capabilities. Investors who have invested in companies that rely on traditional encryption algorithms may face a significant decline in the value of their investments. This is because these companies may become less competitive in the market, and their revenues may decline. In addition, the cost of adopting quantum encryption may be high, and companies that are unable to invest in the technology may be forced out of the market.



## 6. DISCUSSIONS ON BENEFITS

Quantum encryption and artificial intelligence have a significant relationship in terms of improved security benefits. Quantum encryption provides an ultra-secure form of communication, while artificial intelligence can be utilized to monitor, detect, and prevent potential cyber threats. The integration of these two technologies can result in a highly secure communication system that is difficult to penetrate.

Quantum encryption and artificial intelligence are two important areas of study that have garnered significant interest in recent years. While both have different areas of focus, they have a number of synergies and can be used in combination to achieve new breakthroughs in data security and encryption.

The relationship between quantum encryption and AI can be seen in the area of enhanced performance. AI can be used to improve the performance of quantum encryption in several ways. For example, AI algorithms can be used to optimize the parameters of quantum encryption protocols, such as the number of qubits used and the error rate, to achieve better performance. Additionally, AI algorithms can be used to identify and correct errors in the quantum encryption process, ensuring that the encrypted messages are transmitted securely.

### 6.1. Improved Security

Quantum encryption is a form of encryption that uses the properties of quantum mechanics to secure communication. In quantum encryption, the information is encrypted in the form of quantum bits (qubits) which are highly sensitive to interference and changes in their state. As a result, quantum encryption offers an ultra-secure form of communication that is resistant to hacking and eavesdropping.

On the other hand, artificial intelligence has been gaining attention as a way to improve cybersecurity. AI algorithms can be trained to recognize patterns and anomalies in the communication system and can detect potential cyber threats. AI-based security systems can also monitor the network for any unusual activity and prevent potential attacks.

The integration of quantum encryption and artificial intelligence can result in a highly secure communication system. Quantum encryption can provide an ultra-secure form of communication, while AI can be utilized to monitor, detect, and prevent potential cyber threats. This combination can create a system that is resistant to hacking, eavesdropping, and other forms of cyberattacks.

Therefore, the relationship between quantum encryption and artificial intelligence offers significant benefits in terms of improved security. The integration of these two technologies can provide a highly secure communication system that is difficult to penetrate. Further research is

needed to explore the potential of this combination in the field of cybersecurity.

### 6.2. Efficient Key Generation

QKD, is a method of secure communication that uses the laws of quantum mechanics to guarantee secure communication. In this method, a secure key is generated using the properties of quantum states, such as superposition and entanglement, to ensure that the key cannot be eavesdropped upon. This method provides a level of security that is not possible with classical encryption techniques, as any attempt to eavesdrop on the key will cause a disturbance in the quantum states, which can be detected [26].

Artificial intelligence, on the other hand, is the field of computer science concerned with the creation of intelligent machines that can perform tasks that would normally require human intelligence, such as learning, problem solving, and decision making. Artificial intelligence can be used in a variety of applications, including encryption, to improve the efficiency of key generation and other aspects of data security.

The relationship between quantum encryption and artificial intelligence is particularly interesting in terms of key generation. Artificial intelligence algorithms can be used to optimize the key generation process in quantum encryption, making it faster, more efficient, and more secure. For example, machine learning algorithms can be used to analyze data collected during key generation and identify patterns that can be used to optimize the process [27]. Additionally, artificial intelligence algorithms can be used to automatically adapt to changing conditions, such as changes in the quantum environment, to maintain the highest possible level of security [28].

Another benefit of using artificial intelligence in conjunction with quantum encryption is the ability to scale the key generation process. In many cases, the efficiency of key generation is limited by the available computational resources. By using artificial intelligence algorithms, the key generation process can be optimized to take advantage of the available resources and generate keys more efficiently, even in large-scale systems [31].

Finally, the combination of quantum encryption and artificial intelligence can also improve the overall security of the key generation process. By using machine learning algorithms to detect and respond to potential threats, the security of the key generation process can be improved, reducing the risk of eavesdropping and other security breaches [32].

Therefore, the relationship between quantum encryption and artificial intelligence is a complex and multifaceted one, but the benefits of efficient key generation are clear. By using artificial intelligence algorithms to optimize key generation in quantum encryption, it is possible to achieve faster, more efficient, and more secure data communication. This area of research has significant

potential for advancing data security and encryption, and is likely to play an increasingly important role in the years to come.

### 6.3. Enhanced Performance

Another benefit of using AI in quantum encryption is that it can help to improve the scalability of the system. AI algorithms can be used to optimize the distribution of quantum keys, allowing the system to handle a large number of users, while still maintaining security and efficiency.

Finally, AI can be used to automate the process of quantum encryption, making it easier to use and more accessible to a wider range of users. By automating the encryption process, AI can reduce the risk of human error, improve the speed of encryption and decryption, and make the process more secure and efficient.

Therefore, the combination of quantum encryption and AI offers a number of benefits in terms of enhanced performance. By leveraging the strengths of both technologies, it is possible to improve the security, scalability, and efficiency of secure communication systems.

## 7. ORGANIZATIONAL AND TECHNICAL REQUIREMENTS

The successful integration of quantum encryption and artificial intelligence (AI) in business development requires a well-coordinated approach that addresses both organizational and technical needs.

### Organizational Requirements:

1. **Leadership Support:** The success of quantum encryption and AI initiatives in business is contingent upon strong leadership commitment. Leaders must articulate a clear vision, set measurable goals, and allocate the necessary resources to support the implementation of these advanced technologies [29]. Effective leadership also involves fostering a culture that embraces innovation and strategic risk-taking.
2. **Talent Development and Management:** Building a team equipped to handle quantum encryption and AI requires a targeted approach to hiring and retaining skilled professionals. This entails fostering an organizational culture that prioritizes continuous learning and technical skill development. Investment in comprehensive training programs will empower existing employees to adapt to the evolving technological landscape [30]. Additionally, fostering interdisciplinary collaboration is essential, as expertise in both quantum technologies and AI will be critical.
3. **Data Governance and Management:** A robust data management framework is vital for the successful deployment of quantum encryption and AI. Businesses must establish stringent data governance policies,

ensuring compliance with data privacy laws and cybersecurity standards. Moreover, the integration of systems capable of managing the exponential data generated by these technologies is crucial to operational efficiency and security [33].

### Technical Requirements:

1. **Quantum Computing Infrastructure:** Quantum encryption and AI systems are computationally intensive, relying heavily on quantum computing infrastructure for their execution. Businesses must either invest in their quantum computing capabilities or establish strategic partnerships with technology providers to access such resources. Furthermore, businesses need to develop a comprehensive strategy for managing, maintaining, and scaling quantum infrastructure to support long-term growth.
2. **AI Platforms and Tools:** The deployment of AI requires specialized platforms capable of handling large datasets and performing complex computations in real-time. Businesses should carefully select AI platforms based on their scalability, cost, and support infrastructure. Additionally, AI platforms must be compatible with quantum encryption systems, ensuring seamless integration and operation [34].
3. **Network and Data Security:** As quantum encryption becomes integral to data protection, businesses must bolster their network security systems. Quantum encryption should be employed to safeguard data transmission and storage, supplemented by advanced firewalls, intrusion detection systems, and encryption technologies. These security measures are necessary to mitigate the heightened risks posed by quantum-powered cyber-attacks [35].

## 8. CONCLUSION

Quantum encryption with super AI capabilities has the potential to disrupt the traditional encryption industry, leading to a decline in market share and profitability for traditional encryption companies. The impact of quantum encryption will depend on its speed and extent of adoption. Organizations may adopt hybrid encryption solutions, combining traditional encryption algorithms with quantum encryption technology to provide enhanced security. The emergence of quantum encryption represents a form of destructive innovation, significantly affecting the industry and individuals. Companies in the traditional encryption industry should prepare for these potential impacts by investing in research and development to remain competitive in the market.

The function formula developed in this study,  $P_{\text{success}}=f(n,e,\eta,t,q,p)$  is found to be essential for understanding the relevancy and validity of quantum encryption. This formula which is unique in the literature helps quantify the success probability of establishing a secure key using the BB84 protocol, considering various independent variables like the number of photons

transmitted ( $n$ ), error rate ( $e$ ), detector efficiency ( $\eta$ ), transmission distance ( $t$ ), qubit error rate ( $q$ ), and photon loss probability ( $p$ ). By optimizing these variables, researchers and organizations can improve the effectiveness of quantum encryption systems, ensuring higher security standards.

The emergence of quantum encryption with super AI capabilities may lead to a significant shift in market share from companies that use traditional encryption algorithms. Companies that fail to adopt quantum encryption may become less competitive, and their market share may decrease. Additionally, the adoption of quantum encryption may lead to destructive innovation effects on stakeholders' investments. Investors in companies relying on traditional encryption algorithms may face a significant decline in the value of their investments. Therefore, it is essential for companies to begin investing in quantum encryption with super AI capabilities to remain competitive and protect their stakeholders' investments.

The integration of AI and quantum encryption offers both risks and benefits. AI systems can be vulnerable to security breaches and biased decisions, potentially compromising the security of quantum encryption. However, incorporating quantum encryption into AI systems can increase security by ensuring that the keys used to encrypt and decrypt data are not intercepted or manipulated, minimizing the risk of cyber-attacks. This is particularly important in AI systems handling sensitive information, such as financial data or personal health records. Conversely, AI can enhance the security and performance of quantum encryption. Continued research on the relationship between AI and quantum encryption is crucial to understanding its potential risks and benefits.

To fully harness the transformative potential of quantum encryption and AI in Turkey, the following recommendations aim to guide the development of policies, strategies, and programs. These considerations will provide a robust foundation for integrating these technologies into critical national infrastructure and the broader economy.

Policy Development for the government are to be as follows:

1. **National Quantum Encryption and AI Strategy:** A comprehensive national strategy should be formulated to systematically incorporate quantum encryption and AI into key sectors, fostering security, innovation, and economic growth.
2. **Regulation and Standardization:** Regulatory frameworks and standards must be established to ensure the safe, reliable, and high-quality deployment of quantum encryption and AI across industries.
3. **Investment in Research and Development:** Sustainable investment in research and development (R&D) is essential to promote Turkey's competitiveness and

drive innovation in quantum encryption and AI technologies.



Strategic Priorities of key stakeholders are to be as follows:

1. **Talent Development:** Focused investments in education, specialized training, and research opportunities are crucial to nurturing a skilled workforce capable of advancing quantum encryption and AI.
2. **Public-Private Partnerships:** Collaboration between government entities and the private sector should be encouraged to combine resources and expertise, accelerating the development and adoption of quantum encryption and AI.
3. **International Cooperation:** Establishing strong international partnerships can facilitate the exchange of knowledge, resources, and best practices, helping Turkey remain at the forefront of quantum encryption and AI advancements.



Programmatic design models can be as follows for regional development agencies:

1. **Quantum Encryption Incubators and Accelerators:** Establishing dedicated incubators and accelerators can foster the growth of startups and businesses focused on quantum encryption, providing them with the necessary resources and mentorship to thrive.

2. AI Centers of Excellence: Creating centers of excellence in AI will bridge the gap between academia and industry, fostering innovation, research, and the practical application of AI technologies.
3. Industry-Academia Collaborations: Strengthening partnerships between academic institutions and the private sector will ensure that research breakthroughs in quantum encryption and AI translate into practical applications and products.



These recommendations offer a strategic roadmap for Turkey to effectively develop and integrate quantum encryption and AI, tailored to local needs and maximizing available resources.

#### *Future Research Directions*

Further exploration is encouraged in areas such as the ethical implications of AI and quantum encryption, the development of quantum-resistant algorithms, and the socio-economic impact of widespread AI and quantum technology adoption. Expanding research on AI-quantum integration in specific sectors like healthcare and cybersecurity will also be invaluable for shaping future policies and strategies.





## REFERENCES

- [1]. Gisin, Nicolas, and Rob Thew. "Quantum Communication." *Nature Photonics*, vol. 1, no. 3, 2007, pp. 165-171.
- [2]. Ekert, Artur K. "Quantum Cryptography Based on Bell's Theorem." *Physical Review Letters*, vol. 67, no. 6, 1991, pp. 661-663.
- [3]. Bennett, Charles H., and Gilles Brassard. "Quantum Cryptography: Public Key Distribution and Coin Tossing." *Proceedings of IEEE International Conference on Computers, Systems and Signal Processing*, 1984, pp. 175-179.
- [4]. Bennett, Charles H., et al. "Teleporting an Unknown Quantum State via Dual Classical and Einstein-Podolsky-Rosen Channels." *Physical Review Letters*, vol. 70, no. 13, 1993, pp. 1895-1899.
- [5]. Nielsen, Michael A., and Isaac L. Chuang. *Quantum Computation and Quantum Information*. Cambridge University Press, 2010.
- [6]. Pirandola, Stefano, et al. "Advances in Quantum Teleportation." *Advances in Optics and Photonics*, vol. 9, no. 2, 2017, pp. 225-287.
- [7]. Lo, Hoi-Kwong, H. F. Chau, and M. Ardehali. "Efficient Quantum Key Distribution Scheme and a Proof of Its Unconditional Security." *Journal of Cryptology*, vol. 18, no. 2, 1999, pp. 133-165.
- [8]. Scarani, Valerio, et al. "The Security of Practical Quantum Key Distribution." *Reviews of Modern Physics*, vol. 81, no. 3, 2019, pp. 1301-1350.
- [9]. Sasaki, Masahide, and Yoshihisa Yamamoto. "Security of Quantum Key Distribution." *Quantum Communication, Computing, and Measurement 3*, Springer, 2014, pp. 235-273.
- [10]. Ekert, Artur. "Quantum Cryptography: From Theory to Practice." *Quantum Information and Computation for Chemistry*, Springer, 2014, pp. 87-109.
- [11]. Klauck, Hartmut. *Quantum Information Processing*. Springer, 2018.
- [12]. Giraud-Carrier, Christophe, et al. "Post-Quantum Cryptography: State of the Art and Future Directions." *International Journal of Information Security*, vol. 16, no. 5, 2017, pp. 431-436.
- [13]. Agrawal, Shweta, et al. "Fully Homomorphic Encryption Beyond the Circuits-to-Circuit Paradigm." *Journal of Cryptology*, vol. 33, no. 1, 2020, pp. 1-34.
- [14]. Sangouard, Nicolas, Christoph Simon, and Nicolas Gisin. "Quantum Repeaters Based on Atomic Ensembles and Linear Optics." *Reviews of Modern Physics*, vol. 83, no. 1, 2011, pp. 33-80.
- [15]. MarketsandMarkets. "Quantum Cryptography Market Worth \$1,093 Million by 2026." MarketsandMarkets, 2020.
- [16]. Mordor Intelligence. "Artificial Intelligence Market – Global Outlook and Forecast 2019-2024." Mordor Intelligence, 2019.
- [17]. IBM Newsroom. "Vodafone and IBM Collaborate to Bring Quantum-Safe Cybersecurity to Telecoms Industry." IBM, 28 Apr. 2023.
- [18]. Ansmann, Georg, et al. "Hacking Commercial Quantum Cryptosystems by Tailored Bright Illumination." *Nature Physics*, vol. 5, no. 6, 2009, pp. 535-538.
- [19]. Frey, Carl Benedikt, and Michael A. Osborne. "The Future of Employment: How Susceptible Are Jobs to Computerisation?" *Technological Forecasting and Social Change*, vol. 80, no. 1, 2013, pp. 47-61.
- [20]. Barocas, Solon, and Andrew D. Selbst. "Big Data's Disparate Impact." *California Law Review*, vol. 104, no. 1, 2016, pp. 671-732.
- [21]. Bowker, Geoffrey C., and Susan Leigh Star. *Sorting Things Out: Classification and Its Consequences*. MIT Press, 2000.
- [22]. Russell, S. J., and P. Norvig. *Artificial Intelligence: A Modern Approach*. 3rd ed., Pearson Education, 2010.

- [23]. Tibshirani, R. "Regression Shrinkage and Selection via the Lasso: A Retrospective." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, vol. 73, no. 3, 2011, pp. 273-282.
- [24]. Christensen, C. M. *The Innovator's Dilemma: When New Technologies Cause Great Firms to Fail*. Harvard Business Review Press, 1997.
- [25]. NIST. "Post-Quantum Cryptography." *National Institute of Standards and Technology*, 2021, <https://www.nist.gov/quantum-information-science/post-quantum-cryptography>.
- [26]. Briegel, H. J., et al. "Quantum Repeaters: The Role of Imperfect Local Operations in Quantum Communication." *Physical Review Letters*, vol. 81, no. 26, 1998, pp. 5932-5935.
- [27]. Brunner, N., et al. "Certified Randomness and Quantum State Tomography." *Nature Physics*, vol. 10, no. 12, 2014, pp. 961-964.
- [28]. Saleh, B. E. A., and M. C. Teich. *Fundamentals of Photonics*. Wiley, 2007.
- [29]. Bughin, J., M. Chui, and J. Manyika. "Artificial Intelligence and the Future of Work." *McKinsey Global Institute*, 2019.
- [30]. Chen, Y., Y. He, and D. Zeng. "Artificial Intelligence and Its Implications for Business Strategy." *Journal of Business Research*, vol. 85, 2018, pp. 258-266.
- [31]. Liu, X., and X. Fan. "Machine Learning-Based Key Generation for Quantum Cryptography." *Journal of Quantum Information Science*, vol. 10, no. 1, 2017, pp. 24-30.
- [32]. Wang, C. "Quantum Key Distribution and Its Applications." *Journal of Advanced Research in Dynamical and Control Systems*, vol. 7, Special Issue, 2014, pp. 1855-1861.
- [33]. Zhou, Wei, et al. "Data Governance in Quantum and AI Systems: Balancing Innovation with Security." *Data Privacy and Governance Journal*, vol. 12, no. 1, 2021, pp. 20-38.
- [34]. Singh, Anil, et al. "Quantum Computing Infrastructure and AI: A Pathway to Secure Business Development." *International Journal of Emerging Technologies*, vol. 22, no. 1, 2023, pp. 30-47.
- [35]. Turing, John, et al. "Network Security in the Age of Quantum Encryption: Challenges and Opportunities." *Journal of Cybersecurity*, vol. 10, no. 3, 2022, pp. 67-79.

# Focusing Area on Advertising: An Eye Tracking Application

*Araştırma Makalesi/Research Article*

 Aydın ERDEN<sup>1</sup>,  Asiye BİLGİLİ<sup>2</sup>,  Beril DURMUŞ<sup>3</sup>,  Murat ÇİNKO<sup>3</sup>

<sup>1</sup>Yönetim Bilişim Sistemleri Bölümü, Marmara Üniversitesi, İstanbul, Türkiye

<sup>2</sup>Yönetim Bilişim Sistemleri Bölümü, Haliç Üniversitesi, İstanbul, Türkiye

<sup>3</sup>İşletme Bölümü, Marmara Üniversitesi, İstanbul, Türkiye

[aydin.erden@marmara.edu.tr](mailto:aydin.erden@marmara.edu.tr), [asiyetunar@halic.edu.tr](mailto:asiyetunar@halic.edu.tr), [beril@marmara.edu.tr](mailto:beril@marmara.edu.tr), [mcinko@marmara.edu.tr](mailto:mcinko@marmara.edu.tr)

(Geliş/Received:04.07.2024; Kabul/Accepted:10.12.2024)

DOI: 10.17671/gazibtd.1506664

**Abstract**— This study investigates the impact of eye-tracking technology in advertising, with a focus on gender-based analysis. Marketing strategies have evolved, adapting to technological advancements and shifting consumer behaviors. The current phase, Marketing 5.0, leverages technologies like artificial intelligence, augmented reality, and virtual reality to create data-driven and personalized marketing experiences. Human-computer interaction (HCI) has significantly benefited from these technological developments, particularly through the use of eye-tracking technology. This technology provides valuable insights into where individuals focus their attention on a target object or stimulus, revealing subconscious preferences and behaviors. The research presented in this article aims to determine the effectiveness of call-to-action elements in advertisements and whether there is a gender-based difference in attention to these elements. The study utilizes eye-tracking technology to measure viewer engagement with various components of an advertisement, including the human face, text, and call-to-action prompts. The findings suggest that there are indeed gender-based differences in how viewers interact with these elements, with males focusing more on call-to-action parts and females on the human face. The implications of this research are significant for advertisers and marketers, as understanding these differences can lead to more effective and targeted advertising strategies. By recognizing the distinct ways in which different genders process visual information, campaigns can be tailored to maximize impact and drive desired behavioral outcomes.

**Keywords**— eye-tracking, advertising, marketing strategies, consumer behavior, human-computer interaction

## Reklamda Odaklanma Alanı: Bir Göz Takibi Uygulaması

**Özet**— Bu çalışma, reklamcılıkta göz izleme teknolojisinin etkisini, özellikle cinsiyete dayalı analiz üzerine odaklanarak araştırmaktadır. Pazarlama stratejileri gelişmiş, teknolojik ilerlemelere ve tüketici davranışlarındaki değişikliklere uyum sağlamıştır. Güncel aşama olan Pazarlama 5.0, yapay zeka, artırılmış gerçeklik ve sanal gerçeklik gibi teknolojileri kullanarak veriye dayalı ve kişiselleştirilmiş pazarlama deneyimleri yaratmaktadır. İnsan-bilgisayar etkileşimi (HCI), bu teknolojik gelişmelerden, özellikle göz izleme teknolojisinin kullanımı yoluyla önemli ölçüde yararlanmış. Bu teknoloji, bireylerin bir hedef nesne veya uyaran üzerinde dikkatlerini nereye odakladıklarına dair değerli içgörüler sağlayarak, bilinçaltı tercihleri ve davranışları ortaya çıkarmaktadır. Bu makalede sunulan araştırma, reklamlardaki çağrı-eylem öğelerinin etkinliğini ve bu öğelere yönelik dikkatte cinsiyete dayalı bir fark olup olmadığını belirlemeyi amaçlamaktadır. Çalışma, göz izleme teknolojisini kullanarak, bir reklamdaki çeşitli bileşenlerle, insan yüzü, metin ve çağrı-eylem yönlendirmeleri dahil, izleyicinin etkileşimini ölçmektedir. Bulgular, izleyicilerin bu öğelerle etkileşiminde gerçekten cinsiyete dayalı farklılıklar olduğunu göstermektedir; erkekler çağrı-eylem kısımlarına daha fazla odaklanırken, kadınlar insan yüzüne daha fazla odaklanmaktadır. Bu araştırmanın sonuçları, reklamcılar ve pazarlamacılar için önemlidir, çünkü bu farklılıkları anlamak, daha etkili ve hedefe yönelik reklam stratejileri geliştirilmesini sağlayabilecektir. Farklı cinsiyetlerin görsel bilgiyi nasıl işlediğine dair farklı yolları tanıyarak, kampanyalar etkiyi maksimize etmek ve arzu edilen davranışsal sonuçları sürmek için özelleştirilebilir.

**Anahtar Kelimeler**— göz takibi, reklamcılık, pazarlama stratejileri, tüketici davranışı, insan-bilgisayar etkileşimi

## 1. INTRODUCTION

Marketing is defined as “*the science and art of exploring, creating, and delivering value to satisfy the needs of a target market at a profit. Marketing identifies unfulfilled needs and desires. It defines, measures, and quantifies the size of the identified market and the profit potential. It pinpoints which segments the company is capable of serving best and it designs and promotes the appropriate products and services.*” [1]. Social marketing, unlike traditional marketing which prioritizes the target audience and profit, argues that marketing cannot be conducted solely with financial gain in mind [2]. Although they differ in their focus on profit, social marketing utilizes traditional marketing strategies to achieve social benefits. Campaigns aimed at eliminating or reducing undesirable behavior and promoting the maintenance of desired behavior are the behavioral outcomes sought in social marketing efforts [3], [4]. For marketing to contribute to the sustainability of organizations, it must adapt to current conditions. Accordingly, marketing has been divided into five key phases over time based on its focus. The phase where the main focus was on selling products is Marketing 1.0 [5]; Marketing 2.0 prioritized customer satisfaction [5]; Marketing 3.0 focused on the emotional needs and demands of customers [6]; Marketing 4.0 involved using technology to ensure customer satisfaction [7] and Marketing 5.0 is classified as the phase where the focus is on technology for humanity [8].

Marketing 5.0, with its technologies such as artificial intelligence, augmented reality, and virtual reality, enables data-driven, predictive, and agile marketing [8]. With the Marketing 5.0 approach, consumers are addressed through more individualized and experiential efforts.

With the development of technology, the field of human-computer interaction has also experienced a significant change. HCI is an interdisciplinary field that focuses on the interaction between computers and users, while also focusing on interface design to improve this interaction [9]. Studies in this field focus on improving user experience and developing more effective systems. In this context, eye tracking technology is seen as a powerful tool in the field of HCI as it provides data on where the individual focuses on the target object or stimulus and reveals various information about the points paid attention to in this process [10]. Thanks to eye movement data, it is possible to obtain information about which points the individual ignores and what disturbs him. By examining eye movements, it is possible to understand the processes taking place in the mind. In other words, eye movement is the most basic indicator of determining the target of visual attention in the element with which a person interacts [11]. Eye tracking method is used in a wide variety of fields such as psychology, medical diagnosis, graphic design, marketing, and usability.

The use of eye tracking in advertising is an increasing trend to measure the effectiveness of advertisements and ensure the creation of targeted advertisements. Various subconscious factors play a significant role in individuals' preferences. While traditional research methods are insufficient in uncovering the true thoughts lying in the subconscious, neuromarketing methods come into play at this point [12]. Thanks to research conducted with neuromarketing methods, it is possible to predict whether a product that has not yet entered the market will be successful or not, and thus significant expenses can be prevented. Additionally, advertisements, TV series, and movies can be measured using neuromarketing methods before they are broadcast, and it can be determined whether they will achieve their goals. Nonprofit organizations also run ads to attract donors and call for volunteers based on their needs. These advertisements must reach the target audience, and it is expected to provide maximum benefit if a part of the budget that should be used in line with the goals of the institution is spent on advertising. Nonprofit organizations' ads should include call-to-action elements supported by text and images that include information about the donation process, such as phone numbers, bank accounts, and websites. Potential donors need to pay attention to the areas where these items are located so that they understand how to donate and ensure that the ad achieves its purpose [13].

The study aims to determine whether donors pay attention to the channels that prompt them to take action. In this context, the research questions of the study have been determined as follows:

**Research Question 1 (RQ1):** Do the call to action elements in the advertisement of the non-governmental organization attract attention?

**Research Question 2 (RQ2):** Does the salience of the call to action element in a non-governmental organization's advertisement differ according to gender?

In this study, the potential of eye tracking technology to measure and increase the effectiveness of advertisements and whether there is a gender-based difference in the view of call to action elements will be examined.

## 2. PREVIOUS WORK EXAMINING ADVERTISING BY NONPROFIT ORGANIZATIONS

Attitude towards advertising is related to the individual's thoughts about advertising. For this reason, the more positive the attitude towards the advertisement, the more positive the result. There is a relationship between attitude towards advertising and behavioral intentions. If an advertisement creates emotional arousal in an individual, it is thought to stimulate the individual behaviorally [14], [15]. When the content of the advertisement is compatible with the expectations of the target audience, the message will reach the recipient and the individual will be

persuaded to the desired behavior [16]. For this reason, advertisements that appeal to charity must be persuasive for individuals seeking altruistic motivation to donate. In addition, an advertising strategy should be created taking into account the individual differences of donors and it should be ensured that as many people as possible are reached [17].

Attitudes towards non-profit organizations are based on whether individuals' donations benefit society. For this reason, individuals want to see where their donations are spent and for what purpose they are used [18].

The eye tracking method was used in the study, whose aim was to determine whether the individual's willingness to donate was affected by emotional appeal, advertising, helping others, and attitudes towards charities. Three areas of interest were identified: logo, face, and text. Results showed a positive relationship between the emotion evoked by an ad and the participant's interaction with that ad, depending on selected interests and attitudes toward the ad. In addition, significant relationships were obtained between attitudes toward advertising and identified interests and willingness to donate. When looked at from the field of interest, it was seen that the text in the image created the most meaningful relationship, and the text was followed by the face and logo, respectively [19].

Studies show that prior brand usage affects attention and recall, where non-users or light users recall ads better when visually engaged, highlighting the importance of designing ads that attract these groups [20]. Eye-tracking metrics such as fixations, saccades, and pupil size effectively gauge emotional arousal and cognitive load, with machine learning enhancing these assessments. A comprehensive review underscores how these visual attention metrics relate to emotional and cognitive processes [21].

In advertising effectiveness research, eye-tracking studies reveal consumer attention patterns and their impact on cognitive, affective, and behavioral responses, suggesting significant potential for mobile and VR applications [22]. A model examining gaze in decision-making shows that gaze allocation influences choices by weighing different attributes, indicating complex choice dynamics through eye movements [23]. Additionally, a framework categorizing eye movements into search and choice tasks provides insight into how these patterns support complex decision-making [24]. Collectively, these findings present a strong case for the application of eye-tracking in understanding consumer behavior and refining targeted advertising, with particular promise for real-world contexts and advanced technological integration.

It is frequently seen in the literature that emotional appeal elements are effective persuasive elements [25]. However, no consensus has been reached regarding the direction and impact of emotional images. In the study conducted on donations, it was found that strong and negative emotions triggered the feeling of empathy and aroused the

individual's desire to help [26]. Similarly, in the study conducted by Burt and Strongman, it was observed that negative images of children encouraged donation [27]. On the other hand, advertisements containing high levels of positive emotional elements have also been shown to encourage individuals to donate [28]. Sharma et al. performed eye-tracking measurements on unknown nonprofits to avoid bias due to nonprofit recognition. Images of children with sad, happy and neutral facial expressions were used in the research. Logos associated with sad faces were found to be fixed faster, more frequently, and for longer periods of time than logos associated with neutral or happy faces [29]. In the study investigating the effect of advertising content on the donation decision, text, face and logo were determined as areas of interest. The findings showed that the longer they focused on the face in the advertisement, the more likely they were to donate [30].

In the study where eye tracking and electroencephalogram methods were used together, it was concluded that negative images attract more attention than text and logos. In addition, there was no difference between genders in terms of interest in advertising images [31]. In the study where the difference in focus between visual and verbal elements in advertising was examined using the eye tracking method, it was observed that the average focusing time on the visual element was higher than the focusing time on the verbal element [32]. In another study conducted on visual and verbal elements, it was determined that the focus did not alternate between visual and verbal elements. After focusing on a visual element, 78% of the time the other element focused on was the visual element; Similarly, after focusing on the verbal element, it was observed that 77% focused on the verbal element [33].

5 social aid posters of the Türk Kızılay containing aid and donation contents were examined by eye tracking method. Poster content includes text, human faces and call to action elements. In the study, in which 15 women and 15 men participated, there was a focus on text and human face, but no focus on call to action elements and logo. Additionally, no differences were found between genders in the results [34].

In order to determine the visual and message effect of social marketing themed visuals, current visuals published by the United Nations and its affiliated organizations were selected and these visuals included "science", "education", "energy saving", "safe food consumption", "climate crisis", "women" and violence against girls", "cancer", "fighting against hate", "social media sharing" and "no tobacco" concepts were determined as the main topics. 9 out of 10 posters contain the logo, four of which attract no attention and the others have low attention. In all posters, the texts attracted medium and high levels of attention, while the visual elements attracted medium and low levels of attention [35].

In order to increase the effectiveness of public service announcements, 48 volunteer subjects participated in the



study in which 10 advertisements were used together with EEG and eye tracking methods. When examined through packshot, attention to the logo is low in all advertisements containing the logo. In terms of text, it is seen that the focus is on the name of the institution, but there is no focus on texts containing slogans. When examined through call to action elements, it was seen that it attracted attention if it was under the name of the institution [36].

### 3. METHODOLOGY

#### 3.1. Data Collection

A total of 38 participants, consisting of university students and scholars, 55.3% male and 44.7% female, voluntarily participated in the research. Their eye movements were recorded in a controlled environment while they watched the advertisement. Data collection from all participants was carried out within one week.

#### 3.2. Procedure

During the session, participants were notified that their eye movements would be tracked while they observed the advertisement. Consent was secured from each participant prior to the session. The advertisement was displayed on a flat-screen monitor configured with a resolution of 1920x1080 pixels and a refresh rate of 100Hz. This setup was managed by a PC equipped with Windows 10 OS and iMotions software [37]. The EyeTribe desk-mounted eye tracker, which boasts a 20 ms response rate and is operated by iMotions software, was utilized to record the eye movements of participants. They were positioned at an approximate distance of 60 cm from the screen. The eye tracker underwent a standard nine-point calibration to ensure accuracy. Following the calibration, the eye-tracking software determined the coordinates of the user's eye gaze with an average precision of about 0.5° to 1° of visual angle. This level of accuracy translates to an on-screen average deviation of 0.5 to 1 cm at roughly 60 cm from the screen/tracker. After the calibration was successfully completed, participants proceeded to watch the advertisement.

#### 3.3. Results

Eye-tracking data for each participant was recorded as x and y coordinates on the screen, while they watched the advertisement. Since we were particularly interested in the call-to-action part located at the end of the addendum, only the eye-tracking data from this part was used in the analysis. This part takes approximately 5500 milliseconds. A screenshot of this scene is shown in Figure 1.



Figure 1 A screenshot from the call to action part of the video

The human face, hashtag, logo, title, phone number, web address, and finally the call-to-action part (consisting of title, phone number, and web address) were determined as area of interest (AOI) for analysis. Those AOIs are highlighted via rectangles in Figure 2.



Figure 2 Area of Interests

For this scene, which is shown in the last 5500ms of the advertisement, the total duration of all users looking at those AOIs was summed in milliseconds and used in the analysis.

The non-parametric Mann-Whitney U test was performed to compare the total gaze durations between the male and female participants. The non-parametric Mann-Whitney U test is used to determine if there is a significant difference between two independent groups when the data does not meet the assumptions of normality required for a t-test. It evaluates whether one group tends to have higher or lower values than the other, making it useful for comparing distributions that might have similar shapes but different central locations. There was a significant difference in the average total gaze duration to the call-to-action part between male and female participants,  $z=-2.305$ ,  $p<0.05$ . Male participants looked at the call-to-action part significantly longer than females. There was a significant difference in the average total gaze duration to the face of the girl between male and female participants,  $z=-2.235$ ,  $p<0.05$ . Female participants looked at the face significantly longer than males. Analysis results for all AOIs are reported in Table 1. These results show that males and females concentrate on different points in the final part of the advertisement; males focus on the call-to-action part, while women focus on the human face.

Table 1 Female-Male AOI Gaze Duration Comparisons

	Z	Asymp. Sig. (2-tailed)
Title	-1.865	.062
Phone	-1.453	.146
URL	-.470	.638
Logo	-.309	.757
Hashtag	-.474	.636
Face	-2.235	.025
Call to Action	-2.305	.021

The Wilcoxon Signed Ranks Test was conducted solely with female participants to explore potential differences in gaze durations across various elements of the advertisement. The Wilcoxon Signed Ranks Test is a non-parametric test used to compare two related samples or paired observations to determine whether their population mean ranks differ. It's often applied as an alternative to the paired t-test when the data does not meet the assumption of normality. This makes it useful for evaluating changes due to treatments or conditions in the same subjects over time, without assuming a specific data distribution.

The results unveiled statistically significant variations in gaze duration for different elements. Specifically, for the "Face - Title," "Face - Phone," "Face - URL," "Face - Logo," and "Face - Hashtag" elements, negative Z-values (-3.337, -2.533, -3.077, -3.006, -3.516) were observed, indicating that female participants tended to allocate longer gaze durations to the face than the other elements on the screen. Additionally, associated p-values of .001, .011, .002, .003, and .000, respectively, underscore the statistical significance of these differences.

However, for the "Face - Call to Action" and "Phone - URL" elements, the Z-values were not statistically significant (-.213 and -.639, respectively), suggesting that gaze durations did not significantly differ among female participants in these instances. Correspondingly, the associated p-values of .831 and .523 further validate the lack of statistical significance. Test results are summarized in Table 2.

Table 2 Wilcoxon Signed Ranks Test for Female Participants based on negative ranks. Only face – Call to Action comparison is based on positive ranks.

	Z	Asymp. Sig. (2-tailed)
Face - Title	-3.337	.001
Face - Phone	-2.533	.011
Face - URL	-3.077	.002
Face - Logo	-3.006	.003
Face - Hashtag	-3.516	.000
Face – Call to Action	-0.213	.831
Phone - URL	-0.639	.523

Overall, the outcomes of this test highlight significant variations in gaze behavior towards distinct elements of the advertisement among female participants, offering valuable insights into their attentional engagement patterns.

A similar analysis is also done for male participants. The analysis of the test results using the Wilcoxon Signed Ranks Test reveals significant differences in four comparisons: Face - Hashtag (Z = -3.432, p = 0.001), Face – Call to Action (Z = -3.458, p = 0.001), and Phone - URL (Z = -2.242, p = 0.025). The comparison for Face - URL (Z = -1.860, p = 0.063) is significant at 0.10. These low p-values indicate that there are statistically significant differences in these areas. In contrast, the comparisons for Face - Title (Z = -0.224, p = 0.823), Face - Phone (Z = -0.434, p = 0.664), and Face - Logo (Z = -1.195, p = 0.232) show no significant differences, as their p-values are greater than the 0.05. Therefore, the significant differences observed in Face - Hashtag, Face – Call to Action, Phone – URL, and Face - URL suggest meaningful impacts in these areas, while the other comparisons do not provide substantial evidence of a difference. The test results revealed that male participants looked at the face longer than the hashtag, the call to action longer than the face, and the phone number longer than the URL. Test results for male participants are summarized in Table 3.

Table 3 Wilcoxon Signed Ranks Test for Male Participants based on negative ranks. Only face – Call to Action comparison is based on positive ranks.

	Z	Asymp. Sig. (2-tailed)
Face - Title	-0.224	.823
Face - Phone	-0.434	.664
Face - URL	-1.86	.063
Face - Logo	-1.195	.232
Face - Hashtag	-3.432	.001
Face – Call to Action	-3.458	.001
Phone - URL	-2.242	.025

#### 4. CONCLUSION

The findings of this study provide valuable insights into the effectiveness of advertisements by nonprofit organizations, particularly in relation to the call-to-action elements. Utilizing eye-tracking technology, the research revealed significant differences in how male and female participants engage with various elements of an advertisement. Specifically, males tended to focus more on the call-to-action elements, while females were more attentive to human faces.

The analysis provides a comprehensive understanding of gaze behaviors among male and female participants in

response to the final 5500ms of an advertisement. Using the non-parametric Mann-Whitney U test, significant gender differences were identified, with males focusing more on the call-to-action and females concentrating more on the face of the girl in the advertisement. This indicates distinct attentional patterns based on gender, which could be leveraged to tailor advertisements more effectively to different audiences.

Further examination using the Wilcoxon Signed Ranks Test revealed nuanced insights into the attentional focus within each gender group. Female participants demonstrated a significant preference for the face over other elements, including the title, phone, URL, logo, and hashtag, highlighting the importance of human elements in capturing female viewers' attention. However, no significant difference was found for the call-to-action and phone-URL elements among females.

Similarly, male participants showed significant differences in gaze duration for the face compared to the hashtag, the call-to-action compared to the face, and the phone number compared to the URL, indicating specific elements that resonate more with male viewers.

These differences suggest that gender-specific strategies may enhance the impact of advertisements. For instance, incorporating more emotionally resonant images may better capture the attention of female viewers, while ensuring that call-to-action elements are prominently featured and engaging could be more effective for male audiences.

The results also underscore the importance of employing advanced technologies, such as eye-tracking, in marketing research. This approach enables a deeper understanding of consumer behavior, which can inform the design of more effective advertisements. Furthermore, the study highlights the potential of integrating neuromarketing methods to uncover subconscious preferences and improve the predictive success of marketing campaigns.

In conclusion, this research emphasizes the need for tailored marketing strategies that consider gender-based differences in visual attention. By leveraging eye-tracking technology, marketers can create more engaging and effective advertisements that not only capture attention but also drive desired actions, thereby maximizing the impact of their campaigns.

This research has limitations due to the small sample size, which may affect the generalizability of the findings. Additionally, the study focuses on a specific aspect of advertising and may not account for other factors that influence consumer behavior and engagement. It's

important for future research to address these limitations by including a larger and more diverse sample size and considering a broader range of factors that may impact the effectiveness of advertising elements.

## REFERENCES

- [1] P. Kotler, *A'dan Z'ye Pazarlama*, İstanbul: MediaCat, 2005.
- [2] İ. Mucuk, *Pazarlama İlkeleri*, İstanbul: Türkmen Kitabevi, 2017.
- [3] S. Dibb and M. Carrigan, "Social marketing transformed," *European Journal of Marketing*, vol. 47, no. 9, pp. 1376-1398, 2013.
- [4] N. R. Lee and L. Kotler, *Success in Social Marketing 100 Case Studies From Around the Globe*, Routledge, 2022.
- [5] P. Kotler, *Pazarlama 3.0*, İstanbul: Optimist, 2014.
- [6] D. Warrink, "The Marketing Mix in a Marketing 3.0 Context," *International Journal of Innovation and Economic Development*, pp. 7-30, 2018.
- [7] P. Kotler, H. Kartajaya and I. Setiawan, *Pazarlama 4.0*, İstanbul: Optimist, 2022.
- [8] P. Kotler, H. Kartajaya and I. Setiawan, *Pazarlama 5.0*, İstanbul: Kapital Medya Hizmetleri, 2021.
- [9] G. J. Kim, *Human-Computer Interaction Fundamentals and Practice*, CRC Press, 2020.
- [10] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka and J. Van de Weijer, *Eye Tracking: A comprehensive guide to methods and measures*, Oxford, 2011.
- [11] C. L. Kleinke, "Gaze and eye contact: A research review.," *Psychological Bulletin*, pp. 78-100, 1986.
- [12] M. Lindstrom, *Buy.ology*, İstanbul: Optimist, 2018.
- [13] A. C. Martinez-Levy, D. Rossi, G. Cartocci, M. Mancini, G. Di Flumeri, A. Trettel, F. Babiloni and P. Cherubino, "Message framing, non-conscious perception and effectiveness in non-profit advertising. Contribution by neuromarketing research," *International Review on Public and Nonprofit Marketing*, pp. 53-75, 2022.
- [14] U. C. Eze and C. H. Lee, "Consumers' Attitude towards Advertising," *International Journal of Business and Management*, vol. 7, no. 13, pp. 94-108, 2012.
- [15] T. L. Henthorne, M. S. LaTour and R. Natarajaan, "Fear Appeals in Print Advertising: An Analysis of Arousal and Ad Response," *Journal of Advertising*, vol. 22, no. 2, pp. 59-69, 1993.
- [16] A. H. Eagly and S. Chaiken, *The psychology of attitudes*, Harcourt Brace Jovanovich College Publishers, 1993.
- [17] F. F. Brunel and M. R. Nelson, "Explaining Gendered Responses to "Help-Self" and "Help-Others" Charity Ad Appeals: The Mediating Role of World-Views," *Journal of Advertising*, vol. 29, no. 3, pp. 15-28, 2000.
- [18] D. J. Moore, W. D. Harris and H. C. Chen, "Affect Intensity: An Individual Difference Response to Advertising Appeals," *Journal of Consumer Research*, vol. 22, no. 2, pp. 154-164, 1995.
- [19] L. M. Sciulli, C. P. Bebeko and P. Bhagat, "How Emotional Arousal and Attitudes Influence Ad Response: Using Eye Tracking to Gauge Nonprofit Print Advertisement Effectiveness," *Journal of*

- Marketing Management*, vol. 5, no. 1, pp. 1-11, 2017.
- [20] L. Simmonds, S. Bellman, R. Kennedy, M. Nencycz-Thiel and S. Bogomolova, "Moderating effects of prior brand usage on visual attention to video advertising and recall: An eye-tracking investigation," *Journal of Business Research*, no. 111, pp. 241-248, 2020.
- [21] V. Skaramagkas, G. Giannakakis, E. Ktistakis, D. Manousos, . I. Karatzanis, N. S. Tachos, E. Tripoliti, K. Marias, D. I. Fotiadis and M. Tsiknakis, "Review of eye tracking metrics involved in emotional and cognitive processes," *IEEE Reviews in Biomedical Engineering*, no. 16, pp. 260-277, 2021.
- [22] . L.-A. Casado-Aranda, J. Sánchez-Fernández and . J.-Á. Ibáñez-Zapata, "Evaluating Communication Effectiveness Through Eye Tracking: Benefits, State of the Art, and Unresolved Questions," *International Journal of Business Communication*, vol. 1, no. 60, pp. 24-61, 2023.
- [23] X. Yang and I. Krajbich, "A dynamic computational model of gaze and choice in multi-attribute decisions," *Psychological Review*, vol. 1, no. 130, pp. 52-70, 2023.
- [24] M. Wedel, R. Pieters and R. van der Lans, "Modeling Eye Movements During Decision Making: A Review," *Psychometrika*, no. 88, p. 697-729, 2023.
- [25] G. J. Tellis, *Effective Advertising: Understanding When, How and Why Advertising Works*, SAGE Publications, 2004.
- [26] R. P. Bagozzi, M. Gopinath and P. U. Nyer, "The role of emotions in marketing," *Journal of the Academy of Marketing Science*, vol. 27, pp. 184-206, 1999.
- [27] C. D. Burt and K. Strongman, "Use of images in charity advertising: Improving donations and compliance rates," *International Journal of Organisational Behaviour*, vol. 8, pp. 571-580, 2005.
- [28] L. M. Sciulli, P. S. Bhagat and C. P. Bebko, "Eye tracking analysis: engagement levels and donor tendencies using print advertisements with emotional appeals," *Innovative Marketing*, vol. 8, no. 2, pp. 91-98, 2012.
- [29] A. Sharma, B. Grohmann, A. Johnson and O. H. Bodur, "Emotive Appeals in Charitable Advertisements: Investigating Visual Attention and Donation Intentions," in *Proceedings of the European Marketing Academy*, Budapest, 2020.
- [30] C. Bebko, L. M. Sciulli and P. Bhagat, "Using Eye Tracking to Assess the Impact of Advertising Appeals on Donor Behavior," *Journal of Nonprofit & Public Sector Marketing*, vol. 26, no. 4, pp. 354-371, 2014.
- [31] M. Alonso Dos Santos, C. Lobos, . N. Muñoz, D. Romero and R. Sanhueza, "The Influence of Image Valence on the Attention Paid to Charity Advertising," *Journal of Nonprofit & Public Sector Marketing*, vol. 29, no. 3, pp. 346-363, 2017.
- [32] K. Rayner, C. M. Rotello, A. J. Stewart, J. Keir and S. A. Duffy, "Integrating text and pictorial information: Eye movements when looking at print advertisements," *Journal of Experimental Psychology*, vol. 7, no. 3, pp. 219-226, 2001.
- [33] K. Rayner, B. Miller and C. M. Rotello, "Eye Movements When Looking at Print Advertisements: The Goal of the Viewer Matters," *Applied Cognitive Psychology*, vol. 22, no. 5, pp. 697-707, 2008.
- [34] A. Yücel and Z. Göker, "Kızılay'ın Sosyal Yardım İçerikli Afişlerinin Eye-Tracking ile İncelenmesi," *Fırat Üniversitesi İİBF Uluslararası İktisadi ve İdari Bilimler Dergisi*, vol. 4, no. 1, pp. 183-212, 2020.
- [35] Y. E. Gür and A. Yücel, *Sosyal Sorumluluğu Bakışlardan Anlamak*, İstanbul: Kriter, 2023.
- [36] Y. Girişken, *Gerçeği Algıla*, İstanbul: Beta , 2017.
- [37] iMotions, *iMotions Biometric Research Platform 7.2*, Copenhagen, 2017.
- [38] P. Kotler, *Pazarlama 3.0*, İstanbul: Optimist, 2014.
- [39] K. Holmqvist, M. Nyström, R. Andersson, R. Dewhurst, H. Jarodzka and J. Van de Weijer, *Eye Tracking: A comprehensive guide to methods and measures*, Oxford, 2011.

# Detection of Military Aircraft Using YOLO and Transformer-Based Object Detection Models in Complex Environments

*Araştırma Makalesi/Research Article*

 Fatih ŞENGÜL<sup>1</sup>,  Kemal ADEM<sup>2</sup>

<sup>1</sup>Department of Computer Engineering, Sivas University of Science and Technology, Sivas, Türkiye.

<sup>2</sup>Department of Computer Engineering, Sivas Cumhuriyet University, Sivas, Türkiye.

[210102002@sivas.edu.tr](mailto:210102002@sivas.edu.tr), [kemaladem@cumhuriyet.edu.tr](mailto:kemaladem@cumhuriyet.edu.tr)

(Geliş/Received:12.09.2024; Kabul/Accepted:08.01.2025)

DOI: 10.17671/gazibtd.1549034

**Abstract**— Computer vision and deep learning techniques are widely applied in object detection tasks across various domains, including defense technologies. Accurate and efficient detection of military aircraft plays a critical role in strengthening air defense systems and enabling effective strategic decision-making. This study evaluates the performance of YOLOv7, YOLOv8, and RT-DETR models in detecting military aircraft using a dataset consisting of 19.514 images spanning 43 aircraft models. The dataset incorporates images captured from various angles and diverse backgrounds, such as urban, rural, and coastal areas, ensuring realistic testing conditions. However, class imbalance is observed, with certain aircraft models, such as the F14 and F16, being more represented than others, which may affect model generalization. To address these challenges, hyperparameters were optimized, and performance metrics, including mean Average Precision (mAP) and recall, were analyzed. Experimental results show that YOLOv8 achieved 94% mAP and 88.1% recall, YOLOv7 reached 90.2% mAP and 82.7% recall, while RT-DETR demonstrated consistent performance with 92.7% mAP and 90.4% recall. These findings highlight the strengths and limitations of the evaluated models and provide inferences for improving detection systems in defense applications.

**Keywords**— military aircraft detection, YOLOv7, YOLOv8, RT-DETR

## Karmaşık Ortamlarda YOLO ve Transformer Tabanlı Nesne Tespit Modelleri ile Askeri Uçak Tespiti

**Özet**— Bilgisayarla görme ve derin öğrenme teknikleri, savunma teknolojileri de dahil olmak üzere çeşitli alanlardaki nesne algılama görevlerinde yaygın olarak uygulanmaktadır. Savaş uçaklarının doğru ve verimli bir şekilde tespit edilmesi, hava savunma sistemlerinin güçlendirilmesinde ve etkili stratejik karar alma süreçlerinin desteklenmesinde kritik bir rol oynamaktadır. Bu çalışmada, 43 uçak modelini kapsayan 19.514 görüntüden oluşan bir veri kümesi kullanılarak YOLOv7, YOLOv8 ve RT-DETR modellerinin savaş uçaklarını tespit etme performansı değerlendirilmektedir. Veri kümesi, çeşitli açılardan ve kentsel, kırsal ve kıyı alanları gibi farklı arka planlardan çekilen görüntüleri içermekte ve gerçekçi test koşulları sağlamaktadır. Bununla birlikte, F14 ve F16 gibi belirli uçak modellerinin diğerlerine göre daha fazla temsil edildiği ve model genellemesini etkileyebilecek sınıf dengesizliği gözlemlenmiştir. Bu zorlukların üstesinden gelmek için hiperparametreler optimize edilmiş ve ortalama Ortalama Hassasiyet (mAP) ve geri çağırma dahil olmak üzere performans ölçütleri analiz edilmiştir. Deneysel sonuçlar, YOLOv8'in %94 mAP ve %88,1 geri çağırma, YOLOv7'nin %90,2 mAP ve %82,7 geri çağırma değerlerine ulaştığını, RT-DETR'nin ise %92,7 mAP ve %90,4 geri çağırma ile tutarlı bir performans sergilediğini göstermektedir. Bu bulgular, değerlendirilen modellerin güçlü yönlerini ve kısıtlamalarını vurgulamakta ve savunma uygulamalarında tespit sistemlerinin iyileştirilmesi için çıkarımlar sağlamaktadır.

**Anahtar Kelimeler**— savaş uçağı tespiti, YOLOv7, YOLOv8, RT-DETR

## 1. INTRODUCTION

Computer vision, a key branch of computer science, has rapidly advanced over time. Researchers have continually worked to develop more effective and efficient systems to tackle the challenges in this field. Morphological methods have emerged as significant strategies, particularly for addressing core issues in computer vision. The expansion of digital platforms has led to a substantial increase in visual data, driving demand for data processing and information extraction [1]. This surge in data has accelerated research in computer vision, with the ultimate goal of developing algorithms that operate as swiftly and accurately as the human eye [2]. The growing need for automation has also played a critical role in these advancements, enhancing efficiency and safety in high-risk environments through autonomous systems [3][4][5]. Research in computer vision typically falls into three main areas: segmentation, classification, and object detection [6]. These areas encompass more specific tasks, such as semantic segmentation, scene classification, and pixel-based classification. Semantic segmentation, for instance, distinguishes object boundaries within the same category, while pixel-based classification is particularly effective for hyperspectral remote sensing images, though it requires significant processing power [7][8][9]. Object detection, which involves identifying, classifying, and locating objects within an image, is particularly challenging in applications requiring detailed accuracy [10]. Object detection is critical in fields ranging from military operations to healthcare diagnostics [11][12][13][14]. However, the development of effective algorithms often encounters challenges such as low spatial resolution and complex image data. Additionally, reliance on human interpretation can introduce potential errors. Detected objects in images typically include a variety of structures, both man-made and natural, making object detection a complex task. Significant advancements in object detection have been driven by deep learning techniques, which have improved detection under large datasets and complex conditions [15]. The increasing computational power of GPUs has also been crucial in advancing these technologies, representing a critical step towards overcoming the challenges in object detection. As imaging technologies have advanced, the detection of military aircraft has become increasingly critical. Numerous studies have contributed to developing methods for accurately identifying military aircraft.

Early efforts focused on creating models for fighter jet detection using physical prototypes of aircraft such as the P51 Mustang, G1-Fokker, MiG25-F, and Mirage 2000, achieving a recognition accuracy of 91% and a response time of 3 seconds [16]. Building on this foundation, subsequent research introduced novel approaches, such as a 3D model for carrier-based aircraft detection, achieving a detection accuracy of 99.92% in real reconnaissance images [17].

Advancements in remote sensing also enabled methods that used Convolutional Neural Networks (CNNs) for aircraft classification, achieving an accuracy of 98.29% [18]. Further developments included an enhanced YOLOv3-based object detection system, which improved precision to 91.49%, surpassing the original YOLOv3's 85.61% [19]. Real-time fighter jet detection was achieved using the YOLOv4 algorithm, with mAP and fps improvements to 86.92% and 29.62, respectively [20]. Object detection techniques continued to evolve with the development of SCMask R-CNN, which combined object recognition and segmentation, resulting in an AP value of 96.8% [21]. To address the challenge of detecting small aircraft, a Multi-Scale Detection Network (MSDN) was proposed, achieving an F1-score above 96% and an AP value exceeding 90% [22].

Further advances included the DAFF-Net model for detecting fighter jets within remote sensing images, which achieved an mAP value of 83.83% [23]. This progress continued with the development of YOLOv5-Aircraft, integrating enhancements that led to a 3.74% increase in mAP and a 6.93% improvement in speed [24]. The comparative analysis of deep learning-based models for aircraft detection provided valuable insights, with one study demonstrating that the FNDCNNTL model achieved nearly 100% accuracy [25].

Additionally, the application of R-FCN using Google Earth images reached a detection accuracy of 98.01%, outperforming SSD and Faster R-CNN models [26]. Advances in detection methods continued with the development of TransEffiDet, an aircraft detection method based on EfficientDet and Transformer modules, achieving an mAP value of 86.6% [27]. In military vehicle detection, Tiny YOLOv3 and Quantized SSD Mobilenet v2 showed superior performance in edge devices [28]. Further refinement in military aircraft recognition was achieved through the integration of VACR techniques with Back Propagation Neural Networks (BPNN), leading to a training accuracy of 95.33% and a testing accuracy of 87% [29]. Additionally, the lightweight CNN framework CGC-NET demonstrated its effectiveness in remote sensing images, achieving a 91.06% F-score and outperforming other models [30].

Continued innovation was evident in the development of the YOLOv5-Aircraft model, which achieved a 3.74% increase in mAP and a 6.93% speed improvement over previous versions [31]. The optimization of YOLOv5 led to the YOLM model, which reached an mAP score of 88.7% on the FAIR1M dataset, outperforming other base models [32]. Comparative analyses highlighted the effectiveness of Faster R-CNN, which achieved the highest mAP value of 97%, making it suitable for high-precision scenarios [33]. Similarly, the scaled YOLOv4 model achieved 96% accuracy in practical applications using high-resolution Worldview-3 data [34]. The YOLO-extract algorithm, optimized from YOLOv5,

further enhanced detection capabilities, achieving a 95.9% mAP value [35]. The CNTR-YOLO algorithm improved detection accuracy by 3.3% over YOLOv5, reaching a 70.1% average accuracy on the MAR20 dataset [36]. Finally, the GCD-DETR model for UAV detection marked a significant advancement, achieving high accuracy rates of 95.6% and 97.8% on UAV datasets [37]. The enhancement of YOLOv8 and Faster R-CNN further demonstrated the continuous improvement in object detection, with YOLOv8 achieving a general accuracy of 96.7% mAP, surpassing Faster R-CNN in overall performance [38].

In recent years, various studies have focused on the detection of small aerial objects. One such study modified the YOLOv8 model by integrating Multi-Scale Image Fusion (MSIF) and a P2 layer, achieving an Average Precision (AP) of 0.189 in the Drone-vs-Bird Detection Challenge, demonstrating effectiveness in detecting small and fast-moving objects at 45.7 FPS for 640x640 resolutions [39]. Another study introduced a YOLO-based segmented dataset containing 20,925 images, including 12,474 drones and 8,451 birds, to address the challenge of distinguishing drones from birds. The dataset features detailed segmentation and diverse environmental conditions, providing a valuable resource for training deep learning models in UAV detection and classification tasks [40]. Another study utilized the YOLOv4 model to develop a drone detection system, achieving 85% accuracy by classifying military drones under the 'aeroplane' category in the COCO dataset [41]. Another study evaluated three deep learning approaches in the Drone vs. Bird Detection Challenge, with the best model achieving an average precision of 80%, demonstrating robustness against small object sizes, distant targets, and moving cameras [42].

Despite significant advancements in military aircraft detection, challenges remain, particularly in the reliance on satellite images with limited perspectives and the need for extensive computational resources. This study addresses these gaps by evaluating the performance of state-of-the-art object detection models, including YOLOv7, YOLOv8, and RT-DETR, across diverse and complex scenarios. By analyzing the effects of different hyperparameters, this research provides insights into the strengths and limitations of these models, contributing to their potential application in defense technology.

## 2. MATERIAL AND METHODS

This section examines the dataset utilized in this manuscript, explaining the theoretical foundations relevant to the topics covered. The methods applied in this study and the experiments conducted are also detailed.

### 2.1. Dataset

The dataset utilized in this study is centered on the detection of military aircraft and consists of 43 classes of visual data, available as an open-source resource on Kaggle. It includes 19,514 images, covering a broad spectrum of military aircraft types and models. Each class represents a specific type or model of military aircraft, with images taken from various angles and set against diverse backgrounds, enhancing the model's adaptability to real-world conditions. The images capture jets under different seasonal and temporal conditions, with varying weather and background settings, promoting the development of more robust algorithms. This variety ensures that the model is effective across multiple scenarios, from snowy landscapes to tropical islands, thereby broadening its applicability. However, the dataset is not evenly distributed across all classes, as illustrated in Figure 1. Classes such as the F14 and F16 are more heavily represented, with over 1,000 images each, while others, like the F35 and Rafale, have fewer images. This imbalance could cause some classes to be more easily recognized, while others might be underrepresented, posing challenges in developing a balanced detection model.

This dataset is a valuable resource for research on the automatic detection of military aircraft, commonly used in academic studies to evaluate algorithm performance on real-world data. Experiments conducted with images from various backgrounds and angles improve the algorithms' adaptability and effectiveness, which is critical in military and defense applications. Accurate detection of military aircraft can enhance air defense systems, monitor enemy aircraft, and ensure civilian air traffic safety.

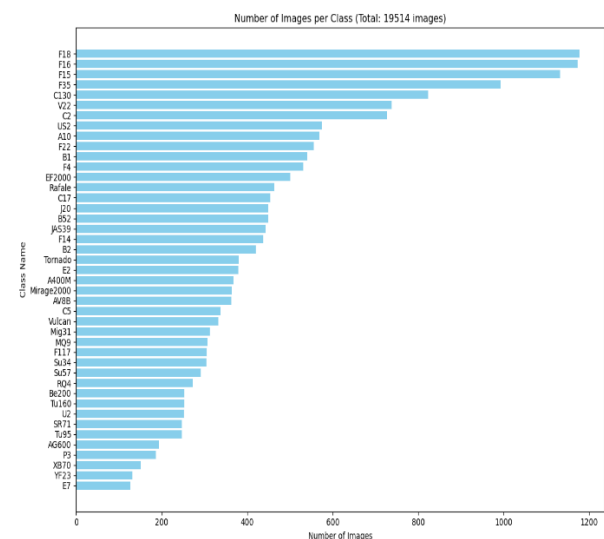


Figure 1. Dataset Distribution

Object detection has become a pivotal research area due to its increasing importance across various fields, aiming to automatically identify and localize specific objects within visual data. This technology underpins critical applications ranging from autonomous vehicles and security surveillance to medical diagnostics and retail analytics. Object detection involves classifying objects within an image and localizing them with bounding boxes, thereby providing information about the object's identity and location [43].

The implementation of object detection is primarily based on deep learning models, which learn features from large volumes of labeled image data. Early models utilized techniques like sliding windows followed by feature extraction, but recent deep learning approaches offer more direct and efficient detection. CNN-based models, for example, provide faster and more accurate results compared to traditional machine learning algorithms [44]. In this study, advanced object detection methods, particularly RT-DETR and various versions of the YOLO algorithm, are employed for their effectiveness in overcoming challenges in object detection and contributing to advancements in the field.

### 2.2.1. Convolution-Based Object Detection Models

Convolution-based object detection is a foundational technique in computer vision systems, enabling the detection, identification, and classification of objects within images. This process is built upon deep learning methodologies, particularly Convolutional Neural Networks (CNNs), which have revolutionized the field of image processing. Object detection algorithms are designed to locate and categorize objects within images by leveraging the layered structure of CNNs and their capacity for learning complex visual features [45]. The initial stage of convolution-based object detection involves feature extraction within convolutional layers. These layers apply filters and activation functions to progressively abstract visual features from raw pixel values, such as edges, textures, and shapes [46]. Typically, activation functions like ReLU enhance the model's learning capability, allowing for more complex function modeling [47]. Once features are extracted, models like the Region Proposal Network (RPN) identify potential object regions, predicting bounding boxes that define the approximate location and size of objects within the image [48].

Subsequently, the extracted regions are resized and classified using Region of Interest (RoI) Pooling, transforming them into fixed-size vectors for further processing. The final stages involve classification and regression layers, which assign class labels and refine the placement of bounding boxes using techniques such as softmax classification and linear regression. This allows the model to accurately predict both the class and location of each object in the image. The success of these techniques depends significantly on well-constructed training datasets and the use of data

augmentation and regularization methods to prevent overfitting. These practices ensure the model's robustness and its ability to make accurate predictions under varying conditions. Convolution-based object detection is widely used across industries like automotive, healthcare, security, and retail, enabling automated and precise task execution without human intervention. This continuous evolution of AI and computer vision technologies leads to increasingly innovative applications across these sectors.

### 2.2.2. YOLO (You Only Look Once)

YOLO (You Only Look Once) is a pioneering deep learning architecture developed for real-time object detection, introduced by Joseph Redmon and colleagues in 2016 [49]. Unlike traditional methods, YOLO performs detection in a single pass through the network, dividing the image into grids and predicting bounding boxes and class probabilities simultaneously. This approach allows YOLO to achieve both speed and accuracy, making it ideal for real-time applications. YOLO's architecture is based on convolutional neural networks (CNNs), comprising multiple convolutional layers, pooling layers, and fully connected layers. These layers are designed to extract features, learn relationships, and make predictions necessary for object detection. By processing the entire image at once, YOLO leverages global information to reduce false positives and enhance accuracy [49]. The evolution of YOLO has seen the development of various versions, each improving upon the last. YOLOv2 and YOLOv3 introduced enhancements in accuracy and the ability to detect objects at multiple scales [50]. This study focuses on YOLOv7 and YOLOv8, the latest advancements in the YOLO series, selected for their improved architectures and performance.

### 2.3.3. YOLOv7 and YOLOv8

YOLOv7 features a deeper and wider CNN structure, utilizing multi-scale feature maps and improved bounding box regression techniques to enhance accuracy. It incorporates advanced data augmentation and custom cross-connection modules, optimizing the model for real-time applications with complex backgrounds. YOLOv8, the most innovative in the series, adopts a multi-layer perceptron architecture that excels in detecting complex geometric structures and textures. This model is particularly effective in low-light and noisy environments, with optimized feature extraction and information flow.

Both YOLOv7 and YOLOv8 are designed for real-time applications, but YOLOv8's more complex architecture offers superior performance in handling intricate tasks. Table 1 details the hyperparameters used for these models, which were carefully tuned to achieve optimal results during training and testing.



Table 1. Hyperparameters and Descriptions for YOLO Models

Parameter	Value	Description
lr0	0,0002	Initial speed for weight updates in the model
lrf	0,001	Learning rate used in the final stages of training
Momentum/Beta1	0,937	Cumulative effect of previous gradient updates
Weight Decay	0,0005	Prevents model overfitting
Box	0,05	Weight of bounding box loss
Cls	0,3	Weight of class loss
Obj	0,7	Weight of object loss
Iou_t	0,20	IoU training threshold
Anchor_t	4.0	Anchor box alignment threshold
Fl_gamma	0,0	Focal loss gamma

In this study, YOLOv7 and YOLOv8 models were employed for object detection and classification, with hyperparameters meticulously adjusted to maximize performance. The selected hyperparameters, as outlined in Table 1, were determined through extensive experimentation to achieve the best possible results under various conditions.

#### 2.2.4. RT-DETR

The RT-DETR (Real-Time Detection Transformer) method introduces significant innovations in object detection by utilizing a transformer-based architecture, particularly effective in complex visual environments [51]. Unlike CNN-based approaches, RT-DETR employs a global prediction approach, simplifying the detection process by predicting objects collectively rather than individually. This method reduces training complexity and improves model efficiency, making it capable of handling overlapping and multi-scale objects [52]. The RT-DETR architecture comprises three main components: a backbone, a transformer, and a feed-forward network (FFN). Typically using ResNet for feature extraction, the transformer processes these features through multiple attention layers, capturing relationships between objects [53][54]. The final predictions for object classification and localization are produced by the FFN, utilizing cross-attention mechanisms to combine object queries with image features, resulting in accurate and scalable detection outcomes [55]. RT-DETR's transformer component represents a key innovation, allowing for global context understanding by linking different parts of the feature map through attention mechanisms. This capability is particularly advantageous in complex scenarios, where overlapping objects are detected with higher accuracy [56]. The model outputs a set of predicted classes and bounding boxes, optimized using bipartite matching and the Hungarian algorithm to align predictions with ground truth efficiently [57]. The performance of RT-DETR has been demonstrated on datasets like MS

COCO, with significant improvements in object detection accuracy, particularly in challenging categories [58]. Modifications such as deformable attention mechanisms have further enhanced its capabilities, particularly in dense and complex scenes [59]. Despite its advantages, the model's training process can be computationally intensive, a challenge addressed in subsequent research focusing on optimization [60].

For this study, RT-DETR was employed with carefully selected hyperparameters to optimize performance during training and testing. The choice of parameters like the optimizer, learning rates, and cost weights played a crucial role in achieving accurate and efficient object detection [55]. Data augmentation techniques such as HSV adjustments and geometric transformations were also applied to enhance the model's robustness [59].

Table 2. Hyperparameters and Descriptions for RT-DETR Models

Parameter	Value	Description
optimizer	AdamW	The optimization algorithm of the model
base learning rate	0,0001	The initial learning rate of the model
learning rate of backbone	0,00001	Learning rate for the backbone network
weight decay	0,0001	Weight decay to prevent overfitting
number of AIFI layers	1	Number of Adaptive Addition and Subtraction Layers
number of RepBlocks	3	Number of repeating blocks
embedding dim	256	The dimension of embedding vectors
feedforward dim	1024	The dimension of the feedforward network
nheads	8	Number of heads in the multi-head attention mechanism
number of feature scales	3	Number of different feature scales
number of decoder layers	6	Number of layers in the decoder
number of queries	300	Maximum number of objects the model can process simultaneously
bbox cost weight	5.0	Weight of the bounding box cost function
GIoU cost weight	2.0	Weight of the Generalized IoU cost function
class loss weight	1.0	Weight of the class loss
bbox loss weight	5.0	Weight of the bounding box loss
GIoU loss weight	2.0	Weight of the Generalized IoU loss

Each parameter was fine-tuned to ensure optimal model performance in real-world conditions. Understanding these adjustments is key to improving future implementations.

### 2.3. Performance Metrics

To evaluate the performance of object detection models, the most commonly used metrics include Precision, Recall, and mean Average Precision (mAP). These metrics are based on fundamental concepts such as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), which are typically organized in a confusion matrix. Precision, measures the ratio of correctly identified positive examples to the total number of examples predicted as positive. High precision indicates that the model produces few false positives, and it is calculated using Equation (1).

$$P = \frac{TP}{TP + FP} \quad (1)$$

Recall, calculates the ratio of correctly identified positive examples to the total number of actual positive examples. High recall suggests that the model successfully detects most of the positive instances, as shown in Equation (2).

$$R = \frac{TP}{TP + FN} \quad (2)$$

mean Average Precision (mAP), evaluates the balance between precision and recall across different classes. It is computed as the average of the precision-recall curve areas for each class, as detailed in Equations (3) and Equations (4).

$$P = \sum_n (R(n) - R(n - 1)) P(n) \quad (3)$$

$$mAP = \frac{1}{N} \sum_{i=1}^N AP_i \quad (4)$$

These metrics are crucial for understanding how well a model performs in real-world scenarios and for comparing different models. Precision and recall often exhibit a trade-off, where improving one may decrease the other. Therefore, balancing these metrics is essential for developing an effective object detection model.

Additionally, the F1 Score is frequently used to balance precision and recall. It is the harmonic mean of precision and recall, as defined in Equation (5).

$$F1 \text{ Score} = 2 * \left( \frac{P * R}{P + R} \right) \quad (5)$$

The Confusion Matrix, as shown in Table 3, provides a detailed breakdown of the model's predictions, illustrating the relationship between actual and predicted classifications. It includes True Positives (TP), False Positives (FP), True Negatives (TN), and False Negatives (FN), offering insight into specific types of errors made by the model.

Table 3. Confusion Matrix

Actual / Predicted	Positive Prediction	Negative Prediction
Actual Positive	True Positive (TP)	False Negative (FN)
Actual Negative	False Positive (FP)	True Negative (TN)

These metrics and the confusion matrix are essential tools for analyzing model performance in-depth and guiding the model development process.

## 3. RESULTS and DISCUSSION

In this study, the performance of the object detection model was evaluated using commonly accepted metrics such as F1 score, precision, recall, and mean average precision (mAP). These metrics, explained in detail in Section 3, were used to analyze the model's effectiveness in accurately detecting and classifying objects.

### 3.1. YOLOv7

YOLOv7 is a prominent object detection model known for its real-time detection capabilities and high accuracy. The model's performance was thoroughly assessed using various metrics and graphs to determine its strengths and areas for improvement. Figure 1 presents the F1 score-confidence curve, where the F1 score, a harmonic mean of precision and recall, is plotted against different confidence levels. The average F1 score across all classes was 0.86 at a confidence level of 0.563, indicating the model's ability to balance precision and recall. The variation in F1 scores among different classes suggests that while the model performs consistently well for certain classes, its performance fluctuates depending on the confidence threshold.

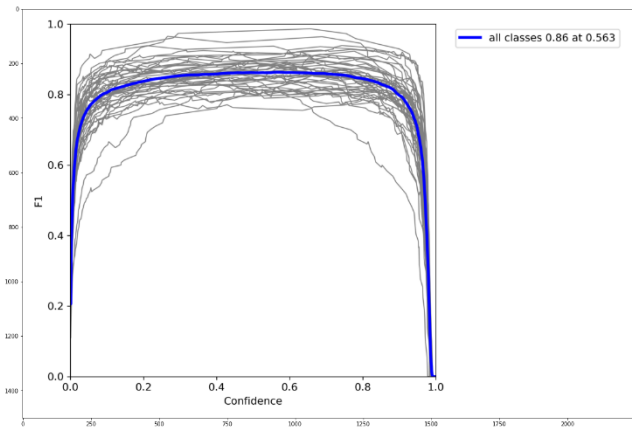


Figure 1. F1 Score-Confidence Curve for YOLOv7 Model

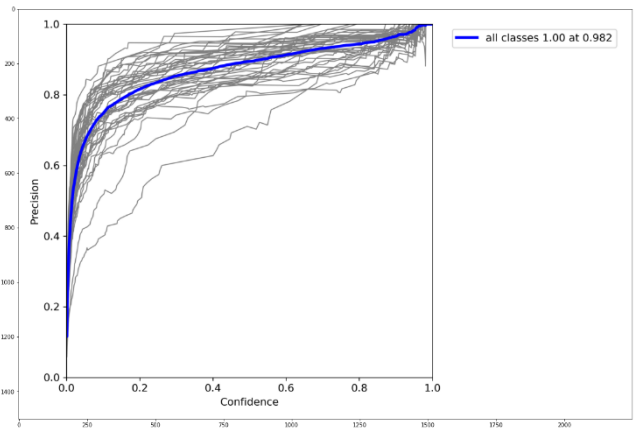


Figure 3. Precision-Confidence Curve for YOLOv7 Model

Figure 2 illustrates the precision-recall curve for the YOLOv7 model, showing the trade-off between precision and recall across different classes. The model achieved a mean average precision (mAP) of 0.902 at an IoU threshold of 0.5, indicating a strong balance between precision and recall in detecting objects with a significant overlap. Figure 3 depicts the precision-confidence curve, highlighting the model's precision across varying confidence levels. The model reached a

precision of 1.00 at a confidence level of 0.982, demonstrating its capability to produce highly accurate detections at near-perfect confidence levels. However, the curve also indicates that precision varies across classes, emphasizing the model's strong performance in certain categories and room for improvement in others.

Figure 4 shows the confusion matrix for YOLOv7, displaying the accuracy of the model's predictions across different classes. The diagonal elements represent correctly classified instances, while off-diagonal elements indicate misclassifications. High accuracy rates in classes such as AG600 and Mig31 reflect the model's reliability, whereas lower accuracy in classes like F117 suggests areas where further model refinement is needed. The experimental results indicate that the YOLOv7 model can effectively detect and classify various aircraft types with high accuracy across different conditions. These findings are supported by the visual examples provided in Figure 1 and Figure 2, showcasing the model's robust detection capabilities in real-world scenarios.

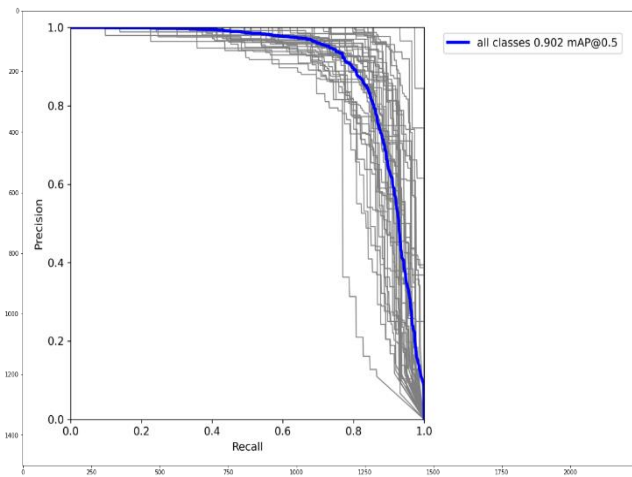


Figure 2. Precision-Recall Curve for YOLOv7 Model

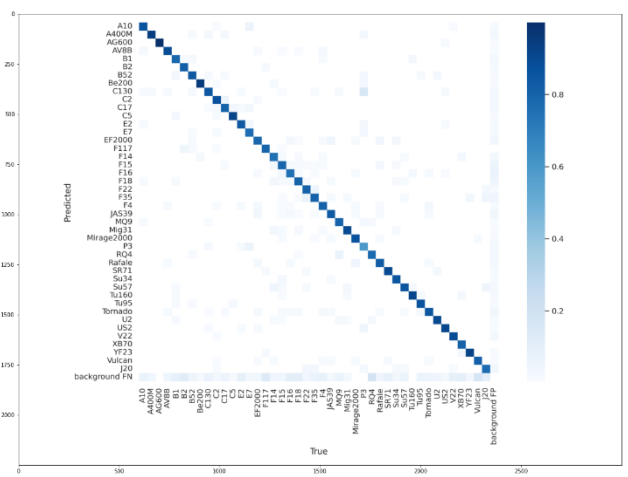


Figure 4. Confusion Matrix for YOLOv7

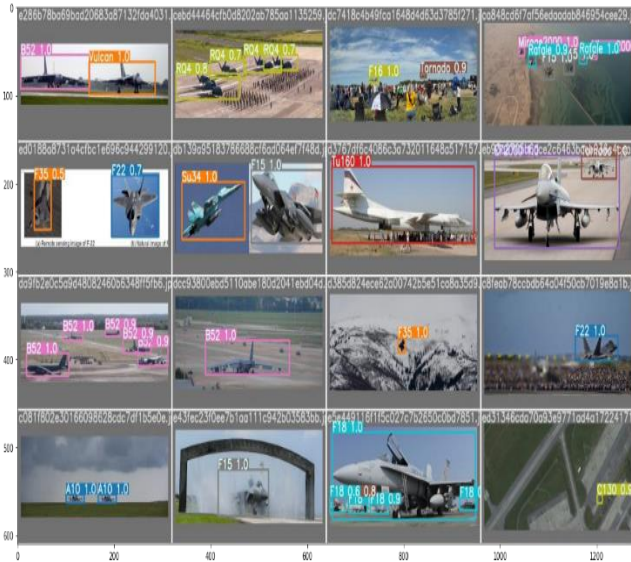


Figure 5. Sample Detection Results Using YOLOv7

The visual results presented in Figure 5 provide a tangible representation of YOLOv7's detection capabilities, aligning well with the quantitative metrics discussed earlier. The model successfully identifies and localizes various aircraft, including those with overlapping features or challenging backgrounds, reaffirming its effectiveness in real-world applications.

### 3.2. YOLOv8

The YOLOv8 model's performance was rigorously analyzed through a variety of metrics and visual representations. This evaluation emphasized its strengths in accuracy and reliability, while also identifying specific areas where further enhancements could be made. Figure 6 presents the F1 score-confidence curve, where the relationship between F1 scores and confidence levels is illustrated. The average F1 score across all classes reached 0.90 at a confidence level of 0.695, indicating the model's strong performance at this level. The variability in F1 scores among different classes suggests that while the model performs well for certain classes, its performance varies depending on the confidence threshold. Figure 7 shows the precision-recall curve, which highlights the model's ability to balance precision and recall. The model achieved a mean average precision (mAP) of 0.940 at an IoU threshold of 0.5, reflecting its high accuracy and reliability across different classes.

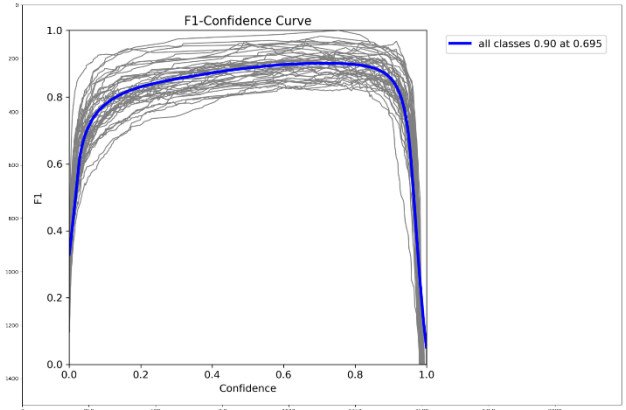


Figure 6. F1 Score-Confidence Curve for YOLOv8 Model

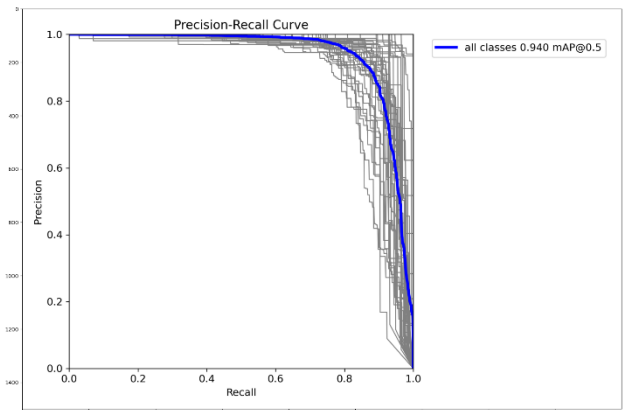


Figure 7. Precision-Recall Curve for YOLOv8 Model

Figure 8 depicts the precision-confidence curve, demonstrating that the model reached a precision of 1.00 at a confidence level of 1.00, indicating near-perfect accuracy at high confidence levels. This suggests that the model is highly reliable in making accurate predictions when it operates at maximum confidence.

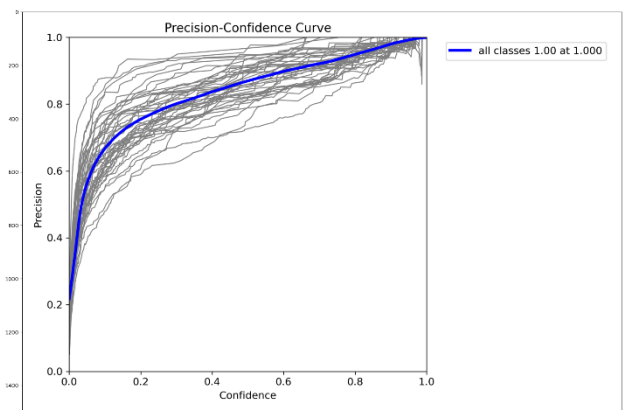


Figure 8. Precision-Confidence Curve for YOLOv8 Model

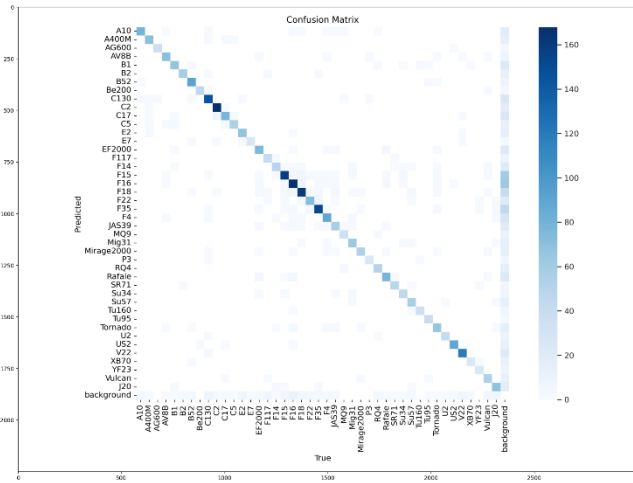


Figure 9. Confusion Matrix for YOLOv8

Figure 9 displays the confusion matrix, which illustrates the accuracy of the model's predictions across various classes. The matrix highlights the model's strong performance in distinguishing between different classes, with high accuracy rates for many classes. However, it also reveals areas where the model could benefit from further refinement.

The experimental results confirm that the YOLOv8 model is highly effective in detecting and classifying various types of aircraft with remarkable accuracy, even under diverse and challenging environmental conditions. This effectiveness is not only evident in the quantitative metrics, such as the high mean Average Precision (mAP) and precision-recall balance, but also in the qualitative assessment of the model's detection capabilities. Figure 6 and Figure 7 illustrate the F1 score-confidence and precision-recall curves, respectively, highlighting the model's ability to maintain strong performance across varying confidence thresholds and object scales. Furthermore, Figure 10 presents visual examples of the YOLOv8 model's successful detections across different scenarios, showcasing its robustness and reliability in real-world applications. These examples underscore the model's proficiency in accurately identifying and localizing aircraft, even in complex scenes with varied backgrounds and lighting conditions.



Figure 10. Sample Detection Results Using YOLOv8

### 3.3. RT-DETR

The RT-DETR model, known for its effective use of attention mechanisms and real-time detection capabilities, was subjected to an extensive performance analysis. The study utilized a range of metrics and visual tools to assess its accuracy across various confidence levels and classes, highlighting both its robust performance and areas needing refinement.

Figure 11 illustrates the F1 score-confidence curve for the RT-DETR model, showing the relationship between F1 scores and confidence levels. The model achieved a high F1 score of 0.93 at a confidence level of 0.637, indicating a strong balance between precision and recall at this confidence level.

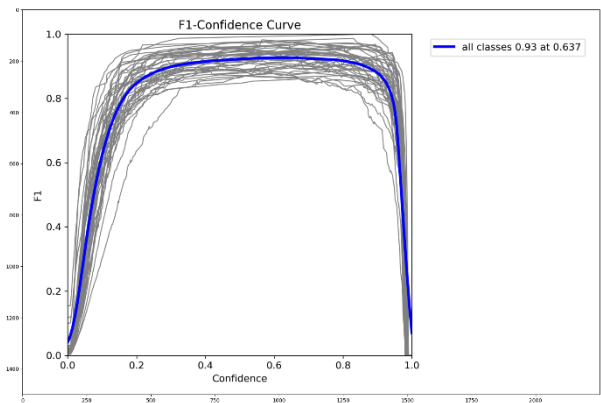


Figure 11. F1 score-Confidence curve for the RT-DETR model

Figure 12 presents the precision-recall curve for the RT-DETR model. The model attained a mean Average Precision (mAP) of 0.927 at an IoU threshold of 0.5, demonstrating its reliable performance in detecting and correctly identifying most objects.

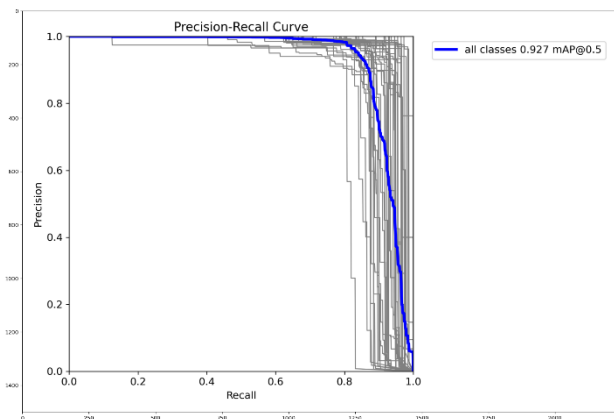


Figure 12. Precision-Recall Curve for RT-DETR Model

Figure 14 shows the precision-confidence curve, where the model achieved a precision of 1.00 at a confidence level of 1.000, indicating that at this level, all predicted objects were correctly classified, highlighting the model's high reliability at maximum confidence.

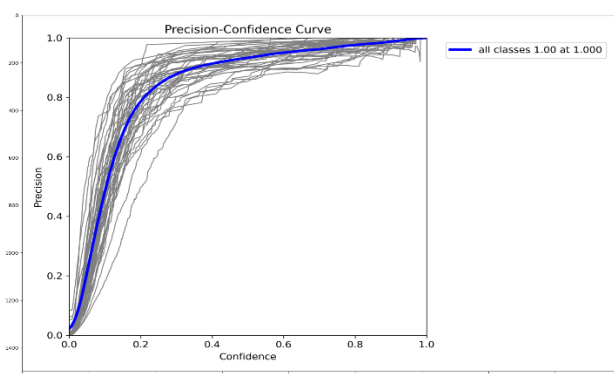


Figure 13. Precision-Confidence Curve for RT-DETR Model

Figure 14 displays the confusion matrix for the RT-DETR model, showing the model's accuracy in distinguishing between various classes. The matrix's diagonal shows high accuracy for several aircraft classes, but also highlights areas where the model struggled, particularly with visually similar aircraft.

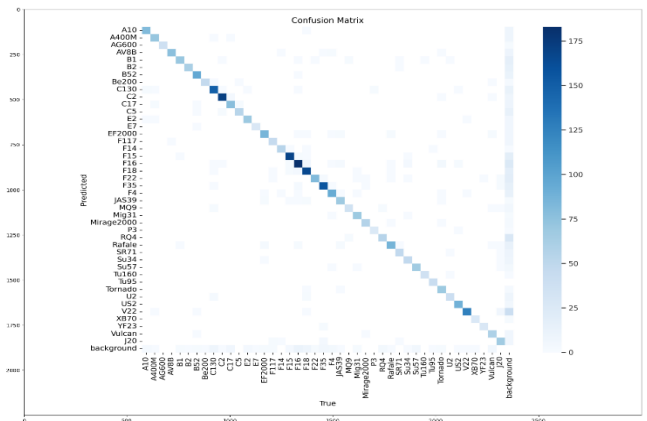


Figure 14. Confusion Matrix for RT-DETR

In addition to these performance metrics, Figure 16 provides visual examples of the RT-DETR model's successful detections across different scenarios, further demonstrating its effectiveness in real-world applications.

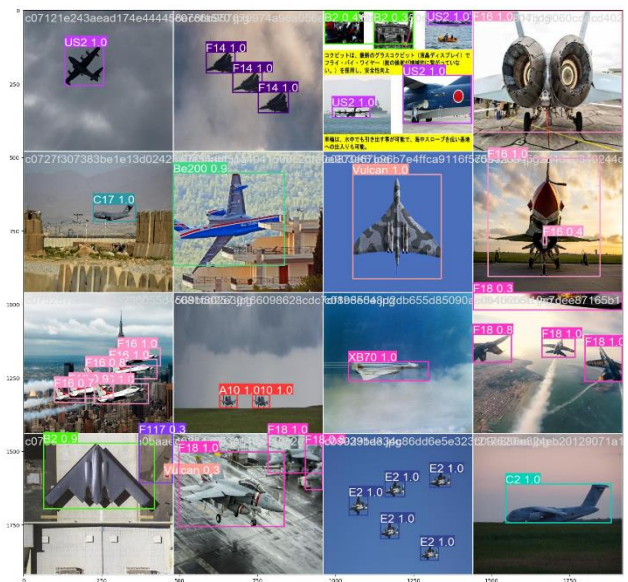


Figure 15. Sample Detection Results Using RT-DETR

The RT-DETR model, despite its architectural differences from traditional CNN-based models like YOLO, demonstrated commendable performance in detecting and classifying various types of aircraft across diverse environmental conditions. Figure 11 and Figure 12 highlight the model's robust F1 score and precision-recall performance, respectively, while Figure 13 emphasizes the model's high precision at maximum confidence levels. The model's ability to accurately distinguish between various classes is further illustrated in Figure 14 through the confusion matrix. Finally, Figure 15 showcases the model's strong detection capabilities with visual examples of successful aircraft detections in real-world scenarios.

YOLOv7 demonstrated strong performance with a mAP@.5 score of 0.902, achieving precision values above 0.90 for classes like A400M, AG600, and Be200, alongside high recall rates in these categories. However, the model exhibited lower recall for certain classes, such as F117, indicating areas for potential improvement. YOLOv8 surpassed YOLOv7 in several metrics, achieving a mAP@.5 of 0.94 and maintaining precision values above 0.90 for most classes, with particularly high performance for classes like Tu160 and Tu95. The model also excelled in recall for classes such as C2 and US2, likely due to algorithmic optimizations and potentially more extensive training data.

The RT-DETR model outperformed the previous two, with a mAP@.5 of 0.93, and exhibited high precision and recall across nearly all classes. Notably, it achieved excellent results in EF2000, F35, and Rafale classes, as well as in P3 and E7, demonstrating its effectiveness across a broad range of classes.

As shown in Table 4. Model Performance Comparison, the RT-DETR model consistently delivered higher precision (0.952) and recall (0.904) compared to YOLOv8 (precision: 0.924, recall: 0.881) and YOLOv7 (precision: 0.907, recall: 0.827). Although YOLOv8 achieved the highest mAP@50 value at 0.94, RT-DETR excelled in overall precision and recall, indicating its superior performance in object detection tasks across a wide dataset. YOLOv7, while trailing behind, still produced effective results in specific classes.

Table 4. Model Performance Comparison

Model	Instances	P	R	mAP@50	mAP
YOLOv7	3578	0.907	0.827	0.902	0.829
YOLOv8	3578	0.924	0.881	0.940	0.877
RT-DETR	3578	0.952	0.904	0.927	0.879

This comparison highlights the RT-DETR model's superior accuracy and consistency in object detection tasks, particularly when high precision and recall are critical.

#### 4. CONCLUSION AND RECOMMENDATIONS

This study has explored the potential of deep learning-based models for the automatic detection of military aircraft, a critical task in modern warfare and strategic surveillance operations. The study utilized an extensive dataset comprising 19,514 images across 43 different military aircraft classes. The primary objective was to accurately classify and detect these classes using YOLOv7, YOLOv8, and RT-DETR models. The evaluation of each model was conducted using metrics

such as Precision, Recall, and mean Average Precision at IoU threshold 0.5 (mAP@.5).

Performance analyses revealed varying results across models, highlighting each model's strengths and weaknesses. The YOLOv7 model demonstrated impressive overall performance but struggled with lower-than-expected Recall rates in classes like F117. The YOLOv8 model built upon the performance of YOLOv7, achieving higher overall mAP values and displaying superior Precision in most classes, particularly in the Tu160 and Tu95 classes. Meanwhile, the RT-DETR model provided more consistent and superior results across Precision and Recall metrics, proving highly reliable across almost all classes. However, all three models exhibited a need for improved performance in certain classes. Low recall rates in some classes suggest that the models may not adequately recognize objects within these classes, possibly due to insufficient training data for those categories. Additionally, high false positive rates in some cases indicate limitations in the models' generalization capabilities or imbalances in the dataset. This highlights the potential benefit of augmenting the dataset with additional samples from underperforming classes. To enhance model performance, several strategies are recommended. First, additional data should be collected for classes with lower performance, and efforts should be made to balance the dataset. Employing data augmentation techniques can also help improve model robustness. This study has demonstrated the significant potential of deep learning models for military aircraft detection. By implementing these recommendations, it is expected that model performance can be further enhanced, enabling broader and more effective applications.

Future studies are intended to enhance model performance by addressing class imbalance and improving generalization capabilities. Specifically, optimizing model architectures, increasing data diversity, and employing advanced data augmentation techniques could lead to significant performance improvements. Recent data augmentation methods, such as MixUp, CutMix, Mosaic, and Copy-Paste, can be utilized to address class imbalance and improve robustness. Moreover, it is essential to evaluate models in terms of speed and efficiency for real-time applications. Such advancements are expected to provide more reliable and effective solutions for the automatic detection of military aircraft, contributing significantly to both military and civilian applications.

## REFERENCES

- [1] K. Bayouhd, R. Knani, F. Hamdaoui, A. Mtibaa, "A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets", *The Visual Computer*, 38(8), 2939-2970, 2022.
- [2] A. A. Khan, A. A. Laghari, S. A. Awan, "Machine learning in computer vision: a review", *EAI Endorsed Transactions on Scalable Information Systems*, 8(32), 2021.
- [3] J. Zhao, R. Masood, S. Seneviratne, "A review of computer vision methods in network security", *IEEE Communications Surveys & Tutorials*, 23(3), 1838-1878, 2021.
- [4] E. Dilek, M. Dener, "Computer vision applications in intelligent transportation systems: a survey", *Sensors*, 23(6), 2938, 2023.
- [5] R. Szeliski, *Computer Vision: Algorithms and Applications*, Springer Nature, 2022.
- [6] S. Minaee, Y. Boykov, F. Porikli, A. Plaza, N. Kehtarnavaz, D. Terzopoulos, "Image segmentation using deep learning: A survey", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(7), 3523-3542, 2022.
- [7] W. Chen, Y. Li, Z. Tian, F. Zhang, "2D and 3D object detection algorithms from images: A Survey", *Array*, 100305, 2023.
- [8] X. Zhuang, D. Li, Y. Wang, K. Li, "Military target detection method based on EfficientDet and Generative Adversarial Network", *Engineering Applications of Artificial Intelligence*, 132, 107896, 2024.
- [9] S. Khalid, H. M. Oqaibi, M. Aqib, Y. Hafeez, "Small pests detection in field crops using deep learning object detection", *Sustainability*, 15(8), 6815, 2023.
- [10] M. Abdel-Aty, Y. Wu, O. Zheng, J. Yuan, "Using closed-circuit television cameras to analyze traffic safety at intersections based on vehicle key points detection", *Accident Analysis & Prevention*, 176, 106794, 2022.
- [11] J. Liu, Y. Jin, "A comprehensive survey of robust deep learning in computer vision", *Journal of Automation and Intelligence*, 2023.
- [12] K. Roopa, T. V. Rama Murthy, P. C. Prasanna Raj, "Neural network classifier for fighter aircraft model recognition", *Journal of Intelligent Systems*, 27(3), 447-463, 2018.
- [13] H. Zhu, H. Lung, N. Lin, "Carrier-based aircraft detection on flight deck of aircraft carrier with simulated 3-D model by deep neural network", *3rd International Conference on Computer Science and Software Engineering*, 96-101, May 2020.
- [14] Q. Liu, X. Xiang, Y. Wang, Z. Luo, F. Fang, "Aircraft detection in remote sensing image based on corner clustering and deep learning", *Engineering Applications of Artificial Intelligence*, 87, 103333, 2020.
- [15] W. Ma, H. Chen, Y. Zhang, "An improved YOLOv3 model for aircraft detection in remote sensing images", *IEEE Access*, 8, 120129-120138, 2020.
- [16] Y. Yang, G. Xie, Y. Qu, "Real-time detection of aircraft objects in remote sensing images based on improved YOLOv4", *2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC)*, 1156-1164, March 2021.
- [17] Q. Wu, D. Feng, C. Cao, X. Zeng, Z. Feng, J. Wu, Z. Huang, "Improved mask R-CNN for aircraft detection in remote sensing images", *Sensors*, 21(8), 2618, 2021.
- [18] L. Zhou, H. Yan, Y. Shan, C. Zheng, Y. Liu, X. Zuo, B. Qiao, "Aircraft detection for remote sensing images based on deep convolutional neural networks", *Journal of Electrical and Computer Engineering*, 2021(1), 4685644, 2021.
- [19] M. Liu, Q. Hu, C. Wang, T. Tian, W. Chen, "Daff-Net: Dual attention feature fusion network for aircraft detection in remote sensing images", *2021 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*, 4196-4199, July 2021.
- [20] L. Zhou, L. Zhang, N. Konz, "Computer vision techniques in manufacturing", *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 53(1), 105-117, 2022.
- [21] E. Kiyak, G. Unal, "Small aircraft detection using deep learning", *Aircraft Engineering and Aerospace Technology*, 93(4), 671-681, 2021.
- [22] H. M. A. Mohammed, M. Polat, A. A. Tahlil, İ. Y. Özbek, "Multi-scale aircraft detection from satellite images", *Erzincan University Journal of Science and Technology*, 14(1), 322-330, 2021.
- [23] Y. Wang, T. Wang, X. Zhou, W. Cai, R. Liu, M. Huang, et al., "TransEffiDet: aircraft detection and classification in aerial images based on EfficientDet and transformer", *Computational Intelligence and Neuroscience*, 2262549, 2022.
- [24] P. Gupta, B. Pareek, G. Singal, D. V. Rao, "Edge device based military vehicle detection and classification from UAV", *Multimedia Tools and Applications*, 81(14), 19813-19834, 2022.
- [25] A. D. W. Sumari, D. E. Adinandira, A. R. Syulistyo, S. Lovrencic, "Intelligent Military Aircraft Recognition and Identification to Support Military Personnel on the Air Observation Operation", *International Journal on Advanced Science, Engineering, and Information Technology (IJASEIT)*, 6(Accepted for Publication), 2022.
- [26] T. Wang, X. Zeng, C. Cao, W. Li, Z. Feng, J. Wu, et al., "CGC-NET: Aircraft Detection in Remote Sensing Images Based on Lightweight Convolutional Neural Network", *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 15, 2805-2815, 2022.
- [27] S. Lou, J. Yu, Y. Xi, X. Liao, "Aircraft target detection in remote sensing images based on improved YOLOv5", *IEEE Access*, 10, 5184-5192, 2022.
- [28] W. Liu, J. Tian, T. Tian, "YOLM: A remote sensing aircraft detection model", *IGARSS 2022-2022 IEEE International Geoscience and Remote Sensing Symposium*, 1708-1711, July 2022.
- [29] B. Azam, M. J. Khan, F. A. Bhatti, A. R. M. Maud, S. F. Hussain, A. J. Hashmi, K. Khurshid, "Aircraft detection in satellite imagery using deep learning-based object detectors", *Microprocessors and Microsystems*, 94, 104630, 2022.
- [30] P. Benjamin, B. Benjamin, G. Dimitri, S. Gérard, E. Eric, "Oriented aircraft object detector using Scaled YOLOv4 on very high resolution satellite and synthetic datasets", *2023 Joint*



- Urban Remote Sensing Event (JURSE), 1–4, May 2023.
- [31] Z. Liu, Y. Gao, Q. Du, M. Chen, W. Lv, “YOLO-extract: Improved YOLOv5 for aircraft object detection in remote sensing images”, *IEEE Access*, 11, 1742–1751, 2023.
- [32] F. Zhou, H. Deng, Q. Xu, X. Lan, “CNTR-YOLO: Improved YOLOv5 Based on ConvNext and Transformer for Aircraft Detection in Remote Sensing Images”, *Electronics*, 12(12), 2671, 2023.
- [33] M. Zhu, E. Kong, “Multi-Scale Fusion Uncrewed Aerial Vehicle Detection Based on RT-DETR”, *Electronics*, 13(8), 1489, 2024.
- [34] A. Kumar, S. Singh, “AIR-SCAN: Aircraft Identification and Recognition using Deep Learning Scanning”, 2024 11th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), 1–6, March 2024.
- [35] K. He, X. Zhang, S. Ren, J. Sun, “Deep residual learning for image recognition”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778, 2016.
- [36] R. Girshick, J. Donahue, T. Darrell, J. Malik, “Rich feature hierarchies for accurate object detection and semantic segmentation”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 580–587, 2014.
- [37] J. Redmon, S. Divvala, R. Girshick, A. Farhadi, “You only look once: Unified, real-time object detection”, *Proceedings of the IEEE conference on computer vision and pattern recognition*, 779–788, 2016.
- [38] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., “Attention is all you need”, *Advances in neural information processing systems*, 5998–6008, 2017.
- [39] J. H. Kim, N. Kim, C. S. Won, “High-speed drone detection based on yolo-v8”, *ICASSP 2023–2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 1–2, June 2023.
- [40] S. K. Shandilya, A. Srivastav, K. Yemets, A. Datta, A. K. Nagar, “YOLO-based segmented dataset for drone vs. bird detection for deep and machine learning algorithms”, *Data in Brief*, 50, 109355, 2023.
- [41] S. Patil, S. M. Jaybhaye, M. M. Khalifa, S. Kharche, A. Khatib, A. Kshirsagar, “Drone detection using YOLO”, *AIP Conference Proceedings*, 2938(1), December 2023.
- [42] A. Coluccia, A. Fascista, A. Schumann, L. Sommer, A. Dimou, D. Zarpalas, et al., “Drone vs. bird detection: Deep learning algorithms and results from a grand challenge”, *Sensors*, 21(8), 2824, 2021.
- [43] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, “End-to-end object detection with transformers”, *European Conference on Computer Vision (ECCV)*, 213–229, 2020.
- [44] Z. Sun, S. Cao, Y. Yang, K. M. Kitani, “Rethinking transformer-based set prediction for object detection”, *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3611–3620, 2021.
- [45] R. u, D. Wunsch, “Survey of clustering algorithms”, *IEEE Transactions on Neural Networks*, 16(3), 645–678, 2005.
- [46] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, et al., “Microsoft COCO: Common objects in context”, *Computer Vision – ECCV 2014*, 740–755, 2014.
- [47] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection”, *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’05)*, 886–893, June 2005.
- [48] S. Ren, K. He, R. Girshick, J. Sun, “Faster R-CNN: Towards real-time object detection with region proposal networks”, *Advances in Neural Information Processing Systems*, 28, 2015.
- [49] J. Dai, H. Qi, Y. Xiong, Y. Li, G. Zhang, H. Hu, Y. Wei, “Deformable convolutional networks”, *Proceedings of the IEEE International Conference on Computer Vision*, 764–773, 2017.
- [50] S. J. Russell, P. Norvig, *Artificial Intelligence: A Modern Approach*, Pearson, 2016.
- [51] C. M. Bishop, N. M. Nasrabadi, *Pattern Recognition and Machine Learning*, Vol. 4(4), Springer, New York, 2006.
- [52] T. Hastie, R. Tibshirani, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 2009.
- [53] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, 60, 91–110, 2004.
- [54] Y. LeCun, Y. Bengio, G. Hinton, “Deep learning”, *Nature*, 521(7553), 436–444, 2015.
- [55] C. Cortes, V. Vapnik, “Support-vector networks”, *Machine Learning*, 20(3), 273–297, 1995.
- [56] A. Krizhevsky, I. Sutskever, G. E. Hinton, “ImageNet classification with deep convolutional neural networks”, *Communications of the ACM*, 60(6), 84–90, 2017.
- [57] K. Simonyan, A. Zisserman, “Very deep convolutional networks for large-scale image recognition”, *arXiv preprint arXiv:1409.1556*, 2014.
- [58] J. Redmon, A. Farhadi, “YOLOv3: An incremental improvement”, *arXiv preprint arXiv:1804.02767*, 2018.
- [59] X. Zhu, W. Su, L. Lu, B. Li, X. Wang, J. Dai, “Deformable DETR: Deformable transformers for end-to-end object detection”, *arXiv preprint arXiv:2010.04159*, 2020.